

*CENTRO
PER LA STORIA DELL'UNIVERSITÀ DI PADOVA*

FRONTIERE DELLA CONOSCENZA

**Big Data nelle scienze fisiche, sociali,
umanistiche e della vita**

a cura di

**FILIBERTO AGOSTINI, PIERDANIELE GIARETTA,
GIORGIO MORO, GIOVANNI SILVANO**

FrancoAngeli
OPEN  ACCESS

CONTRIBUTI
ALLA STORIA DELL'UNIVERSITÀ DI PADOVA
Nuova serie

59



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

1222 • 2022
800
A N N I



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Comitato scientifico

Antonella Barzani, Anna Bettoni, Giovanni Luigi Fontana, Marta Nezzo,
Giulio Peruzzi, Chiara Maria Valsecchi, Fabio Zampieri

*Il comitato assicura attraverso un percorso di double blind peer review la
validità scientifica dei volumi pubblicati*

*CENTRO
PER LA STORIA DELL'UNIVERSITÀ DI PADOVA*

FRONTIERE DELLA CONOSCENZA

**Big Data nelle scienze fisiche, sociali,
umanistiche e della vita**

a cura di

**FILIBERTO AGOSTINI, PIERDANIELE GIARETTA,
GIORGIO MORO, GIOVANNI SILVANO**

FrancoAngeli
OPEN  ACCESS

Il volume è stato pubblicato con il contributo del Centro per la Storia dell'Università – Università degli Studi di Padova.

Copyright © 2021 by FrancoAngeli s.r.l., Milano, Italy.

L'opera, comprese tutte le sue parti, è tutelata dalla legge sul diritto d'autore ed è pubblicata in versione digitale con licenza *Creative Commons Attribuzione-Non Commerciale-Non opere derivate 4.0 Internazionale* (CC-BY-NC-ND 4.0)

L'Utente nel momento in cui effettua il download dell'opera accetta tutte le condizioni della licenza d'uso dell'opera previste e comunicate sul sito
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.it>

Indice

Introduzione, di <i>Filiberto Agostini, Pierdaniele Giaretta, Giorgio Moro, Giovanni Silvano</i>	pag.	7
Natura e ruolo dei dati in Mach e Carnap, di <i>Fabio Grigenti, Pierdaniele Giaretta</i>	»	11
Big Data. Contesto Tecnologico, di <i>Andrea Pietracaprina, Gippino Pucci</i>	»	36
Il ruolo della statistica nell'era dei Big Data, di <i>Bruno Scarpa, Alessandra Salvan</i>	»	50
Le sfide della biologia dei Big Data, di <i>Sabina Leonelli</i>	»	68
Predizione e comprensione dei fenomeni. Fisica e Machine Learning a confronto, di <i>Marco Zanetti</i>	»	77
Quanta informazione può elaborare il cervello di un uomo? E quello di un'ape?, di <i>Aram Megighian</i>	»	94
Che cos'è un «dato» storico? Una riflessione di metodo fra Big Data, approcci quantitativi e scale di analisi, di <i>Andrea Caracausi</i>	»	108
Bo2022: un database online di studenti e docenti dell'Università di Padova (1222-secolo XX), di <i>Pierluigi Terenzi</i>	»	128
I laureati in Scienze politiche dell'Università di Padova (1926-2011). Lettura e interpretazione del Database, di <i>Filiberto Agostini</i>	»	143
Indice dei nomi	»	183

Introduzione

di *Filiberto Agostini, Pierdaniele Giaretta, Giorgio Moro, Giovanni Silvano*

La crescita accelerata negli ultimi anni della cosiddetta ICT (*Information and Communication Technology*) ha modificato profondamente la nostra organizzazione sociale, fino a influenzare le modalità di produzione culturale e i processi di creazione delle conoscenze sia in ambito scientifico che umanistico. Un ruolo determinante è svolto dal mondo convenzionalmente denominato come Big Data: la disponibilità di estese banche dati concernenti vari aspetti delle attività umane, dalle informazioni a carattere personale (ad esempio dati medico-clinici o dati relativi ai consumi-acquisti) a informazioni climatico-ambientali, fino ai risultati di avanzati esperimenti scientifici. Tutto ciò è stato reso possibile dall'avanzamento tecnologico dei sistemi digitali per la memorizzazione dei dati e dalla possibilità di trasmettere e quindi di acquisire efficacemente le informazioni. Parallelamente, e si potrebbe dire necessariamente, per fare uso di una tale quantità di informazioni sono enormemente cresciute le capacità di elaborazione dei dati grazie ai progressi dell'informatica e della *Computer Science*, con particolare riferimento all'insieme delle metodologie che costituiscono la cosiddetta Intelligenza Artificiale.

C'è un aspetto da considerare: l'avvento dei Big Data non evita alcuni problemi che da sempre si sono presentati nella raccolta e nella rappresentazione dei dati. Anche se la capacità di raccogliarli e rappresentarli è enormemente aumentata, può ancora succedere che i dati raccolti abbiano un impatto negativo per la loro manipolazione, per l'influenza di qualche distorsione cognitiva (cognitive bias) o di circostanze contingenti delle quali non si è tenuto conto. Se, ad esempio, i dati relativi a un problema patologico riguardano soprattutto malati con certe caratteristiche, la sovrarappresentazione di questi malati può condizionare negativamente la ricerca di soluzioni valide anche per i malati sottorappresentati.

Può essere difficile valutare l'adeguatezza della raccolta e rappresentazione dei dati, ma è ovvio che si tratta di un obiettivo al quale non si può rinunciare. E un tale proposito non è rilevante solo dal punto di vista epistemico:

una rappresentazione adeguata del fatto che, in posizioni di alta responsabilità, ci siano molti più uomini che donne mette in evidenza l'esistenza di un problema da affrontare sul piano educativo e sociale.

In questo volume molti contributi sono, però, rilevanti per l'impatto che i Big Data possono avere sulla nostra concezione della conoscenza, e in particolare riguardo al modo di intendere la scoperta scientifica. Da un punto di vista molto generale – se facciamo riferimento alla fondazione della scienza moderna da parte di Galileo Galilei e Isaac Newton – due sono i modi nei quali si può intendere la scoperta scientifica. Può essere innanzitutto un fatto nuovo – precedentemente non noto – come, ad esempio, la velocità di caduta di un grave, che è stato predetto sulla base di una teoria, o di una ipotesi teorica, e verificato con successo mediante opportuni esperimenti. Tuttavia, può essere intesa come scoperta anche una teoria, o una ipotesi teorica che permetta di spiegare fatti già noti, relativi, ad esempio, ai moti terrestri e celesti, che prima si spiegavano in modo meno soddisfacente o in modi separati. Generalmente le scoperte di nuove teorie esplicative permettono anche predizioni riuscite di fatti nuovi, quindi scoperte nel primo senso. Peter V. Coveney, Edward R. Dougherty, Roger R. Highfield (2016) (*Big data need big theory too*. Phil. Trans. R. Soc. A 374: 20160153. <http://dx.doi.org/10.1098/rsta.2016.0153>) considerano la capacità di fare predizioni sulla base di teorie una caratteristica specifica della scienza nata con Galileo e Newton e vedono nell'avvento dei Big Data il ritorno a un tipo di ricerca prebaconiana e pregalileiana, che non è guidata né da disegni sperimentali, né da ipotesi teoriche. Senza condividere del tutto l'analisi di Coveney, Dougherty, Highfield e, in particolare, la loro contestazione della piena scientificità di alcune discipline, si può prendere spunto dalle loro discutibili considerazioni per formulare un problema su cui vale la pena di riflettere. Ci si può chiedere se con l'avvento dei Big Data sia diventato possibile scoprire senza bisogno di fare riferimento a teorie.

L'esperienza umana si è continuamente trovata di fronte a fatti nuovi, non prima osservati, e non c'è ragione di escludere che sarà così anche in futuro. Sembra inoltre naturale pensare che più sono i dati che si raccolgono, più siano anche i dati di tipo nuovo che si possono incontrare. Ma ci sono anche dati molto rari che, soprattutto, richiedono di essere prima concepiti per poter essere poi scoperti. Ad esempio, la capacità di generare ed esaminare una grande quantità di dati ha certamente favorito l'individuazione sperimentale del bosone di Higgs, ma si fa fatica a pensare che questo bosone avrebbe potuto essere sperimentalmente rilevato indipendentemente dalla predizione teorica della sua esistenza.

Considerazioni generali come quelle proposte devono essere approfondite, ed eventualmente corrette, tenendo conto del fatto che il mondo dei Big Data sta avendo una forte ricaduta su tutte le discipline culturali, dal mondo scientifico in senso stretto agli ambiti umanistici e sociali. Dato il rilevante impatto, sia in ambiti scientifici in senso stretto sia in quelli sociali e umanistici, è

indispensabile un approccio multidisciplinare, come prospettato nel presente volume: l'obiettivo è quello di riconoscere il ruolo dei Big Data nella nostra cultura e i loro effetti specifici nei diversi campi. Un riflesso importante di questa "rivoluzione tecnologica" è evidente anche nelle discipline storiche, in particolare per quanto riguarda il rapporto tra storia e dato, metodo quantitativo e qualitativo, microanalisi e generalizzazioni, intelligenza umana e capacità analitica dei computer di fronte a una massa enorme di dati. Big Data fornisce metodologie e strumenti innovativi che possono contribuire in maniera rilevante al progresso delle nostre conoscenze. Viste le potenzialità, ogni disciplina tende ad acquisire gli strumenti forniti da Big Data per sviluppare i propri obiettivi specifici. I ricercatori e i cultori delle singole discipline sono sicuramente testimoni di tale processo – soprattutto negli ultimi anni – ma rimangono spesso prigionieri di un'ottica limitata al solo loro campo d'indagine disciplinare.

Può essere invece utile guardare all'impatto dei Big Data da una prospettiva più generale di tipo interdisciplinare, allo scopo di individuarne le caratteristiche comuni e le eventuali criticità. E questo è l'obiettivo del presente volume, che raccoglie un insieme di saggi predisposti da autori, specialisti in ambiti diversi, che hanno partecipato come relatori alla Scuola Estiva di Storia e Filosofia della Scienza, tenutasi a Padova nei giorni 15-17 luglio 2019. Su iniziativa di tre Centri di ricerca dell'Università di Padova, promotori dell'iniziativa, e precisamente il Centro interdipartimentale di ricerca in Storia e Filosofia della Scienza (Cisfis), il Centro interdipartimentale di ricerca in Storia della Medicina (Cism) e il Centro per la storia dell'Università di Padova (Csup), la riflessione sull'impatto Big Data è stata individuata quale obiettivo della Scuola Estiva, dedicata alle *Frontiere della conoscenza: Big Data nelle scienze fisiche, sociali, umanistiche e della vita*, titolo ripreso anche per il presente volume.

L'intento era di avviare una riflessione interdisciplinare circa gli effetti delle metodologie Big Data sull'insieme delle discipline esistenti. Grazie alle analisi e alle riflessioni presentate in tale sede, riteniamo che la Scuola Estiva 2019 abbia avuto successo, come peraltro testimoniato dalla partecipazione e dalla risposta degli studenti. Abbiamo quindi chiesto ai relatori – che in maggioranza hanno accettato – di pubblicare i loro contributi, che costituiscono i diversi capitoli del volume che sottoponiamo ora alla vostra attenzione.

Avvertenza

I criteri di elaborazione del testo e di citazione sono uniformati compatibilmente con le peculiarità degli apparati critici dei singoli testi. I contributi di natura biologica, tecnologica e statistica sono privi di note a piè di pagina, ma riportano la bibliografia essenziale a fine testo.

Natura e ruolo dei dati in Mach e Carnap

di *Fabio Grigenti, Pierdaniele Giaretta**

Ernst Mach (1838-1916) e Rudolf Carnap (1891-1970) sono agli estremi di un ampio movimento di pensiero che tra la metà del secolo XVIII e primi decenni del successivo ha posto a tema la questione dei fondamenti conoscitivi del sapere scientifico. Il primo, scienziato di professione, diede importanti contributi all'acustica, all'elettrologia, all'idrodinamica, alla meccanica e alla psicologia della percezione. La sua carriera si compì tutta all'interno dell'Impero Asburgico, nelle cui Università si formò e divenne poi docente tra i più prestigiosi. Si laureò a Vienna nel 1860 e si dedicò subito a ricerche di fisiologia delle sensazioni; fu in seguito professore di fisica nell'università di Graz e di Praga (1867-1895), dove conseguì i suoi più importanti risultati in campo scientifico. Concluse la sua attività all'Università di Vienna (1895-1901), dove fu titolare della cattedra di Filosofia della scienza. Dal punto di vista della sua visione epistemologica, fu, assieme a Richard Avenarius, autorevole esponente dell'*empiriocriticismo*, per il quale i costrutti teorici della scienza non rispecchiano la natura in sé delle cose, ma si configurano come elaborazioni di tipo pragmatico-strumentale volti alla miglior organizzazione economica dei dati derivati dall'esperienza. Sottolineò l'importanza dell'indagine storico-critica sulla nascita e sullo sviluppo delle teorie scientifiche, opponendosi a ogni dogmatismo. Com'è noto, le sue posizioni esercitarono un notevole influsso su Albert Einstein e sulla genesi del neopositivismo.

Più variato e complesso risulta il contributo filosofico di Carnap. Dal 1909 al 1914 studiò matematica e fisica a Friburgo e Jena, dove seguì le lezioni di logica di Gottlob Frege. Nel 1921, dopo aver conseguito il dottorato, si trasferì a Vienna chiamato da Moritz Schlick come istruttore di filosofia presso l'Università. Negli anni di permanenza nella capitale austriaca (1926-1931) pubblicò diversi lavori, tra i quali *Der logische Aufbau der Welt*, in cui pre-

* Dipartimento di Filosofia, Sociologia, Pedagogia e Psicologia applicata, Università degli Studi di Padova.

sentò in forma sistematica una concezione fenomenistica che cercava di dar conto dell'intero edificio del linguaggio dotato di senso in termini di strutture logiche e dati immediati dell'esperienza vissuta. Da questa iniziale posizione Carnap si staccherà sotto vari punti di vista. Durante il periodo praghese egli abbandonò la pretesa di unicità del linguaggio logico, affermando l'esistenza di molteplici strutture linguistiche in grado di rendere conto degli universi di discorso delle scienze, e rimettendo la scelta tra le varie forme possibili a una libera convenzione (principio di tolleranza). Nell'ultima fase del suo pensiero, dopo il trasferimento negli Stati Uniti, Carnap giunse a sostenere che le teorie scientifiche sono ipotesi interpretative dell'esperienza sottoponibili a controllo empirico, ma non riducibili integralmente ad asserti osservativi.

Anche se in un senso diverso dall'attuale dibattito sui Big Data, sia Mach sia Carnap pongono al centro della loro riflessione due questioni epistemologiche strettamente interrelate tra loro: la prima attiene alla natura dei dati – che cos'è un dato? – la seconda alla forma e al ruolo della loro organizzazione nella conoscenza che, diversamente da altri orientamenti presenti nella stessa epoca, era unitariamente intesa. In aderenza al contesto storico in cui i due filosofi operarono, in larga parte dominato dai dibattiti intorno alla logica e alla psicologia scientifica, la loro concezione della datià fa inevitabile riferimento all'esperienza, intesa come modo privilegiato di accesso al mondo da parte di un essere dotato di capacità di percezione e di successiva sistemazione in strutture più complesse degli elementi primari.

In generale, quindi, sia per Mach che per Carnap, prima di parlare di dati si deve introdurre il tema delle sensazioni e delle osservazioni: le prime prevalentemente connesse a esperienze percettive dirette e le seconde più spesso intese come tali da non escludere regimentazioni e quantificazioni ulteriori. Riguardo alle sensazioni, in genere si suppone che esse abbiano contenuti, che poi sono stati denominati *sense-data*, i quali si darebbero in se stessi prima di ogni misurazione. Riguardo a tali contenuti si è discusso se per concepirli sia essenziale il riferimento alle esperienze di un soggetto o essi siano già individuati indipendentemente dal presentarsi come soggettivi o oggettivi. Mach era di questa opinione, ma non il primo Carnap, il quale, come abbiamo visto, parlava di esperienze vissute elementari. Un'altra questione riguarda la presenza nell'esperienza di dati ultimi, così qualificati in quanto non ulteriormente analizzabili. Mach la assumeva, ma non Carnap.

Oltre al tema relativo al rapporto tra i dati e la soggettività conoscitiva, vi era il problema di comprendere se i dati stessi possano costituire la base a partire dalla quale, attraverso la logica e la matematica, si possono costruire nozioni equivalenti di concetti che nella scienza sono applicati ai dati e posti in relazioni tra loro mediante leggi. Si riteneva che la questione dell'organizzazione successiva alla acquisizione degli elementi primi della conoscenza fosse altrettanto essenziale per la costruzione o giustificazione di più ampie

e coordinate dimensioni di senso, senza le quali era difficile sostenere la pretesa della scienza di costituire un'impresa capace di restituire una visione generale del mondo.

Queste questioni sono in parte estranee al dibattito riguardante i Big Data. Tali elementi sono considerati in una forma quantificata, spesso mediante operazioni di misura, o, comunque, in forme standardizzate e regimentate. Inoltre, non vi è alcuna pretesa di ridurre ad essi i concetti che vengono usati nella scienza. Talora, però, sono ritenuti tali da ridimensionare il ruolo delle teorie. Certamente i problemi sollevati dall'accumulo e dallo sfruttamento dei Big Data hanno una loro specificità. Tuttavia pensiamo che il modo in cui se ne parla ora possa essere utilmente confrontato con problemi e punti di vista proposti in un passato non troppo lontano. In particolare, da una prospettiva epistemologica, a noi sembra che le idee di Ernst Mach e del Rudolf Carnap del 1928 conservino tuttora un notevole interesse per la diversa individuazione di ciò che nell'esperienza sensibile si può considerare primitivo o fondamentale, e, come mostreremo, anche per una loro parziale concordanza e i problemi connessi con l'aspetto sotto il quale concordano.

Parleremo di Mach nella prima parte di questo contributo e poi, nella seconda parte, di Carnap e di Mach e Carnap. Affronteremo per prima la questione dell'esistenza di dati ultimi nella prospettiva di Mach.

I

1. Esistono dati ultimi?

Questa domanda costituisce lo sfondo non sempre nettamente espresso, ma comunque presente, di tutte le analisi svolte da Ernst Mach in *Die Analyse der Empfindungen und das Verhältnis des Physischen zum psychischen* (1896). Il problema dei fondamenti del conoscere è ricorrente e costituisce uno dei tratti specifici dell'epoca di Mach; nell'opera dello scienziato moravo, tuttavia, esso si arricchisce di almeno due connotazioni particolari. La prima concerne l'ampliamento del campo epistemico di indagine, che nella ricerca machiana non comprende più solo la fisica, ma si estende alla psicologia. La seconda attiene al coinvolgimento nella discussione di tre ordini di entità: i corpi, le modificazioni fisiologiche e quello dei cosiddetti "stati interni". Sul piano degli obiettivi, Mach intende determinare con precisione le basi conoscitive del sapere scientifico per poi procedere – anticipando anche qui sviluppi novecenteschi – alla eliminazione di pseudo-nozioni e concetti ambigui che in vario modo minerebbero la saldezza e la credibilità del discorso scientifico. Nel capitolo iniziale del volume, intitolato *Osservazioni preliminari*

antimetafisiche, il momento centrale del ragionamento machiano circa i dati ultimi è occupato dalla considerazione di questa figura:



Mach commenta così l'immagine:

Se io ad esempio sono sdraiato su un divano e chiudo l'occhio destro, al mio occhio sinistro si presenta l'immagine riprodotta nella figura. In una cornice formata dall'arcata sopraccigliare, dal naso e da un baffo appaiono una parte del mio corpo, nella misura in cui esso è visibile, e gli oggetti che lo circondano. Il mio corpo si distingue dagli altri corpi umani, oltre che per la circostanza che ogni vivace idea di movimento si estrinseca subito nel movimento corrispondente e che ogni contatto con esso condiziona mutamenti molto più vistosi di quelli che percepiamo avvenire in altri corpi, anche per il fatto che esso viene visto solo in parte, e in particolare senza la testa. Se osservo un elemento A che si trovi nel mio campo visivo e ne indago la connessione con un altro elemento B, che si trovi nel medesimo campo visivo, passo dal campo della fisica a quello della fisiologia o della psicologia quando B, per usare la vivace impressione usata occasionalmente da un mio amico mentre osservava questo disegno, mi trapassi la pelle. Considerazioni simili a quelle che abbiamo fatto per il campo visivo si possono fare per il campo tattile e per i campi di percezione degli altri sensi¹.

Proviamo ad approfondire il commento di Mach. L'immagine mostra chiaramente una stanza in una prospettiva di profondità. Sul lato opposto,

1. Ernst Mach, *Die Analyse der Empfindungen und das Verhältnis des Physischen zum psychischen*, Jena 1896, p. 15; trad. it. di Libero Sosio, *L'analisi delle sensazioni e il rapporto fra fisico e psichico*, Milano 1975, p. 50.

per quanto lontano da noi, vediamo, prima di tutto e più chiaramente, una finestra. Oltre questa finestra catturiamo un paesaggio a malapena distinguibile. Il complesso di oggetti costituito dalla finestra e dal paesaggio ci viene incontro e richiama per primo la nostra attenzione, nonostante la sua lontananza. In questo venir verso di noi, sono da distinguere due aspetti: quello per il quale sono la finestra e il paesaggio a farsi avanti, e quello dello sguardo, che sembra cogliere e trarre a sé ciò che risiede più in là. Questa stretta implicazione di soggetto e oggetto nella costituzione della rappresentazione è la matrice epistemica che regge tutti i passaggi dell'analisi machiana. Tuttavia, e nonostante questo doppio movimento di approssimazione rispetto al punto di osservazione, la finestra e il paesaggio si mantengono “là in fondo”, appartengono certamente alla dimensione più esterna del campo di visione, sono parte di un mondo che si dirige “verso il fuori”. Deve essere sottolineata l'immediatezza e la subitanea precisione con cui, pur dalla distanza prospettica in cui sono localizzati, questi oggetti si mostrano ogni volta che – dopo aver distolto lo sguardo – torniamo a fissare la scena. Di questo fatto abbiamo conferma indiretta se spostiamo la nostra attenzione sulla parete a sinistra: i particolari dell'arredo che la copre sono più nitidi verso il fondo della stanza e sfumano, divenendo sempre meno riconoscibili, man mano che si procede verso l'osservatore. Anche in questo caso, ciò che “sta più in là”, in “prospettiva” e a “distanza”, è riconosciuto immediatamente e senza incertezza. Ma vi sono altri dati che appaiono quasi contemporaneamente: si tratta delle scarpe e delle gambe incrociate di un individuo che giace sdraiato su una poltrona. A lato – dapprima staccata dal corpo – vediamo una mano che impugna una penna e, procedendo all'indietro – ce ne accorgiamo un attimo dopo – scorgiamo un lato del naso, l'arco delle sopracciglia e il profilo di un paio di baffi. A questo punto la scena ci è divenuta definitivamente chiara: l'immagine non è altro che la porzione di mondo accessibile visivamente a un osservatore (forse Mach stesso) che ha chiuso l'occhio destro e che sta sdraiato su di una poltrona rivolta contro il fondo di una stanza su cui si apre una finestra.

Giustamente Mach osserva che, se consideriamo attentamente la nostra descrizione, risulta chiaro che essa ha portato in primo piano tre insiemi ben delineati di dati: quelli contenuti nella stanza fuori di noi – finestra, paesaggio e mobilio – e quelli legati al corpo dell'osservatore, il quale appartiene al mondo esterno ma contemporaneamente possiede una relazione privilegiata con l'individuo che sta guardando. In effetti, mentre gambe, mano, naso, sopracciglia e baffi sono visti esattamente come la finestra – cioè come dati esterni contenuti nella scena – la vera istanza “guardante”, l'occhio organico, ma soprattutto ciò che sta dietro a questo – ossia l'io, il soggetto o la coscienza rappresentativa – non appartiene alla porzione di mondo raffigurata e non è percepibile in essa. Tuttavia, anche in questo dominio – quello psichico – sono presenti dati, i quali sono strettamente correlati a quelli costituiti dalle

modificazioni fisiologiche del corpo organico e a quelli costituiti dagli oggetti “reali” appartenenti al mondo esterno. Secondo Mach, in definitiva, l’esperienza in generale – e quella scientifica in particolare – si strutturano a partire da tre campi diversificati di entità ultime – fisiche, fisiologiche e psichiche – tra loro strettamente correlate e sempre implicate in ogni atto percettivo, non solo di tipo visivo, ma afferente alla sfera dei sensi nel suo complesso (suoni, tatto, gusto, ecc.).

2. Concatenazione

L’apparente non omogeneità dei dati ultimi non inficia, secondo Mach, la strutturale coordinazione che li lega. La nozione usata per significare il legame tra le differenti sfere dell’esperienza è quella di concatenazione (*Verknüpfung*); essa viene introdotta così:

Colori, suoni, calore, pressioni, spazio, tempo ecc. sono fra loro in vari modi concatenati e con loro sono connesse disposizioni, sentimenti, volontà. Da tale connessione risulta il *quid relativamente persistente* e permanente che si imprime nella memoria e si manifesta con la parola. Come relativamente più persistenti anzitutto, nel tempo e nello spazio, ci si presentano i complessi di colori, suoni pressioni ecc., che per ciò portano nomi diversi e sono distinti come corpi (*Körper*). Ma in nessun modo tali complessi sono assolutamente persistenti. Ad esempio: il mio tavolino ora più ora meno illuminato, ora più ora meno riscaldato; può essere macchiato d’inchiostro; può mancare di un piede².

E in un testo di qualche anno dopo:

Il mondo è fatto di colori, di suoni, di calori, di pressioni, di spazi, di tempi, ecc., *ai quali noi per ora non vogliamo dare il nome di sensazioni né di fenomeni*, perché l’uno e l’altro vocabolo contengono già una teoria unilaterale ed arbitraria. *Li diremo senz’altro elementi*. Conoscere come procedano questi elementi, sia che a ciò si giunga direttamente o indirettamente, è il vero scopo della scienza³.

Nei testi di Mach la definizione del concetto di dato viene cercata all’incrocio dei piani fisico e psichico e nella tensione tra l’idea di *sensazione* e una nuova nozione, significata col termine *elemento*. Il dominio del fisiolo-

2. Ivi, p. 9, trad. it., p. 37.

3. Ernst Mach, *Populär-wissenschaftliche Vorlesungen*, Leipzig 1896, p. 177, trad. it., *L’evoluzione della scienza. Nove “Lezioni popolari”*, a cura di Massimo Debernardi, Milano 2010, p. 162. Su questo vedi: Pietro Gori, *Il darwinismo di Ernst Mach. Riflessioni sul principio di economia della scienza*, «Annali dell’Istituto Italiano per gli Studi Storici», XXII (2006/2007), pp. 223-252.

gico scorre sullo sfondo, per essere affrontato successivamente. Ora, colori, suoni, calori, pressioni, tempo e spazio – che immediatamente saremmo propensi a intendere come sensazioni o come proprietà degli oggetti – sono trattati come dati “neutri”, né soggettivi né oggettivi, e perciò indipendenti sia dal fisico puramente considerato sia dallo psichico puramente considerato. Mach rifiuta anche la prospettiva di ascendenza kantiana, la quale individua nel fenomeno il mattone costruttivo fondamentale dell’edificio cognitivo. Le ragioni di tale presa di posizione sono dichiaratamente epistemiche: assumere nozioni condivise e naturali non mette al riparo da presupposizioni di natura teorica o, peggio, metafisica. Nei casi segnati Mach ritiene che la sovrastruttura filosofica legata ad alcuni concetti ammessi senza alcuna previa discussione possa indebolire fin dall’inizio una ricerca che invece vuole determinarsi come attenta a non imbarcare surrettiziamente presupposti non adeguatamente fondati. Va poi considerato che per Mach, a differenza di quanto in genere si presuppone, una delle caratteristiche preponderanti dei dati intesi come elementi non è la saldezza e l’immutabilità, ma, al contrario, la non persistenza. Su come debba essere intesa questa mutevolezza del dato, aiutano alcuni passaggi che Mach dedica alla descrizione dei rapporti tra i domini di elementi:

Indicheremo con le lettere A B C... K L M... $\alpha \beta \gamma...$ gli elementi poco fa stabiliti. Indicheremo, per chiarezza, con A B C... i complessi di colori, suoni, ecc. che vengono chiamati abitualmente corpi (*Körper*); indicheremo con K L M... il complesso che viene designato come il nostro corpo [*Leib*] e che è una parte dei primi contraddistinta in virtù di talune particolarità; indicheremo infine con $\alpha \beta \gamma...$ il complesso di volizioni, ricordi, ecc. Di solito il complesso $\alpha \beta \gamma...$ K L M... viene contrapposto come io al complesso A B C... come mondo fisico; a volte viene considerato come io il complesso $\alpha \beta \gamma...$ mentre i complessi A B C... K L M... vengono considerato globalmente come mondo fisico. Dapprima A B C... appare indipendente dall’io e rispetto a esso autonomo e contrapposto. Quest’autonomia è solo relativa e non resiste a una considerazione attenta. Nel complesso $\alpha \beta \gamma...$ possono peraltro prodursi vari mutamenti senza che la cosa diventi molto percepibile in A B C..., e vale anche l’inverso. Molti mutamenti che hanno luogo in $\alpha \beta \gamma...$ trapassano, attraverso mutamenti in K L M..., in A B C... e viceversa (Quando ad esempio pensieri vivaci si estrinsecano in azioni oppure quando l’ambiente induce nel nostro corpo mutamenti considerevoli)⁴.

In relazione a questo e ad altri passi, è importante notare che la classica distinzione tra io e corpo (o tra anima e materia) viene complicata per la duplicazione del campo soggettivo in elementi fisiologici e psichici in senso più stretto. Secondo Mach, il termine io può essere riferito all’unione dei

4. Mach, *Die Analyse*, p. 7; trad. it., p. 42.

complessi $\alpha \beta \gamma \dots$ e K L M..., che in questo caso si contrapporrebbero ad A B C... inteso come mondo fisico, ma “io” può essere considerato il solo complesso $\alpha \beta \gamma \dots$, mentre K L M... A B C... vengono globalmente identificati col mondo fisico. In effetti, il corpo individuale appartiene anche alla realtà esterna, per cui è possibile includerlo sia in una indagine fisica, attenta alle relazioni dei corpi tra di loro, sia in uno studio psicologico, che presterà attenzione ai complessi di elementi che riguardano pulsioni e istinti. Mach non sembra dare troppo peso a questa distinzione, dal momento che nella sua visione il corpo e l'io rappresentano *complessi di elementi* (e perciò di dati) che possono essere trattati senza fare riferimento all'unità di cui farebbero parte, specialmente se si considera che ogni aggregato non possiede una stabilità definita.

Ciò che a Mach preme invece evidenziare è il venir meno del contrasto tra l'io e il mondo come entità globali, perché ciò che importa è «unicamente la connessione degli elementi»⁵ nel rapporto tra domini differenti. Un esempio di come deve essere trattata una correlazione tra dati appartenenti ai complessi ABC..., KLM... e $\alpha \beta \gamma \dots$, è il seguente: un cubo se vicino ci appare grande, piccolo se lontano; visto con l'occhio destro ci appare in un modo; con l'occhio sinistro in un altro modo, talvolta addirittura duplicato; se, infine, chiudiamo gli occhi, il cubo scompare. È evidente che il complesso delle qualità KLM... modifica la propria configurazione in rapporto al mutamento del complesso ABC...; la distanza fisica degli oggetti influisce sul profilo funzionale degli organi di senso, ma è anche vero il contrario una modificazione dell'assetto percettivo del corpo (cambiare posizione, chiudere un occhio) ha un'immediata conseguenza sulla conformazione fisica del dato esterno. Tuttavia, nei diversi modi in cui cogliamo il cubo noi sappiamo sempre riconoscere un nucleo persistente, che chiamiamo appunto il “cubo”. Questo *quid* appartiene però a un altro dominio di datità, che corrisponde al complesso $\alpha \beta \gamma \dots$, ossia al dominio degli elementi mentali. In quest'ultimo, il dato “cubo”, già presente come “oggetto” nel mondo fisico e “modificazione fisiologica” nel corpo, emerge come un dato psichico, che a sua volta può collegarsi ad altri dello stesso tipo. Nel caso in esame possiamo ipotizzare che una certa rappresentazione o idea sia immediatamente connessa a un elemento di natura fono-simbolica, la parola “cubo”, appunto.

Fino a questo punto, può ancora sembrare che Mach pensi all'esistenza di complessi di datità che sono differenti in maniera intrinseca. In effetti l'uso di lettere diverse per designare gli oggetti compresi nei diversi domini potrebbe lasciar intendere questa possibilità interpretativa. Tuttavia, come vedremo meglio subito, lo scienziato moravo intende sostenere il contrario: non ci sono

5. Ivi, pp. 11, 46.

mondi composti da entità dotate di proprietà eterogenee, ma un complesso di dati ultimi provvisti di caratteristiche semplici e generali, che si arricchiscono di ulteriori connotazioni nel passaggio da un ambito a un altro. Nel caso appena visto dobbiamo ipotizzare che si dia un elemento “cubo” il quale, trapassando dal dominio fisico a quello psichico, da un lato permane come dato in sé, indipendente dal contesto, dall’altro assume via via determinazioni dipendenti dall’ambiente funzionale in cui si trova immerso. Il “cubo fisico” è tale solo nel mondo fisico, e per questo avrà dimensioni, massa, orientamento, mentre il “il cubo sensazione” e la “immagine mentale” del cubo, pur essendo ancora il “cubo”, assumeranno configurazioni particolari dipendenti dalla fisiologia dei sensi e dalle dinamiche che presiedono alla formazione delle rappresentazioni mentali. Del resto, se non fosse così, Mach non potrebbe in alcun modo sostenere – come vedremo sotto – una teoria in cui il fisico e lo psichico non sono separati da un muro invalicabile, ma interconnessi in uno scambio di reciproca co-determinazione.

Visto il ruolo di intermediazione svolto tra mondo fisico e psichico, le modificazioni appartenenti a K L M..., ossia i dati derivanti dal mutamento dello stato corporeo sembrano possedere una sorta di primato nella determinazione di nuove riconfigurazioni di elementi sia in ABC... sia in $\alpha \beta \gamma...$; questa circostanza ha certamente favorito le letture “empiriste” o “fenomeniste” di Mach, ma i testi e le intenzioni dello scienziato moravo sembrano configurare un altro orizzonte di pensiero:

La grande distanza che sussiste fra la ricerca fisica e quella psicologica è dunque tale solo per il comune modo stereotipato di considerare le cose. Un *colore* diventa un *oggetto fisico* non appena consideriamo ad esempio la sua dipendenza dalla sorgente di luce che illumina (altri colori, calore, spazi ecc.) Se invece poniamo mente alla sua *dipendenza dalla retina* (dagli elementi KLM...) esso diventa un oggetto *psicologico*, una *sensazione*. Non *l'oggetto* bensì *l'orientamento della ricerca* è diverso nei due campi⁶.

In effetti, se consideriamo la nozione di dato coestensiva a quella di elemento, vediamo che per Mach il *medesimo dato* assume una diversa *datità* in relazione al campo elementare da cui dipende funzionalmente. Il colore è un dato fisico in relazione alla luce, ma è un dato sensibile in relazione alla retina. Da questo punto di vista, tra i dati non è possibile stabilire alcuna gerarchia di valore in ordine alla conoscenza, ciò che conta è la correlazione che mutuamente si stabilisce tra singoli o interi gruppi di elementi appartenenti a domini diversi. Tuttavia, i dati sensibili per la loro vicinanza e familiarità – in quanto sono modificazioni del corpo proprio – paiono di nuovo rivestire

6. Ivi, p. 14; trad. it., p. 48.

un ruolo preponderante nel mutamento dei campi di esperienza. Una trasformazione dello stato corporeo produce sempre modificazioni di tutta la massa degli elementi in gioco⁷.

In realtà Mach sembra talora voler esprimere – e ciò sarebbe in linea con la posizione di prevalenza prospettica assunta dall’osservatore nell’immagine che abbiamo riprodotto in apertura – un’asimmetria di tipo spaziale e fisiologico. In primo luogo è solo dal punto del soggetto di conoscenza che può essere esperito un mondo, in secondo luogo è anche evidente che nessun atto cognitivo può prescindere dal *medium* corporeo. Tuttavia, nessuna di queste condizioni implica che il nostro mondo debba essere considerato come esclusivamente formato da sensazioni, ma solo che è impossibile metterci in relazione con i dati dell’universo fisico prescindendo dalla mediazione fisiologico-percettiva.

3. Monismo

Mach non intende attribuire alle sensazioni il valore costitutivo di dati primi in ordine alla conoscenza della realtà, pur riconoscendo loro l’imprescindibile ruolo di elementi “immediati” nel processo di rappresentazione del mondo. La prospettiva di Mach va piuttosto nella direzione di ammettere l’esistenza di una sola realtà, composta da un insieme di dati che formano complessi più o meno stabili, ma comunque mai isolati gli uni dagli altri; la nostra conoscenza di questa struttura viene certamente attuata da un *punto di vista* particolare, quello del soggetto senziente, ma nel complesso l’analisi riposa su di una concezione monistica, che fa riferimento ad entità elementari prive di qualsiasi caratterizzazione propria, le quali tuttavia si modificano trapassando da un dominio all’altro eppur mantenendo un elemento un aspetto strutturale inalterato. Va di nuovo ribadito che, nonostante venga sottolineato il ruolo delle sensazioni, la soggettività per Mach non va mai intesa in senso idealistico: l’occhio che vede non possiede lo statuto di un fondamento auto-costituito capace di mettere in essere un mondo come prodotto di atti di coscienza, ma vale solo come istanza puramente prospettica, dalla quale può essere aperta la scena dell’esperienza. Il dominio delle sensazioni non ha dietro di sé un modo interno, che pur rispecchiandolo, si configura come un dominio altro e del tutto irriducibile a quello che chiamiamo “esterno”. Alcuni passaggi dei testi sono inequivocabili:

7. Espressioni quali «in tal caso sarebbe giusto dire che il mondo sussiste soltanto per le nostre sensazioni. Ma noi appunto conosciamo soltanto le sensazioni» (cfr. *ivi*, p. 14; trad. it., p. 48) introducono certamente elementi di incertezza circa il reale pensiero di Mach, che in taluni punti può apparire ambiguo o contraddittorio.

Non esistono interiorità ed esteriorità, né sensazioni che corrispondano a cose esterne da sé distinte. Esistono solo elementi di una medesima specie i quali compongono sia la presunta interiorità sia l'altrettanto presunta esteriorità⁸.

I dati ultimi, in definitiva, sarebbero entità elementari prive di qualsiasi determinazione qualitativa, tanto che alcuni critici hanno voluto leggere la filosofia di Mach come un monismo neutrale⁹. I componenti fondamentali della realtà non manifestano alcun tipo di qualità se non a partire dalla relazione con gli elementi del dominio a cui appartengono e con i dati degli altri complessi corporei; la loro particolare qualità fisica o psichica che sia, dipende dal modo in cui vengono considerati, e ogni elemento può valere nei diversi ambiti in maniera differente. È l'insieme della struttura derivante dalle relazioni funzionali tra i dati a determinare la qualità complessiva di una certa esperienza del mondo.

Questo modo di considerare la realtà naturale fa naturalmente cadere la distinzione tra la ricerca fisica e quella psicologica, ponendosi su un piano differente, equidistante da entrambe. La presunta distanza tra le due discipline di studio risulta essere per Mach nient'altro che un «comune modo stereotipo di considerare le cose»¹⁰ dal momento che dal suo punto di vista si tratta semplicemente di interpretare attraverso canoni differenti i medesimi elementi, i quali possono venir considerati come dati ultimi in quanto si situano in una dimensione più astratta e generale rispetto ai fatti fisici e psichici, essendo appunto privi di qualsiasi carattere che riconduca ad uno dei due ambiti.

II

Non c'è un'unica concezione di dato empirico che sia stata formulata e accettata nell'ambito del neopositivismo. Quella di Mach era nota ai neopositivisti, ma nei loro scritti, anche negli scritti di Rudolf Carnap che del neopositivismo è stato uno degli esponenti più significativi, se ne trovano anche altre riguardo alle quali si sviluppò una vivace discussione.

Lo scopo di questa parte è di prendere in considerazione la concezione di dato presente in *Der logische Aufbau der Welt*¹¹ (d'ora in poi menzionato

8. Ivi, p. 16; trad. it., p. 49.

9. Per un inquadramento generale di Mach anche dal punto di vista filosofico e per ulteriori indicazioni bibliografiche rimandiamo a Pietro Gori, *E. Mach, tra scienza e filosofia*, Pisa 2018.

10. *Ibidem*.

11. Rudolf Carnap, *Der logische Aufbau der Welt, Scheinprobleme in der Philosophie: das Fremdpsychische und der Realismusstreit*, Hamburg 1961; originariamente: *Der logische Aufbau der Welt*, Berlin-Schlachtensee 1928, e *Scheinprobleme in der Philosophie: das*

come *Aufbau*) ed esaminare se essa sia davvero inconciliabile con quella di Mach, come a prima vista potrebbe sembrare. La nostra risposta sarà che vi è una certa convergenza tra le due concezioni, anche se Carnap percepiva un contrasto netto tra la sua nozione soggettivista del dato e quella neutralista di Mach, secondo la quale il dato non è in sé né soggettivo né oggettivo. Poiché mettere in evidenza questa convergenza è l'obiettivo principale di questa seconda parte del nostro contributo, ci si limita a delineare in modo sommario la cornice nella quale la nozione di dato si colloca come punto di partenza e, almeno in parte, anche come base fondazionale. Seguirà un accenno ad una problematica filosofica che più recentemente è stata sollevata riguardo alla rappresentazione dei dati e alla concezione strutturalista della scienza.

1. Scopo e articolazione del progetto di Carnap

Scopo dichiarato dell'*Aufbau* è delineare, e in parte realizzare, il progetto di una costruzione logica del mondo. Di fatto l'*Aufbau* è – più esattamente – il tentativo di ricostruire la scienza del mondo, dando rilievo preminente alla distinzione tra componente empirica e componente logica, e quindi cercando di limitare il più possibile gli aspetti indipendenti dai dati e dalla logica. La natura precisa e le difficoltà del progetto di Carnap sono state ampiamente discusse nella letteratura¹². Qui ci limitiamo a considerare le caratteristiche generali del suo progetto.

Punto di partenza e fondamento della sua ricostruzione sono le esperienze vissute elementari del singolo soggetto (*Elementare Erlebnisse*, d'ora in poi *eve*), che costituiscono la componente empirica di base. Di esse Carnap dice che appartengono al campo psichico proprio o ad un campo psichico altrui. Quelle appartenenti al campo psichico proprio sono considerate conoscitivamente primarie, perché le seconde sono conosciute per mezzo delle prime¹³.

Fremdpsychische und der Realismusstreit, Berlin-Schlachtensee 1928; trad. it. di Emanuele Severino: *La costruzione logica del mondo, Pseudoproblemi nella filosofia*, Torino 1997. Le citazioni sono tratte dalla traduzione italiana, talora con qualche modifica segnalata mettendo il testo modificato tra le parentesi []. Per facilitare l'individuazione in altre edizioni dei passi citati abbiamo aggiunto l'indicazione della sezione.

12. I principali e più noti promotori di tale discussione sono stati Nelson Goodman (*A Study of Qualities*, Ph.D. dissertation thesis, Harvard University 1941, pubblicata da Garland (Harvard Dissertations in Philosophy Series), New York 1990; *The Structure of Appearance*, Harvard University Press, Cambridge, MA, 1951, seconda ed. Bobbs-Merrill, Indianapolis 1966; terza ed. Reidel, Boston 1977) e Willard Van Orman Quine (*Two Dogmas of Empiricism*, «Philosophical Review», 60 (1951), pp. 20-43, poi in *From a Logical Point of View*, Harvard University Press, Cambridge, MA, 1953, seconda ed. 1980, pp. 20-46, trad. it., *Il problema del significato*, Roma 1966, pp. 20-44).

13. Carnap, *Aufbau*, p. 202, sez. 54.

Carnap si preoccupa di limitare le assunzioni ontologiche riguardo a tali esperienze, precisando che non si deve contrapporle a quelle altrui considerando le prime reali e le seconde irreali. Nemmeno si dovrebbe assumere che nel concetto di esperienza psichica sia implicito il riferimento a un io. La qualificazione stessa di psichico dovrebbe essere considerata attribuibile al dato solo dopo il compimento della costituzione logica e non originariamente presente nella sua acquisizione.

La logica in senso lato, comprensiva di almeno una parte della teoria degli insiemi, interviene nella fondazione della parte concettuale e teorica della scienza del mondo, innanzitutto della scienza fisica. A questo proposito è bene ricordare che punto di riferimento erano per Carnap i vari tipi di fondazione o riduzione che erano già state fornite in ambito matematico. Consideriamo, come esempio non troppo complesso, la ricostruzione dei numeri interi a partire dai numeri naturali (gli interi positivi non nulli). Data la relazione di equivalenza tra coppie di naturali (m, n) e (m', n') tale che (m, n) è equivalente a (m', n') se solo se $m + n' = m' + n$, intendendo come classe di equivalenza della coppia (m, n) la classe delle coppie a questa equivalenti nel senso definito, i numeri interi risultano identificabili con tali classi di equivalenza, in particolare -1 con la classe di equivalenza $[0,1]$, alla quale appartengono $(0,1), (1,2), (2,3) \dots$, -2 con la classe di equivalenza $[0,2]$, alla quale appartengono $(0,2), (1,3), (2,4) \dots$ e così via. L'identificazione è possibile perché si può dimostrare che tali classi di equivalenza soddisfano le proprietà strutturali che caratterizzano i numeri interi. La relazione d'ordine può essere definita facendo riferimento all'ordine dei naturali. L'addizione e la moltiplicazione tra interi, rappresentati come classi di coppie equivalenti di naturali, possono essere definite mediante le operazioni corrispondenti sui numeri naturali in modo tale che il risultato rappresenti il numero intero "giusto". In sintesi si possono definire l'ordine e le operazioni tra le suddette classi di equivalenza in modo tale che esse abbiano le proprietà attribuite ai numeri interi. Così si ottiene una classe di numeri più ampia dei naturali perché si può identificare il numero naturale $n > 0$ con la classe $[(n, 0)]$ e assumere la classe $[(0, n)]$, con $n > 0$, come l'intero negativo $-n$. L'intero 0 è rappresentato dalla classe $[(0,0)]$. È importante notare che essenziale è il ruolo della relazione di equivalenza per permettere di identificare un numero intero con una classe di coppie equivalenti.

Analogamente, in modo più o meno complesso, ogni classe numerica più ampia può essere ottenuta da una classe numerica meno ampia, in particolare i razionali dagli interi e i reali dai razionali. Se alla fine anche per i numeri naturali si riesce a fornire una riduzione a nozioni primitive più generali, come ad esempio quella di insieme, si ottiene una fondazione della matematica. È noto che i progetti fondazionalisti della matematica concepiti da Frege e da Russell non sono stati realizzati in modo soddisfacente. Il progetto fonda-

zionalista che Carnap ha abbozzato per la scienza, ispirandosi parzialmente alla fondazione logicista della matematica, va incontro a difficoltà ancora maggiori, anche per la vastità del progetto che riguarda qualunque cosa possa essere oggetto di conoscenza scientifica, a partire dalle qualità sensibili. Qui, mediante i seguenti punti 1-3 di questa sezione, ci limitiamo a fornire una rassegna sintetica delle caratteristiche molto generali e delle difficoltà principali del progetto carnapiano.

1) Le *eve* sulle quali si basa la ricostruzione progettata sono entità empiriche. Carnap parte da esse per cercare di ricostruire le qualità come entità estensionali, in particolare come opportune classi di *eve*, ad esempio la qualità rosso come classe di *eve* simili tra loro. Ma la ricostruzione va incontro alla difficoltà che nessuna relazione che egli introduce a questo scopo, sulla base della più basilare relazione di ricordo di similarità, è tale da rendere un'attribuzione di rosso equivalente all'affermazione di somiglianza con un qualunque membro di una classe che si possa considerare rappresentazione adeguata della proprietà di essere rosso. Inoltre, riguardo alle *eve* e a differenza dei numeri naturali, non si può ovviamente pensare ad una loro reinterpretazione in termini logici. La componente empirico-descrittiva fornita dalle *eve* appare ineliminabile.

2) Le *eve* e le qualità non bastano da sole a costituire il mondo di cui parla la fisica. La costruzione dello spazio fisico richiede qualcosa di diverso e di più complesso. Inizialmente Carnap considera come attribuire qualità a punti e identifica le cose visibili come opportune classi di punti ai quali sono attribuite qualità. Ma il valore conoscitivo dell'attribuzione di qualità è limitato: «*La formazione concettuale* (e quindi anche la costituzione che la [segue]) *del mondo della percezione ha una validità soltanto provvisoria: essa deve far posto, nello sviluppo della conoscenza (e, rispettivamente, della costituzione) al mondo della scienza fisica, rigorosamente univoco, ma completamente privo dell'aspetto qualitativo*»¹⁴. Il mondo della scienza fisica si ottiene attraverso l'attribuzione di proprietà fisiche ai punti di uno spazio matematico astratto. Carnap dice:

[Nel mondo della fisica] ai punti dello spazio numerico quadridimensionale sono attribuiti dei puri numeri, le “*grandezze fisiche*”. Lo scopo di questa costituzione consiste nel costruire un campo che sia determinato da *leggi matematicamente concepibili*. Tali leggi devono essere matematicamente concepibili affinché col loro aiuto possano venire *calcolate* certe determinazioni a partire da quelle altre mediante le quali esse sono determinate. La necessità della costituzione del mon-

14. *Ibidem*.

do fisico dipende inoltre dalla circostanza che solo questo e non il mondo della percezione [...] fornisce la possibilità di una intersoggettivazione univoca e esente da contraddizione¹⁵.

Carnap non nega che esistano rapporti di regolarità tra qualità sensibili e quindi l'intersoggettività sia possibile anche al livello della percezione, ma aggiunge: «la via di gran lunga più semplice per pervenire a un campo caratterizzato da una regolarità e calcolabilità generali consiste comunque nella costituzione del mondo fisico come mondo puramente numerico»¹⁶.

3) Quali grandezze scegliere per costituire il mondo fisico dipende dall'«empiria», ma anche dall'adozione di principi metodologici fondamentali, per esempio quello della maggior semplicità¹⁷. Le leggi naturali, a loro volta, dipendono dalla scelta delle grandezze. In ogni caso, quali che siano le scelte fatte e i principi metodologici assunti, il mondo fisico è costituito coordinandolo a quello della percezione (coordinazione fisico-qualitativa). Di tale coordinazione Carnap dice che al punto cosmico coordinato del mondo percepibile è attribuita la qualità coordinata a una certa struttura numerica di grandezza fisiche, mentre «mediante l'attribuzione di una qualità a un punto cosmico del mondo percepibile non resta determinato quale singola struttura di grandezza sia attribuita all'ambito del punto cosmico-fisico coordinato, ma resta determinata solo una classe, alla quale questa struttura deve appartenere»¹⁸.

Complessivamente, senza entrare in dettagli che non sempre sono chiari, si può dire che il mondo della scienza fisica ricostruito da Carnap non è totalmente ridotto alla base delle *eve* che Carnap pone a fondamento della ricostruzione. W.V.O. Quine ha enfatizzato l'incompletezza della riduzione osservando, in particolare, l'assenza di una definizione eliminativa per la relazione 'is at' tra gli oggetti della base esperienziale e lo spazio fisico¹⁹. È tuttavia improprio parlare di fallimento del progetto riduzionista di Carnap perché il progetto di Carnap non voleva essere completamente riduzionista.

15. Ivi, p. 331, sez. 136.

16. Ivi, p. 332, sez. 136.

17. Vedi ivi, p. 332, sez. 136.

18. Ivi, p. 333, sez. 136.

19. Quine, *From a Logical Point of View*, p. 40; *Il problema del significato*, p. 39.

2. La tensione tra la soggettività della base iniziale e l'oggettività perseguita

I seguenti due punti sono essenziali per comprendere la natura del progetto di Carnap. Il primo riguarda la difficoltà di assumere a fondamento dati soggettivi e quindi l'esigenza di superare la soggettività dei dati assunti a fondamento, mentre il secondo riguarda la possibilità di considerare ininfluente tale soggettività e quindi superarla.

1) Carnap ritiene che i dati di base, le *eve*, siano soggettivi ma lo scopo finale è quello di dar ragione di come a partire da essi si possa pervenire al mondo oggettivo della scienza. Carnap avverte un pericolo e una tensione delle quali è pienamente consapevole:

Se la base del sistema di costituzione viene posta in ciò che è psichicamente proprio, sembra che non si possa evitare il pericolo del soggettivismo. Si presenta quindi il problema di come sia possibile [...] raggiungere l'oggettività della conoscenza. [...] l'oggettività sta anche a significare l'indipendenza dal soggetto giudicante, la validità anche per gli altri soggetti. È proprio questa *intersoggettività ad essere una proprietà essenziale della "realtà"*, che serve insieme a distinguere la realtà dal sogno e dall'illusione. Il problema si presenta dunque in questo modo: come perverrà la scienza ad asserti intersoggettivamente validi se tutti i suoi oggetti sono costituiti a partire da un soggetto individuale, e dunque tutti gli asserti della scienza hanno fondamentalmente per oggetto solamente dei rapporti tra "i miei stati vissuti"? Poiché la corrente dei dati vissuti è diversa per ogni uomo, come potrà essere obiettiva, in questo senso, anche una sola proposizione della scienza, e cioè come potrà valere per ogni individuo, allorché uscirà dalla sua corrente individuale di dati vissuti?²⁰.

2) Secondo Carnap è possibile pervenire a un mondo oggettivo a partire da dati soggettivi quali sono le *eve*, perché queste presentano relazioni strutturali, dette anche formali, che sono indipendenti dal soggetto. Dice Carnap:

La soluzione di questo problema consiste nel fatto che, certamente, il *materiale* delle correnti individuali di dati vissuti è completamente diverso, o, piuttosto, è affatto incomparabile, giacché è assurdo un confronto tra due sensazioni o tra due sentimenti appartenenti a soggetti diversi e considerati come qualità immediate del dato; ma certe *proprietà strutturali* concordano in tutte le correnti di dati vissuti. La scienza deve limitarsi agli asserti intorno a queste proprietà strutturali, giacché essi sono oggettivi. Ed essa, anche, *può* limitarsi agli asserti di struttura [...], poiché tutti gli oggetti della conoscenza non sono contenuto, ma forma, e possono venir rappresentati come formazioni strutturali. Solo sul fondamento di questa conoscenza, che cioè

20. Carnap, *Aufbau*, pp. 221-222, sez. 66.

la scienza è per essenza scienza della struttura, e che pertanto esiste una via per costituire l'oggettivo uscendo dalla corrente individuale dei dati vissuti, solo su questo fondamento la forma sistematica con base psichica propria è accettabile²¹.

«Costituire l'oggettivo uscendo dalla corrente individuale dei dati vissuti» vuol dire costituire un mondo intersoggettivo di cui Carnap dice che «forma l'autentico campo di oggetti della scienza», un campo di oggetti da intendersi come obiettivo da raggiungere, poiché secondo Carnap

la scienza non contiene solamente asserti intersoggettivi, ma anche asserti non-intersoggettivi, tali che ad essi corrispondono asserti intersoggettivi o che possano essi venir trasformati in asserti intersoggettivi. Questa trasformazione spetta ai compiti della scienza; gli sforzi della scienza mirano a pervenire ad una situazione in cui non vi siano altri asserti che quelli intersoggettivi²².

L'aspetto strutturale attribuito alle *eve*, presentato e discusso nella prossima sezione, permette a Carnap di iniziare un percorso attraverso il quale pervenire a costituire la scienza come sapere oggettivo unitario. In questo percorso convenzione ed esperienza hanno ruoli fondamentali. In una delle ultime sezioni, la 179 intitolata «il compito della scienza», Carnap ricorda ed enfatizza tali ruoli nell'ambito di una visione unitaria del sapere scientifico:

Inserendo gli oggetti della scienza in un unico sistema unitario costituzionale, anche le diverse scienze vengono contemporaneamente rilevate come rami dell'unica scienza unitaria e organizzate in un unico sistema. Orbene, come si determina *il compito della scienza unitaria* dal punto di vista della teoria della costituzione? Il fine della scienza consiste nel trovare e nell'ordinare gli asserti veri intorno agli oggetti della conoscenza. (Non tutti gli asserti veri, bensì una scelta di essi, da effettuare secondo determinati principi). [...] tali oggetti devono essere costituiti. *La costruzione del sistema di costituzione* è pertanto *il compito primo della scienza*. [...] Al primo compito, quello della costituzione degli oggetti, è concatenato, come *secondo compito*, *quello della ricerca delle restanti proprietà* e rapporti non costituzionali degli oggetti. Il primo compito viene assolto mediante una convenzione, questo secondo, invece, mediante *l'esperienza*. (Secondo la concezione della teoria della costituzione, nella conoscenza non esistono altre componenti oltre queste due: quella convenzionale e quella empirica; non c'è dunque posto per una sintesi a priori). Come già si è detto, nel processo reale della scienza i due compiti sono quasi sempre fra loro collegati²³.

21. Ivi, p. 222, sez. 66.

22. Ivi, p. 353, sez. 149.

23. Ivi, pp. 410-411, sez. 179. È importante tenere presente che qui Carnap si riferisce sempre alla scienza unitariamente considerata, anche quando usa solo il termine "scienza".

In realtà, se si considera attentamente la teoria della costituzione così come viene precedentemente delineata da Carnap, anche la costituzione degli oggetti presenta aspetti che sono determinati dall'esperienza e quindi i due aspetti della convenzione e dell'esperienza risultano già compresenti a livello di costituzione, contro ciò che alla fine Carnap sembra suggerire nel passo qui sopra citato. Più precisamente, il convenzionalismo dell'*Aufbau* riguarda la scelta dello spazio matematico astratto e delle grandezze fisiche, quindi anche in parte delle leggi. Non però tutto ciò che è formale dipende da convenzioni perché Carnap dice anche che «la forma [di sistema] e le forme [di] oggetto del sistema di costituzione sono determinate empiricamente»²⁴.

Integralmente convenzionalista è invece la sua concezione della logica e della matematica, poiché Carnap dice che «la logica (matematica inclusa) non consiste in altro che in convenzioni sull'uso dei segni e in tautologie [...] sulla base di tali convenzioni»²⁵.

3. L'identificazione strutturale del dato e la convergenza con Mach

Carnap ritiene che il percorso da lui delineato di costituzione degli oggetti della conoscenza (e della scienza) possa essere fatto a partire da una base diversa da quella psichica propria, in particolare «dai dati vissuti di tutti i soggetti o anche dalla base fisica»²⁶. Ovviamente l'adozione di una base fisicalistica, cioè una base osservativa costituita da oggetti intesi come entità dello stesso tipo di quelle delle quali si parla in fisica (o in una parte della fisica) o anche da una base neutrale nel senso di Mach, secondo il quale gli elementi costitutivi ultimi dell'esperienza non sono di per sé qualificabili né come soggettivi né come oggettivi, avrebbe reso meno difficile il problema di come pervenire al livello di oggettività richiesto dalla scienza. Perché il Carnap del 1928 non ha fatto una di queste scelte? E come, avendo scelto di adottare una base di dati soggettivi, ha affrontato il problema di superare la loro soggettività per pervenire all'oggettività della scienza?

Alla prima domanda si può rispondere osservando che Carnap è interessato a partire da esperienze conoscitivamente primarie, quali secondo lui sono le *eve*. È da questo punto di vista che egli motiva la sua scelta e critica alcune scelte alternative:

24. Ivi, p. 287, sez. 103.

25. Ivi, p. 295, sez. 107. Sotto questo aspetto Carnap anticipa la concezione che esporrà estesamente in *Die logische Syntax der Sprache*, Wien 1934, trad. it. di Alberto Pasquelli, *Sintassi logica del linguaggio*, Milano 1966.

26. Ivi, p. 406, sez. 176.

noi scegliamo come base, per il progetto del sistema di costituzione, il campo psichico proprio. La ragione più importante della qual cosa risiede nell'intenzione di rappresentare, mediante questo sistema di costituzione, non solo un ordinamento logico-costituzionale degli oggetti, ma anche il loro ordinamento conoscitivo [...]. Per la stessa ragione resta certamente escluso anche l'impiego della forma sistematica con base fisica, per la quale sono state rilevate possibilità logiche diverse. Ora, è vero che, talvolta, vien sostenuta la concezione che non già il campo psichico proprio, bensì il campo psichico generale formi il campo fondamentale anche dal punto di vista dell'ordinamento conoscitivo; ma tale concezione non può essere mantenuta, se si tiene presente il fatto che la conoscenza di ciò che è psichicamente estraneo non è possibile senza la mediazione costituita dalla conoscenza del campo fisico [...]. La seconda ragione della preferenza per la forma sistematica con base psichica propria è di carattere logico-formale. Anche qualora un sistema di costituzione con base psichica generale fosse in grado di pervenire alla rappresentazione anche dell'ordinamento conoscitivo degli oggetti, tuttavia un sistema di quell'altra forma avrebbe il vantaggio di costituire la stessa totalità di oggetti sul fondamento di una base considerevolmente più ristretta²⁷.

