

Metriche citazionali e mutazioni strategiche della ricerca scientifica: narrative ed evidenze

Luca Guerra
[orcid: 0000-0002-8903-7744]

Ancora a lei,
Eleonora

Abstract

A seguito della diffusione anche all'interno delle università del modello gestionale promosso dal New public management, settori fino ad allora dotati di una particolare protezione, come quello accademico, si sono progressivamente visti sottoposti a nuovi controlli e vincoli, che hanno progressivamente assunto la forma di rilevazioni quantitative, con un ruolo crescente delle metriche citazionali. Questi processi di valutazione hanno suscitato diverse e importanti prese di posizione critiche sul piano internazionale.

Se è vero che, secondo la "legge" di Goodhart "quando una misura diviene un obiettivo smette di essere una buona misura, ancora più significativo è rilevare che quando una misura diviene un obiettivo, il misurato smette di essere quello che era prima. La misurazione quantitativa delle performance accademiche ha innescato infatti forme di *gaming* tali da alterare il gioco stesso della ricerca scientifica, dei suoi scopi e delle forme della sua condivisione. Nell'articolo svolgeremo un confronto tra le forme della mutazione e le narrative che le accompagnano per vedere fino a quale punto siamo legittimati oggi a parlare della trasformazione della ricerca scientifica in *ricerca scientifica strategica*.

Insorgenze

Da alcuni anni assistiamo a insorgenze planetarie avverso le forme della misurazione della ricerca, dei ricercatori e le correlate politiche di assegnazione dei fondi, di promozione nelle carriere accademiche e delle abilitazioni scientifiche. Queste mobilitazioni sono cresciute in particolare in seguito all'affermazione del new public management che, dall'inizio degli anni Novanta, dopo un contesto di crisi economica, ha cominciato a investire le prassi manageriali delle pubbliche amministrazioni e ha posto al proprio centro il tema dell'*accountability*. Proprio attorno a questa esigenza di valutazione e misurazione, da qualche tempo entrata al cuore delle università, sono sorti fenomeni impreveduti, controfinalità e conflitti. Le forme del disagio sono molto cresciute negli ultimi anni all'interno della comunità dei ricercatori e hanno portato per l'appunto a forme organizzate di critica e di opposizione.

Una delle più importanti e influenti prese di posizione in particolare contro la valutazione centrata sugli indici bibliometrici è quella conosciuta come dichiarazione DORA, ovvero *la San Francisco Declaration on research assessment*, proposta durante il convegno annuale della Società Americana di Biologia Cellulare nel dicembre 2012. Nella dichiarazione si prende specificamente di mira l'*Impact factor* (IF) come fonte di perniciose conseguenze ad ampio raggio nel mondo della ricerca scientifica. In questa dichiarazione non si contesta tanto, su un piano generale, l'uso di metodi quantitativi¹ nella valutazione della ricerca, ma specificamente l'impiego distorto dell'IF, il primo e il più noto indicatore elaborato dall'Institut for scientific information (ISI) nel 1961 e impiegato su base sistematica nel Journal of citation report (JCR) dal 1974. Si tratta effettivamente di un indice per il quale – nota Biagioli – un articolo paradossalmente "nasce valutato", già dotato di impatto, ancor prima che qualcuno l'abbia letto e ammesso che qualcuno mai lo leggerà. Eugene Garfield stesso, fondatore di ISI, aveva ritenuto di assumere una posizione critica² rispetto a come venga mal impiegata la sua creatura. Quell'indice – spiega – l'aveva inventato per aiutare i bibliotecari nella selezione dei titoli di periodici da sottoscrivere. In termini generali – potremmo pure considerare – è plausibile

¹ Si suggerisce per esempio di adottare al posto del solo IF, un set di indicatori numerici quali: impact factor a cinque anni, EigenFactor, SCImago, h-index (combinati con la valutazione della non meglio precisata "influenza su politiche e prassi").

² Per esempio, pur all'interno di una difesa del complessivo lavoro di ISI contro le critiche di due professori tedeschi, nel 1998 Garfield scriveva "The source of much anxiety about Journal Impact Factors comes from their misuse in evaluating individuals, e.g. during the Habilitation process".

Qui il testo integrale della lettera all'editore:

[http://www.garfield.library.upenn.edu/papers/derunfallchirurg_v101\(6\)p413y1998english.html](http://www.garfield.library.upenn.edu/papers/derunfallchirurg_v101(6)p413y1998english.html)

pensare a periodici molto citati come fonti scientifiche di alta qualità anche se questo non ci dà, per converso, la sicurezza che, per esempio, periodici orientati allo sviluppo di nuovi promettenti filoni di ricerca ancora poco frequentati e dunque con basso IF, contengano articoli di scarsa qualità.

Dunque, gli estensori di DORA, dopo avere specificato alcune ragioni di contrarietà al modo in cui viene impiegato l'IF, si rivolgono a cinque interlocutori: Agenzie di finanziamento, Istituzioni, Editori, Organizzazioni che forniscono metriche, Ricercatori affinché non facciano affidamento nelle loro rispettive attività all'IF e invitano tutti a esplorare nuovi indicatori quantitativi, a dare misura di impatto del singolo *articolo* (anziché del suo contenitore) e ad assegnare una maggiore importanza alla valutazione qualitativa. Da quel manifesto è nato il progetto operativo, svolto in collaborazione con il *Centre for Science and Technology Studies (CWTS)* di Leida, una delle più note agenzie di elaborazione di indici, chiamato *Tools to advance research metrics*³ (TARA).

