

Gli archivi ad accesso aperto e l'impatto delle pubblicazioni

Danilo Deana

Divisione Coordinamento biblioteche
Università degli studi di Milano
danilo.deana@unimi.it

Indicatori e tecniche di misurazione

Negli ultimi anni si è assistito a una crescita dell'interesse di governi, imprese e altre istituzioni verso le iniziative volte a misurare l'impatto della ricerca.¹ Non sempre però ci si è soffermati sulle tecniche utilizzate per ottenere tali misure. L'obiettivo di questo articolo è più modesto. Non intendiamo infatti occuparci di scientometria, ma solo di quella parte di essa che utilizza tecniche matematiche e statistiche per misurare l'impatto delle pubblicazioni, e che va sotto il nome di bibliometria.² Le domande cui cercheremo di rispondere sono due:

- se i pre-print o i post-print³ degli articoli depositati negli archivi aperti comportino un vantaggio dal punto di vista dell'impatto;
- se le tecniche della bibliometria possano essere applicate (o lo siano già state) anche a questi archivi.

L'Impact Factor

Per misurare l'impatto di un articolo ci si dovrebbe basare sull'autorevolezza del periodico su cui è comparso e sul numero delle citazioni che ha ricevuto. Naturalmente, si dovrebbe anche tener conto dell'autorevolezza dei periodici su cui sono apparsi gli articoli che citano l'articolo in questione, per non dare legittimità a pubblicazioni appartenenti, ad esempio, a quella che



Eco ha chiamato la quarta dimensione.⁴

Sino a qualche tempo fa, nel misurare l'impatto di un articolo, si tendeva per comodità a privilegiare l'autorevolezza del periodico a scapito del numero di citazioni,⁵ nonostante fossero state addotte molte buone ragioni per non farlo.⁶

L'indicatore più diffuso per misurare l'autorevolezza di un periodico è senza dubbio l'Impact Factor. Esso nasce da una pubblicazione, il "Science Citation Index", il cui scopo era quello di eliminare i riferimenti a dati falsi, erronei o incompleti, dando la possibilità ai ricercatori di conoscere quanto era stato scritto a proposito degli articoli che intendevano citare.⁷ Una volta individuati tutti gli articoli che ne citano un determinato altro, è stato relativamente facile confrontare il numero degli articoli pubblicati su un periodico con quello delle citazioni che questi

hanno ricevuto e attribuire ad ogni periodico un Impact Factor.⁸

L'enorme mole di dati raccolta e la semplicità di utilizzo hanno decretato il successo di questo indicatore, tanto che anche nelle *Linee guida per la valutazione della ricerca* del Comitato di indirizzo per la valutazione della ricerca (CIVR) era raccomandato, per gli articoli, di riportarlo.⁹

Da qualche anno a questa parte le critiche verso l'Impact Factor (e verso il prodotto che da questo ha preso origine, il *Journal Citation Report*) sono andate moltiplicandosi: esse riguardano il criterio con cui sono selezionati i periodici spogliati (meno di 9.000 sulle decine di migliaia esistenti e in prevalenza di lingua inglese)¹⁰ e quello con cui sono prese o meno in considerazione determinate tipologie di articoli (le rassegne bibliografiche, ad esempio).¹¹ Nessuna di queste critiche, però, mette in questione il fondamento su cui si basa il calcolo dell'Impact Factor, né la necessità di valutare l'autorevolezza di un periodico per poter stabilire l'impatto sia degli articoli pubblicati, sia di quelli citati nel periodico stesso.

In tempi recenti sono stati messi a punto degli indicatori alternativi il cui scopo è quello di risolvere i problemi legati all'utilizzo dell'Impact Factor e che mettono l'accento più sul numero di citazioni rice-

vute che sull'autorevolezza del periodico. I più noti tra questi nuovi indicatori sono l'Eigenfactor e l'*h*-index.

Alla base dell'Eigenfactor c'è un'idea della letteratura accademica come di una rete di contributi connessi tra loro attraverso le citazioni, idea espressa per la prima volta da de Solla Price.¹² La struttura di questa rete riflette i milioni di decisioni prese dai singoli ricercatori relativamente ai contributi che essi considerano importanti per il loro lavoro. La misura dell'importanza di un determinato articolo è quindi direttamente proporzionale al tempo che ad esso hanno dedicato i ricercatori, ed è appunto questo che si propone di misurare l'Eigenfactor, utilizzando le tecniche messe a punto da Sergey Brin e Lawrence Page, i creatori di Google.¹³

L'*h*-index, che prende nome dal suo inventore, Jorge H. Hirsh, è stato invece messo a punto per misurare l'impatto che la produzione scientifica di un determinato ricercatore ha avuto nel suo ambiente disciplinare.¹⁴

Tanto l'Eigenfactor quanto l'*h*-index prendono comunque in considerazione solo una parte dei periodici esistenti, il che ci riporta al problema cui accennavamo in precedenza, ossia la necessità di valutare l'autorevolezza di un periodico per poter stabilire l'impatto sia degli articoli pubblicati, sia di quelli citati nel periodico stesso.¹⁵

Archivi aperti e impatto

Harnad – che con la sua proposta sovversiva¹⁶ è diventato uno degli alfieri del movimento che vuole la letteratura scientifica liberamente disponibile “per tutti, dappertutto e per sempre” – parla a proposito dell'accesso aperto di due strade: 1) la “gold road”, ossia la pubblicazione in un periodico ad accesso aperto;¹⁷

2) la “green road”, ossia la pubblicazione dell'articolo in un periodico non ad accesso aperto e il successivo deposito del pre-print o del post-print in un archivio ad accesso aperto.

