

LA VALUTAZIONE DELLA RICERCA

di Paolo Rossi

Ogni prospettiva di significativo e stabile sviluppo della ricerca pubblica italiana richiederà necessariamente una stretta combinazione tra una maggior disponibilità di risorse (umane e finanziarie) e un aumento di efficienza nel loro utilizzo. Premessa indispensabile a tale aumento di efficienza è certamente l'acquisizione di una reale capacità di valutazione *ex post* dei risultati, sulla cui base impostare politiche premiali e, ove necessario, disincentivanti.

Le metodologie valutative da adottarsi sono tuttavia ancora oggetto di un vasto dibattito. È quindi necessario formulare ipotesi di lavoro, al momento evidentemente provvisorie, e metterle concretamente alla prova per misurarne la validità, sempre mantenendo una chiara consapevolezza del fatto che nel breve-medio periodo, a fronte di evidenze negative, occorrerà operare revisioni anche sostanziali della impostazione inizialmente adottata.

Prima di qualunque proposta operativa occorre però identificare le principali opzioni metodologiche e i loro possibili domini di applicazione, notando che in contesti differenti possono risultare adeguate opzioni anche molto diverse tra loro. La varietà delle situazioni cui può riferirsi la valutazione della ricerca può essere sommariamente classificata secondo i seguenti principali parametri:

- a) finalità della valutazione (finanziamenti, carriera, retribuzioni);
- b) soggetti della valutazione (individui, dipartimenti, Atenei);
- c) metodologia valutativa (*peer review*, criteri bibliometrici, metodi misti);
- d) oggetti della valutazione (articoli, monografie, brevetti, altri prodotti di ricerca);
- e) grandi aree disciplinari (scientifico-tecnologica, scienze della vita, umanistico-sociale).

In merito alle finalità della valutazione una prima e fondamentale distinzione è volta a separare le azioni che hanno come obiettivo quello di formulare graduatorie (*ranking*), in genere mediante attribuzione di punteggi (*rating*), al fine di reclutamento, avanzamento di carriera, finanziamento di progetti, ripartizione di risorse e più in generale con obiettivi di incentivazione

e premialità, dalle azioni valutative finalizzate alla pura e semplice verifica di superamento di prefissati requisiti minimi di produttività scientifica individuale (richiesti ad esempio per il conseguimento di idoneità o per la progressione stipendiale).

Non insisteremo sul secondo aspetto, pur denso di importanti conseguenze pratiche, in quanto ci pare che nella stragrande maggioranza dei casi il carattere non comparativo del giudizio consenta di formulare la valutazione sulla base di indicatori di natura prevalentemente quantitativa, ancorché certamente da scegliersi con la massima attenzione all'accuratezza e all'efficacia.

La valutazione comparativa vera e propria comporta invece un'analisi più sofisticata e impegnativa, che dovrà necessariamente fare attento e distinto riferimento a ognuno dei parametri sopra elencati. Restando fin d'ora inteso che anche la natura degli oggetti della valutazione e l'area disciplinare di riferimento comportano distinzioni talvolta cruciali e richiedono analisi appropriate e proposte specifiche, sembra tuttavia opportuno organizzare il nucleo della discussione sulla base di due soli grandi assi di riferimento, quello relativo ai soggetti della valutazione, che possono variare dal singolo studioso ai grandi numeri di un Ateneo, e quello relativo alla metodologia, che può variare con una certa continuità dalla valutazione "calda" e puramente qualitativa (*peer review*) a quella "fredda" e prevalentemente quantitativa degli indicatori bibliometrici.

Per stabilire quale relazione debba esistere tra la tipologia (e dimensione) dei soggetti valutati e il peso relativo che nella valutazione dovrebbero avere le due differenti metodologie dovremo preliminarmente esaminare in dettaglio l'origine e il significato dei criteri bibliometrici.

Notiamo che ogni criterio quantitativo trae il proprio fondamento concettuale da considerazioni di tipo statistico, sia che si tratti di un riferimento a valori medi (come nel caso dell'*Impact Factor* o di qualunque altro tipo di *ranking* di riviste o di sedi di pubblicazione) sia che si tratti di valori assoluti individuali (come per il numero

delle pubblicazioni e delle citazioni, l'indice h e simili) la cui interpretazione dipende comunque dalla comparazione con un grande numero di casi simili, ovvero dalla posizione occupata all'interno di una distribuzione. Ma nelle distribuzioni caratterizzate da una probabilità che, in percentuale, decresce al crescere del valore del parametro misurato (le cosiddette distribuzioni prive di scala) il significato della media è assai poco pregnante (non potendosi spesso definire la varianza), ed è stato più volte dimostrato, mediante il calcolo delle probabilità, che il rischio che un articolo pubblicato su una rivista di minor rango sia migliore di uno pubblicato su una rivista più "qualificata" non è in genere trascurabile. Sulla base di questa fondamentale premessa è facile argomentare che l'adozione di parametri bibliometrici (in modo esclusivo o prevalente) ai fini di una valutazione comparativa di individui è di per sé un errore concettuale, anche indipendentemente dalla qualità (spesso opinabile) del parametro stesso.