Come opportunamente segnalava già Figà-Talamanca (2000), in un'aspra critica dell'IF (qualificato come "virus"), non solo è assai poco sensato utilizzare le citazioni di un ente astratto come l' "articolo medio"⁴ da nessuno mai scritto, ma deve anche essere preso in conto l'effetto trasformativo che il fatto della diffusione dell'IF ha prodotto e produce nel sistema editoriale e della distribuzione. L'orientamento preferenziale e concorrenziale a pubblicare sul nucleo delle riviste a maggiore IF determinava e determina infatti la concentrazione degli articoli più interessanti in quel nucleo, facendo avverare la previsione a mezzo della previsione stessa. Questo spostamento ha così determinato un *focus sempre maggiore sulle riviste*, ha cioè amplificato enormemente l'importanza del "contenitore" e ha portato, in fatto, a ridurre relativamente il focus sull'articolo. Questo accadimento ha generato in breve tempo una rigida gerarchizzazione delle testate e ha posto le basi per una crescita sempre più poderosa del potere nelle mani degli editori. Ulteriormente, quella che Robert Merton qualificava come "moneta della scienza", cioè le citazioni che si impiegano per pagare tributo intellettuale, entra in una deriva iperinflattiva e perde valore. Non solo il numero delle citazioni-moneta esplode parallelamente all'impiego sempre più diffuso dell'IF, ma inizia a circolare moneta falsa⁵: ci si accorda tra autori, riviste, gruppi per citarsi a vicenda con abbondanza. Se riconosciamo che nell'atto stesso di citare già intervengono motivi strategici (per appoggiare certi autori e correnti o al contrario per cercare di emarginarli, pure talvolta per questioni di rivalità personale), dobbiamo riconoscere che attualmente la componente strategica delle citazioni ne risulta enormemente accresciuta, laddove esse non mirano più soltanto al riconoscimento intellettuale ma a condizionare incarichi, carriere e finanziamenti⁶.

Nella letteratura sulla valutazione della ricerca viene molto citata la legge di Goodhart che recita <quando una misura diviene un obiettivo, cessa di essere una buona misura>. Il punto più importante però da considerare è che *quando una misura diviene un obiettivo, il misurato cessa di essere quel che era prima*. L'IF ha infatti generato e diffuso comportamenti adattivi che gli anglosassoni chiamano *gaming behaviours*. Secondo la classica distinzione tra *play* e *game*, quest'ultimo è un gioco autotrasformativo, nel quale le regole divengono esse stesse oggetto del gioco. Le regole che governano i giochi della ricerca scientifica e della *scholarly communication* sono entrate così in una fase di mutazione, i cui contorni è opportuno cominciare a delineare.

A distanza di pochi anni, in seguito all'*International conference on science and technology indicators*, svoltosi a Leida nel settembre 2014, viene pubblicato nel 2015 su *Nature* il **Manifesto di Leida per le metriche della ricerca**⁷, articolato in dieci principi. La constatazione in premessa fa riferimento alla prassi ormai divenuta prevalente e "ossessiva" di valutare la ricerca a mezzo di dati, spesso male raccolti e male interpretati. I principi esortano a: utilizzare le metriche solo in funzione ausiliaria rispetto alla valutazione qualitativa condotta da esperti; correlare le prestazioni valutate alle finalità specifiche dei vari progetti di ricerca (di base, frontiera, sviluppo); non penalizzare le ricerche di interesse locale e quelle sociali ed umanistiche; rendere trasparenti indici e loro applicazioni; consentire ai valutati di verificare dati e

³ <https://www.cwts.nl/projects/current-projects/tools-to-advance-research-assessment-tara>

⁴ Ancor più in un contesto di asimmetria. Garfield, a sostegno della significatività dei suoi indici citazionali pur basati su un ristretto nucleo di riviste, poteva citare Bradford e il suo "moltiplicatore" (di applicazione settoriale) per cui risultava che gli articoli più "rilevanti" interni ad una disciplina sono concentrati in poche riviste. Similmente Lotka aveva rilevato che, all'interno dei settori scientifici, pochi autori contano per la gran parte degli articoli "rilevanti" di quei settori. Per non farsi mancare niente, anche Garfield si è confezionato la sua "legge di concentrazione", che è per lui anche una benedizione economica perché gli evitava il lavoro di dover indicizzare montagne di riviste. (Ci risparmiamo qui la legge di Zipf).

⁵ Marco Biagioli, in relazione a questo traffico di citazioni, parla di "post-production misconduct" (Biagioli, 2016).

⁶ Si vedano ad esempio le considerazioni di Paul Wouters (2020).

⁷ <https://www.nature.com/articles/520429a>

analisi; utilizzare una *suite* di indicatori diversificati per ambito disciplinare e utilizzare normalizzazioni⁸ di collocazione delle riviste per settore; valutare i ricercatori secondo più dimensioni qualitative del lavoro di ricerca; non fingere precisioni numeriche che sarebbero solo millantate; utilizzare *suite* di indicatori anziché uno solo, che rischia di condurre a comportamenti “sistemic”; esaminare e aggiornare regolarmente gli indici che devono correlarsi alle trasformazioni delle prassi di ricerca. In conclusione, gli estensori riconoscono che “research metrics can provide crucial information that would be difficult to gather or understand by means of individual expertise” e invitano i valutatori ad un saggio mix di dati quantitativi e valutazione qualitativa.