Per quanto riguarda la “gold road”, la domanda cui occorre rispondere è se un articolo, per il fatto di essere pubblicato su un periodico ad accesso aperto, goda di qualche vantaggio dal punto di vista dell'impatto. L'indagine di Eysenbach ha posto a confronto articoli pubblicati sul medesimo periodico, alcuni liberamente disponibili, altri no. La conclusione cui è giunto è che i primi sono di gran lunga più citati dei secondi.¹⁸ Come ulteriore conferma si potrebbe portare l'atteggiamento di molti ricercatori che, di fronte ad un numero di articoli continuamente crescente, adottano la strategia dell'abbastanza buono, privilegiando ciò che è immediatamente disponibile rispetto a ciò che non lo è.¹⁹

Nei caso dei pre-print o dei post-print depositati in archivi ad accesso aperto, ossia della “green road”, la domanda è diversa. In questo caso, infatti, si tratta di stabilire se la possibilità di consultare liberamente queste versioni dell'articolo abbia o meno un effetto positivo sull'impatto dell'articolo stesso.

Il primo studio dedicato all'argomento è stato quello di Lawrence, dove si sosteneva che la disponibilità dell'articolo aumentava significativamente il numero di citazioni.²⁰ Successivamente, si sono individuati altri motivi che possono portare a un incremento delle citazioni: il fatto che gli articoli siano disponibili prima e il fatto che gli autori rendano disponibili di preferenza i loro articoli migliori.²¹ Di qui gli inviti alla prudenza nel considerare la correlazione tra la disponibilità e l'aumento del numero delle citazioni,²² inviti che comunque devono riguardare più l'entità dell'effetto che l'effetto stesso.²³

Da questo punto di vista, non si può quindi che condividere l'esortazione di Meho:

Scientists now need to make it their job to disseminate their work on as many platforms and in as many different ways as possible, such as publishing in open access and high-impact journals, and posting their work in institutional repositories, personal homepages and e-print servers, if they want their peers to be aware of, use and ultimately cite their work. Publishing a journal article is now only the first step in disseminating or communicating one's work; the Web provides a multitude of methods and tools to publicize its scholarly worth.²⁴

Questo anche in considerazione del fatto che, come già detto, sebbene non sia vero che solo ciò che è digitale esiste, pure gli studiosi adottano sempre più nelle loro ricerche bibliografiche l'approccio dell'“abbastanza buono”: se un articolo non è immediatamente disponibile, lo si ignora e se ne cerca un altro che lo sia.

Nuovi vestiti per vecchi numeri

Il titolo del paragrafo è ripreso da un libro di Irvin Adler dedicato ai numerali e mi è sembrato particolarmente adatto a descrivere la situazione delle bibliometria a confronto con l'accesso aperto. I periodici ad accesso aperto rappresentano infatti un modello alternativo all'editoria tradizionale, pensato per risolvere quella che viene chiamata la crisi dei periodici,²⁵ ma sono valutati nello stesso modo di quelli a pagamento, tanto è vero che molti hanno un Impact Factor.²⁶ Recentemente, si è cercato di affiancare nuovi indicatori a quelli in uso, come ad esempio l'Usage Factor. La nascita di questo indica-

tore è stata resa possibile dall'aumento delle versioni elettroniche dei periodici e dal lancio di un'iniziativa come COUNTER.²⁷ Esso infatti risulta dal rapporto tra "total usage (COUNTER JR1 data for a specified period)" e "total number of articles published on line (during a specified period)".

Sheperd, il direttore di COUNTER, sostiene che il concetto di Usage Factor è un concetto significativo, che esso è relativamente facile da implementare e che fornisce elementi aggiuntivi per giudicare il valore e la qualità dei periodici elettronici.²⁸ Già non tutti considerano possibile stabilire l'impatto attraverso le citazioni;²⁹ utilizzare poi indicatori basati su misure messe a punto per tutt'altri scopi³⁰ e che possono, tra l'altro, essere facilmente alterate, è una strada che deve essere percorsa con prudenza, anche se l'obiettivo è solo quello di un'integrazione che dovrebbe portare ad una maggiore affidabilità delle valutazioni.