Esplicitamente Carnap considera già refutata nella letteratura la «posizione antisolipsistica» di Mach²⁸. Inoltre, insistendo che la base esperienziale deve essere data in modo originario, prima di qualunque sua rielaborazione o interpretazione, Carnap così critica la concezione atomistica, condivisa da Mach, degli elementi di base dell'esperienza:

Si potrebbe forse pensare di assumere, come elementi fondamentali, gli ingredienti ultimi che risultano dall'analisi psicologica e fenomenologica dei dati vissuti, come potrebbero essere le sensazioni più semplici (come propone Mach [in *Die Analyse der Empfindungen*, Jena 1886]), o, più in generale, gli elementi psichici di specie diverse, con i quali possono venir costruiti i dati vissuti. Ma ad una considerazione più adeguata, dobbiamo tuttavia riconoscere che in questo caso non viene assunto come elemento fondamentale il dato stesso, bensì delle astrazioni ottenute da quest'ultimo, e quindi qualcosa di conoscitivamente secondario. I sistemi di costituzione che muovono da tali elementi fondamentali sono certamente corretti e realizzabili tanto quanto potrebbero esserlo sistemi con base fisica. Ma poiché noi esigiamo che il nostro sistema di costituzione tenga conto anche dell'ordine conoscitivo degli oggetti [...], dobbiamo muovere allora da ciò che è conoscitivamente primario rispetto a ogni altra cosa, e cioè dal «dato» [*Gegebene*] e il dato sono *i dati vissuti* [*Erlebnisse*] *stessi nella loro totalità e unità conchiusa*²⁹.

La seconda questione, riguardante il passaggio dal soggettivo all'oggettivo, è affrontata da Carnap fin dalle prime sezioni dell'*Aufbau*. La soggettività

27. Ivi, pp. 215-6, sez. 64.

28. Ivi, p. 218, sez. 64.

29. Ivi, p. 223, sez. 67.

delle *eve* è ritenuta innegabile, ma anche costituisce una prima difficoltà da superare per giungere a giustificare l'oggettività della scienza. Carnap affronta il problema chiedendosi sotto quale condizione le *eve* siano utilizzabili allo scopo della ricostruzione logica del sapere scientifico, innanzitutto quello della fisica. La sua risposta è che esse lo siano solo nella misura in cui esibiscono delle relazioni strutturali con le seguenti caratteristiche:

1) Tali relazioni sono dette strutturali perché di esse si considerano solo proprietà riguardanti la loro estensione, prescindendo dai contenuti specifici con cui gli elementi dell'estensione possono essere esperiti da un particolare soggetto. In altre parole, si considerano solo le proprietà logico-formali. Per la relazione di somiglianza, che Carnap cerca di ottenere da quella di ricordo di similarità, sono proprietà logico-formali la riflessività e la simmetria.

2) Le relazioni strutturali – nelle intenzioni di Carnap – dovrebbero permettere di specificare classi di dati che siano controparti oggettive delle qualità soggettive. Tali classi possono essere considerate oggettivamente determinate solo se i loro elementi sono oggettivamente identificabili e a questo scopo, secondo Carnap, l'identificazione oggettiva dei loro elementi deve essere possibile solo per mezzo di descrizioni, intese nel senso di Russell, che affermino l'esistenza di un'unica entità che soddisfa una data condizione, e siano tali che la condizione non contenga alcun termine che si riferisca a un'entità in virtù della presentazione o ostensione dell'entità alla quale si riferisce. Descrizioni di questo genere sono chiamate da Carnap «puramente strutturali».

Naturalmente c'è una certa tensione tra l'esigenza di limitarsi a considerare solo le proprietà strutturali e ciononostante pervenire a identificare entità e classi di entità. Carnap si preoccupa innanzitutto di mostrare che l'identificazione strutturale è possibile e, poi, che ad essa si può sempre ricorrere, assumendo – in modo discutibile – che entità che non risultino strutturalmente distinguibili siano identiche. Carnap mostra con un esempio che identificazioni puramente strutturali sono possibili. L'esempio da lui dato è quello di una rete ferroviaria le cui stazioni sono identificabili in virtù delle proprietà strutturali della rete. Naturalmente si può obiettare che stazioni diverse possono essere connesse allo stesso modo, ma Carnap sembra aver presente la difficoltà, o almeno la possibilità che nel dominio di riferimento ci siano oggetti non identificabili sulla base delle proprietà strutturali considerate. Non potendo dimostrare che, per qualunque dominio, tutti gli elementi del dominio siano strutturalmente identificabili, Carnap suggerisce come si potrebbe procedere per cercare di ottenere un dominio nel quale tutti gli elementi lo siano, e anche come reagire di fronte all'eventuale impossibilità di una differenziazione strutturale:

Se [una tale descrizione definita] non è ancora realizzabile in modo univoco, il campo deve essere allargato oppure si devono prendere in ausilio altre relazioni. Se, nonostante l'utilizzazione di tutti i rapporti scientifici disponibili, non è rilevabile alcuna differenza tra due determinati oggetti di un campo di oggetti, allora essi sono per la scienza completamente uguali, anche se soggettivamente essi possono venir riguardati come diversi. (Si accenni soltanto che, qualora le condizioni indicate fossero completamente soddisfatte, i due oggetti non solo dovrebbero essere considerati come uguali, ma come identici in senso rigoroso; ma non ci si può qui addentrare nella giustificazione di questa affermazione dall'apparenza paradossale.) Il risultato è dunque che *[una descrizione definita mediante asserti] puramente strutturali è possibile in generale per quel tanto che è possibile una distinzione scientifica in generale*: per due oggetti [una tale descrizione] non può essere effettuata solo se essi non sono affatto distinguibili con mezzi scientifici³⁰.

Appare dunque evidente che l'identificazione strutturale degli oggetti è relativa a un dominio e non è sempre disponibile, ma Carnap ritiene che, quando non lo sia, si possa estendere il dominio o prendere in considerazione ulteriori relazioni in modo da renderla possibile.

Pertanto – egli prosegue – viene a cadere la necessità di indicare in ogni asserto scientifico il campo di oggetti, e noi otteniamo come risultato che *ogni asserto scientifico può essere per principio trasformato in modo tale che esso non venga ad essere altro che un asserto strutturale*. Ma questa trasformazione non solo è possibile, bensì è da esigersi. Giacché la scienza intende parlare dell'oggettivo; ma tutto ciò che non appartiene alla struttura, bensì al materiale, tutto ciò che viene [indicato mediante una concreta definizione ostensiva], è da ultimo un che di soggettivo. Questo svincolamento d[al] soggettivo è facilmente riscontrabile nella *fisica*, che ha già ricondotto quasi tutti i concetti fisici a puri concetti strutturali³¹.

Detto kantianamente l'identificabilità strutturale degli oggetti, e quindi il superamento della soggettività della base di partenza, è una condizione della possibilità della scienza che Carnap, come Kant, dà per scontata. Chiudiamo questa sezione con il suggerimento di un modo nel quale il contrasto tra Mach e Carnap riguardante la concezione dei dati può essere attenuato, se non addirittura superato, quando si consideri come Mach spiega la distinzione tra oggettivo e soggettivo e cosa Carnap ritiene rilevante nella costituzione della conoscenza scientifica.

Secondo Mach vi sono elementi costitutivi ultimi dell'esperienza, che di per sé non sono qualificabili né come soggettivi né come oggettivi e dai quali qualunque fatto, fisico o psichico, è costituito. La differenza che è usuale fare

30. Ivi, p. 134, sez. 15.

31. Ivi, p. 136, sez. 15.

tra oggettivo e soggettivo è da lui spiegata come una differenza di prospettiva. Così Mach si esprime in un passo già citato:

La grande distanza che sussiste fra la ricerca fisica e quella psicologica è [...] tale solo per il comune modo stereotipato di considerare le cose. Un *colore* diventa un *oggetto fisico* non appena consideriamo ad esempio la sua dipendenza dalla sorgente di luce che illumina (altri colori, calore, spazi ecc.) Se invece poniamo mente alla sua *dipendenza dalla retina* (dagli elementi KLM...) esso diventa un *oggetto psicologico*, una *sensazione*. Non l'*oggetto* bensì l'*orientamento della ricerca* è diverso nei due campi³².

Carnap, che all'epoca dell'*Aufbau* è convinto della soggettività dei dati iniziali disponibili ad un soggetto, è tuttavia interessato al problema di come a partire da essi si possa pervenire al livello di oggettività richiesto dalla scienza. Abbiamo visto come in questo percorso un ruolo decisivo sia svolto dalle proprietà strutturali. Secondo Carnap, l'oggettività della scienza richiede che questa sia essenzialmente conoscenza di proprietà strutturali, quindi oggettive, anche quando riguardano dati soggettivi.

Nella concezione di Mach i dati non sono soggettivi per loro natura, ma lo sono in virtù della connessione con un apparato che elabora informazione esterna. È in virtù della prospettiva dell'apparato ricevente che sono considerati soggettivi. Ma esiste anche una prospettiva, data dalla connessione con la fonte, che li rende oggettivi. La differenza di prospettiva è una differenza di proprietà relazionali: la soggettività dell'elemento e per l'apparato ricevente x è data dalle relazioni tra e e x e l'oggettività dell'elemento e con fonte f è data dalle relazioni tra e e f . Questi diversi insiemi di relazioni potrebbero essere intesi come proprietà strutturali diverse, ciascuna delle quali, nelle circostanze e con le specificazioni opportune, è tale da identificare la stessa entità. Supponendo che una tale identificazione possa essere resa puramente strutturale e prescindendo dalla questione della atomicità dei dati iniziali, rispetto alla quale avevano idee diverse, un certo parziale accordo sembra dunque possibile tra Mach e Carnap.

Riassumendo, benché la soggettività dei dati sia affermata e motivata da Carnap, mentre è negata da Mach, non solo risulta evidente che Carnap non procede a sviluppare una analisi fenomenologica alternativa a quella di Mach e la sua preoccupazione di rendere la soggettività dei dati compatibile con la costituzione di una scienza oggettiva finisce per ridimensionarne il ruolo. Appare anche possibile trovare un punto di contatto tra le due concezioni, poiché il modo di Mach di spiegare l'attribuzione della soggettività ai dati può essere inteso e chiarito mediante l'idea dell'identificazione struttu-

32. Mach, *Die Analyse*, p. 14; trad. it., p. 48.

rale che Carnap ritiene che debba essere applicata anche ai dati allo scopo – si potrebbe dire – di neutralizzare la loro soggettività.

4. Uno sguardo oltre Mach e Carnap

Nelle sezioni 153-155 dell'*Aufbau* Carnap solleva il problema della identificazione strutturale delle stesse relazioni fondamentali e in particolare di quella di ricordo di similarità. L'analisi del problema e le indicazioni che egli dà per giustificare una risposta positiva non sono chiare e suscitano qualche perplessità riguardo all'esito a cui *prima facie* sembrano condurre. Per queste ragioni, relativamente agli scopi di questa presentazione, possiamo – ed è opportuno – trascurare l'analisi piuttosto confusa e le incerte indicazioni di Carnap.

Tuttavia non è inopportuna una riflessione sul significato del problema che Carnap sembra sollevare. Cosa significa identificare strutturalmente una relazione strutturale? Se ciò vuol dire affermare l'esistenza e l'unicità di una relazione che soddisfi certe condizioni formali, l'identificazione strutturale non risulta quasi mai possibile. Per la maggior parte dei domini, la maggior parte delle condizioni puramente formali sono soddisfatte da molte relazioni. Condizioni formulate usando una logica del secondo ordine, che ammette la quantificazione di variabili predicative, possono permettere l'identificazione di certe relazioni a meno di isomorfismo nel senso che esse possono essere unicamente soddisfatte in domini diversi isomorficamente strutturati. Ma l'isomorfismo non implica l'identità e ciò induce a concludere che il mondo che è oggetto della scienza non possa essere caratterizzato in modo unico sulla base di proprietà puramente strutturali. In particolare le relazioni che interessano la scienza non sono isolabili sulla base di caratteristiche puramente formali senza il ricorso a ulteriore informazione, anche se la natura di tale ulteriore informazione è difficile da stabilire.

Nei suoi termini essenziali questa difficoltà è stata sollevata dal matematico Maxwell Herman Alexander Newman nei confronti della concezione che Bertrand Russell ha esposto in *The analysis of Matter*³³. Senza fare riferimento alla concezione più articolata ma simile di Carnap, Bastiaan Cornelis van Fraassen, noto come Bas van Fraassen, la riprende in uno scritto del 1997³⁴, ponendola in relazione con un argomento contro il realismo presentato da Putnam e con lo strutturalismo semantico secondo il quale una teoria è iden-

33. Bertrand Russell, *The analysis of Matter*, London 1927.

34. Bas Cornelis van Fraassen, *Structure and perspective: philosophical perplexity and paradox*, pp. 511-530, in Maria Luisa Dalla Chiara et al. (eds.), *Logic and Scientific Methods*, vol. 1, Dordrecht 1997.

tificata con una famiglia di modelli. Secondo van Fraassen la difficoltà non può essere superata senza tenere conto che nella scienza i modelli sono intesi come modelli di qualcosa e il modo effettivo in cui lo sono è rilevante: «the actual way matters and fixes the relevant relation between model and nature – relevant, that is, to the evaluation as well as application of that theory»³⁵. Inoltre la natura è confrontata con un modello della teoria via un modello di dati: «To be matched are two models, a data model and a theoretical model»³⁶. Per completezza, tra le cose che dovrebbero essere approfondite e discusse, aggiungiamo che Van Fraassen sostiene anche la tesi, da alcuni ritenuta problematica, che un modello di dati, in quanto rappresentazione, ha condizioni di verità oggettive, indipendenti da cosa e come conosciamo.

La questione sollevata può essere più esplicitamente e direttamente connessa con i punti di vista di Mach e di Carnap. Entrambi danno più importanza alle relazioni tra dati che ai dati in se stessi. Carnap lo fa per quanto riguarda la loro identificazione, ma la tesi generale che si possa fare in modo che le entità di un dominio siano tutte identificate strutturalmente è problematica e ancora di più lo è la tesi, che pure egli sembra sostenere, che le relazioni stesse tra le entità del dominio siano identificabili strutturalmente. Le due difficoltà possono essere riassunte in termini non carnapiani, dicendo che l'identificazione strutturale è relativa a un modello che non è strutturalmente identificabile: al più il modello è unico a meno di isomorfismo e quindi non è unico. Ciò non sembra compatibile con quanto Carnap richiede alla scienza, se la richiesta è del tutto generale e non limitata a casi particolari:

una scienza della realtà deve pur sapere se essa parla di persone o di villaggi. Il punto decisivo è qui: *la scienza della realtà deve certamente poter distinguere [queste varie entità]*, e [inizialmente] essa fa questo [per lo più mediante descrizioni definite che utilizzano altre entità, *ma alla fine le descrizioni definite sono formulate esclusivamente mediante descrizioni strutturali*]³⁷.

L'esigenza così espressa non può essere soddisfatta nel modo da lui indicato, cioè via descrizioni strutturali che identifichino tutte le cose delle quali parla la scienza. Ne segue che se carnapianamente si intende la prospettività, nella quale Mach fa consistere la soggettività di una esperienza individuale, come data da relazioni che strutturalmente identificano le entità coinvolte, abbiamo che in domini diversi tali relazioni possono essere soddisfatte da entità diverse. L'aggancio ad uno specifico mondo e in particolare a uno specifico soggetto, non dipende solo dalla rappresentazione strutturale basata su

35. Ivi, p. 523.

36. Ivi, p. 524.

37. Carnap, *Aufbau*, p. 129, sez. 12.

tali relazioni. Come suggerito da Van Fraassen³⁸ nel 2007, la comunicazione di come il soggetto *a* prende una data direzione in un dato luogo non permette ad *a* di orientarsi nel modo descritto se *a* non si identifica con il soggetto presente nel luogo al tempo della comunicazione. In uno scritto precedente Van Fraassen dice analogamente: «a theory says nothing to us unless we can locate ourselves, in our own language, with respect to its content»³⁹. Secondo Van Fraassen c'è bisogno, a questo scopo, di qualcosa di indessicale che è coinvolto nella stessa rappresentazione dei dati.

Come conclusione generale, senza addentrarci in dettagli e approfondimenti, possiamo dire che le analisi di Mach e Carnap non risolvono il problema della connessione tra la nostra esperienza e la scienza. Tuttavia esse permettono di comprenderlo un po' meglio. In particolare, per quanto riguarda il tema di questo volume, mostrano che c'è almeno un aspetto della rappresentazione dei dati rispetto al quale la loro dimensione quantitativa non appare rilevante.

Abstract

The sensations (Empfindungen), according to Ernst Mach, and the elementary experiences (Elementarerlebnisse), according to Rudolf Carnap (1928), are the primary and direct sources of the data that science takes into consideration and in various ways represents. Compared to Mach, Carnap has a more complex conception of what is involved in the transition from the level of perceptual experience to the objective level of scientific knowledge. However, Mach and Carnap do not seem so far away, if we look, on the one hand, at Mach's way of explaining the attribution of subjectivity to sensations and, on the other hand, at Carnap's need for a structural identification of the data of experience that can be placed at the basis of the construction of science. The aspect under which they could converge allows us to re-examine and deepen the problem of the connection between experience and science.

38. Bas Cornelis van Fraassen, *Structuralism(s) About Science: Some Common Problems*, «Proceedings of the Aristotelian Society Supplementary», LXXXI (2007), pp. 45-61.

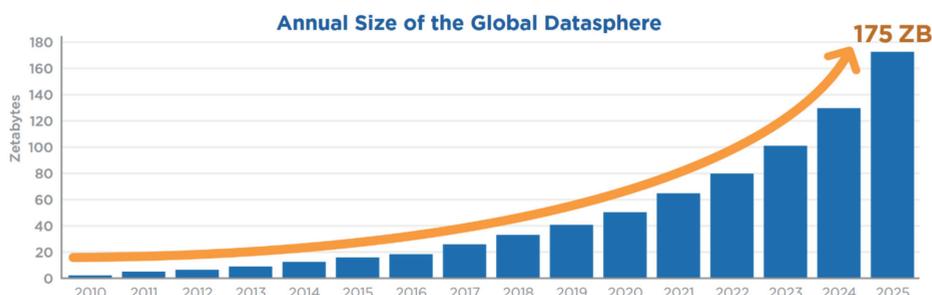
39. Bas Cornelis van Fraassen, *Weyl's Paradox: the Distance between Structure and Perspective*, p. 34, in Andreas Berg-Hildebrand and Christian Suhm (eds.), *Bas Cornelis van Fraassen: The Fortunes of Empiricism*, Frankfurt 2005, pp. 13-34.

Big Data. Contesto Tecnologico

di Andrea Pietracaprina, Geppino Pucci*

1. Quantificazione e caratterizzazione del fenomeno dei Big Data

Il termine *Global DataSphere* è stato introdotto recentemente per descrivere il complesso dei dati che sono disponibili in formato digitale, a livello mondiale. La *International Data Corporation (IDC)*, società specializzata in ricerche di mercato e consulenze nel settore dell'*Information and Communication Technology (ICT)*, porta avanti da diversi anni uno studio sulla crescita della *Global DataSphere*. Nel 2018 la IDC ha rilasciato un *white paper*¹ in collaborazione con l'azienda *Seagate* (azienda leader nel settore della produzione di supporti di memorizzazione), in cui, tra le altre cose, si stima la crescita della quantità di dati prodotta ogni anno nel mondo sino al 2025. Tale stima è riportata nel grafico seguente, estratto dal white paper:



Facciamoci un'idea concreta delle quantità in gioco. Uno ZettaByte (ZB) equivale a mille miliardi di GigaByte (GB), unità di misura più nota. Ad

* Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Padova.

1. David Reinsel, John Gantz, John Rydning, *The Digitization of the Worlds*. White paper IDC-Seagate, November 2018.

esempio, dal grafico si evince che la stima per il 2020 è di circa 50 ZB. Se tale quantità di dati fosse trasferita su dischi da un TeraByte (TB), ovvero mille GB, considerati oggi abbastanza capienti, e si riempisse un disco al giorno, sarebbero necessari 137 milioni di anni per trasferire tutti i dati! Un'altra quantificazione efficace era stata fatta in precedenza da Eric Schmidt nel 2010, quando era *Chief Executive Officer* (CEO) di Google, che aveva stimato che i dati digitali generati in due giorni sulla terra, a quel tempo, equivalevano in quantità a tutta l'informazione (opportunamente digitalizzata) prodotta dalla comparsa della specie umana sino al 2003.

La crescita straordinaria della quantità di dati digitali che si è osservata negli ultimi due decenni, e che si stima in ulteriore aumento per i prossimi anni, è il frutto di una combinazione di diversi fattori. Da un lato, l'avanzamento tecnologico, che si traduce in un continuo aumento della potenza di calcolo, della capacità di memorizzazione e della banda di comunicazione disponibili per le applicazioni, accompagnato da una riduzione drastica dei costi dell'ICT. Dall'altro, la pervasività delle tecnologie informatiche e di comunicazione nella quasi totalità delle attività umane. Si pensi, ad esempio, all'uso massiccio delle *cloud* come depositi di dati a cui si è assistito negli ultimi anni. Il white paper citato in precedenza prevede che nei prossimi anni più del 50% dei dati digitali sarà memorizzato nelle cloud, mentre si assisterà a una progressiva riduzione del numero e della capacità dei data center tradizionali legati a singole realtà aziendali e a un minore ricorso a dispositivi di memorizzazione individuali (fenomeno, peraltro, già in atto).

2. I Big Data e le sfide computazionali che essi pongono

Grandi moli di dati vengono prodotte in una varietà molto ampia di contesti, e diversi sono gli ambiti applicativi dove l'analisi e lo sfruttamento di tali dati è in grado, potenzialmente, di dare un netto vantaggio competitivo. Per fare alcuni esempi, non certo esaustivi, si pensi all'uso di dati nell'ambito del commercio elettronico, per dare suggerimenti mirati ai (potenziali) clienti; in ambito bancario, per la scoperta di frodi e per la valutazione del rischio; in ambito finanziario, per migliorare l'accuratezza delle previsioni; in ambito scientifico, per la scoperta e la validazione di nuovi risultati; in ambito medico, per la ricerca delle cause di specifiche patologie e per la previsione dell'impatto di pandemie (tema quanto mai attuale), o, infine, in ambito sociologico, per l'analisi del ruolo dei *social media* nella società.

La locuzione Big Data è diventata estremamente popolare in questi ultimi anni, ma viene usata per denotare diversi aspetti legati alla produzione di grandi quantità di dati e al loro utilizzo nei diversi ambiti applicativi. È opportuno allora fare un po' di chiarezza, anche per definire meglio lo scopo

del presente contributo. Quando si parla di Big Data ci si riferisce di solito a insiemi di dati che contengono informazioni potenzialmente utili, la cui estrazione, però, non può essere fatta “manualmente” da un essere umano, a causa della taglia dei dati in ingresso e della complessità delle operazioni da svolgere. I problemi che possono emergere relativamente all'utilizzo dei Big Data sono di varia natura ed è opportuno distinguerli.

In alcuni casi, per quanto la taglia dei dati sia abbastanza grande da rendere impossibile un'analisi manuale, essa è gestibile facilmente anche da un semplice PC di potenza limitata. In tali casi, i problemi principali da affrontare non riguardano gli aspetti computazionali ma, piuttosto, la scelta degli strumenti di analisi più efficaci, l'eventuale sviluppo di strumenti ad hoc, la scelta e la preparazione dei dati da fornire come input dell'analisi, e l'accertamento della significatività delle informazioni ottenute in relazione agli obiettivi dell'analisi.

In altri casi invece, anche se il tipo di analisi da effettuare sui dati è chiara, la mole massiva dei dati da analizzare, la tipologia stessa dei dati, il protocollo con il quale i dati vengono forniti in input, ed eventuali requisiti stringenti sui tempi di risposta, rendono il calcolo particolarmente impegnativo, tanto da richiedere spesso un radicale cambio di paradigma rispetto alle metodologie di calcolo tradizionali. Le sfide computazionali che emergono in questo contesto sono efficacemente descritte dalle cosiddette “4 V”: *Volume*, *Velocità*, *Varietà* e *Veridicità*. Il Volume fa riferimento alla enorme quantità di dati da elaborare, che richiede algoritmi efficienti e sistemi di calcolo potenti e con grandi quantità di memoria. La Velocità fa riferimento alla frequenza elevata e, a volte, irregolare, con la quale i dati da analizzare vengono generati, dando origine a un flusso continuo (potenzialmente illimitato), per cui diventa impossibile il loro salvataggio per una successiva elaborazione *off-line*, soprattutto se si richiedono tempi di risposta brevi. Si pensi, ad esempio, all'analisi di *tweet* (più di 500 milioni di *tweet* sono generati ogni giorno) o di dati forniti da potenti telescopi (si stima che il radiotelescopio SKA, in fase di progetto, raccolga diversi TeraByte di dati al secondo). La Varietà fa riferimento alle diverse tipologie dei dati (ad es., dati strutturati provenienti da database, o non strutturati come file di testo o immagini) che spesso coesistono nell'input di specifiche analisi. Ma essa fa anche riferimento al fatto che per lo stesso tipo di analisi, strumenti diversi possono essere richiesti a seconda di specifiche caratteristiche dei dati, come ad esempio la loro dimensionalità. Infine, la Veridicità fa riferimento alla possibilità, concreta per molte applicazioni reali, che i dati da analizzare siano rumorosi o non del tutto affidabili, richiedendo perciò un maggiore sforzo computazionale per rendere i risultati dell'analisi robusti e significativi.

Diamo un esempio concreto per illustrare la necessità di nuovi approcci quando si elaborano grandi moli di dati. Nel 1929 lo scrittore ungherese

Frigyes Karinthy aveva formulato, nel suo racconto “Catene”, la teoria, diventata poi celebre, dei *sei gradi di separazione*, secondo la quale qualsiasi coppia di persone nel mondo è collegata da una catena di al più sei amicizie, come illustrato nella seguente figura.



La congettura di Karinthy ha stimolato diversi studi ed esperimenti che hanno cercato di verificarne la validità. Nel 2011 alcuni informatici dell’Università di Milano hanno deciso di misurare sperimentalmente i gradi di separazione nella rete di amicizie definita da Facebook che, al tempo, conteneva circa 700 milioni di profili utente (detti *nodi* della rete) e 70 miliardi di amicizie tra coppie di profili (detti *archi* della rete). Essi hanno riscontrato un grado di separazione medio di 3,74 e quindi decisamente più basso di quello ipotizzato da Karinthy². Il calcolo del grado di separazione medio in una rete può essere fatto tramite algoritmi standard che vengono insegnati nei primi anni dei corsi universitari in Informatica. Tuttavia, l’esecuzione di tali algoritmi richiede un numero di operazioni almeno proporzionale al prodotto del numero di nodi e del numero di archi. Se per reti di piccole dimensioni, il calcolo è eseguibile in poco tempo su un PC di fascia bassa, al crescere del numero di nodi e di archi, la complessità del calcolo diventa rapidamente non affrontabile in tempi ragionevoli, anche su piattaforme potenti. Ad esempio, determinare il grado di separazione medio con gli algoritmi standard sulla rete di Facebook attuale, dove il numero di profili è dell’ordine dei miliardi e il numero di amicizie è dell’ordine delle centinaia di miliardi, richiederebbe centinaia di anni di calcolo su un PC e risulterebbe comunque estremamente impegnativo anche per i più potenti sistemi di elaborazione al mondo.

Ci troviamo quindi davanti a un tipo di analisi che algoritmi tradizionali sono in grado di effettuare senza alcun problema quando la mole di dati è contenuta, ma che si trasforma in una sfida computazionale formidabile

2. Paolo Boldi, Marco Rosa and Sebastiano Vigna, *Four Degrees of Separation, Really*. Proc. of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2012.

per quantità di dati massive. In questi casi, si ricorre a un compromesso tra complessità computazionale e accuratezza del calcolo: si rinuncia da avere una risposta esatta (nell'esempio, il grado di separazione medio esatto) e ci si accontenta di un'approssimazione, accompagnata da una previsione sull'errore commesso, per rendere ragionevoli i tempi di risposta. Lo studio sopra citato ha utilizzato questo approccio, riuscendo a ottenere, su un computer di medio-bassa potenza, una stima molto accurata del grado di separazione medio di Facebook nel 2011, con un tempo di calcolo che è ordini di grandezza inferiore rispetto a quello richiesto dal calcolo esatto.

3. Piattaforme di elaborazione

Per rispondere alle sfide poste dai Big Data, è necessario disporre di piattaforme adatte alla loro memorizzazione ed elaborazione. Lo spettro di soluzioni architetturali adottate copre una vasta gamma di scenari, caratterizzati da diversi rapporti tra costi e prestazioni. A una estremità dello spettro si collocano *sistemi di elaborazione convenzionali* dotati di una memoria secondaria (*memoria esterna*) molto capiente, che costituiscono soluzioni di basso costo ma prestazioni limitate. Il ruolo della memoria esterna nella computazione diventa importante in quanto, se il volume dei dati da elaborare è molto elevato, non è possibile utilizzare l'approccio degli algoritmi tradizionali (*in-core*) caricando *tutti i dati di ingresso* contemporaneamente nella memoria centrale dell'elaboratore (*core memory*), la cui capienza potrebbe facilmente risultare insufficiente, ma è necessario mantenere tali dati nella memoria esterna durante l'elaborazione e sviluppare tecniche algoritmiche che operino su essi *a blocchi*, caricando i blocchi progressivamente, quando servono, nella core memory. In questo caso, è di primaria importanza organizzare la computazione in modo da mascherare il più possibile l'alta *latenza*, ovvero il tempo necessario a trasferire un blocco di dati dalla memoria esterna alla core memory, un tempo che può essere di diversi ordini di grandezza più elevato rispetto al tempo di accesso a un dato residente in core memory. Per ammortizzare la latenza associata a tale trasferimento, è necessario cercare di massimizzare il numero di operazioni utili da eseguire su un dato blocco prima che il prossimo blocco venga caricato nella core memory per proseguire l'elaborazione.

All'altra estremità dello spettro si trovano i *sistemi di supercalcolo (supercomputer)* in cui numerosi processori (da qualche centinaio sino a milioni) sono integrati in un unico sistema di calcolo tramite una rete di interconnessione ad altissime prestazioni. La potenza necessaria ad alimentare un supercomputer è enorme (nell'ordine dei milioni di watt) e richiede spesso la costruzione di centrali elettriche dedicate. Tali sistemi sono molto costosi,

caratterizzati da un alto tasso di obsolescenza e la loro utilizzazione efficiente richiede abilità di programmazione estremamente specializzate. Per tali motivi i supercomputer vengono prevalentemente utilizzati per pochi problemi fondamentali della scienza e dell'ingegneria ad alto impatto economico (le cosiddette *grand challenge*), le cui soluzioni richiedono enormi potenze computazionali. Tra le *grand challenge*, ricordiamo la quantocromodinamica (computazioni intensive di fisica delle alte energie), la chimica computazionale per il progetto di nuove molecole, la modellizzazione del clima globale, la simulazione del cervello umano e la cosmologia.

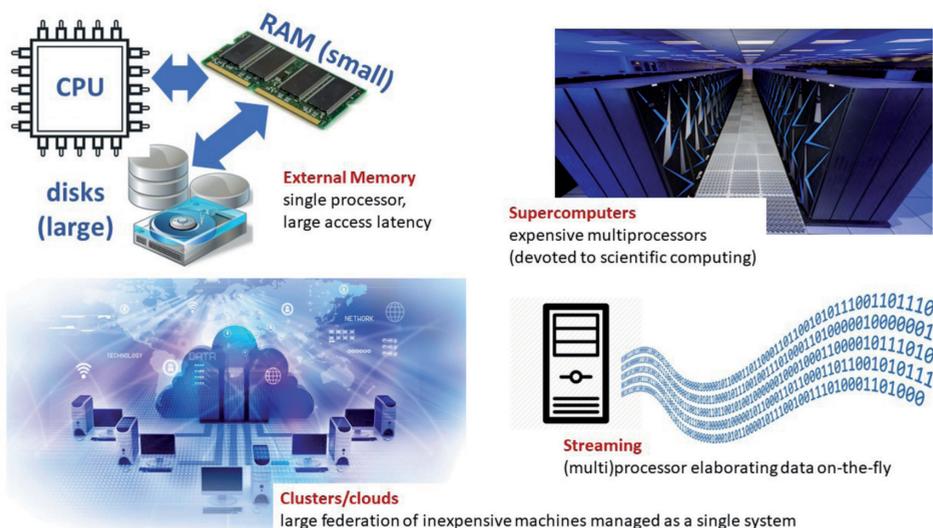
I sistemi di elaborazione forse più adeguati alla gestione e all'elaborazione dei Big Data si trovano a metà dello spettro: la varietà degli scenari applicativi e la pervasività dei Big Data richiedono infatti infrastrutture sufficientemente potenti ma di costo moderato. Queste architetture prendono solitamente la forma di *cluster* di componenti di largo impiego, ad esempio, processori standard interconnessi da reti locali commerciali ad alta banda. Un passo rivoluzionario verso la creazione di un ecosistema di calcolo in cui la gestione dell'infrastruttura viene totalmente sottratta all'utente è la *virtualizzazione* offerta dal paradigma di *cloud computing*. Una cloud è una grande federazione di macchine (solitamente economiche) che vengono gestite come un singolo sistema. Gli utenti della cloud possono “noleggiare” segmenti dell'infrastruttura, dimensionati alle loro esigenze. La virtualizzazione permette a tutti gli utenti di interagire con il loro segmento come se fosse una macchina isolata e dedicata.

L'offerta di servizi cloud forniti dai *provider* (ad esempio, i ben noti AWS di Amazon e Azure di Microsoft) può essere di varia natura. Nel servizio *IaaS* (*Infrastructure-as-a-Service*) il provider fornisce ai clienti accesso con metodo di pagamento a consumo (*PAYG - Pay-As-You-Go*) alla memoria, alle connessioni di rete, ai server e ad altre risorse di calcolo nel cloud. Nel servizio *PaaS* (*Platform-as-a-Service*) il provider offre l'accesso a un ambiente completo basato su cloud, in cui gli utenti possono sviluppare, distribuire ed eseguire applicazioni. È previsto infine il servizio *SaaS* (*Software-as-a-Service*), in cui il provider fornisce software e applicazioni già sviluppate. Gli utenti si abbonano al software e accedono a esso via internet, tramite il web o con specifiche interfacce.

Come si è osservato nel paragrafo precedente, oltre al Volume, un altro importante attributo dei Big Data che costituisce una sfida computazionale e può influenzare l'architettura del sistema di elaborazione, è la Velocità. In presenza di flussi di dati da analizzare in *tempo reale*, è richiesta la capacità di elaborare ogni elemento del flusso in maniera sufficientemente veloce per consentire l'allineamento tra la generazione dei dati e l'attività di analisi. In tale contesto, l'elemento cruciale, oltre a un'infrastruttura di calcolo reattiva (processore singolo o multiprocessore), è la disponibilità di algoritmi di

analisi efficienti che siano in grado di estrarre informazioni dal flusso di dati entranti molto velocemente.

La figura seguente riassume le diverse soluzioni architetturali descritte sopra, che coprono un ampio spettro di rapporti tra costo e prestazioni.



4. Framework computazionali

Come descritto nel paragrafo precedente, i sistemi di calcolo più comunemente utilizzati per l'elaborazione di Big Data sono cluster di computer gestiti *in-house* o noleggiati attraverso fornitori di servizi cloud. Il vantaggio di questi sistemi è, da un lato, il costo relativamente moderato, che amplia enormemente la platea dei possibili fruitori e, dall'altro, una capacità di memoria e una potenza di calcolo aggregata che, per quanto lontane da quelle raggiunte dai sofisticati sistemi di supercalcolo, possono superare di gran lunga quelle di singoli server e sono sufficienti per molte delle applicazioni. Tuttavia, lo sviluppo di software che sfrutti appieno le potenzialità di un cluster, soprattutto se non si ha un controllo diretto dell'infrastruttura come nel caso del noleggio tramite cloud, richiede competenze di programmazione molto sofisticate. Ad esempio, aspetti importanti come la distribuzione dei dati e delle attività (*task*) da svolgere tra le diverse unità di calcolo, il bilanciamento del loro carico di lavoro, la gestione dei guasti, la cui frequenza cresce all'aumentare delle componenti del sistema, se non sono gestiti efficientemente, rischiano di ridurre considerevolmente le prestazioni del sistema, sottoutilizzandone la potenza.

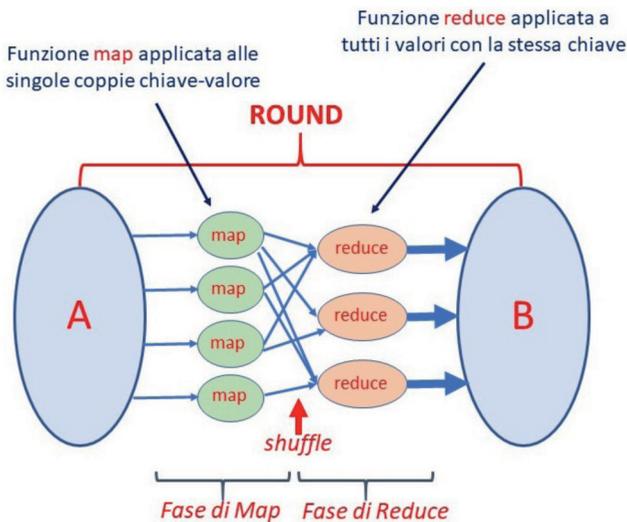
Per consentire lo sviluppo di applicazioni che possano elaborare efficientemente grandi quantità di dati senza però richiedere un'eccessiva complessità di programmazione, sono stati creati dei *framework computazionali*, ovvero dei modelli per l'organizzazione del calcolo e per la programmazione, orientati specificamente all'elaborazione di Big Data. Tali framework si basano sull'utilizzo di uno strato di software intermedio e di apposite interfacce di programmazione (*Application Programming Interface* o *API*) che permettano al programmatore di concentrarsi sulle operazioni da fare sui dati, sollevandolo dalla gestione di tutti gli aspetti di più basso livello menzionati prima, importanti per assicurare elevate prestazioni, dei quali si occuperà lo strato software intermedio in modo trasparente al programmatore, cercando di sfruttare al meglio le potenzialità della infrastruttura hardware sottostante. Così facendo si perde la possibilità da parte del programmatore di un controllo fine sulle prestazioni del software, semplificandone però enormemente lo sviluppo. Dato che molte delle applicazioni per Big Data richiedono prestazioni elevate ma non così estreme come quelle richieste dalle grand challenge del supercalcolo menzionate in precedenza, questa soluzione rappresenta il giusto compromesso tra prestazioni e costi di sviluppo.

MapReduce. L'esempio più noto di framework computazionale per l'elaborazione di Big Data è rappresentato da *MapReduce*, introdotto da Google nel 2004³. L'obiettivo di MapReduce è quello di supportare lo sviluppo di applicazioni per l'elaborazione distribuita di grandi moli di dati su cluster di computer. La formulazione originale del framework contiene la descrizione di un modello di programmazione per la scrittura del software e la descrizione di una possibile implementazione dello strato di software di supporto all'esecuzione dei programmi. La caratteristica principale di MapReduce è la sua *visione data-centrica*, dove il programma specifica una sequenza di trasformazioni da applicare a partire dai dati di input, senza però esplicitare come i dati e le operazioni da eseguire vengano distribuite tra le unità di calcolo e senza prevedere esplicite azioni di ripristino nel caso di malfunzionamenti o guasti di singole componenti: la gestione dei guasti è quindi del tutto trasparente e il programmatore opera su un'astrazione computazionale libera da guasti (*fault-free*).

Vediamo il framework MapReduce più in dettaglio. I dati da elaborare (e quelli eventualmente generati durante l'elaborazione) sono rappresentati da *coppie chiave-valore*. Le chiavi costituiscono, da un lato, una sorta di identificatori, benché non univoci, per i singoli dati, e dall'altro permettono di partizionare il dataset da elaborare in segmenti più piccoli sui quali si opera

3. Una descrizione dettagliata del framework originale e della sua implementazione si trova nell'articolo: Jeffrey Dean and Sanjay Ghemawat, *MapReduce: Simplified Data Processing on Large Clusters*, «Communications of the ACM», 51(1) (2008), pp. 107-113.

individualmente e in parallelo in modo da distribuire il carico di lavoro e ridurre la memoria richiesta a ciascuna unità di calcolo, che costituisce la tecnica fondamentale per l'elaborazione efficiente di Big Data. Una computazione MapReduce si articola in una sequenza di uno o più *round*. Ciascun round trasforma un insieme A di coppie chiave-valore in un altro insieme di B , sempre di coppie chiave-valore. La trasformazione avviene in due fasi: nella *Fase di Map* ciascuna coppia di A viene elaborata individualmente da una procedura definita dal programmatore, che chiameremo *map*, che crea come risultato zero, una o più nuove coppie chiave-valore, dette *coppie intermedie*; nella successiva *Fase di Reduce*, le coppie intermedie vengono raggruppate in base al valore della chiave e ciascun gruppo di valori viene elaborato individualmente da un'altra procedura definita dal programmatore, che chiameremo *reduce*, che crea come risultato zero, una o più nuove coppie chiave-valore. L'output del round, ovvero l'insieme B risultato della trasformazione, è costituito da tutte le coppie chiave-valore prodotte dalle Fase di Reduce. Quanto appena descritto è schematizzato nella seguente figura.



È importante sottolineare che nella Fase di Map viene eseguita la stessa procedura *map* su ciascuna coppia chiave-valore in A . Ogni istanza di *map* può essere eseguita in parallelo sulle diverse coppie, da diverse unità di calcolo. Analogamente, nella Fase di Reduce, la stessa procedura *reduce* viene eseguita su ciascun gruppo di valori con la stessa chiave, e può essere eseguita in parallelo sui diversi gruppi, da diverse unità di calcolo. Il dataset A è inizialmente suddiviso in blocchi memorizzati nelle varie unità di calcolo,

che quindi possono eseguire la procedura map sulle coppie chiave-valore appartenenti ai blocchi a loro assegnati senza bisogno di interagire tra loro. Tuttavia, nella successiva Fase di Reduce, il raggruppamento per chiave delle coppie intermedie richiede una riallocazione delle coppie tra le unità di calcolo, che è un'operazione di comunicazione potenzialmente onerosa chiamata *shuffle*.

Descriviamo un esempio semplice che illustra come realizzare tramite un round MapReduce una primitiva di pratica utilità. Si supponga di avere una collezione molto grande di documenti testuali, D_1, D_2, \dots, D_n , e di dover calcolare il numero totale di occorrenze di ciascuna parola nella collezione. Ad esempio, per la collezione:

D_1	mela, arancia, mela
D_2	melone, mela, banana, pesca
D_3	pesca, melone, banana
D_4	banana, pesca, mela

il calcolo darebbe il seguente risultato: (arancia, 1), (banana, 3), (mela, 4), (melone, 2), (pesca, 3). Conoscere la frequenza delle parole in una collezione di documenti è di primaria importanza per l'analisi dei testi e per la caratterizzazione dei loro contenuti. Se però la collezione contiene un numero di documenti molto elevato e/o documenti molto lunghi, è opportuno distribuire il conteggio su più unità di calcolo. In MapReduce il calcolo potrebbe essere eseguito in un round. Si consideri un documento come una coppia chiave-valore, dove la chiave è un identificatore del documento, e il valore è la sequenza di parole che compare nel documento. Nella Fase di Map, ciascun documento D viene elaborato individualmente producendo, per ogni parola w che occorre in D , la coppia intermedia (w,c) , dove c è il numero di occorrenze di w in D . Ad esempio, dal documento D_1 dell'esempio precedente si produrrebbero le coppie (mela,2) e (arancia,1). A questo punto, considerando w come chiave della coppia (w,c) , nella successiva Fase di Reduce, si raggruppano tutte le coppie relative alla stessa parola e si sommano il loro conteggi parziali. Ad esempio, nell'esempio precedente, per la parola "mela" la Fase di Reduce raccoglierebbe le coppie (mela,2), (mela,1) e (mela,1) prodotte dai documenti D_1, D_2 e D_4 , per generare la coppia di output (mela,4).

Nell'esecuzione dell'algoritmo appena descritto su un cluster, i documenti della collezione verrebbero distribuiti tra le varie macchine che compongono il cluster, che nelle Fase di Map lavorerebbero in parallelo per estrarre da ciascun documento le coppie chiave-valore intermedie. Analogamente, nella

Fase di Reduce, le coppie intermedie sarebbero distribuite tra i computer in modo da assegnare ogni gruppo di coppie relative alla stessa parola a una singola macchina che ne aggrega poi i conteggi parziali. In questo modo si sfrutterebbe appieno la potenza di calcolo del cluster, dato che le operazioni da svolgere sono distribuite tra le diverse macchine che possono lavorare in parallelo con una riduzione del tempo di esecuzione circa proporzionale al loro numero. Inoltre, la capacità di memoria richiesta nelle singole macchine del cluster è considerevolmente inferiore a quella necessaria per memorizzare tutta la collezione.

La semplicità d'uso del modello di programmazione MapReduce, che è alla base della sua popolarità, risiede nel fatto che si richiede al programmatore solo di concentrarsi sull'organizzazione della computazione in round, specificando, per ciascun round, come suddividere i dati in segmenti sufficientemente piccoli da elaborare in parallelo tramite le opportune funzioni *map* e *reduce* da applicare. Per quanto questi elementi siano molto importanti per assicurare l'efficienza del calcolo e spesso necessitino di molta esperienza e conoscenze approfondite riguardo al progetto di algoritmi, il poter demandare allo strato di software sottostante la gestione delle risorse di calcolo disponibili e il ripristino da eventuali guasti semplifica enormemente il lavoro di programmazione, riducendone sensibilmente costi e tempi di sviluppo.

Hadoop e Spark. L'implementazione più nota di MapReduce è costituita da *Hadoop*⁴, una collezione di componenti software *open-source* sviluppata dalla *Apache Software Foundation* a partire dal 2006, che fornisce lo strato di software necessario per poter eseguire programmi sviluppati secondo il modello di programmazione MapReduce. Un elemento chiave di Hadoop è l'*Hadoop Distributed File System* (HDFS), che permette la gestione e la memorizzazione distribuita dei grandi file di dati che le applicazioni MapReduce devono elaborare. Tra le caratteristiche salienti dell'HDFS vanno evidenziate: la possibilità di utilizzare componenti hardware (dischi) di basso costo; un meccanismo efficace di rilevamento dei guasti e di ripristino dei dati perduti o corrotti; e un'elevata banda di accesso ai dati.

Per quanto riguarda i tempi di accesso ai dati, nonostante la banda elevata assicurata dall'HDFS, è bene tenere presente che il sistema si basa sull'utilizzo di dischi che sono dispositivi di memorizzazione molto più lenti dei componenti di memoria utilizzati all'interno dei processori, come RAM, cache e registri. Nella implementazione proposta inizialmente per il modello MapReduce, in ogni round di computazione, i dati di input e di output e i dati intermedi prodotti tra la Fase di Map e la Fase di Reduce dovevano essere memorizzati nell'HDFS. Ben presto però ci si è resi conto che tale approc-

4. <https://hadoop.apache.org>.

cio penalizzava pesantemente le prestazioni, proprio a causa dell'accesso ai dati sull'HDFS. Inoltre, il continuo ricorso alla memorizzazione sull'HDFS imponeva una sincronizzazione globale tra un round e il successivo, con il conseguente sottoutilizzo delle unità di calcolo con minor carico di lavoro in un round. Per far fronte a tali inefficienze sono state sviluppate versioni successive di Hadoop e ulteriori implementazioni che hanno dato origine a un vero e proprio *ecosistema* (l'*Hadoop ecosystem*) di framework software per Big Data. Tra le diverse varianti di Hadoop merita sicuramente attenzione il framework *Spark*⁵, sviluppato nel 2009 in ambito accademico, presso l'*AMPLab* della *University of California at Berkeley*, e successivamente donato alla Apache Software Foundation. Negli ultimi anni Spark ha avuto un'ampia diffusione anche in ambito aziendale ed è diventato uno dei framework di riferimento per lo sviluppo di applicazioni per Big Data. Per quanto Spark permetta l'implementazione di programmi scritti secondo il modello MapReduce, in realtà offre molto di più, consentendo un utilizzo più ottimizzato delle risorse di calcolo a disposizione. In particolare, per ottenere un'elevata efficienza computazionale, Spark cerca di mantenere i dati, il più possibile, nelle memorie principali delle unità di elaborazione, limitando i trasferimenti da e verso i dischi, e organizza l'esecuzione del calcolo in modo flessibile, per migliorare l'utilizzo della potenza aggregata offerta dalle diverse unità.

5. Tecniche algoritmiche

Nei paragrafi precedenti abbiamo argomentato che le problematiche poste dalla elaborazione dei Big Data, riassunte efficacemente dalle quattro "V", richiedono piattaforme e framework computazionali specifici. Serve tuttavia anche un cambio di paradigma nelle tecniche algoritmiche utilizzate per la loro analisi.

Le caratteristiche dei Big Data che inducono la maggiore differenziazione rispetto all'algoritmica tradizionale sono certamente Volume e Veridicità. Grandi collezioni di dati possono nascondere elevati livelli di replicazione, e contenere porzioni affette da rumore, incomplete o addirittura erranee. Per questo motivo, diventano poco significative le cosiddette analisi *esatte*, che sono pensate per dati strutturati, omogenei e coerenti. Un'analisi esatta prevede di determinare, per ogni insieme di dati di ingresso, una e una sola soluzione (il risultato dell'analisi) che, solitamente, soddisfa un qualche criterio di ottimalità definito sui dati. Tali analisi, che possono essere molto gravose per grandi volumi di dati, perdono senso in presenza di rumore, che

5. <https://spark.apache.org>.

rende poco significativo il concetto di soluzione unica. Per l'analisi di Big Data, diventano allora cruciali *algoritmi approssimati*, in cui viene rilassato il vincolo dell'ottimalità, al fine di rendere il calcolo più veloce, e *algoritmi a campione*, in cui l'analisi esatta non viene applicata all'intero insieme dei dati ma soltanto a un sottoinsieme piccolo ma "rappresentativo", ottenuto utilizzando approcci statistici e/o sofisticate metodologie di selezione (la cosiddetta *coreset selection*), che variano con il tipo di analisi da condurre. I due approcci possono essere utilizzati in maniera sinergica: è infatti spesso possibile ottenere risultati approssimati di qualità controllabile operando un'analisi a campione.

A titolo di esempio, illustreremo brevemente la semplice tecnica del *polling* (la tecnica utilizzata per le proiezioni elettorali). Si supponga di avere un insieme molto grande di N dati, appartenenti a K categorie distinte (ad esempio, l'insieme degli N voti espressi per K partiti politici a una elezione). Per valori contenuti di K , è possibile stimare con elevata accuratezza la percentuale di dati in ogni categoria (ovvero, la percentuale dei voti per ogni partito), senza elaborare l'intero insieme di N dati ma semplicemente estraendo un campione di taglia *esponenzialmente più piccola di N* e calcolando le percentuali su tale campione. Basta un campione di poche migliaia di voti per ottenere stime accurate per valori di N dell'ordine delle decine di milioni. La qualità di questi metodi dipende crucialmente dalla capacità di estrarre un campione che sia il più possibile rappresentativo dell'intero insieme.

6. Conclusioni

Le enormi quantità di dati digitali che vengono, ormai da anni, prodotte globalmente in un ampio spettro di contesti, racchiudono una straordinaria ricchezza di informazioni con un elevato potenziale per il progresso della società. Lo sfruttamento dei cosiddetti Big Data richiede, da un lato, strumenti di analisi efficaci e, dall'altro, piattaforme hardware e software e tecniche algoritmiche adeguate. Il presente contributo si è soffermato maggiormente su questo secondo aspetto. Si è argomentato che, considerata la pluralità di ambiti applicativi in cui emerge l'esigenza di analizzare grandi moli di dati, un buon compromesso tra costi e prestazioni può essere ottenuto utilizzando piattaforme hardware basate su cluster di macchine (eventualmente noleggate tramite servizi cloud) e framework computazionali che prevedano una gestione delle risorse della piattaforma di calcolo il più possibile ottimizzata e trasparente al programmatore, il cui compito è quello di pianificare le trasformazioni da operare sui dati, secondo una visione data-centrica. Dal punto di vista algoritmico, è necessario discostarsi da paradigmi tradizionali per poter far fronte all'elevato volume dei dati in ingresso, alla velocità con i

quali, in determinati contesti, essi possono essere generati, alla loro varietà e alla possibile presenza di rumore ed errori nei dati. L'attenzione della comunità scientifica internazionale e degli organismi finanziatori su questi temi è elevatissima. Inoltre, negli ultimi anni molte università stanno offrendo corsi e programmi di studio rilevanti per l'ambito dei Big Data, che dovrebbero consentire la formazione di figure professionali con le competenze necessarie per affrontare le sfide (computazionali e non) che emergono in questo ambito.

Abstract

In this essay, we provide a brief introduction to the Big Data phenomenon. We quantify the impressive growth observed (and projected) in the amount of digital data available world-wide, and point out the computational challenges that their analysis poses. Also, we outline the hardware-software platforms and algorithmic techniques which are most suited to efficiently process datasets characterized by high volumes, high velocity, high variability, and non-uniform veracity.

Il ruolo della statistica nell'era dei Big Data

di Bruno Scarpa*, Alessandra Salvan**

1. Statistica e statistiche

L'approccio culturale e i modi secondo cui la scienza statistica viene utilizzata per analizzare e interpretare i dati hanno subito una rapida evoluzione in anni recenti. La versione inglese di Wikipedia definisce la statistica come quel “ramo della matematica che si occupa di raccogliere, organizzare, analizzare, interpretare e presentare i dati” (Wikipedia, 2020). Sulla stessa linea anche diversi eminenti statistici contemporanei (traduzione ed enfasi sono nostre): Barnett (1999) la definisce “lo studio di come l'*informazione* dovrebbe essere utilizzata per comprendere situazioni concrete caratterizzate da *incertezza* e orientare le decisioni relative ad esse” e Brad Efron, secondo quanto riportato da Friedman (2001), parla di “*scienza dell'informazione* di maggior successo”. Nell'introduzione al libro *Computer Age Statistical Inference*, Efron e Hastie (2016) più semplicemente la definiscono come “la scienza di *apprendere dall'esperienza*”, e Bühlmann e van de Geer (2018) ricordano che “quando ci sono dati, è necessario anche un metodo statistico”. In definitiva, come afferma Senn (2003):

Statistica, al singolare, non ha nulla che vedere con la percezione comune che la descrive come una scienza che ‘riguarda i fatti’. Essa, invece, si occupa di come noi conosciamo o sospettiamo o crediamo che qualcosa sia un ‘fatto’ e, poiché la conoscenza delle cose richiede che esse vengano contate e misurate, allora è certamente vero che anche le statistiche, al plurale, sono parte di ciò di cui si occupa la scienza statistica, che è invece la *scienza del ragionamento quantitativo*. Una scienza che ha molto più in comune con la filosofia (in particolare con l'epistemologia) che con la contabilità. Gli statistici sono *filosofi applicati*. I filosofi discutono su quanti angeli possono danzare sulla capocchia di uno spillo, gli statistici li contano. O piuttosto,

* Dipartimento di Scienze Statistiche e Dipartimento di Matematica “Tullio Levi Civita”, Università degli Studi di Padova.

** Dipartimento di Scienze Statistiche, Università degli Studi di Padova.

contano quanti vi potrebbero *probabilmente* danzare [...]. Non si può prevedere nulla con certezza, ma si può prevedere quanto saranno incerte, in media, le nostre previsioni. La statistica è la scienza che dice come farlo.

2. Big Data

Questa scienza, che si basa quindi sui dati, deve ovviamente confrontarsi con la tipologia di informazione potenzialmente disponibile. La capacità di creare, raccogliere e memorizzare quantità enormi di dati caratterizza la materia grezza che la statistica oggi deve essere in grado di analizzare. Una delle conseguenze, infatti, dell'innovazione tecnologica consiste nell'assoluta facilità di disporre di masse di dati e di informazioni su svariati fenomeni di interesse. Tale semplicità è legata a due fattori: da una parte lo sviluppo di metodi automatici di rilevazione dei dati che rende sempre meno necessarie indagini campionarie, di per sé costose, e che permette di raccogliere, con modesti errori di rilevazione, ingenti quantità di informazione; dall'altra, la capacità di memorizzare a bassissimo costo quantità sempre più grandi di dati che diventano disponibili facilmente a chiunque (si veda, ad esempio, Azzalini e Scarpa, 2012, par. 1.1).

“Ogni volta che la quantità di dati aumenta di un fattore di dieci, dovremmo completamente ripensare come analizzarli”: così Friedman (1997) ribadisce che la scienza statistica è sempre chiamata a rinnovarsi e a rimanere al passo con le esigenze dei tempi. Brad Efron è anche più categorico nell'affermare che “chi ignora la statistica è condannato a reinventarla” (frase attribuita a Efron da Friedman, 2001, p. 6), richiamando come la maggioranza delle nuove metodologie che nascono nei diversi ambienti culturali, di fatto siano reinvenzioni della statistica, la scienza dei dati.

Pur non essendoci una definizione univoca e universalmente accettata per Big Data, vi sono alcuni aspetti su cui pare esserci accordo e che caratterizzano questi “nuovi” tipi di dati. Big Data sono certamente osservazioni disponibili in grandi quantità (*volume*) e salvate e immagazzinate in forme non strutturate o con una *varietà* di strutture. Spesso poi si tratta di dati generati ad una *velocità* molto elevata da strumenti automatici e da fonti di diverso tipo, come ad esempio i *social media*. Infine, vengono considerati *Big Data* quei dati che richiedono nuovi strumenti per essere immagazzinati, processati, analizzati, visualizzati e integrati.

Quando si è in presenza di Big Data, la cosiddetta matrice dei dati, dove le righe raccolgono le unità di osservazione e le colonne le variabili di interesse, può assumere diverse forme. Possono presentarsi dati detti “larghi”, in cui sono disponibili migliaia o milioni di variabili, ma molte meno unità statistiche (a volte dette ‘campioni’), ad esempio nell'ordine delle centinaia.

Esempi classici sono i problemi della genomica (si veda, ad esempio, Giovanni Parmigiani *et al.*, 2003) o dell'analisi dei testi scritti (ad esempio, Silge e Robinson, 2017). In questi casi si presenta la situazione in cui sono disponibili troppe variabili ed è forte il rischio di sovradattamento di qualsiasi modello ai dati disponibili. È quindi necessario pensare a metodi per selezionare le variabili rilevanti, o regolarizzare le procedure di adattamento dei modelli o attuare entrambe queste procedure contemporaneamente, ad esempio utilizzando strumenti di selezione delle variabili (si veda, ad esempio, Hastie *et al.*, 2015) e controllando la significatività di test d'ipotesi multiple (ad esempio, Goeman e Solari, 2014). Una seconda tipologia di dati sono quelli detti “alti”, ovvero in cui sono presenti decine o centinaia di variabili e milioni di unità. Spesso in questi casi, i modelli più semplici non hanno un buon adattamento, ma è disponibile un numero sufficiente di unità statistiche per adattare modelli non lineari anche complessi che possono prevedere molte interazioni. Molti metodi automatici sono stati sviluppati per affrontare questo tipo di problema, dalle semplici generalizzazioni dei modelli lineari a modelli più complessi, come le foreste casuali, il *boosting* o il *deep learning*. Si presentano anche situazioni in cui i dati sono sia “larghi” che “alti” e molta letteratura degli ultimi anni si è sviluppata per affrontare i problemi, soprattutto computazionali, che si generano in questi casi, con un ampio sviluppo di metodologie di calcolo parallelo, o basate su campionamenti, come ad esempio, strumenti di *MapReduce* (si veda, ad esempio, Lämmel, 2008), algoritmi come l'*Alternating Direction Method of Multipliers* (si veda, ad esempio, Hastie *et al.*, 2015, par. 5.7), metodologie di tipo *divide et impera* (*divide and combine*, Cleveland e Hafen, 2014) o metodi basati su proiezioni casuali (*hashing*, si veda ad esempio Leskovec *et al.*, 2020, cap. 3) o che sfruttano la sparsità dei dati (ad esempio Hastie *et al.*, 2015).

3. L'analisi dei dati basata su modelli statistici

Le basi concettuali della statistica classica furono gettate nel fondamentale lavoro di Ronald Fisher (1922), *On the mathematical foundations of theoretical statistics*, definito da George Joseph Stigler (2005) come “senza dubbio il più influente contributo sull'argomento del ventesimo secolo”. Nel paragrafo intitolato “Obiettivo del metodo statistico”, Fisher afferma che obiettivo della statistica è sostituire l'intero insieme di dati con opportune quantità riassuntive e, subito dopo, definisce la cornice entro cui si sviluppa la statistica per molti decenni successivi:

Questo obiettivo è raggiunto ipotizzando una popolazione infinita, di cui i dati osservati siano considerati un campione casuale. La legge probabilistica di questa ipoteti-

ca popolazione è specificata da un numero relativamente piccolo di parametri, sufficienti per la descrizione sintetica di tutte le quantità di interesse.

Nel linguaggio attuale si tratta della definizione di modello statistico (Romeijn, 2014; Spanos, 2010; Davison, 2003; Pace e Salvan, 1997, cap. 1). Si tratta di un'astrazione utile in molti casi per rappresentare la variabilità accidentale dei dati. Poiché diverse leggi probabilistiche possono generare i dati effettivamente osservati, il modello statistico è un *insieme* di leggi di probabilità, che possono essere pensate come diverse *ipotesi* sul meccanismo generatore dei dati. In questo contesto, un'ipotesi non potrà mai essere falsificata con certezza, ma si può valutare, tramite il valore p (*p-value*), la conformità dei dati con una particolare ipotesi. Ciò può far ritenere (Candès, 2017) la costruzione di Fisher (1935), sebbene ideata indipendentemente, come uno sviluppo del falsificazionismo di Popper (1934). Sempre Fisher (1935) sottolinea che un esperimento scientifico dovrebbe essere disegnato solo per dare ai fatti la possibilità di confutare l'ipotesi nulla. Il ricercatore può essere più o meno esigente su quanto piccola debba essere la probabilità (valore p) che ritiene sufficiente per ammettere che le osservazioni abbiano confutato l'ipotesi nulla.