Alla metà del 2015 viene pubblicato il rapporto ***The metric tide: report of the independent review of the role of metrics in research assesment and management*** (Wilsdon et al. 2015). Si tratta di uno studio che sviluppa considerazioni sotto molteplici prospettive a partire dai risultati di una “call of evidence” nella quale la comunità accademica del Regno Unito era chiamata ad esprimere le proprie posizioni, in particolare in relazione al Research excellence framework (il processo di valutazione della ricerca accademica finalizzato all’efficiente allocazione delle risorse per la ricerca). Il gruppo di lavoro fa proprie le raccomandazioni di DORA, prende atto della “marea” crescente costituita dall’impiego sempre più diffuso e sistematico delle metriche (degli “indici”, come preferiscono chiamarli) e, dopo avere evocato il rischio di una “new barbery”, auspica l’adozione di un approccio pragmatico che conduca alla realizzazione di “metriche responsabili⁹”. “Proceeding cautiously, in an exploratory way, seems an appropriate approach” (Wilsdon et al. 2015, 29) - scrivono gli autori. Poiché il rapporto tra metriche citazionali e giudizi di qualità è problematico in quanto assai variabile all’interno delle diverse discipline¹⁰, nel rapporto viene anche presentata l’“informed peer review” “the idea that the judicious application of specific bibliometric data and indicators may inform the process of peer review, depending on the exact goal and context of the assessment” (Wilsdon et al. 2015, 64). Ma anche questa proposta di valutazione deve essere assunta con cautela poiché – suggerisce il rapporto – sono varie le interpretazioni circa la sua concreta attuazione ed inoltre è difficile escludere che la *peer review* non sia influenzata a sua volta dalle metriche relative a prodotti scientifici oggetto di valutazione.

Anziché opporsi alla marea, quasi fosse un’evenienza naturale, allo scopo di evitare o minimizzare “strategic behaviours and gaming” e, ancor peggio, forme di “goal displacement” (formula mertoniana per cui, semplicemente, il fine della ricerca scientifica di produrre nuova conoscenza viene sostituito dal fine di produrre impatto bibliometrico), nel rapporto vengono definiti cinque principi che dovrebbero informare queste un po’ misteriose entità che si chiamano “metriche responsabili”. I principi proposti si articolano in: “robustness” (le metriche devono basarsi sulla migliore raccolta possibile di dati in termini di accuratezza e di ambito); “humility” (va riconosciuto che la valutazione quantitativa dovrebbe essere di supporto alla valutazione qualitativa condotta dagli esperti, e non sostitutiva); “transparency” (raccolta dei dati e processi di analisi dovrebbero essere mantenuti aperti e trasparenti, affinché i valutati possano controllare e verificare i risultati); “diversity” (va tenuto conto delle diversità tra ambiti e andrebbe utilizzato un range di indicatori che riflettano e sostengano una pluralità di percorsi di ricerca e di carriera); “reflexivity” (vanno riconosciuti e anticipati gli effetti potenziali e sistemici degli indicatori, che vanno aggiornati in risposta a quegli effetti).

Da allora si sono moltiplicati manifesti e pubbliche prese di posizione da parte di comunità scientifiche disciplinari, finanziatori, editori e vari altri soggetti interessati alla produzione e alla circolazione delle conoscenze scientifiche che, con accenti più o meno convergenti, prospettano i rischi dell’utilizzo delle metriche citazionali, sottolineano i pericoli di una crescente pressione all’impatto e alla produttività e auspicano l’adozione di diversi criteri di valutazione dei prodotti scientifici e in generale delle performance dei ricercatori al fine di preservare l’integrità della scienza.

Un passo importante dal piano della denuncia a quello della proposta è stato recentemente compiuto dall’internazionale *Coalition for advancing research assesment* (CoARA) con l’***Agreement on reforming research assesment***¹¹ (sottoscritto anche dall’Anvur alla fine del 2022) mediante il quale si apre infine un cantiere operativo che impegna i membri a condividere entro un anno dalla sottoscrizione le loro soluzioni operative e a fissare come data la fine del 2027 per verificarne idoneità ed efficacia. Articolato per principi e impegni, l’agreement invita a partire dai

⁸ In realtà indici normalizzati per quartili nei vari settori disciplinari al fine della loro comparabilità già esistono sia in Web of science che in Scopus (naturalmente ogni indice è una “costruzione” e dunque la delimitazione dei *confini* delle discipline tra riviste e l’afferenza disciplinare degli articoli è un’opzione esercitata dai compilatori).

⁹ È disponibile un blog per favorire il confronto all’indirizzo www.ResponsibleMetrics.org

¹⁰ “The correlation between bibliometrics and peer review is weaker in most fields in the humanities, the applied fields, the technical sciences, and the social sciences. This is partly caused by lower coverage in the citation databases, but also by varying citation and publication cultures” (Wilsdon et al. 2015, 64)

¹¹ <https://coara.eu/agreement/the-agreement-full-text>

cantieri di TARA per le STEM e da Humetrics¹² (progetto finanziato dalla Mellon Foundation) per le HSS e in cui si combinano l'affermazione di principi generali con l'invito alla *rule of thumb* (ad un approccio pragmatico). Il passaggio ad un piano operativo è certamente da giudicare positivamente e quasi sicuramente cotanto dispiegamento di uomini e mezzi porterà a nuove proposte, tuttavia almeno rileviamo come già ora la bibliometria¹³ metta a disposizione una vasta panoplia di indici (variamente "responsabili"), che il riferimento allo SCOPE framework¹⁴ (estremamente impegnativo e laborioso per gli istituti che lo vorranno adottare) rischia di creare una babele di ulteriori indici, e che il *peer reviewing* vi compaia come stella della salvezza, senza ravvisarne limiti e costi¹⁵.

Mutazioni e casi teratologici

Nel Manifesto di Leida si fa riferimento a forti pressioni sui dottorandi perché pubblichino in gran fretta e su riviste ad alto IF al fine di attirare finanziamenti, si cita anche il caso di sociologi spagnoli che per spuntare indici più elevati di IF si concentrano su modelli astratti o su dati statunitensi.

Quanto in generale l'obiettivo dell'impatto orienta e plasma la ricerca scientifica stessa, nei suoi contenuti e nei suoi stili? In che misura muta il processo di produzione e condivisione della conoscenza, nelle sue prassi, nei suoi oggetti e nelle sue finalità? Fino a quale punto si può parlare di una *trasformazione della ricerca scientifica in ricerca scientifica strategica*?