Si è pensato di utilizzare il numero degli scaricamenti anche nel caso degli archivi ad accesso aperto. Secondo alcuni, esso permetterebbe di prevedere il futuro impatto dell'articolo e potrebbe essere affiancato agli indicatori basati sul numero della citazioni.³¹ Si tratta però di stime che risentono tanto del campione considerato, quanto dei mutamenti che potrebbero intervenire nel mondo dell'accesso aperto³² e che vanno anch'esse prese con molta cautela.³³

Un recente articolo di Davis e Fromerth è appunto dedicato a dimostrare quanto sia complesso utilizzare questo tipo di indicatori, soprattutto se li si vuole considerare insieme. Essi hanno notato come gli articoli depositati in ArXiv siano meno scaricati dai siti degli editori rispetto a quelli non depositati. Ora, poiché gli articoli depositati risultano più citati rispetto a quelli non depositati, come spie-

gare questa diminuzione? Se la citazione è una misura della qualità, infatti, gli articoli depositati dovrebbero essere più scaricati, non meno.³⁴

Archivi tematici e archivi istituzionali

Ci sono ormai numerosi esempi di archivi che rappresentano un punto di riferimento per una particolare disciplina, primo fra tutti ArXiv. Si tratta di realtà consolidate inserite da tempo tra quelle considerate dai motori di ricerca specializzati, ma diffuse soprattutto in ambito scientifico.

Gli archivi istituzionali assolvono ad altre funzioni. Dal punto di vista della valutazione della ricerca, il fatto di contenere tutte le registrazioni bibliografiche delle pubblicazioni di docenti e ricercatori di un'università permetterebbe però di valutare non solo la produzione più propriamente scientifica dell'università stessa, ma anche quella prodotta nel campo delle scienze umane e sociali. Tra questi campi, infatti, esistono notevoli differenze per quanto riguarda le pratiche relative alle pubblicazioni e alle citazioni;³⁵ queste differenze fanno sì che i metodi utilizzati per valutare la ricerca nel primo non possano essere applicati senza modifiche ai secondi.³⁶ Un archivio istituzionale, con l'elenco completo dei risultati ottenuti in termini di pubblicazioni, permetterebbe di apportare i necessari aggiustamenti.³⁷ In più, cosa che van Raan ha dimostrato non essere possibile utilizzando le registrazioni bibliografiche presenti nella banche dati,³⁸ si potrebbe valutare non solo l'università nel suo complesso, ma anche i singoli dipartimenti che la compongono.

Nel 2004 l'allora Ministero dell'istruzione, dell'università e della ricerca aveva indicato come possibi-

le candidato a ricoprire il ruolo di Archivio nazionale della ricerca le procedure informatiche che il CINECA aveva predisposto per ogni docente.³⁹ La banca dati CINECA è però poco popolata rispetto alle anagrafi della ricerca "locali". La ragione sta nel fatto che i docenti inseriscono nella banca dati solo le registrazioni bibliografiche di pubblicazioni relative al progetto di ricerca per il quale chiedono un finanziamento. Ciò è un freno alla completezza bibliografica che, al contrario, le anagrafi dei singoli atenei riescono a raggiungere tramite un più costante aggiornamento da parte dei docenti in occasione della distribuzione dei fondi dipartimentali di ricerca.⁴⁰

Conclusioni

È fondamentale che le registrazioni bibliografiche all'interno degli archivi istituzionali permettano di collegare il pre-print o il post-print alla versione definitiva dell'articolo e forniscano sufficienti elementi ai motori di ricerca per aggiungere servizi a quelli previsti dall'archivio stesso.⁴¹

Definire chiaramente i contenuti dell'archivio e assicurare la qualità dei dati ha permesso l'inserimento di Wageningen Yeld, l'archivio istituzionale della Wageningen University,⁴² in Google Scholar e in Scopus.⁴³ Questo è molto importante in quanto il vantaggio degli articoli liberamente disponibili non deriva tanto dalle interrogazioni compiute direttamente nell'archivio all'interno del quale sono inseriti, quanto da quelle effettuate dai motori di ricerca che rendono le registrazioni dei periodici ad accesso aperto e degli archivi istituzionali visibili all'interno di grandi aggregazioni;⁴⁴ e sono eventualmente in grado di offrire una serie di servizi aggiuntivi.⁴⁵

Abbiamo lasciato per ultima una considerazione che riguarda gli sviluppi futuri delle iniziative legate all'accesso aperto e alla valutazione. Sarebbe troppo lungo elencare tutte le difficoltà che si oppongono alla realizzazione di un progetto di questo tipo, ma esso merita comunque di venire menzionato.

L'inserimento all'interno degli archivi istituzionali potrebbe risultare di particolare importanza soprattutto, anche se non solo, nel caso di pubblicazioni relative ad aree disciplinari non coperte da banche dati utilizzabili per la valutazione della ricerca. Sebbene infatti esistano molte banche dati per l'area delle scienze umane e sociali, esse hanno caratteristiche che le rendono poco adatte o addirittura del tutto inutili ai fini del calcolo di indicatori bibliometrici.⁴⁶ Se si riuscisse ad estendere l'utilizzo di uno strumento come Citebase a tutti gli archivi istituzionali,⁴⁷ come auspicato dagli autori di *Incentiving the Open Access research Web*, si disporrebbe di una base dati che permetterebbe di superare i limiti degli strumenti utilizzati attualmente.⁴⁸

Note

¹ "In Italia è stata istituita l'Agenzia nazionale di valutazione del sistema universitario e della ricerca (ANVUR), il cui scopo è quello di svolgere attività di valutazione e di proporre al Ministro dell'università e della ricerca i parametri di ripartizione delle quote dei finanziamenti statali alle università e agli enti di ricerca che dipendono dalla qualità dei risultati delle attività svolte" (*Regolamento recante disposizioni relative alla struttura e al funzionamento dell'Agenzia nazionale di valutazione del sistema universitario e della ricerca (ANVUR)*, bozza del 12 marzo 2007).