Assunto un modello statistico, la teoria della probabilità può venire usata per dedurre il comportamento di dati generati dal modello. La probabilità gioca quindi, in primo luogo, il ruolo di canale di comunicazione tra il modello e i dati, fornendo un legame *deduttivo*. La statistica *inverte* il processo, utilizzando i dati osservati per ottenere informazioni su quali elementi del modello risultino più plausibili alla luce dei dati, o anche per convalidare il modello stesso (Barnett, 1999).

Come disegnato in Figura 1, la teoria matematica della probabilità, tramite deduzioni logiche, permette di descrivere le proprietà probabilistiche dei dati generati da leggi di probabilità entro il modello. Queste proprietà, se il modello è appropriato, dovrebbero riflettere le caratteristiche osservate nei dati reali. Al contrario, la teoria della statistica considera i dati reali che si presentano in una situazione concreta, raccolti tramite disegno sperimentale o tramite semplice osservazione, e li usa per convalidare il modello, per formulare congetture razionali o stime di quantità numeriche rilevanti, per verificare ipotesi, per suggerire un nuovo modello, come pure per prendere decisioni in situazioni concrete.

Anche questo processo inverso, *induttivo*, si avvale tipicamente della teoria della probabilità, che permette di quantificare l'incertezza presente nelle affermazioni basate sui dati osservati. Va detto che non esiste un'unica via, o paradigma, per il processo induttivo (Barnett, 1999), ma entrambi i principali paradigmi, frequentista e bayesiano, utilizzano, sia pure in modo differente, il calcolo delle probabilità per valutare l'incertezza. In questo senso, la probabi-

lità assume un duplice ruolo: a) descrittivo della variabilità campionaria e b) *epistemologico*, di valutazione dell'incertezza della conoscenza acquisita dai dati (Cox, 1990; Reid e Cox, 2015).

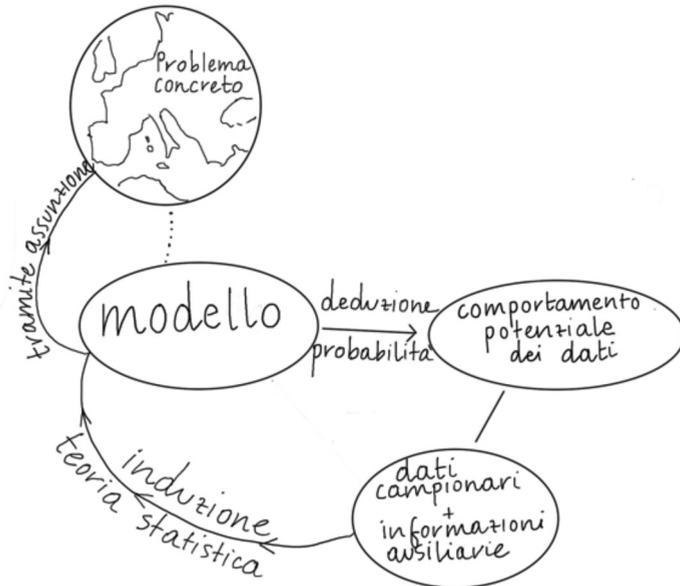


Fig. 1 - Rappresentazione grafica dell'analisi di dati basata su un modello statistico

4. Big Data: un nuovo paradigma scientifico?

Nell'ambito dei Big Data, quando i dati hanno struttura complessa e sono spesso raccolti senza seguire un piano sperimentale, è necessario specificare con particolare attenzione alcuni aspetti del modello proposto. Sono infatti di fondamentale importanza, in questo contesto, da una parte la quantificazione dell'incertezza, essere cioè in grado di dire quanto affidabili possono essere le affermazioni basate sui dati che si sono ottenute e, dall'altra, l'identificazione di quanto rilevanti siano le assunzioni che legano il modello statistico alla realtà o, nel linguaggio statistico, la quantificazione di quanto robuste sono le procedure di inferenza proposte.

Il contesto attuale di analisi di Big Data, tuttavia, sembra anche essere caratterizzato da un nuovo approccio che, secondo alcuni costituirebbe addirittura un nuovo paradigma scientifico (Anderson, 2008; Hey *et al.*, 2009; Kitchin, 2014). Già nel 2001, Leo Breiman discute sulla legittimità di una cultura statistica che partisse dai dati, non conoscendo il meccanismo che li ha generati:

Ci sono due culture nell'uso dei modelli statistici per raggiungere conclusioni dai dati. Una assume che i dati siano generati da un modello stocastico noto. L'altro utilizza modelli algoritmici e tratta il meccanismo generatore dei dati come ignoto" (Breiman, 2001).

L'articolo di Breiman è corredato alla fine da una vivace discussione con replica finale di Breiman stesso, segno che il dibattito è già presente fin dai primi anni 2000.

In moltissime situazioni, i dati preesistono alle ipotesi scientifiche: prima si raccolgono i dati per finalità anche molto diverse dalla ricerca scientifica, e solo in seguito ci si pongono le domande e si formalizzano le ipotesi scientifiche.

In un contesto in cui i dati sono raccolti prima della formulazione delle ipotesi di ricerca, diventa fondamentale quantificare l'affidabilità delle ipotesi stesse, formulate dopo aver 'visto' i dati (*data snooping*), e delle relative conclusioni. Si tratta di un approccio molto diverso rispetto alla *ricerca guidata dalle ipotesi*, e per questo un filone della letteratura statistica si occupa della cosiddetta *post-selection inference*, in cui si considera l'effetto "*look elsewhere*", per cui risultati che apparentemente sono statisticamente significativi, potrebbero in realtà esserlo in misura inferiore o non esserlo affatto, a causa dell'elevato numero di ipotesi verificate contemporaneamente sugli stessi dati. In queste situazioni è necessario misurare l'affidabilità delle procedure inferenziali nel contesto di tutte le ipotesi che vengono esplorate (per una rassegna introduttiva di questi metodi si veda, ad esempio, Goeman e Solari, 2011).

5. La scienza di apprendere dall'esperienza

In un contesto di analisi di Big Data, è spesso di interesse prevedere accuratamente nuovi casi non osservati sulla base di dati disponibili. La previsione "puntuale" è, di solito, il principale obiettivo dell'analisi, e spesso si preferisce trascurare ogni possibilità di interpretazione della realtà. In molti contesti, ci si concentra unicamente sulla costruzione di una procedura automatizzata per analizzare i dati. Si pensi, ad esempio, nella teoria dei segnali e nell'elaborazione di immagini, video o registrazioni audio, al problema di classificare automaticamente un'immagine con i soggetti rappresentati nell'immagine stessa, o a quello di identificare specifici oggetti presenti in una certa sequenza di immagini in un video, o convertire una registrazione audio in un testo.

Tuttavia, in altri contesti, spesso legati alla ricerca scientifica, è di fondamentale importanza ottenere idee e suggerimenti dalle grandi quantità di dati disponibili. In un contesto predittivo, risulta importante comprendere anche quali variabili hanno effetto sulla quantità di interesse e in che modo questa relazione viene caratterizzata.

In un contesto scientifico, risulta anche fondamentale misurare la qualità della previsione e delle inferenze che vengono svolte, per garantire riproducibilità, replicabilità e affidabilità dei risultati ottenuti (Dunson, 2018). Si tratta infatti di predisporre procedure caratterizzate da possibilità di generalizzazione e replicabilità. I risultati basati sul medesimo modello e algoritmo applicati a un nuovo insieme di dati, somiglianti a quello già analizzato, dovrebbero essere gli stessi, o almeno “simili”. È necessario, poi, definire quando due risultati possono essere considerati simili tra loro ed è cruciale misurare l’incertezza relativa alle conclusioni che si ottengono dalle procedure di previsione e di inferenza.

6. Misurare l’incertezza: un esempio

Si supponga di avere osservato x_1, \dots, x_n , numeri che descrivono un fenomeno di interesse, ad esempio il tasso di incidenti automobilistici in Italia, negli $n=7904$ comuni italiani. La *media aritmetica*

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

riassume in un unico numero quanto osservato. Siamo, tuttavia, interessati ad avere una misura dell’accuratezza di questo numero. Ad esempio, è possibile calcolare l’*errore standard*

$$\widehat{se} = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

In questo semplice esempio, l’operatore media aritmetica \bar{x} corrisponde all’“algoritmo” di stima, mentre l’errore standard fornisce un’inferenza sull’accuratezza dell’algoritmo. L’aspetto importante della teoria statistica, qui evidente, consiste nel fatto che gli stessi dati che forniscono una stima, possono anche valutarne l’accuratezza. Chiaramente l’algoritmo precede la valutazione inferenziale ed è un’attività certamente più affascinante e “avventurosa” rispetto alla misura della sua qualità che, invece, sembra rivestire un ruolo minoritario, in un’ottica di comunicazione e divulgazione (Efron e Hastie, 2016, cap. 1).

In questo contesto diviene importante, soprattutto in ambito scientifico, non amplificare i risultati ottenuti dall’applicazione di un approccio solo al-

goritmico, per quanto sofisticato, mentre è fondamentale quantificare in maniera appropriata l'incertezza, che spesso è enorme, per evitare di riempire la letteratura scientifica di risultati falsi o imprecisi. Infatti, quando l'incertezza sui risultati ottenuti viene quantificata in modo appropriato, si può osservare un opportuno ridimensionamento della valenza delle deduzioni scientifiche ottenute da un approccio solo algoritmico. La replicabilità degli esperimenti è, al giorno d'oggi, ampiamente riconosciuta come una sfida chiave per la scienza. *Replicabilità*, non nel senso di rendere il codice disponibile in modo da riprodurre i medesimi risultati per qualche altro esempio, che è preferibile chiamare *riproducibilità*, ma nel senso che un altro scienziato può replicare l'esperimento, ottenere nuovi dati e giungere alle medesime conclusioni scientifiche (ad esempio, Schwalbe, 2016).

I metodi basati unicamente su algoritmi, come ad esempio il *deep learning* e molti metodi di intelligenza artificiale, sono particolarmente adatti in situazioni in cui i dati hanno una chiara struttura spaziale o temporale (immagini, video, audio, ecc.), e quando sono disponibili grandissimi insiemi di dati in cui è presente la variabile risposta e l'interesse principale riguarda una previsione di tipo *black box*, in cui cioè non è di interesse fare inferenza e quantificare l'incertezza.

Tuttavia, in molti ambiti scientifici i dati sono complessi, senza una chiara struttura spaziale o temporale, sono disponibili moltissime variabili, ma non altrettante unità statistiche (si pensi, ad esempio, al mondo della genomica o dell'analisi di testi). In questi contesti, gli scienziati non sono quasi mai soddisfatti da previsioni a scatola nera (*black box*), gli studi scientifici sono spesso incentrati sulla comprensione della struttura del problema in esame, il meccanismo generatore dei dati, anche se una previsione ottima (ad esempio di un *outcome* medico) è chiaramente un aspetto cruciale della ricerca.

7. Un esempio: le connessioni nel cervello umano

Consideriamo un esempio in ambito medico, legato alle connessioni all'interno del cervello umano (*human brain connectomes*). Tramite risonanza magnetica e *diffusion tensor imaging* è possibile stimare le posizioni di tratti di fibra di materia bianca (che hanno il ruolo di grandi vie di comunicazione per l'attività neurale e per la comunicazione all'interno del cervello) per ciascun individuo. Molti diversi tratti di fibra vengono misurati su ogni individuo. In uno studio di neuroscienza, sono stati raccolti dati relativi a circa un migliaio di individui, e l'interesse scientifico principale consiste nel capire come le reti ottenute dalle connessioni nel cervello si collegano a delle variabili risposta di interesse (*outcome*). Si vuole cioè prevedere una quantità di interesse (ad esempio, il quoziente di intelligenza) usando le caratteristiche delle connes-

sioni nel cervello. In questo esempio il numero n di soggetti nello studio è molto più piccolo del numero p di variabili misurate, i dati sono *complessi* e *strutturati geometricamente*, e la scelta di p è in qualche modo arbitraria. Infatti la tecnologia usata per le misurazioni raccoglie dati ad una risoluzione troppo elevata per essere utilizzata nelle analisi, perché comporterebbe problemi computazionali e statistici irrisolvibili. Non è possibile analizzare i dati a livello di singolo *voxel* per l'impossibilità di allineare i cervelli di persone diverse a quel livello di risoluzione, dato l'enorme numero di coppie di *voxel*. Si deve perciò scegliere la risoluzione adeguata per analizzare i dati e, di conseguenza, il valore di p .

Un tipico approccio 'algoritmico' moderno consiste nell'effettuare una fase di pre-analisi dei dati, per ridurre la complessità e rendere semplice e trattabile con strumenti automatici la struttura geometrica osservata. Ad esempio, si possono ridurre le informazioni relative all'individuo i -esimo in una variabile risposta y_i e un vettore di variabili esplicative $x_i = [x_{i1}, \dots, x_{ip}]$. La variabile risposta è la quantità di interesse per l'individuo i , mentre x_{ij} sono, ad esempio, variabili indicatrici di qualche struttura di connessione tra la j -esima coppia di regioni del cervello, per $j = 1, \dots, p = R(R-1)/2$, dove R è il numero di regioni in cui il cervello è stato suddiviso. Per prevedere y_i come funzione di x_i si possono poi usare strumenti flessibili di regressione o classificazione, come *boosting*, foreste casuali, o reti neurali, anche se questi strumenti non sono di facile interpretazione. È tipico, inoltre, includere nella modellazione strumenti di selezione delle variabili, ad esempio con metodi come il *lasso* (si veda, ad esempio, Hastie *et al.*, 2015), in modo da identificare le coppie di regioni del cervello che hanno connessioni rilevanti per variabile risposta. Tuttavia, questa selezione dipende in maniera critica dalla scelta della risoluzione usata, cioè da R (e quindi da p). I risultati di questa analisi, per un qualche valore di R , potrebbero essere, perciò, inaffidabili e rischierebbero di produrre errori nella pratica portando a sovrainterpretare quanto ottenuto. Inoltre, il tipico *output* di queste analisi non fornisce misure di affidabilità e della quantità di incertezza presente nei risultati.

Tipicamente in problemi a elevata dimensionalità come quello appena descritto, ci si concentra sulla stima puntuale, risolvendo il problema di ottimizzazione tramite una penalizzazione che riduca efficacemente la dimensionalità del modello. Questi metodi sono giustificati da considerazioni di tipo asintotico, dove si ipotizza non solo che $n \rightarrow \infty$ ma anche che $p \rightarrow \infty$ con n , spesso anche ad un tasso più elevato. Sotto vincoli abbastanza restrittivi (ad esempio che la verità è sparsa, che la matrice dei dati è quasi ortogonale e che i segnali non nulli sono abbastanza grandi) è possibile fornire dei risultati asintotici in termini di capacità di identificare la vera struttura di dimensione ridotta, cioè i coefficienti effettivamente nulli, e stimare accuratamente i parametri.

Tuttavia questi metodi, da una parte, producono solo una stima puntuale senza misura di incertezza e, dall'altra, vengono giustificati in campioni finiti solo tramite studi di simulazione limitati a valutare il tasso di errore relativamente ad altri metodi di stima puntuale. Applicare risultati ottimi ottenuti assumendo che la dimensione tenda ad infinito sotto condizioni eccessivamente idealizzate rassicura poco in piccoli campioni. In realtà, è evidente che si osserveranno errori elevati. Piuttosto che focalizzarsi su forti risultati positivi basati su assunzioni irrealistiche avrebbe certamente maggior impatto disporre di una teoria che cerchi di descrivere i comportamenti positivi e negativi sulla base di assunzioni realistiche che nascono dal problema specifico che si sta analizzando. Se il problema è troppo complicato per essere risolto con i dati disponibili, allora è meglio essere consapevoli che si sta cercando un obiettivo impossibile.

Nell'esempio delle connessioni cerebrali non si è solo interessati ad ottenere un algoritmo predittivo con la semplice aggiunta di una stima delle variabili importanti. Si vorrebbero invece identificare strumenti che dicano quanto affidabili sono le decisioni basate sulla selezione delle variabili (connessioni) data la dimensione campionaria, la dimensionalità e la struttura di correlazione dei dati disponibili. Saremmo interessati ad ottenere anche eventuali risultati che ci suggeriscano, se necessario, di essere meno ambiziosi sul tipo di inferenza da fare – ad esempio, forse non possiamo pretendere di esaminare le reti cerebrali ad una risoluzione troppo fine, date le limitazioni statistiche – e quindi R dovrebbe essere scelto inferiore a una qualche soglia. Abbiamo bisogno di misure accurate e interpretabili di incertezza dei risultati, anche utilizzando strumenti che includano maggiore conoscenza e struttura nelle analisi, in modo da migliorare le prestazioni. Infine, i dati, nella loro forma “nativa”, spesso non hanno la struttura di matrici e in molti casi includono vincoli geometrici che potrebbero essere sfruttati, limitando la fase di preparazione dei dati e migliorando l'efficienza delle stime oltre che i suggerimenti ai medici.

Spesso, in un contesto scientifico, i ricercatori conoscono molto bene i vincoli che dovrebbero essere imposti nelle analisi, oltre alla sparsità, al rango ridotto e alle altre semplici assunzioni che si impongono in un approccio puramente algoritmico. Inoltre, è difficile fare progressi nel mondo scientifico utilizzando generici algoritmi predittivi (*agnostic*), che non tengano in considerazione come i dati sono stati raccolti e il bagaglio di conoscenze già disponibili nel contesto che si sta analizzando. Di conseguenza, c'è ancora bisogno di approcci statistici sviluppati per applicazioni scientifiche, pensati con cura e finalizzati agli obiettivi.

8. Distorsioni ed errori nell'analisi di Big Data

Distorsione da selezione del campione

Un tema fondamentale, da considerare con attenzione, è legato al processo di selezione con cui grandi basi di dati vengono raccolte. Nel condurre inferenza statistica, ci si concentra sulla stima di un particolare parametro di interesse, θ , che rappresenta una caratteristica di una qualche popolazione di interesse P . Ad esempio, θ potrebbe essere la proporzione di cittadini obesi in Italia o la percentuale di cittadini statunitensi che hanno votato Trump nelle ultime elezioni. Se dalla popolazione P si estraesse un campione *casuale* $x = [x_1, \dots, x_n]$, con $x_i = 1$ se l' i -esimo individuo del campione è obeso e $x_i = 0$ altrimenti, potremmo semplicemente stimare la proporzione di cittadini obesi usando la media del campione.

Tuttavia, quando si ha a che fare con *Big Data*, non si hanno praticamente mai campioni casuali dalla popolazione di interesse, ma generalmente campioni di dati raccolti tramite un processo complesso, tipicamente ignoto. Ad esempio, quando i dati vengono raccolti tramite indagini via web, o tramite *social media* come Twitter, l'inclusione nel campione non corrisponde ad una estrazione casuale, ma sono gli individui stessi che scelgono se rispondere o meno alla domanda di interesse posta via web, o di 'postare' su quel tema. A causa della distorsione dovuta alla selezione (*selection bias*) e agli errori di misura (*measurement errors*), la media delle risposte nel campione potrebbe essere anche molto diversa dalla vera proporzione nella popolazione di interesse, anche quando la dimensione campionaria è enorme.

La cosiddetta "distorsione per selezione" deriva dal fatto che gli individui che rispondono alla domanda rappresentano una popolazione Q che potrebbe essere anche molto diversa e in maniera sostanziale dalla popolazione di interesse P (considerando l'esempio sull'obesità in Italia, anche in termini di obesità). Nella popolazione Q di chi sceglie di rispondere all'indagine via web, potrebbe ad esempio essere presente una maggiore proporzione di maschi con più di 40 anni, rispetto alla popolazione di interesse P , provocando effetti rilevanti nella stima della percentuale di obesi. Nell'esempio, "errori di misura" si presentano quando gli individui non sanno con precisione il loro stato di "obesità" o quando chi risponde vuole o può fornire risposte non esatte o non precise. Senza sapere in che modo Q è diverso da P , o senza conoscere la dimensione dell'errore di misura, la stima di θ basata sul campione disponibile può essere completamente errata e quindi inutile.

Avere a disposizione enormi quantità di dati (Big Data) non risolve il problema. Diminuisce l'incertezza statistica nello stimare una quantità che è, di fatto, sbagliata (nell'esempio, un attributo della popolazione Q anziché della popolazione di interesse P). In questo contesto Meng (2018) definisce un

“paradosso dei Big Data” per cui, più grande è la quantità di dati, più è facile farsi ingannare. Se si considera il semplice problema di stimare la media di una variabile X di una popolazione P con N unità, di cui si osserva un campione, non necessariamente casuale, e si considera la differenza tra la media aritmetica \bar{x}_n ottenuta nel campione con n osservazioni e \bar{x}_N la media nell'intera popolazione di interesse, è possibile scrivere una identità che lega qualità dei dati (*data quality*), quantità dei dati (*data quantity*) e difficoltà del problema specifico considerato. Sia, infatti, $R = [R_1, \dots, R_N]$ il vettore di variabili indicatrici della presenza di ciascuna unità della popolazione nel campione, cioè $R_i = 1$, per $i = 1, \dots, N$, se l' i -esima osservazione della popolazione è stata selezionata nel campione e 0 altrimenti. Vale la seguente identità:

$$\bar{x}_n - \bar{x}_N = \underbrace{\rho_{R,X}}_{\text{qualità dei dati}} \times \underbrace{\sqrt{\frac{1-f}{f}}}_{\text{quantità dei dati}} \times \underbrace{\sigma_X}_{\text{difficoltà del problema}},$$

dove la *difficoltà del problema* è colta da σ_X , che è la deviazione standard di X (più cresce più è difficile stimare \bar{x}_N in maniera accurata), la *quantità dei dati* è colta da $\frac{1-f}{f}$ con $f = \frac{n}{N}$ (quando il campione coincide con l'intera popolazione vale 0, $f = 1$, e cresce all'infinito quando non si hanno dati, cioè $f = 0$) e la *qualità dei dati* è rappresentata dalla correlazione tra la variabile nella popolazione e l'indicatore della selezione del campione R , $\rho_{R,X} = \text{corr}(R,X)$. In assenza di distorsione da selezione (*selection bias*), cioè con campionamento probabilistico, la probabilità di osservare o meno un particolare valore di X non dipende dal suo valore. Di conseguenza $\rho_{R,X}$ sarà mediamente nulla. Quando si perde il controllo del campionamento probabilistico, ciò che determina l'errore di stima non è più la numerosità campionaria n , ma la dimensione della popolazione N . A parità di correlazione media $\mathbb{E}_R\{\rho_{R,X}\} \neq 0$, l'errore di \bar{x}_n rispetto all'analogo errore sotto campionamento casuale semplice senza reinserimento (in blocco), cresce con la dimensione della popolazione N ad un tasso di \sqrt{N} , ossia

$$\frac{\bar{x}_n - \bar{x}_N}{\sqrt{\text{Var}_{CCS}\{\bar{x}_n\}}} = \frac{\rho_{R,X} \sqrt{\frac{1-f}{f}} \sigma_X}{\sqrt{\frac{1-f}{n} \frac{N}{N-1}} \sigma_X} = \sqrt{N-1} \rho_{R,X},$$

dove $\text{Var}_{CCS}\{\bar{x}_n\}$ è la varianza della media campionaria nel caso di campionamento in blocco. Questa formula dice che se si usa la media campiona-

ria come stima della media della popolazione, l'errore che si commette è $\sqrt{N - 1}\rho_{R,X}$, sulla scala dell'errore standard del campionamento casuale semplice. Quando la numerosità della popolazione N cresce, se si vuole mantenere l'errore legato alla qualità dei dati ad un ordine di grandezza confrontabile con le altre componenti dell'errore complessivo, la correlazione tra R e X non deve essere superiore a $N^{-\frac{1}{2}} = 1/\sqrt{N}$.

Si pensi ad esempio alla previsione di voto negli USA del 2016, quando è stato eletto il Presidente Ronald Trump. La popolazione USA votante era di circa $N \approx 230.000.000$ persone. Per raggiungere $\rho_{R,X} \approx 1/\sqrt{N}$, il massimo valore di $\rho_{R,X}$ ammissibile è $\rho_{R,X} \approx 0.000084$, un coefficiente di correlazione minuscolo per essere 'coperto' da un meccanismo di selezione autoregolato. Le stime ottenute dal Cooperative Congressional Election Study (CCES) relative alle elezioni americane nel 2016 suggeriscono una correlazione di $\rho_{R,X} \approx -0.005$ relativa alla (preventiva) autodichiarazione di votare per Donald Trump, dovuta a una consistente quantità di non rispondenti tra i votanti Trump. Ciò significa che, i risultati inferenziali che si ottengono usando un campione di numerosità 1% ($n = 2.300.000$), a causa di questa piccolissima correlazione tra selezione del campione e voto per Trump, corrispondono a quanto si otterrebbe con campione di $n = 400$ unità campionate casualmente (una riduzione della "numerosità campionaria effettiva" del 99,98%).

Decisioni corrette (fair decisions)

Negli ultimi anni si è osservato un crescente interesse nello sviluppo di algoritmi automatici e strumenti statistici per assistere, ed eventualmente sostituire, gli esseri umani nei processi di decisione ad "alto rischio". Alcuni classici casi (Corbett-Davies *et al.*, 2017) sono legati all'assunzione di lavoratori in un'azienda sulla base del curriculum vitae, alle decisioni sulle politiche ambientali considerando il livello di inquinamento, alle scelte di accordare o meno una linea di credito utilizzando il profilo creditizio del richiedente, alle decisioni sulla colpevolezza di un imputato sulla base del comportamento precedente o alla scelta di ammettere uno studente all'università sulla base dei voti ottenuti nella scuola superiore. L'utilità dell'introduzione di algoritmi in questi processi decisionali deriva da diverse considerazioni relative al tentativo di ridurre comportamenti scorretti da parte dei decisori oltre alla volontà di ridurre i costi ed aumentare l'efficienza di questi processi (ad esempio, Friedler *et al.*, 2019). Si ritiene, infatti, che decisioni che si basano su sistemi automatici siano imparziali, efficienti ed obiettive, assumendo che i computer non siano in grado di avallare pregiudizi od opinioni soggettive. Correttezza ed oggettività sono caratteristiche fondamentali nei processi

decisionali, soprattutto quando le scelte hanno un impatto importante sulla società o sui singoli individui. Tuttavia, gli algoritmi automatici vengono “addestrati” su grandi insiemi di dati raccolti in studi osservazionali in cui il meccanismo di selezione campionaria è generalmente complesso o non noto, rendendo i dati soggetti a qualche forma di distorsione sistematica che potrebbe influenzare l'imparzialità e l'oggettività anche delle decisioni basate su strumenti automatici. Quando gli algoritmi di *machine learning* vengono “addestrati” su dati raccolti tramite procedure distorte, i risultati riprodurranno e amplificheranno la distorsione presente nei dati (ad esempio, Angwin *et al.*, 2016).

Ad esempio, nel processo di selezione di candidati per l'assunzione in aziende *high-tech*, ci si basa su liste ridotte di potenziali candidati, selezionate da algoritmi che utilizzano curriculum vitae, e che sono stati “addestrati” sui dati relativi ai precedenti colloqui di lavoro dell'azienda stessa (ad esempio, Hoffman *et al.*, 2017). Se nel passato era presente un diverso atteggiamento relativamente al genere del candidato, questa tendenza verrebbe propagata negli algoritmi e nelle previsioni futuri, amplificando il fenomeno invece di mitigarlo.

Un altro tipico esempio riguarda gli algoritmi, che si cominciano ad usare in alcuni paesi, per aiutare la giustizia criminale. I modelli che valutano il rischio criminale si basano, tipicamente, su dati demografici e sulla storia criminale passata per prevedere la probabilità che una persona che è stata arrestata possa venire arrestata di nuovo nel futuro. Tali previsioni hanno effetti sulla detenzione preventiva, sulla libertà vigilata, e sulla condanna. I dati usati per addestrare gli algoritmi si basano sulle informazioni relative ai precedenti arresti e fermi della polizia che, ad esempio negli Stati Uniti, hanno una forte distorsione razziale (Simoiu *et al.*, 2017; Rudovsky, 2001). Le informazioni nei *database* della polizia americana non sono chiaramente rappresentative della criminalità, a causa della discrezionalità degli agenti di polizia, che possono scegliere le aree da pattugliare e gli individui da trattenere (Lum e Isaac, 2016). Di conseguenza, utilizzando algoritmi “addestrati” su questi dati, persone di colore vengono fermate più spesso dei caucasici, saranno quindi sovrarappresentate nel campione e verranno quindi, mediamente, assegnate a categorie di rischio più elevate.

Per cercare di mitigare queste distorsioni, può essere utile usare metodi statistici che correggono l'automatismo di questi algoritmi, aggiustando in modo opportuno gli aspetti che non funzionano. Il pannello di sinistra della Figura 2 mostra l'istogramma della distribuzione della probabilità di un nuovo arresto, prevista tramite un algoritmo automatico basato sulle informazioni relative ai precedenti arresti. È evidente che l'algoritmo prevede un rischio di un nuovo arresto per persone di colore sistematicamente più elevato rispetto ai caucasici. Se si effettua un'opportuna correzione nei dati (Aliverti *et al.*,

2020) per contrastare la distorsione presente, legata al colore della pelle, e si applica il medesimo algoritmo, si ottiene la distribuzione nel pannello di destra della Figura 2, in cui si vede chiaramente che, a parità di altre caratteristiche, l'algoritmo non assegna più un diverso rischio di un nuovo arresto a seconda che la persona sia caucasica o nera.

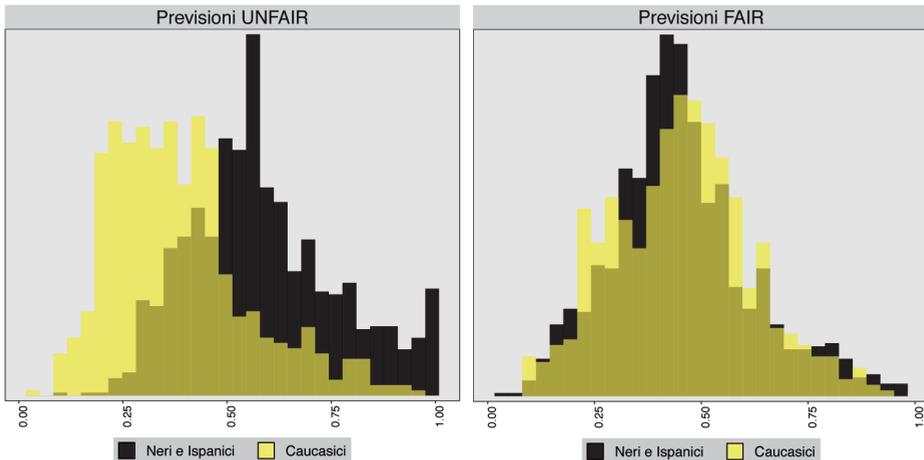


Fig. 2 - Istogramma della distribuzione del rischio di un nuovo arresto per neri e ispanici da un lato e caucasici dall'altro. A sinistra previsione con algoritmo automatico; a destra previsione con algoritmo automatico con correzione per imparzialità

9. Aprire le scatole nere

L'idea di “aprire le scatole nere” degli algoritmi per capire cosa succede dentro e migliorarne il contenuto, come nell'esempio della recidiva di arresto, viene ben riassunto da una frase di Brad Efron: “[...] and I sometimes think that the AI crowd is not critical enough – is a little too facile. The whole point of science is to open up black boxes, understand their insides and build better boxes for the purposes of mankind” (Brad Efron, An Interview with Brad Efron of Stanford, www.b-eye-network.com/view/9947).

Dai diversi esempi che abbiamo citato, è evidente che tipicamente i *Big Data* sono soggetti a distorsioni da selezione o ad errori di misura e spesso non sono disponibili informazioni nei dati sulla grandezza, la fonte e il tipo di questi errori. In questi casi, l'impiego di dati per accrescere la conoscenza scientifica e per supportare importanti decisioni richiede particolare attenzione e consapevolezza. È cruciale identificare gli aspetti di debolezza dei dati

e tentare, per quanto possibile, di tenerne conto in fase di analisi, includendo queste fonti di errore nella quantificazione dell'incertezza.

Per concludere, ricordiamo una frase di Leonardo Da Vinci, che già cinquecento anni fa metteva in guardia dai pericoli di ottenere risultati pratici senza voler realmente comprendere il fenomeno di interesse: *Quelli che s'innamoran di pratica senza scienza son come 'l nocchier ch'entra in navilio senza timone o bussola, che mai ha certezza dove si vada* (Leonardo da Vinci, *Trattato della Pittura*, parte seconda, 77).

Ringraziamenti

Il materiale presentato è frutto di un'attività condivisa con diversi laureandi, dottorandi e ricercatori di questi anni, molti dei quali hanno ora posizioni rilevanti nel mondo accademico e aziendale italiano. In particolare, siamo debitori a Mariangela Guidolin, Emanuele Aliverti, Annalisa Balata, Antonio Canale, Daniele Durante, Sally Paganin, Tommaso Rigon, Massimiliano Russo e Anna Zanin.

Abstract

We look at the role of statistics in analyzing Big Data. After introducing the view of data analysis based on statistical models, we discuss its limits and the challenges coming from big data. Besides the need of developing appropriate concepts, methodology and algorithms, we discuss interpretability, uncertainty quantification, replicability and selection bias as ingredients of any new method introduced to analyze big data.

Bibliografia

- Aliverti E., Lum K., Johndrow J.E., Dunson D.B. (2018), *Removing the influence of a group variable in high-dimensional predictive modelling*, «Journal of the Royal Statistical Society, Series A», 184, pp. 791-811.
- Anderson C. (2008), *The end of theory: The data deluge makes the scientific method obsolete*, «Wired», 23 giugno 2008. Disponibile su: www.wired.com/science/discoveries/magazine/16-07/pb_theory (consultato il 29 febbraio 2020).
- Angwin A., Larson J., Mattu S., Kirchner L. (2016), *Machine bias: there's software used across the country to predict future criminals. And it's biased against blacks*, «ProPublica», 23 maggio 2016. Disponibile su: www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (consultato il 29 febbraio 2020).
- Azzalini A., Scarpa B. (2012), *Data Analysis and Data Mining*, Oxford University Press, New York.
- Barnett V. (1999), *Comparative Statistical Inference*, John Wiley & Sons, Londra.
- Breiman L. (2001), *Statistical modeling: The two cultures*, «Statistical Science», 16, pp. 199-215.
- Bühlmann P., van de Geer S. (2018), *Statistics for big data: A perspective*, «Statistics and Probability Letters», 136, pp. 37-41.
- Candes E. (2017), *Around the reproducibility of scientific research in the big data era: what statistics can offer?* J.L. Lions Colloquium, UPMC, Parigi, 17 marzo 2017. Disponibile su www.ljll.math.upmc.fr/contenu/article/lecons-j-l-lions-2017-17-03-2017-14h00-colloquium-e-candes (consultato il 29 febbraio 2020).
- Cleveland W.S., Hafen R. (2014), *Divide and recombine: Data science for large complex data*, «Statistical Analysis and Data Mining», 7, pp. 425-433. Doi: 10.1002/sam.11242
- Corbett-Davies S., Pierson E., Feller A., Goel S., Huq A. (2017), *Algorithmic decision making and the cost of fairness*, in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 797-806.
- Cox D.R. (1990), *Role of models in statistical analysis*, «Statistical Science», 5, pp. 169-174.
- Davison A.C. (2003), *Statistical Models*. Cambridge University Press, Cambridge.
- Dunson D.B. (2018), *Statistics in the big data era: failures of the machine*, «Statistics & Probability Letters», 136, pp. 4-9.
- Efron B., Hastie T. (2016), *Computer Age Statistical Inference*, Cambridge University Press, New York.
- Fisher R.A. (1922), *On the mathematical foundations of theoretical statistics*, «Philosophical Transactions of the Royal Society A», 222, pp. 309-368.
- Fisher R.A. (1935), *The Design of Experiments*, Oliver and Boyd, Edimburgo.
- Friedler S.A., Scheidegger C., Venkatasubramanian S., Choudhary S., Hamilton E.P., Roth D. (2019), *A comparative study of fairness-enhancing interventions in machine learning*, in *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 329-338. Disponibile su arXiv:1802.04422v1.
- Friedman J.H. (1997), *Data Mining and Statistics: What's the Connection?*, in *Proceedings of the 29th Symposium on the Interface Between Computer Science and Statistics*. Disponibile su: <http://statweb.stanford.edu/~jhf/ftp/dm-stat.pdf>.
- Friedman J.H. (2001), *The role of statistics in the data revolution?*, «International Statistical Review», 69, pp. 5-10.

- Goeman J.J., Solari A. (2014), *Tutorial in Biostatistics: Multiple hypothesis testing in genomics*, «Statistics in Medicine», 33, pp. 1946-1978.
- Goeman J.J., Solari A. (2011), *Multiple testing for exploratory research*, «Statistical Science», 26, pp. 584-597. Doi: 10.1214/11-STS356
- Hastie T., Tibshirani R., Wainwright M. (2015), *Statistical Learning with Sparsity: The Lasso and Generalizations*, CRC Press, Boca Raton.
- Hey T., Tansley S., Tolle K. (2009), *The Fourth Paradigm: Data Intensive Scientific Discovery*, Microsoft Research.
- Hoffman M, Kahn L.B., Li D. (2018), *Discretion in Hiding*, «The Quarterly Journal of Economics», 133, pp. 765-800. Doi: 10.1093/qje/qjx042
- Kitchin R. (2014), *Big Data, new epistemologies and paradigm shifts*, «Big Data & Society», 1, pp. 1-12. Doi: 10.1177/205395171452848
- Lämmel R. (2008), *Google's map reduce programming model - revisited*, «Science of Computer Programming», 70, pp. 1-30. Doi: 10.1016/j.scico.2007.07.001
- Leskovec J., Rajaraman A., Ullman J. (2010), *Mining of Massive Datasets*, Cambridge University Press.
- Lum K., Isaac W. (2016), *To predict and serve?*, «Significance», 13, pp. 14-19.
- Meng X.-L. (2018), *Statistical paradises and paradoxes in big data (I): Law of large populations, big data paradox, and the 2016 US presidential election*, «The Annals of Applied Statistics», 12, pp. 685-726.
- Pace L., Salvan A. (1997), *Principles of Statistical Inference from a neo-Fisherian Perspective*, World Scientific, Singapore.
- Parmigiani G., Garrett E.S., Irizarry R.A., Zeger S.L. (2003), *The Analysis of Gene Expression Data: Methods and Software*, New York.
- Popper K. (1934), *Logik der Forschung. Zur Erkenntnistheorie der modernen Naturwissenschaft*, Verlag Von Julius Springer, Vienna.
- Reid N., Cox D.R. (2015), *On some principles of statistical inference*, «International Statistical Review», 83, pp. 293-308.
- Romeijn J.W. (2014), *Philosophy of Statistics*, Stanford Encyclopedia of Philosophy.
- Rudovsky D., *Law enforcement by stereotypes and serendipity: racial profiling and stops and searches without cause*, «University of Pennsylvania Journal of Constitutional Law», p. 296.
- Senn S. (2003), *Dicing with Death*, Cambridge University Press, Cambridge.
- Silge J., Robinson D. (2017), *Text Mining with R. A Tidy Approach*, O'Reilly Media, Inc., Sebastopol (Usa).
- Simoiu C., Corbett-Davies S., Goel S. (2017), *The problem of infra-marginality in outcome tests for discrimination*, «The Annals of Applied Statistics», 11, pp. 1193-1216.
- Spanos A. (2010), *Akaike-type criteria and the reliability of inference: Model selection versus statistical model specification*, «Journal of Econometrics», 158, pp. 204-220.
- Stigler S. (2005), *Fisher in 1921*, «Statistical Science», 20, pp. 32-49.
- Schwalbe M. (2016), *Statistical Challenges in Assessing and Fostering the Reproducibility of Scientific Results: Summary of a Workshop*, National Academies Press, Washington DC (Usa).
- Wikipedia (2020), *Statistica*, in Wikipedia, *L'enciclopedia libera*. Disponibile su <https://en.wikipedia.org/wiki/Statistics> (consultato il 29 febbraio 2020).

Le sfide della biologia dei Big Data

di Sabina Leonelli*

La disponibilità di Big Data è in grado di trasformare numerosi settori delle scienze della vita e di inaugurare nuovi modi di fare ricerca. In questo articolo sostengo che la biologia dei Big Data solleva degli interrogativi importanti anche per la filosofia della scienza: ad esempio, cosa costituisce un buon *dataset*, e come è possibile ricavare conoscenza affidabile dai Big Data? Collaborazioni tra biologi, *data scientists* e filosofi della scienza ci aiuteranno a rispondere a queste e a ulteriori domande.

Le scienze della vita hanno una lunga esperienza nel trattare grandi quantità di dati, e recenti avanzamenti nelle possibilità sperimentali hanno largamente ampliato la mole dei dati che necessita di essere archiviata e analizzata. Anche il potere computazionale a disposizione dei ricercatori è cresciuto nel tempo, ma il volume e l'eterogeneità dei dati eccedono puntualmente le correnti strategie e gli strumenti a disposizione per la loro collezione e analisi. Inoltre, il volume dei dati oggi a disposizione, specialmente nelle scienze "omiche", solleva degli interrogativi fondamentali sui processi alla base della ricerca, come ad esempio il ruolo della teoria, l'importanza del contesto, la funzione delle abilità operative (cosiddette di *know-how*) nell'interpretazione dei dati. Ad esempio, esiste un diffuso dibattito sul livello di familiarità che uno scienziato è tenuto ad avere – per poter essere in grado di interpretare i dati – con i protocolli e le strumentazioni utilizzate per generare dati, e con i rilevanti aspetti biologici degli organismi utilizzati. Fino a che punto gli algoritmi riescano a identificare in modo attendibile i legami causali all'interno dei dati è ugualmente oggetto di controversia: scoprire che un determinato meccanismo di regolazione genica viene di frequente associato a un particolare tratto fenotipico non significa comprendere perché questo accade e se è proprio tale meccanismo a causare la comparsa del tratto.

* Dipartimento di Sociologia, Filosofia e Antropologia, Università di Exeter. La traduzione dall'inglese è di Andra Meneganzin.

Ci sono molti altri interrogativi che sono d'interesse per i filosofi della scienza. Far affidamento sui Big Data cambia forse la stessa idea di scoperta biologica e di cosa debba essere inteso come conoscenza biologica? Che ruolo svolgono le teorie all'interno di una ricerca che fa un utilizzo intensivo di dati, e come si rapporta la biologia dei Big Data alla ricerca guidata dall'ipotesi, dall'osservazione e dall'esplorazione? In che modo l'automatizzazione dell'analisi dei dati influenza l'affidabilità dei risultati? Qual è la differenza tra dati e rumore, e soprattutto, che cosa sono in primo luogo i dati? I biologi potrebbero pensare che queste domande, nonostante siano in sé interessanti e importanti, siano state in qualche modo rimosse dal proprio lavoro quotidiano, e vi possano perciò risultare irrilevanti. In questo articolo mi propongo di contrastare criticamente questa percezione, dando rilievo a considerazioni filosofiche che possano essere d'aiuto nel far fronte ad alcune delle sfide-chiave derivanti dell'impiego dei Big Data in biologia.

2. La biologia dei Big Data incontra il pluralismo biologico

La biologia è notoriamente frammentata nei suoi diversi metodi, fini, strumentazioni, e nelle sue intelaiature concettuali. Spesso diversi gruppi – addirittura all'interno della medesima sotto-branca – sono in disaccordo sulla terminologia da preferire, sugli organismi di ricerca e sui metodi e sui protocolli sperimentali (Leonelli, 2012). Di conseguenza, lo stesso termine può venir impiegato per riferirsi a processi differenti, oppure definizioni diverse possono far capo al medesimo termine. Questa profonda frammentazione, che i filosofi chiamano “pluralismo” (Kellert *et al.*, 2006) si rispecchia nelle numerose tecnologie e negli standard specifici per ogni settore che vengono utilizzati per generare, archiviare, condividere e analizzare dati (O'Malley e Soyer, 2012). Metter a punto delle strategie per affrontare questo pluralismo è una sfida cruciale per la biologia dei Big Data.

Risulta facile licenziare queste difficoltà come questioni puramente tecniche, che possono essere risolte, per esempio, utilizzando *database* inter-operativi e formati file per integrare dati provenienti da fonti differenti, affinché questi possano essere utilizzati e riutilizzati in svariati contesti di ricerca. Tuttavia, ci sono delle difficoltà concettuali e filosofiche più profonde. I *database* devono essere consultati (o “interrogati”, “*query*”) attraverso un sistema di ricerca comune, e questo solleva il problema di quali terminologie debbano essere impiegate per classificare i dati e per integrarli con altri dati, nonché delle implicazioni che queste scelte comportano. L'enorme sforzo dedicato all'elaborazione di sistemi di recupero delle informazioni per i *database* biologici è rappresentativo della difficoltà di questo compito: questo problema è ben illustrato dagli accesi dibattiti che circondano definizioni di termini come

“patogeno” o “metabolismo” nella banca dati di *Gene Ontology* (*The Gene Ontology Consortium*, 2019).

Le implicazioni per la biologia dei Big Data sono sostanziali. Lungi dal costituire la “morte della teoria”, lo sforzo computazionale di estrazione di Big Data (“*data mining*”) si basa su assunzioni teoretiche significative. La scelta e la definizione delle parole-chiave che vengono utilizzate per classificare e recuperare i dati hanno una rilevanza enorme per la loro successiva interpretazione. Collegare tra loro *dataset* diversi comporta prendere delle decisioni sui concetti che meglio rappresentano e indagano la natura. In altre parole, le reti concettuali associate ai dati nelle infrastrutture dei Big Data dovrebbero essere viste come teorie: modi di intendere il mondo biologico che guidano il ragionamento scientifico e direzionano la ricerca, e che vengono spesso riesaminati per render conto di nuove scoperte (Leonelli *et al.*, 2011). Il tentativo di integrare dati su larga scala rende necessario, per tutte le discipline biologiche, identificare tali teorie, e discutere delle relative implicazioni per la modellizzazione e l’analisi dei Big Data (Leonelli, 2016).

I filosofi hanno a lungo dibattuto dell’importanza teoretica della classificazione e delle pratiche di nomenclatura in biologia (Dupré, 2001), spesso in collaborazione con tassonomisti, e occasionalmente con biologi molecolari e biologi dello sviluppo. Per esempio, i ricercatori hanno attribuito una molteplicità di significati al concetto di “gene”, che i filosofi hanno documentato e articolato come parte di una più ampia analisi delle fondamenta intellettuali e delle implicazioni del “carrozzone molecolare” che ha dominato gli ultimi cinquant’anni della ricerca biologica (Griffiths e Stotz, 2013; Rheinberger e Müller-Wille, 2017). Questi studi dimostrano che i concetti biologici, anche se genericamente definiti, sono sempre parte integrante di prospettive teoretiche più ampie su come la natura funzioni (Callebaut, 2012).

Con ciò non si intende dire che la biologia dei Big Data sia completamente determinata da ipotesi pre-esistenti. Piuttosto, la biologia poggia su teorie e ipotesi esistenti, ma non permette a queste di predeterminare gli esiti della ricerca (Waters, 2007). È altrettanto importante osservare come osservazioni e misurazioni, indipendentemente dal metodo utilizzato per generarle, si collocano sempre entro uno specifico quadro di riferimento (Bogen, 2013). Non importa quanto standardizzate possano essere: le strumentazioni utilizzate per generare tali dati vengono ideate per soddisfare precise agende di ricerca (Rheinberger, 2011). Questo significa che dobbiamo riconoscere che nessun dato è “grezzo”, nel senso di essere indipendente dall’interpretazione umana. Inoltre, i dati possono essere elaborati in maniera differente. È dunque importante comprendere le scelte concettuali che hanno plasmato la produzione e la classificazione dei dati. I ricercatori che fanno uso di Big Data devono riconoscere che le strutture teoretiche, che hanno informato la produzione e l’elaborazione dei dati influenzeranno anche il loro utilizzo futuro.

Ci si potrebbe domandare se il pluralismo non sia d'ostacolo all'integrazione di dati provenienti da fonti differenti e al ricavare da questi dati una conoscenza affidabile e accurata. I filosofi della scienza hanno sostenuto che il pluralismo può addirittura risultare vantaggioso quando si tenta di ricavare conoscenza dai processi altamente complessi e variabili delle scienze della vita (Duprè, 1993; Mitchell, 2003). La frammentarietà delle tradizioni di ricerca deriva da secoli di perfezionamento degli strumenti di ricerca, perché si potesse studiare un dato processo o una specie nel maggior dettaglio possibile. Mentre questo pluralismo rende sempre più difficile generalizzare questi strumenti e la conoscenza che ne deriva (Levins, 1984; Wimsatt, 2007), garantisce tuttavia che i dati raccolti siano robusti e le inferenze accurate (Longino, 2013; Wylie, 2013). È di fondamentale importanza per la biologia dei Big Data far affidamento su questo lascito, ideando modalità di lavoro con dati di origine diversa, senza mal interpretarne la provenienza o perdendo le informazioni che i dati riescono a offrire sulla complessità della vita.

2. Valutare la qualità dei dati

I biologi sono spesso preoccupati della qualità dei dati e dei metadati che si ritrovano nei *database* online, soprattutto quando le banche dati d'interesse non sono curate da esperti dello specifico ambito disciplinare e/o dello specifico organismo. Molte banche dati non godono di servizi di revisione tra pari (*peer review*) e non hanno curatori, e qualora li avessero, le valutazioni relative alla qualità e all'affidabilità dei dati sono spesso specifiche di ogni campo di ricerca, e difficilmente possono essere applicate ad altri ambiti o ad altre tipologie di studio, anche all'interno del medesimo campo di ricerca (Floridi e Illari, 2014; Leonelli, 2017). Il rischio di perdita di qualità del dato aumenta man mano che le banche dati diventano sempre più inter-operative, poiché un più esteso collegamento tra i dati rende possibile a fonti inaffidabili di contaminare l'affidabilità complessiva delle raccolte dati online.

Questo è un altro ambito in cui il pluralismo sembra risultare problematico per la biologia dei Big Data. La mancanza di consenso riguardo alle modalità di valutazione della qualità dei dati è forse indice di una distintiva limitatezza del modo, in cui la biologia può (e dovrebbe) rapportarsi alla ricerca basata sui Big Data? Un modo per rispondere a questa domanda consiste nel sottoporre a vaglio critico la concezione dei dati su cui è fondata. Pensare che i dati siano intrinsecamente buoni o cattivi – indipendentemente dal contesto e dagli obiettivi di ricerca – significa concepirli come rappresentazioni statiche della natura, utili in quanto documentano accuratamente e con obiettività una caratteristica del mondo a un dato luogo e in un dato momento. Questa prospettiva sicuramente promuove la ricerca di modalità definitive, univer-

sali e indipendenti dal contesto, di valutare quali dati sono affidabili e quali no. Ma non tiene in considerazione il fatto che i dati sono spesso artefatti largamente rielaborati, e che risultano da interazioni col mondo ampiamente pianificate; non rende nemmeno giustizia all'osservazione che i biologi hanno visioni differenti riguardo a cosa costituisca un dato affidabile, o un dato in generale (Borgman, 2015; Leonelli, 2016). Per tale ragione, ciò che viene considerato “rumore” da una certa comunità e/o per determinati obiettivi di ricerca, può a volte rappresentare un dato per un'altra (McAllister, 2011; Loettgers, 2009; Woodward, 2010).

Sulla base di queste riflessioni, ho argomentato che i dati sono “relazionali”: in altre parole, gli oggetti che svolgono la funzione di “dati” possono variare sulla base di standard, obiettivi e metodi usati per generare, elaborare e interpretare tali oggetti come evidenze (Leonelli, 2016). Questo spiega perché valutare la qualità dei dati dipende sempre dalla ricerca che si sta conducendo. Spiega anche la riluttanza dei ricercatori a fidarsi di sorgenti di dati la cui cronologia non sia chiaramente documentata, e delle relative unità di memoria (*drive*) per raccogliere metadati sulla provenienza dei dati stessi.

I *data scientists* sottovalutano a volte l'importanza di collegare le banche dati ai campioni fisici da cui i dati sono stati originariamente ricavati (quali esemplari, tessuti, cellule e colture microbiche). È stato dimostrato che la possibilità di accedere ai campioni originali incrementa la riproducibilità dei dati e fornisce ai ricercatori maggiori opportunità di replicare esperimenti e di riutilizzare i dati (Dietrich *et al.*, 2014). L'accessibilità ai campioni originali offre inoltre un punto di contatto concreto tra le diverse tradizioni e approcci di ricerca, attraverso cui possono essere identificate ed esaminate criticamente le rispettive differenze (Leonelli e Ankeny, 2012).

Accogliere una visione relazionale dei dati significa allontanarsi da generici approcci di conservazione dei dati e muoversi verso approcci più sensibili al contesto, che includano descrittori dei dati più dettagliati, anche se questo può rallentare il ritmo della ricerca (Leonelli e Tempini, 2018). Allo stesso tempo, riconoscere la natura locale e localizzata della selezione di Big Data aiuta a comprendere quali conclusioni possono essere ricavate dall'analisi dei dati, e a valutare come generalizzare alcune inferenze particolari. Questo è particolarmente utile quando ci si ritrova a stabilire se una data estrapolazione può essere estesa da una specie a un'altra (ad esempio, dai topi agli esseri umani).

Non c'è dubbio che l'estrazione di Big Data abbia una funzione euristica potente: spesso costituisce il primo passo di ogni ricerca biologica, contribuendo a definire la direzione e la portata della ricerca stessa (Nickles, 2018). I Big Data permettono ai biologi di riconoscere *pattern* e *trend* più facilmente, e i filosofi stanno sicuramente iniziando a comprendere come l'estrazione di dati possa aiutare a esplorare, sviluppare, e verificare ipotesi

meccanicistiche (Pietsch, 2016; Ratti, 2015; Canali, 2019). Allo stesso tempo, la concezione relazionale mette in luce come l'interpretazione e l'affidabilità delle inferenze ricavate dai Big Data dipenda da due fattori cruciali: primo, un regolare confronto con altri metodi, modelli e approcci di ricerca (Elliott *et al.*, 2016); e secondo, un'attenta contestualizzazione dei dati rispetto ai cambiamenti di prospettive, metodi e obiettivi dei ricercatori (Shavit e Grieseimer, 2009). Adottare una visione relazionale dei dati significa considerare seriamente la storia dei dati, che è carica di valori e di teoria. La prospettiva relazionale promuove inoltre l'impegno volto a documentare tale storia all'interno delle banche dati, dimodoché futuri fruitori dei dati possano stabilire autonomamente, e in base ai propri standard, la qualità dei dati. Un caso esemplificativo è una recente collaborazione tra un tassonomista e un filosofo sul valore dell'ambiguità nelle etichette utilizzate per i dati nella ricerca sulla biodiversità (Sterner e Franz, 2017).

L'analisi automatizzata dei dati è una prospettiva entusiasmante per la scoperta biologica. Lungi dal rendere il giudizio umano superfluo, la crescente forza computazionale degli algoritmi richiede una proporzionale crescita in termini di pensiero critico. Collaborazioni tra filosofi e biologi possono promuovere una riflessione di fondamentale importanza su quali parti della navigazione e dell'integrazione tra dati debbano essere condotte con l'ausilio di algoritmi, e come debbano essere interpretati i risultati. La collaborazione tra filosofi e bioinformatici (ed altre tipologie di *data scientists*) è in grado di promuovere lo sviluppo di infrastrutture per i dati che possano opportunamente registrarne la provenienza, e di incoraggiare gli utilizzatori a valutare la loro qualità e rilevanza in relazione alle singole domande di ricerca¹.

Abstract

The use of big data is transforming the life sciences and ushering in new ways of doing research, which can complement and strengthen existing methods. Here, I argue that big data biology also raises fundamental questions in the philosophy of science: for example, what is a good dataset, how can reliable knowledge be extracted from big data and what are the repercussions of relying on existing data for future work? Collaborations between biologists, data scientists and philosophers of science can play an essential role in finding answers to these and other questions.

1. Questa ricerca è stata generosamente finanziata dall'European Research Council, grant numero 335925.

Bibliografia

- Bogen J. (2013), *Theory and observation in science*, «The Stanford Encyclopedia of Philosophy». <http://plato.stanford.edu/archives/spr2013/entries/science-theory-observation> [Accessed March 22, 2019].
- Borgman C. (2015), *Big Data, Little Data, No Data*, Mit Press.
- Callebaut W. (2012), *Scientific perspectivism: a philosopher of science's response to the challenge of big data biology*, «Studies in History and Philosophy of Science, Part C: Studies in History and Philosophy of Biological and Biomedical Sciences», 43, pp. 69-80. Doi: <https://doi.org/10.1016/j.shpsc.2011.10.007>, Pmid: 22326074
- Canali S. (2019), *Evaluating evidential pluralism in epidemiology: mechanistic evidence in exposome research*, «History and Philosophy of the Life Sciences», 41, p. 4. Doi: <https://doi.org/10.1007/s40656-019-0241-6>, Pmid: 30756196
- Dietrich M.R., Ankeny R.A., Chen P.M. (2014), *Publication trends in model organism research*, «Genetics», 198, pp. 787-794. Doi: <https://doi.org/10.1534/genetics.114.169714>, Pmid: 25381363
- Dupre J.A. (1993), *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*, Harvard University Press.
- Dupré J. (2001), *In defence of classification*, «Studies in History and Philosophy of Science, Part C: Studies in History and Philosophy of Biological and Biomedical Sciences», 32, pp. 203-219. Doi: [https://doi.org/10.1016/S1369-8486\(01\)00003-6](https://doi.org/10.1016/S1369-8486(01)00003-6)
- Elliott K.C., Cheruvilil K.S., Montgomery G.M., Soranno P.A. (2016), *Conceptions of good science in our data-rich world*, «BioScience», 66, pp. 880-889. Doi: <https://doi.org/10.1093/biosci/biw115>, Pmid: 29599533
- Floridi L., Illari P. (2014), *The Philosophy of Information Quality*, Springer.
- Griffiths P., Stotz K. (2013), *Genetics and Philosophy: An Introduction*, Cambridge University Press. Doi: <https://doi.org/10.1017/CBO9780511744082>
- Kellert S.H., Longino H.E., Waters C.K. (2006), *Introduction: the pluralist stance*, in Kellert S.H., Longino H.E., Waters C.K. (Eds.), *Scientific Pluralism*. University of Minnesota Press.
- Leonelli S., Diehl A.D., Christie K.R., Harris M.A., Lomax J. (2011), *How the gene ontology evolves*, «BMC Bioinformatics», 12, p. 325. Doi: <https://doi.org/10.1186/1471-2105-12-325>, Pmid: 21819553
- Leonelli S. (2012), *When humans are the exception: cross-species databases at the interface of biological and clinical research*, «Social Studies of Science», 42, pp. 214-236. Doi: <https://doi.org/10.1177/0306312711436265>, Pmid: 22848998
- Leonelli S. (2016), *Data-Centric Biology: A Philosophical Study*, Chicago University Press.
- Leonelli S. (2017), *Global data quality assessment and the situated nature of “best” research practices in biology*, «Data Science Journal», 16, p. 32. Doi: <https://doi.org/10.5334/dsj-2017-032>
- Leonelli S., Ankeny R.A. (2012), *Re-thinking organisms: the impact of databases on model organism biology*, «Studies in History and Philosophy of Science, Part C:

- Studies in History and Philosophy of Biological and Biomedical Sciences», 43, pp. 29-36. Doi: <https://doi.org/10.1016/j.shpsc.2011.10.003>
- Leonelli S. (2018), *La Ricerca scientifica nell'Era dei Big Data*, Meltemi.
- Leonelli S., Tempini N. (2018), *Where health and environment meet: the use of invariant parameters in big data analysis*, *Synthese*, 29. Doi: <https://doi.org/10.1007/s11229-018-1844-2>
- Levins R. (1984), *The strategy of model building in population biology*, in Sober E. (Ed.), *Conceptual Issues in Evolutionary Biology*, MIT Press, pp. 18-27.
- Loettgers A. (2009), *Synthetic biology and the emergence of a dual meaning of noise*, «Biological Theory», 4, pp. 340-356. Doi: https://doi.org/10.1162/BIOT_a_00009
- Longino H. (2013), *Studying Human Behaviour*, University of Chicago Press.
- McAllister J.W. (2011), *What do patterns in empirical data tell us about the structure of the world?*, «Synthese», 182, pp. 73-87. Doi: <https://doi.org/10.1007/s11229-009-9613-x>
- Mitchell S. (2003), *Biological Complexity and Integrative Pluralism*, Cambridge University Press. Doi: <https://doi.org/10.1017/CBO9780511802683>
- Nickles T. (2018), *Alien reasoning: is a major change in scientific research underway?*, «Topoi», 16. Doi: <https://doi.org/10.1007/s11245-018-9557-1>
- O'Malley M.A., Soyer O.S. (2012), *The roles of integration in molecular systems biology*, «Studies in History and Philosophy of Science, Part C: Studies in History and Philosophy of Biological and Biomedical Sciences», 43, pp. 58-68. Doi: <https://doi.org/10.1016/j.shpsc.2011.10.006>, PMID: 22326073
- Pietsch W. (2016), *The causal nature of modeling with big data*, «Philosophy & Technology», 29, pp. 137-171. Doi: <https://doi.org/10.1007/s13347-015-0202-2>
- Ratti E. (2015), *Big data biology: between eliminative inferences and exploratory experiments*, «Philosophy of Science», 82, pp. 198-218. Doi: <https://doi.org/10.1086/680332>
- Rheinberger H.-J. (2011), *Infra-experimentality: from traces to data, from data to patterning facts*, «History of Science», 49, pp. 337-348. Doi: <https://doi.org/10.1177/007327531104900306>
- Rheinberger H.-J., Muller-Wille S. (2017), *The Gene From Genetics to Postgenomics*, University of Chicago Press.
- Shavit A., Griesemer J. (2009), *There and back again, or the problem of locality in biodiversity surveys*, «Philosophy of Science», 76, pp. 273-294. Doi: <https://doi.org/10.1086/649805>
- Sterner B., Franz N.M. (2017), *Taxonomy for humans or computers? Cognitive pragmatics for big data*, «Biological Theory», 12, pp. 99-111. Doi: <https://doi.org/10.1007/s13752-017-0259-5>
- The Gene Ontology Consortium (2019), *The gene ontology resource: 20 years and still GOing strong*, «Nucleic Acids Research», 47, pp. D330-D338. Doi: <https://doi.org/10.1093/nar/gky1055>, Pmid: 30395331
- Waters C.K. (2007), *The nature and context of exploratory experimentation: an introduction to three case studies of exploratory research*, «History and Philosophy of the Life Sciences», 29, pp. 275-284. Pmid: 18822658

- Wimsatt W. (2007), *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*, Harvard University Press. Doi: <https://doi.org/10.1007/s10539-010-9199-1>
- Woodward J. (2010), *Data, phenomena, signal, and noise*, «Philosophy of Science», 77, pp. 792-803. Doi: <https://doi.org/10.1086/656554>
- Wylie A. (2017), *How archaeological evidence bites back: strategies for putting old data to work in new ways*, «Science, Technology & Human Values», 42, pp. 203-225. Doi: <https://doi.org/10.1177/0162243916671200>

Predizione e comprensione dei fenomeni. Fisica e Machine Learning a confronto

di *Marco Zanetti**

1. Introduzione: conoscenza empirica e metodo scientifico

L'apprendimento è alla base dell'interazione tra le forme di vita più evolute e l'ambiente. In particolare l'uomo impara – sia in prima persona che da esperienze pregresse dei suoi predecessori – a sfruttare a proprio vantaggio le caratteristiche dell'ambiente e ciò che in esso è presente. L'osservazione di eventi naturali e i tentativi di modificare il proprio habitat hanno portato nel corso dei secoli a un accumulo di conoscenze *empiriche* che hanno consentito all'uomo di migliorare le proprie condizioni di vita. Tali conoscenze derivano dal constatare la presenza di correlazioni tra fenomeni. Tuttavia, è solo l'avvento del metodo scientifico e la sua applicazione che hanno portato a progressi rapidi e a uno sviluppo tecnologico costante. I cardini del metodo scientifico sono la modellizzazione matematica dei fenomeni, che consenta in merito a questi la formulazione di predizioni quantitative accurate, e la verifica rigorosa di queste ultime con esperimenti riproducibili.