Vediamo un po' di queste mutazioni o almeno apparenti tali. Retraction watch, che cura un database di articoli ritirati (e anche solo corretti) dopo la pubblicazione ci aggiorna di un numero crescente di pubblicazioni scientifiche che sono state ritirate. Oransky, co-fondatore di Retraction watch, si riferisce a un "retraction boom" (Oransky 2022) registrato negli ultimi anni. L'autore ci dice che nel 2010 gli articoli ritirati erano in media 45 al mese e che nel 2021 sono stati quasi 300. Il tasso complessivo di ritiro è un po' inferiore allo 0,1% ma aggiunge che, sulla base di alcuni rapporti e indagini condotti dai "cacciatori" di pubblicazioni errate o fraudolente, il tasso di articoli che andrebbero ritirati sarebbe di circa il 2%. A cosa è dovuta questa crescita dei ritiri? Ad un progressivo peggioramento del livello qualitativo della ricerca? Non parrebbe: per Oranski la comunità dei ricercatori è diventata sempre più attenta alle ricerche falsificate e dunque è stata in grado di attivare un più efficiente sistema di autocontrollo. Dunque se il numero delle pubblicazioni ritirate cresce non è tanto perché crescono quelle fasulle ma soprattutto perché vengono più efficacemente intercettate. Resta tuttavia il fatto inquietante che un numero niente affatto trascurabile di ricercatori falsifica scientemente e antiscientificamente i dati per poter pubblicare e guadagnare impatto. La cosa ulteriormente preoccupante è che spesso "researches continue to cite retracted papers" e che il processo di ritiro "remains comically clumsy, slow and opaque — often taking years, if it ever happens at all" (Oransky 2022, 9). Può essere di interesse sapere che esiste una correlazione (Fang e Casadevall, 2011) positiva tra frequenza di ritiro e IF, interpretabile con il fatto che su quelle riviste si pubblica più facilmente quando in questione ci siano scoperte di grande rilevanza (e dunque gli autori che vi vogliono pubblicare sentono l'incentivo a gonfiare indebitamente o anche fraudolentemente la portata dei loro risultati) e anche perché esiste un tasso di controllo e sorveglianza sugli articoli, dopo la pubblicazione, molto più elevato rispetto agli altri periodici. Sul JAMA è apparso uno studio (Gaudino et al. 2021) specifico per l'ambito biomedico relativo al periodo 1971-2020 da cui è risultato che il tasso delle pubblicazioni ritirate è crescente, specialmente dal 1980 al 2014 (in particolare dopo il 2004) ed inizia a calare a partire dal 2015, laddove il numero assoluto è in continua crescita almeno fino a circa il 2018. Nell'ambito degli studi oncologici Pantziarka e Meheus (2019), oltre a confermare l'andamento crescente dei ritiri anche per questa specializzazione, evidenziano come gli articoli esaminati avessero grande impatto potenziale anche su pazienti colti in cerca di trattamenti alternativi a quelli convenzionali, con connesso aumento del rischio di mortalità. La buona notizia che ci comunicano è che il tempo intercorrente tra la pubblicazione di un articolo di ambito oncologico e la sua identificazione come erroneo o fraudolento è in significativo costante calo, con riduzione del rischio che questi articoli entrino nel circolo delle citazioni. Particolare effetto fa sapere, come risulta da diverse note sul blog di Retraction watch che, in più di un caso, le industrie farmaceutiche avevano preso in considerazione per la produzione di farmaci alcune ricerche smascherate poi come fraudolente.

La fantasia dei ricercatori impegnati nel *goal displacement* è molto prolifica. Tra le modalità entrate maggiormente in voga, per esempio, c'è il "p-hacking" (conosciuto anche come "dredging" o "snooping") che consiste in almeno due varianti: operare molti test statistici sui dati e riportare solo quelli che danno risultati significativi di correlazione (i falsi

¹² <https://humetricshss.org/about>

¹³ Per un'introduzione in lingua italiana si possono vedere Nicola De Bellis (2014) e Simona Turbanti (2018)

¹⁴ <https://inorms.net/scope-framework-for-research-evaluation>

¹⁵ Si vedano per esempio: Brezis e Birukou (2020); Dondio et al. (2019); Richard Smith (2006). Biagioli parla di "fake peer-review rings" a indicare la permeabilità del modello da parte di condotte fraudolente miranti allo scambio "collaborativo" di citazioni.

Restano ancora assai istruttive le considerazioni di Figà-Talamanca che ricostruisce per l'ambito biomedico italiano il conflitto *peer reviewing* / metriche citazionali come conflitto di potere intergenerazionale interno all'accademia.