² Parlando di bibliometria, occorre sempre tener presente che "peer review undoubtedly is and has to re-

main the principal procedure of quality judgement. But peer review and related expert-based judgements may have serious shortcoming and disadvantages [...]. For a substantial improvement of decision-making on matters of scientific activities, advanced bibliometric indicators have to be used in parallel to a peer-based evaluation procedure" (A.F.J. VAN RAAN, *Fatal attraction: conceptual and methodological problems in the ranking of universities by bibliometric methods*, "Scientometrics", 62 (2005), 1, p. 133-143; 135).³ I termini "pre-print" e "post-print" sono usati come nel sito SHERPA (*Securing a Hybrid Environment for Research Preservation and Access*): con pre-print si intende quindi "the version of the paper before peer review"; con post-print "the version of the paper after peer-review, with revisions having been made. This means that in terms of content, post-prints are the article as published. However, in terms of appearance this might not be the same as the published article, as publishers often reserve for themselves their own arrangement of typesetting and formatting".

⁴ La quarta dimensione è quella dell'editoria a pagamento che, come notava Umberto Eco, ha anch'essa "le sue riviste e i suoi editori, il proprio giro di recensioni (rarissime stroncature), le proprie vie del successo e del pubblico riconoscimento" (U. Eco, *L'industria del genio italico*, in *Il costume di casa*, Milano, Bompiani, 1973, p. 59-72; 61).

⁵ Vale la pena di ricordare, a questo proposito, che "some 90% of papers that have been published in academic journals are never cited. Indeed, as many as 50% of papers are never read by anyone other than their author, referees and journal editors" (L.I. MEHO, *The rise and rise of citation analysis*, "Physics World", 20 (2007), 1, p. 32-36; 32).

⁶ La critica più radicale a questo approccio è quella di Seglen. Egli inizia col rilevare che "although journal impact factors are rarely used explicitly, their implicit counterpart, journal prestige, is widely held to be a valid evaluation criterion and is probably the most used indicator besides a straightforward count of publications". Tutta-

via "the journal cannot in any way be taken as representative of the article" e anche se lo fosse, "the journal impact factor would still be far from being a quality indicator: citation impact is primarily a measure of scientific utility rather than of scientific quality, and authors' selection of references is subject to strong biases unrelated to quality" (P.O. SEGLEN, *Why the impact factor of journals should not be used for evaluating research*, "British Medical Journal", 314 (1997), p. 498-502; 500).

⁷ E. GARFIELD, *Citation indexes for science. A new dimension in documentation through association of ideas*, "Science", 122 (1955), p. 108-111.

⁸ ID., *Citation analysis as a tool in journal evaluation*, "Science", 178 (1972), p. 471-479. La formula con cui questo indicatore viene calcolato è la seguente: "A journal's impact factor is based on 2 elements: the numerator, which is the number of citations in the current year to any items published in a journal in the previous 2 years, and the denominator, which is the number of substantive articles (source items) published in the same 2 years" (ID., *Journal impact factor: a brief review*, "Canadian Medical Association Journal", 161 (1999), 8, p. 979-980; 979).

⁹ "Ogni prodotto selezionato deve essere accompagnato da una scheda illustrativa che riporti: area e disciplina di riferimento, un *abstract*, e ogni notizia utile a definire il collocamento internazionale, l'eventuale carattere interdisciplinare, la notorietà del prodotto, nonché l'autorevolezza della rivista / editore / manifestazione / ecc. con cui è stato reso pubblico, inclusi, se applicabili, gli indici bibliometrici (in particolare *impact factor* e *citation analysis*)" (COMITATO DI INDIRIZZO PER LA VALUTAZIONE DELLA RICERCA, *Linee guida per la valutazione della ricerca*, [Roma], Comitato di indirizzo per la valutazione della ricerca, 2003, p. 24-25).

¹⁰ Va detto che Garfield non considera un difetto la mancanza di una copertura completa: "Lack of complete coverage is not necessary an argument against a citation index. It is in fact an argument in its favor" (E. GARFIELD, *Citation indexes for science...*, cit., p. 109).

¹¹ P. DONG – M. LOH – A. MONDRY, *The "impact factor" revisited*, "Biomedical Digital Library", 2 (2005), 7.

¹² D.J. DE SOLLA PRICE, *Networks of scientific papers*, "Science" 169 (1965), p. 510-515.