I modelli scientifici ancora non falsificati dalle misure sperimentali sono caratterizzati da una *capacità predittiva* ben più elevata di quella ottenuta dalle conoscenze empiriche¹, e consentono applicazioni pratiche non realizzabili altrimenti. Le teorie scientifiche hanno poi spesso la proprietà di generalizzare ed unificare la descrizione di fenomeni apparentemente scollegati, si pensi alla caduta di una mela dall'albero e al moto di rivoluzione della terra attorno al sole. In tal senso, la conoscenza che ne consegue può dirsi di un livello più alto, può essere definita *comprensione*.

* Dipartimento di Fisica e Astronomia, Università degli Studi di Padova.

1. In questo testo 'modello', 'teoria', 'modello teorico', 'modello matematico' vengono utilizzati come sinonimi; in particolare, in Fisica tali termini possono avere accezioni diverse.

2. La Fisica Fondamentale come esempio di predittività e comprensione

Tra le varie discipline, la Fisica è tra quelle che più fedelmente incarnano il metodo scientifico. Qui ne consideriamo l'ambito più rappresentativo, ovvero la Fisica Fondamentale, che ha come oggetto di studio i costituenti fondamentali della materia e le forze elementari che ne determinano le interazioni. In un tempo relativamente breve (circa un secolo rispetto alle migliaia di anni di storia dell'uomo) si è arrivati alla formulazione di un modello matematico, chiamato "Modello Standard" (MS), che inquadra tre delle quattro interazioni fondamentali a tutt'oggi note². L'accuratezza con cui i fenomeni vengono descritti non ha uguali in alcuna altra disciplina scientifica: sono infatti molti i casi in cui il modello teorico riesce a fare delle predizioni con accuratezza molto migliore della parte per miliardo (ad esempio, il valore del momento giromagnetico dell'elettrone o i rapporti di decadimento di diverse particelle instabili). Da ciò deriva la grande *capacità predittiva* del Modello Standard: ad oggi sono state effettuate migliaia di misure di precisione riguardanti i fenomeni elementari e nessuna è stata confermata in disaccordo con le predizioni della teoria; ciò a fronte di un numero di parametri liberi molto limitato (26). Anche dal punto di vista matematico, il Modello Standard è solido e consistente, non vi sono infatti vincoli formali che ne limitino il campo di validità.

Nonostante siano molte le questioni fondamentali non inquadrate nel modello (la natura microscopica della materia oscura, l'interazione gravitazionale, la massa non nulla dei neutrini, ecc.), si può senza esitazione affermare che il MS sia una delle teorie scientifiche più di successo. Con altrettanta confidenza si può concludere che l'attuale grado di *comprensione* dei fenomeni elementari è molto elevato. I passi che hanno portato alla formulazione del MS sono quelli tipici del circolo virtuoso che caratterizza il metodo sperimentale, ovvero il susseguirsi di ipotesi e relative predizioni confrontate con le osservazioni, con queste che a loro volta ispirano correzioni o aggiunte alla teoria.

3. I modelli teorici come antitesi della ricerca di correlazioni

Oltre al costante confronto con le misure sperimentali, il successo del Modello Standard è dovuto anche ai principi che hanno guidato la sua formulazione matematica: il modello deve riflettere proprietà generali dettate

2. Si tratta delle forze elettromagnetica, nucleare debole e nucleare forte, mentre la gravità non è ancora stata inclusa con successo in una modello unificato basato sulla Meccanica Quantistica.

dal buon senso, come l'indipendenza dalle caratteristiche dell'osservatore o il rispettare principi di simmetria; le caratteristiche essenziali dei fenomeni descritti devono poi emergere evidenti, a ciò è legata la comprensione degli stessi. Tale approccio è caratteristico della Fisica in generale e viene applicato in contesti molto diversi tra loro, dalle dinamiche dell'Universo, allo studio dei sistemi biologici. Ciò dà ottimi risultati quando i sistemi studiati sono "semplici" (il MS, ad esempio, descrive l'interazione tra due corpi puntiformi), mentre è più difficile inquadrare fenomeni "complessi", quelli in cui i corpi interagenti sono moltissimi e la dinamica dipende criticamente dalle condizioni al contorno: ad esempio discipline come la Medicina o la Biologia sono molto lontane dallo sviluppare un insieme di leggi matematiche che spieghi nel dettaglio il funzionamento degli organismi viventi. Tuttavia, un modello quantitativo che faccia delle previsioni falsificabili rimane l'obiettivo principale dell'approccio scientifico ai problemi.

L'alternativa allo sviluppo teorico di un modello matematico è la ricerca di correlazioni: per cercare di capire i meccanismi che regolano un dato fenomeno, si fa affidamento a eventi che sembrano correlati, assumendo che tale correlazione sia frutto di una relazione di causa ed effetto. Se è vero però che una relazione di causa-effetto implica una correlazione, non è sempre vero il contrario; ovvero correlazione non implica causalità. Esempi di correlazioni spurie (ovvero non causali) sono ovunque; Tyler Vigen raccoglie in un sito diventato oramai famoso (Vigen, 2011), svariati esempi paradossali di tali correlazioni. In Figura 1 ne vengono riportati un paio.

Le conoscenze empiriche che ci consentono di interagire con l'ambiente si fondano sulla constatazione di correlazioni, apprese nel tempo grazie alle nostre esperienze: non tocchiamo una pentola sul fuoco a prescindere dalle nostre conoscenze di termodinamica, così come chi tira a canestro lo fa senza risolvere le equazioni di balistica della palla. La fiducia che ciascuno di noi ha nella propria esperienza ci espone tuttavia al rischio di assumere causali correlazioni che in realtà non lo sono. Per capirne le ragioni, consideriamo l'approccio bayesiano alla statistica:

$$p(\text{modello}|\text{dati}) \propto p(\text{dati}|\text{modello}) p(\text{modello})$$

Posterior probability Likelihood Prior probability

dove si evidenzia come il nostro grado di confidenza in merito all'affidabilità di un certo modello ("posterior probability") dipende sia da quanto bene i dati siano in accordo con le previsioni del modello ("likelihood"), ma anche dal preconcetto soggettivo che abbiamo riguardo al modello stesso ("prior probability").

Gli esempi di Figura 1 risultano paradossali. Considerando ad esempio il primo dei due, la probabilità che assegniamo a priori a un modello in cui la spesa statunitense in scienza, spazio e tecnologia sia causalmente correlata al numero di suicidi per impiccagione è essenzialmente nulla; perciò anche dopo aver osservato i dati, tale probabilità, proporzionale al prodotto di prior e likelihood, rimane molto bassa. Il problema sorge quando in merito a un certo fenomeno la nostra prior è “piatta”, ovvero non abbiamo informazioni *ab origine* o preconcetti: in quei casi una likelihood che esibisce una correlazione ci porta a credere che tale correlazione non sia casuale. Per mettersi al riparo serve dunque avere dei preconcetti corretti, ovvero una solida conoscenza dei fenomeni, cosa che risulta impossibile per tutti i fenomeni con cui potremmo doverci confrontare.

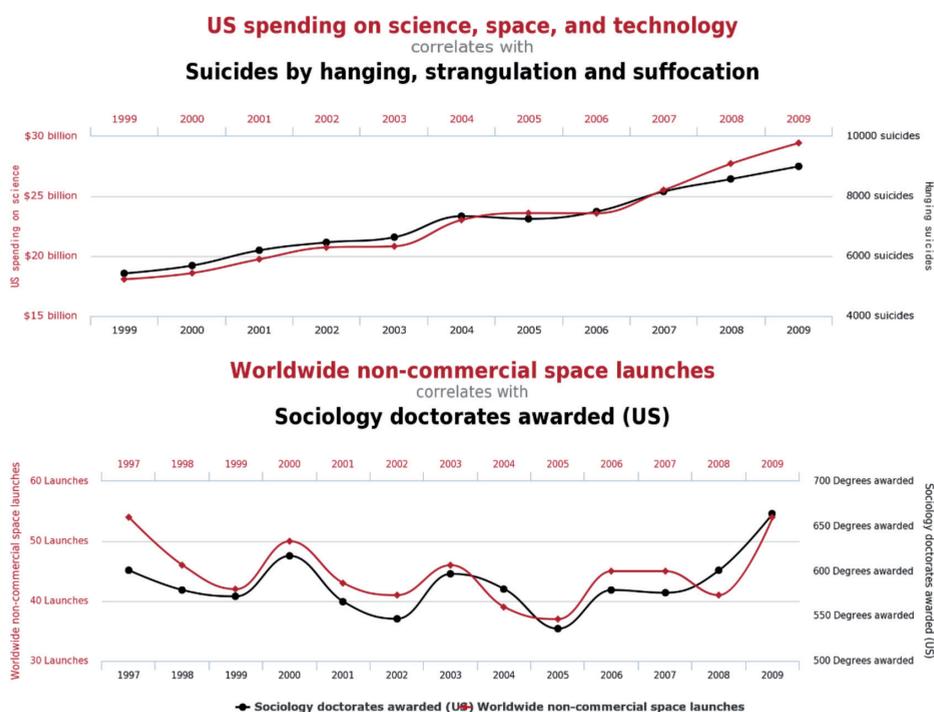


Fig. 1 - Esempi di correlazioni paradossali, dal sito edito da Tyler Vigen (ref)

Oltre a correlazioni accidentali, altrettanto problematico è il caso in cui i dati non rendono manifesta una relazione di causa-effetto in realtà presente nel fenomeno osservato. Ciò si verifica quando la raccolta dei dati è condizionata da “rumore”, ovvero quando la significanza statistica del campione ana-

lizzato è piccola o la risoluzione dello strumento utilizzato per la misura non è sufficiente. In generale dedurre dai dati il modello matematico sottostante è un problema estremamente complesso, se non addirittura difficile da porre in termini statistici. Randall Munroe, l'autore dei fumetti satirici-scientifici XKCD (Munroe, 2015), nella striscia in Figura 2, esemplifica questo problema con ironia ma nel suo essenziale.

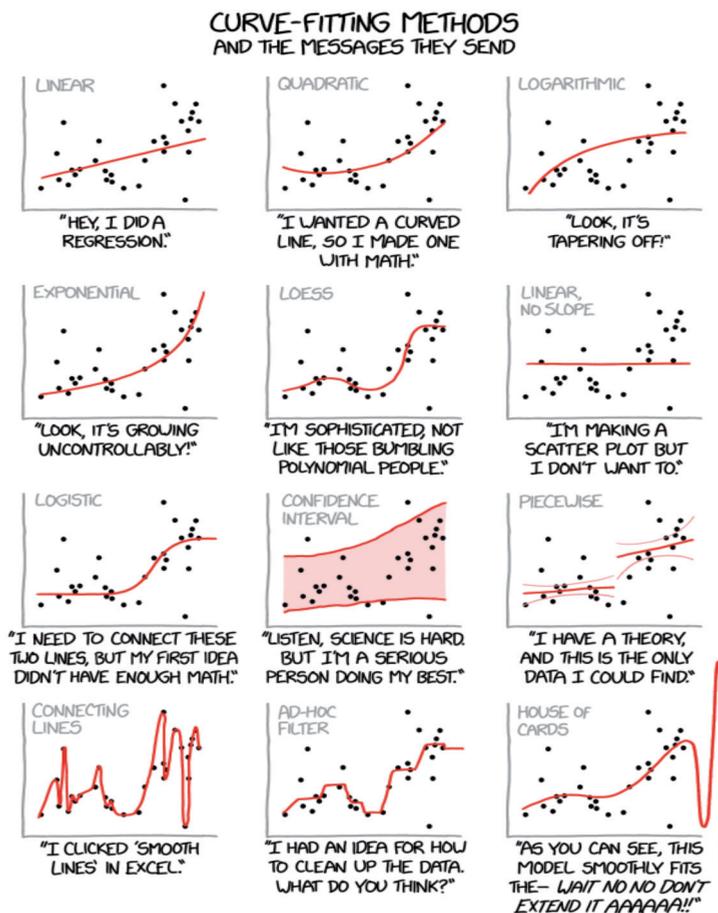


Fig. 2 - Esempi delle varie opzioni che si hanno quando si cerca di estrarre un modello da dati "rumorosi", da XKCD

Quali delle curve in figura – ognuna rappresentante un possibile modello – descrive ("fitta", per usare un inglesismo) meglio la distribuzione dei dati? Come evidenziato nella striscia, la scelta del modello non è univoca e

può portare a conclusioni completamente opposte! Quanto appena discusso rende evidente la difficoltà di estrarre informazioni dai dati senza un modello teorico che faccia da guida. L'approccio che risulta perciò più efficiente nello sviluppo della conoscenza consiste nel cercare un legame di causa-effetto solo dove tale legame viene predetto da un modello teorico ben giustificato, in sostanza dove la prior bayesiana ha valori non troppo distanti dall'unità. In tali condizioni, la teoria suggerisce quali siano le misure da compiere, quali dati analizzare e sotto quali ipotesi, innescando appunto il circolo virtuoso proprio del metodo scientifico.

Quando si arriva ad ottenere un modello teorico che descrive in maniera generalizzata sotto uno schema comune varie tipologie di fenomeni, si ha la sensazione che di tali fenomeni sia stata colta l'essenza; in effetti una definizione empirica di *comprensione* potrebbe essere proprio questa. La comprensione dei fenomeni, oltre ad essere di per sé un'innata pulsione dell'uomo, è di fondamentale importanza per la società, rendendo possibile un efficiente sviluppo tecnologico. Le capacità *predittive* di un modello teorico infatti consentono di sfruttare i fenomeni a proprio vantaggio; le predizioni del modello possono cioè essere "ingegnerizzate". Il progresso straordinario che l'umanità sta vivendo dal diciannovesimo secolo è reso possibile solamente grazie alle conoscenze prodotte dall'applicazione del metodo scientifico.

4. La deriva della Fisica Fondamentale: dalla bellezza alla perdita di predittività

Nel paragrafo precedente abbiamo discusso il ruolo essenziale dei modelli teorici nello sviluppo della conoscenza e nel progresso tecnologico. Torniamo ora a prendere in considerazione il caso della Fisica Fondamentale, il cui soggetto di studio ha senz'altro condizionato l'approccio utilizzato per sviluppare i relativi modelli: è del tutto lecito infatti assumere che i fenomeni fondamentali della Natura siano regolati da leggi non involute o inutilmente complicate, ma che anzi risultino in qualche misura semplici ed eleganti. Ciò ha trovato pieno riscontro nel Modello Standard, una teoria governata dalle fondamenta da principi di simmetria. Un caso ancora più eclatante è la teoria della relatività generale che descrive la forza di gravità ed il tessuto spazio-temporale dell'universo. Tale teoria ha solamente due parametri liberi, la velocità della luce e la costante di Newton³, ed ha al tempo stesso una capacità predittiva straordinaria, sia in ambito scientifico (dalle onde gravitazionali ai buchi neri)

3. In versioni generalizzate della teoria potrebbero essere presenti altri parametri, come la costante cosmologica.

che applicativo (Gps). Il suo sviluppo non ha essenzialmente richiesto input sperimentale, erano infatti famosi i “Gedankenexperiment”, di cui Einstein faceva uso per argomentare le assunzioni alla base della teoria; si pensi in particolare all’equivalenza tra un corpo sulla terra e a un altro in un razzo in accelerazione nello spazio profondo (principio di equivalenza).

Il successo di questi sviluppi teorici ha portato ad avere grande confidenza sulle capacità dell’uomo di cogliere l’essenza delle Leggi Fondamentali. Negli anni si è fatta strada la convinzione che il nostro innato senso di *bellezza* possa con efficacia farci da guida per scoprire l’intimo della Natura. Tale convinzione ha fatto presa sulla comunità scientifica, ma si è imposta anche nella cultura popolare e nella percezione che la società ha della Fisica; molti sono i libri divulgativi di successo che hanno passato questo messaggio, si possono citare ad esempio best-seller come “L’universo elegante” di Brian Green o “La Teoria del tutto” di Stephen Hawking:

Physicists have come to realize that mathematics, when used with sufficient care, is a proven pathway to truth (Green, 2003).

My goal is simple. It is a complete understanding of the universe, why it is as it is and why it exists at all (Hawking, 2002).

Se da un lato tali considerazioni risultano esaltanti, infondendo una fiducia straordinaria sulle capacità intellettive dell’uomo, dall’altro ci espongono al rischio di una pericolosa deriva, ovvero di concentrarsi troppo su questioni estetiche, sottovalutando il ruolo delle misure sperimentali. Nel seguito vengono discussi i casi più rappresentativi della moderna Fisica Fondamentale, l’esplorazione agli acceleratori di particelle dei fenomeni microscopici ed i tentativi di formulare una teoria quantistica della gravità.

Ricerche di estensioni del Modello Standard a Lhc

L’acceleratore LHC del Cern è attualmente lo strumento più potente per l’esplorazione del mondo microscopico. In operazione dal 2010, ha avuto fin da subito un grande successo scientifico (e mediatico), esperimenti di Lhc infatti sono stati i responsabili della scoperta del bosone di Higgs, l’elemento essenziale del Modello Standard predetto negli anni Settanta, ma di cui si è avuta conferma per la prima volta appunto solo nel 2012. Con la conferma dell’esistenza del bosone di Higgs, il MS risulta una teoria matematicamente consistente, priva di limiti di validità intrinseci. La particella di Higgs stessa però fa sospettare l’incompletezza del modello, in particolare il valore della sua massa rende quest’ultimo non “naturale”, una questione nota come

problema di gerarchia. Il valore misurato della massa del bosone di Higgs (ovvero la sua energia a riposo) richiederebbe un aggiustamento ritenuto eccessivamente fine se la validità delle previsioni del MS si estendesse a scale di energia molto più alte. Di per sé, tale aggiustamento fine non inficia né la consistenza né la predittività del modello, ma, appunto principi di “bellezza” o “naturalzza”, spingono a credere che deviazioni dalle predizioni del MS debbano manifestarsi già a scale di energia prossime a quella della particella di Higgs. In particolare, tali fenomeni non inquadrati dal MS dovrebbero essere prodotti da LHC.

Sono stati formulati molti modelli teorici che estendono il MS, tutti sviluppati per risolvere il problema di gerarchia. Una classe di modelli, la Supersimmetria (SUSY), era in particolare ritenuta particolarmente promettente: oltre infatti a fornire una spiegazione naturale per il valore della massa della particella di Higgs, la Supersimmetria implica l’unificazione delle tre interazioni fondamentali del MS (ovvero il raggiungimento ad alte energie di un valore comune delle costanti di accoppiamento caratteristiche delle tre forze) e fornisce anche un candidato alla interpretazione microscopica della Materia Oscura. La prior bayesiana per SUSY era, a dire il vero, molto vicina ad uno; gran parte della comunità era essenzialmente certa che si sarebbe osservata a LHC⁴. Purtroppo, né la Supersimmetria, né alcun altro modello alternativo sono stati confermati dagli esperimenti. Alcuni dei risultati di tali ricerche sono mostrati in Figura 3.

I dettagli della Figura 3 sono volutamente di difficile lettura, il messaggio che si vuole passare è un altro; si vogliono mettere in evidenza infatti quelli che sono i veri problemi che da quelle ricerche emergono. Innanzitutto, i risultati sono espressi come “limiti”, ovvero i modelli di per sé non sono falsificati, ma sono stati posti solamente dei vincoli sui rispettivi intervalli di validità: quasi tutti questi modelli hanno un grande numero di parametri liberi, giocando con i quali si può regolare la scala di energia alla quale le previsioni alternative al MS entrano in gioco. La falsificabilità di un modello è però legata alla sua qualità scientifica: un modello che non è possibile falsificare è inutile perché *non predittivo* e perché non porta ad una migliore *comprensione* dei fenomeni.

Dalla Figura 3 si intuisce poi che i modelli alternativi al MS sono molti, la Supersimmetria stessa è declinabile in moltissime forme diverse. Il lavoro di verifica sperimentale è stato fatto secondo la prescrizione standard, ovvero, come menzionato nel capitolo precedente, facendosi guidare dal modello sulla tipologia di dati da guardare e sulle modalità della loro analisi. Se tuttavia le

4. Sono diverse le testimonianze dirette (compresa quella dell’autore) di fisici dall’altissimo profilo scientifico che sull’esistenza della SUSY hanno scommesso parti del proprio corpo funzionali alla riproduzione.

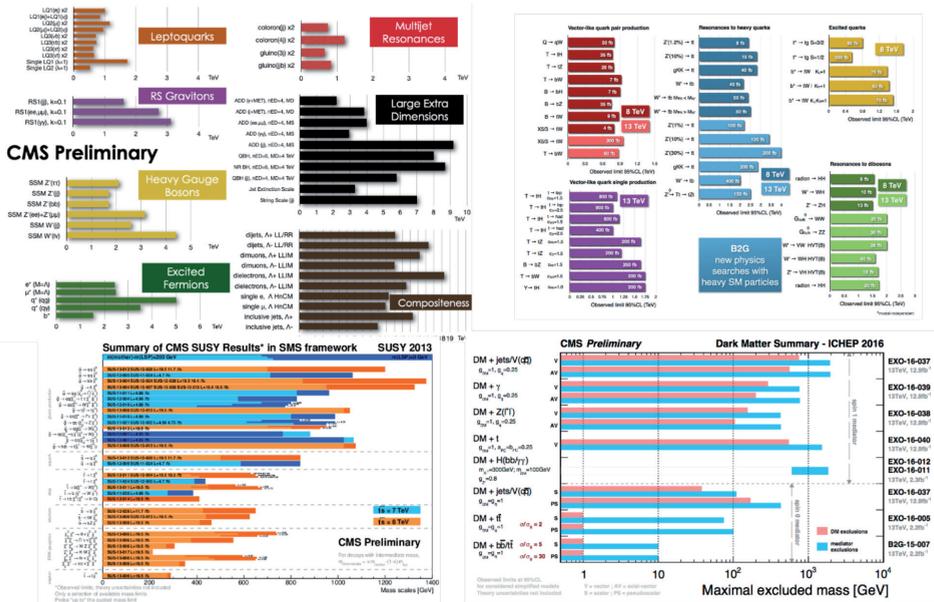


Fig. 3 - Alcuni dei risultati ottenuti dall'esperimento Cms a Lhc sulle ricerche di modelli estensione del Modello Standard

alternative sono molte (troppe), il pregiudizio (prior) sui singoli modelli diminuisce e nella pratica la guida teorica perde di efficacia.

Lo stato dell'arte non lascia che due scenari possibili: o la validità del Modello Standard si estende ben oltre la scala di energia sondata da LHC (rendendo il problema di gerarchia un fatto prettamente estetico e dunque irrilevante) oppure il modello che correttamente estende il MS non è ancora stato trovato nei dati. Un episodio, in particolare, risulta illuminante per dimostrare quanto questo secondo scenario non sia da trascurare: nel 2015 i due principali esperimenti di LHC mostrarono entrambi una deviazione dalle predizioni del MS nella stessa regione dello spazio delle fasi, ovvero qualcosa che poteva essere interpretato come un segnale di nuova fisica. L'analisi dei dati era considerata solida, essendo del tutto analoga a quella utilizzata per scoprire il Bosone di Higgs 3 anni prima (ricerca di un picco nello spettro di massa invariante di una coppia di fotoni). I risultati delle due analisi sono confrontabili nei grafici di Figura 4.

La reazione dei membri delle collaborazioni sperimentali è stata tiepida, per non dire scettica, non per possibili dubbi sulla qualità dell'analisi, ma perché non era noto alcun modello teorico che prevedesse un nuovo segnale di quel tipo (la mancanza di guida teorica ha implicato giustamente una prior

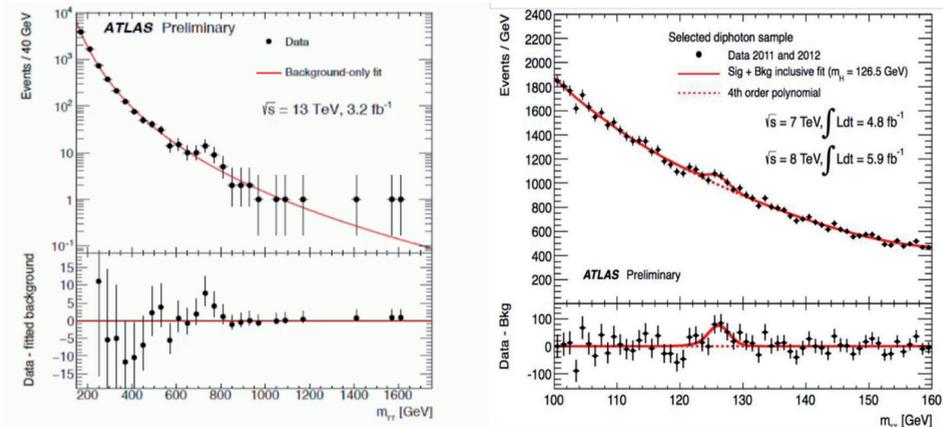


Fig. 4 - Esempi di segnali nello spettro di massa di una coppia di fotoni. A sx la fluttuazione osservata nel 2015, a dx una delle evidenze che ha portato alla scoperta del Bosone di Higgs nel 2012

collettiva piatta). La comunità di fisici teorici (gli esperti nella formulazione di modelli) ha invece reagito con grandissimo entusiasmo, producendo in brevissimo tempo una vasta gamma di nuovi modelli che inquadrassero tale segnale (Backovic, 2016). Ciò ha dimostrato come le pure numerose teorie estensione del MS finora formulate, siano lontane dall'esaurire tutte le possibilità! Ad onor di cronaca, il segnale si dimostrò in seguito una fluttuazione statistica. A tutt'oggi questa rimane una delle pochissime analisi condotte dagli esperimenti senza una iniziale guida teorica.

La Quantizzazione della Gravità

Veniamo ora alla questione più controversa relativa agli sviluppi degli ultimi decenni della Fisica Fondamentale. Nonostante siano entrambe due teorie di grandissimo successo, il Modello Standard e la Teoria della Relatività Generale (RG) non hanno alcun punto in comune. Questo è dovuto al fatto che le scale di energia proprie delle due teorie sono completamente diverse, la scala della RG (massa di Planck, circa 10^{19} GeV/c²) è di molti ordini di grandezza superiore alla scala del MS. Ciò rende estremamente complicato inquadrare le tue teorie in una descrizione unificata. Allo stesso tempo è impensabile al momento realizzare un esperimento per sondare direttamente la scala di Planck: l'attuale strumento più potente, LHC (che arriva a sondare i fenomeni fisici ad energie dell'ordine di 10^4 GeV), è stato realizzato in non meno di 15 anni grazie all'impegno di migliaia di scienziati e con un investi-

mento di diversi miliardi di dollari; a meno di rivoluzioni tecnologiche, non è ovvio quanto più in là si possa ancora estendere la frontiera dell'energia. La questione però è della massima rilevanza, probabilmente la più importante della Fisica Fondamentale. È per questo che gli sforzi per formulare una teoria quantistica della gravità continuano senza perdere vigore da decine di anni. Lo schema ritenuto più promettente e più noto al pubblico è la Teoria delle Stringhe.

Fin dai primi sviluppi la difficoltà del fare predizioni confrontabili coi dati è risultata evidente; tuttavia i progressi della teoria, misurati in termini di consistenza matematica e, ancora una volta, di "bellezza", hanno generato uno slancio che ancora oggi non si è spento. Lee Smolin a tal proposito racconta:

There was a sense that the one true theory had been discovered. Nothing else was important or worth thinking about. Seminars devoted to string theory sprang up at many of the major universities and research institutes. At Harvard, the string theory seminar was called the Postmodern Physics seminar. This appellation was not meant ironically (Smolin, 2006).

In totale antitesi con le conoscenze euristiche acquisite sulla base dell'esperienza, la teoria delle stringhe o le sue alternative sono un esempio in cui la prior soggettiva su un modello – costruita appunto sulla base di criteri di consistenza e bellezza matematica – conta più del termine di likelihood nel processo di definizione della posterior probability, anzi tale termine, ovvero il confronto del modello con i dati, viene completamente ignorato. Paradossalmente, alla base di questa deriva c'è lo stesso tipo di implicazione logica fallace riportata nella sezione 3: se è plausibile che la teoria corretta sia bella, non è detto che una teoria bella sia vera. La bellezza è l'archetipo di attributo soggettivo: seppur sia possibile convenire sulla bellezza di una cosa, è certamente impossibile mettere tutti d'accordo su quale sia la più bella di tutte! La teoria delle Stringhe e in generale tutte le teorie finalizzate a riconciliare meccanica quantistica e relatività generale sono lungi dall'essere pratiche inutili, avendo prodotto innumerevoli sviluppi in Matematica e trovando applicazioni nei contesti più diversi, in particolare nella descrizione di stati della materia fortemente accoppiati.

Per quanto riguarda invece la vera finalità di questi studi, come c'era da aspettarsi, i risultati non hanno rispettato le aspettative (né giustificato gli investimenti in termini di capitale umano e finanziario): il metodo scientifico è nei fatti non più applicato; la teoria perde ogni capacità predittiva; la comprensione dei fenomeni fondamentali diventa una mera illusione. Basare gli sviluppi di una teoria unicamente sul gusto personale senza avere un confronto con i dati è risultato non proficuo. La vera questione a questo punto è se

abbiamo effettivamente raggiunto il limite delle nostre potenzialità intellettive e tecnologiche o se esistono maniere alternative per speculare sui fenomeni fondamentali.

5. La riscossa della Scienza dei Dati

Nella sezione 3 abbiamo discusso i limiti dell'estrarre dai dati informazioni di alto livello, in particolare relazioni causali, senza avere un modello matematico che faccia da guida. Va sottolineato ancora però che in alcuni contesti, *l'esperienza* risulta più efficace dell'applicazione di una teoria scientifica: ciò vale ad esempio per qualsiasi nostro movimento nello spazio, ma anche per l'interpretazione di ciò che percepiscono i nostri sensi. Paradossalmente ciò che viene associato al concetto di intelligenza – e che ci differenzia dagli altri animali – è proprio la capacità di cogliere le correlazioni, di sviluppare conoscenza empirica dei fenomeni. Non è perché abbiamo sviluppato il Modello Standard o le altre teorie scientifiche che ci definiamo intelligenti, tale definizione sussiste a prescindere.

I limiti della nostra esperienza sono tuttavia intrinseci alla nostra fisiologia, in particolare la quantità di informazione che riusciamo a processare così come la velocità a cui lo facciamo sono naturalmente limitate. Cosa succederebbe se tali limiti fossero in qualche maniera superati? Non ci stiamo ovviamente riferendo a congetture balzane in stile Dianetics, ma alla possibilità di utilizzare a tal fine calcolatori elettronici, ovvero di sviluppare una “Intelligenza Artificiale” (AI).

La versione moderna dell'AI, quella che sta avendo successi straordinari in moltissime applicazioni, si fonda sull'idea di replicare il processo dell'accumulo di esperienze, ovvero dell'apprendimento. Oggigiorno “apprendimento automatico” (Machine Learning, ML) e AI sono essenzialmente sinonimi.

I calcolatori seppur intrinsecamente “stupidi”, hanno capacità di elaborazione e di stoccaggio di informazioni ordini di grandezza superiori a quelle umane. La differenza è che l'uomo, fin dalla nascita, impara costantemente sulla base degli esempi che gli vengono forniti e sulle esperienze che accumula. L'obbiettivo è dunque di sviluppare un metodo che consenta ai calcolatori di fare lo stesso.

Al termine del processo di apprendimento si dovrebbe ottenere l'analogo di quello che per l'uomo abbiamo chiamato modello empirico; naturalmente per un calcolatore quest'ultimo non può che essere altro che una funzione matematica $f(\#)$:

$$f(x; w) = y$$

Dove x corrisponde ai dati di input, una lista di n numeri (in termini matematici è un vettore n -dimensionale), y è un numero (continuo o discreto), l'output, e w è il vettore di parametri. La funzione (matematicamente uno scalare $f: \mathbb{R}^n \rightarrow \mathbb{R}$) costituisce un modello, ovvero esprime la relazione di causa-effetto tra x e y . Sono moltissimi i problemi che possono essere espressi in questi termini: ad esempio x potrebbe essere l'insieme dei valori dei pixel di una foto (per ciascun pixel, un valore tra 0 e 1 per indicare la scala di grigio se la foto è in bianco o nero, oppure 3 valori tra 0 e 1 per le tre scale di rosso verde e blu se la foto è a colori) e y potrebbe essere una variabile discreta i cui valori etichettano delle categorie, ad esempio 0 per indicare la categoria "cane" ed 1 la categoria "gatto". In questo modo il modello, una volta "allenato", è in grado di classificare le fotografie a seconda rappresentino cani o gatti. In modo analogo è possibile allenare dei modelli di ML per interpretare i comandi vocali o leggere un manoscritto, oppure, come vedremo in seguito, distinguere un segnale di un processo fisico rispetto al rumore di fondo.

A tutti gli effetti, anche in questo caso l'obiettivo è quello di sviluppare un corretto modello matematico, qual è dunque la differenza con i modelli matematici delle discipline scientifiche? La differenza sta nel tipo di funzione matematica che implementa la relazione di causalità del fenomeno studiato: nel caso del Machine Learning le funzioni sono molto particolari, involute e caratterizzate da un numero enorme di parametri (anche decine di migliaia). L'esempio più noto di tali funzioni sono le reti neurali artificiali, ovvero un tentativo di realizzare matematicamente lo schema di funzionamento del cervello, con neuroni che comunicano tra loro e si attivano se ricevono un segnale sopra una determinata soglia. Senza entrare nel dettaglio, va sottolineato che grazie alle loro proprietà, in particolare al grande numero di parametri, tali funzioni possono approssimare qualsiasi data funzione (analitica o meno).

Il processo di apprendimento è l'aspetto cruciale delle applicazioni di ML. In sostanza consiste nel fornire degli esempi di quale dovrebbe l'output dato un certo input, ovvero coppie (x,y) . Questi esempi vengono poi utilizzati per trovare i valori dei parametri che minimizzano la distanza tra l'output del modello date le x e le rispettive y . Ci sono molti modi di definire tale distanza, ad esempio i migliori parametri potrebbero essere definiti tramite:

$$\min_w \sum_{i=1}^N [f(x_i; w) - y_i]^2$$

La rivoluzione sopra citata è avvenuta grazie a tre principali fattori che hanno reso l'apprendimento, codificato nella formula di cui sopra, estrema-

mente efficiente: innanzitutto il numero degli esempi (N nella sommatoria, da non confondersi con il sopra citato n , la dimensione del vettore di input) da fornire al calcolatore, che può essere grande praticamente a piacere vista la disponibilità di dati, soprattutto dalla rete internet; la velocità di elaborazione dei calcolatori (in particolare utilizzando Gpu) che consente di analizzare la grande mole di esempi fornita; le tecniche matematiche di minimizzazione che consentono di esplorare con efficacia spazi dei parametri ad elevatissima dimensionalità⁵.

L'analizzare enormi quantità di dati (Big Data), fa sì che la probabilità di cadere nella trappola delle correlazioni spurie sia drasticamente ridotta. Infatti, false relazioni di causa-effetto sono destinate a sparire all'aumentare della taglia del dataset considerato (ovvero al crescere della sua significanza statistica), così come è successo al segnale poi non confermato dell'aneddoto riportato nel paragrafo 4.1. Al crescere degli esempi forniti, diminuisce anche l'effetto del rumore, cosicché la relazione causale sottostante risulta più evidente.

Ciò spiega il successo del Machine Learning, che ha rivoluzionato innumerevoli applicazioni non solo pratiche (dall'auto a guida autonoma alla finanza), ma anche scientifiche. Anche in Fisica il Machine Learning viene sempre più utilizzato. Una sfida pubblica lanciata qualche anno fa (Kaggle, 2014) chiedeva ai concorrenti di classificare, in un dataset sintetico, gli eventi di collisione di LHC; in tale dataset erano presenti processi coinvolgenti il bosone di Higgs e i rispettivi processi di rumore. A tale sfida hanno partecipato sia esperti di fisica delle particelle (inclusi alcuni degli stessi responsabili della scoperta della particella di Higgs), ma anche degli "scienziati dei dati" assolutamente ignoranti di fisica. I primi hanno applicato le loro conoscenze di fisica per analizzare i dati, i secondi tecniche avanzate di ML. I risultati sono stati sorprendenti, il team di fisici che si è meglio classificato non ha superato la venticinquesima posizione⁶!

Come menzionato precedentemente, l'obiettivo del processo di apprendimento non è tuttavia di imparare la vera relazione di causalità tra input ed output, ma di approssimarla al meglio. L'approccio è estremamente pragmatico, l'unica cosa che conta è essere il più predittivi possibile.

5. Quello descritto in questo paragrafo va sotto il nome di allenamento "supervisionato" in quanto, durante questa fase, gli esempi forniti includono il valore di output corrispondente ad ogni input. Esistono versioni in cui l'allenamento è solo parzialmente supervisionato o non supervisionato; in questi casi il modello deve rendersi conto autonomamente della presenza di diverse categorie nei dati.

6. Ad onor del vero, i dataset forniti ai partecipanti della competizione erano estremamente semplificati; contenevano infatti solo informazioni di alto livello e corrispondenti a una frazione piccola di quella prodotta dal rivelatore (<1kByte per evento contro i 2MB di un evento "grezzo"). Il grosso del lavoro di analisi dei fisici consiste proprio nell'elaborare l'output del rivelatore, ovvero nella "data reduction" nel gergo della scienza dei dati.

Per illustrare meglio la questione descriviamo come vengono applicati i modelli di ML e come ne vengono giudicate le performance. Facciamo riferimento ai grafici di Figura 5. Assumiamo il caso di classificazione di due categorie che chiameremo “segnale” e “rumore”. Per ogni evento fornitogli il modello produce un output (mostrato nel grafico di sinistra), eventi di tipo “segnale” tendono ad avere valori di y più vicini a 1 mentre eventi di “rumore” più vicini a 0. Più la separazione tra gli eventi di segnale e rumore è netta, migliore è la loro classificazione; la distinzione non è mai perfetta, per classificare le performance del modello si utilizza la cosiddetta curva “Receiver Operating Characteristic” (Roc), ovvero l’efficienza di selezione degli eventi di segnale contro la reiezione degli eventi di rumore al variare della soglia applicata all’output del modello (grafico di destra). Un modello che assegna casualmente il valore di y agli eventi corrisponderebbe ad una diagonale inclinata di 45 gradi, mentre un modello perfettamente discriminante presenterebbe una curva completamente squadrata; per quantificare le capacità predittive di un modello ML tipicamente si utilizza il valore dell’area sotto la Roc.

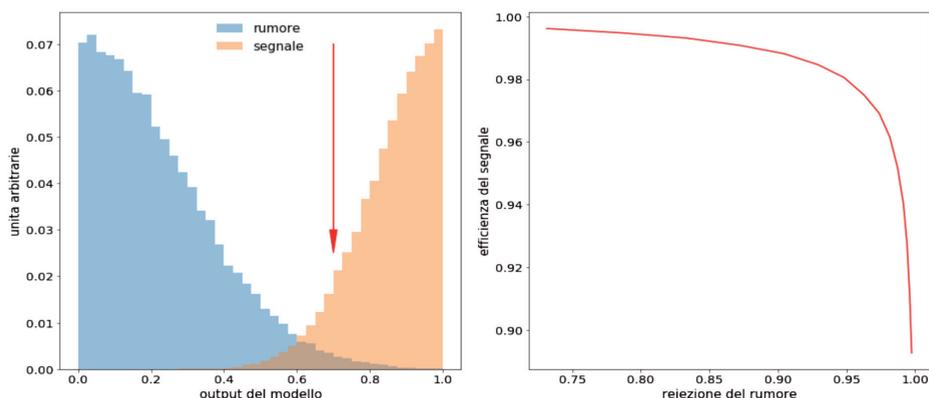


Fig. 5 - A sx l’output di un modello ML per gli eventi di due categorie, segnale e rumore; a dx la Roc corrispondente

Quasi mai si ottiene una classificazione perfetta, ma d’altra parte anche noi umani a volte fraintendiamo ciò che ci viene detto (a livello sintattico) o non riconosciamo un oggetto in un’immagine. Quello che conta è avere una frequenza di errore bassa; per la verità i modelli di ML hanno performance superiori all’uomo in moltissimi contesti, dal riconoscimento facciale agli scacchi. Rimane però il problema che il ML non porta ad alcuna *comprensione*. I modelli ML rimangono a tutt’oggi delle scatole nere.

6. Conclusioni e prospettive

La sfida dell'uomo alla comprensione dei fenomeni a un loro utilizzo per migliorare le proprie condizioni di vita è ad un punto di svolta. Il metodo scientifico, nella sua declinazione classica, si è dimostrato cruciale per lo sviluppo della conoscenza e della tecnologia, tuttavia l'approccio più di "forza bruta" introdotto dalla moderna Intelligenza Artificiale ha dimostrato di avere grandissime potenzialità. Il caso della Fisica Fondamentale è forse la disciplina in cui il contrasto tra i due approcci è più evidente: l'entusiasmo legato alle capacità dell'uomo di cogliere la *bellezza* delle leggi di Natura ha portato ad una deriva che stride fortemente con il sempre più importante ruolo giocato dai (big) dati.

C'è tuttavia una ragione per essere ottimisti e puntare a risolvere le difficoltà che attualmente limitano gli sviluppi in Fisica Fondamentale utilizzando tecniche ispirate al Machine Learning. Si dovrebbe andare tuttavia ben oltre l'attuale utilizzo del ML negli esperimenti, che tipicamente si riduce alla discriminazione tra i segnali e richiede un allenamento ancora basato completamente sulle previsioni dei modelli teorici. La maggior parte degli esperimenti, infatti, produce moli enormi di dati complessi; alla loro analisi guidata dai modelli teorici, potrebbe idealmente essere complementare un approccio *unbiased*, che utilizzi algoritmi di ML per ricercare nei dati schemi o strutture di alto livello, fornendo input per indirizzare gli sviluppi teorici.

In conclusione, l'obiettivo virtuoso da porsi nei prossimi anni è di non stravolgere il metodo scientifico, ma di arricchirlo e potenziarlo con l'apporto di un nuovo strumento quale il Machine Learning.

Abstract

Physics, in particular in its description of the fundamental phenomena of Nature, offers a remarkable example of the effectiveness of the scientific method, reaching astonishing accuracy in the comparison of the mathematical models and the experimental measurements. Such models have been developed by inference on the basis of little, high level information. In contrast, empirical knowledge, based on un-systematic observation of correlations among events, can lead to factious causal relations. In recent years however the amount of information digitally available in conjunction with the exponentially increasing computing power and the development of machine learning techniques, allowed a systematic chasing of correlations in the data, reducing to negligible levels the false causal connections and bringing "comprehension" of highly complex phenomena. In this contribution I will discuss how Physics could profit from this brute-force, "Big-Data" approach to further investigate the laws of Nature.

Bibliografia

- Backovic M. (2016), *A Theory of Ambulance Chasing*. Retrieved from <https://arxiv.org/abs/1603.01204>
- Green B. (2003), *The Elegant Universe*, W.W. Norton & Company.
- Hawking S.W. (2002), *The Theory of Everything*, New Millennium Entertainment.
- Kaggle (2014), *Higgs Boson Machine Learning Challenge*. Retrieved from <https://www.kaggle.com/c/higgs-boson>
- Munroe R. (2015), *XKCD*. Retrieved from <https://xkcd.com>
- Smolin L. (2006), *The Trouble with Physics: The Rise of String Theory, the Fall of a Science, and What Comes Next*, Houghton Mifflin Harcourt.
- Vigen T. (2011), *Spurious Correlations*. Retrieved from <http://tylervigen.com/spurious-correlations>

Quanta informazione può elaborare il cervello di un uomo? E quello di un'ape?

di *Aram Megighian**

Quando parliamo di Big Data o di Intelligenza artificiale, direttamente o indirettamente pensiamo sempre al nostro cervello. Esso rappresenta contemporaneamente sia un punto di riferimento costante di una capacità di calcolo e di elaborazione enorme che ci ha fornito la natura, sia un punto da oltrepassare per ottenere un qualcosa, al di là dell'immaginabile, che possa superare quello che la natura ha forgiato nel corso dell'evoluzione.

La storia degli elaboratori elettronici è relativamente recente rispetto alla storia dell'evoluzione del sistema nervoso negli organismi biologici. Eppure, in questo pur breve periodo di tempo si sono osservati dei passi giganteschi che hanno riguardato principalmente la dimensione degli elaboratori e soprattutto dei loro elementi *core*, i chip e le Cpu, e la loro capacità e velocità di calcolo. Due elementi in apparente antitesi; difficile concepire un aumento incredibile ed esponenziale della capacità di calcolo degli elaboratori elettronici, mentre al contempo si assiste a una loro progressiva e altrettanto esponenziale riduzione nelle dimensioni.

Appartengo alla generazione di coloro che da bambini hanno vissuto la conquista della luna. Moltissimi, tra di noi, avrebbero aspirato da grandi di fare l'astronauta per poter vedere da vicino i *crateri e i mari* della luna come fecero gli astronauti dell'Apollo 8, i primi ad abbandonare la relativa sicurezza dell'orbita terrestre per dirigersi verso il nostro satellite. Per non parlare poi di coloro che calpestarono il suolo lunare. Essi hanno rappresentato per tutti noi bambini della mia generazione degli idoli da seguire e di cui leggere le imprese. Lo scorso anno, in occasione del cinquantesimo anniversario della conquista della luna, ho rivissuto quell'impresa con gli occhi di un adulto ormai nel pieno della sua vita lavorativa (se non già verso la parabola discendente), leggendo alcuni dei numerosi libri pubblicati in occasione dell'evento.

* Dipartimento di Scienze Biomediche e Padova Neuroscience Center (Università degli Studi di Padova).

Fa particolarmente impressione leggere che quel *gioiello di tecnologia* che era il Lem dell'Apollo, fosse gestito da un computer dal potere di calcolo di un ormai vetusto orologio digitale, per non parlare dell'impietoso confronto con uno *smart watch* attuale. Così come mi aveva fatto impressione conoscere che quel computer era andato in crash per sovraccarico proprio nel momento del primo allunaggio, parlando personalmente con Buzz Aldrin qualche anno fa alla Nasa, in una di quelle incredibili e rare occasioni in cui un *bambino* adulto incontra direttamente l'idolo della sua giovinezza.

Se andiamo a visitare uno dei numerosi musei della scienza che vi sono in giro per il mondo, possiamo direttamente osservare come la dimensione dei computer stessi si sia ridotta, mentre al contempo la capacità di calcolo è aumentata. L'Eniac utilizzato negli anni '40-'50 occupava un'intera stanza ed aveva una capacità di calcolo minore di un attuale modulo Arduino. Nel mondo dei computer, quindi, dimensione e capacità di elaborazione non vanno necessariamente di pari passo.

Ma nel mondo biologico succede esattamente così? Se osserviamo le dimensioni del sistema nervoso rispetto alle dimensioni di un animale, notiamo come il primo aumenti di volume rispetto all'aumento del volume corporeo. Una cosa abbastanza logica, considerato il *maggior carico* di lavoro del sistema nervoso per elaborare i segnali sensitivi raccolti dalla periferia e inviare i segnali motori ai sistemi motori interni ed esterni. Questo aumento progressivo delle dimensioni volumetriche del sistema nervoso si associa a un aumento del numero delle cellule nervose. Il volume di quest'ultime, tuttavia, rimane sostanzialmente invariato tra i differenti organismi. Si osserva, semmai, un aumento della lunghezza complessiva delle ramificazioni dendritiche (cioè di quelle strutture che assieme al soma neuronale *ricevono* i segnali degli altri neuroni) e degli assoni (cioè le strutture neuronali che *inviando* informazioni ad altri neuroni).

Se consideriamo il sistema nervoso di *C. Elegans*, un piccolo verme molto utilizzato nel campo della ricerca biologica, esso è composto da 302 neuroni e 56 cellule gliali, tutti identificati molecolarmente, sia per quel che riguarda le loro caratteristiche come cellule differenziate che per quel che riguarda il loro sviluppo a partire da precursori. Passando all'insetto adulto del moscerino della frutta (*Drosophila melanogaster*) il numero di neuroni cresce a circa 100.000, anch'essi in gran parte identificati molecolarmente, sia come cellule differenziate sia per quel che riguarda il loro differenziamento. Nell'adulto di un altro insetto come *Apis mellifera*, il numero di neuroni del sistema nervoso centrale aumenta ancora di un fattore dieci, raggiungendo il valore di circa 1.000.000.

Passando ai vertebrati, *Mus musculus*, il topolino spesso utilizzato nella ricerca biomedica e nelle neuroscienze come *modello sperimentale* anche per lo studio della funzione di geni specifici mediante linee geneticamente

modificate, ha un sistema nervoso con circa 71 milioni di cellule nervose. Quest'ultime aumentano a 32 miliardi in uno dei primati a noi vicini (Urang U tan) per arrivare a 86 miliardi in *Homo sapiens sapiens*. Ad una prima analisi, quindi, sembrerebbe che l'aumento delle capacità cognitive, cioè della capacità di elaborare in modo profondo e complesso le informazioni ricevute dai sistemi sensoriali e da inviare ai sistemi motori, sia correlata all'aumento progressivo del numero di neuroni del sistema nervoso. Difficile infatti non pensare che le *capacità cognitive* di un verme o di un moscerino (ammesso esistano, ma invito il lettore a procedere oltre riguardo a questo argomento) siano inferiori a quelle di un topolino e che quelle di quest'ultimo non siano inferiori a quelle di un primate o dell'uomo (anche se modelli sperimentali murini di malattie psichiatriche umane, vorrebbero affermare questo). Questa linearità così precisa appena descritta, presenta tuttavia delle marcate eccezioni. Il piccione ha circa 4 volte più neuroni di un topolino, la qual cosa potrebbe far sorgere qualche dubbio sul fatto che quest'ultimo sia considerabile un buon modello per studiare la neurofisiologia delle funzioni complesse del cervello. Ancor più considerevole è il numero dei neuroni del cervello di un invertebrato cefalopode come il polpo: 500 milioni (sette volte più di un topolino).

Se vogliamo, tuttavia, rimanere nell'ambito di un rapporto lineare tra dimensione e numero di neuroni del sistema nervoso, è da considerare che il numero dei neuroni di un elefante è tre volte quello di un uomo. Appare quindi evidente che esistono variabili ulteriori, oltre alle dimensioni e al numero di neuroni, in grado di influire e regolare la capacità di elaborazione di un sistema nervoso. Ad esempio, la sua struttura con aree specifiche come possono essere la corteccia cerebrale o i nuclei della base, e/o la presenza di microcircuiti specifici, cioè circuiti nervosi costituiti da neuroni particolari che nel loro insieme svolgono delle funzioni distinte.

Strutture anatomiche macroscopiche e microcircuiti nervosi, poi non vanno a loro volta a braccetto. L'unità funzionale della corteccia cerebrale dei mammiferi è costituita dalla colonna corticale, cioè da un insieme di neuroni correlati funzionalmente tra di loro nell'elaborare un singolo input e un singolo output informativo. Una colonna (microcolonna corticale) della neocorteccia di topo è formata da circa 1000 neuroni, di cui si conoscono la geometria e le caratteristiche molecolari e biofisiche. Ciò ha permesso, nel progetto Blue Brain diretto dal prof. Markram del Politecnico di Losanna, di ricostruirla *in silico* per studiarne le caratteristiche funzionali tramite computer. Si tratta di un vero esempio di *Big Data* dal momento che le risposte bioelettriche dei neuroni dipendono dal loro corredo di canali ionici, neurotrasmettitori e recettori sinaptici, così come dalle loro proprietà biofisiche. Un'enorme massa di dati e variabili che devono essere computati assieme per riprodurre la funzione come unità dei neuroni della colonna stessa. La ricostruzione al

computer di questa struttura unitaria funzionale, potrà permettere di esplorarne la funzione. Siamo ancora lontani, però dal capire la funzione della nostra corteccia, se consideriamo che essa è formata da milioni di queste colonne corticali.

Se consideriamo il cervello di un piccione, non si osserva invece la presenza di una struttura anatomica che possa richiamare la nostra corteccia cerebrale. Tuttavia, analizzando i circuiti neuronali localizzati nelle regioni dove vengono elaborate le informazioni che arrivano dalla periferia, è possibile evidenziare dei microcircuiti che sono organizzati in modo molto simile alle colonne corticali: una fibra afferente che porta le informazioni raccolte in periferia dai recettori sensitivi, un insieme di neuroni correlati tra di loro funzionalmente che elaborano questa informazione, e un neurone di *uscita* che dà origine a un assone tramite il quale l'informazione elaborata dal microcircuito viene inviata ad altre regioni (centrali o periferiche) del sistema nervoso.

Analogamente è difficile rintracciare all'interno del sistema nervoso di un insetto un qualcosa che assomigli ai gangli della base che si osservano nella regione centrale dell'encefalo di un mammifero. Queste strutture, funzionalmente, vengono attivate dalle regioni della corteccia cerebrale coinvolte nella preparazione del movimento volontario prima dell'attivazione della corteccia motoria primaria, dove il movimento è tradotto in impulsi da inviare ai motoneuroni del midollo spinale. Il ruolo dei gangli della base è sostanzialmente quello di scegliere il modulo motorio (cioè l'insieme di atti motori) più adatto (in termini di vantaggio per l'individuo) per generare una determinata risposta comportamentale, basandosi su delle valutazioni tra vantaggi e svantaggi anche sulla base delle informazioni passate (memorie). Tuttavia, gli esperimenti neuroanatomici e funzionali (neurofisiologici e comportamentali) hanno rivelato che i microcircuiti fondamentali che si trovano nei gangli della base dei mammiferi sono presenti anche nel cervello degli insetti, dove svolgono sostanzialmente la stessa funzione (ovviamente scalata in termini di complessità) che svolgono nel sistema nervoso dei vertebrati.

Ma come funziona in termini generali e specifici, un sistema nervoso? È possibile paragonarlo in tutto e per tutto alla funzione di un computer o presenta delle differenze? Sostanzialmente il nostro sistema nervoso genera e regola il nostro comportamento. Secondo Levitis e altri, il comportamento è l'insieme di risposte (azioni e inazioni) coordinate internamente degli interi organismi viventi (individui o gruppi di individui) a stimoli interni o esterni, escludendo le risposte più facilmente interpretabili come modifiche dello sviluppo. Questa definizione di comportamento può essere adattata a tutti gli organismi, sia individualmente che come gruppi sociali.

Il sistema nervoso genera il comportamento, adattandolo alle informazioni sensoriali interne (ambiente interno) ed esterne (ambiente esterno, in prossimità con l'individuo o lontano dall'individuo), sulla base delle informazioni

precedentemente memorizzate e sul risultato (positivo/vantaggio o negativo/svantaggio) di analoghe risposte comportamentali passate. È chiaro quindi che le informazioni sensoriali presenti o passate, la loro elaborazione cognitiva come un *unicum* che porta alla percezione e l'analisi della risposta comportamentale intesa come adattamento dell'organismo allo stimolo sensoriale e cioè all'ambiente in cui si trova e del conseguente vantaggio che questa risposta comportamentale porta all'organismo stesso, sono il fondamento della funzione nervosa.

Questa attività di elaborazione deve sempre comportare il miglior vantaggio in termini evolutivi (cioè in termini di sopravvivenza e possibilità di accoppiamento) all'organismo stesso. Le informazioni sono elaborate, o meglio, codificate dai neuroni sotto forma di segnali elettrici che possiamo definire "digitali", i potenziali d'azione, che vengono inviati agli altri neuroni dei circuiti nervosi attraverso le connessioni sinaptiche e l'attività sinaptica. Il segnale elettrico nervoso, o potenziale d'azione, è un segnale stereotipato definito anche "tutto o niente". Ha quindi le caratteristiche di un bit di informazione, un 1 che si distingue nettamente dalla condizione in cui il segnale stereotipato è assente, o 0. Il neurone può generare un potenziale d'azione, cioè un segnale elettrico caratterizzato da una depolarizzazione completa del potenziale di membrana con inversione della polarità (fino a +30 mV all'interno della membrana rispetto che all'esterno) di durata (circa 1 ms) ed ampiezza (circa 100 mV) stabile, non appena il potenziale di membrana viene ridotto ad un valore di circa -55 mV, definito come valore soglia per l'innescò del potenziale d'azione.

Hodgkin e Huxley nei loro fondamentali studi sull'assone gigante del calamaro osservarono che, quando il potenziale di membrana raggiungeva il valore soglia del potenziale d'azione la permeabilità della membrana stessa per gli ioni fondamentali potassio e sodio, si modificava drammaticamente. Si passava cioè da una situazione "a riposo" in cui la permeabilità della membrana per il potassio era nettamente superiore a quella per il sodio, ad una condizione diametralmente opposta in cui la permeabilità per il sodio diveniva nettamente superiore a quella per il potassio. Tale situazione era transitoria e determinava, appunto transitoriamente, la modifica del potenziale di membrana che caratterizza il segnale elettrico stereotipato del potenziale d'azione, una modifica temporanea del potenziale di membrana del neurone.

Il numero di potenziali d'azione generati dal neurone, la durata del periodo in cui il neurone genera una *scarica* o *burst* di potenziali d'azione in successione, la frequenza con cui i potenziali d'azione vengono generati, la variazione della frequenza dei potenziali d'azione all'interno di un *burst*, sono tutte caratteristiche che permettono la codifica dell'informazione da parte del neurone.

Tutte le informazioni elaborate dal sistema nervoso sono *codificate* in sequenze di potenziali d'azione e periodi in cui il neurone è *a riposo*. Siano queste informazioni legate all'elaborazione di uno stimolo sensoriale come uno stimolo tattile, o alla sua elaborazione percettiva che porta al riconoscimento dell'oggetto toccato attraverso la combinazione dell'informazione tattile con altre modalità sensitive e con memorie passate. Siano informazioni che riguardano un particolare atto motorio volontario (prendere una mela da un tavolo) o l'informazione che uno specifico motoneurone invia alle fibre muscolari di cui controlla la contrazione (ad esempio, un motoneurone che innerva le fibre del muscolo estensore dell'avambraccio sul braccio, cioè il muscolo tricipite). Siano infine le informazioni più complesse elaborate dalla nostra mente come le memorie, il tono dell'umore, le pulsioni sessuali, l'attenzione, il sonno e la veglia o, infine, la coscienza.

Tuttavia, i neuroni sono poi connessi tra di loro a formare i circuiti nervosi. Questi non sono separati tra di loro, ma uniti, attraverso le connessioni sinaptiche, a formare un'intricata rete attraverso cui le informazioni codificate in sequenze di potenziali d'azione viaggiano e si distribuiscono nelle differenti regioni cerebrali. Ogni neurone rappresenta perciò, al tempo stesso, un nodo in cui afferiscono le informazioni provenienti da altri neuroni (in media circa 1000 per ciascun neurone), un nodo di elaborazione di queste informazioni, ed infine un nodo di distribuzione ad altri neuroni delle informazioni da esso elaborate. All'interno di questa intricata rete di trasmissione delle informazioni, i neuroni assumono un ruolo fondamentale non solo come *ripetitori*, ma anche di *integratori* ed *elaboratori* delle informazioni ricevute, che vengono quindi successivamente inviate ad altri neuroni in cui il processo di elaborazione ricomincia di nuovo.

Le sinapsi svolgono un ruolo fondamentale in questi meccanismi neurofisiologici. È attraverso la funzione sinaptica che l'informazione elaborata da un neurone (il neurone pre-sinaptico) viene trasmessa a un altro neurone (il neurone post-sinaptico), e qui viene integrata con le informazioni che lo stesso neurone post-sinaptico riceve da altri neuroni pre-sinaptici. Il processo di integrazione (integrazione sinaptica) è l'essenza stessa della capacità di elaborazione del neurone post-sinaptico, cioè della sua capacità di combinare le informazioni che riceve e generare a sua volta una risposta da inviare ad altri neuroni.