positivi) oppure operare test statistici su dati presentati come casuali e invece raccolti sulla base di alcuni criteri non dichiarati. Un altro modo scorretto di produrre risultati è lo HARKing (*hypothesizing after the results are known*) consistente nel presentare come ipotesi di correlazione successiva alla raccolta e ai test sui dati quanto invece è ipotesi precedente, assecondando così il “publication bias”, laddove risulterebbe meno profittevole pubblicare risultati che diano nullità di rilevanza statistica. Un altro preoccupante fenomeno all’interno del mondo della ricerca è quanto viene definita “crisi della riproducibilità” che mette per l’appunto in questione uno dei principi della scientificità stessa e dovuto a diversi fattori, di cui è difficile dire quanti siano intenzionali e malevoli e quali no. La mancanza o la parzialità dei dati messi a disposizione, l’impiego di licenze che inibiscano od ostacolino il riuso, l’applicazione di software proprietari o con versionamento non dichiarato, tutto ciò incide negativamente sulla possibilità del controllo e della validazione dei risultati da parte di altri gruppi di ricerca. Altre forme perniciose per l’integrità della scienza derivano dalla diffusione dei “predatory journals¹⁶”, ovvero di quelle riviste che millantano crediti scientifici, comitati editoriali illustri (i cui membri spesso ignorano di farne parte), rigore nel *peer reviewing*, livelli di IF inverosimili o altri indici fantasmagorici inventati di sana pianta, riconoscimenti internazionali etc., laddove si tratta in realtà di riviste che in diversi casi neppure svolgono il lavoro di controllo e selezione delle pubblicazioni ma mirano semplicemente ad incassare quanti più introiti dall’*article processing charge* (APC) secondo il modello pay-to-play. Naturalmente ci si può legittimamente chiedere: chi sarebbero queste prede così indifese e sprovvedute? Forse è più verosimile immaginare che tali periodici siano non di rado opportunisticamente utilizzati dai ricercatori non tanto per pubblicare articoli finalizzati alla lettura, ma al mero incremento del punteggio bibliometrico. Sul web viene aggiornata la straordinariamente lunga Bealls’list¹⁷ dei “potential predatory journals and editors”. Esiste poi quell’emergente multiforme fascio di “post-production misconducts” (ipoteticamente senza pregiudizio della qualità intrinseca della ricerca) che spazia dalla vera e propria frode al comportamento adattivo e che muta al variare delle metriche, a cui è associabile quella esplosione citazionale a cui abbiamo già accennato. Si tratta del circolo del “never ending tuning” (degli indici) e “hacking” (degli stessi) sul quale Biagioli (2016) insiste molto, rappresentandolo però quasi come un processo che si autogenera, senza – a nostro giudizio – richiamare adeguatamente il fatto che quel circolo si salda ogni volta perché un ricercatore viene meno ai principi fondativi dell’integrità della ricerca e della pubblica condivisione dei suoi risultati (poiché, come noto, non esiste virtualmente nulla che non possa essere manipolato).

Frequenti, ancora, sono state le denunce di casi per i quali l’autore aveva suggerito come valutatori alcuni reviewer e comunicato indirizzi mail di cui questi stesso era titolare.

Resta poi il caso, al limite dell’incredibile, del sw SCiGen, inventato da alcuni goliardi studenti del MIT per creare pseudo articoli di computer science dall’aspetto a prima vista credibile, e che è stato utilizzato in particolare da alcuni accademici cinesi per pubblicare articoli “a getto” (pubblicati anche da editori blasonati). Il più recente ChatGpt (e chissà per i suoi fratelli a venire), bot basato sull’intelligenza artificiale, promette nuove chicche pseudoscientifiche essendo in grado, oltre a rispondere a domande, di produrre testi, codici sw, contratti, immagini etc.. Che il sistema dei controlli e delle metriche possa infatti essere ampiamente giocato, è testimoniato anche dal signor Ike Antkare (I can’t care) dotato di un h-index superiore a quello di Einstein, pur senza essere mai esistito, se non nella fantasia dell’impertinente Cyril Labbé.

Meno preoccupante ma significativa è la pratica denunciata del “salami-slicing” consistente nel frazionare in più articoli la comunicazione dei risultati al fine di aumentare la produttività ma al costo di rendere più laborioso e farraginoso il lavoro di ricostruzione da parte del lettore. Va tuttavia evidenziato come statisticamente risulti un trend di lunga lena nella crescita media della lunghezza delle pubblicazioni, talché dal *salami-slicing* possiamo inferire un allungamento del brodo contenutistico (e non una riduzione della lunghezza degli articoli). La pratica crescente della *co-authorship*¹⁸, ulteriormente, per cui “the average number of authors of publications in the scholarly literature continues to increase” (Wilsdon et al. 2015, 32) è ancora più ambigua da interpretare perché può costituire risposta all’incentivo di mostrare alti livelli individuali di prolificità scientifica aumentando strategicamente i gruppi di ricerca con cui si collabora, oppure può indicare la scelta di una collaborazione meramente figurativa e negoziata come risposta alla pressione alla

¹⁶ In realtà pochi sono i casi in cui esistono sentenze definitive di condanna per falso nei confronti di editori di riviste scientifiche. Esiste una zona grigia assai ampia entro cui gli editori possono adottare condotte predatorie o di spamming e ugualmente pubblicare anche articoli scientifici validati. Verrebbe però forse più facile identificare quali sono gli editori *non predator*, date le oltraggiose politiche di prezzo (ed ora, in seguito all’affermazione degli “accordi trasformativi”, politiche di APC) da parte di editori blasonati che legalmente minano il processo di condivisione della conoscenza scientifica.

¹⁷ <https://beallslist.net>

¹⁸ Probabilmente il record di co-autorialità spetta ad un articolo pubblicato nel 2014 dalla *Physical review letters* firmato da 5.154 ricercatori

pubblicazione oppure ancora indicare un aumento del livello di collaborazione proprio della ricerca attuale (multidisciplinare e transdisciplinare).

Alcuni autori ritengono che, complessivamente, le mutazioni occorse al mondo della ricerca abbiano profondamente alterato la fisionomia del processo della *scholarly communication* nell'arco di pochi decenni, più di quanto sia accaduto dalla metà del Seicento. Mark Edwards e Siddharta Rey, in un articolo dal titolo suggestivo "Science is broken", dopo avere considerato che "Competition among researchers for funding has never been more intense, entering an era with the worst funding environment in half a century", scrivono (riferendosi prudentemente però ad un ipotetico futuro): "We believe that this system presents a real threat to the future of science. If immediate action is not taken, we risk creating a corrupt professional culture akin to that revealed in professional cycling (ie, 20 out of 21 Tour de France podium finishers during 1999-2005 were conclusively tied to doping), where an uncontrolled perverse-incentive system created an environment in which athletes felt that they had to cheat to compete" (Edwards e Rey, 2005). I risultati di un questionario sottoposto ad alcune migliaia di ricercatori negli Stati Uniti portavano gli autori a ipotizzare che "certain features of the research working environment may have unexpected and potentially detrimental effects on the ethical dimensions of scientists' work" (Martinson 2005). Ancora, scrivono altri autori "some violations are considered to be related to the increased pressures to publish" (Hayer 2013).