¹³ S. BRIN – L. PAGE, *The anatomy of a large-scale hypertextual Web search engine*, "WWW7 / Computer Networks", 30 (1998), 1-7, p. 107-117. Il metodo con cui è calcolato l'Eigenfactor è illustrato in *Eigenfactor: detailed methods*, <<http://eigenfactor.org>>.

¹⁴ "A scientist has index b if b of his/her N_p papers have at least b citations each, and the other ($N_p - b$) papers have no more than b citations each" (J.E. HIRSCH, *An index to quantify an individual's scientific research output*, "Proceedings of the National Academy of Science of the United States of America", 102 (2005), 46, p. 16596-16572; 16596). L' b -index ha già alcune varianti: il *contemporary b-index* (A. SIDIROPOULOS – D. KATSAROS – Y. MANOLOPOULOS, *Generalized b-index for disclosing latent facts in citation networks*, <arXiv:cs.DL/0607066>); l'*individual b-index* (P.D. BATISTA – M.G. CAMPITELI – O. KINOCHI – A.S. MARTINEZ, *Is it possible to compare researchers with different scientific interests?*, "Scientometrics", 68 (2006), 1, p. 179-189), l'*Egghe's g-index* (L. EGGHE, *Theory and practice of the g-index*, "Scientometrics", 69 (2006), 1, p. 131-152) e l'*AW-index* (BIHUI JIN, *The AR-index: complementing the b-index*, "ISSI Newsletter", 3 (2007), 1, p. 6). Devo queste segnalazioni ad Antonella De Robbio e Paola Gargiulo, che ringrazio per la collaborazione.

¹⁵ "Ma se è vero, com'è vero, che l'*impact factor* di limiti ne ha parecchi, la conferma della validità dell'*b-index* parametrizzata sull'*impact factor* questi limiti se li porta con sé" (R. MORRIELLO, *L'indice di Hirsch (b-index) e altri indici citazionali dopo l'impact factor*, "Biblioteche oggi", 25 (2007), 1, p. 23-31; 29).

¹⁶ S. HARNAD, *Universal FTP archives for esoteric science and scholarship: a subversive proposal*, in *Scholarly journals at the cross-roads: a subversive proposal for electronic publishing*, edited by A. Okerson and J. O'Donnell, Association of Research Libraries, Washington DC, 1995.

¹⁷ S. HARNAD, T. BRODY, F. VALLIERES, L. CARR, S. HITCHCOCK, Y. GINGRAS, C. OPENHEIM, H. STAMERJOHANN, E.R. HILF, *The access/impact problem and the green and gold roads to Open Access*, "Serials Review", 30 (2004), 4, p. 310-314; 310. L'aggettivo "green" non è utilizzato a caso. Il sito SHERPA, infatti, classifica gli editori in bianchi, gialli, blu e, appunto, verdi. Questi ultimi sono quelli che consentono il deposito in un archivio ad accesso aperto tanto dei pre-print quanto dei post-print.

¹⁸ "The 1.159 papers which were neither self-archived nor immediate journal OA articles had an average 4.4 citations, whereas the 334 papers which were either self-archived or published originally as OA (or both) had 5.9 [...] citations. The risk of not being cited for papers which were either published originally as OA or self-archived was only 6.9%, while it was 13.81% for articles in the non-OA group" (G. EYSENACH, *Citation advantage of open access articles*, "PLOS Biology", 4 (2006), 5, p. 692-698; 695).

¹⁹ "Because libraries cannot provide comprehensive access to all the scholarly literature, it is inevitable that researchers will, at some point, need to source the articles they need elsewhere. The survey indicates that researchers are most likely to turn to inter-library loan (ILL): 80% of them use this service at least occasionally. The second most popular option – used by 40% of researchers – is to contact the author to request a copy of the article; and we know from the focus groups that they also contact colleagues in other institutions. But many researchers, if they come up against a toll barrier when they are searching for an article, say that they 'give up and find something else I can get for nothing'. [...] But we detected in our interviews and focus groups another aspect of the 'good enough' approach: if an article or book will take some time to obtain – and that time varies between disciplines – then some researchers will abandon the chase and go without" (*Researchers' use of academic libraries and their services, a report commissioned by The Research Information Network and the Consortium of Research Libraries*, 2007, p. 32-33).

²⁰ "We analysed 119,924 conference articles in computer science and related disciplines [...]. The mean number of citations to offline articles is 2.74, and the mean number of citations to online articles is 7.03, an increase of 157%" (S. LAWRENCE, *Free online availability substantially increases a paper's impact*, "Nature", 411 (2001), p. 521).

²¹ "There are (at least) three possible, and non-exclusive, explanations for the effect noted by Lawrence and Brody, et al. 1) Because the access to the articles is unrestricted by any payment mechanism authors are able to read them more easily, and thus they cite them more frequently; the Open Access (OA) postulate. 2) Because the article appears sooner it gains both primacy and additional time in press, and is thus cited more; the Early Access (EA) postulate. 3) Authors preferentially tend to promote (in this case by posting to the internet) the most important, and thus most citable, articles; the Self-selection Bias (SB) postulate" (M.J. KURTZ, G. EICHORN, A. ACCOMAZZI, C. GRANT, M. DEMLEITNER, E. HENNEKEN, S.S. MURRAY, *The effect of use and access on citations*, "Information Processing and Management", 41 (2005), p. 1395-1402; 1396).