Il processo di integrazione sinaptica dipende da moltissimi fattori, tra cui la forma geometrica del neurone e le sue dimensioni, nonché la capacità che hanno gli ioni di attraversare la membrana cellulare (caratteristiche biofisiche), il potenziale di membrana e la posizione della sinapsi sul neurone stesso. La trasmissione sinaptica è un processo complesso in cui sostanzialmente il potenziale d'azione del neurone presinaptico viene trasformato in un segnale chimico attraverso la liberazione nello spazio tra il neurone presinaptico

e il neurone post-sinaptico di molecole di neurotrasmettitore, che si legano su degli specifici recettori posti sulla membrana del neurone post-sinaptico, dove il segnale chimico è ritrasformato in un segnale elettrico: il potenziale post-sinaptico. Va da sé quindi che oltre all'integrità del neurone, per il buon funzionamento del sistema nervoso e della capacità di elaborazione delle informazioni codificate in potenziali d'azione, è necessario un buon funzionamento delle sinapsi. Sostanze chimiche esterne come la morfina che si legano a recettori postsinaptici endogeni in modo esagerato e prolungato nel tempo possono portare a un esagerato funzionamento dei circuiti nervosi, in cui esse agiscono e quindi alla comparsa di tutti i fenomeni neurofisiologici e cognitivi che caratterizzano l'azione della morfina. Analogamente, è noto come l'uso di composti farmacologici sia in grado di stimolare o bloccare particolari tipi di recettori post-sinaptici per il controllo dei sintomi più gravi, complessi e disabilitanti delle malattie psichiatriche, e come questi farmaci siano divenuti, nel corso degli ultimi decenni, sempre più flessibili da utilizzare presentando sempre meno effetti collaterali disabilitanti.

Sembra chiaro, quindi, che l'enorme varietà e l'affascinante flessibilità delle risposte del cervello risiedano nella complessità della sua struttura anomo-funzionale. Ma quanto complesso deve essere un cervello per essere in grado di generare tale varietà e flessibilità di risposta, per essere in grado di poter scegliere la miglior risposta o comportamento adattativo – valutando tutti i pro e i contro – a quello che intende fare anche sulla base delle precedenti esperienze acquisite? A una prima analisi sembrerebbe che questo è di sola pertinenza del cervello dell'uomo, anche se Darwin (1871) ebbe a scrivere riguardo alle formiche:

It is certain that there may be extraordinary activity with an extremely small absolute mass of nervous matter; thus the wonderfully diversified instincts, mental powers, and affections of ants are notorious, yet their cerebral ganglia are not so large as the quarter of a small pin's head. Under this point of view, the brain of an ant is one of the most marvellous atoms of matter in the world, perhaps more so than the brain of man.

E pertanto, quanto deve essere grande un cervello per mostrare capacità "cognitive", cioè per mostrare di essere in grado di *selezionare attentivamente le informazioni, valutare, memorizzare, percepire*? Se consideriamo il sistema nervoso di un'ape, questo è composto da circa un milione di neuroni, decisamente meno degli 83 miliardi di cellule nervose dell'uomo. Eppure, questo piccolo *cervello* è in grado di sviluppare straordinarie capacità, come comunicare agli altri individui la distanza e la direzione rispetto all'arnia dei fiori con il loro nettare (danza delle api), o di orientarsi distinguendo i colori o la forma degli oggetti e scegliendo l'appropriata in base a una decisione.

Tutte abilità che possono essere considerate e studiate nell'uomo come abilità cognitive e che sono state osservate come "cognitive" nell'ape. Quindi la dimensione del cervello e/o la massa corporea per quanto secondo alcuni studi sembrano essere in qualche modo predittori di capacità cognitive, secondo altri studi non appaiono correlati. Un problema nasce poi quando consideriamo gli insetti cosiddetti sociali, dove ad un piccolo cervello si associa un grosso *repertoire* comportamentale.

La velocità di apprendimento può essere un indice di "intelligenza"? Ancora una volta le api sembrano contraddire questa ipotesi, dal momento che presentano una velocità di apprendimento superiore a quella di gran parte dei vertebrati, per non parlare dei neonati umani. Di nuovo, lo studio approfondito della danza delle api ci svela l'estrema ed elevata attività cognitiva delle api. Come sopra detto, la danza ha lo scopo di informare gli altri individui della posizione della fonte di cibo, rispetto all'arnia. Il territorio coperto dalla danza è di parecchi km² attorno all'arnia. Un'area di cui devono essere memorizzati il luogo, i fiori o le zone più ricche di cibo, le informazioni provenienti da più fuchi, il momento della giornata in cui l'informazione di ciascun fuco possa portare maggior vantaggio. Inoltre gli individui che ricevono le informazioni devono anche associarle in modo ordinato e stabile, memorizzare il momento della giornata in cui l'informazione di ciascun fuco possa portare al maggior vantaggio e utilizzare la posizione spaziale come segnale contestuale per generare delle risposte. Infine imparare le sequenze di posizioni spaziali (*landmarks*) nel territorio, associandole a istruzioni di direzione per ciascuna posizione spaziale.

Quale è, quindi, il significato, in termini di convenienza, di aumentare le dimensioni di un cervello, dal momento che un cervello così piccolo come quello di un'ape, è già in grado di compiere delle operazioni così complesse? L'aumento di dimensioni di determinate aree del cervello può portare a miglioramenti nelle abilità cognitive (vantaggi *qualitativi*), e/o può portare ad un maggiore dettaglio o risoluzione, cioè a migliore sensibilità o precisione nell'elaborazione dell'informazione (vantaggi *quantitativi*; spesso lo si osserva nei sistemi sensitivi o motori). Si può ipotizzare che la ridondanza di particolari microcircuiti nervosi implicati in determinate attività cognitive (ad esempio i circuiti a *loop* che coinvolgono la corteccia cerebrale e i gangli della base) sia alla base delle migliorate attività cognitive (vantaggi *qualitativi*).

Un esempio di miglioramento della capacità di risoluzione (vantaggi *quantitativi*) si può osservare confrontando il numero di unità funzionali dell'occhio del moscerino della frutta (700 unità funzionali; ommatidi) con il numero di quelle presenti nella retina umana (circa 400.000 unità funzionali). Di quest'ultime, 60.000 sono localizzate nella fovea che rappresenta l'1% della retina, ma che è la regione in cui focalizziamo gli oggetti che hanno catturato la nostra attenzione visiva selettiva. Questo aumento della capacità di risol-

zione della retina umana è ottenuto aumentando oltre che le unità funzionali, anche i circuiti nervosi che a ciascuna di esse fanno capo. C'è quindi un incremento *parallelo* delle vie di inoltro dell'informazione visiva alle aree centrali, in cui essa viene ulteriormente elaborata. L'incremento anatomico delle vie nervose si traduce in un aumento delle dimensioni delle regioni cerebrali interessate. Tuttavia, è necessario che questo aumento nel numero dei neuroni e dei circuiti nervosi, si accompagni anche a un aumento della velocità di propagazione dell'informazione nervosa. A tale scopo, è comparsa da un punto di vista evolutivo, la guaina mielinica che avvolge i prolungamenti dei neuroni che terminano nei bottoni sinaptici. L'aumento della velocità di propagazione dell'informazione nervosa si è associata a un aumento della risoluzione temporale di elaborazione delle differenti informazioni. Questo vantaggio è combinato con la comparsa di sinapsi in grado di trasmettere il segnale nervoso da un neurone all'altro ad alta frequenza. In tal modo si è creato il correlato di funzioni in grado di permettere al sistema nervoso l'elaborazione ad alta frequenza di informazioni.

In ultima istanza, la maggiore richiesta di risoluzione delle informazioni si accompagna pure a una maggiore richiesta di riaggiornare più frequentemente le informazioni stesse, e alla fine ad una maggiore quantità di informazioni da immagazzinare. Tutti questi fattori richiedono una maggiore quantità di tessuto nervoso. Ciò implica che non abbiamo bisogno *creare* nuove forme di elaborazione delle informazioni, ma semplicemente di utilizzare quelle che abbiamo, cioè i neuroni e circuiti nervosi che già ci sono rendendoli semmai ridondanti. Un esempio possono essere i glomeruli olfattivi, cioè i circuiti primari di elaborazione dell'informazione olfattiva. Essi risultano sostanzialmente simili tra drosophila, topo ed uomo. La differenza tra queste specie sta nel numero: 43 in drosophila, 350 nell'uomo e 1000 nel topo. Nelle tre specie osserviamo lo stesso circuito nervoso, che però è presente quantitativamente in modo differente.

Riassumendo, la maggiore dimensione del corpo permette di avere un maggior numero di organi sensoriali (e anche organi sensoriali di dimensione maggiore) che facilitano una migliore conoscenza dell'ambiente (interazione con l'ambiente; comportamento adattativo). Una configurazione di questo genere presuppone di aumentare in parallelo i circuiti primari e centrali che elaborano le informazioni e ciò determina una ridondanza che rende ragione della capacità plastica di adattamento dei circuiti nervosi che si osserva molto bene in risposta alle lesioni del tessuto nervoso stesso. In questi casi, come ad esempio a seguito di un *ictus* nell'uomo, la fase di deficit funzionale che si osserva immediatamente all'instaurarsi della lesione, è seguita da una fase di recupero funzionale parziale dovuta alla *presa in consegna* della capacità elaborativa della regione lesa da parte delle regioni cerebrali circostanti rimaste intatte. Ciò porta anche ad una maggiore diversificazione funzionale dei

neuroni migliorandone la capacità di risoluzione temporale delle informazioni e quindi la capacità di elaborare in modo più profondo l'informazione.

L'aumento nel numero e nella dimensione dei neuroni aumenta però l'energia necessaria per il loro funzionamento, riducendo l'efficienza energetica dell'elaborazione nervosa. Un chiaro esempio di questa situazione si osserva nelle cosiddette malattie mitocondriali, caratterizzate dalla disfunzione dei mitocondri, organelli intracellulari deputati alla produzione di energia sotto forma di Atp a partire dal consumo di Ossigeno. Ne deriva che aspetti negativi promuovono, al contrario, la riduzione delle strutture nervose a un minimo funzionale. Se consideriamo organismi come gli insetti, con un sistema nervoso *in miniatura*, notiamo come essi siano in grado di generare un numero di comportamenti etologici non di molto inferiore a quello dei mammiferi.

Un così ampio *repertoire* di comportamenti richiede anche un'ampia capacità di elaborazione delle informazioni motorie in uscita per l'apparato motorio e quindi *regioni motorie* di grandi dimensioni nel sistema nervoso. Se consideriamo il numero di muscoli di una locusta essi sono addirittura maggiori di quelli di un ratto o di un opossum e simili a quelli di un macaco, eppure il cervello della locusta è decisamente più piccolo di quello dei mammiferi appena considerati. E allora perché i cervelli di questi mammiferi non sono più piccoli, essendo anche più costosi da un punto di vista energetico? Una possibile ipotesi a spiegazione di questa differenza potrebbe essere la presenza di capacità cognitive complesse, assenti negli insetti. In effetti questa era l'ipotesi più accreditata fino a circa una quarantina di anni fa, quando un gruppo di ricercatori di matrice culturale differente (biologi, informatici, fisiologi) iniziò a studiare la visione negli insetti con l'intento di descriverla da un punto di vista cibernetico. Da quegli studi pionieristici sfociarono una serie di ulteriori analisi quantitative approfondite del comportamento negli invertebrati, in cui una parte fondamentale ebbero Martin Heisenberg con i suoi esperimenti sulla visione nel moscerino della frutta e Seymour Benzer con i suoi studi sulla neurogenetica del comportamento. Il primo introducendo un'analisi comportamentale quantitativa basata sull'utilizzo di paradigmi sperimentali *open loop* e *closed-loop*, il secondo sfruttando l'estrema manipolabilità genetica del moscerino della frutta per studiare l'impatto di specifici geni sul comportamento. Usando una combinazione dei due approcci sperimentali è stato possibile ricostruire i circuiti responsabili di determinati comportamenti complessi (ad esempio l'orientamento). Da questi studi ne sono originati altri, in cui l'approccio di un'analisi quantitativa del comportamento è stato utilizzato anche in altri insetti (ape mellifera, formica), producendo una grande quantità di evidenze sperimentali sulle capacità cognitive degli insetti. Dall'osservazione di capacità attentive selettive, alla dimostrazione della capacità di categorizzare gli stimoli visivi, all'osservazione di forme di apprendimento contestuale, di sequenze o sociale o temporale

(cioè previsione di quando avverrà un determinato evento in conseguenza dell'esperienza passata). Interessanti poi gli studi sulla capacità di contare o *path-integration*. Questi *mini-cervelli* sono perciò in grado di generare risposte "cognitive" complesse anche in presenza di un numero di neuroni decisamente inferiore rispetto a cervelli più grandi.

In conclusione quindi possiamo dire che la dimensione del sistema nervoso non è un indice in grado di discriminare la capacità di generare e regolare complesse risposte di tipo cognitivo. Piuttosto, l'aumento delle dimensioni di un cervello, oltre ad associarsi alla comparsa della mielina per velocizzare la trasmissione delle informazioni anche a distanze più lunghe, determina la comparsa di nuove regioni che agiscono *in parallelo* con quelle già presenti. Regioni nelle quali si ritrovano microcircuiti ridondanti, cioè strutturati in modo simile a quelli già presenti precedentemente. Ciò si associa a un notevole aumento delle connessioni a breve distanza che hanno lo scopo di favorire la *disseminazione* delle informazioni e la *segregazione* delle regioni cerebrali stesse. Tutto questo favorisce l'emergenza di nuovi comportamenti e/o una più fine modulazione attraverso la varietà di risposte, di comportamenti già presenti precedentemente.

Attualmente, quindi, per quanto sia argomento di discussione anche al di fuori del campo delle neuroscienze se possa essere applicabile il termine "cognitivo" a questi comportamenti osservati nei *minibrains* degli insetti, l'idea è che queste attività complesse presentino delle similarità molto elevate nei loro meccanismi di base condivisi con le medesime risposte osservate nei vertebrati e nei mammiferi compreso l'uomo, e che quindi ne condividano i microcircuiti che generano questi meccanismi di base. La differenza sta quindi nella *ricchezza* dell'attività complessa analizzata, che permette ai cervelli di dimensioni maggiori di poter accedere a un *repertorio* di differenti moduli di risposta molto più ampio per meglio soddisfare la ricerca di quel *vantaggio* che è alla base del comportamento adattativo all'ambiente circostante (vide supra). Un chiaro esempio è la capacità di memorizzazione visiva che è limitata nelle api (arrivano a memorizzare fino a 6 siti di approvvigionamento) rispetto ad animali più distanti come il piccione (800 immagini) o l'uomo (fino a 10.000 immagini).

In conclusione, l'aumento di dimensioni di un cervello porta a un aumento nel numero dei suoi neuroni e dei suoi circuiti. Questo però non porta alla comparsa di un qualcosa di nuovo, da un punto di vista strutturale e funzionale, rispetto ai cervelli di dimensioni minore. Porta invece a un aumento della capacità di elaborazione di un *sistema* che nei suoi aspetti di base non è poi così dissimile da prima. Aumenta la capacità di raccogliere e definire l'informazione attraverso i canali sensoriali, e di disseminarla in parecchie aree cerebrali, la cui attività in parallelo migliora la capacità di elaborare l'informazione, permettendo di generare anche dal punto di vista efferente delle

risposte comportamentali molto più dettagliate e adeguate all'input ambientale. Tutto questo nella prospettiva che è vantaggio dell'organismo, in termini evolutivi, adattarsi nel miglior modo all'ambiente in cui vive.

L'insetto con il suo *minibrain* non potrà mai ovviamente competere con un uomo. Comunque la sua attività nervosa deve essere osservata da una prospettiva differente da quella rigidamente meccanicistica con cui è stata osservata fino a ora. Parafrasando il titolo di un lavoro pubblicato dal gruppo di M. Heisenberg, un moscerino non potrà mai andare in bicicletta, ma ha in sé le potenzialità di poterlo fare se solo avesse potuto contare su di un cervello grande.

Abstract

When we discuss big data, we are directly or indirectly thinking about our brain. It is at the same time a reference point showing an enormous computational ability provided us by nature and a boundary that could be crossed for reaching unexpected and inconceivable fields far beyond what nature gave us during evolution. In the biological world, is it only our brain showing these marvelous abilities or can we find them also in the tiny brains of smaller and simpler organisms like insects? These aspects are critically discussed in the following contribution.

Bibliografia

- Avarguès-Weber A., Dyer A.G., Giurfa M. (2010), *Conceptualization of above and below relationships by an insect*, «Proc Biol Sci», 278, pp. 898-905.
- Bührlé C.P., Sonnhof U. (1985), *The ionic mechanism of postsynaptic inhibition in motoneurons of the frog spinal cord*, «Neuroscience», 14, pp. 581-592.
- Burns J.G., Foucaud J., Mery F. (2010), *Costs of memory: lessons from 'mini' brains*, «Proc Biol Sci», 278, pp. 923-929.
- Callister R.J., Brichta A.M., Schaefer A.T., Graham B.A., Stuart D.G. (2020), *Pioneers in CNS inhibition_ 2. Charles Sherrington and John Eccles on inhibition in spinal and supraspinal structures*, «Brain Res», 1734, 146540.
- Chittka L., Niven J. (2009), *Are Bigger Brains Better?*, «Curr Biol», 19, R995-R1008.
- Chittka L., Skorupski P. (2011), *Information processing in miniature brains*, «Proc Biol Sci», 278, pp. 885-888.
- Collins L.T. (2019), *The case for emulating insect brains using anatomical 'wiring diagrams' equipped with biophysical models of neuronal activity*, «Biological Cybernetics», 113, pp. 465-474.
- Computers may take the place of parliament* | Comment | The Times. 1-9 (2019). Available at: (Accessed: 10 October 2020)
- da Costa N.M., Martin K.A.C. (2010), *Whose Cortical Column Would that Be?*, «Front. Neuroanat.», 4, 16.
- Dacke M., Srinivasan M.V. (2008), *Evidence for counting in insects*, «Anim Cogn», 11, pp. 683-689.
- Dicke U., Roth G. (2016), *Neuronal factors determining high intelligence*, «Philos. Trans. R. Soc. Lond., B, Biol. Sci.», 371, 20150180-9.
- Hawkins J., Ahmad S., Cui Y. (2017), *A Theory of How Columns in the Neocortex Enable Learning the Structure of the World*, «Front Neural Circuits», 11, 14608-18.
- Healy S.D., Rowe C. (2013), *Costs and benefits of evolving a larger brain: doubts over the evidence that large brains lead to better cognition*, «YANBE», 86, e1-e3.
- Herculano-Houzel S. et al. (2014), *The elephant brain in numbers*, «Front. Neuroanat.», 8, p. 46.
- Hesse J., Schreiber S. (2015), *Externalization of neuronal somata as an evolutionary strategy for energy economization*, «Curr Biol», 25, R324-R325.
- Hofman M.A. (2014), *Evolution of the human brain: when bigger is better*, «Front. Neuroanat.», 8, p. 15.
- Hustert R. (2012), *Giant and dwarf axons in a miniature insect, Encarsia formosa, (Hymenoptera, Calcididae)*, «Arthropod Struct Dev», 41, pp. 535-543.
- Ito K. et al. (2014), *A Systematic Nomenclature for the Insect Brain*, «Neuron», 81, pp. 755-765.
- Kotrschal A. et al. (2013), *The benefit of evolving a larger brain: big-brained guppies perform better in a cognitive task*, «YANBE», 86, e4-e6.
- Laughlin S.B., Sejnowski T.J. (2003), *Communication in neuronal networks*, «Science», 301, pp. 1870-1874.

- Levitis D.A., Lidicker W.Z., Freund G. (2009), *Behavioural biologists don't agree on what constitutes behaviour*, «YANBE», 78, pp. 103-110.
- Menzel R., Giurfa M. (2001), *Cognitive architecture of a mini-brain: the honeybee*, «Trends in Cognitive Sciences», 5, pp. 62-71.
- Minelli A., Fusco G. (2019), *No limits: Breaking constraints in insect miniaturization*, «Arthropod Struct Dev», 48, pp. 4-11.
- Molnar Z. (2013), *Cortical columns. Neural Circuit Development and Function in the Brain: Comprehensive Developmental*, «Neuroscience», vol. 3. <http://dx.doi.org/10.1016/B978-0-12-397267-5.00137-0>
- Niven J.E., Farris S.M. (2012), *Miniaturization of Nervous Systems Review and Neurons*, «Curr Biol», 22, R323-R329.
- Peng F., Chittka L. (2017), *A Simple Computational Model of the Bee Mushroom Body Can Explain Seemingly Complex Forms of Olfactory Learning and Memory*, «Curr Biol», 27, pp. 224-230.
- Perry C.J., Chittka L. (2019), *Science Direct How foresight might support the behavioral flexibility of arthropods*, «Curr Opin Neurobiol», 54, pp. 171-177.
- Polilov A. (2012), *A. The smallest insects evolve anucleate neurons*, «Arthropod Struct Dev», 41, pp. 29-34.
- Rivera-Alba M., Peng H., de Polavieja G.G., Chklovskii D.B. (2014), *Wiring economy can account for cell body placement across species and brain areas*, «Curr Biol», 24, R109-R110.
- Roth G., Dicke U. (2005), *Evolution of the brain and intelligence*, «Trends in Cognitive Sciences», 9, pp. 250-257.
- Sane S.P. (2016), *ScienceDirect Neurobiology and biomechanics of flight in miniature insects*, «Curr Opin Neurobiol», 41, pp. 158-166.
- Sengupta B., Stemmler M., Laughlin S.B., Niven J.E. (2010), *Action Potential Energy Efficiency Varies Among Neuron Types in Vertebrates and Invertebrates*, «PLoS Comput. Biol.», 6, e1000840-16.
- Shettleworth S.J. (2012), *Modularity, comparative cognition and human uniqueness*, «Philos. Trans. R. Soc. Lond., B, Biol. Sci.», 367, pp. 2794-2802.
- Smith D., Wessnitzer J., Webb B. (2008), *A model of associative learning in the mushroom body*, «Biological Cybernetics», 99, pp. 89-103.
- Theobald J. (2014), *Insect Neurobiology: How Small Brains Perform Complex Tasks*, «Curr Biol», 24, R528-R529.
- van Gerven D.J.H., Ferguson T., Skelton R.W. (2016), *Acute stress switches spatial navigation strategy from egocentric to allocentric in a virtual Morris water maze*, «Neurobiol Learn Mem», 132, pp. 29-39.
- Williams R.W., neuroscience, K.H.A.R.O. (1988), *The control of neuron number*, «annualreviews.org», 11, pp. 423-453.
- Wolf R., Voss A., Hein S., Heisenberg M. (1992), *Can a fly ride a bicycle?*, «Phyl Trans R S B», 337, pp. 261-269.

Che cos'è un «dato» storico? *Una riflessione di metodo fra Big Data,* *approcci quantitativi e scale di analisi*

di *Andrea Caracausi**

L'irrompere dei Big Data nella società contemporanea ha stimolato un'ampia riflessione non solo in merito al diritto alla *privacy*, ma anche alle implicazioni che l'utilizzo di una simile massa di informazioni può avere nell'influenzare il comportamento individuale e collettivo in termini di consumi, investimenti e, non ultimo, scelte politiche. Questo fascino per i Big Data ha coinvolto – sebbene a geografie variabili – anche le scienze storiche, dando luogo a una rinnovata discussione sul rapporto fra storia e dato e fra metodi quantitativi e qualitativi, riprendendo le fila di un discorso iniziato già a metà Novecento e interrottosi qualche decennio dopo¹.

Tornare a riflettere sul metodo storico e sulle implicazioni del rapporto fra quantità e qualità è una questione centrale. Al di là del fatto che ogni discipli-

* Dipartimento di Scienze Storiche, Geografiche e dell'Antichità, Università degli Studi di Padova. Ho avuto l'occasione di iniziare a ragionare su questi temi all'interno di una giornata di studio organizzata da Daniele Andreozzi presso l'Università degli Studi di Trieste (*Quantità/qualità. La storia tra sguardi micro e generalizzazioni*, 26 novembre 2015). Come spesso accade, i tempi di scrittura (in questo caso miei) non erano maturi per essere pubblicati nel volume successivo, ma il sottoscritto ha potuto beneficiare ampiamente, come si vedrà anche nelle pagine seguenti, dello scambio di idee con i partecipanti dell'incontro, in primo luogo Daniele Andreozzi, Alida Clemente, Giovanni Favero e Luca Mocarelli. Ringrazio inoltre, per la loro generosità e i loro commenti a una prima versione del testo, Lucio Biasiori, Benoît Maréchaux, Federico Mazzini, Michele Nani, Andrea Pase e Denny Solera. Il capitolo è pubblicato nell'ambito del progetto d'eccellenza Mobility & Humanities del Dipartimento di Scienze Storiche Geografiche e dell'Antichità (DiSSGeA) dell'Università degli Studi di Padova.

1. Kris James Mitchener, *The 4D Future of Economic History: Digitally-Driven Data Design*, «The Journal of Economic History», 75 (2015), pp. 1234-1239; *Quantità/qualità: la storia tra sguardi micro e generalizzazioni*, a cura di Daniele Andreozzi, Palermo 2017; *Nouvelles cuisines de l'histoire quantitative*, edité par Karine Karila-Cohen, Claire Lemercier, Isabelle Rosé et Claire Zalc, «Annales. Histoire, Sciences Sociales», 73 (2018), pp. 773-783; Myron P. Gutmann, *Quantifying Interdisciplinary History: The Record of (Nearly) Fifty Years*, «The Journal of Interdisciplinary History», 50 (2020), pp. 517-545; Michele Nani, *La metà perduta. Appunti su studi storici e metodi quantitativi a partire da pubblicazioni recenti*, «Italia contemporanea», 293/2020, pp. 177-189.

na deve confrontarsi con i concetti e gli strumenti che utilizza, le recenti giravolte storiografiche (i molti «tornanti» sono lì a testimoniarlo) e i numerosi tentativi di accreditarsi al novero delle scienze utili, in forma pubblica o meno, testimoniano una certa debolezza epistemologica di fondo che non è tanto di contenuti, quanto invece di domande poste e di metodi con cui affrontarle². Il rischio è altrimenti quello di vedere la storia svuotata di un suo senso e di un suo statuto autonomo, ridotta a un semplice contenitore di informazioni buone per testare teorie o modelli elaborati da altre discipline e, quindi, in piena subalternità alle altre scienze sociali, fisiche e della vita³.

Affronterò questo tema con il seguente percorso. Un primo paragrafo mostrerà i principali stimoli, emersi nel recente dibattito storiografico, a dialogare in maniera più stretta non soltanto con le scienze sociali, ma anche con quelle fisiche e della vita, grazie alle potenzialità che grandi banche dati e nuovi strumenti informatici possono offrire. Un secondo paragrafo ripercorrerà gli snodi principali delle discussioni intorno alla storia quantitativa dagli anni Settanta, per rimettere al centro i motivi che hanno – a suo tempo – provocato un distacco da essa. Nel terzo paragrafo aprirò una riflessione sull'importanza di sviluppare una metodologia di analisi storica che tenga presente la natura del dato storico, mentre nel quarto verranno indagati i processi di quantificazione e il loro nesso con l'analisi qualitativa. L'accettazione di una siffatta concezione del dato storico e l'applicazione di un metodo scientifico coerente per trattarlo condiziona inevitabilmente la scala di osservazione, come si vedrà nel paragrafo quinto. L'ultima sezione, invece, si sofferma sul rapporto fra storia e altre scienze, a partire dalle ipotesi e generalizzazioni che la prima può produrre per contribuire a un'analisi più complessa e critica del mondo reale.

1. La storia fra legittimità scientifica, Big Data e lunga durata

In un articolo apparso su «Il Sole 24 Ore» nel maggio del 2019, in una recensione al volume del filosofo della scienza Alexander Rosenberg, *How history gets things wrong. The neuroscience of our addiction to stories*, Gilberto Corbellini sosteneva la «falsità» di molti studi storici, con l'eccezione di quelli che usavano metodi quantitativi o «approcci controllabili». L'utilizzo

2. Vedi in particolare: *AHR Forum: Historiographic "Turns" in Critical Perspective*, «American Historical Review», 117 (2012), pp. 698-813.

3. Daniele Andreozzi, *Senso e potere: alla ricerca della storia tra dimensioni, confini e rilevanze*, in *Quantità/qualità*, pp. 17-34. Questo è ciò che fanno, a buon diritto dal loro punto di vista, gli economisti storici: vedi Ran Abramitzky, *Economics and the Modern Economic Historian*, «The Journal of Economic History», 75 (2015), pp. 1240-1251; Luca Mocarelli, *Storia ed economia: un matrimonio impossibile?*, in *Quantità/qualità*, pp. 71-91.

della teoria della mente, volta a interpretare i presunti ragionamenti degli individui del passato, sarebbe causa di continue discussioni che non porterebbero mai a un accordo, né a qualcosa di fondato. Al centro di quell'intervento vi è dunque l'assenza di scientificità e di utilità sociale della disciplina storica, che dipenderebbe troppo dalle narrazioni e dalla ricerca di un modello di comportamento esplicativo, portando così i lavori scritti dagli storici a perdere credibilità e a essere falsi⁴.

La voce di Corbellini non era isolata. Qualche anno prima aveva visto la luce un libro intitolato *The History Manifesto*, a opera di David Armitage e Jo Guldi (2014). Il lavoro si proponeva di rimettere al centro del dibattito scientifico storico una serie di dicotomie, quali breve/lungo periodo, qualitativo/quantitativo, micro/macro. Gli autori insistevano affinché la storia tornasse ad accreditarsi nel panorama pubblico attraverso tre elementi: l'utilizzo di banche dati e strumenti digitali, in costante dialogo con le scienze informatiche; l'ampiezza temporale della prospettiva di ricerca, riprendendo la lunga durata braudeliana, in un'ottica oggi di *deep* o *big history* (una storia quindi «profonda» o «grande»); la necessità di affrontare grandi questioni, come l'ambiente, la disuguaglianza, la democrazia⁵. Il *Manifesto per la Storia*, così tradotto in Italia da Donzelli, suscitò subito un acceso confronto fra gli studiosi, con una certa risonanza nelle pagine delle riviste di Storia più prestigiose al mondo: negli Stati Uniti, *The American Historical Review* venne meno alla sua tradizionale ritrosia a pubblicare articoli polemici, facendone un vero e proprio forum di dibattito, mentre in Francia *Annales. Histoire, Sciences Sociales* vi dedicò la sezione di un fascicolo⁶. Le pagine del volume

4. Gilberto Corbellini, *Questa storia è davvero molto falsa*, «Il Sole 24 Ore», 14 maggio 2019, www.ilsole24ore.com/art/questa-storia-e-davvero-molto-falsa--ABAXajuB (ultimo accesso 30 maggio 2020). Su molte questioni ritorneremo nelle pagine seguenti. Si veda già però l'intervento successivo degli stessi storici del Cnr che sottolineano l'aver ignorato, ad esempio, la rilevanza che la questione della critica della fonte e della distinzione fra storia e memoria ha avuto nella riflessione storiografica e la riflessione epistemologica che ha visto gli storici impegnati nel superare le rigide dicotomie quantitativo/qualitativo o struttura/soggettività: *Al Cnr la Storia è una scienza? Una risposta all'intervento di Gilberto Corbellini*, www.alfabeta2.it/2019/05/26/al-cnr-la-storia-e-una-scienza-una-risposta-allintervento-di-gilberto-corbellini (ultimo accesso 30 maggio 2020).

5. Jo Guldi e David Armitage, *The History Manifesto*, Cambridge 2014.

6. Fra le molte, segnalo Deborah Cohen e Peter Mandler, *The History Manifesto: A Critique*, «The American Historical Review», 120 (2015), pp. 530-542; Claire Lemerrier, *Une histoire sans sciences sociales?*, «Annales. Histoire, Sciences Sociales», 70 (2015), pp. 345-357; Francesca Trivellato, *Un nouveau combat pour l'histoire au XXIe siècle?*, «Annales. Histoire, Sciences Sociales», 70 (2015), pp. 333-343; Giulia Bassi, *Storia, storiografia, manifesto: alcune considerazioni in merito a una sintesi difficile*, «Studi storici» (2016), pp. 297-313. La discussione, soprattutto in Italia, si è soffermata più sui temi della breve o lunga durata e del digitale/pubblico, mentre un dibattito sui metodi quantitativi è ben lontano. Vedi in proposito le osservazioni di Nani, *La metà perduta*.

si articolavano del resto intorno ad alcuni snodi importanti: la maggiore rilevanza della lunga durata rispetto al breve periodo, l'utilizzo dei Big Data per illuminare le ombre della storia su temi «grandi», una scrittura che guarda al passato *in direzione del futuro* e *in pubblico* per favorire maggiormente la condivisione di idee (corsivo mio). Lo storico dovrebbe tornare a fianco dei governanti per suggerire politiche concrete al fine di risolvere i grandi problemi delle società attuali dal momento che la storia modellerebbe il presente e influirebbe sul futuro⁷.

La nuova era aperta dai Big Data – continuano gli autori – inviterebbe infatti a decifrare miliardi di informazioni riguardanti il genoma umano o lemmi e parole mescolate ogni anno in report e altri documenti ufficiali, dando così la possibilità di effettuare nuovi calcoli sulle possibilità di risolvere vecchie questioni e porne di nuove. I Big Data spingerebbero la storia verso più grandi problemi e verso periodi di tempo sempre più lunghi, grazie all'accumulo di dati operato nella seconda metà del Novecento dalle scienze naturali e umane. L'emergere dell'umanistica informatica come campo di studi ha contribuito a mettere a disposizione sempre più strumenti per chi vuole cercare di fare ricerche nel passato anche con le proprie mani, facilitandone la visualizzazione e la sintesi grazie a strumenti come *Google Books Ngram Viewer* o *Paper Machines*⁸.

Non vorrei soffermarmi troppo in questa sede sulle critiche al libro, che sono già state discusse altrove, ma è importante ricordare alcuni aspetti: l'aver sottovalutato la profondità e le conseguenze degli studi sul breve periodo; l'aver confuso l'importanza che lunga e breve durata congiuntamente avevano in Fernand Braudel; l'assenza di una prospettiva concreta sui metodi e sugli strumenti per una simile ricerca; l'eccessivo peso attribuito all'utilità o all'esistenza di un «mercato» per una conoscenza del passato finalizzata a scopi meramente pratici. Punto centrale per il nostro ragionamento è il fatto che il libro portava alla ribalta diverse correnti di studio che, fin dagli anni Novanta del secolo scorso, non avevano mancato di alimentare nuove ricerche. Fra queste vi era, in primo luogo, la *big history*.

Apparsa all'inizio nel contesto accademico statunitense, questo particolare filone, fortemente multidisciplinare, si proponeva di essere «una nuova storia e scienza della natura», richiamando le vecchie storie universali, ma andando al di là delle frontiere tradizionali, grazie a una collaborazione con le scienze dure, in particolare la geologia o la cosmologia. Attraverso una vera e propria rivoluzione cronologica, spingendosi indietro fino a 13 milioni di anni, l'obiettivo è di storicizzare le scienze naturali e mostrare il cambiamento come

7. Guldi e Armitage, *The History Manifesto*, pp. 61-116.

8. Guldi e Armitage, *The History Manifesto*, pp. 103-106.

un risultato di lenti mutamenti in un vasto periodo di tempo e non di leggi generali⁹. I temi della *big history* sono molti, ma in sostanza si ritrovano insieme nello studio del ruolo dell'uomo all'interno del reame biologico al fine di mostrare come l'umanità abbia alterato il processo di selezione naturale e come i cambiamenti della biosfera non siano recenti, ma di lunghissimo periodo. Una simile storia unirebbe il vasto arcipelago di conoscenze fra discipline e traccerebbe così «una *singola e coerente* storia del cambiamento del tempo» (corsivo mio)¹⁰.

La lunga durata della prospettiva della *big history* si accompagna, e talvolta si lega, a un secondo aspetto, l'aumento di dati a disposizione per studiare quei fenomeni e la crescita di ricerche storiche con metodi quantitativi. Come ha mostrato Patrick Manning in un manuale introduttivo alla ricerca storica incentrata sui Big Data, questi ultimi, a differenza dei dati oggi raccolti grazie ai cellulari o alla navigazione sui computer, devono essere selezionati, digitalizzati, documentati e trasformati, grazie a un processo di identificazione, approvazione, integrazione e standardizzazione che permette una successiva visualizzazione e analisi. Se l'obiettivo principale è quello di arrivare a comparare i modelli del passato con quelli del presente, il rischio – ricordato dallo stesso autore – è la *selettività* con cui il ricercatore (o meglio, il gruppo di ricerca) predilige alcuni dati (sul commercio, le migrazioni, la popolazione o i prezzi) su altri¹¹.

Ciononostante, l'accumulo di informazioni è in continuo aumento. La grande disponibilità di dati sulle popolazioni del passato, sulle transazioni finanziarie, sulla sicurezza sociale o sui fenomeni ambientali (temperature, precipitazioni, ecc.) ha portato oggi ad avere miliardi di micro-dati, suggerendo nuove piste di ricerca e interpretazioni su determinati comportamenti, dai processi di *gentrificazione* ai percorsi occupazionali, dalla mobilità economica ai cicli di vita, con analisi multi-livello (demografico, familiare, lavorativo) e intergenerazionali, grazie anche alla possibilità di geo-referenziare gli stessi e legare elementi quali il genere, la provenienza e lo status socio-economico¹². A seguito dei contemporanei sviluppi nell'umanistica digitale,

9. David Christian, *Big History: The Longest "Durée"*, «Österreichische Zeitschrift Für Geschichtswissenschaften», 20 (2009), pp. 91-106, in particolare p. 98.

10. Christian, *Big History*, p. 104; David Christian, *The Case for "Big History"*, «Journal of World History», 2 (1991), pp. 223-238. Il campo è già molto ampio, vedi in particolare: Fred Spier, *The structure of big history: from the Big Bang until today*, Amsterdam 1996; David Christian, *Maps of time: an introduction to big history*, Berkeley 2005; Fred Spier, *Big history and the future of humanity*, Oxford 2010.

11. Patrick Manning, *Big Data in History*, Londra 2014, pp. 2-5.

12. Steven Ruggles e Diana L. Magnuson, *The History of Quantification in History: The JIH as a Case Study*, «The Journal of Interdisciplinary History», 50 (2018), pp. 363-381, p. 294; Myron Gutmann, *Quantifying Interdisciplinary History: The Record of (Nearly) Fifty Years*, «The Journal of Interdisciplinary History», 50 (2020), pp. 517-545, soprattutto pp. 541-

poi, l'incontro fra econometria, statistica e linguistica computazionale ha permesso l'utilizzo anche di fonti quali giornali o riviste che in precedenza non erano impiegate e l'ampliamento dello spettro di domande da porre e di temi da analizzare: dal comportamento economico al pensiero, dai gusti alle motivazioni fino ai valori degli individui¹³. Vi è così chi propone una storia (economica) 4D, ovvero *Digitally-Driven Data Design* (concepita su dati animati digitalmente): grazie ad ampie raccolte di documenti storici e all'aumento della potenza dei computer sarebbe possibile riesaminare domande classiche con nuovi dati e strumenti tecnici, per immaginare e raffigurare il passato in maniera diversa (grazie in particolare alla tecnologia Gis), aprendosi a inedite collaborazioni interdisciplinari, insegnando ai computer a riconoscere modelli nel linguaggio, impararli e generare nuove banche dati di interesse¹⁴. Il passato diviene così un utile campo per testare e informare la teoria, per comprendere i meccanismi di cambiamento e per rispondere a grandi questioni¹⁵. Il punto centrale è che questi discorsi non sono del tutto nuovi, ma riportano alla mente discussioni già avvenute intorno alla metà degli anni Settanta.

2. Quantità: un ritorno?

Verso la metà del secolo scorso, l'incontro fra storia e scienze sociali aveva provocato l'ambizione a elaborare una storia scientifica, una «nuova storia» che mirasse a produrre leggi generali per comprendere il cambiamento storico¹⁶. La scuola delle *Annales* in Francia, la storia sociale marxista nel Regno Unito e la nuova storia economica negli Stati Uniti (altre volte detta cliometria), vedevano nel dialogo con i modelli delle scienze sociali (economia e sociologia *in primis*) l'aspirazione alla costruzione di lunghe serie di dati, raccolti geograficamente su aree sempre più ampie e in maniera cumulativa, grazie ai quali poter testare modelli paradigmatici, talvolta controfattuali, grazie all'applicazione di formule matematiche¹⁷. Una storia che non è quantificabile non può ritenersi scientifica, scriveva Le Roy Ladourie, e lo storico avrebbe dovuto trasformarsi in un programmatore, perché le serie storiche

543; Myron Gutmann, Emily Klancher Merchant e Evan Roberts, "*Big Data*" in *Economic History*, «The Journal of Economic History», 78 (2018), pp. 268-299, p. 272-274.

13. Gutmann, *Quantifying Interdisciplinary History*, pp. 541-543; Gutmann, Klancher Merchant e Roberts, "*Big Data*", pp. 283-284.

14. Mitchener, *The 4D Future of Economic History*, pp. 1238-1389

15. Abramitzky, *Economics*, p. 1248 (si riferisce alla teoria economica, ma la riflessione è estendibile anche ad altri ambiti).

16. Lawrence Stone, *The revival of narrative: reflections on a new old history*, «Past & Present», 85 (1979), pp. 3-24, p. 5.

17. Stone, *The revival of narrative*, pp. 5-6.

erano necessarie per far vedere il «peso delle strutture» nel lungo periodo e i modelli oggettivi che condizionavano il comportamento, le azioni e i pensieri di (inconsi) attori sociali¹⁸.

Alla metà degli anni Settanta, tuttavia, si arrivò a una profonda divaricazione: se da un lato gran parte degli storici si orientò verso un approccio culturale e linguistico, che negava la vera accessibilità a ciò che vi era dietro al testo che proponeva invece una centralità a discorsi e rappresentazioni, la storia scienza sociale (*social science history*) intraprese un percorso sempre più scientifico. Gli effetti sono stati a volte drammatici, come in certi filoni della storia economica cliometrica. Quest'ultima è diventata in alcuni ambiti una branca dell'economia che applica la teoria (in particolare neoclassica) al passato per testarla grazie i fenomeni occorsi. Il problema principale di una simile prospettiva, però, è quello di identificare dati storici che contengano esperimenti naturali utili a verificare fenomeni di causalità e risultati differenti in popolazioni osservate nel corso dei secoli¹⁹.

In altri ambiti l'impatto è stato diverso. Alcuni recenti bilanci apparsi sul *Journal of Interdisciplinary History* mostrano bene come, a fronte di una diminuzione nel periodo 1990-2010, nell'ultimo decennio la percentuale di articoli che utilizzano metodi quantitativi sia tornata ad aumentare, così come si sia ampliata la provenienza degli autori fuori dagli Stati Uniti²⁰. Quali sono le principali caratteristiche di questa storia quantitativa, almeno alla luce degli articoli pubblicati in quella sede, rivista che comunque ha insistito molto sul dialogo fra storia e scienze sociali e sull'applicazione dei metodi di queste ultime in ambito storico? Da un lato vi è un utilizzo maggiore di una statistica descrittiva e non inferenziale (seppur in aumento siano l'uso di correlazioni, regressioni e della statistica spaziale); dall'altro si riscontra sempre di più la presenza di nuovi strumenti per la raccolta di dati insieme a un uso massivo di Big Data e a nuovi approcci alle fonti, come la storia e l'umanistica digitale²¹. Questi approcci tornerebbero poi in auge non solo per una crescente sfiducia verso il post-modernismo, ma anche per il dialogo con nuovi strumenti e nuovi metodi che – sempre a detta di alcuni studiosi – renderebbero più

18. Stone, *The revival of narrative*, p. 5; William H. Sewell, *Logics of History: Social Theory and Social Transformation*, Chicago 2005, p. 28; Jan de Vries, *Changing the Narrative: The New History That Was and Is to Come*, «The Journal of Interdisciplinary History», 48 (2017), pp. 313-234; Nani, *La metà perduta*.

19. de Vries, *Changing the Narrative*, pp. 322 e 327; Mocarelli, *Storia ed economia*; Naomi Lamoreaux, *The Future of Economic History Must Be Interdisciplinary*, «The Journal of Economic History», 75 (2015), pp. 1251-1257.

20. Ruggles e Magnuson, *The History of Quantification*: anche se vi è da sottolineare come l'ampliamento sia un effetto della richiesta dei processi di valutazione in alcuni paesi, come l'Italia.

21. Steven Ruggles, *Big Microdata for Population Research*, «Demography», 51 (2014), pp. 287-297.

semplice e meno dispendiosa la ricerca, legando spazialmente e tematicamente i dati al fine di interrogare una miriade di argomenti²².

Mi sembra tuttavia che gli aspetti all'origine dell'allontanamento dalla storia quantitativa allora siano presenti, irrisolti e talvolta ignorati ancor oggi in questa recente apertura verso i Big Data o l'umanistica digitale. Mi riferisco in particolare a una discussione su che cosa sia un dato storico e su come si possa interrogare e utilizzare²³. Si tratta di un aspetto essenziale, perché riguarda la metodologia della ricerca e l'epistemologia stessa della disciplina. La crescita dei Big Data ha portato di certo la ricerca storica al di là della statistica inferenziale, grazie a un universo di dati a disposizione che permetterebbe di evitare di desumere le caratteristiche di una popolazione da quelli di un campione²⁴. Ma questo fatto porta ancor più al centro il problema non tanto degli strumenti con cui trattiamo un dato storico, quanto invece della concezione stessa di dato e del processo che guida la ricerca. La critica centrale alla prima stagione di storia quantitativa fu proprio questa: l'assenza di una riflessione sui dati, sulle categorie utilizzate e sulla loro storicità²⁵.

Molte ricerche che guardano al quantitativo o a prospettive di lungo periodo fanno ricorso a dati ricavati da fonti aggregate, definizioni uniformi e tassonomie costruite ex-ante da un singolo o da un gruppo di ricercatori, anche attraverso discussioni e condivisioni da più parti del mondo per cercare di superare le spesso lamentate barriere culturali e linguistiche e per interpretare processi che interrogano almeno cinque secoli del recente passato, se non oltre. Mi riferisco in primo luogo ai progetti avviati nell'ambito della storia globale delle migrazioni e delle relazioni di lavoro, ad alcune delle quali lo scrivente ha partecipato in prima persona e ha condiviso le difficoltà delle fasi concettuali e di raccolta delle informazioni²⁶. Soffermarsi sul processo di produzione dei dati e quindi delle tassonomie o categorie è quindi fondamentale.

22. Ruggles e Magnuson, *The History of Quantification*, pp. 380-382.

23. In parte, quelle procedure di codificazione, programmazione, inserimento e processo dei dati di cui parlava Stone, *The revival of narrative*, pp. 11-12.

24. Gutmann, *Quantifying Interdisciplinary History*, p. 523.

25. Lemerrier, *Une histoire*, pp. 351-354.

26. Jan Lucassen e Leo Lucassen, *The Mobility Transition Revisited, 1500-1900: What the Case of Europe Can Offer to Global History*, «Journal of Global History», 4 (2009), pp. 347-377; *Globalising Migration History: The Eurasian Experience (16.-21. Centuries)*, edited by Jan Lucassen and Leo Lucassen, Leiden 2014; Leo Lucassen, *Working Together: New Directions in Global Labour History*, «Journal of Global History», 11 (2016), pp. 66-87; Elise van Nederveen Meerkerk, *Big Questions and Big Data: The Role of Labour and Labour Relations in Recent Global Economic History*, «International Review of Social History», 62 (2017), pp. 95-121; Karin Hofmeester e C. Moll-Murata, *Big Questions and Big Data: A Reply from the Collaboratory*, «International Review of Social History», 62 (2017), pp. 123-130.

3. I dati storici

Una riflessione sulle premesse metodologiche che caratterizzano il lavoro dello storico è fondamentale non solo per arrivare a una coerenza interna dell'analisi, ma anche per smentire la percezione, diffusa in molte scienze sociali e naturali, che la storia sia una semplice disciplina descrittiva, un repertorio di esempi validi per ogni utilizzo e il campo di applicazione di teorie elaborate altrove²⁷. La scienza storica non può esimersi dall'elaborare un suo metodo d'analisi coerente con i dati che analizza. Questo percorso dovrà partire dal confronto con le domande provenienti dalle scienze sociali e naturali, dalla conoscenza critica dei dati con cui entra in relazione e dalla loro successiva quantificazione; dalla scelta di scala, o meglio prospettiva, che vorrà adottare nell'analisi; fino all'elaborazione di (nuove) ipotesi passibili a loro volta di essere testate. Nelle pagine seguenti cercherò di discutere questo percorso, a partire dalla conoscenza critica dei dati, lasciando il rapporto con le altre scienze per l'ultima parte.

Una riflessione su cosa siano i dati è fondamentale per qualsiasi disciplina che si propone di raccogliarli, analizzarli, quantificarli. Che cos'è un dato *storico*? La domanda non è banale. Lo scienziato storico infatti si rapporta con dati desunti da fonti che, a loro volta, sono *tracce* che ci sono state lasciate dal passato, in maniera parziale e il cui livello di deformazione è assai variabile a causa dei rapporti di potere che le hanno prodotte e le traversie che le hanno portate a noi²⁸. Questa circostanza è diametralmente opposta a molte scienze sociali dove invece il ricercatore può costruire i propri dati, seppur in maniera controllata attraverso la selezione del campione e con procedure formalizzate di raccolta delle informazioni²⁹. La parzialità e il livello di deformazione delle fonti storiche, invece, pone il ricercatore storico di fronte al suo primo obiettivo, che è quello di interpretare la fonte, leggerla in contropelo o al di là della superficie³⁰. Non può, in sostanza, considerare la fonte come un semplice repertorio o contenitore da cui trarre le informazioni ivi presenti³¹. Bisogna invece intraprendere un'operazione di analisi del processo di produzione della fonte che è necessaria per comprendere ed estrarre dati

27. Favero, *Sul metodo storico*, p. 58.

28. Edoardo Grendi, *Micro-analisi e storia sociale*, «Quaderni storici», 12 (1977), pp. 506-520; *Il piccolo, il grande, il piccolo. Intervista a Giovanni Levi*, «Meridiana», 10 (1990), pp. 211-234; Carlo Ginzburg, *Il filo e le tracce: vero, falso, finto*, Milano 2006 (4. ed.).

29. Favero, *Sul metodo storico*, p. 62.

30. Ginzburg, *Il filo e le tracce*; Giovanni Levi, *Frail Frontiers?*, «Past & Present», 242 (2019 Supplement), pp. 37-49.

31. Simona Cerutti e Isabelle Grangaud, *Sources and Contextualizations: Comparing Eighteenth-Century North African and Western European Institutions*, «Comparative Studies in Society and History», 59 (2017), pp. 5-33, soprattutto p. 9.

neutri e separati dalla specificità in cui sono stati costruiti. Solo a quel punto – e non prima – è possibile estrarre e porre le informazioni collettive in una struttura predefinita³².

Questo fatto è ben evidente nelle ricerche che si occupano di prezzi, salari e standard di vita, le quali si accingono a raccogliere le informazioni, talvolta in maniera parziale, per costruire lunghe serie di dati, anche plurisecolari. Il problema è che una fonte contabile riporta sì un salario o un prezzo, ma la registrazione effettuata e il numero quindi rinvenibile nel documento nascondono in realtà molti altri significati e transazioni che non sono utilizzabili acriticamente per calcolarne l'andamento sul lungo periodo. Lo stesso può dirsi per le fonti notarili utilizzate per studiare il mercato della terra³³; per gli inventari *post-mortem* (giudiziari o notarili) con cui si calcola l'andamento dei consumi; o per le statistiche e le fonti aggregate che si usano nelle ricerche in ambito migratorio. In quest'ultimo campo di recente si è proposto di calcolare i fenomeni migratori a partire dalla percentuale di migranti coinvolti in «un'importante *esperienza migratoria*» nel corso della loro vita in confronto a coloro che invece non la compiono, senza considerare che le registrazioni burocratiche dei migranti operate dalle singole autorità si concentrano più sui *movimenti migratori*, registrando l'atto di arrivo o di partenza da una comunità (definita in base a un villaggio, una città, una provincia o uno stato). Ciò può provocare forti distorsioni, soprattutto su una prospettiva plurisecolare, portando a sotto-stimare le migrazioni complessive e le mobilità permanenti o frequenti³⁴.

L'analisi qualitativa della fonte è funzionale per capire se il processo che l'ha generata è cambiato o meno, con conseguenze decisive per la successiva analisi quantitativa³⁵. Lo studio delle cause giudiziarie è un esempio lampante, perché un minimo cambiamento nella procedura o nella gestione del tribunale può avere un grosso impatto nel tipo di cause discusse e nel relativo

32. Cerutti e Grangaud, *Sources and Contextualizations*, p. 9; Angelo Torre, *Un commento*, «Quaderni storici», 2 (2017), pp. 577-584.

33. Si vedano le osservazioni di Giovanni Levi, *L'eredità immateriale. Carriera di un esorcista nel Piemonte del Seicento*, Torino 1985 (nuova edizione 2020). Sui salari cfr. Francesca Trivellato, *Salaires et justice dans les corporations vénitiennes au XVIIe siècle: le cas des manufactures de verre*, «Annales. Histoire, Sciences Sociales», 54 (1999), pp. 245-273; Luca Mocarrelli, *Wages and the Labour Market in the Building Trade in 18th Century Milan*, «Jahrbuch für Wirtschaftsgeschichte», 2 (2004), pp. 83-96; Andrea Caracausi, *I giusti salari nelle manifatture della lana di Padova e Firenze (secoli XVI-XVII)*, «Quaderni storici», 3 (2010), pp. 857-884.

34. Si veda il confronto fra Lucassen e Lucassen, *The Mobility Transition Revisited* (per il primo metodo) e Josef Ehmer, *Quantifying Mobility in Early Modern Europe: The Challenge of Concepts and Data*, «Journal of Global History», 6 (2011), pp. 327-338, pp. 329-330 per le successive osservazioni.

35. Andrè Carus e Sheilagh Ogilvie, *Turning Qualitative into Quantitative Evidence: A Well-Used Method Made Explicit*, «The Economic History Review», 62 (2009), pp. 893-925.

svolgimento, che talvolta avviene comunque attraverso più cause e in via extragiudiziale³⁶. La fonte è infatti un'azione in se stessa e bisogna capire l'intenzionalità che vi è dietro dal momento che la sua produzione è situata in un contesto specifico dal quale è modellata, ma che modifica a sua volta³⁷.

Un dato storico è dunque qualcosa di diverso dal dato prodotto in laboratorio, in un esperimento o in gran parte della ricerca sociale, perché, come si è detto, è desunto da tracce, parziali e deformate, che il passato ci ha lasciato³⁸. Questo aspetto ci pone di fronte a un elemento ineludibile: la nostra conoscenza dell'oggetto di studio sarà giocoforza incompleta e questo ci spingerà il più delle volte a spiegare le fonti per verificare la loro affidabilità, per collegarle e per estrarre informazioni e dati passibili di essere inseriti in categorie più ampie, attraverso un paradigma che è stato definito indiziario³⁹. La presenza di un nome nel registro delle cariche di un'organizzazione politica può essere meno rilevante della produzione di quello stesso registro che ci spiega invece che cosa e perché un determinato individuo viene registrato. Un dato che ricaviamo da una fonte è un'azione che deve essere interpretata e spesso si lega a un'altra diversa e inserita in un contesto differente. Bisogna domandarsi, ad esempio, il senso di mettere insieme i dati numerici tout court (durata, salario) desunti da contratti d'apprendistato lontani nel tempo e nello spazio per crearne una media, quando il singolo contratto è legato alle controparti che lo stipulano e quindi a relazioni (famigliari, amicali o comunitari) che vanno al di là del contratto d'apprendistato stesso⁴⁰. Questo processo di connettere fra loro dati simili, ma provenienti da rapporti diversi, è presente in molti lavori che si propongono di costruire grandi banche dati. Il rischio è di soffermarsi su medie e comparare situazioni radicalmente differenti. L'influenza di legami presenti in altre sfere, come mostrato qualche decennio fa dall'antropologo Fredrick Barth, vale da monito: l'azione (e quindi il dato) di acquisire di una carica, ad esempio politica, in una determinata istituzione può essere motivata da ragioni esterne a quell'istituzione e da legami precedenti o da intenzioni successive in altre sfere della medesima società⁴¹. Secondo queste considerazioni, quindi, un elenco – anche il più completo –

36. Renata Ago e Simona Cerutti, *Premessa*, «Quaderni storici», 2 (1999), pp. 307-314.

37. Torre, *Un commento*, p. 582.

38. Ginzburg, *Il filo*, cap. 13.

39. Favero, *Sul metodo storico*, p. 62. Su questo punto vedi ovviamente Carlo Ginzburg e Carlo Poni, *Il nome e il come: scambio ineguale e mercato storiografico*, «Quaderni storici», 14 (1979), pp. 181-190.

40. Andrea Caracausi, *The price of an apprentice: contracts and trials in the woollen industry in sixteenth century Italy*, «Mélanges de l'École française de Rome - Italie et Méditerranée modernes et contemporaines» [En ligne], 128-1, 2016, mis en ligne le 3 mars 2016, consulté le 29 mai 2020.

41. Fredrik Barth, *Balinese Worlds*, Chicago 1993, p. 111.

dei nomi di coloro che hanno ricoperto delle cariche può non rivelarci queste motivazioni, a meno che non sia incrociato con altre fonti (o tracce), talvolta meno ricche d'informazioni.

4. Quantità vs. qualità?

Le considerazioni sopra esposte non vogliono essere un segno di sfiducia verso la costruzione di banche dati o verso gli approcci quantitativi. Tutt'altro. L'obiettivo è di invitare a un'analisi più critica di ciò che è un dato storico e del suo utilizzo, che può essere potenzialmente più rilevante di quanto non lo sia oggi. Si vuole anche far riflettere sull'inconsistenza dell'opposizione fra qualitativo e quantitativo poiché i due approcci si basano reciprocamente l'uno sull'altro. Il processo d'immersione nelle fonti è necessariamente di tipo qualitativo, per operare quella critica della produzione delle fonti e afferrare le intenzioni dei partecipanti. Per avere una conoscenza del significato delle categorie e dei parametri di base del contesto sociale di produzione della fonte, però, è necessaria un'operazione comparativa (anche con altre società) e statistica che è gioco-forza quantitativa, perché fa riferimento alla distribuzione di alcune variabili su uno spettro di valori possibili⁴². È spesso difficile dire dove inizia un momento e dove termina l'altro: sia perché forme discorsive e quantitative sono prodotte in maniera simile, affidandosi a sistemi simbolici di comunicazione e fissando le osservazioni di qualche sorta, sia perché testi, in larga parte discorsivi, sono sempre usati per interpretare i dati e le categorie su cui si fondano le analisi quantitative⁴³. Un'operazione dunque – quella quantitativa e comparativa – che, a opinione di chi scrive, tutti gli storici fanno sempre, seppur alle volte inconsciamente e senza rappresentarla chiaramente con tabelle o grafici.

L'operazione quantitativa è necessaria anche per altri aspetti. Ci consente di interrogarci sul senso dei dati e delle dichiarazioni dell'epoca analizzata, andando talvolta a risolvere dubbi circa l'affidabilità o l'inconsistenza fra le fonti analizzate in un primo momento in maniera qualitativa; ci permette di effettuare un primo test sulle ipotesi e sulle teorie provenienti dalle scienze sociali, per capire ad esempio quali leggi erano applicate e quali norme influenzavano i comportamenti⁴⁴. L'approccio quantitativo ha poi il pregio di portarci a esplicitare le nostre procedure di ricerca, definendo e ridefinendo le ipotesi; a prendere coscienza dei limiti delle fonti e a vederle sotto una luce

42. Carus e Ogilvie, *Turning Qualitative*, p. 894.

43. Sewell, *Logics*, p. 370.

44. Claire Lemercier e Claire Zalc, *Méthodes quantitatives pour l'historien*, Paris 2008, pp. 17-8; Carus e Ogilvie, *Turning Qualitative*, p. 902.

diversa a seconda dei primi risultati; a interrogarci nuovamente se le codificazioni e le categorie elaborate dopo l'immersione qualitativa erano troppo riduttive⁴⁵. L'operazione qualitativo/quantitativo non è infatti unidirezionale, ma è costellata da andate e ritorni, proprio per la necessità di avvicinarsi sempre di più al linguaggio degli attori che sono sotto indagine, dipartendosi il più possibile dal meta-linguaggio del ricercatore. Contare nel tempo, infatti, ci pone sempre più di fronte al confronto con la storicità del vocabolario e delle categorie utilizzate⁴⁶.