Parallelamente a queste denunce allarmate e allarmanti, sono sorte, accanto ad istituzioni consolidate come per esempio l'US Office of Research Integrity, iniziative spesso informali gestite da gruppi di ricercatori (talvolta anonimi e con il gusto dello sberleffo) come Retraction watch, PubPeer e altri blog in funzione di "watchdogs" proprio per intercettare queste inedite forme di *gaming*. È tornata conseguentemente di grande attualità anche la questione dell'*etica* della scienza, della responsabilità e dell'integrità di coloro che vi partecipano. Un recente documento particolarmente significativo è costituito dai cinque Principi di Hong Kong (Moher et al. 2020) come guida per i ricercatori e per i valutatori al fine di preservare l'"integrità" della scienza. L' "Allea European code of conduct for research integrity" propone principi e buone prassi per la preservazione della libertà della ricerca e la promozione di forme di autoregolazione e autocorrezione¹⁹. Lo "Slow science manifesto"²⁰ denuncia i rischi per la qualità della ricerca e richiama alla natura di *common* della conoscenza, oggi considerata a rischio a causa delle recenti forme di science management. Le linee guida del Committee on publication ethics²¹ (COPE) stabiliscono alcuni principi etici a cui dovrebbe inderogabilmente informarsi la *scholarly communication*.

Il tema dell'*etica* della scienza è probabilmente oggi più radicale di quello che generalmente si ritiene e – a nostro giudizio – costituisce il nucleo più importante e ancora in gran parte opaco interno a quanto in Italia oggi si definisce "terza missione" delle università. Sottratta alla retorica e allo stucchevole paternalismo che spesso accompagna le riflessioni sulla terza missione e anche sulla *citizen science*²², agli scienziati si chiede un'interrogazione radicale sul proprio esercizio, sulle conseguenze che esso produce nella società, nel pianeta e in generale su tutti gli oggetti di significato, fino a interrogare in radice il metodo scientifico stesso, nelle sue secolari e congiunte strutture epistemologiche del disinteresse e della osservatività.

Produzione di dati e cultura publish or perish

Abbiamo detto, all'inizio, che l'applicazione dei principi del new public management e in particolare dell'*accountability* ha avuto un ruolo sostanziale nel riplasmare strutture e prassi proprie delle pubbliche amministrazioni. Almeno un altro aspetto va però messo in linea di conto in questa nostra ricostruzione che vede l'impiego sempre crescente di metodi quantitativi nella raccolta, elaborazione e valutazione dei dati. E cioè il fatto che, per l'appunto, mai come oggi ci siamo ritrovati sommersi in un gigantesco mare di dati o, meglio, mai come oggi le società del terziario avanzato hanno *prodotto* tale quantità industriale di dati. Non è la sede per entrare nel merito della nozione di "dato", che per lo più viene assunto oggettivisticamente, ma è bene almeno avere contezza che il dato non è mai un oggetto, bensì una *relazione*, si trova cioè collocato nel punto di precipitazione di un complesso intreccio di pratiche, operazioni e

¹⁹ <https://allea.org/code-of-conduct>

²⁰ <https://slowscience.be/the-slow-science-manifesto-2>

²¹ <https://publicationethics.org>

²² Nelle dichiarazioni di principio della *European citizen science association* e dell'americana *Citizen science association* si trovano esortazioni a considerare la ricerca a cui partecipano gli appassionati di scienza come di pari livello della ricerca condotta internamente agli istituti di ricerca. Basta però leggerli gli statuti di alcune iniziative specifiche, e si ritrova pressoché sistematicamente (e – a nostro giudizio – inevitabilmente) una netta distinzione tra *researchers* e *practitioners*. È bastata una fetta del salame piccante *chorizo*, recentemente spacciata per un'immagine di Proxima Centauri, a raffreddare gli entusiasmi sulla *crowd science*.

tracciamenti che lo mettono e lo tengono all'opera. Il dato non sta, per richiamarci ad una celebre affermazione di Foucault, in un limbo, in attesa di un atto di attenzione e dunque la nota filastrocca dato-informazione-conoscenza non ci aiuta granché. Questa crescita esponenziale di dati è dovuta principalmente al fatto che ai nostri gesti sono state connesse scritture elettroniche automatiche e dunque, nell'esplicarsi, essi producono registrazioni sempre più fitte e "granulari", molto spesso senza che ci sia né intenzione né consapevolezza. Oltre ad essere automatiche, tali scritture avvengono a costi molto bassi e dunque si espandono con grande facilità a formare quella società che variamente è definita "audit society" (Power 1997) o "società della sorveglianza" e che – ancora per richiamarci alla prospettiva foucaultiana, vede *sovrapporsi* razionalità disciplinare (basata sulla registrazione e sul controllo del dettaglio e che si avvale di "quadrettature" (*quadrillages*) epistemologiche, organizzative, spaziali, architettoniche, elettroniche) e razionalità biopolitica che individua e plasma normali e trend statistici di ordine sociale e biologico). Le nuove tecnologie rendono possibile *contestualmente* il dettaglio e il massivo e dunque offrono al computo (agli algoritmi) un raggio di azione sociale inedito e crescente. Dunque - è bene preliminarmente averlo chiaro - la questione della valutazione non può risolversi in una trattazione che si limiti alle metriche citazionali e alle performance dei ricercatori solo perché investe gruppi sociali perlopiù privilegiati e maggiormente protetti come quelli dell'accademia. Al fine della sua comprensione essa non può restare su un piano corporativo ma richiede di essere collocata entro un più ampio orizzonte.