Dovrebbe quindi apparire chiaro che la valutazione della ricerca individuale, a qualunque fine essa sia effettuata, deve comunque fare necessariamente ricorso a un giudizio di merito sulla produzione scientifica, formulato da valutatori indipendenti e dotati della necessaria competenza disciplinare (*peer review*), che potranno servirsi anche dei parametri bibliometrici interpretandoli alla luce delle proprie esperienze e conoscenze (*informed peer review*). Il riferimento a parametri quantitativi da parte dei revisori, pur non potendo condizionare il giudizio qualitativo, è importante al fine di evitare che elementi di assoluta soggettività possano portare a valutazioni totalmente arbitrarie e a discrepanze gravi tra i giudizi espressi da soggetti differenti.

Viceversa, è altrettanto vero che, quando il numero dei soggetti, valutati in modo aggregato, è sufficiente a giustificare l'ipotesi che le fluttuazioni di qualità non possano incidere significativamente sull'interpretazione del dato quantitativo, le stesse leggi della statistica ci assicurano un grado di attendibilità dei risultati dell'analisi quantitativa che cresce naturalmente al crescere del numero dei valutati. La comparazione di differenti Dipartimenti relativi alla stessa disciplina, una volta opportunamente standardizzata per tener conto delle dimensioni dei Dipartimenti stessi, può essere quindi plausibilmente effettuata in modo prevalente sulla base di indicatori bibliometrici. Con opportune operazioni di nor-

malizzazione tra discipline differenti, anch'esse possibili su base statistica, l'applicazione dei metodi bibliometrici di valutazione potrebbe costituire anche la base per ripartizioni di risorse almeno parzialmente ancorate al "merito" delle singole strutture all'interno di un Ateneo dato, e anche per la formulazione di un giudizio complessivo sui singoli Atenei.

L'uso degli indicatori bibliometrici per le valutazioni collettive e aggregate non può comunque ridursi a una mera misurazione della quantità di produzione scientifica, che inevitabilmente indurrebbe a comportamenti opportunistici, privilegiando una produzione intensiva, scarsamente meditata e di conseguenza dotata di scarso impatto sulla comunità scientifica nazionale e internazionale. Per questo motivo si è andata sviluppando in alcune comunità di ricerca, soprattutto nelle aree scientifico-tecnologica e bio-medica, la ricerca di criteri e di parametri anch'essi di natura quantitativa, ma volti a misurare il grado di interesse con cui le pubblicazioni scientifiche vengono accolte. Malgrado le numerose e anche facili critiche, il conteggio del numero delle citazioni ricevute dagli articoli su rivista sembra generalmente costituire la base di partenza per la stima di questi parametri e indicatori di "impatto scientifico".

A partire dal conteggio delle citazioni individuali si è poi evoluto il concetto di *Impact Factor* (IF) di una rivista, che in sostanza consiste nella valutazione del numero medio annuo delle citazioni per articolo ricevute dagli articoli pubblicati sulla rivista stessa. Come già spiegato più sopra è tuttavia assai improprio convertire l'IF di una rivista in un giudizio sul singolo articolo, e quindi sulla qualità della produzione scientifica individuale. Ma esistono anche fondati motivi, più volte discussi nella letteratura sull'argomento, per ritenere che un'applicazione meccanica dell'IF non produca un'adeguata parametrizzazione della qualità scientifica neanche nel caso di valutazioni collettive, in quanto l'ampiezza dell'intervallo di valori empirici dell'IF (da frazioni dell'unità a diverse decine) è del tutto sproporzionata all'effettiva differenza di qualità media degli articoli pubblicati, e risponde pesantemente a logiche di diffusione editoriale e alle importanti differenze esistenti nelle dimensioni e nelle modalità comunicative delle diverse aree e settori disciplinari, con discrepanze rilevanti anche tra sottogruppi appartenenti allo stesso settore.

Nel campo delle scienze umane e sociali la situazione è resa ancor più complessa dall'assenza di basi di dati sufficientemente ampie e adeguate per il conteggio delle citazioni, e dal ruolo rilevante della produzione monografica, per la quale tale conteggio è intrinsecamente più difficile, se non addirittura impossibile, almeno con gli attuali strumenti di analisi.

La definizione di criteri bibliometrici per la valutazione della ricerca richiede quindi un ripensamento complessivo, con l'auspicabile obiettivo di identificare criteri che da un lato siano per quanto possibile omogenei tra le differenti discipline, e dall'altro non penalizzino artificialmente differenti tradizioni culturali e modalità di comunicazione scientifica.

Un'indicazione che appare emergere in misura abbastanza largamente condivisa è quella di operare una prima sommaria distinzione riconducibile in pratica alla verifica del soddisfacimento dei criteri di scientificità (presenza di revisori oppure, in subordine, esistenza di un autorevole Comitato Scientifico), così come specificati anche nelle proposte del C.U.N.