Nelle scienze storiche si è ragionato forse troppo poco sulla dicotomia emico/etico, sull'utilizzo del linguaggio degli attori nel primo caso e sull'imporre agli attori le nostre categorie, nel secondo⁴⁷. Bisognerebbe in primo luogo discutere più a fondo fino a che punto sia possibile accedere alla prospettiva degli attori e avere categorie interne al loro contesto culturale o se invece non sia necessaria una «genuina sintesi», con categorie d'analisi il più «emico» possibili, ma che consentano di rispondere in maniera diretta alle domande del presente⁴⁸. Come è stato discusso nelle scienze geografiche, ad esempio, la conoscenza sviluppata da un ricercatore/ricercatrice è sempre parziale, dato il suo posizionamento e la sua collocazione nel tempo e nello spazio, nonché le sue «mappe di coscienza»⁴⁹. E del resto ogni categoria implica una scelta e tale scelta non è mai neutra: dare forza al caos dei dati, continuare a codificare e decodificare e chiarire il processo svolto può essere una via⁵⁰. Gran parte dei lavori di *big history*, costruiti a partire da larghissime banche dati e su ampi periodi di tempo, fanno però largo uso di un approccio «etico», con concetti più vicini al punto di vista del ricercatore o al gruppo che li ha elaborati. Il grosso problema è quello di non vedere quali sono le vere domande che possono emergere dall'analisi di un determinato problema: il concetto di «migrazione» o di «mobilità», e le forme che esso assume, possono differire non poco se si guarda ai movimenti migratori anziché ai migranti; ma, come ricordato più sopra, il problema della mobilità che per

45. Sewell, *Logics*, p. 370-371; Lemerrier e Zalc, *Méthodes*, pp. 17-18.

46. Lemerrier, *Une histoire*, p. 351.

47. Ma vedi Simona Cerutti, *Histoire pragmatique, ou de la rencontre entre histoire sociale et histoire culturelle*, «Tracés. Revue de sciences humaines», 15 (2008), pp. 147-168; Carus e Ogilvie, *Turning Qualitative*; Lemerrier e Zalc, *Méthodes*; Cerutti e Grangaud, *Sources and Contextualizations*. Come sottolinea lo stesso Carlo Ginzburg, *Our Words, and Theirs: A Reflection on the Historian's Craft, Today*, «Cromohs - Cyber Review of Modern Historiography», 18 (2014), pp. 97-114, la dicotomia etico/emico è comunque semplicistica, perché entrambi sono campi di discussione e conflitto per gli attori e per gli storici. Essere consoci di questa distinzione ci porta a liberarci da prospettive etnocentriche.

48. Lemerrier e Zalc, *Méthodes*, p. 42; Carus e Ogilvie, *Turning Qualitative*, p. 900.

49. Marina Bertoncin, Andrea Pase e Daria Quatrada, *Geografie di prossimità: prove sul terreno*, Milano 2014, pp. 20-21.

50. Lemerrier e Zalc, *Méthodes*, p. 47; Carus e Ogilvie, *Turning Qualitative*, p. 902.

alcune società può essere diverso nel tempo e nel contesto, così da rendere complesso ogni tentativo di quantificazione su concetti uniformi⁵¹.

Questa circostanza è ancor più delicata se si considerano gli sviluppi recenti dell'informatica, soprattutto umanistica, che ha portato a un utilizzo sempre più ampio di fonti usate in maniera qualitativa, grazie alla disponibilità di strumenti di analisi del testo e del suo contenuto. La ripetibilità effettiva di un dato storico, però, costituisce un problema: non solo per la singolarità dell'azione riportata nella fonte a cui si è fatto riferimento, ma anche perché fatti definiti semanticamente in maniera uguale da documenti simili in diversi momenti o contesti possono avere significati assai differenti. Le informazioni stesse sono registrate e classificate sulla base di convenzioni linguistiche ancor prima che formali e sono queste ultime che rendono i dati ricavati dalle fonti comparabili fra loro⁵². Lo storico deve interrogarsi sulla loro mutevolezza nel tempo, poiché il modo e il significato con cui gli oggetti vengono descritti e misurati varia e in maniera alle volte repentina e altre impercettibile. Gli studi sul lavoro hanno mostrato bene come le parole «maestro», «lavorante» e «garzone» (per non dire «lavoro») possono nascondere significati diversi a seconda non solo del secolo, ma anche della situazione in cui vengono impiegati nello stesso contesto spazio/temporale⁵³. Gli esempi potrebbero moltiplicarsi, includendo concetti quali disoccupazione, costo della vita, democrazia, repubblica e così via.

Ancor una volta, questo non significa che lo studio quantitativo dei fatti storici sia inutile, tutt'altro. Le serie quantitative, così come le categorie, devono essere costruite e ricostruite secondo i contesti di produzione delle fonti e delle domande di ricerca; allo stesso tempo, la tabella, il grafico o anche la visualizzazione dei dati non sono il fine della ricerca, ma ne rappresentano un punto intermedio, dal quale ripartire, alla luce dei limiti della fonte già ricordati in precedenza⁵⁴. Quello che uno storico deve fare è un esame sensibile delle fonti e delle ipotesi da cui è partito, esplicitando le procedure operate e senza esitare a riformulare le tesi o a cambiare le categorie se non c'è coerenza con la banca dati. Come è stato notato in altre sedi, si tratta di un'o-

51. Vedi ancora Ehmer, *Quantifying mobility*, p. 337, ma anche Simona Cerutti, *Étrangers: étude d'une condition d'incertitude dans une société d'Ancien régime*, Montrouge 2012.

52. Alain Desrosières, *Séries longues et conventions d'équivalence*, «Genèses. Sciences sociales et histoire», 9 (1992), pp. 92-97, pp. 92-93; Lemerrier, *Une histoire*, p. 351; Favero, *Sul metodo storico*, p. 63.

53. Andrea Caracausi, *Dentro la bottega. Culture del lavoro in una città d'età moderna*, Venezia 2008; *What Is Work? Gender at the Crossroads of Home, Family, and Business from the Early Modern Era to the Present*, edited by Anna, Bellavitis, Manuela Martini e Raffaella Sarti, New York 2018.

54. Lemerrier, *Une histoire*, p. 348; Karila-Cohen, Lemerrier, Rosé e Zalc, *Nouvelles Cuisines*, pp. 776 e 781.

perazione che spesso non si fa per la paura di essere accusati di «cucinare i dati», ma è un procedimento centrale per dare un senso alla ricerca e leggere in maniera efficace i dati quantitativi⁵⁵. La stessa messa in linea del testo integrale del documento può essere un'operazione parzialmente efficace, poiché prendere contatto con la *materialità* del documento è fondamentale per non perdere eventuali indici che dobbiamo tenere presente per quantificare⁵⁶.

Poiché le fonti sono tracce, il nostro oggetto di studio non è mai raggiungibile in modo completo e i nostri esperimenti sono diversi da quelli di altri scienziati, come storici dobbiamo rassegnarci al fatto che quando siamo di fronte a una statistica non è la media, né la deviazione standard, il cuore della nostra ricerca, come invece ama sostenere chi lavora sui Big Data⁵⁷, ma è talvolta ciò che è fuori riga, «l'anomalo». Come è già stato ricordato in molti contesti, un evento eccezionale, una testimonianza involontaria può essere rivelatrice di qualcosa di più generale⁵⁸. Ma questo «eccezionale normale», così come i processi di critica della produzione della fonte e di estrapolazione e trattamento dei dati che abbiamo più sopra ricordato, ci pone di fronte alla scelta degli sguardi e delle prospettive da adottare⁵⁹.

5. Giochi di scala

Molto si è discusso sulla dicotomia micro/macro e sulle scale da adottare per giungere a un livello pertinente di osservazione. Troppo spesso viene creata un'opposizione fra micro e macro che sarebbe superata solo da una combinazione delle due prospettive per produrre una sintesi più ragionata, «sensibile ed etica» rispetto ai dati storici. Anche nel *Manifesto per la Storia* con cui si è aperto questo intervento, il «micro» è visto come elemento utile a destabilizzare le narrative e altre forme di pensiero teleologico, ma è allo stesso tempo necessario combinarlo con una più ampia macro-storia. Questo approccio è spesso presente in molte storie globali o mondiali, dove il micro è equiparato talvolta al «locale» ed è speculare all'analisi macro (appunto, «globale»), un elemento che concorre a creare la struttura generale, divenendone l'eccezione se si discosta dal modello generale.

Tuttavia, questa prospettiva è erronea e fuorviante, come è stato più volte osservato in molti interventi che hanno discusso sui rapporti fra microstoria e storia globale, evidenziandone i possibili punti d'incontro o sottolineando-

55. Sewell, *Logics*, p. 371.

56. Karila-Cohen, Lemerrier, Rosé e Zalc, *Nouvelles Cuisines*, p. 778.

57. Gutmann, Klancher Merchant e Roberts, "*Big Data*", p. 287.

58. Grendi, *Micro-analisi*, p. 512.

59. Ginzburg e Poni, *Il nome*, pp. 187-198.

ne le incommensurabili distanze⁶⁰. Lo sguardo proposto dalla microstoria, almeno nella sua tradizione italiana, non aveva nulla a che vedere con la dimensione, piccola, grande o media, dell'oggetto. E non ha nulla a che fare con le domande, come gli stessi autori della *big history* vorrebbero sostenere con riferimento alla necessità di indagare il rapporto fra gli umani con altre specie o il ruolo degli stessi nell'universo⁶¹. Il *micro* della *microstoria* – non bisogna stancarsi di ricordarlo – era lo sguardo al microscopio, operazione necessaria per individuare che cosa vi fosse di sovra-rappresentato rispetto a uno sguardo a occhio nudo, dalle domande di partenza alle dinamiche economico-sociali, evitando tautologie, categorie o modelli prestabiliti. La stessa costruzione dell'unità spaziale è diversa nel procedimento microstorico: lo spazio non è il risultato oggettivo di una situazione geografica (o, meglio, geometrica), ma di una rete di significati e legami costruiti su quel luogo, inteso come costruzione sociale e culturale⁶². Le relazioni di prossimità vengono a giocare un ruolo fondamentale, sia nel rapporto con l'oggetto di studio, sia nel momento della sua analisi, per avvicinarsi a comprendere più a fondo i processi stessi di territorialità⁶³.

Il gioco di scala, inteso qui come l'azione di avvicinarsi il più possibile al proprio oggetto di studio alla stregua di uno sguardo al microscopio, è una diretta conseguenza delle premesse più sopra esplicitate riguardo alle fonti, ai dati e ai processi di qualificazione e quantificazione. Dal momento che siamo di fronte a tracce, il più delle volte deformate e la cui eccezionalità può essere rivelatrice di fenomeni più generali; dal momento che dobbiamo capire l'effettiva intenzionalità della fonte e del dato ricavato per poterlo utilizzare e legarlo ad altri; dal momento che la comparazione quantitativa può avvenire solo all'interno di una sua coerenza linguistica e concettuale; allora, se si accettano tutti questi punti, solo un approccio che sia il più vicino possibile agli individui può permetterci di comprendere se il nostro percorso è coerente con i limiti posti alla conoscibilità del nostro oggetto. A venir meno è quindi

60. La riflessione è stata molto ampia. Vedi, fra i più recenti, Christian G. De Vito, *Verso una microstoria translocale (micro-spatial history)*, «Quaderni storici», 3 (2015), pp. 815-833; Andreozzi, *Senso e potere*; Torre, *Un commento*; Bertrand Romain et Guillaume Calafat, *La microhistoire globale: affaire(s) à suivre*, «Annales. Histoire, Sciences Sociales», 73 (2018), pp. 1-18; John-Paul A. Ghobrial, *Introduction: Seeing the World like a Microhistorian*, «Past & Present», 242 (2019), Supplement, pp. 1-22; Levi, *Frail Frontiers?*; Jan de Vries, *Playing with Scales: The Global and the Micro, the Macro and the Nano*, «Past & Present», 242 (2019), Supplement, pp. 23-36.

61. Christian, *The Case for "Big History"*, pp. 226-227.

62. Levi, *Frail Frontiers?*, p. 40. Su questi temi il riferimento è ad Angelo Torre, *Luoghi: la produzione di località in età moderna e contemporanea*, Roma 2011.

63. Questi ragionamenti sono stati del resto oggetto di analisi approfondita nell'ambito delle scienze geografiche: vedi Bertocin, Pase e Quatrida, *Geografie*, in particolare pp. 12-14 e 94-96.

il problema della «rappresentatività», così caro alle scienze sociali, ma così poco rilevante e anzi fuorviante per le scienze storiche. Per l'analisi storica, infatti, c'entra poco se un caso di studio è rappresentativo o meno, poiché lo storico si confronta con tracce che gli sono state lasciate (e non con dati che può costruire e campionare), con fonti che sono deformate da rapporti di potere e dal passato (e che solo un'attenta ricostruzione e interrogazione può sciogliere) e con l'impossibilità di conoscere pienamente l'oggetto di studio. La rappresentatività di un caso di studio e la media (o la deviazione standard) prodotta da un'analisi statistica ha quindi poca rilevanza: se il problema che la storia si pone è, oltre alla comprensione del funzionamento di un sistema, il cambiamento dello stesso, allora il ricercatore dovrà essere interessato non soltanto alle medie, ma anche – e per certi versi soprattutto – a quando le interferenze diventano determinanti⁶⁴.

5. Teorie, ipotesi, generalizzazioni

Come ho ricordato all'inizio, la ricerca storica deve partire da un confronto con le domande che le altre scienze si pongono. Ecco che temi quali l'ineguaglianza, la *governance* mondiale, il cambiamento climatico, la crescita economica o i livelli di povertà, così come molti altri, sono centrali e con questi dobbiamo confrontarci. Il dialogo deve tuttavia essere aperto, costruttivo e coerente con il processo metodologico ed epistemologico più sopra delineato. Due punti però distinguono la storia dalle altre scienze, anche con rapporto all'analisi di grandi banche dati. Il primo, come si è visto, è la necessità di una metodologia che sia coerente con i propri dati e senza un'applicazione passiva di concetti e metodi elaborati altrove. Questi metodi vanno invece discussi e, se necessario, contestati. La storia non può essere una semplice operazione dedita a testare la teoria e non è possibile, ad esempio, comprendere i meccanismi di un fenomeno associando semplicemente una o più variabili a una variabile dipendente⁶⁵: bisogna comprendere come la totalità delle variabili agisce per costruire un sistema di relazioni.

Il secondo punto riguarda la parzialità e la provvisorietà delle risposte che possono essere date alle domande di ricerca. Siccome la realtà è più complessa di ogni sistema ideato per descriverla, e siccome il passato è inconoscibile nella sua interezza, il problema dello storico è quello di trovare un legittimo livello di semplificazione, identificato fra una infinità di casi differenti e il

64. Sigurdur Gylfi Magnússon, "The Singularization of History": *Social History and Microhistory within the Postmodern State of Knowledge*, «Journal of Social History», 36 (2003), pp. 701-735; Favero, *Sul metodo storico*, p. 70.

65. Abramitzky, *Economics*, p. 1248.

modello unico. Questo aspetto non c'entra nulla con il relativismo, ma è un richiamo crescente alla complessità e alla critica degli strumenti euristici usati dagli storici⁶⁶. Come è stato messo in luce dalle scienze geografiche, non è possibile rappresentare tutto il mondo per intero e una volta per tutte, ma bisogna cercare di vederlo da punti di vista molteplici per restituire una pluralità di immagini, metafore e categorie concettuali⁶⁷.

Questo non significa il rigetto delle generalizzazioni, come sostenuto dalla *big history* e, talvolta, anche dalla macrostoria o dalla *social science history*⁶⁸. La microstoria si propone anch'essa come una nuova forma di generalizzazione contro le semplificazioni delle spiegazioni macrostoriche, in particolare contro l'idea di un'uniformità del comportamento degli individui, della catena causale degli eventi (che osserviamo *ex-post*), della linearità del processo e dell'adattabilità del modello a ogni contesto spazio-temporale. Lo sforzo dello storico deve essere rivolto a mostrare la presenza di diverse possibilità di scelta e comportamento, di diverse reazioni a seconda di luoghi, contesti, culture, usanze e volontà: una generalizzazione che non deve negare il disordine della realtà e che deve evitare di cadere nella tentazione di voler per forza offrire delle statistiche verificabili per spiegarla⁶⁹. Il rapporto con il dato quantitativo non può essere meccanico e riduttivo, ma deve essere molto più riflessivo delle implicazioni epistemiche della costruzione delle nostre banche dati e delle interrogazioni ermeneutiche delle stesse⁷⁰. Il contributo della storia davanti ai Big Data non può essere quindi limitato alla raccolta di miliardi di dati per operare delle interrogazioni o delle correlazioni al fine di testare una determinata teoria, dando risposte generali e uniformizzanti. A partire da un problema dobbiamo costruire delle banche dati, anche grandissime, ma coerenti con le fonti utilizzate, per contribuire all'elaborazione di una nuova teoria e di nuove domande, anche in presenza di un numero sempre più ampio di fonti, dati e oggetti. I Big Data dovrebbero spingere le frontiere della conoscenza storica non tanto per elaborare una nuova «singola e coerente storia del cambiamento del tempo»⁷¹, quanto invece per offrire molteplici narrazioni utili a proporre nuovi sguardi, nuove domande e nuove prospettive al presente.

66. Giovanni Levi, *The Origins of the Modern State and the Microhistorical Perspective*, in *Mikrogeschichte Makrogeschichte. Komplementar oder inkommensurabel*, a cura di Giovanni Levi, Charles Tilly, Maurizio Gribaudi, Jürgen Schlumbohm, Göttingen 1998, pp. 53-82.

67. Bertocin, Pase, Quatrada, *Geografie*, p. 37 (con riferimento ai risultati del lavoro di campo).

68. Christian, *The Case for "Big History"*, p. 224; De Vries, *Playing with Scales*, p. 24.

69. Levi, *The Origins*, p. 55; Levi *Frail Frontiers?*, pp. 43-34; Sewell, *Logics*, p. 372.

70. Sewell, *Logics*, p. 375.

71. Christian, *Big History*, p. 104.

Il rischio più grave, ben presente nei progetti di *big history*, è quello di produrre una generalizzazione unica, di ricostruire l'immagine di un progresso unico e di una storia lineare, quando invece nel mondo reale esistono false partenze e incongruità. Lo stesso dato che si riferisce a un evento recente può legarsi non a quello che lo precede immediatamente, ma a cause remote, in maniera quindi non lineare⁷². La storia dovrebbe essere quindi funzionale a mettere in crisi i concetti presi per assodati e a far emergere nuove domande: a partire da un tema generale si procede per testare molteplici risultati, porre nuove questioni e far emergere nuovi problemi⁷³.

Se la caratteristica dei Big Data è quella di eccedere le capacità analitiche umane e si spinge fino ai confini delle capacità analitiche dei computer⁷⁴, la nostra sfida è quella di operare un metodo d'analisi coerente con la produzione di quei dati e analizzarli di conseguenza. Anziché ridurre a una piccola visualizzazione un archivio di dati altrimenti troppo grande da leggere e condensarli per raccontare meglio la storia⁷⁵, il nostro obiettivo dovrebbe essere quello di identificare, fra un'infinità di casi differenti e il modello unico, una serie di generalizzazioni legittime e utili a comprendere il vasto spettro di possibilità, passate e presenti⁷⁶. Queste generalizzazioni non sono una tipologia, ma sono ipotesi che sono a loro volta essenziali nell'indirizzare la ricerca mediante le prospettive e le scale di osservazione, portando alla luce oggetti altrimenti trascurati. Lunghi dall'applicare la teoria allo studio dei fatti, gli storici usano i loro casi di studio per costruire nuove ipotesi teoriche⁷⁷. Tenute presenti queste premesse, anche una sintesi potrà essere operata, a patto che sia frutto di un preliminare lavoro sulle fonti primarie e di un'attenta critica e selezione di fonti secondarie, e il cui obiettivo non sia di chiudere il problema, ma anzi contribuire ad aprire nuove ipotesi.

Un simile processo, che nasce dall'identificazione di problemi generali, ma che procede attraverso un metodo di ricerca coerente con il proprio oggetto di studio e che giunge alla formulazione di nuove ipotesi e domande generali, può contribuire a ridare alla storia la sua legittimità. Quest'ultima non deriva tanto dalla sua capacità di essere compresa e sposata da un più vasto pubblico di non esperti; né da semplici visualizzazioni, realizzate grazie all'utilizzo acritico di strumenti digitali; né dalla creazione di sintesi politicamente informate per «parlare al potere» e incidere nelle questioni riguardanti il mondo attuale o quello futuro. L'agenda di ricerca non può essere dettata

72. Levi, *Frail Frontiers?*, p. 41.

73. Levi, *Frail Frontiers?*, p. 45.

74. Gutmann, Klancher Merchant e Roberts, "*Big Data*", p. 270.

75. Armitage e Guldi, *The History Manifesto*, p. 89.

76. Levi, *The Origins*, p. 62.

77. Favero, *Sul metodo storico*, p. 65.

dall'esterno da un gruppo elitario di scienziati. Sta invece a noi trovare la rilevanza delle nostre domande durante il processo di ricerca (e non *a priori*): domande generali che servono a capire le differenze di situazioni distinte, ma anche a fornire risposte particolari che possano contribuire alla costruzione di alternative⁷⁸.

Anziché raccogliere migliaia di dati per creare nuove narrazioni o visualizzazioni, il contributo degli storici dovrebbe essere quello di esercitare a riflettere sulla costruzione dei dati e sulle molteplici interrogazioni possibili, per poi quantificare di conseguenza; anziché semplificare per comunicare, l'obiettivo centrale dovrebbe essere quello di comunicare la complessità dei modelli e delle diverse teorie possibili; anziché formare un unico pensiero che appiattisca le differenze, è necessario allenare menti critiche capaci di vedere che dietro a slogan o a messaggi all'apparenza semplici vi è una realtà che è molto più complessa. Solo così potremmo tornare a dare legittimità alla scienza storica e a offrire strumenti concreti per agire sul presente.

Abstract

This paper aims to discuss the possible encounters between big data analysis and historical sciences. After a brief presentation of the new fields of big history and deep history, I will resume the discussion arose around the role of quantitative history during the 1970s. Later I will open a discussion on the importance of using an historical method that takes into account the nature of the historical data, avoiding any misunderstanding of the sources or a divide between quantitative and qualitative analysis and between scale of analysis. Finally, I focus on the relationship between history and other sciences and the type of historical "generalization" that can contribute to a more critical analysis of the world.

78. Levi, *Frail Frontiers?*, p. 45.

Bo2022: un database online di studenti e docenti dell'Università di Padova (1222-secolo XX)

di Pierluigi Terenzi*

1. Introduzione

Nel 2022 l'Università di Padova festeggerà gli ottocento anni della sua esistenza. Nel 1222 un gruppo di studenti bolognesi, a causa di un controllo del comune ritenuto eccessivo, si trasferì nella città veneta fondando un nuovo *studium*. Diversamente da altre esperienze simili dell'epoca, lo studio padovano riuscì a durare nel tempo, anche se le tracce della sua esistenza si fanno flebili durante la dominazione di Ezzelino da Romano (1237-1256). Subito dopo, però, l'istituzione fu capace di riprendere vita e rafforzarsi, ottenendo il riconoscimento papale nel 1264 – necessario perché i titoli accademici, conferiti dal vescovo, avessero validità – quando gli statuti cittadini includevano già norme di protezione per gli studenti e sostegno all'attività universitaria. La solidità dello *studium* fu tale che prima i signori Da Carrara (dal 1318) e poi Venezia, che nel 1405 inglobò Padova, lo sostennero permettendogli di diventare nei secoli seguenti uno dei più importanti atenei d'Europa¹. Oltre che dal sostegno politico, questo risultato dipese ovviamente da studenti e docenti provenienti dall'Italia e dall'Europa per apprendere e insegnare discipline giuridiche, grammaticali, filosofiche, mediche e teologiche. Fra loro, ci

* Dipartimento di Storia, Archeologia, Geografia, Arte e Spettacolo, Università degli Studi di Firenze.

1. Fra i numerosi studi sulla storia dell'ateneo, mi limito a richiamare: Nancy G. Siraisi, *Arts and Science at Padua. The Studium Before 1350*, Toronto 1973; Girolamo Arnaldi, *Le origini dello Studio di Padova. Dalla migrazione universitaria del 1222 alla fine del periodo ezzeliano*, «La Cultura. Rivista di Filosofia, Letteratura e Storia», 15 (1977), pp. 388-431; Paolo Marangon, *Ad cognitionem scientiae festinare. Gli studi nell'Università e nei conventi di Padova nei secoli XIII e XIV*, a cura di Tiziana Pesenti, Trieste 1997; Donato Gallo, *Università e signoria a Padova dal XIV al XV secolo*, Trieste 1998; *L'Università di Padova. Otto secoli di storia*, a cura di Piero Del Negro, Padova 2002; *L'università di Padova nei secoli (1222-1600). Documenti di storia dell'Ateneo*, a cura di Piero Del Negro e Francesco Piovan, Treviso 2017.

furono personaggi di rilievo assoluto nei vari ambiti, come Jacopo dell'Arena per il diritto, Pietro d'Abano per la filosofia e la medicina, Galileo Galilei per la scienza e molti altri che insegnarono o studiarono a Padova².

Per valorizzare il contributo di queste persone e conoscere meglio la “popolazione accademica” padovana nel corso dei secoli, in occasione dell'ottocentesimo dell'ateneo il Dipartimento di Scienze storiche, geografiche e dell'antichità (Dissgea), il Centro per la storia dell'Università di Padova (Csup) e il Centro di ateneo per i musei (Cam) dell'Università di Padova hanno promosso la realizzazione di un database online che raccolga dati su studenti, docenti e altre figure dello *studium* dalla sua fondazione al secolo XX. L'iniziativa si connette ad altre attività previste per le celebrazioni, coordinate dal Rettorato. In particolare, il database fungerà da bacino di informazioni per la redazione dei volumi previsti sulla storia dell'ateneo, che saranno pubblicati nella collana *Patavina libertas. Una storia europea dell'Università di Padova*³. Il database offrirà un contributo fondamentale soprattutto ai volumi incentrati sui temi della libertà religiosa, politica e del sapere, delle donne, e della mobilità accademica. Su quest'ultimo tema il database realizza un'ulteriore connessione, poiché sostiene alcune delle ricerche in corso nell'ambito del progetto di eccellenza Dissgea *Mobility and the Humanities*, dedicato allo studio delle varie forme e dei soggetti della mobilità (persone, oggetti, idee, informazioni, teorie e metodi)⁴.

Un unico database si trova dunque al centro di una rete di attività e gruppi di ricerca, coltivando pertanto l'ambizione di offrire strumenti per indagini di diverso tipo e su diversi periodi. Ciò è possibile grazie alla flessibilità dello strumento adottato e all'attenta configurazione che se n'è fatta, volta anche a commisurare le esigenze di normalizzazione tipiche di un database con il rispetto delle sfumature che caratterizzano le fonti storiche.

2. Il database: aspetti basilari

Il database online – <https://patavini.800anni.unipd.it>, attualmente ad accesso riservato per i collaboratori, ma che sarà reso pubblico – è stato costruito non soltanto per essere popolato direttamente, ma anche per importare i dati già raccolti in diversi progetti da vari studiosi, con metodi differenti. Una delle sfide nel configurare lo strumento web era proprio questa: definire una

2. Cfr. *Clariores. Dizionario biografico dei docenti e degli studenti dell'Università di Padova*, a cura di Piero Del Negro, Padova 2015.

3. Per altre iniziative, si consulti la pagina web dedicata: www.ottocentariouniversitadi-padova.it.

4. Per approfondimenti: www.mobilityandhumanities.it.

struttura e dei meccanismi che funzionassero bene per tutti i bacini di dati già esistenti, oltre che per il nuovo data input. A questa esigenza avrebbe potuto rispondere un web database costruito *ad hoc*, ma la necessità di mettere lo strumento rapidamente a disposizione dei vari gruppi di ricerca e l'esistenza di piattaforme preimpostate ma configurabili ha spinto verso un'altra direzione. La scelta è ricaduta su *nodegoat*, una piattaforma utilizzata anche da altri progetti di storia dell'università sui quali torneremo.

Nodegoat (<https://nodegoat.net>) è una piattaforma web per la gestione dei dati, la network analysis e la visualizzazione dei dati stessi, creata da Pim van Bree e Geert Kessels nel 2011. Permette la creazione e la gestione di un numero illimitato di set di dati usando un'interfaccia grafica, sulla base di un modello creato dall'utente. È pensato specificamente per l'ambito umanistico e storico in particolare, perché è dotato di funzionalità in grado di riferire al tempo e allo spazio ogni informazione immessa e di visualizzare dinamicamente queste informazioni cronoreferenziate e georeferenziate. L'approccio di Nodegoat è *object-oriented*: ciò significa che persone, eventi, oggetti e fonti sono considerati e trattati allo stesso modo, come *objects*. Le gerarchie fra *objects* dipendono solo dalle relazioni fra loro, non dalla loro natura intrinseca (cioè dalla qualità degli *objects*). Ciò permette, da un lato, di costruire un modello di dati *ex novo* basato sulle relazioni fra gli oggetti che trattiamo nella nostra ricerca, ma anche e soprattutto, dall'altro lato, di costruire un modello a partire da un classico database relazionale realizzato in locale, con MySQL o altri sistemi, con MS Access o altri software. L'importante è costruire un modello coerente con quello che si può evincere da tabelle e relazioni del database locale. È quello che si è fatto per il database di ateneo, a partire dal set di dati esistenti più complesso, quello riguardante il periodo 1222-1405, realizzato da chi scrive attraverso un database relazionale MySQL, popolato usando *forms* di LibreOffice Base⁵.

Il database è organizzato in quattro *projects* cronologici: 1222-1405, 1406-1605, 1606-1805, 1806-1989⁶. La scansione dipende da eventi della storia di Padova e dell'università che hanno marcato una svolta (ad esempio il passaggio sotto il dominio veneziano nel 1405), ma soprattutto dal diverso tipo di informazioni che le fonti offrono per ciascun periodo. Per il primo segmento disponiamo prevalentemente di atti notarili, che mostrano docenti, studen-

5. Il database è stato realizzato nell'ambito del progetto di ricerca *PADU-A. Prosopographical-Access-Database of University-Agenda. Verso una banca dati di studenti e docenti nei primi due secoli dell'Università*, diretto da Donato Gallo. Ulteriori dettagli in Pierluigi Terenzi, *University and urban society. A prosopographical database of the Paduan "Studium" in the Middle Ages (1222-1405)*, «Studia Universitatis Babeş-Bolyai, Historia», 64/1 (2019), pp. 39-58.

6. Il *terminus ad quem* non è ancora definitivo. L'immissione dati per il 1222-1405 è curata da chi scrive, che inoltre coordina l'intero database e gestisce le importazioni automatiche di set di dati già esistenti; il 1406-1605 è curato da Giulia Zornetta e Dennj Solera; quest'ultimo si occupa anche del 1606-1805; Andrea Martini cura il periodo 1806-1989.

ti, dottori e altri appartenenti allo *studium* in azione nella società. Ecco un esempio tradotto dal latino:

1373. 17 novembre. A Padova, nella contrada di San Canziano, nella casa in cui abita il sapiente uomo signor Bartolomeo Capodivacca, dottore in leggi, figlio del fu nobiluomo signor Francesco Paradisi [...], alla presenza del sapiente uomo signor Michele da Marostica, dottore in leggi, figlio del fu signor Martino⁷.

Da questo atto veniamo a sapere che Michele da Marostica e Bartolomeo Capodivacca erano dottori in leggi nel 1373, e che il secondo abitava in contrada San Canziano: tutti elementi che vengono registrati, in forma sia normalizzata che restituendo la fonte, nel database. Non abbiamo elenchi di studenti o docenti che possano far luce sull'intero corpo accademico, permettendo una schedatura essenziale dei dati meramente universitari, ma questi atti offrono più informazioni sulla vita di quelle persone, ed è per questo che il database originario e il corrispondente *project* di Nodegoat sono più complessi degli altri⁸. Il secondo segmento (1406-1605) è coperto soprattutto dagli *Acta graduum academicorum Gymnasii Patavini*, una serie di volumi in cui sono editi i verbali di conferimento di *licentia* e *doctoratus* agli studenti⁹. Il tipo di informazioni offerte è più ristretto e maggiormente aderente al profilo universitario, per cui il *project* corrispondente è appena più semplice del primo. Ma i verbali sono discorsivi, per così dire, e non una semplice giustapposizione di elementi essenziali a definire il conferimento del grado accademico. Ecco un altro esempio tradotto dal latino:

[1497] aprile 15. A Padova, nel solito luogo degli esami. [...] Esame privato e dottorato in diritto canonico del signor presbitero Paolo Biancuzzi da Schio, esaminato in presenza del signor Antonio Brancio dei Luschi vicentino rettore, davanti al signor Leonardo Contarini vicario [del vescovo], approvato, senza alcun contrario, sotto la promozione del signor Antonfrancesco dei Dottori professore in diritto canonico [ecc.]¹⁰.

In questo caso, oltre ai dati puramente accademici, veniamo a sapere che Paolo Biancuzzi era presbitero. Si è quindi deciso di mantenere la possibilità di registrare eventuali informazioni di altro tipo che dovessero emergere dalle

7. *Monumenti della Università di Padova (1318-1405)*, a cura di Andrea Gloria, Padova 1888, vol. II, p. 100.

8. La base documentaria è fornita da *Monumenti della Università di Padova (1222-1318)*, a cura di Andrea Gloria, Venezia 1884, e dal già citato *Monumenti della Università di Padova (1318-1405)*, 2 volumi, cui si aggiungono altre pubblicazioni e alcuni documenti d'archivio.

9. Se ne veda l'elenco alla pagina www.centrostoria.unipd.it/publicazioni-fonti.

10. *Acta graduum academicorum Gymnasii Patavini*, II/6, a cura di Elda Martellozzo Forin, Roma-Padova 2001, p. 1354.

fonti, come nel *project* 1222-1405. Del tutto simile è il ragionamento fatto per il 1606-1805, anch'esso coperto dagli *Acta graduum* ma con una consistente disponibilità di documentazione d'archivio che aggiunge molte informazioni sugli studenti. Al momento, la configurazione del *project* è in via di perfezionamento, proprio sulla base dello studio della qualità delle informazioni disponibili, già raccolte in una serie di schedature in vari formati (database, fogli di calcolo, testo semplice) in seno al Csup. L'ultimo *project*, 1806-1989, è quello più semplice dal punto di vista della struttura, perché si basa su verbali di laurea in stile moderno, che offrono informazioni regolate sugli studenti e i docenti e, man mano che si procede nel tempo, si avvicinano alla logica normalizzatrice dei database, poi utilizzati dalle segreterie universitarie per registrare i laureati. A fronte di una minore varietà di informazioni, sta ovviamente il numero enormemente più alto di persone coinvolte, dovuto alla sempre maggiore accessibilità all'università ma anche alla stessa disponibilità di fonti. Per dare una misura: il *project* 1222-1405 conta al momento poco più di 1.500 individui, il *project* 1406-1605 ne conta circa 10.200, il *project* 1806-1989 oltre 16.300 (le cifre sono in continuo aggiornamento).

Per gestire le differenze fra i vari periodi, consistenti in non pochi campi diversi fra l'uno e l'altro, non si sono dovute costruire strutture differenti, perché Nodegoat offre la possibilità di selezionare quali *objects*, campi o gruppi di campi devono essere visualizzati – e quindi valorizzati e/o modificati – per ciascun *project*. La struttura del database – o modello, per usare il linguaggio Nodegoat – è unica, ciò che cambia è la visibilità dei suoi elementi. Ciò permette, per fare un esempio banale, di far figurare il campo *Faculty/School* solo nel *project* 1806-1989, omettendolo negli altri perché in quei periodi non esistevano tali strutture. Viceversa, un campo come *Degree*, nel quale scegliere il tipo di grado accademico medievale e moderno (*licentia*, *doctoratus*, ecc.) non è utile per l'età contemporanea. Naturalmente, altri campi e gruppi di campi figurano in tutti i *projects*, come ad esempio *Position* per indicare se la persona era studente o docente a una certa data.

3. Il database: struttura e funzionamento

La complessità del database, con le sue varianti, è tale da sconsigliare in questa sede di illustrarne per intero la struttura: meglio concentrarsi sugli aspetti principali e sul funzionamento.

Nodegoat prevede in primo luogo la creazione di un *type*, cioè di un contenitore di *objects* recanti informazioni coerenti fra loro. Nel nostro caso, il *type* principale è *Person*, destinato a raccogliere tutte le informazioni sugli accademici padovani. Altri *types* sono funzionali al sistema, come quelli *source* che servono a immettere e poi selezionare le fonti storiche da cui ogni dato è tratto (Fig. 1). Per Nodegoat ogni *type* è allo stesso livello.



Fig. 1 - I types

Nel *type Person* è presente un *object*, cioè un insieme di campi – detti *descriptions* – che servono a fornire informazioni basilari sull'individuo, come nome e cognome, sesso, famiglia di appartenenza, profili biografici *online* ecc. Di questi campi, solo alcuni sono ripetibili, quelli che non concorrono a definire l'identità della persona ma che offrono informazioni aggiuntive su di essa (*Family* – nel caso in cui le fonti indichino famiglie diverse di appartenenza, specialmente per il medioevo e l'età moderna – *Bio-graphy (external)*, per indicare un URL di voci biografiche come quelle del *Dizionario biografico degli italiani*). Alcuni campi sono obbligatori, come *Gender* (Fig. 2).

Object **Sub-Object**

Name Fixed Field In Overviews

Descriptions del add

Surname String
 Multiple Required Default Unique Name Quick Search Overview

Given name String
 Multiple Required Default Unique Name Quick Search Overview

Standard Name String
 Multiple Required Default Unique Name Quick Search Overview

Father name String
 Multiple Required Default Unique Name Quick Search Overview

Mother name String
 Multiple Required Default Unique Name Quick Search Overview

Gender Reference: Classification Gender
 Multiple Required Default Unique Name Quick Search Overview

Family Reference: Classification Family Name
 Multiple Required Default Unique Name Quick Search Overview

Separator Default

Fig. 2 - Object del type Person

Per ciascun *object* si può creare un numero illimitato di *sub-objects*, secondo una relazione gerarchica che non permette di assegnare lo stesso *sub-object* a diversi *objects*. Nel nostro caso, i *sub-objects* corrispondono a gruppi di informazioni distinte riguardanti la vita e il percorso accademico di ciascun individuo, per loro natura ripetibili e dunque destinate a tabelle gerarchicamente inferiori. In altri termini, essi descrivono attributi di un *object* che possono essere ricorrenti o cambiare nel tempo e nello spazio.

Alcuni *sub-objects* compaiono in tutti o nella gran parte dei *projects*, perché sono quelli fondamentali:

1. *Birth - date & place*: per indicare luogo e data di nascita;
2. *Origin - mention*: per la provenienza geografica attestata a una certa data, quando non si conosce la data di nascita;
3. *Latin names*: per i nomi e le varianti in latino (non compare in 1806-1989);
4. *Position*: per l'attribuzione dello status di studente o professore a una certa data;
5. *Graduation*: per registrare l'ottenimento di un grado accademico (medievale, moderno, contemporaneo);
6. *Degree mention*: per le attestazioni di un grado accademico posseduto a una certa data.

Il lungo lasso cronologico del database impone una distinzione fra azioni di cui si conosce la data, da un lato, e menzioni di uno status del quale non si conosce il momento di acquisizione da parte della persona, dall'altro. Lo stesso vale per i luoghi di provenienza. Ciò riflette bene la diversa qualità delle informazioni offerte dalle fonti, che soltanto in età contemporanea, ad esempio, indicano luogo e data di nascita. Per la stessa ragione, non tutti i campi dei *sub-objects* compaiono in tutti i *projects*. In *Graduation*, ad esempio, per il 1806-1989 compare una serie di campi che sono omessi negli altri *projects*: *Dissertation title*, *Final mark* e *Cum laude* non avrebbero alcuna utilità per il medioevo e l'età moderna (Fig. 3).

Ogni *sub-object* è dotato di funzionalità per indicare il tempo e il luogo, che possono essere attivate o meno. Se attivate, ciascun set presenterà i campi *Date* e *Location*: valorizzandoli, sarà possibile fare ricerche filtrando sui periodi desiderati e costruire mappe e rappresentazioni di tutto ciò che ha coordinate spazio-temporali. *Location* è infatti un tipo campo che offre automaticamente la geolocalizzazione, sulla base di due serbatoi di luoghi normalizzati inclusi nella piattaforma Nodegoat: *City* e *Country*, entrambi basati a loro volta sugli elementi di georeferenziazione e di classificazione offerti dal database geografico GeoNames (www.geonames.org). In altre parole, l'utente potrà scegliere fra *locations* già registrate e, nel caso mancassero, potrà crearle rispettando gli standard GeoNames.

The image shows a web-based form for editing a 'Person' sub-object. The form is organized into several sections:

- Person:** Contains text input fields for Surname, Given name, Father name, Mother name, Gender, and Biographical Notes. There are also 'ver' buttons next to each field.
- Biography (external):** A text input field with a 'ver' button.
- ID:** A text input field with a 'ver' button.
- Sub-Objects: Editor:** A list of sub-objects with expand/collapse icons: [Birth - date & place], [Origin - mention], [Death], [Religion], [Position], and [Graduation].
- [Graduation] (selected):** A detailed view of the Graduation sub-object with the following fields:
 - Period:** A dropdown menu set to 'Point', followed by two date pickers (d-m-y).
 - Dissertation Title:** A text input field with a rich text editor toolbar (H1, B, I, etc.) and a 'ver' button.
 - Written / Oral:** A text input field with a 'ver' button.
 - Final Mark:** A text input field with a 'ver' button.
 - Cum Laude:** Radio buttons for 'Yes', 'No', and 'None' with a 'ver' button.
 - Tutor 1:** A text input field with a 'ver' button.
 - Tutor 2:** A text input field with a 'ver' button.
 - Buro Annotations:** A text input field with a 'ver' button.

Fig. 3 - I sub-objects del *type Person* per il project 1806-1989, con il sub-object *Graduation* in modalità modifica

La stessa possibilità è offerta per altri campi i cui valori andranno scelti da altri set di record. Nella costruzione del modello è infatti possibile scegliere, oltre a quelli di tipo *string*, *text*, *date* ecc., campi di tipo *reference*, che permettono appunto di scegliere fra valori già registrati nel database. In un database relazionale non web, questa funzione è assolta da quelle tabelle da cui si selezionano valori predefiniti per più campi di altre tabelle. Un database web come Nodegoat offre però il vantaggio di gestire in modo molto più rapido l'immissione delle informazioni nella tabella-bacino di valori, perché evita di doverlo fare prima di operare sulla tabella che ospiterà quei valori per attribuirli a una data persona. Impostando un campo *reference: type*, e selezionando su quale *type* operare, l'utente potrà inserire rapidamente un valore fra quelli disponibili, ma anche aggiungerne direttamente di nuovi, se necessario, cliccando su *new* (Fig. 4).

La relazione fra *types* così creata può essere sfruttata agevolmente attraverso la funzione *cross-referenced*. Per rimanere sullo stesso esempio, cliccando sul valore di *Faculty/School* immesso, come "Giurisprudenza", otterremo un elenco di tutte le persone alle quali abbiamo attribuito il medesimo valore: in altri termini, una lista di studenti e docenti della facoltà di Giurisprudenza. Se ci fossero relazioni fra persone (cioè fra due *types Person*), ad esempio fra uno studente e il docente tutor della sua tesi di laurea, cliccando

Fig. 4 - La selezione fra i valori del *type Faculty/School*, con la possibilità di crearne di nuovi (*new*)

sul nome del docente – selezionato in un apposito campo del *sub-object Graduation* – avremo un elenco di tutti gli studenti che si sono laureati con lui.

Il meccanismo è identico per il tipo *reference: classification*. Una *classification* è un insieme di valori, coerenti fra loro per il contenuto che trasmettono, costituiti da semplici *strings* di testo. Sono insomma elenchi di valori normalizzati fra i quali scegliere, come ad esempio la *position* dell'individuo: l'utente potrà scegliere fra “professore” e “studente”. Anche in questo caso, la lista può essere sia interamente preconstituita sia aggiornabile via via, a seconda delle necessità. Nel caso appena menzionato, tuttavia, non c'è bisogno di aggiungere altre voci alla *classification* per cui la funzione *new* è disabilitata per gli utenti che non hanno privilegi di amministratore (Fig. 5). Dopo aver immesso i dati e salvato, cliccando su “professore” nel *sub-object Position* avremo la lista di tutti i docenti del *project* in cui ci troviamo.

Le funzioni descritte sin qui sono armonizzate nella struttura del database, tenendo conto delle differenze fra i *projects*, per cui alcuni campi *reference* compaiono solo per alcuni periodi. Ma il funzionamento del database non riguarda soltanto gli aspetti tecnici, perché ad essi sono legati quelli metodologici. Si tratta di un database prosopografico storico, pertanto il suo utilizzo è sottoposto ad alcune pratiche per garantire la coerenza, la validità e la correttezza di rappresentazione dei dati raccolti. Trattandosi di informazioni storiche, è necessario interpretarle, sia in fase di data input sia in fase di analisi: non parlano da sole, insomma. In termini generali – non avendo qui lo spazio per trattare tutte le implicazioni di questo aspetto – la questione principale che riguarda l'uso di database da parte degli storici è la relazione fra dato

Fig. 5 - La selezione fra i valori della *classification Position*, senza la possibilità di crearne di nuovi

normalizzato e informazione fornita dalla fonte¹¹. Il grado di normalizzazione delle informazioni è tanto minore quanto più si va indietro nel tempo, fino ad arrivare ai tanti modi in cui una stessa persona è chiamata nel medioevo. Ciò rende necessario registrare le varianti onomastiche, da rapportare a un nome di base, univoco, che può essere scelto nella sua versione latina oppure reso in italiano, come nel nostro caso (Fig. 6).

Perché registrare tutti questi nomi? Una lista onomastica permette di verificare la bontà dell'identificazione fra due individui che non hanno esattamente lo stesso nome, inteso come insieme di elementi onomastici (nome di battesimo, patronimici, ecc.). Altri dati, incrociati a questi, permettono di confermare o escludere l'identità di due individui: il fatto che *Nicolaus Laurencius* (ultima riga di Fig. 6) corrisponda a *Nicolaus filius domini Mathei Laurencii* (prima riga) è provato dagli altri dati in nostro possesso, che sono coerenti per tutte le attestazioni. Lo storico – ma questa è un'ovvietà – deve vagliare le informazioni prima di poterle attribuire a un individuo, non può limitarsi a inserire dati acriticamente. Uno strumento come un web database rende l'operazione più agevole per via della rapidità con cui è possibile cercare i dati, incrociarli, conoscerli. In questo caso, peraltro, considerando il gran numero di individui poco noti che sono censiti, risulta di poca utilità avvalersi di identificativi comuni a livello internazionale, come quelli offerti da Vial (http://vial.org).

11. Fra gli studi disponibili, mi limito a segnalare Julien Alerini e Stéphane Lamassé, *Données et statistiques. L'avenir du travail en ligne pour l'historien*, in *Les historiens et l'informatique: un métier à réinventer*, a cura di Jean-Philippe Genet e Andrea Zorzi, Roma 2011, pp. 171-187, e suoi riferimenti bibliografici.

PIERLUIGI TEREZI

Overview **Cross-Referenced**

Niccolò di Matteo Lorenzo da Padova [edit](#)
(ngHd4G67qGNK1GzPzIRUSNL6RA4tdZbK)

StandardName Niccolò di Matteo Lorenzo da Padova
Gender M

Sub-Objects: Overview **[Origin - mention]** **[Latin Names]** **[Social Status]** **[Degree Mention]** **[Domicile]** **[Personal Ties]**

25 1 - 16 of 16

Date Start	Date End	Given Name	Particle 1	Element 1	Particle 2	Element 2	Toponymic	Source
01-07-1361	-	Nicolaus	filius	domini Mathei Laurencii				Gloria 1888 1214 52
25-06-1363	-	Nicolaus		domini Mathei Laurencii				Gloria 1888 1224 56
30-07-1369	-	Nicolaus	filius	domini Matheilaurentii				Gloria 1888 1293 79
27-05-1372	-	Nicolaus	filius quondam	domini Mathei Lorencii				Gloria 1888 1337 94
17-09-1372	-	Nicolaus	filius quondam	domini Mathei Laurencii				Gloria 1888 1338 94
23-06-1372	-	Nicolaus	filius quondam	domini Mathei Laurencii			de Padua	Gloria 1888 1338 94
25-06-1372	-	Nicolaus	quondam	domini Mathei Laurencii			de Padua	Gloria 1888 1338 94
01-01-1369	-	Nicolaus					de Padua	Gloria 1888 1283 75
23-06-1372	-	Nicolaus					de Padua	Gloria 1888 1338 94
04-08-1372	-	Nicolaus	filius quondam	domini Mathei Laurencii			de Padua	Gloria 1888 1340 95
15-03-1373	-	Nicolaus	quondam	domini Mathei Laurencii			de Padua	Gloria 1888 1347 97
13-09-1373	-	Nicolaus					de Padua	Gloria 1888 1353 99
13-01-1374	-	Nicolaus	filius quondam	domini Matheilaurentii				Gloria 1888 1358 100
05-06-1374	-	Nicolaus	filius quondam	domini Matheilaurentii			de Padua	Gloria 1888 1366 103
13-01-1375	-	Nicolaus	quondam	domini Mathei Laurencii			de Padua	Gloria 1888 1374-a 106
11-07-1376	-	Nicolaus		Laurencius				Gloria 1888 1407-b 117

Fig. 6 - I *Latin names* di Niccolò di Matteo Lorenzo da Padova

Lo storico è anche tenuto a restituire la fonte da cui le informazioni sono tratte. Come si vede sempre dalla Figura 6, ogni attestazione onomastica è riferita a una fonte precisa, e così è per ogni *sub-object*. Si tratta di una pratica metodologicamente corretta, che però in un web database di questo tipo diventa uno strumento di analisi in più. Cliccando sulla fonte, infatti, attraverso *cross-referenced* si potranno visualizzare tutte le persone e i *sub-objects* in cui quella precisa fonte è stata utilizzata, permettendo di raggruppare individui e dati per nuclei informativi.

Gli automatismi offerti dalla piattaforma Nodegoat, opportunamente configurati, permettono inoltre di avere rappresentazioni immediate di alcune caratteristiche dei dati raccolti. Particolarmente utile ai fini della ricerca storica è la rappresentazione su mappa delle provenienze degli individui censiti. Attraverso l'indicazione di una località georeferenziata nel campo *Location* (nei *sub-objects Birth - date & place* e *Origin - mention*) la mappa è già pronta per essere visualizzata, ma si possono impostare dei filtri per limitare la visualizzazione, per esempio soltanto a certi periodi o a certe categorie di persone (ad esempio studenti o professori). Un semplice colpo d'occhio, rispetto a quello che si vuol sapere, è sufficiente per avere un'idea efficace e immediata dei fenomeni riguardanti la mobilità delle persone accademiche.

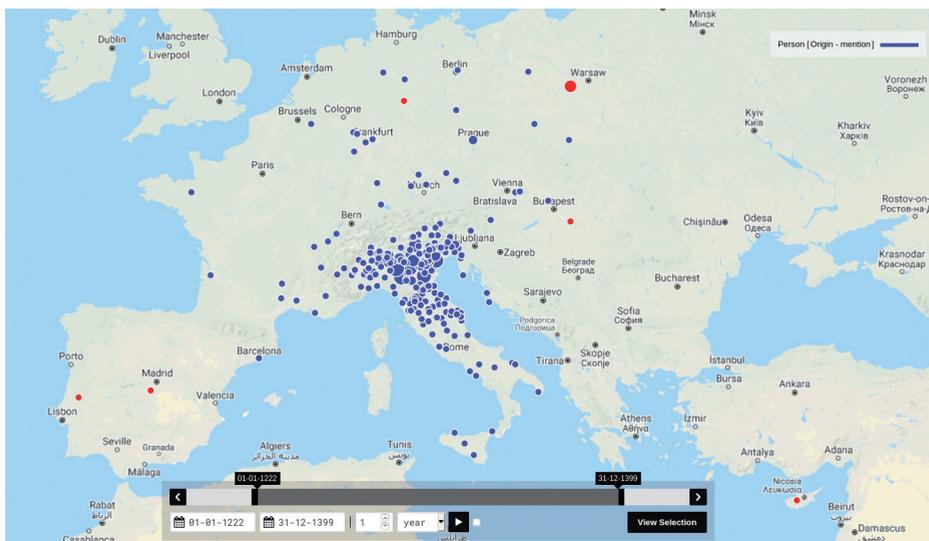


Fig. 7 - Le provenienze degli individui censiti nel *project* 1222-1405. In blu le località, in rosso le aree indicate nelle fonti (es. *Hispanus*). La dimensione dei punti si basa sul numero di attestazioni

È tuttavia essenziale approcciarsi a queste rappresentazioni con consapevolezza: per il 1222-1405, ad esempio, su 1.527 individui censiti coloro di cui si conosce la provenienza sono 1.228 (80,4%). È una percentuale alta, ma comunque non coincidente con l'intera popolazione censita, per cui sarà bene parlare di tendenze più che di dati consolidati: in questo caso, la prevalenza di individui provenienti dall'Italia centro-settentrionale è evidente, ma non è affatto trascurabile l'apporto dei transalpini. Tutto ciò potrà essere oggetto di analisi più approfondite – per esempio incrociando le provenienze con la posizione (studente, professore) e la disciplina – per comprendere quali flussi si verificavano verso Padova e perché.

4. Da un network di progetti a un cross-search database

Le questioni cui ho accennato sono solo parte di quelle che si possono affrontare nella storia dell'università e in ricerche prosopografiche in genere. Proprio la dimensione comune delle questioni ha spinto un gruppo di studiosi a mettere insieme le forze per tentare un'interessante sperimentazione. Alla base, vi è la costituzione di una rete europea di progetti di storia dell'università: *Héloïse - European Network on Digital Academic*

*History*¹². Nato nel 2012, il *network* promuove incontri scientifici sulla storia delle università europee e intende predisporre strumenti tecnici e soluzioni per una collaborazione fattiva nella raccolta, gestione e pubblicazione digitale di dati. A questo scopo, in seno a *Héloïse*, alcuni progetti hanno avviato la definizione (metodologica e tecnica) di un *cross-search database*, cioè di uno strumento che permetta di interrogare contemporaneamente più di un database online, senza dover intervenire sulla struttura di questi ultimi per renderla omogenea rispetto ad altri o a un modello unico.

La realizzazione di una struttura digitale “sovrapposta” permette, senza costringere a interventi tecnici che potrebbero anche interessare l’approccio metodologico adottato dai singoli gruppi di ricerca, di tenere insieme progetti nati e sviluppati in tempi e con obiettivi in parte diversi (anche per la diversa natura delle fonti disponibili)¹³. Facciamo qualche esempio. Il progetto *Studium Parisiense* (<http://studium.univ-paris1.fr/home>), sviluppato in seno al Laboratoire de Médiévistique Occidentale de Paris (LaMOP) dell’Université Paris 1 Panthéon-Sorbonne, riguarda i membri delle *scholae* e dello *studium* di Parigi fra il XII e il XVI secolo, e punta a realizzare schede bio-bibliografiche complete di ciascun individuo, di cui non si registrano solo i dati biografici minimi e di carriera universitaria, ma tutto ciò che possiamo saperne: status sociale, opere prodotte, ecc. Ciò richiede quindi un lavoro che non può limitarsi alle sole fonti prodotte in ambito universitario (come ad esempio le liste degli studenti iscritti), ma implica una ricerca ad amplissimo raggio su ciascuno degli individui.

Anche il progetto *Bo2022* registra alcuni aspetti extra accademici (per l’età medievale e moderna), ma lo fa in modo più limitato perché – visto lo spettro cronologico amplissimo, unico nel suo genere – sarebbe controproducente allargare la ricerca ad altre fonti quando vanno ancora “smaltite” quelle prodotte in ambito universitario. Un approccio simile, cioè concentrato su un ateneo e legato strettamente alle fonti connesse allo *studium*, è adottato dal progetto *Onomasticon* dell’Università di Perugia (<https://onomasticon.unipg.it>), che censisce studenti e docenti di quell’ateneo dalla fondazione nel 1308 in avanti, e che è stato fondamentale per impostare il database *Bo2022*¹⁴. Un’impostazione diversa riguarda il progetto RAG – *Repertorium Academicum Germanicum* delle università di Berna e Giessen ([12. Il *network* dispone di un sito web \(<http://heloise.ish-lyon.cnrs.fr>\) e di un blog \(<https://heloise.hypotheses.org>\), dove è possibile conoscere i molti progetti partecipanti, che non si possono elencare qui.](https://rag-online.</p>
</div>
<div data-bbox=)

13. Si può fare riferimento a *Per una storia digitale delle università. Alcune esperienze del network Héloïse*, a cura di Willem Frijhoff, sezione monografica di «Annali di storia delle università italiane», 21/1 (2017).

14. Colgo l’occasione per ringraziare il team di *Onomasticon* e in particolare Stefania Zucchini e Giulio Quaresima.

org)¹⁵, che censisce i titolari di gradi accademici del Sacro Romano Impero fra il 1250 e il 1550, dovunque li abbiano conseguiti, offrendo informazioni anche su ciò che fecero prima e dopo aver ottenuto la licenza o il dottorato.

Ciascun progetto, com'è facile immaginare, presenta set di dati strutturati diversamente, in base ai propri scopi e alle informazioni offerte dalle fonti. Ma l'insieme di dati comuni è tale da poter permettere una condivisione: si pensi agli elementi onomastici, come pure alle discipline e ai titoli accademici. A partire da questa constatazione, proprio il team del Rag si è reso protagonista dell'avvio della sperimentazione di un *cross-search database*, insieme ai progetti di Parigi, Perugia, Bologna, Poitiers e altri, nell'ambito del Caad – *Collaborative approach on medieval and early modern academic databases*, sviluppato all'interno di *Héloïse*. Il lavoro è condotto ora nell'ambito del progetto Snsf Spark *Dynamic Data Ingestion (Ddi): Server-side data harmonization in historical research*, diretto da Kaspar Gubler del Rag (<https://histdata.hypotheses.org/spark>).

La sfida è quella di armonizzare i differenti progetti, senza modificarli, attraverso una struttura digitale sovrapposta che interroghi contemporaneamente i database coinvolti, attraverso corrispondenze fra i campi e un'ontologia generale definita a monte, che permetta di classificare sotto voci univoche la varietà di informazioni registrata nei diversi progetti¹⁶. Con intenzioni simili, nell'ambito di una partnership con il Caad, Christian Hesse e Kaspar Gubler del Rag stanno sviluppando il progetto Repac – *Repertorium Academicum* (<http://repac.ch>), che consiste in una collezione di dati provenienti dal Rag, dal *Repertorium Academicum Helveticum* (sugli studenti svizzeri, 1300-1600) e dal *Repertorium Bernense* (su persone e conoscenza a Berna, 1350-1600). Il Repac offre inoltre la possibilità di includere i dati di altri progetti in un *Virtual research environment* (Vre) costituito proprio da Nodegoat.

Come si vede, la comunità di storici dell'università sta investendo molto nelle potenzialità digitali e in particolare dei database, per restituire in termini moderni e in formati analizzabili quella rete di studenti e professori che si creò in Europa con la nascita e lo sviluppo degli *studia*, che furono un motore di altissima mobilità di persone e idee, e dunque di avanzamento delle conoscenze. Padova, nello spirito internazionale che la contraddistingue da sempre, si è inserita in questo flusso con il database *Bo2022* che, anche grazie alla scelta di adottare Nodegoat come piattaforma, entrerà presto a far parte dei progetti che stanno sperimentando il *cross-research database*.

15. Dal 1° gennaio 2020 è la sola Università di Berna a gestirlo.

16. Ulteriori informazioni alla pagina <http://heloise.ish-lyon.cnrs.fr/forum>.

Abstract

This essay presents the project 'Bo2022', an online database of students and professors of the University of Padua, from its foundation in 1222 to the 20th century. After a brief history of this institution, the contribution focusses on the features of the web-platform used for data structuring and input (*nodegoat*), as well as on the treatment of information provided by historical sources, which is needed to both adapt it to the database structure and keep its information capacity. The essay also illustrates the project of a cross-search European database, currently under development.

*I laureati in Scienze politiche
dell'Università di Padova (1926-2011).
Lettura e interpretazione del Database*

di *Filiberto Agostini**

1. Introduzione

Nell'ambito delle ricerche in cantiere per “celebrare” il traguardo degli 800 anni dell'Università di Padova (1222-2022), occupa un posto di rilievo il progetto *Patavina libertas*, che si propone di pubblicare una serie di volumi relativi alla “storia europea” dell'Ateneo. La pagina web di presentazione della collana precisa che «tra i progetti è prevista inoltre l'implementazione di una banca dati [...], un lavoro che includerà il censimento di chi ha studiato e insegnato nell'Ateneo»¹: questo patrimonio documentario – in continua crescita – dovrebbe fungere come bacino di informazioni utilizzabile da diversi gruppi di ricerca². Oggi i dati – rinvenibili solo attraverso un accesso riservato al sito “patavini.800anni.unipd.it” – vengono gestiti tramite *Nodegoat*, un

* Dipartimento di Scienze Politiche, Giuridiche e Studi Internazionali, Università di Padova.

1. Si veda il seguente indirizzo internet: www.ottocentenariouniversitadipadova.it/progetti/racontarela-nostra-storia/patavina-libertas-una-storia-europea-delluniversita-di-padova [accesso effettuato il 31 ottobre 2020].

2. L'ingresso dei Big Data nel mondo contemporaneo ha alimentato una significativa riflessione sulla implicazione che una sterminata massa di informazioni può avere nell'influenzare il comportamento individuale e collettivo in termini di investimenti, consumi e scelte politiche. L'implementazione dei dati in un contenitore elettronico coinvolge anche le scienze storiche, dando luogo a dibattiti sui metodi quantitativi/qualitativi, al rapporto tra storia e dato, riprendendo le linee maestre di discorsi già avviati decenni addietro. Il database relazionale viene a trovarsi al centro di una rete di attività e gruppi di ricerca, offrendo strumenti per indagini di diverso tipo e su diversi periodi. Ciò è possibile grazie alla flessibilità dello «strumento adottato e all'attenta configurazione che se n'è fatta, volta anche a commisurare le esigenze di normalizzazione tipiche di un database con il rispetto delle sfumature che caratterizzano le fonti storiche». Cfr. a titolo esemplificativo: *Quantità/qualità: la storia tra sguardi micro e generalizzazioni*, a cura di Daniele Andreozzi, Palermo 2017; Michele Nani, *La metà perduta. Appunti su studi storici e metodi quantitativi a partire da pubblicazioni recenti*, «Italia contemporanea», 293/2020, pp. 177-189.

software specifico che consente di trasformare il semplice elenco prosopografico in una ricca piattaforma, nella quale le singole “entrate” possono essere intersecate, classificate e rielaborate, anche in chiave europea³.