²² "The benefits of self-archiving for an individual article or the work of an individual author are uncertain and could be as much affected by subject area, inherent variations in publication, and citation patterns generally, and the presence and/or importance of a specialized online pre-print archive. Scientific citation is influenced, overwhelmingly, by the relevance and importance of a given scholarly work to other scholars in the field. While other factors might have moderate effects, the process of science is driven not by access, but by discovery" (I. CRAIG, A. PLUME, M. McVEIGH, J. PRINGLE, M. AMIN, *Do open access articles have greater citation impact? A critical review of the literature*, "Journal of Informetrics", 1 (2007), 3, p. 239-248; 248).

²³ "We shouldn't be surprised to see that good relevant literature that is easier to find and retrieve is cited more often than good relevant literature that is harder to find and retrieve. Or, if a careful study concluded that this view is false, then one might expect it to be

more careful in summarizing the reason why” (P. SUBER, *Publishers doubt the OA impact advantage*, “Open Access News”, 18 May 2007. Il riferimento è proprio all’articolo di Craig Plume, McVeigh, Pringle, e Amin).

²⁴ L.I. MEHO, *The rise and rise of citation analysis...*, cit., p. 36.

²⁵ “According to Ulrichsweb, about 24,000 peer-reviewed research journals exist worldwide, across all disciplines and languages, publishing about 2.5 million articles per year. But because journal prices keep rising and library budgets are limited, each university can afford only a small portion of that total. [...] even if all 24,000 peer-reviewed research journals were sold to universities at cost – i.e., with not a penny of profit – it would still be true that almost no university has anywhere near enough money to afford all or even most of the 24,000 journals” (S. HARNAD, T. BRODY, F. VALLIERES, L. CARR, S. HITCHCOCK, Y. GINGRAS, C. OPPENHEIM, H. STAMERJOHANN, E.R. HILF, *The access/impact problem and the green and gold roads...*, cit., p. 311).

²⁶ Non senza qualche polemica, come nel caso di “Journal of the Pancreas”, il cui comitato scientifico ha polemizzato con quella che a loro modo di vedere era un’esclusione immeritata (*Considerations on a “pancreatic” electronic journal at the beginning of its 8th year of publication*, “Journal of the Pancreas”, 8 (2007), 3, p. 263-267).

²⁷ “COUNTER (Counting Online Usage of Networked Electronic Resources), launched in 2002, is an international initiative to improve the reliability of online usage statistics. It is supported by the vendor, intermediary and librarian communities. COUNTER’s objective is to ensure that vendor online usage reports are credible, compatible and consistent. This is achieved by the publication of Codes of Practice that specify the content, format, delivery mechanisms and data processing rules for a set of core usage reports that are easily implemented by vendors and easily understood by librarians.”

²⁸ P.T. SHEPHERD, *The feasibility of developing and implementing journal usage factor: a research project sponsored by UKSG*, “Serial”, 20 (2007), 2, p. 117-123.

²⁹ “The basic assumptions of citation analysis are clearly false. The assump-

tion ‘that the research cited by scientists in their own papers represents a roughly valid indicator of influence on their work’ has been falsified repeatedly, as has the assumption on which it is based: that scientists are motivated to cite their influences or to give credit where credit is due. Neither of these assumptions is supported by evidence. Rather, scientists do not cite most of their influences; citing is biased, secondary sources replace primary sources, and informal sources, which are the lion’s share, are not credited. And authors definitely are not motivated to bestow credit” (M.H. MACROBERTS – BARBARA R. MACROBERTS, *Problems of citation analysis*, “Scientometrics”, 36 (1996), 3, p. 435-444; 442).

³⁰ La rilevazione del numero degli scaricamenti è stata pensata per stabilire se il prezzo pagato da un’istituzione per accedere ad un determinato periodico elettronico fosse proporzionale all’uso che ne veniva fatto.

³¹ “Download impact is useful for at least two reasons; 1) The portion of download variance that is correlated with citation counts provides an early-days estimate of probable citation impact that can begin to be tracked from the instant an article is made Open Access and that already attains its maximum predictive power after 6 months. 2) The portion of download variance that is uncorrelated with citation counts provides a second, partly independent estimate of the impact of an article, sensitive to another form of research usage that is not reflected in citations” (T. BRODY – S. HARNAD – L. CARR, *Earlier Web usage statistics as predictors of later citation impact*, “Journal of the American Society for Information Science and Technology”, 57 (2006), 8, p. 1060-1072).

³² “The papers used for testing how well downloads can be used to predict citation impact are from the high energy physics sub-ArXivs and deposited between 2000 and 2002 (a 2 year period). [...] This study found a significant and sizeable correlation of approximately 0.4 between the citation and download impact of articles in physics and mathematics. [...] The true correlation may in fact prove somewhat higher once more download

sites are monitored and automatic linking becomes more accurate. It will no doubt vary from field to field, and may also change as the proportion of Open Access content (now 10-20%) approaches 100%” (*ibidem*, p. 1068).