Tale processo capillare ed invasivo, nel mondo della ricerca, ha iniziato ad innescare condotte elusive, resistenze, giochi. Da tempo si sostiene che nell'accademia si sia introdotta una cultura ipercompetitiva legittimata dall'assunto che quanto più intensa è la competizione tra ricercatori e tanto maggiori saranno i risultati in termini di produttività e miglioramento della qualità della ricerca. Molta letteratura – viceversa – sostiene che quanto essa definisce *cultura del publish or perish* (a cui andrebbe aggiunto *impact or perish*) produce effetti opposti a quelli attesi e genera incentivi perversi che anziché migliorare l'impresa scientifica, la mina nei fondamenti, innesca mutazioni e teratologie. Il rischio - come si è detto - è che dal *publish or perish* esca una *broken science*. Anche sul piano del vissuto si raccolgono convergenti testimonianze. Per esempio, i risultati di un questionario sul tema sottoposto a ricercatori olandesi di ambito biomedico porta gli autori della ricognizione a considerare che "substantial proportion of the responding medical professors judge publication pressure as having become excessive, and a substantial part believes that this affects the validity and credibility of medical science" (Tijdink et al. 2013). Secondo le testimonianze, il problema si sposta anche sul piano del proprio livello di benessere e di salute, che ha ancora, a sua volta, conseguenze sul piano dell'efficienza scientifica, clinica e didattica. Scrivono infatti ancora gli autori "Apart from general effects on medical science and practice, our study suggests that excessive publication pressure has detrimental effects on personal well-being. [...] publication pressure causes them to develop a cynical view on medical science, and may be associated with increased risk of developing burn-out" (Tijdink et al. 2013).

Non c'è ragione di dubitare che quelle testimonianze circa un crescente disagio personale e collettivo non siano autentiche e conseguenti a trasformazioni effettive. Eppure non è affatto facile riconoscere un sicuro e diretto legame tra uso delle metriche citazionali, cultura POP e condotte scientifiche discutibili o fraudolente. In un modello comportamentale molto complesso (e molto astratto), già nel 2011 i ricercatori Lacetra e Zirulia mettevano in questione il legame causale tra cultura POP e condotte manipolatorie o fraudolente. Anzi, le pressioni all'impatto e alla produttività e l'alto livello di competizione possono sortire – ad avviso degli autori – un effetto dissuasivo rispetto alle condotte scorrette. Leggiamo: "This pressure can actually serve as a powerful mechanism to deter fraud as it increases the incentives of peers to scrutinize each other's work. This is due both to the increase in the reward from discovering fraud and to the anticipation of the author's higher temptation to cheat" (Lacetra e Zirulia 2011, 588). Più recentemente, Fanelli, sulla base di un ampio studio²³ statistico che si concentra sui primi 15 anni di carriera di un ampio numero di ricercatori, tende ad escludere alcuni effetti del POP e conclude "In sum, according to our findings, scientists today are not publishing, on an individual basis, at higher rates than their colleagues in the 1950s" (Fanelli 2020, 117). È ulteriormente difficile discriminare tra comportamenti (subiti) indotti dalla pressione istituzionale del POP (per cui o hai elevato impatto e alta produttività o sei tagliato fuori) e quelli invece (cercati) motivati soggettivamente dal carrierismo e dalla volontà di incamerare bonus (opzione lasciata alla libertà dei singoli). Scrive Fanelli "The pressure to publish hypothesis is not supported because the only significant patterns we detected in these and previous analyses support a role for cash incentive policies, which do not constitute real 'pressures' on researchers, but rather an incentive that is misaligned with the scientific norms of disinterestedness"²⁶ (Fanelli 2022, 15). È vero che le rilevazioni di Fanelli riguardano un limitato campione di riviste, si concentrano su casi specifici di cattive condotte, concernono limitati ambiti disciplinari, selezionano (necessariamente) un numero limitato di variabili, ipotizzano "intenzioni" e dunque non sono né interamente generalizzabili né conclusive, tuttavia ci ammoniscono a non accedere a facili semplificazioni sui

²³ In relazione a questi studi si è impiegato il pomposo nome di *metascience*, laddove più prosaicamente si tratta dell'applicazione di metodi statistici alla *scholarly communication*

fenomeni in atto nell'ambito della ricerca e della comunicazione scientifiche. Aggiungiamo che quanto abbiamo definito insorgenze riguarda piccole *minoranze* di ricercatori: per esempio la dichiarazione DORA, tra i documenti più conosciuti, ad oggi risulta sottoscritta da circa 20.000 singoli ricercatori e 2.700 organizzazioni, numeri niente affatto trascurabili, laddove però il numero globale dei ricercatori è di poco inferiore ai nove milioni²⁴. Al netto delle ingenuità di alcuni commentatori paladini del *peer reviewing* e fieramente ostili alle metriche, è bene sapere che questa è anche una partita *interna* alla comunità scientifica, non solo una battaglia di puri principi. La misura del reciproco peso tra le due modalità di valutazione si polarizza entro un campo di forze e può determinare mutamenti per l'appunto in termini di carriere, distribuzione di fondi, assegnazione di incarichi etc.

Insomma non ci è consentito semplicisticamente dire che competizione, misurazione delle performance, impiego delle metriche citazionali abbiano in sé effetti distruttivi. Quanto possiamo ipotizzare, dai preoccupanti segnali di talune mutazioni, dalle molteplici insorgenze e dalle testimonianze di disagio, è che ci stiamo verosimilmente appressando ad una *soglia* (forse già in più ambiti oltrepassata) che può indurre la trasformazione della ricerca scientifica in ricerca scientifica strategica. È allora oggi il tempo dello studio attento e critico di *scholarly production* e *scholarly communication* e non delle dichiarazioni caricaturali su burocrati trinariciuti in cospirazione contro la scienza.

References

Biagioli, Marco. 2016. "Watch out for cheats in citation game". *Nature* 535: 201 <https://doi.org/10.1038/535201a>.