Per quanto riguarda poi specificamente gli articoli su rivista dovrà essere valutata con attenzione l'ipotesi di una classificazione sintetica (e sufficientemente elastica) che identifichi con procedure ampiamente condivise, all'interno dell'insieme delle riviste ritenute scientifiche, differenti livelli di qualificazione, garantendo comunque un'equilibrata ripartizione delle riviste tra i suddetti livelli. Il "punteggio" qualitativo da attribuire ai diversi livelli dovrebbe in ogni caso essere riferito a un intervallo di valori non troppo ampio.

La determinazione di questo *pool* di riviste dovrebbe comunque risultare da un processo interattivo, nel quale le comunità scientifiche, anche tramite le proprie associazioni, giochino un ruolo determinante, e nel quale i criteri di riferimento, oltre quelli già in precedenza definiti in relazione alla scientificità, siano l'impatto (nazionale e meglio ancora internazionale) della rivista, la presenza nelle principali banche dati internazionali, l'autorevolezza della direzione scientifica, l'affidabilità della gestione organizzativa.

Per quanto riguarda le monografie non sembra oggi possibile immaginare criteri di classificazione "oggettiva" del loro impatto scientifico che si basino sulla loro collocazione editoriale e su un qualche tipo di *ranking* delle Case Editrici,

una pratica del tutto sconosciuta in tutti i Paesi che hanno affrontato il problema della valutazione della produzione scientifica.

Qualora si intenda, in sede valutativa, graduare in qualche misura il valore delle opere pubblicate in volume sembra quindi ineludibile stabilire, anche nel caso di valutazione aggregata, una qualche forma di giudizio comparativo mediante *peer reviewing*, malgrado l'impegno straordinariamente più elevato insito in tale procedura.

Restando dunque confermato che una valutazione "fredda" delle strutture è in linea di principio possibile, e per certi aspetti anche auspicabile (soprattutto in relazione a costi, tempi e sistematicità della valutazione stessa), l'impegno valutativo sarebbe comunque sostanzialmente sterile qualora esso non fosse accompagnato da meccanismi capaci di tradurre le differenze di produttività scientifica così misurate in un effettivo aumento delle risorse per i più meritevoli, erogato a partire dal livello dipartimentale: meccanismi incentivanti o disincentivanti che operino soltanto al livello degli Atenei ben difficilmente otterrebbero effetti positivi sul comportamento delle singole aree e settori scientifici. Una strategia di incentivazione dipartimentale è anche preconditione per una concreta valorizzazione dei nuclei d'eccellenza esistenti anche in quelle realtà che, per ragioni storiche, territoriali e/o organizzative non fossero in grado di raggiungere nel loro complesso i livelli qualitativamente più elevati di produttività scientifica.

Le soluzioni proposte fino a questo punto sono comunque da considerarsi largamente contingenti. Se infatti proviamo a collocarci in una prospettiva temporale più ampia possiamo facilmente immaginare che nel medio periodo (ovvero nei prossimi 5-10 anni) l'evoluzione delle modalità di comunicazione, anche scientifica, associata alla diffusione e allo sviluppo di Internet e dell'editoria elettronica, e soprattutto delle moderne forme di interazione collettiva (*social network* e simili) identificate anche dalla locuzione Web 2.0, porterà a nuovi meccanismi valutativi, al momento ancora in embrione nel campo della ricerca, ma già efficaci in altri campi di attività. Ci riferiamo ad esempio all'idea del "controllo aperto" lanciata da Nature, e consistente nell'idea di "esporre" i manoscritti per un certo tempo su Internet, con la possibilità di essere commentati da tutti i lettori interessati, prima che il comitato editoriale decida di "pubblicarli"

(probabilmente soltanto in forma elettronica), anche sulla base dei giudizi che sono stati espressi. Possiamo anche aspettarci sostanziali raffinamenti di quella forma di controllo *ex post* che è data dalle citazioni, se si supererà l'attuale meccanismo centralizzato (fondato quasi esclusivamente su basi di dati proprietarie) in favore di un'analisi a tutto campo mediante motori di ricerca (sul modello, per quanto certamente ancora molto rozzo, proposto da Google *Scholar*), che permetterebbe facilmente di estendere questo tipo di valutazione anche alle monografie,

soprattutto se le forme di pubblicazione elettronica, come prevedibile, andranno generalizzandosi, auspicabilmente con modalità *open access*.

In quest'ottica è più che mai importante mantenere un atteggiamento aperto nei confronti dei meccanismi e delle regole per la valutazione, evitando irrigidimenti normativi e burocratici che rischierebbero di risultare superati in un breve arco di tempo, e quindi di frenare pericolosamente il pieno dispiegarsi delle potenzialità offerte alla ricerca realmente creativa dalle nuove forme della comunicazione scientifica.

PAOLO ROSSI

Contatti:

Dipartimento di Fisica "E. Fermi"

Email: rossi@df.unipi.it

Largo Pontecorvo, 3

56127 Pisa