All’interno di questo vasto orizzonte programmatico, lo scopo del presente saggio è quello di fornire alcuni spunti di riflessione sui dati concernenti i laureati a Padova in Scienze politiche, che risulta essere «una delle più antiche e prestigiose Facoltà» presenti in Italia⁴. Fondata nel 1924 come Scuola di Scienze Politiche e Sociali, dal 1933 si trasformò in Facoltà e tale rimase fino alla cosiddetta riforma Gelmini, quando, in un esteso sforzo riformatore dell’intero sistema accademico italiano, al posto delle Facoltà furono attivati i Dipartimenti, tuttora vigenti.

Tuttavia, proprio in virtù del fatto che l’Ateneo patavino si approssima a festeggiare il suo ottocentesimo anno di apertura, pare opportuno offrire un brevissimo inquadramento storico per contestualizzare in un’ottica più ampia il periodo oggetto di analisi (1926-2011)⁵. L’Università di Padova era sorta oltre sette secoli prima, nel 1222, come «un’associazione corporativa» nella quale lo studio delle arti liberali rivestiva un ruolo importante⁶, rappresentando – nell’ambito cittadino – in tempi antichi come in epoca recente, un «centro vitale del sistema di potere e chiave dell’egemonia culturale»⁷. Nel Medioevo e nel Rinascimento essa si impose a livello europeo per la fama negli studi filosofici di contenuto aristotelico⁸; successivamente fu il turno della scienza – con esponenti quali il celebre Galileo, protetto e incoraggiato dalla

3. Un ampio e collaudato progetto europeo di storia delle università è rappresentato da Héloïse, network europeo *on digital academic history*. La documentazione relativa a Scienze politiche è conservata nell’Archivio dell’Università di Padova, Segreteria studenti, Facoltà di Scienze Politiche, Verbali di laurea, volumi 1-188. L’implementazione dei dati è stata effettuata da Rossella Bortolotto, che qui ringrazio.

4. Antonio Varsori, *Prefazione*, in Giulia Simone, *Fascismo in cattedra. La Facoltà di Scienze politiche di Padova dalle origini alla Liberazione (1924-1945)*, Padova 2015, p. 9.

5. Il termine *ad quem* della presente ricerca è rappresentato dalla trasformazione delle Facoltà in Dipartimenti nel 2011 (c.d. legge Gelmini), come vedremo nei paragrafi successivi.

6. La citazione è tratta da Giovanni Reale e Dario Antiseri, *Storia della filosofia dalle origini a oggi*, III. *Patristica e Scolastica*, Milano 2008, p. 282. Sullo Studio patavino si vedano: *L’Università di Padova nei secoli*, a cura di Piero Del Negro e Francesco Piován, I-III, Treviso 2002-2017; *L’Università di Padova. Otto secoli di storia*, a cura di Piero Del Negro, Padova 2001.

7. Angelo Ventura, *Padova*, Roma-Bari 1989, p. 68 (a p. 11 definisce Padova «capitale intellettuale» del Veneto). Si veda *L’Università di Padova nei secoli (1806-2000)*, a cura di Piero Del Negro e Francesco Piován, III, Treviso 2017.

8. Per una semplice introduzione cfr., tra i molti studi, Antonino Poppi, *Profilo storico-istituzionale della teologia nello studio di Padova (1363-1806)*, «Quaderni per la storia dell’Università di Padova», XXXV, 2002, pp. 3-46.

Repubblica di Venezia⁹ – a tenere alto il nome di Padova. Caduta la Serenissima, dopo la parentesi napoleonica Padova venne inglobata nei domini di Vienna; l'Ateneo, parificato agli altri dell'Impero asburgico, visse appieno la fiammata rivoluzionaria del 1848, cui seguì un periodo di forzata chiusura¹⁰. Nel 1866, quando il Veneto entrò a far parte del Regno di Vittorio Emanuele II, la transizione fu «traumatica e dolorosa» e comportò – tra i numerosi cambiamenti – la soppressione della storica Facoltà di teologia¹¹.

Per quanto qui ci interessa, va rilevato che il picco minimo di iscrizioni complessive si registrò nel 1860-1861, con la presenza di soli 673 studenti¹². L'Ateneo riprese ben presto ad espandersi in termini di iscritti, mentre – sul piano culturale – veniva sempre più configurandosi come una roccaforte del pensiero laico, «di una cultura giuridica e politica di chiaro profilo laicista», protrattasi fino alla Conciliazione¹³. Il tetto dei duemila studenti fu superato soltanto dopo la fine della prima guerra mondiale¹⁴; nel frattempo, già nella seconda metà dell'Ottocento, all'interno della Facoltà di Giurisprudenza era stata prevista la possibilità di istituire corsi specifici per lo studio delle cosid-

9. Enrico Berti, *Storia della filosofia dal Quattrocento al Settecento*, Roma-Bari 1991, p. 55. Fondamentale è il volume *Clariores. Dizionario biografico dei docenti e degli studenti dell'Università di Padova*, a cura di Piero Del Negro, Padova 2015.

10. La chiusura si protrasse fino al novembre 1850. All'inizio della Restaurazione, nel 1815, si contavano quasi 2.000 iscritti a Padova: cfr. Giulia Simone, *Gli studenti dell'Università di Padova dal 1850 al 1870*, in *L'Ateneo di Padova nell'Ottocento. Dall'Impero asburgico al Regno d'Italia*, a cura di Filiberto Agostini, Milano 2019, p. 277. Sul periodo sono utili anche: Giampietro Berti, *L'Università di Padova dal 1814 al 1850*, Treviso 2011; Maria C. Ghetti, *Da Venezia a Vienna. I poteri politici e l'Università*, in *Istituzioni culturali, scienza, insegnamento nel Veneto dall'età delle riforme alla Restaurazione (1761-1818). Atti del convegno di studi, Padova 28-29 maggio 1998*, a cura di Luciana Sitran Rea, Trieste 2000, pp. 1-14; Angelo Ventura, *L'8 febbraio 1848 nella storia dell'Università di Padova*, in *Studenti, Università, città nella storia padovana. Atti del convegno Padova 6-8 febbraio 1998*, a cura di Francesco Piovan e Luciana Sitran Rea, Trieste 2001, pp. 707-720.

11. Cfr. Filiberto Agostini, *La transizione dall'Impero asburgico al Regno d'Italia. I docenti dello Studio di Padova nel 1866*, in Id., *L'Ateneo di Padova nell'Ottocento*, p. 241; Liliana Billanovich scrive che «l'intero tornante storico che vide compiersi il processo di unificazione nazionale» rappresentò un momento di «passaggio epocale»: Ead., *La cesura di un rapporto plurisecolare. Università e Chiesa di Padova a seguito del 1866*, in *L'Ateneo di Padova nell'Ottocento*, p. 179. Per un quadro d'insieme si veda anche *Il Veneto nel Risorgimento. Dall'Impero asburgico al Regno d'Italia*, a cura di Filiberto Agostini, Milano 2018.

12. Simone, *Gli studenti dell'Università di Padova dal 1850 al 1870*, p. 283.

13. Cfr. Billanovich, *La cesura di un rapporto plurisecolare*, pp. 193 e 198. L'autrice sottolinea che le manifestazioni e le iniziative organizzate in onore di Galileo divennero l'occasione per «esaltare i diritti dell'Università e della scienza laica contro le gesuitiche usurpazioni».

14. Piero Del Negro, *Otto secoli di storia dell'Università di Padova: alcuni spunti di riflessione a proposito di élite e di "massa" studentesca*, in *Dall'università d'élite all'università di massa. L'Ateneo di Padova dal secondo dopoguerra alla contestazione sessantottesca*, a cura di Alba Lazzaretto e Giulia Simone, Padova 2017, p. 28.

dette scienze sociali. A partire da quegli anni, infatti, lo studio della sociologia e delle scienze politiche acquisì una propria indipendenza, anche grazie ai notevoli approfondimenti epistemologici dovuti all'imporsi della filosofia positivista in Europa¹⁵.

2. La fondazione e lo sviluppo della Scuola/Facoltà di Scienze politiche

In Italia sorsero Scuole per le scienze politiche e sociali a Firenze, Pavia, Roma e, dal 1924, a Padova¹⁶. Qui nel primo anno accademico gli iscritti furono poche decine, su un totale complessivo di 2.490 matricole; si trattava, in massima parte, di studenti veneti, già laureati in Giurisprudenza, che, con pochissimi esami, potevano ottenere il diploma in una delle due sezioni della Scuola¹⁷. Nell'anno 1926-1927 si laurearono i primi 8 studenti; nel corso del 1929-1930 gli iscritti salirono a 100 e conseguirono il titolo 20 immatricolati. Numeri così bassi sono spiegati da Piero Del Negro, secondo il quale, a deprimere le iscrizioni in tutte le Facoltà era stata essenzialmente «la stretta classista e maschilista operata agli inizi degli anni Venti dalla riforma del ministro Gentile»¹⁸.

Successivamente si entrò nel periodo di consolidamento del regime fascista e la presenza del partito si fece sempre più pervasiva, sia all'interno dell'Università sia – soprattutto – della Facoltà di Scienze politiche: di sicura fede governativa erano diversi docenti, tra i quali Alfredo Rocco, Donato Donati, Corrado Gini, Adolfo Ravà¹⁹; della stessa impronta erano i programmi di studio, le materie d'esame – si pensi a Diritto coloniale o a Storia e dottrina del fascismo – e gli argomenti delle tesi di laurea²⁰. È noto che, a livello nazionale, tutti i docenti universitari, nel 1931, dovettero prestare giuramento

15. *Il positivismo a Padova tra egemonia e contaminazioni, 1880-1940*, a cura di Giampietro Berti e Giulia Simone, Treviso 2016. Su Roberto Ardigò, docente di filosofia positivista a Padova, si vedano gli atti del convegno tenutosi il 21 ottobre 1999, «Quaderni per la storia dell'Università di Padova», XXXIV, 2001, pp. 3-230.

16. Simone ricorda che «quella di Padova è, dunque, in ordine cronologico, la terza Scuola a nascere in Italia»: Simone, *Fascismo in cattedra*, p. 26.

17. Simone, *Fascismo in cattedra*, pp. 31-32, 34, 59-60.

18. Del Negro, *Otto secoli di storia dell'Università di Padova*, p. 30.

19. Sul rettore dell'Università Carlo Anti: cfr. Simone, *Fascismo in cattedra*, p. 82; Angelo Ventura, *Carlo Anti rettore magnifico e la sua università*, in *Carlo Anti. Giornate di studio nel centenario della nascita*, Verona, Padova, Venezia, 6-8 marzo 1990, Trieste 1992, pp. 155-222.

20. Norberto Bobbio, *La cultura e il fascismo*, in *Fascismo e società italiana*, a cura di Guido Quazza Torino 1973, p. 241; Simone conclude che la Facoltà divenne «cassa di risonanza dell'operato del regime e dell'impero»: Giulia Simone, «La Facoltà Cenerentola». *Scienze politiche a Padova dal 1948 al 1968*, Milano 2017, p. 19.

di fedeltà al regime; due anni più tardi furono obbligati a possedere la tessera del partito per poter continuare a insegnare²¹. L'8 maggio 1932 il neo vescovo di Padova, mons. Carlo Agostini, si recò in visita ufficiale dal rettore, contrassegnando così l'avvenuta pacificazione tra potere politico e potere religioso, già sancita a livello nazionale dai Patti lateranensi del 1929. Commentando l'avvenimento, Mario Isnenghi non esita a rimarcare l'importanza e a definirlo «una data periodizzante, che fece epoca»²². È in questo clima che, con il regio decreto n. 1592 dell'agosto 1933, la Scuola assunse la denominazione di Facoltà di Scienze politiche. Qualche anno dopo, quando l'ideologia razziale tedesca venne fatta propria dalla classe dirigente italiana²³, la Facoltà di Padova si distinse per capacità di conformarsi al nuovo clima politico-filosofico²⁴.

Quando, nell'autunno del 1938, all'Ateneo fu chiesto di applicare le leggi razziali e dunque espellere il personale «di razza ebraica», Scienze Politiche subì una considerevole epurazione: venne allontanato innanzitutto Donato Donati, insigne giurista nonché preside della Facoltà; cui seguirono l'economista Marco Fanno e il filosofo del diritto Adolfo Ravà. I tre docenti così espulsi, si allontaneranno da Padova durante la guerra. Solamente Marco Fanno farà ritorno, una volta terminato il conflitto, per riprendere l'insegnamento nella sua Facoltà.

A livello di Ateneo stava intanto lievitando il numero degli iscritti, dato da imputare sia alla marcata crescita degli studenti nelle scuole secondarie, sia a un fenomeno più generale, che vede l'ascesa sociale della piccola borghesia italiana: si passò dai 4.000 studenti del 1934-1935 ai 6.000 del 1940-1941. A Scienze politiche i numeri erano molto più contenuti: alla vigilia della seconda guerra mondiale la Facoltà contava 135 iscritti, peraltro superata in Italia solo da Roma e da Firenze²⁵. Nell'anno accademico 1944-1945 gli stessi calarono a 88²⁶, cifra davvero esigua, soprattutto se si considera che l'Ateneo

21. Sul tema si veda, a titolo di esempio: Helmut Goetz, *Il giuramento rifiutato. I docenti universitari e il regime fascista*, Firenze 2000.

22. Mario Isnenghi, *I luoghi della cultura*, in *Storia d'Italia. Le regioni dall'Unità a oggi. Il Veneto*, a cura di Silvio Lanaro, Torino 1984, pp. 290-291.

23. Tra i molti contributi, si veda Michele Sarfatti, *Mussolini contro gli ebrei. Cronaca dell'elaborazione delle leggi del 1938*, Torino 1994.

24. Simone scrive che «nella campagna razziale la Facoltà di Scienze politiche ha un ruolo di punta» e che, nel 1939, essa «appare "fascistissima" e del tutto conforme al volere del regime dal punto di vista della politica razziale» (Simone, *Fascismo in cattedra*, pp. 105 e 108).

25. Archivio Generale di Ateneo (Padova), *Atti del Rettorato*, b. 369, fasc. 55 «Facoltà di Scienze politiche 1939-1940», lettera non firmata inviata ad Anton Maria Bettanini, 6 aprile 1940.

26. Simone, *La Facoltà di Scienze politiche (1945-1968)*, in Lazzaretto e Simone, *Dall'università d'élite all'università di massa*, p. 88.

– nel suo complesso – superò quota 9.000 matricole negli anni 1942-1943, 1943-1944 e 1944-1945²⁷. L'aumento generale di studenti durante il conflitto si può spiegare in due modi: da un lato, per quanto riguarda i maschi, l'iscrizione all'università comportava la possibilità di chiedere il rinvio del servizio militare; dall'altro, la percentuale di donne iscritte, salita considerevolmente, contribuiva all'incremento complessivo degli immatricolati²⁸.

La situazione dell'Ateneo patavino si trasformò radicalmente dopo la caduta del fascismo e l'8 settembre 1943, quando il rettorato passò da Carlo Anti a Concetto Marchesi, che inaugurò il nuovo anno accademico, ripromettendosi di salvaguardare l'indipendenza dell'Università da qualsiasi ingerenza politica²⁹. Furono mesi molto difficili, tanto che – a riprova del coraggio dimostrato da molti docenti, studenti e personale amministrativo, l'Ateneo padovano figura – unico in Italia e tra i pochissimi in Europa – insignito della medaglia d'oro al valor militare³⁰. Nel frattempo però, nelle zone sotto controllo alleato, una circolare del ministro della Pubblica istruzione Guido De Ruggiero aveva abolito le Facoltà di Scienze politiche, in non poche occasioni dimostrate pronte «a servire l'ideologia fascista»³¹. Alla fine del conflitto il provvedimento soppresivo venne applicato anche a Padova e gli studenti iscritti, così come i docenti, furono costretti a trasferirsi a Giurisprudenza³².

Solo nell'estate del 1948, in un clima politico completamente diverso, il Governo italiano permise la riapertura dei corsi di Scienze politiche, in un Ateneo che, negli anni successivi, si allargò notevolmente³³: la Facoltà di Agraria fu istituita nel 1951; Magistero nel 1952; Economia e Commercio – con sede distaccata a Verona – nel 1963; Scienze statistiche, demografiche e

27. Del Negro, *Otto secoli di storia dell'Università di Padova*, pp. 30-31.

28. Alba Lazzaretto, *L'Ateneo patavino tra il secondo dopoguerra e il Sessantotto: istituzioni, docenti, studenti*, in Lazzaretto e Simone, *Dall'università d'élite all'università di massa*, p. 51 (nel 1945-1946 si registrarono 1.980 giovani donne iscritte: cfr. nota 63).

29. Ezio Franceschini, *Concetto Marchesi. Linee per l'interpretazione di un uomo inquieto*, Padova 1978, pp. 172-175; Giuseppe Zaccaria (a cura di), *Concetto Marchesi e l'Università di Padova 1943-2003*, Padova 2004.

30. Angelo Ventura, *L'Università di Padova nella Resistenza*, «Quaderni per la storia dell'Università di Padova», XXVIII, 1995, pp. 157-172; Id., *Padova nella Resistenza*, in Giuliano Lenci e Giorgio Segato (a cura di), *L'amministrazione comunale di Padova nel 1943. Dalla crisi del regime fascista alla Resistenza*, Padova 1996, pp. 17-34; Chiara Saonara, *Il rettorato degli anni difficili*, in Giuseppe Gola, *Il mio rettorato (1943-1945)*, Treviso 2015, pp. 7-27; Chiara Saonara, *Studenti tra Fascismo e Resistenza*, in *Gli studenti nella storia dell'Università di Padova. Cinque conferenze*, a cura di Francesco Piovan, Padova 2002, pp. 65-78.

31. Simone, *Fascismo in cattedra*, p. 144. Si tratta della circolare n. 1120 del 27 novembre 1944.

32. Si veda la circolare n. 863 del 15 febbraio 1945, a firma del ministro Vincenzo Arancio-Ruiz.

33. Rimase però una «continuità accademica nella discontinuità politica»: Simone, «*La Facoltà Cenerentola*», p. 20.

attuariali nel 1968³⁴. Nel 1947-1948 Scienze politiche contava una presenza ridottissima – solo 46 iscritti – dei quali 41 fuori corso; l'avvicinamento al mondo cattolico, che dominava politicamente con l'affermazione elettorale della Democrazia cristiana di De Gasperi, fu rappresentato anche dalla scelta del nuovo preside, il sacerdote – e docente di storia – Anton Maria Bettanini, formatosi all'Università Cattolica con padre Gemelli³⁵. In questo contesto, mentre le iscrizioni complessive nell'Ateneo rimanevano stabili negli anni Cinquanta, Scienze politiche continuò a rappresentare un'entità modestissima nel panorama patavino, sia in termini di iscritti sia di prestigio: i docenti, in particolare, percepivano l'insegnamento presso la Facoltà come «una tappa di passaggio per una carriera che sarebbe stata completata in sedi più appetibili e prestigiose»³⁶. Nel 1959, quando si contavano 285 studenti, si attuò il passaggio di presidenza dal citato Bettanini a Ettore Anchieri, storico di tempra nello studio delle relazioni internazionali, che avrebbe guidato la Facoltà fino al 1968.

3. Dagli anni Sessanta agli anni Novanta

Negli anni Sessanta si verificò in Italia una vera «rivoluzione demografica», che determinò una impennata di iscritti all'università, trasformandola in «un organismo di massa»³⁷. Ma, oltre all'aumento della platea di immatricolati, ben altro attendeva il sistema accademico italiano, ancora ingabbiato in un apparato legislativo progettato nel ventennio fascista. Da Berkeley, prestigiosa università americana gemellata con Padova, giungeva eco in Europa delle manifestazioni studentesche. Nel clima caldo del Sessantotto, la ventata di proteste che divampò in tutta Italia si alimentò non solo di legittime richieste di democratizzazione del sistema universitario, ma anche di «aspirazioni palingenetiche», che miravano a cambiare l'intera società³⁸. Citando Angelo Ventura, che definì Padova «città di vipere», Alba Lazzaretto ricorda che nell'Ateneo si verificarono numerose occupazioni. Venne fatta esplodere una

34. Lazzaretto, *L'Ateneo patavino (tra il secondo dopoguerra e il Sessantotto)*, p. 43.

35. Si veda Simone, «*La Facoltà Cenerentola*», pp. 47-60. Sulla città, amministrata dal sindaco Cesarino Crescente: Ventura, *Padova*, p. 380; Monica Fioravanzo, *Élites e generazioni politiche. Democristiani socialisti e comunisti veneti (1945-62)*, Milano 2003, pp. 243-244; Paolo Giaretta e Francesco Jori, *La Padova del sindaco Crescente (1947-1970)*, Padova 2017.

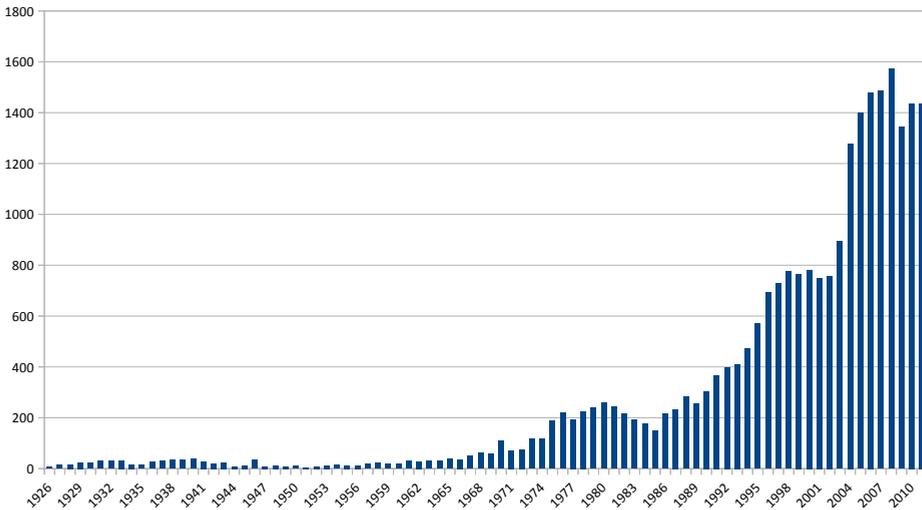
36. Simone, «*La Facoltà Cenerentola*», p. 73.

37. Lazzaretto, *L'Ateneo patavino (tra il secondo dopoguerra e il Sessantotto)*, pp. 52 e 54.

38. Sia Lazzaretto che Simone usano il termine «palingenesi» per indicare la radicalità delle rivendicazioni studentesche e operaie del periodo: cfr. Simone, «*La Facoltà Cenerentola*», p. 158; Lazzaretto, *L'Ateneo patavino (tra il secondo dopoguerra e il Sessantotto)*, p. 58.

bomba nello studio del rettore Enrico Opocher, mentre la città nel suo complesso si apprestò a diventare «un laboratorio politico particolarmente inquietante e violento»³⁹. A Scienze politiche, così come a Magistero, gli studenti ottennero il permesso di partecipare alle sedute del Consiglio di Facoltà; nel primo incontro allargato, si discusse anche della necessità di nuovi spazi per l'aumento della popolazione studentesca⁴⁰, al di là dell'uso dei Palazzi Dottori e Wollemborg da poco acquistati. Per quanto concerne il numero degli iscritti alla Facoltà, questo crebbe costantemente fino al 1967, quando si registrarono 612 studenti; l'anno successivo si avvertì un modesto calo, ma fu a partire dall'entrata in vigore della legge Codignola del 1969⁴¹ – che apriva l'accesso a tutte le Facoltà con qualsiasi tipo di diploma superiore – che si passò a una “università democratica”. Nel 1971-1972, infatti, si immatricolarono quasi 1.000 studenti in più a Scienze politiche, portando il numero degli iscritti a 1.618.

Numero di laureati e diplomati per anno (1926-2011)



39. Lazzaretto, *L'Ateneo patavino tra il secondo dopoguerra e il Sessantotto*, p. 58. La citazione di Ventura è tratta dal volume *Padova*, p. 385. Sul tema si veda anche il volume *Le istituzioni universitarie e il Sessantotto*, a cura di Alessandro Breccia, Bologna 2013, specialmente i saggi di Alba Lazzaretto (pp. 185-198) e di Paola Caldognetto (pp. 199-212).

40. Tra il 1957 e il 1967 gli studenti della Facoltà erano aumentati del 150%, passando da 244 a 612: Simone, «*La Facoltà Cenerentola*», p. 148.

41. Legge n. 190 dell'11 dicembre 1969.

Di fronte al grafico dei laureati in Scienze politiche nel corso degli anni, si percepisce chiaramente come la Facoltà sia transitata dall'essere una «università d'élite» a una «di massa», per parafrasare il titolo del volume di Lazzaretto e Simone sull'Ateneo patavino⁴². Dal 1926 al 1930 si addottorarono a Padova 87 persone; nel decennio compreso tra il 1931 e il 1940 sono contati 287 laureati, con una punta massima di 37 nell'anno 1940. Successivamente, nei cinque anni di guerra, il numero cala drasticamente, tanto che nel 1944 solo 9 persone conclusero il ciclo di studio. Nel dopoguerra i numeri tardano a risalire, complici sia la chiusura della Facoltà a opera degli Alleati che le difficili condizioni socio-economiche dell'Italia. Non stupisce insomma che nel decennio 1941-1950 solo 169 persone abbiano conseguito un titolo universitario a Scienze politiche, facendo registrare una contrazione di oltre il 40% rispetto all'analogo periodo precedente. Anche negli anni Cinquanta si sono rilevate cifre molto contenute in termini di laureati⁴³. Come detto, è a partire dagli anni Sessanta che i laureati hanno cominciato a superare la media dell'anteguerra: alla fine del decennio – con le riforme della scuola e dell'università in atto – i laureati a Scienze politiche balzarono da 59 – dato del 1969 – a 108 nel 1970, censendo un aumento dell'83%. Se si fa riferimento all'intero decennio, l'incremento supera il 270%.

Negli anni Settanta Padova e il Veneto vissero, come altre città italiane, la dolorosa esperienza del terrorismo, sia di matrice fascista che marxista-leninista⁴⁴. Nella seconda metà del decennio, quando sembrava imminente il superamento della maggioranza politica a guida democristiana da parte del blocco social-comunista e l'entrata nel governo del Pci⁴⁵, quando l'intero Paese fu scosso dall'assassinio di Aldo Moro, le nefaste previsioni sul collasso

42. Lazzaretto e Simone, *Dall'università d'élite all'università di massa*, pp. 68-112.

43. Il 1951 è stato l'anno con il minor numero di laureati, solo 5. Rispetto al decennio 1941-1950, negli anni Cinquanta il numero di laureati si è contratto del 15%.

44. Si vedano, tra gli altri, il ricordo del rettore dell'Università di Padova: Luciano Merigliano, *Eventi e risultati più significativi del mio Rettorato (1972-1984)*, Treviso 2000; *Il mondo della guerra fredda e l'Italia degli anni di piombo. Una regia internazionale per il terrorismo?*, a cura di Valentine Lomellini, Milano 2017: in particolare i saggi di Alba Lazzaretto (*Il Senato accademico dell'Università di Padova negli "anni di piombo"*) e di Giulia Simone (*Violenza ed eversione negli anni Settanta: la risposta istituzionale della Facoltà di Scienze politiche di Padova*); inoltre Francesco Jori, *La stagione del terrorismo a Padova, in Il Veneto nel secondo Novecento. Politica e istituzioni*, a cura di Filiberto Agostini, Milano 2015, pp. 228-237. Tra i docenti che vissero quelle drammatiche esperienze: Angelo Ventura, *Per una storia del terrorismo italiano*, Roma 2010; Guido Petter, *I giorni dell'ombra. Diario di una stagione di violenza italiana*, Padova 2011.

45. I diplomatici statunitensi, usando la metafora del «ponte verso l'abisso» per descrivere la situazione italiana, preconizzavano una deriva politica senza ritorno: cfr. Umberto Gentiloni Silveri, *L'Italia sospesa. La crisi degli anni Settanta vista da Washington*, Torino 2009, pp. 190-229; Roberto Fornasier, *The Dove and the Eagle*, Cambridge Scholars Publishing, Newcastle upon Tyne, 2012, pp. 132-138.

della società non si avverarono e, soprattutto, quegli eventi non riuscirono a scardinare i meccanismi democratici della Repubblica italiana. Anzi, va ricordato che negli anni Settanta furono intrapresi importanti provvedimenti riformatori della scuola, come i “quattro decreti delegati” del 1974 che democratizzarono gli organi collegiali scolastici⁴⁶. All’Università di Padova, negli anni Settanta, e in particolare dal 1973 in avanti, si nota un *trend* di iscritti e di laureati in irrefrenabile ascesa: i 116 neo-dottori in Scienze politiche del 1973 diventarono addirittura 258 nel 1980. L’aumento percentuale più significativo (del 60%) si registrò tra il 1974 e il 1975, mentre l’incremento tra inizio e fine decennio superò il 260%.

Contenuta la spinta eversiva, lo Stato italiano si trovò – nel corso degli anni Ottanta – a patire le conseguenze di un sistema politico ingessato in una partitocrazia irriformabile e in una crescente difficoltà economica. Tuttavia, negli anni Ottanta la crescita del numero dei laureati in Scienze politiche continuò, seppur riportando una diminuzione nel triennio 1983-1985; percentualmente, si rileva un ulteriore aumento del 25% tra inizio e fine del decennio⁴⁷.

<i>Periodo</i>	<i>Numero lauree/diplomi</i>	<i>Percentuale sul totale</i>
1926-1930	87	0,35%
1931-1940	287	1,15%
1941-1950	169	0,68%
1951-1960	143	0,57%
1961-1970	471	1,89%
1971-1980	1.705	6,84%
1981-1990	2.268	9,10%
1991-2000	5.965	23,93%
2001-2010	12.398	49,73%
2011	1.436	5,76%
Totale	24.929 ⁴⁸	

46. Così il Dpr n. 416, che si riprometteva di dare alla scuola il carattere «di una comunità che interagisce con la più vasta comunità sociale e civica», prevedeva il diritto di assemblea sia per gli studenti delle secondarie superiori, sia per i genitori: si vedano in particolare gli articoli 1, 42-44 e 45. Altro provvedimento degno di nota è la legge n. 748 dell’11 ottobre 1977, che introdusse la pubblicità delle sedute degli organi collegiali scolastici (art. 1).

47. Un rinnovamento legislativo avvenne con la legge n. 168 del 9 maggio 1989, che concesse «autonomia didattica, scientifica, organizzativa, finanziaria e contabile» alle università italiane (art. 6).

48. Le tabelle relative ai laureati in Scienze politiche forniscono la cifra complessiva di 20.928 persone laureate, ma 24.929 titoli conseguiti, perché, con l’introduzione del sistema “3+2”, una persona viene conteggiata due volte alla fine del suo percorso accademico quinquennale.

4. L'introduzione delle lauree triennali e specialistiche/magistrali

Il processo di democratizzazione del sistema educativo e accademico italiano fece un balzo in avanti negli anni Novanta, anche grazie allo stimolo offerto dalle leggi Bassanini⁴⁹. Per dare un'idea della profondità dei cambiamenti in atto, basti notare che, dal momento dell'istituzione della Scuola di Scienze politiche fino al 1990, ossia in 65 anni, hanno conseguito un titolo universitario 5.130 studenti; ebbene, nel solo decennio successivo, dal 1991 al 2000, si sono laureate 5.965 persone⁵⁰.

In tale contesto una trasformazione epocale si verificò quando venne previsto il passaggio dalle lauree quadriennali (o quinquennali) al cosiddetto 3+2, ossia un percorso che congiungeva una laurea triennale a una specialistica biennale (o magistrale). La legge n. 127 del 15 maggio 1997 (art. 17, co. 95-101) fu attuata con decreto ministeriale n. 509 del 3 novembre 1999, che garantì agli Atenei la libertà di “costruire” percorsi di studio adeguati alle esigenze della locale realtà economica e sociale, nel rispetto dei criteri generali – definiti a livello nazionale – in termini di obiettivi da raggiungere e di aspetti generali delle attività formative. A Padova, presso la Facoltà di Scienze politiche, vennero attivati sette corsi di laurea triennale⁵¹; accanto a questi vennero previsti sei corsi di laurea specialistica⁵². Il sistema così riformato ha prodotto laureati a partire dal 2001-2002 e fino al 2007-2008⁵³.

Dai dati numerici si ricava che 15.071 studenti hanno ottenuto una laurea quadriennale, pre-riforma del 1999; 5.642 sono invece i laureati nei sette di-

49. Merita menzione l'art. 21 della legge n. 59 del 15 marzo 1997, che concedeva autonomia organizzativa e didattica alle istituzioni scolastiche. Nel 1994 il testo unico del decreto legislativo n. 297 portava ordine nell'insieme di leggi di riforma della scuola italiana, prevedendo al suo titolo primo norme per gli organi collegiali e per le assemblee di studenti e di genitori.

50. L'aumento di laureati nel corso degli anni Novanta è stato del 113%; nel solo 1995 si è registrato un incremento del 25% sull'anno precedente e nel 1996 di circa il 22%.

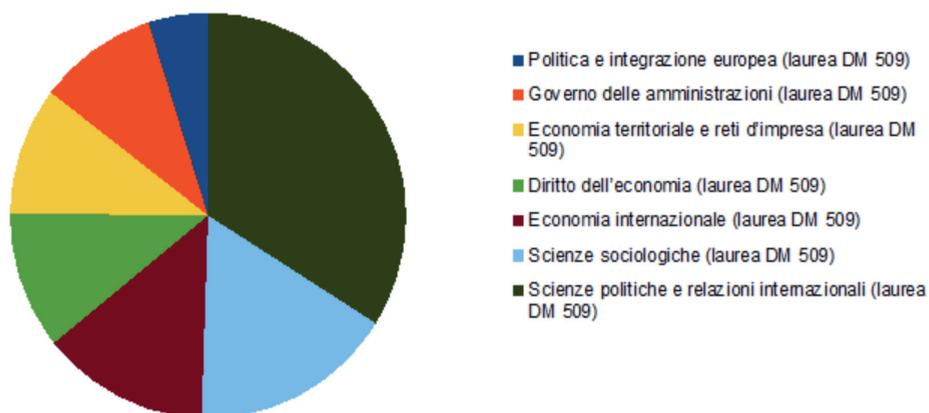
51. “Diritto dell'economia” (DEC - Classe II); “Economia internazionale” (ECI - Classe XXVIII); “Economia territoriale e reti d'impresa” (ETR - Classe XXVIII); “Governo delle amministrazioni” (GAM - Classe XIX); “Politica e integrazione europea” (PIE - Classe XV); “Scienze politiche e relazioni internazionali” (SPR - Classe XV); “Scienze sociologiche” (SCS - Classe XXXVI).

52. In “Sociologia” (SOC - Classe 89/S); “Economia dei sistemi produttivi” (ESP - Classe 64/S); “Istituzioni e politiche dei diritti umani e della pace” (IPD - Classe 60/S); “Politica internazionale e diplomazia” (PID - Classe 60/S); “Diritto dell'integrazione europea” (DIE - Classe 99/S); “Politiche dell'Unione Europea” (PUE - Classe 99/S). Cfr. il “Bollettino-Notiziario” della Facoltà di Scienze politiche, Università degli Studi di Padova, a.a. 2003-2004, pp. 7-8.

53. Gli studenti che si erano iscritti al corso di laurea in Scienze politiche in anni accademici precedenti al 2001-2002 mantennero invece l'ordinamento quadriennale fino al completamento del loro percorso accademico.

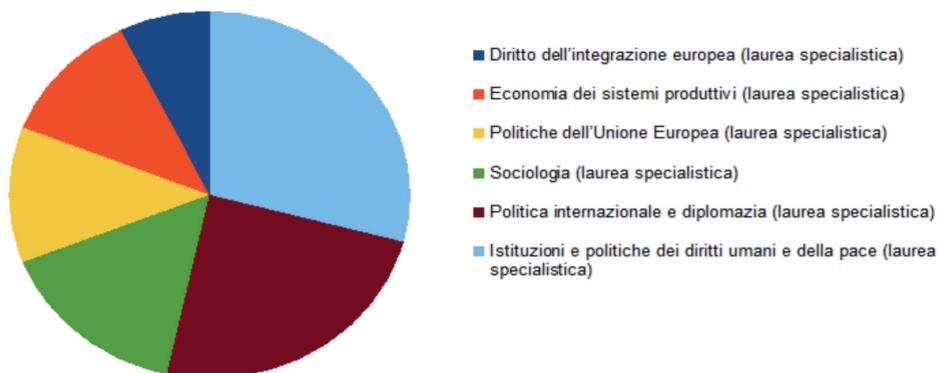
versi corsi triennali offerti con l'entrata in vigore del decreto 509. Di questi, uno su tre ha optato per il tradizionale percorso in Scienze politiche e relazioni internazionali; il 16,6% ha preferito l'indirizzo sociologico e il 13,6% quello economico internazionale. A conseguire una delle sei lauree specialistiche sono stati 1.346 studenti: oltre il 29% di costoro ha scelto il percorso in Istituzioni e politiche dei diritti umani e della pace, mentre il 24,3% quello in Politica internazionale e diplomazia.

Laureati per corso triennale - riforma 1999



<i>Lauree triennali D.M. 509</i>	<i>N. laureati</i>	<i>%</i>
Politica e integrazione europea	275	4,87%
Governo delle amministrazioni	549	9,73%
Economia territoriale e reti d'impresa	577	10,23%
Diritto dell'economia	620	10,99%
Economia internazionale	770	13,65%
Scienze sociologiche	937	16,61%
Scienze politiche e relazioni internazionali	1.914	33,92%
Totale	5.642	100,00%

Laureati per corso specialistico - riforma 1999



<i>Lauree specialistiche D.M. 509</i>	<i>N. laureati</i>	<i>%</i>
Diritto dell'integrazione europea	98	7,28%
Economia dei sistemi produttivi	158	11,74%
Politiche dell'Unione Europea	161	11,96%
Sociologia (laurea specialistica)	210	15,60%
Politica internazionale e diplomazia	327	24,29%
Istituzioni e politiche dei diritti umani e della pace	392	29,12%
Totale	1.346	100,00%

Nel 2004 è intervenuta una ulteriore riforma attraverso il decreto ministeriale 270, che ha rimesso mano all'offerta dei corsi di laurea universitari, prevedendo nuovi e diversi corsi triennali e percorsi magistrali biennali⁵⁴. Complessivamente, l'effetto combinato delle due riforme ha permesso a 12.398 studenti di conseguire un titolo in Scienze politiche nel decennio 2001-2010, molti di più della cifra cumulativa dei laureati dagli anni Venti del Novecento al 2000 (11.095 laureati). Quello che ne deriva è che, nel giro di pochissimo tempo, i laureati sono raddoppiati rispetto all'inizio del nuovo millennio; la

54. La Facoltà di Scienze politiche offrì quattro corsi triennali, "Diritto dell'economia e Governo delle organizzazioni" (DOR - Classe L-14); "Economia internazionale" (ECI - Classe L-33); "Scienze politiche, Studi internazionali ed europei" (SSE - Classe L-36); "Scienze sociologiche" (SCS - Classe L-40). Otto furono invece i corsi di laurea magistrale proposti: "Istituzioni e Politiche dei Diritti umani e della Pace" (IPD - Classe LM-52); "Politica internazionale e Diplomazia" (PID - Classe LM-52); "Economia internazionale" (MEI - Classe LM-56); "Scienze del governo e politiche pubbliche" (SGP - Classe LM-63); "Scienze del Servizio sociale" (SSS - LM-87); "Sociologia" (SOC - Classe LM-88); "Diritto, Istituzioni e Politiche dell'integrazione europea" (DIP - Classe LM-90); "Studi europei" (SEU - Classe LM-90).

punta massima si è raggiunta nel 2008, quando in dodici mesi sono stati censiti ben 1.573 neodottori⁵⁵. La media annua dei laureati è cresciuta dai circa 170 nel decennio 1971-1980 a 227 in quello successivo; negli anni Novanta tale cifra è più che raddoppiata, balzando a 596, per poi replicare nuovamente – con 1.240 laureati/anno – nel corso della prima decade del Duemila. Ciò vuol dire che, considerando tutti i laureati in Scienze politiche, la metà di loro ha conseguito il titolo nel decennio 2001-2010.

Va infine rilevato che il legislatore ha apportato trasformazioni radicali alla forma strutturale stessa degli atenei italiani. Così, se lo statuto dell'Università di Padova del 1995 prevedeva che «le strutture didattiche fondamentali dell'Università» fossero le Facoltà, intese come complessi «di appartenenza dei professori e dei ricercatori per lo svolgimento della loro attività didattica»⁵⁶, in seguito all'entrata in vigore della legge n. 240 del 2010 venne previsto un nuovo ordinamento statutario, che invece riconosceva nei Dipartimenti «le strutture deputate in via primaria alla promozione e all'organizzazione della ricerca scientifica, oltre che delle attività didattiche e formative»⁵⁷. I Dipartimenti, «articolazioni interne dell'Università», a loro volta avrebbero potuto raggrupparsi «in strutture di raccordo denominate Scuole di Ateneo»⁵⁸. In tale contesto riformatore, una volta abolita la Facoltà di Scienze Politiche, venne creato il Dipartimento di Scienze Politiche, Giuridiche e Studi Internazionali, che riunisce studiosi di area politologica, sociologica, giuridica, storica, filosofica e linguistica e che, con il Dipartimento di Scienze economiche e aziendali, fa parte della Scuola di economia e Scienze politiche⁵⁹.

5. L'ascesa della componente femminile

È necessario ora indugiare sul fatto – già citato – che dopo la seconda guerra mondiale il numero di studentesse universitarie è cresciuto notevolmente. Su un totale di 20.928 persone laureate in Scienze politiche a Padova fino al 2011, la percentuale tra i due sessi è egualmente distribuita, con una leggerissima prevalenza maschile (50,94%). Ma il dato va “spalmato” nel corso dei decenni.

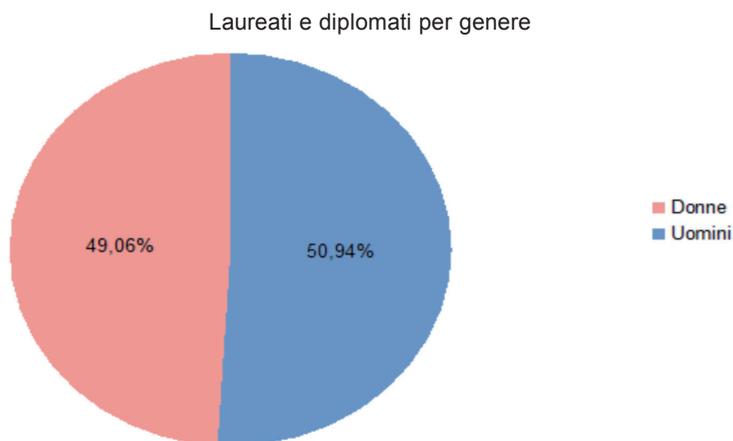
55. Complessivamente, nel corso del decennio, i laureati sono aumentati del 91% circa.

56. Statuto emanato con decreto rettorale n. 94 dell'8 novembre 1995: si vedano gli articoli 28, co. 1, e 29, co. 1.

57. Statuto emanato con decreto rettorale n. 3276 del 16 dicembre 2011, art. 2, co. 3.

58. Statuto del 2011, art. 39, co. 1, e art. 43, co. 1.

59. Va rilevato che, in seguito alla riforma del 2004, ma prima della soppressione della Facoltà, sono state discusse 113 lauree triennali e 90 magistrali del nuovo ordinamento didattico.



Nei primi cinque anni di apertura nessuna donna si diplomò alla Scuola patavina di Scienze politiche. Nel 1931, a fronte di 30 colleghi maschi laureati, comparve la prima dottoressa della Facoltà. In tutti gli anni Trenta, solo 12 conseguirono il titolo universitario, rappresentando il 4,1% del totale. Negli anni Quaranta la difficile situazione contingente fece diminuire drasticamente i numeri dei laureati, sia maschi che femmine; solo 5 donne riuscirono nell'impresa di discutere la loro tesi nell'arco dell'intero decennio⁶⁰. Dal 1947, per sei anni consecutivi nessuna conseguì la laurea in Scienze politiche. Negli anni Cinquanta 22 donne "conquistarono" il titolo: il numero, che in termini assoluti appare ridottissimo, rappresenta comunque il 15,4% del totale dei laureati. A partire dal 1961, con l'unica eccezione del 1963, il numero delle donne laureate ha superato sempre la decina; nell'intero arco degli anni Sessanta, 179 donne pervennero al titolo di studio (il 38% dei neodottori). Negli anni Settanta le laureate furono 457, il 26,8% del totale; nel solo anno 1979 ben 70 studentesse riuscirono a discutere la loro tesi⁶¹. Il *trend* di crescita di iscritti – e di laureati – subì un rallentamento nella prima metà degli anni Ottanta. Nel 1984 le laureate furono poco più della metà rispetto al 1979⁶². È a

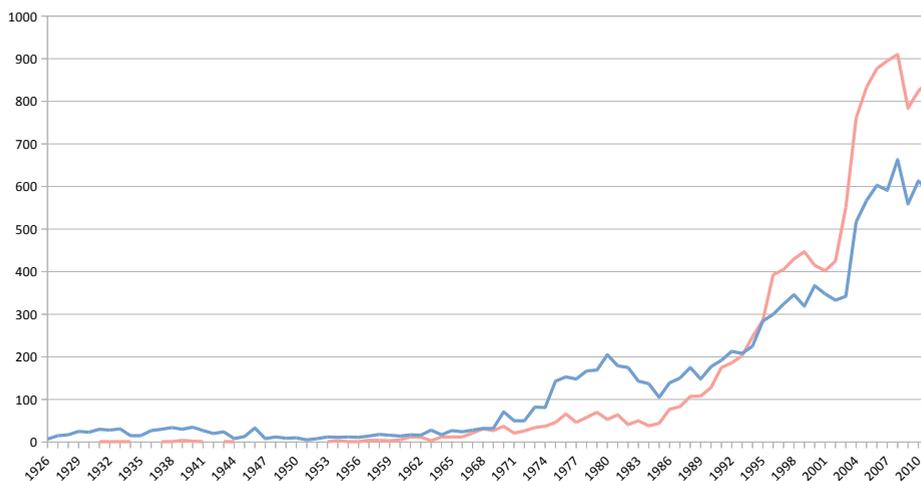
60. Se il numero di laureati maschi, nel decennio 1931-1940, fu di 275, in quello successivo scese a 164, facendo registrare una contrazione del 40%; la riduzione del numero di donne laureate (un calo di quasi il 60%) è meno significativa statisticamente considerando l'estrema esiguità dei numeri in esame (12 contro 5).

61. Si nota che l'impennata di laureati maschi a partire dalla metà del decennio, per le cause sopra ricordate, non è segnata da un eguale aumento delle laureate. Il 1975, in particolare, registra un incremento del 76% di iscritti rispetto all'anno precedente. Nel 1977, ad esempio, le donne laureate (46) sono meno di un terzo rispetto agli uomini (148).

62. Analoga diminuzione si registra tra i maschi e, a fronte di 205 laureati nel 1980, si scende a 105 dopo cinque anni.

partire dalla seconda metà degli anni Ottanta che il numero di donne laureate è aumentato costantemente: dal 1985 al 2011 si sono registrati quasi unicamente scarti incrementali positivi da un anno al successivo⁶³, e la forbice tra il numero di uomini e di donne laureati si è fatta sempre più stretta nel corso degli anni Novanta. Il numero complessivo di donne laureate negli anni Ottanta è di 457 unità, poco più di un quarto (il 26,8%) del totale; la cifra sale a 740 negli anni Novanta e rappresenta quasi un terzo della totalità dei laureati.

Numero di lauree e diplomi per anno e genere (1926-2011)



A partire dai primi anni Novanta, come mostra il grafico qui sopra riportato, i laureati in Scienze politiche cominciano a essere equamente distribuiti tra i due sessi: il 1994 è l'anno in cui il numero delle donne laureate supera quello dei colleghi maschi. A partire da allora, il divario si è allargato progressivamente a favore della componente femminile⁶⁴. Il decennio 1991-2000 conta così 3.187 laureate (53% del totale) contro 2.778 laureati: ciò vuol dire che, ogni 100 uomini laureati, si sono celebrate quasi 115 lauree femminili. È interessante notare che negli anni Novanta il numero delle laureate è cresciuto del 330% rispetto al dato cumulativo del decennio precedente, percentuale che fa “impallidire” quella – comunque molto positiva (ossia un aumento di quasi l'82%) – dei laureati complessivi di sesso maschile⁶⁵.

63. Le uniche eccezioni sono quelle relative agli anni 2000, 2001 e 2009.

64. Nel 1999 le donne hanno superato di 128 unità il numero dei laureati di sesso maschile.

65. Nel solo anno 2000 si sono celebrate 415 lauree con candidate femminili e 367 con studenti maschi: rispetto a dieci anni prima, ossia al dato del 1991, si rileva un incremento del 137% per il sesso femminile.

Successivamente lo scarto in termini numerici è costantemente cresciuto, di anno in anno, a favore della componente femminile: più 54 nel 2001; 209 nel 2003; 266 nel 2005; addirittura 304 nel 2007. Complessivamente, nella prima decade del Duemila, si sono laureate 7.262 donne (il 58,6% del totale) e 5.136 uomini, il che vuol dire che – ogni 100 studenti maschi laureati – sono state registrate 141 lauree di donne⁶⁶. I dati mostrano chiaramente come la Facoltà di Scienze politiche sia diventata progressivamente una sede prediletta dalle studentesse universitarie, soprattutto per quanto riguarda alcuni corsi triennali (ex D.M. 509), come Scienze politiche e relazioni internazionali (1.204 laureate contro 710 uomini) e Scienze sociologiche, dove le laureate quasi doppiano il numero dei colleghi maschi (623 lauree contro 314). In tutte le lauree specialistiche le donne prevalgono, più o meno nettamente, rispetto ai colleghi maschi, con una punta massima in Politiche dei diritti umani e della pace⁶⁷. Il *gender gap*, per molti decenni fortemente sbilanciato a sfavore della componente femminile, si è completamente rovesciato e oggi Scienze politiche è una Facoltà, o meglio un Dipartimento, a “piena tinta rosa”. Anche se i dati non ci permettono di avanzare ipotesi sulle ragioni recondite di tale scelta, va aggiunto che il fenomeno, visto in prospettiva, non deve certo meravigliare; senza voler entrare nel merito rispetto al processo di emancipazione femminile, è innegabile che la società veneta e italiana sia stata interessata da un cambiamento radicale dei rapporti sociali e di genere e da una profonda trasformazione della mentalità stessa delle nuove generazioni, che hanno portato a reputare il conseguimento di una laurea triennale come un normale proseguimento del liceo e a considerare le donne, sempre più istruite e occupate, come delle vere *breadwinner*⁶⁸.

6. Veneti, italiani e stranieri

Passando all'analisi dei laureati per luogo di nascita, risulta che gli italiani rappresentano il 95,56% del totale⁶⁹; tra loro svettano i residenti nel comune di Padova, che raccoglie ben 3.692 dottori in Scienze politiche, il 18,5% di

66. Il numero dei laureati nel corso del decennio 2001-2010 è aumentato del 127%, rispetto a quello precedente, per le donne; ma solo dell'85% per gli uomini.

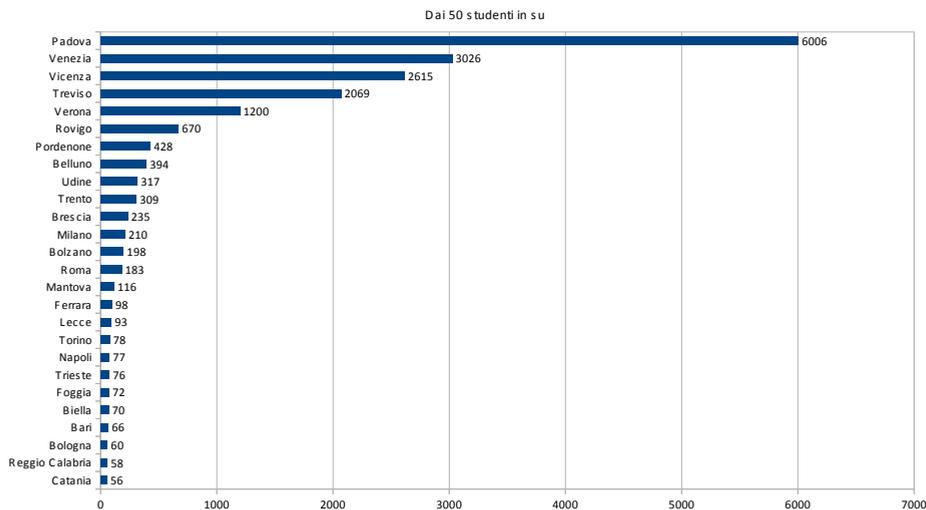
67. Elevatissimo è anche lo scarto ravvisabile nella laurea specialistica in Sociologia, che conta 149 donne laureate a fronte di 61 uomini.

68. Il termine è usato da Alisa Del Re e Lorenza Perini, *Le donne nel Veneto dallo sviluppo alla crisi*, in Agostini, *Il Veneto nel secondo Novecento*, p. 412. Delle stesse autrici si veda anche *Uno straordinario impasto di vecchio e di nuovo: le donne, il lavoro, la famiglia nel Veneto degli anni Duemila*, in Filiberto Agostini e Giovanni Silvano (a cura di), *Il Veneto dopo il Novecento. Politica e società*, Milano 2019, pp. 151-171.

69. In cifre assolute, si tratta di 19.998 nati in Italia e 930 all'estero.

tutti i laureati di nazionalità italiana. Se poi si considerano i residenti nell'intera provincia (6.006 persone), la percentuale sale al 30% di tutti i laureati italiani della Facoltà.

Provenienza degli studenti italiani per provincia attuale (1921-1989)



Continuando il vaglio dei laureati per province, a Padova segue il Veneziano, dove sono registrati 3.026 dottori in Scienze politiche, il 15% del totale; il Vicentino, con 2.615 laureati (13%); il Trevigiano, con 2.069 (10,3%). Molto distanziati sono il Veronese, che con 1.200 laureati raccoglie solo il 6% del totale e il Rodigino, con 670 dottori in Scienze politiche (3,3%). La settima provincia veneta, il Bellunese, si vede scavalcata, in termini assoluti di numero di laureati, dal Pordenonese (428 contro 394 laureati). Seguono altre province limitrofe al Veneto, come Udine (317 laureati), Trento (309) e Brescia (235)⁷⁰.

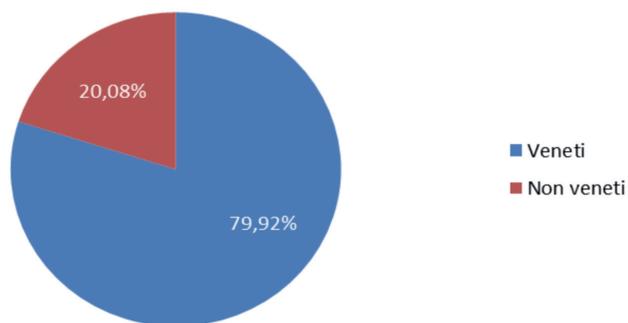
70. Seguono anche Bolzano (198), Mantova (116) e Ferrara (98).

I LAUREATI IN SCIENZE POLITICHE DELL'UNIVERSITÀ DI PADOVA (1926-2011)

<i>Laureati per regione italiana</i>		
<i>Regione</i>	<i>Laureati</i>	<i>Percentuale</i>
Veneto	15.980	79,90%
Friuli Venezia Giulia	856	4,28%
Lombardia	689	3,45%
Trentino Alto Adige	507	2,54%
Emilia-Romagna	279	1,40%
Piemonte	207	1,04%
Liguria	76	0,38%
Valle d'Aosta	3	0,01%
Regioni nord Italia	18.597	93%
Lazio	225	1,13%
Campania	162	0,81%
Toscana	88	0,44%
Abruzzo	76	0,38%
Marche	62	0,31%
Umbria	17	0,08%
Molise	9	0,04%
Regioni Centro Italia	639	3,2%
Puglia	308	1,54%
Sicilia	217	1,09%
Calabria	130	0,65%
Sardegna	73	0,36%
Basilicata	31	0,16%
Regioni Sud Italia	759	3,8%

Complessivamente i veneti – intesi come residenti nelle sette province della regione – assommano a 15.980 unità, circa l'80% di tutti i laureati con nazionalità italiana usciti dalla Facoltà di Scienze politiche.

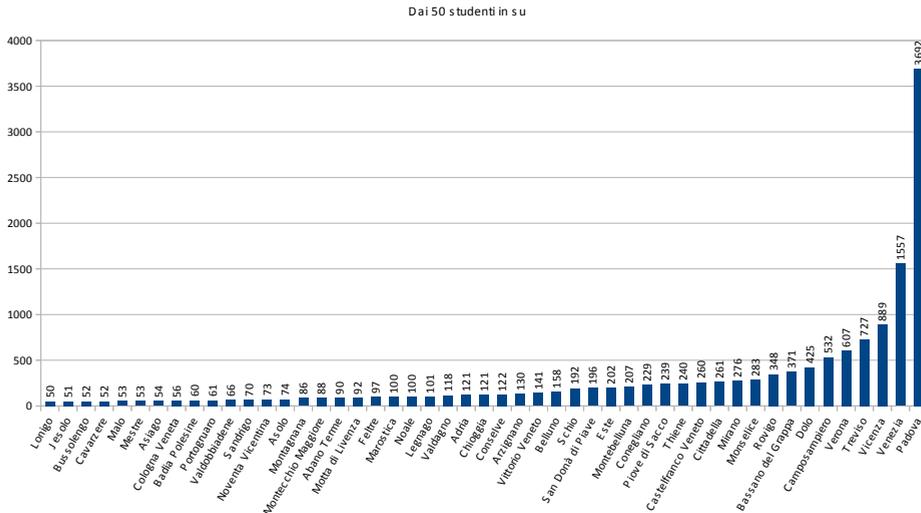
Percentuali di studenti provenienti dal Veneto attuale e dal resto d'Italia



Tra i “non veneti”, alto è il contingente di studenti provenienti dalle regioni limitrofe, come il Friuli-Venezia Giulia (856 laureati) e il Trentino-Alto Adige, con 507⁷¹; numerosi sono anche i lombardi (689). 207 sono i piemontesi, 279 gli emiliani e 76 i liguri. Nel complesso, il Nord Italia – Veneto escluso – accoglie 2.617 laureati in Scienze politiche (13,1% del totale degli italiani)⁷². Nel Centro Italia il Lazio vanta 225 laureati a Padova; la Campania 162; la Toscana 88; l’Abruzzo 76; le Marche 62; l’Umbria 17; il Molise 9: in totale si tratta di 639 persone, pari al 3,2% del totale. A Sud, invece, è la Puglia la regione che ha fornito più laureati in Scienze politiche presso l’Ateneo patavino, ben 308, seguita dalla Sicilia con 217, dalla Calabria (130), dalla Sardegna (73) e dalla Basilicata (31). Nel loro insieme, il Mezzogiorno e le Isole contano 759 laureati, il 3,8% del totale.

Tornando all’analisi dei laureati veneti, a Padova seguono altri quattro comuni capoluogo di provincia: Venezia, con 1.557 laureati (il 7,8%); Vicenza, con 889 (il 4,4%); Treviso, con 727 (il 3,6%); Verona, con 607 (il 3%). Le altre due città capoluogo – Rovigo e Belluno – sono invece presenti al nono e al ventunesimo posto tra i centri veneti, poiché ospitano rispettivamente 348 e 158 laureati in Scienze politiche.

Località di provenienza degli studenti dall’attuale Veneto (1926-2011)



71. È interessante notare che – nel 1952 – sono stati inaugurati i corsi estivi presso la sede di Bressanone (Bz); cfr. Simone, *La Facoltà di Scienze politiche (1945-1968)*, p. 135; Ead., «*La Facoltà Cenerentola*», p. 53.

72. Si considerano anche 3 laureati dalla Valle d’Aosta. Si veda inoltre, poco più avanti, la mappa della penisola che mostra, con dei punti proporzionali al numero di studenti, il luogo di nascita dei laureati italiani in Scienze politiche a Padova.

Sembra naturale che i comuni geograficamente vicini a Padova registrino ragguardevoli numeri di laureati, sia in termini assoluti che relativi alla loro consistenza anagrafica. Camposampiero, paese a nord del capoluogo provinciale, ha raccolto 532 laureati in Scienze politiche, un numero impressionante se si considera che ospita una popolazione di poco più di 10.000 abitanti. Stesse considerazioni valgono per Dolo, comune sulla riviera veneziana del Brenta, nel quale 425 residenti si sono laureati in Scienze politiche a Padova.

Tra i comuni di dimensioni medio-grandi, sopra i 20.000 abitanti, compaiono diverse cittadine vicentine, ossia Bassano del Grappa (con 371 laureati), Thiene (240), Schio (192), Arzignano (130), Valdagno (118) e Montebelluna Maggiore (88). Tra i comuni del Padovano: Albignasego (23 laureati) e Selvazzano Dentro (15)⁷³; per quanto riguarda il Veneziano, primeggia Mirano (con 276 laureati), seguita da San Donà di Piave (196), Chioggia (121), Portogruaro (61), Mestre (53), Jesolo (51) e Mira (22). Nel Trevigiano quattro città rivelano numeri interessanti: Castelfranco Veneto (260 laureati), Conegliano (229), Montebelluna (207) e Vittorio Veneto (141)⁷⁴. Infine, nel Veronese, dopo Legnago (101 laureati), seguono San Bonifacio (42) e Villafranca (25).