Biagioli, Mario e Lippman, Alexandra (cura), ed. 2020. *Gaming the metrics: misconduct and manipulation in academic research*. Cambridge: The Mit press.

<https://mitpress.mit.edu/9780262537933/gaming-the-metrics>

Brezis, Elise S., e Aliksandr Birukou. 2020. "Arbitrariness in the Peer Review Process". *Scientometrics*, 2123: 393–411. <https://doi.org/10.1007/s11192-020-03348-1>;

Cari-Ann, Hayer (et al.). 2013. "Pressures to publish: catalysts for the loss of scientific writing integrity?". *Papers in Natural Resources* 689. Doi: 10.1080/03632415.2013.813845
<https://digitalcommons.unl.edu/natrespapers/689>

De Bellis, Nicola. 2014. *Introduzione alla bibliometria*. Roma: AIB.

Dondio, Pierpaolo (et al.) 2019. "The 'Invisible hand' of peer review: the implications of author-referee networks on peer review in a scholarly journal". *Journal of Informetrics* 13: 708–716. <https://doi.org/10.1016/J.JOI.2019.03.018>.

Edwards, Mark A. e Roy Siddhartha. 2017. "Science is broken. Perverse incentives and the misuse of quantitative metrics have undermined the integrity of scientific research". Aeon.
<https://aeon.co/essays/science-is-a-public-good-in-peril-heres-how-to-fix-it>

Fanelli, Daniele. 2020. "Pressures to publish: what effects do we see? In *Gaming the metrics: misconduct and manipulation in academic research*, a cura di Biagioli, Mario e Lippman, Alexandra, 111-122. Cambridge: The Mit press. Doi: <https://doi.org/10.7551/mitpress/11087.003.0011>

Fanelli, Daniele (et al.) 2022. "Do individual and institutional predictors of misconduct vary by country? Results of a matched-control analysis of problematic image duplications". *PLoS One* 17 (3) e0255334. <https://doi.org/10.1371/journal.pone.0255334>

Fang C., Ferric e Casadevall, Arturo. 2011. "Retracted science and the retraction index". *Infective Immunology* 79 (10): 3855-9. Doi: 10.1128/IAI.05661-11

²⁴ Dato ricavabile da Science business del 2021 <https://sciencebusiness.net/news/number-scientists-worldwide-reaches-88m-global-research-spending-grows-faster-economy>

Figà-Talamanca, Alessandro. 2000 "L'impact factor nella valutazione della ricerca" [Bozza dell'intervento, Bologna 27 giugno 2000].

http://siba.unipv.it/biblioteche/info/SISSCOWEB_A_Fig_Talamanca_L%27impact_factor_nella_valutazione_della_ricerca_SISSCOWEB.htm

Garfield, Eugene. 1998. "The impact factor and using it correctly". *De Unfallchirurg*, 48 (2): 424.

http://garfield.library.upenn.edu/papers/derunfallchirurg_v101%286%29p413y1998english.html

Gaudino, Mario (et al.). 2021. "Trends and characteristics of retracted articles in the biomedical literature, 1971 to 2020". *JAMA international medicine* 181 (8): 1118-1121

Doi: [10.1001/jamainternmed.2021.1807](https://doi.org/10.1001/jamainternmed.2021.1807)

Lacetera, Nicola e Zirulia, Lorenzo. 2011. "The economics of scientific misconduct". *The Journal of Law, Economics, and Organization*. 27: 568–603. <https://doi.org/10.1093/jleo/ewp031>

Martinson, Brian C. (et al.). 2005. "Scientists behaving badly". *Nature*. 435: 737-738

Moher, David, et al. (2020). "The Hong Kong Principles for assessing researchers: fostering research integrity". *PLoS Biology* 18 (7). <https://doi.org/10.1371/journal.pbio.3000737>.

Naujokaitytė, Goda. 2021. "Number of scientists worldwide reaches 8.8M, as global research spending grows faster than the economy" in *Science business*

<https://sciencebusiness.net/news/number-scientists-worldwide-reaches-88m-global-research-spending-grows-faster-economy>

Oranski, Ivan. 2022. "Retractions are increasing, but not enough". *Nature* 608: 9. <https://doi.org/10.1038/d41586-022-02071-6>

Pantziarka, Pan e Meheus, Lydie. 2019. "Journal retractions in oncology: a bibliometric study". *Future oncology* 15 (31): 3597-3608. <https://doi.org/10.2217/fon-2019-0233>

Power, Michael. 1997. *The Audit Society*. Oxford: Oxford University Press

Smith, Richard. 2006. "Peer review: a flawed process at the heart of science and Journals". *Journal of the Royal society of medicine*, 199: 178-182.

Tijdink, Joeri K (et al.). 2013. "Publication pressure and burn out among Dutch medical professors: a nationwide survey". *PLoS One* 8 (9):e7338. Doi: [10.1371/journal.pone.0073381](https://doi.org/10.1371/journal.pone.0073381).

Turbanti, Simona. 2018. *Strumenti di misurazione della ricerca*. Milano: Editrice Bibliografica.

Wilsdon, James et al., ed. 2015. *The metric tide: report of the Independent review of the role of metrics in research assessment and management*. Thousand Oaks: SAGE publications.

doi: [10.13140/RG.2.1.4929.1363](https://doi.org/10.13140/RG.2.1.4929.1363). <https://sk.sagepub.com/books/the-metric-tide>

Wouters, Paul. 2020. "The mismeasurement of quality and impact". In *Gaming the metrics: misconduct and manipulation in academic research*, a cura di Mario Biagioli e Alexandra Lippman, 66-75. Cambridge, The Mit press.

<https://doi.org/10.7551/mitpress/11087.003.0006>

Tutti i siti sono stati verificati in data 15/02/2023