³³ Di parere opposto è Antonella De Robbio, secondo la quale “è altamente prevedibile che un lavoro scaricato ai fini di una fruizione, sia in un prossimo futuro citato da altri articoli e quindi gli indicatori di log statistici sono sostanzialmente indicatori di prevedibilità verso un fattore di impatto basato da una parte su una reale fruizione, dall’altra su future citazioni” (*Analisi citazionale e indicatori bibliometrici nel modello Open Access*, “Bibliotime”, 10 (2007), 2. L’articolo contiene il testo della relazione presentata alla conferenza “Institutional archives for research: experiences and projects in Open Access”, tenutasi a Roma presso l’Istituto superiore di sanità il 1° dicembre 2006).

³⁴ Una possibile spiegazione è che questi due tipi di risorsa soddisfino bisogni differenti: “The publisher’s website may be better for information discovery and browsing, especially for recent published articles. In contrast, the ArXiv may provide some competition for known article searches” (P.M. DAVIES – M.J. FROMERTH, *Does the ArXiv lead to higher citation and reduced publisher downloads for mathematics articles?*, “Scientometrics”, 71 (2007), 2, p. 203-215; 214).

³⁵ Le differenze tra i due campi derivano da: “A more pronounced national and regional orientation. Less publication in serials; more in books. A different pace of theoretical development. Single scholar approach versus team research. A greater share of publications directed at the non-scholarly public” (l’elenco è tratto da A.J. NEDERHOF, *Bibliometric monitoring of research performance in the social sciences and the humanities. A review*, “Scientometrics”, 66 (2006), 1, p. 81-100).

³⁶ “A first and good indication whether bibliometric analysis is applicable to a specific field is provided by the publication characteristics of the field, in particular the role of international, refereed journals. If international journals are the dominating or at least a major means of communication in a

field, then in most cases bibliometric analysis is applicable. Therefore it is important to study first the 'publication practices' of a research group, department, or institute, in order to establish whether bibliometric analysis can be applied. A practical measure here is the share of CI-covered publications [Citation Index] in the total research output" (A.F.J. VAN RAAN, *Challenges in ranking of universities*, relazione presentata alla First International Conference of World Class Universities tenutasi a Shanghai dal 16 al 18 giugno 2005).

³⁷ "In general, to monitor research performance at the international research frontier in social sciences and humanities, the same bibliometric methods can be applied as in science, but with several extensions. In particular, a broader range of both publications (including non-ISI journals and monographs) and indicators is needed in many social sciences and humanities" (A.J. NEDERHOF, *Bibliometric monitoring of research performance in the social sciences and the humanities*, cit., p. 96).

³⁸ A.F.J. VAN RAAN, *Fatal attraction*, cit. Van Rann parla dei molti errori contenuti nelle registrazioni delle banche dati relativamente al nome delle università. Nel caso dei dipartimenti, le cose naturalmente sono di gran lunga peggiori.

³⁹ "Appare [...] prioritaria ed essenziale la costituzione di una banca dati della ricerca scientifica universitaria che abbia come unità di rilevazione il singolo soggetto impegnato in tale attività. Una tale banca dati può essere facilmente creata avvalendosi delle procedure informatiche che il CINECA ha già predisposto per ogni docente (sito docente) e che sono già state utilizzate da molte università per la trasmissione telematica di dati in relazione a richieste di finanziamento, da attribuire previa valutazione delle documentate attività svolte dai proponenti" (MINISTERO DELL'ISTRUZIONE, DELL'UNIVERSITÀ E DELLA RICERCA. COMITATO NAZIONALE PER LA VALUTAZIONE DEL SISTEMA UNIVERSITARIO, *Proposte per la costruzione di un nuovo modello per la ripartizione "teorica" del FFO alle università statali*, 2004, p. 19-20. Il docu-

mento è riportato in allegato al decreto ministeriale 146 del 28 luglio 2004, *Approvazione nuovo modello di valutazione sistema universitario*).

⁴⁰ Una possibile soluzione a questo problema potrebbe essere quella "di far dialogare tali sistemi locali con un data base centrale (che potrebbe essere il già esistente Sito docente CINECA) a garanzia di uno standard e della normalizzazione degli input che verrebbero forniti al sistema di valutazione. Il data base centrale, interoperabile con i sistemi locali, si popolerebbe automaticamente raggiungendo l'eshaustività dei dati contenuti" (F. VALENTINI, *L'editoria elettronica e i sistemi di valutazione della ricerca nelle università*, in *Raccomandazioni per lo sviluppo dell'editoria elettronica negli atenei italiani*, a cura di P. Cotoneschi e G. Pepeu, Firenze, Firenze University Press – CRUI, 2006, p. 89).