Passando all'analisi dei comuni di dimensioni medio-piccole, compresi tra i 10.000 e i 20.000 abitanti, oltre ai sopraccitati Camposampiero e Dolo, si presentano numerosi altri centri della pianura veneta, tra cui Monselice (con 283 laureati), Cittadella (261 laureati), Piove di Sacco (239), Este (202), Conselve (122), Noale (100), Marostica (100), Motta di Livenza (92), Abano Terme (90), Malo (53), Cavarzere (52), Lonigo (50), Oderzo (31), Piazzola sul Brenta (19), Vigodarzere (13), Pianiga (11), Campodarsego (11). Per quanto riguarda i comuni del Basso Veneto, appaiono Adria (con 121 laureati), Badia Polesine (60) e Lendinara (35); nel Veronese Bussolengo (52 laureati), Negrar (46), Zevio (45), Isola della Scala (26) e Peschiera del Garda (12). Nella zona settentrionale della regione, spiccano Feltre (97 laureati), Valdobbiadene (66) e Pieve di Soligo (13). Tra i comuni piccoli o piccolissimi, diversi appartengono a realtà collinari, pedemontane o montane, come Asolo (74 laureati), Pederobba (16), Farra di Soligo (14) e Crespano del Grappa nell'alta Marca Trevigiana (14); Pieve di Cadore (45), Agordo (30) e Lamon (15) tra le montagne bellunesi; Asiago nel Vicentino (54)⁷⁵.

Volendo semplificare, i dati sembrano indicare che l'Ateneo di Padova – e con esso la Facoltà di Scienze politiche – sia stato scelto in maggioranza da

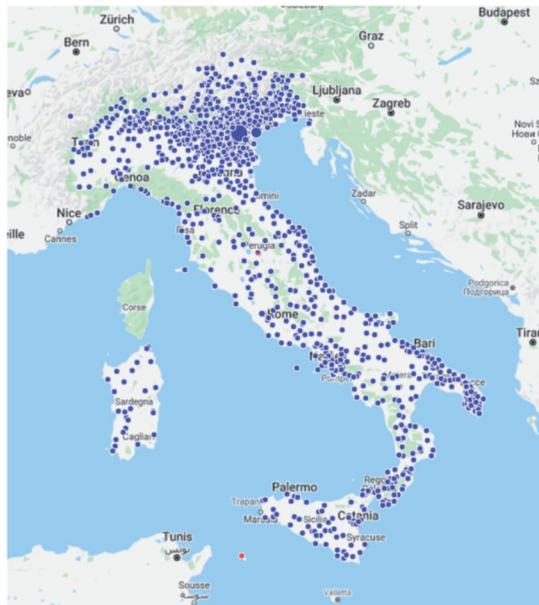
73. I dati relativi ai Comuni veneti con meno di 50 laureati in Scienze politiche non sono riportati nella loro interezza nel presente saggio.

74. Molto più contenuti i laureati provenienti da Mogliano Veneto (15).

75. Tra i Comuni sotto i 10.000 abitanti andrebbero ricordati anche alcuni centri in località pianeggianti del Padovano – Montagnana (che si distingue con i suoi 86 laureati), Campo San Martino (27), Anguillara Veneta, Santa Giustina in Colle – e del Vicentino, con Noventa Vicentina e Sandrigo (73 e 70 laureati, rispettivamente).

studenti delle zone centrali del Veneto, quelle del “quadrilatero bianco”, per usare un’espressione coniata dai politologi Ilvo Diamanti e Gianni Riccamboni⁷⁶; si tratta di un territorio nel quale la Chiesa ha offerto gli orientamenti di fondo della popolazione; una terra politicamente dominata dalla Democrazia cristiana, «garante e interprete della società locale», e dalle sue associazioni collaterali⁷⁷; un luogo passato in una manciata di anni dalla povertà post-bellica alla micro-impresa diffusa⁷⁸.

Dagli elementi in nostro possesso non siamo però in grado di vagliare ulteriormente il dato, distinguendo chi ha volontariamente e consapevolmente scelto Padova come sede qualificata del proprio percorso di studi da chi ha seguito il più prosaico criterio della vicinanza geografica alla propria residenza – il che spiegherebbe l’alto numero di studenti del Padovano – e della comodità dei trasporti verso la città del Santo. I numeri, qualunque possa essere la ragione, testimoniano comunque la percentuale relativamente bassa di veneti provenienti da zone nelle quali altri Atenei si sono posti in concorrenza più o meno diretta con Padova, come quelli di Venezia, Ferrara e Brescia e le più recenti sedi di Verona e Trento.



76. Ilvo Diamanti e Gianni Riccamboni, *La parabola del voto bianco. Elezioni e società in Veneto (1946-1992)*, Vicenza 1992.

77. Marco Almagisti e Nicola Scarnera, 1946-2014. *Capitale sociale e partiti nel Veneto*, in Agostini, *Il Veneto nel secondo Novecento*, pp. 146-147.

78. Paolo Giaretta, *Identità e rappresentanza politica nel Veneto*, in Agostini, *Il Veneto nel secondo Novecento*, p. 170.

L'ultima serie di considerazioni può riguardare i laureati «nati all'estero» per capire il grado di internazionalizzazione della Facoltà. L'Ateneo, nel suo complesso, ha rappresentato nel corso dei secoli la meta di numerosi studenti stranieri; è sufficiente entrare nella Sala dei Quaranta, vicina alla storica Aula Magna di Palazzo Bo, per stupirsi di fronte al numero e alla qualità degli studenti stranieri transitati per Padova e rappresentati lungo le pareti⁷⁹.

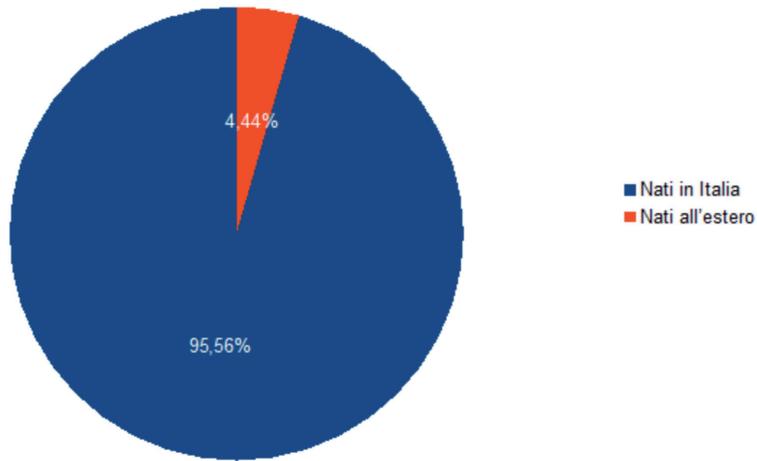


Un dato importante da evidenziare riguarda i 20.928 laureati complessivi in Scienze politiche: 930 di questi provengono dall'estero, vale a dire il 4,44% del totale⁸⁰.

79. Il sito dedicato a celebrare gli ottocento anni dell'Università citato nell'introduzione sottolinea, a proposito degli "anni Duemila", che l'università di Padova «del nuovo millennio è sempre più orientata all'internazionalizzazione», oltre che allo sviluppo della ricerca, alla specializzazione delle competenze e alla diversificazione dell'offerta formativa. Si veda inoltre: www.ottocentenariouniversitadipadova.it/timeline/gli-anni-duemila [accesso effettuato il 30 luglio 2020].

80. La mappa dei continenti qui sopra riportata esprime, attraverso dei punti proporzionali al numero di studenti, le località di provenienza degli stranieri laureati in Scienze politiche a Padova.

Laureati e diplomati per luogo di nascita (Italia/estero) (1926-2011)



Tra i Paesi di origine dei laureati “non italiani” svetta nettamente l’Albania, che conta 147 studenti che hanno concluso il loro percorso accademico a Padova. Già la semplice constatazione di una simile cifra induce a riflettere su un Paese demograficamente contenuto, ma capace di “sforare” una tale messe di dottori. Il dato è ancor più interessante se si tiene conto che – in numeri assoluti – l’Albania è seguita rispettivamente dalla Svizzera con 71 laureati, dalla Croazia con 66, dalla Romania con 52 e dalla Grecia con 46. Va aggiunto che la frequenza di università italiane da parte di studenti albanesi è un fenomeno relativamente recente, successivo alla morte del dittatore Enver Hoxha nel 1985, quando il regime socialista albanese – fino ad allora isolato dal resto del mondo – si è avviato alla sua dissoluzione. Emblema della crisi politica ed economica del Paese fu l’approdo di 17.000 profughi il 7 agosto 1991, presso il porto di Bari⁸¹. È solo dopo tale data che i flussi emigratori verso l’Italia sono cresciuti esponenzialmente, considerando che per molti abitanti del piccolo Paese balcanico la conoscenza dell’italiano era generalmente molto buona⁸².

Dopo la seconda guerra mondiale l’Ateneo di Padova, a ridosso della “cortina di ferro” e alle porte del mondo slavo, divenne «un punto di riferimento

81. Luigi Vittorio Ferraris (a cura di), *Manuale della politica estera italiana 1947-1993*, Roma-Bari 1996, p. 458; Jean-Baptiste Duroselle, *Storia diplomatica dal 1919 ai nostri giorni*, Milano 1998, p. 730.

82. Tra le ipotesi che si possono avanzare, ma che andrebbero suffragate da ulteriori ricerche e dalla comparazione di dati inter-Facoltà, vi è quella che molti abbiano “ripiegato” verso Scienze politiche come scelta contingente rispetto ad altri corsi di laurea a numero chiuso, che prevedevano test d’ingresso più o meno selettivi.

culturale e istituzionale per le popolazioni giuliano-dalmate»⁸³. Se si considerano complessivamente tutti i Paesi dell'ex-Jugoslavia, i numeri sono ancora più indicativi: dopo la Croazia con 66 laureati in Scienze politiche, la Slovenia ne vanta 7, la Bosnia-Erzegovina 5, la Serbia, la Macedonia e il Kosovo 2, il Montenegro 1, per un totale di 85 dottori. È sicuramente rivelatore delle dinamiche politiche e migratorie contemporanee il dato per cui ex-Jugoslavia e Albania hanno fornito, cumulativamente, circa un quarto di tutti i laureati stranieri della Facoltà⁸⁴. Va detto che l'Europa orientale tutta ha rappresentato, nel corso dei secoli, un importante bacino di matricole – e di laureati – per l'Università di Padova. Limitandoci a considerare Scienze politiche, ai 232 dottori albanesi ed ex-jugoslavi possono essere aggregati 2 romeni e 15 moldavi, 11 bulgari, 8 ungheresi, 4 polacchi, 4 cechi e 1 slovacco⁸⁵. Il mondo ex-sovietico, invece, è stato in maniera minima attratto dalla realtà patavina: solo 13 laureati rappresentano quell'importante realtà geopolitica, ma ciò si spiega se si tiene in considerazione che la “cortina di ferro” si è dissolta assai lentamente dopo il formale scioglimento dell'Unione Sovietica nel dicembre 1991 e che molti degli Stati creatisi hanno vissuto un lungo periodo di crisi politica ed economica⁸⁶.

Gli Stati europei al di qua del mondo socialista esprimono invece, nel loro complesso, una consistenza ragguardevole in termini di laureati: ai 71 svizzeri vanno aggiunti 46 greci, 39 tedeschi e 11 austriaci, 27 francesi, 23 belgi, 3 olandesi e 2 lussemburghesi, 5 inglesi, 3 spagnoli e 2 portoghesi. Anche qui qualche semplice considerazione può essere suggerita: il lontano mondo scandinavo è quasi assente tra i laureati della Facoltà, così come poca attrazione verso Padova hanno subito gli anglosassoni e gli iberici, popoli residenti in Stati nei quali sono presenti Atenei con un passato tanto lungo e prestigioso quanto quello patavino. È normale invece che si conti una massiccia presenza di studenti provenienti da nazioni confinanti con l'Italia, dalla Svizzera ad

83. Benedetto Zaccaria, *L'Ateneo padovano e l'Europa centro-orientale, 1945-68. Un caso di diplomazia culturale*, in Lazzaretto e Simone, *Dall'università d'élite all'università di massa*, p. 103.

84. In chiave storico-documentaria, cfr. Luciana Sitran Rea e Giuliano Piccoli (a cura di), *Studenti istriani e fiumani all'Università di Padova dal 1601 al 1974*, Treviso 2004.

85. Si perviene a 327 laureati, il 35% del totale di stranieri addottorati a Padova. Va ricordato pure che l'Ateneo di Padova aveva rappresentato uno «strumento di diplomazia culturale» nei confronti dei regimi di destra instauratisi nell'Europa centro-orientale nel corso degli anni Trenta: Zaccaria, *L'Ateneo padovano e l'Europa centro-orientale*, p. 104. Per un quadro storico più lungo, *Relazioni tra Padova e la Polonia. Studi in onore dell'Università di Cracovia nel VI centenario della sua fondazione*, Padova 1964.

86. Si tratta di 4 georgiani, 2 ucraini, 2 kazaki, 2 armeni, 1 russo, 1 bielorusso, 1 azero. Se si tenesse conto anche dei 15 laureati moldavi, riportati qui tra gli europei dell'Est, il numero di laureati dall'ex Unione Sovietica salirebbe a 27.

esempio⁸⁷. E ancora: si possono considerare i laureati provenienti dall'intero bacino mediterraneo. In questa prospettiva, agli ex-jugoslavi, agli albanesi e ai greci si possono aggiungere altri Paesi affacciati sul *Mare nostrum*, come Cipro, Malta e il tradizionale nemico della Città lagunare, ossia la potenza turca⁸⁸. Queste nazioni hanno offerto quasi trecento laureati alla Facoltà e rappresentano il 32% del totale degli stranieri laureati. Se poi aggiungiamo tutti gli altri Paesi che guardano il Mediterraneo, ossia quelli asiatici, quelli della sponda sud e i due Paesi europei, si arriverebbe a un totale di 373 laureati, il 40% degli stranieri laureatisi in Scienze politiche a Padova⁸⁹. Il riferimento alla Libia dà modo di esprimere un commento sul mondo delle ex-colonie italiane: ben 54 laureati in Scienze politiche, il 5,8% del totale, provengono da quattro Paesi, la Libia (20), la Somalia (17)⁹⁰, l'Etiopia (9) e l'Eritrea (8), nazioni il cui passato è fittamente intrecciato con le imprese coloniali del Regno d'Italia.

Qualche altra osservazione potrebbe riguardare il mondo musulmano: accanto al contingente turco, libico e somalo, diversi altri Stati a stragrande maggioranza islamica hanno contribuito ad incrementare il *pool* di laureati stranieri di Scienze politiche: l'Iran con ben 40 nomi; gli altri paesi africani e asiatici affacciati sul Mediterraneo; l'Arabia Saudita e anche i lontani Afghanistan e Pakistan⁹¹. Se a questi sommiamo i laureati provenienti da una realtà come l'Albania – i cui abitanti sono formalmente musulmani, sebbene in una società ex-socialista e profondamente laica – quasi il 30% dei laureati stranieri di Scienze politiche sono di fede islamica⁹².

<i>Laureati stranieri</i>		
<i>Stato</i>	<i>Laureati</i>	<i>Percentuale</i>
Albania	147	15,81%
Svizzera	71	7,63%
Croazia	66	7,10%
Romania	52	5,59%
Grecia	46	4,95%
Germania	39	4,19%

87. Nel complesso, considerando 168 europei occidentali e 71 svizzeri, si contano 239 laureati, un 25% circa del totale degli stranieri.

88. La Turchia raccoglie 15 laureati in Scienze politiche presso l'Ateneo patavino.

89. Si tratta del Libano con 6 laureati a Padova, cui vanno aggiunti 2 dalla Siria e 1 da Israele. L'Egitto vanta 4 laureati, la Libia 20, la Tunisia 3, l'Algeria 2 e il Marocco 8. Si arriva al numero complessivo includendo i laureati da Francia e Spagna.

90. Sugli stretti contatti tra la Facoltà di Scienze politiche e la Somalia, quando l'Italia ne deteneva l'amministrazione fiduciaria, si veda Simone, «*La Facoltà Cenerentola*», pp. 83-89.

91. Includendo anche 1 azero e 1 iracheno si perviene a 124 laureati, il 13% del totale.

92. Si considerano i 147 laureati albanesi e i 2 kosovari; si giunge a 273 persone totali.

I LAUREATI IN SCIENZE POLITICHE DELL'UNIVERSITÀ DI PADOVA (1926-2011)

<i>Laureati stranieri</i>		
<i>Stato</i>	<i>Laureati</i>	<i>Percentuale</i>
Francia	27	2,90%
Belgio	23	2,47%
Moldavia	15	1,61%
Bulgaria	11	1,18%
Austria	11	1,18%
Ungheria	8	0,86%
Slovenia	7	0,75%
Bosnia-Erzegovina	5	0,54%
Gran Bretagna	5	0,54%
Polonia	4	0,43%
Repubblica Ceca	4	0,43%
Spagna	3	0,32%
Olanda	3	0,32%
Svezia	3	0,32%
Malta	3	0,32%
Altri Stati europei ⁹³	17	1,83%
Stati europei	570	61,29%
Iran	40	4,39%
Turchia	15	1,61%
Libano	6	0,65%
Cina	6	0,65%
Pakistan	5	0,54%
Georgia	4	0,43%
Altri Stati asiatici ⁹⁴	18	1,94%
Stati asiatici	94	10,11%
Camerun	28	3,01%
Libia	20	2,15%
Somalia	17	1,83%
Etiopia	9	0,97%
Nigeria	8	0,86%
Marocco	8	0,86%
Eritrea	8	0,86%
Congo	6	0,65%
Egitto	4	0,43%
Angola	4	0,43%
Tunisia	3	0,32%
Altri Stati africani ⁹⁵	18	1,94%

93. Nell'ordine sono: Serbia, Portogallo, Macedonia, Lussemburgo, Kosovo, Slovacchia, Russia, Montenegro, Cipro, Bielorussia.

94. Si tratta di Siria, Arabia Saudita, Kazakistan, India, Armenia, Vietnam, Tailandia, Filippine, Israele, Iraq, Bangladesh, Azerbaijan e Afghanistan.

95. Nell'ordine Uganda, Togo, Ghana, Ciad, Algeria, Zambia, Tanzania, Senegal, Repubblica del Congo, Mauritius, Madagascar, Costa d'Avorio, Burundi.

<i>Laureati stranieri</i>		
<i>Stato</i>	<i>Laureati</i>	<i>Percentuale</i>
Stati africani	133	14,30%
Argentina	19	2,04%
Brasile	18	1,94%
Venezuela	17	1,83%
Colombia	10	1,08%
Perù	7	0,75%
Cile	4	0,43%
Uruguay	3	0,32%
Panama	3	0,32%
Messico	3	0,32%
Ecuador	3	0,32%
Altri Stati centro-sudamericani ⁹⁶	7	0,75%
Stati centro-sudamericani	94	10,11%
Canada	18	1,94%
Australia	11	1,18%
USA	9	0,97%
Bermuda	1	0,11%
Stati nordamericani/Oceania	39	4,19%

Una parola infine meritano i continenti extra-europei: il database mostra 133 laureati africani (14% circa del totale); 94 asiatici; 94 centro-sudamericani; 38 nordamericani e 11 australiani. Il numero di studenti nordamericani può apparire basso, soprattutto se si considerano gli stretti legami dell'Università di Padova con gli Stati Uniti, sviluppatasi attraverso le borse di studio e di ricerca Fullbright e la creazione, nel 1963, del Centro padovano di studi dell'Università della California⁹⁶. Ma probabilmente, come rilevato per gli inglesi, anche per i nordamericani la presenza di prestigiose università in patria, congiunta alla lontananza geografica dell'Italia, ha contribuito a contenere particolarmente l'emigrazione studentesca verso Padova.

95. Sono Puerto Rico, Honduras, Guatemala, El Salvador, Repubblica Dominicana, Bolivia. Si è optato per inserire Puerto Rico, ispanofono, tra i Paesi centro-sudamericani, mentre le Bermuda, a maggioranza inglesofona, tra gli Stati nordamericani.

96. Zaccaria, *L'Ateneo padovano e l'Europa centro-orientale*, p. 107; Simone, «*La Facoltà Cenerentola*» cit., p. 100. Per un'analisi delle Facoltà universitarie di Padova alla vigilia della loro trasformazione in Dipartimenti, si veda il primo volume della nuova collana "Rettori dell'Università di Padova. Documenti di vita accademica": *Giuseppe Zaccaria, Lasciare un'impronta. Sei anni di rettorato (2009-2015)*, a cura di Filiberto Agostini, Venezia 2020.

7. Conclusioni

Da quanto sopra esposto, non si può non essere d'accordo con l'appunto dell'ex rettore dell'Ateneo di Padova, Vincenzo Milanese, secondo il quale la riforma del 3+2 ha letteralmente «cambiato i connotati all'università italiana, modificando radicalmente l'impianto complessivo della didattica negli Atenei»⁹⁷. Il forte incremento del numero di laureati alla Facoltà di Scienze politiche nel nuovo millennio trova una facile spiegazione nell'aumento e nella diversificazione dei percorsi didattici, che hanno permesso alle matricole di scegliere tra un'ampia offerta di corsi triennali, in un'ottica più orientata alla spendibilità del titolo di studio nel mercato del lavoro, fattosi ormai globale.

Il sistema creatosi con l'entrata in vigore del D.M. 509 – e continuato con il D.M. 270 – ha modificato non solo la curva del numero dei laureati, ma anche quella dell'età media dei neodottori – che comprensibilmente si è abbassata, considerando che da un percorso quadriennale si è passati a uno triennale – e quella del voto medio di laurea. Se quest'ultimo era attestato su 100,5/110 per i laureati in Scienze politiche del vecchio ordinamento⁹⁸, quattro delle sette neonate lauree triennali hanno fatto registrare voti medi significativamente inferiori⁹⁹. Se però si considerano le lauree specialistiche, i voti sono tutti considerevolmente superiori al dato medio dei laureati quadriennali, fino ad una votazione superiore a 108/110 per i dottori in Istituzioni e Politiche dei diritti umani e della pace; in Politiche dell'Unione Europea e in Politica internazionale e diplomazia.

La Facoltà di Scienze politiche, come testimoniano i dati esposti nel corso del saggio, da scelta elitaria di pochi studenti è diventata, nel corso dei decenni, un centro attrattivo capace di sfornare migliaia di laureati ogni anno. Nel corso del tempo, la marcata politicizzazione della Facoltà, prima in chiave filo-fascista e poi, dagli anni Sessanta, come scelta di militanza politica di sinistra, è venuta meno, portandola a diventare una meta prediletta in primo luogo dalle studentesse venete neodiplomate. Al di là di quanto esposto nel presente scritto, in chiusura si vuole ribadire l'importanza del database conte-

97. Vincenzo Milanese, *Sette anni al Bo*, in Lazzaretto e Simone, *Dall'università d'élite all'università di massa*, p. 237.

98. Considerando invece tutti i laureati nel periodo in esame (22.114 esami di laurea), il voto medio conseguito è stato di 100,6/110, con una forte disparità tra maschi (voto medio 99,5) e femmine (voto medio 106,4). Se si prende come voto di laurea il 110 e lode, conseguito in 2.374 occasioni, ad ottenerlo sono state in larga maggioranza donne, il 58,2% del totale.

99. 95,3/110 in Economia territoriale e reti d'impresa; 96,4/110 in Diritto dell'economia; 97,8/110 in Economia internazionale; 98,7/110 in Governo delle amministrazioni. I laureati in scienze sociologiche hanno conseguito invece una media del 100,2; quelli in Scienze politiche e relazioni internazionali del 101,4 ed infine quelli in Politica ed integrazione europea del 102,9. Proprio gli ultimi due sono i corsi che hanno attirato più iscritti e sfornato più laureati.

nente l'anagrafe dei laureati (non solo in Scienze politiche) presso l'Ateneo di Padova, uno strumento che si rivelerà essere fecondo di nuove piste di ricerca. Si pensi – solo per avanzare qualche ipotesi – alla correlazione tra crescita economica del Veneto e numero di immatricolati¹⁰⁰; al tema del diritto allo studio e alle scelte culturali dei giovani veneti; all'influenza della globalizzazione nell'indirizzare i "gusti" degli studenti verso un percorso triennale piuttosto che un altro; alla capacità dell'Ateneo di Padova di attirare "clienti" grazie a mirate politiche di *marketing*, sottraendoli ad altre realtà accademiche della regione, come Venezia e Verona; alla comparazione tra le curve della popolazione studentesca di Scienze politiche e quella di Giurisprudenza; e altro ancora.

100. Si veda Mario Pomini, *La crisi e l'economia veneta*, in Agostini e Silvano, *Il Veneto dopo il Novecento*, pp. 201-213.

Pag. 4 del Registro



N. _____

R. UNIVERSITÀ DEGLI STUDI DI PADOVA

Scuola di Scienze Politiche e Sociali

PROCESSO VERBALE PER ESAME DI LAUREA in Scienze Politiche

Essendo stato accertato che il Sig. Agustino Ettore figlio
di Antonio nato a Cereguano Provincia di _____

percorso lo studio di **Scienze Politiche** negli anni scolastici 1924-25 1925-26 1926-27
anche laureato in legge
e superò gli esami di profitto su tutte le materie d'obbligo come segue:

INDICAZIONE DELLE MATERIE	Punti ottenuti su trenta	INDICAZIONE DELLE MATERIE	Punti ottenuti su trenta
MATERIE OBBLIGATORIE		16. Procedura civile	
1. Istituzioni di Diritto privato		17. Storia delle dottrine politiche e scienza politica generale	27
2. Diritto costituzionale		18. Scienza politica dell'amministrazione nei rapporti interni ed esteri	
3. Diritto ecclesiastico		19. Costituzioni straniere	30
4. Diritto e procedura penale		20. Storia delle relazioni internazionali	28
5. Istituzioni politiche dell' antichità classica		21. Sociologia	29
6. Istituzioni politiche dell' Evo medio e moderno		22. Ragioneria delle amministrazioni pubbliche e private	30
7. Economia politica		MATERIE FACOLTATIVE	
8. Scienza delle Finanze		1. <u>Metin 29</u>	
9. Demografia		2. _____	
10. Diritto amministrativo		3. _____	
11. Diritto coloniale	30	4. _____	
12. Diritto finanziario		5. _____	
13. Diritto internazionale pubbl. e priv.		6. _____	
14. Diritto civile		7. _____	
15. Diritto commerciale			

fu ammesso all' esame di laurea a norma delle vigenti disposizioni.

Presentatosi il giorno 12 luglio 1926 nell'Aula _____ ai sottoscritti membri della Commissione esaminatrice, sostenne la disputa intorno alla dissertazione scritta liberamente da lui sopra il tema da lui scelto La costituzione politica della Repubblica slovacca

ed intorno alle tesi parimente da lui scelte sulle materie seguenti :

1. Diritto civile
2. Diritto amministrativo
3. _____

Procedutosi alla votazione l'esame fu classificato con punti centodieci sopra centotrenta e il Candidato fu dichiarato promosso nell'esame di Laurea in **Scienze Politiche**.

Padova, addì 12 luglio 1926

LA COMMISSIONE

1. Lebonati Presidente
2. Lando Landucci
3. Menotamagna
4. Mericala Pellani
5. Menziani
6. Borsi
7. Alf. Pava
8. A. Pava
9. Giampietrangeli
10. Spadaro
11. Peloni

Pag. 160 del Registro



N. _____

R. UNIVERSITÀ DEGLI STUDI DI PADOVA

Scuola di Scienze Politiche e Sociali

**PROCESSO VERBALE PER ESAME DI LAUREA
in Scienze Politiche**

Essendo stato accertato che Luigi Sigma Bertolini Paola figlio
di Francesco natq a Conegliano Provincia di Treviso
percorso lo studio di **Scienze Politiche** negli anni scolastici
e superò gli esami di profitto su tutte le materie d'obbligo come segue:

INDICAZIONE DELLE MATERIE	Punti ottenuti su trenta	INDICAZIONE DELLE MATERIE	Punti ottenuti su trenta
MATERIE DELLA SCUOLA		15. Diritto del lavoro	<i>in esito</i>
1. Storia delle dottrine politiche e scienza politica generale	<i>23</i>	16. Economia e legislazione bancaria	
2. Scienza dell'amministrazione		17. Economia e legislazione agraria	<i>24</i>
3. Costituzioni straniere	<i>25</i>	18. Economia e legislazione dei mezzi di comunicazione e di trasporto	<i>24,70</i>
4. Diritto coloniale	<i>26</i>	19. Diritto militare	<i>27</i>
5. Istituzioni politiche dell'antichità classica	<i>22</i>	MATERIE DI ALTRE FACOLTÀ	
6. Istituzioni politiche dell'Evo medio e moderno	<i>22</i>	1. <i>Diritto costituzionale</i>	<i>27</i>
7. Storia delle relazioni internazionali	<i>24</i>	2. <i>Lezioni di diritto privato</i>	<i>25</i>
8. Matematica per le scienze sociali		3. <i>Statistica metodologica</i>	<i>25</i>
9. Biologia per le scienze sociali	<i>30</i>	4. <i>Economia pubblica</i>	<i>28</i>
10. Demografia	<i>27</i>	5. <i>Scienze finanziarie</i>	<i>27</i>
11. Storia economica	<i>25</i>	6. <i>Diritto e procedura penale</i>	<i>30</i>
12. Politica economica		7. <i>Diritto internazionale</i>	<i>23</i>
13. Statistica economica	<i>20</i>	8. <i>Scienze delle Finanze</i>	<i>18</i>
14. Geografia politica ed economica	<i>25</i>	9. <i>Filosofia del diritto</i>	<i>22</i>
<i>Medicina legale</i>	<i>25</i>	10. <i>Diritto amministrativo</i>	<i>18</i>
		<i>Diritto commerciale</i>	<i>22</i>

fu ammesso all'esame di laurea a norma delle vigenti disposizioni.

Presentatosi il giorno 1. XII. 22 nell'Aula _____ ai sottoscritti membri della Commissione esaminatrice, sostenne la disputa intorno alla dissertazione scritta liberamente da lui, sopra il tema da lui scelto

La dinamica dei prezzi all'ingrosso nel mercato di Padova

ed intorno alle tesi parimenti da lui scelte sulle materie seguenti:

1. Biologia
2. Economia politica
3. _____

Procedutosi alla votazione l'esame fu classificato con punti centoquattro sopra centodieci e il Candidato fu dichiarato promosso nell'esame di Laurea in **Scienze Politiche**.

Padova, addì 1. XII 192XI

LA COMMISSIONE

- | | |
|--------------------------------|---------------------------|
| 1. <u>Luigi Bonatti</u> | Presidente |
| 2. <u>M. Fasano</u> | 7. <u>C. F. F. F.</u> |
| 3. <u>Luigi Catellani</u> | 8. <u>Alvaro</u> |
| 4. <u>Alto Ferrarino</u> | 9. <u>Paolo Ecc. ecc.</u> |
| 5. <u>G. Salvi</u> | 10. <u>Alvaro</u> |
| 6. <u>Antonio M. Bettinini</u> | 11. <u>A. Finzi</u> |

Lpo

Studente protetto da Norma
Laurea
condizionata

Matricola N° 98/9

REGISTRATO

R. UNIVERSITÀ DEGLI STUDI DI PADOVA

FACOLTÀ DI SCIENZE POLITICHE 18

Processo verbale per esame di laurea

Il Sig. Scotto Guido figlio di Su Giuseppe
 nato a Costanzano provincia di ha frequentato i corsi della
 facoltà per il numero di anni prescritto dallo Statuto Universitario ed ha superato gli esami di profitto
 sulle materie d'obbligo come segue: Media = 28,30 = 103,96

INDICAZIONE DELLE MATERIE	Punti ottenuti su trenta	INDICAZIONE DELLE MATERIE	Punti ottenuti su trenta
Insegnamenti Fondamentali		Lingua <u>Polizara</u>	<u>28</u>
Dottrina dello Stato	<u>30 e lode</u>	Insegnamenti complementari	
Istituzioni di diritto privato	<u>24</u>	Filosofia del diritto	
Istituzioni di diritto pubblico	<u>(1)</u>	Legislazione del lavoro	
Diritto amministrativo (biennale)	<u>30</u>	Demografia generale e demografia comparata della razza	
Diritto internazionale	<u>26</u>	Sociologia	
Diritto corporativo	<u>27</u>	Diritto pubblico romano	<u>30</u>
Diritto costituzionale italiano e comparato	<u>30</u>	Contabilità di Stato	<u>30 e lode</u>
Storia moderna (biennale)	<u>30 e lode</u>	Storia delle dottrine economiche	<u>30</u>
Storia e dottrina del Fascismo	<u>28</u>	Economia coloniale	
Storia delle dottrine politiche	<u>26</u>	Geografia ed etnografia coloniale	
Storia politica coloniale	<u>30</u>	<u>Diritto pubblico italiano</u>	<u>(1) 27</u>
Storia dei trattati e politica internazionale	<u>30</u>	<u>Storia e politica navale</u>	<u>30</u>
Geografia politica ed economica	<u>27</u>		
Economia politica corporativa	<u>30</u>		
Politica economica e finanziaria	<u>25</u>		
Statistica	<u>23</u>		
Cultura militare	<u>30 e lode</u>		
Lingua <u>francese</u>	<u>30 e lode</u>		

È stato quindi ammesso all'esame di laurea a norma delle vigenti disposizioni.

Presentatosi il giorno 28 febbraio nell'Aula Giustiniano
ai sottoscritti Membri della Commissione esaminatrice, sostenne la disputa intorno alla dissertazione scritta
sopra il tema da lui scelto: "Quale carattere ebbe la critica del froel
al marxismo"

ed intorno ai temi orali, sulle materie seguenti.

- 1 Diritto amministrativo
- 2 Storia delle dottrine economiche
- 3 _____

Procedutosi alla votazione, l'esame fu classificato con punti Centotré
sopra centodieci ed il candidato fu dichiarato approvato
nell'esame di laurea in **Scienze Politiche**, otto condizioni.

Padova, addì 28 febbraio 1945 - X

LA COMMISSIONE

- | | |
|--|-------------------------------|
| 1° <u>Anton M. Rikamin</u> ✓, presidente | |
| 2° <u>Albert Turchetti</u> ✓ | 7° <u>Giuseppe Indurcin</u> ✓ |
| 3° <u>Checcini</u> ✓ | 8° <u>Caldarone</u> ✓ |
| 4° <u>Giuseppe ...</u> ✓ | 9° <u>Bentile</u> ✓ |
| 5° <u>F. D. ...</u> ✓ | 10° <u>... Laurani</u> ✓ |
| 6° <u>Amisbram</u> ✓ | 11° <u>Antonio Checcini</u> ✓ |

Matricola N° 4.10

R. UNIVERSITÀ DEGLI STUDI DI PADOVA

FACOLTÀ DI SCIENZE POLITICHE

f.c.

Processo verbale per esame di laurea

Il Sig. D'Arbela Serena ^{25,47 = 93,39} figlio di Felice
nato a Firenze ^{provincia di} 30.6.30 ha frequentato i corsi della
facoltà per il numero di anni prescritto dallo Statuto Universitario ed ha superato gli esami di profitto
sulle materie d'obbligo come segue:

INDICAZIONE DELLE MATERIE	Punti ottenuti su trenta	INDICAZIONE DELLE MATERIE	Punti ottenuti su trenta
<u>Insegnamenti Fondamentali</u>		Lingua <u>inglese</u>	24 [✓]
Dottrina dello Stato	24 [✓]	<u>Insegnamenti complementari</u>	
Istituzioni di diritto privato	27 [✓]	Filosofia del diritto	25 [✓]
Istituzioni di diritto pubblico	27 [✓]	Legislazione del lavoro	
Diritto amministrativo (biennale)	27 [✓]	Demografia generale e demografia comparata della razza	
Diritto internazionale	27 [✓]	Sociologia	
Diritto corporativo <u>del lavoro</u>	25 [✓]	Diritto pubblico romano	26 [✓]
Diritto costituzionale italiano e comparato	26 [✓]	Contabilità di Stato	
Storia moderna (biennale)	26 [✓]	Storia delle dottrine economiche	26 [✓]
Storia e dottrina del Fascismo	—	Economia coloniale	27 [✓]
Storia delle dottrine politiche	25 [✓]	Geografia ed etnografia coloniale	
Storia politica coloniale	22 [✓]		
Storia dei trattati e politica internazionale	27 [✓]		
Geografia politica ed economica	25 [✓]		
Economia politica corporativa	26 [✓]		
Politica economica e finanziaria	25 [✓]		
Statistica	27 [✓]		
Cultura militare	—		
Lingua <u>francese</u>	27 [✓]		

È stato quindi ammesso all'esame di laurea a norma delle vigenti disposizioni.

Presentatosi il giorno 9 luglio 1953 nell' Aula Jaccolli
ai sottoscritti Membri della Commissione esaminatrice, sostenne la disputa intorno alla dissertazione scritta
sopra il tema da lui scelto: Lo sviluppo del Porto Marghera

ed intorno ai temi orali, sulle materie seguenti.

- Diritto amministrativo
- Politica economica
- /

Procedutosi alla votazione, l'esame fu classificato con punti 97/ novantasette.
sopra Centodici ed il candidato fu dichiarato approvato
nell'esame di laurea in **Scienze Politiche.**

Padova, addì 9 luglio 1953 - X

Luca Maria Beckarini Presidente ✓
LA COMMISSIONE

1° <u>Chechin</u> ✓	presidente
2° <u>W...</u> ✓	7° <u>G. P...</u> ✓
3° <u>...</u> ✓	8° <u>...</u> ✓
4° <u>...</u> ✓	9° <u>Enide Lucatello</u> ✓
5° <u>Leopoldo...</u> ✓	10° <u>...</u> ✓
6° <u>Mario F...</u> ✓	11° <u>...</u> ✓

TIP. L. SCUDIER - 14004

Abstract

My essay aims to begin to reflect on the available data about Paduan Political science graduates, the eponymous Faculty being one of the oldest and most prestigious ones in Italy. It was founded in 1924 as a School of Political and Social Sciences; it became a Faculty in 1933, only to be turned into a Department following the so-called Gelmini Reform in 2010. Through a wealth of archival sources (especially viva voce exam transcripts) I have been able to reconstruct a prosopography of Political science graduates whereby to detail the transformation of the Faculty in the 1960s from an elite into a popular institution. One can get an insight for each graduate about the following data: name; and, whenever possible, date ad place of birth; sex; which undergraduate and postgraduate courses they attended; degree date and grade; dissertation title. My essay thus contributes to reconstruct the Faculty's internal workings at a time of important historical and political upheavels.

*Indice dei nomi**

- Abramitzky Ran 109, 113,124
Ago Renata 118
Agostini Carlo 147
Agostini Filiberto 5, 7, 143, 145, 151, 159,
164, 170,172
Aldrin Buzz 95
Alerini Julien 137
Aliverti Emanuele 63, 65,66
Almagisti Marco 164
Anchieri Ettore 149
Anderson Chris 54, 66
Andreozzi Daniele 108, 119, 123,143
Angwin Julia 63, 66
Ankeny Rachel A. 72, 74
Anti Carlo 146 148
Antiseri Dario 144
Antonfrancesco dei Dottori 131
Antonio Brancio dei Luschi 131
Arangio Ruiz Vincenzo 148
Armitage David 110, 111, 126
Arnaldi Girolamo 128
Avenarius Richard 11
Azzalini Adelchi 51, 66
- Backovic Mihailo 86, 93
Barnett Vic 50, 53, 66
Barth Fredrik 118
Bassanini Franco 153
Bassi Giulia 110
Bellavitis Anna 121
Benzer Seymour 103
- Berg Hildebrand Andreas 35
Berti Enrico 145
Berti Giampietro 145, 146
Bertoncin Marina 120, 123, 125
Bettanini Anton Maria 147, 149
Biasiori Lucio 108
Billanovich Liliana 145
Bobbio Norberto 146
Bogen James 70, 74
Boldi Paolo 39
Borgman Christine L. 72, 74
Braudel Fernand 111
Breccia Alessandro 150
Breiman Leo 54, 55, 66
Bühlmann Peter 50, 66
- Calafat Guillaume 123
Caldognetto Paola 150
Callebaut Werner 70, 74
Canali Stefano 73, 74
Candès Emmanuel 53, 66
Capodivacca Bartolomeo 131
Caracausi Andrea 5, 108, 117, 118, 121
Carnap Rudolf 5, 11, 12, 13, 21, 22, 23, 24,
25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35
Carus Andrè 117, 119, 120
Cerutti Simona 116, 117, 118, 120, 121
Christian David 112, 123, 125
Clemente Alida 108
Cleveland William S. 52, 66
Codignola Tristano 150

* Indice a cura di Franca Cosmai.

INDICE DEI NOMI

- Cohen Deborah 110
 Contarini Leonardo 131
 Corbellini Gilberto 109, 110
 Corbett Davies Sam 62, 66, 67
 Coveney Peter V. 8
 Cox David R. 54, 66, 67
 Crescente Cesarino 149
- Da Carrara, famiglia 128
 Dalla Chiara Maria Luisa 33
 Darwin Charles 100
 Davison Anthony C. 53, 66
 De Gasperi Alcide 149
 Del Re Alisa 159
 De Ruggiero Guido 148
 De Vito Christian G. 123
 Dean Jeffrey 43
 Debernardi Massimo 16
 Del Negro Piero 128, 129, 144, 145, 146, 148
 Desrosières Alain 121
 Diamanti Ivo 164
 Dietrich Michael R. 72, 74
 Donati Donato 146, 147
 Dougherty Edward R. 8
 Dunson David B. 56, 66
 Dupré John 70, 71, 74
 Duroselle Jean Baptiste 166
- Ehmer Josef 117, 121
 Einstein Albert 11, 83
 Elliott Kevin C. 73, 74
 Evan Roberts 113
 Ezzelino da Romano 128
- Fanno Marco 147
 Favero Giovanni 108, 116, 118, 121, 124, 126
 Ferraris Luigi Vittorio 166
 Fioravanzo Monica 149
 Fisher Ronald Aylmer 52, 53, 66, 67
 Floridi Luciano 71, 74
 Fornasier Roberto 151
 Franceschini Ezio 148
 Franz Nico M. 73, 75
 Frege Gottlob 11, 23
 Friedler, sorelle A. 62, 66
 Friedman Jerome H. 50, 51, 66
 Frijhoff Willem 140
- Galilei Galileo 8, 129
 Gallo Donato 128, 130
- Gantz John 36
 Gelmini Mariastella 144, 181
 Gemelli Agostino 149
 Genet Jean-Philippe 137
 Gentile Giovanni 146
 Gentiloni Silveri Umberto 151
 Ghemawat Sanjay 43
 Ghetti Maria Cecilia 145, 183
 Ghobrial John-Paul A. 123
 Giaretta Paolo 149 164
 Giaretta Pierdaniele 5, 7, 11
 Gini Corrado 146
 Ginzburg Carlo 116, 118, 120, 122
 Gloria Andrea 131
 Goeman Jelle J. 52, 55, 67
 Goetz Helmut 147
 Gola Giuseppe 148
 Goodman Nelson 22
 Gori Pietro 16, 21
 Grangaud Isabelle 116, 117, 120
 Green Brian 83, 93
 Grendi Edoardo 116, 122
 Gribaudi Maurizio 125
 Griesemer James 73, 75
 Griffiths Paul 70, 74
 Gubler Kaspar 141
 Guldi Jo 110, 111, 126
 Gutmann Myron 108, 112, 113, 115, 122, 126
 Gylfi Magnússon Sigurdur 124
- Hafen Ryan 52, 66
 Hastie Trevor 50, 52, 56, 58, 66, 67
 Hawking Stephen 83, 93
 Heisenberg Martin 103, 105, 107
 Hesse Christian 106, 141
 Hey Tony 54, 67
 Higgs Peter 8, 83, 84, 85, 86, 90, 93
 Highfield Roger R. 8
 Hodgkin Alan Lloyd 98
 Hofmeester Karin 115
 Hoxha Enver 166
 Huxley Andrew Fielding 98
- Illari Phyllis 71, 74
 Isaac Williams 63, 67
 Isnenghi Mario 147
- Jori Francesco 149, 151

INDICE DEI NOMI

- Kant Immanuel 31
 Karila-Cohen Karine 108
 Karinthy Frigyes 39
 Kellert Stephen H. 69, 74
 Kessels Geert 130
 Kitchin Roy 54, 67
 Klancher Merchant Emily 113, 122, 126
- Lamassé Stéphane 137
 Lämmel Ralf 52, 67
 Lanaro Silvio 147
 Lazzaretto Alba 145, 147, 148, 149, 150, 151, 167, 171
 Le Roy Ladourie Emmanuel 113
 Lemercier Claire 108, 110, 115, 119, 120, 121, 122
 Lenci Giuliano 148
 Leonardo (da Vinci) 65
 Leonelli Sabina 5, 68, 69, 70, 71, 72, 74, 75
 Leskovec Jure 52, 67
 Levi Giovanni 116, 117, 123, 125, 126, 127
 Levins Richard 71, 75
 Levitis Daniel A. 97, 107
 Loettgers Andrea 72, 75
 Lomellini Valentine 151
 Longino Helen E. 71, 74, 75
 Lucassen Jan 115, 117
 Lucassen Leo 115, 117
 Lum Kristian 63, 66, 67
- Mach Ernst 5, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 28, 29, 31, 32, 33, 34, 35
 Magnuson Diana L. 112, 114, 115
 Mandler Peter 110
 Manning Patrick 112
 Marangon Paolo 128
 Marchesi Concetto 148
 Maréchaux Benoît 108
 Markram Henry 96
 Martellozzo Forin Elda 131
 Martini Andrea 130
 Martini Manuela 121
 Mazzini Federico 108
 McAllister James W. 72, 75
 Meng Xiao Li 60, 67
 Merigliano Luciano 151
 Michele da Marostica 131
 Milanese Vincenzo 171
 Mitchell Sandra D. 71, 75
- Mitchener Kris James 108, 113
 Mocarelli Luca 108, 109, 114, 117
 Moll-Murata Christine 115
 Moro Aldo 151
 Müller-Wille Staffan 70, 75
 Munroe Randall 81, 93
- Nani Michele 108, 110, 114, 143
 Newman Maxwell Herman Alexander 33
 Newton Isaac 8, 82
 Niccolò di Matteo Lorenzo da Padova 138
 Nickles Thomas 72, 75
- O'Malley Maureen A. 69, 75
 Ogilvie Sheilagh 117, 119, 120
 Opocher Enrico 150
- Pace Luigi 53, 67
 Paolo Biancuzzi da Schio 131
 Parmigiani Giovanni 52, 67
 Pase Andrea 108, 120, 123, 125
 Pasquinelli Alberto 28
 Perin Lorenza 159
 Pesenti Tiziana 128
 Petter Guido 151
 Piccoli Giuliano 167
 Pietro d'Abano 129
 Pietsch Wolfgang 73, 75
 Piovan Francesco 128, 144, 145, 148
 Planck Max 86
 Poni Carlo 118, 122
 Popper Karl 53, 67
 Poppi Antonino 144
 Putnam Hilary 33
- Quaresima Giulio 140
 Quatrada Daria 120, 123, 125
 Quazza Guido 146
 Quine Willard van Orman 25
- Ravà Adolfo 146, 147
 Reale Giovanni 144
 Reid Nancy 54, 67
 Reinsel David 36
 Rheinberger Hans Jörg 70, 75
 Riccamboni Gianni 164
 Robinson David 52, 67
 Rocco Alfredo 146
 Romain Bertrand 123

INDICE DEI NOMI

- Romeijn Jan Willem 53, 67
 Rosa Marco 24, 39
 Rosé Isabelle 108, 121, 122
 Rosenberg Alexander 109
 Rudovsky David 63, 67
 Ruggles Steven 112, 114, 115
 Russell Bertrand 23, 30, 33
 Rydning John 36
- Salvan Alessandra 5, 50, 53, 67
 Saonara Chiara 148
 Sarfatti Michele 147
 Sarti Raffaella 121
 Scarnera Nicola 164
 Scarpa Bruno 5, 50, 51, 66
 Schlick Moritz 11
 Schlumbohm Jürgen 125
 Schmidt Eric 37
 Schwalbe Michelle 57, 67
 Segato Giorgio 148
 Senn Stephen 50, 67
 Severino Emanuele 22
 Sewell William H. 114, 119, 120, 122, 125
 Shavit Ayelet 73, 75
 Silge Julia 52 67
 Silvano Giovanni 159, 172
 Simoiu Camelia 63, 67
 Simone Giulia 144, 145, 146, 147, 148, 149,
 150, 151, 162, 167, 168, 170, 171
 Siraisi Nancy G. 128
 Sitran Rea Luciana 145, 167
 Smolin Lee 87, 93
 Solari Aldo 52, 55, 67
 Solera Dennj 108, 130
 Sosio Libero 14
 Soyer Orkun S. 69, 75
 Spanos Aris 53, 67
 Spier Fred 112
 Sterner Beckett 73, 75
- Stigler George Joseph 52
 Stigler Stephen 67
 Stone Lawrence 113, 114, 115
 Stotz Karola 70, 74
 Suhm Christian 35
- Tempini Niccolò 72, 75
 Terenzi Pierluigi 5, 130
 Tibshirani Robert 52, 67
 Tilly Charles 125
 Torre Angelo 117, 118, 123
 Trivellato Francesca 110, 117
 Trump Ronald 60, 62
- van Bree Pim 130
 van de Geer Sara 50, 66
 van Fraassen Bas C. (Bastiaan Cornelis) 33,
 34, 35
 van Nederveen Meerkerk Elise 115
 Varsori Antonio 144
 Ventura Angelo 144, 145, 146, 148, 149, 150,
 151
 Vigen Tyler 79, 80, 94
 Vigna Sebastiano 39
 Vittorio Emauele II , re d'Italia 145
 Vries Jan 114, 123, 125
- Wainwright Martin 52, 67
 Waters C. Kenneth 70, 74, 75
 Wimsatt William C. 71, 76
 Woodward James 72, 76
 Wylie Alison 71, 76
- Zaccaria Benedetto 148, 167, 170
 Zalc Claire 108, 119, 120, 121, 122
 Zornetta Giulia 130
 Zorzi Andrea 137
 Zucchini Stefania 140

CONTRIBUTI ALLA STORIA DELL'UNIVERSITÀ DI PADOVA

Presso la Editrice Antenore - Padova

1. *Relazioni tra Padova e la Polonia. Studi in onore dell'Università di Cracovia nel VI centenario della sua fondazione.* 1964.
2. F. Lucchetta, *Il medico e filosofo bellunese Andrea Alpago († 1522) traduttore di Avicenna.* 1964.
- 3-4. A. Favaro, *Galileo Galilei e lo Studio di Padova.* 1966. (I due volumi inseparabili)
5. A. Favaro, *Galileo Galilei a Padova.* 1968.
6. L. Gargan, *Lo studio teologico e la biblioteca dei domenicani a Padova nel Tre e Quattrocento.* 1971.
7. A. Sottili, *Studenti tedeschi e Umanesimo italiano nell'Università di Padova durante il Quattrocento. I. Pietro Del Monte nella società accademica padovana (1430-1433).* 1971.
8. G. Brunetta, *Gli inizi dell'insegnamento pubblico dell'architettura a Padova e a Venezia. Cronaca e storia.* 1976.
9. F. Franceschini, *Concetto Marchesi. Linee per l'interpretazione di un uomo inquieto.* 1978.
10. F.L. Maschietto, *Elena Lucrezia Cornaro Piscopia (1464-1684), prima donna laureata nel mondo.* 1978.
11. T. Pesenti Marangon, *La Biblioteca Universitaria di Padova, dalla sua istituzione alla fine della Repubblica Veneta (1629-1797).* 1979.

Presso la Editrice Lint - Trieste

12. R. Palmer, *The 'Studio' of Venice and its graduates in the sixteenth century.* 1983.
13. F. De Vivo, *L'insegnamento della pedagogia nell'Università di Padova durante il XIX secolo.* 1983.
14. *Il «Catalogo dei libri» di Giambattista Morgagni, a cura di E. Barile e R. Suriano.* 1983.
15. *Scienza e filosofia all'Università di Padova nel Quattrocento, a cura di A. Poppi.* 1983.
16. T. Pesenti, *Professori e promotori di medicina nello Studio di Padova dal 1405 al 1509. Repertorio bio-bibliografico.* 1984.
17. F. Piovan, *Per la biografia di Lazzaro Bonamico. Ricerche sul periodo dell'insegnamento padovano (1530-1552).* 1988.
18. A. Gambasin, *«Theses» in sacra teologia nell'Università di Padova dal 1815 al 1873.* 1984.
19. G.A. Salandin - M. Pancino, *Il «teatro» di filosofia sperimentale di Giovanni Poleni.* 1987.
20. *Rapporti tra le Università di Padova e Bologna. Ricerche di filosofia, medicina e scienza, a cura di L. Rossetti.* 1988.

21. M.L. Soppelsa, *Leibniz e Newton in Italia. Il dibattito padovano (1687-1750)*. 1989.
22. A. Robinet, *L'empire leibnizien, La conquête de la chaire de mathématiques de l'Université de Padoue: Jakob Hermann et Nicolas Bernoulli (1707-1719)*. 1991.
23. A. Favaro, *Scampoli galileiani*, vol. I, a cura di L. Rossetti e M.L. Soppelsa. 1992.
24. A. Favaro, *Scampoli galileiani*, vol. II, a cura di L. Rossetti e M.L. Soppelsa. 1992.
25. A. Favaro, *Adversaria galilæiana*, vol. III, a cura di L. Rossetti e M.L. Soppelsa. 1992.
26. Carlo Anti, *Giornate di studio nel centenario della nascita. Verona-Padova-Venezia, 6-8 marzo 1990*. 1992.
27. M. Minesso, *Tecnici e modernizzazione nel Veneto. La Scuola dell'Università di Padova e la professione dell'ingegnere (1806-1915)*. 1992.
28. M.R. Davi, *Bernardino Tomitano filosofo, medico e letterato (1517-1576). Profilo biografico e critico*. 1995.
29. A. Veggetti - B. Cozzi, *La Scuola di Medicina veterinaria dell'Università di Padova*. 1996.
30. *La storia delle Università italiane. Archivi, fonti, indirizzi di ricerca. Atti del Convegno. Padova, 27-29 ottobre 1994*, a cura di L. Sitran Rea. 1996.
31. P. Marangon, *'Ad cognitionem scientiae festinare'. Gli studi nell'Università e nei conventi di Padova nei secoli XIII-XIV*, a cura di T. Pesenti. 1996.
32. *Istituzioni culturali, scienza, insegnamento nel Veneto dall'età delle riforme alla restaurazione (1761-1818). Atti del convegno, Padova, 28-29 maggio 1998*, a cura di L. Sitran Rea. 2000.
34. *Studenti, Università, città nella storia padovana. Atti del Convegno di studi, Padova 5-8 febbraio 1998*, a cura di F. Piovan - L. Sitran Rea. 2001.
37. *Le 'carte del filosofo'. Il fondo 'R. Ardigò' della biblioteca universitaria di Padova*, a cura di G.P. Mantovani. 2004.

Presso la Tipografia Bertoncetto Artigrafiche - Cittadella

33. *Giuseppe Toaldo e il suo tempo nel bicentenario della morte. Scienza e lumi tra Veneto ed Europa. Atti del Convegno di studi, Padova 10-13 novembre 1997*, a cura di L. Pigatto. 2000.

Presso Edizioni Antilia - Treviso

35. T. Pesenti, *Marsilio Santasofia tra corti e Università. La carriera di un «monarcha medicinae» del Trecento*. 2003.
36. A. Poppi, *Statuti dell'«Universitas theologorum» dello Studio di Padova (1385-1784)*. 2004.
38. *Studenti istriani e fiumani all'Università di Padova dal 1601 al 1974*, a cura di L. Sitran Rea - G. Piccoli. 2004.

39. *Harvey e Padova. Atti del convegno celebrativo del quarto centenario della laurea di William Harvey, Padova 21-22 novembre 2002*, a cura di G. Ongaro - M. Ripa Bonati - G. Thiene. 2006.
40. *Insequimini archivum. Atti della giornata di studio in memoria di Paolo Sambin, Padova 19 novembre 2004*, a cura di F. Piovan. 2007.
41. F. Benucci, *Stemmi di scolari dello Studio Patavino fuori delle sedi universitarie*. 2007.
42. G. Ongaro, *Wirsung a Padova, 1629-1643*. 2010.
43. *Il fondo Marsili nella biblioteca dell'Orto Botanico di Padova*, a cura di A. Minelli - A. Angarano - P. Mario. 2010.
44. J. Benavent - M.J. Bertomeu Masiá, *La familia Granvela en el Estudio de Padua. Edición de documentos inéditos*. 2011.
45. G. Berti, *L'Università di Padova dal 1814 al 1850*. 2011.
46. E. Barile, *Per la biografia dell'umanista Giovanni Marcanova*. 2011.
47. S. Collodo - R. Simonetti, *Filosofia naturale e scienze dell'esperienza fra medioevo e umanesimo*. 2012.
48. G. Gola, *Il mio rettorato (1943-1945)*. 2015.
49. *Il positivismo a Padova tra egemonia e contaminazioni*, a cura di G. Berti - G. Simone. 2016.
50. *Memoria di Paolo Sambin*, a cura di D. Gallo - F. Piovan. 2016.

PROFILI BIOGRAFICI

Presso la Editrice Lint - Trieste

1. *Professori di materie scientifiche all'Università di Padova nell'Ottocento*, a cura di S. Casellato - L. Pigatto. 1996.
2. V. Lazzarini - L. Lazzarini, *Maestri Scolari Amici. Commemorazioni e profili di storici e letterati a Padova e nel Veneto alla fine del l'Ottocento e nel Novecento*, a cura di G. Ronconi - P. Sambin. 1999.

Presso Edizioni Antilia - Treviso

3. *Professori e scienziati a Padova nel Settecento*, a cura di S. Casellato - L. Si-tran Rea. 2003.

Centro per la storia dell'Università di Padova

Ultimi volumi pubblicati:

VITTORIA FEOLA, *Early Modern Universities and the Sciences* (disponibile anche in e-book).

FILIBERTO AGOSTINI (a cura di), *Università e Grande guerra in Europa. Medicina, scienze e diritto* (disponibile anche in e-book).

CARLA LESTANI (a cura di), *Il Fondo Marsili nella Biblioteca universitaria di Padova* (disponibile anche in e-book).

CHIARA MARIA VALSECCHI, FRANCESCO PIOVAN (a cura di), *Diritto, Chiesa e cultura nell'opera di Francesco Zabarella. 1360-1417* (disponibile anche in e-book).

FILIBERTO AGOSTINI (a cura di), *L'Ateneo di Padova nell'Ottocento. Dall'Impero asburgico al Regno d'Italia* (disponibile anche in e-book).

LUCIANA SITRAN REA, *Indici. Quaderni per la storia dell'Università di Padova 1968-2017* (E-book).

GIULIA SIMONE, *La Facoltà Cenerentola. Scienze politiche a Padova dal 1948 al 1968* (disponibile anche in e-book).



Il presente volume è pubblicato in open access, ossia il file dell'intero lavoro è liberamente scaricabile dalla piattaforma **FrancoAngeli Open Access** (<http://bit.ly/francoangeli-oa>).

FrancoAngeli Open Access è la piattaforma per pubblicare articoli e monografie, rispettando gli standard etici e qualitativi e la messa a disposizione dei contenuti ad accesso aperto. Oltre a garantire il deposito nei maggiori archivi e repository internazionali OA, la sua integrazione con tutto il ricco catalogo di riviste e collane FrancoAngeli massimizza la visibilità, favorisce facilità di ricerca per l'utente e possibilità di impatto per l'autore.

Per saperne di più:

http://www.francoangeli.it/come_publicare/publicare_19.asp

I lettori che desiderano informarsi sui libri e le riviste da noi pubblicati possono consultare il nostro sito Internet: www.francoangeli.it e iscriversi nella home page al servizio "Informatemi" per ricevere via e-mail le segnalazioni delle novità.

Lo sviluppo tecnologico ha permesso la memorizzazione di grandi quantità di dati e ha enormemente accresciuto la capacità di elaborarli. È nato un nuovo mondo – quello dei Big Data – che, oltre ad avere un grande impatto sulla nostra vita quotidiana, sta influenzando il progresso delle conoscenze scientifiche in modi del tutto imprevedibili. Questa realtà impone ulteriori analisi e riflessioni sui temi dell'apprendere dai dati e della loro relazione con le conoscenze di origine teorica. Dato il rilevante impatto sia in ambiti scientifici in senso stretto sia in quelli sociali e umanistici, è indispensabile un approccio multidisciplinare, come espresso nel presente volume, per riconoscere il ruolo dei Big Data nella nostra cultura e gli effetti specifici nei diversi campi. Un riflesso importante di questa “rivoluzione tecnologica” è evidente anche nelle discipline storiche, in particolare per quanto riguarda il rapporto tra storia e dato, metodo quantitativo e qualitativo, microanalisi e generalizzazioni, intelligenza umana e capacità analitica dei computer di fronte a una massa enorme di dati.

Filiberto Agostini è docente di Storia della Pubblica amministrazione presso l'Università di Padova. Ha diretto il Centro per la storia dell'Università (2016-2020).

Pierdaniele Giaretta è stato docente di Logica e Filosofia della scienza, presidente della Società Italiana di Filosofia Analitica (Sifa), direttore del Centro Interdipartimentale di ricerca in Storia e Filosofia della scienza (Cisfis) dell'Università di Padova.

Giorgio Moro è docente di Chimica Fisica presso l'Università di Padova. Attualmente ricopre il ruolo di Direttore del Centro Interdipartimentale di ricerca in Storia e Filosofia della Scienza (Cisfis).

Giovanni Silvano è docente di Storia moderna presso il Dipartimento di Scienze Storiche, Geografiche e dell'Antichità dell'Università di Padova. Dirige il Centro interdipartimentale di ricerca in Storia della Medicina.