⁴¹ Se fino ad ora i risultati raggiunti dagli archivi istituzionali sono stati relativamente scarsi lo si deve al fatto che essi nascono da iniziative non coordinate e perciò non sono riusciti a diventare parte integrante delle strate-

gie di ricerca. Stando così le cose è naturale che il loro sviluppo sia di gran lunga inferiore alle aspettative, come sottolinea l'articolo di Davis e Connolly apparso quest'anno su "D-Lib Magazine" e intitolato appunto *Evaluating the reasons for non-use of Cornell University's installation of DSpace* ("D-Lib Magazine", 13 (2007) 3/4). La soluzione proposta da Guédon è di trovare "a way to move from institutional to disciplinary and even speciality repositories. [...] In an institutional repository, the metadata should be organized in a sufficient clear and standardized fashion to allow a quick and disciplinary representation of what is available there. [...] The metadata should also be extended to provide some indication of quality" (*The "green" and "gold" roads to open access: the case for mixing and matching*, "Serial Review", 30 (2004), 4, p. 315-328; 323-324).

⁴² Wageningen Yield (<http://library.wur.nl/way/>) "is the access point to publications from Wageningen University and Research Centre. WaY contains references to publications authored by Wageningen University staff from 1976 onwards. Publications authored by the staff of the Wageningen UR Research Institutes are available from 1995 onwards. Full text documents are added when available. The database is updated daily and currently holds more than 110.000 items".

⁴³ Un esempio di come possa essere condotta un'indagine bibliometrica utilizzando Google Scholar è contenuto nell'articolo di Meho e Yang *A new era in citation and bibliometric analyses: Web of science, Scopus, and Google Scholar*, dove, a togliere l'illusione che tutto possa essere fatto automaticamente, si afferma che "GS data were harvested in March 2006; however, identifying their relevancy and full bibliographic information took approximately 3,000 hours of work over a six months period, which included manually verifying, cleaning, formatting, standardizing, and entering the data into EndNote Libraries and Access databases" ("Journal of the American Society for Information Science and Technology", 58 (2007), 13, p. 1-21; 15).

⁴⁴ D. NICHOLAS – P. HUNTINGTON – H.R.

JAMALI, *The impact of open access publishing (and other access initiatives) on use and users of digital scholarly journals*, "Learned Publishing", 20 (2007), 1, p. 11-15. Nell'articolo si parla del fatto che "Oxford University Press moved one of its most prestigious journals, 'Nucleic Acids Research' (NAR), to a full open access (OA), author-pays publishing model in January 2005. A deep log analysis study was carried out in order to determine the impact of this move to OA on the use and users of the journal. Surprisingly the findings showed that although there was a 143% increase in use from early 2003 to January 2005, it was search engines and robots that accounted for a high proportion of the increased use. Robots were responsible for half of sessions in the second quarter of 2005, compared to 1% in the second quarter of 2003".

⁴⁵ H. VAN DE SOMPEL – C. LAGOZE, *Interoperability for the discovery, use, and re-use of units of scholarly communication*, "CTWatch Quarterly", August 2007.

⁴⁶ "Although many databases are publicly available that cover scholarly output in disciplines of the social and behavioral sciences and the humanities, these databases have several properties that make them less suitable and even useless for calculating bibliometric indicators: it may be unclear which bibliographic sources are processed; criteria for selection of sources may be unclear. Database may have national or other *geographical biases*; large fractions of processed documents do *not* have institutional affiliations of publishing authors; even if documents do contain addresses of publishing authors, the database produces may not include these in the database; important data elements – even journal titles and country names – may not be standardized; to the best of our knowledge none of the major database include cited references [...]; many databases are available only through host computer that offer limited counting and other statistical facilities" (A.F.J. VAN RAAN, *Challenges in the ranking of universities*, cit.).

⁴⁷ Citebase è, secondo la definizione dello stesso autore, "a scientometric navigational and evaluational engine that can rank articles and authors on

the basis of a variety of metrics" (T. BRODY, *Evaluating research impact through Open Access to scholarly communications*, Southampton, University of Southampton, 2006. Si tratta della tesi di dottorato di Brody, disponibile all'indirizzo: <<http://eprints.ecs.soton.ac.uk>>). Attualmente esso è utilizzato solo per le registrazioni contenute in ArXiv.

⁴⁸ "If Citebase were already being applied to the worldwide POA network of IRs, and if that network contained 100% of each institution's research publication output, along with each publication's metric, this would not only maximise research access, usage and impact, as OA is meant to do, but it would also provide an unprecedented and invaluable database for scientometric datamining and analysis. OA scientometrics – no longer constrained by the limited coverage, access tolls and non interoperability of today's multiple proprietary databases for publications and metrics – could trace the trajectory of ideas, findings, and authors across time, across fields and disciplines, across individuals, groups, institutions and nations, and even across languages" (T. BRODY, L. CARR, Y. GINGRAS, C. HAJJEM, S. HARNAD, A. SWAN, *Incentiving the Open Access research Web*, "CTWatch Quarterly", August 2007).

Abstract

The author intends to demonstrate that the articles placed in the open access archives have usually better impact and how this advantage can be measured only by means of indicators based on the number of citations an article has collected. Different kinds of indicators appear less reliable. At the end of the article some applications, developed in order to measure the impact of articles published in open access archives, are described and evaluated.