



Ministero dell'Istruzione dell'Università e della Ricerca
Consiglio Universitario Nazionale

Prot. n. 612

All'On.le Ministro
SEDE

Spedito il 30/3/2010

OGGETTO: Valutazione della ricerca.

Adunanza del 25.03.2010

IL CONSIGLIO UNIVERSITARIO NAZIONALE

APPROVA IL SEGUENTE DOCUMENTO:

Introduzione

Ogni prospettiva di significativo e stabile sviluppo della ricerca pubblica italiana passa necessariamente attraverso una stretta combinazione tra disponibilità di risorse finanziarie e umane e 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 della ricerca, sulla cui base impostare politiche premiali e di incentivazione/disincentivazione.

Le metodologie valutative, sulle quali è comunque indispensabile che si raggiunga un convinto e largo consenso, sono ancora oggetto di un vasto dibattito. Sarà 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, potranno imporsi revisioni anche sostanziali della impostazione inizialmente adottata.

In ogni caso è necessario premettere all'esposizione di qualunque proposta operativa una serie di considerazioni volte a 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.

Premessa generale

La varietà delle situazioni cui può riferirsi la procedura di 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) aree disciplinari (scientifico-tecnologica, scienze della vita, umanistico-sociale)

In riferimento al punto a) (finalità della valutazione) occorre operare una prima e fondamentale distinzione tra le azioni volte a stabilire, anche mediante attribuzione di punteggi (*rating*), una vera e propria valutazione comparativa (*ranking*), sia tra individui sia tra istituzioni o strutture, al fine di reclutamento, avanzamento di carriera, finanziamento di progetti, ripartizione di risorse e più in generale con obiettivi di incentivazione e premialità, e le azioni valutative che sono invece finalizzate alla pura e semplice verifica di superamento di prefissati requisiti minimi di produttività scientifica individuale, richiesti per il conseguimento di idoneità o per la progressione salariale.

Il C.U.N. si è già occupato separatamente di questo secondo aspetto, la cui problematica riveste carattere di minore complessità, riducendosi nei casi più semplici alla formulazione di “criteri identificanti il carattere scientifico delle pubblicazioni” (secondo il dettato dell'art.3-ter comma 2 del D.L.10 novembre 2008 n. 180 convertito con L.9 gennaio 2009 n.1) e all'individuazione di livelli minimi (per lo più quantitativi) di produzione scientifica.

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 i parametri d) ed e) comportano distinzioni talvolta cruciali e richiedono analisi appropriate e proposte specifiche, sembra tuttavia opportuno organizzare in primo luogo la discussione sulla base di due soli assi di riferimento, quello relativo al parametro b), variabile dall'individuo ai grandi numeri (migliaia di docenti) di un ateneo, e quello relativo al parametro c), variabile 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 non possiamo dimenticare che nelle distribuzioni caratterizzate da una probabilità che, in percentuale, decresce con il 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 individuale è di per sé un errore concettuale, anche indipendentemente dalla qualità (spesso opinabile) del parametro stesso.



Viceversa, è altrettanto vero che, quando il numero dei soggetti, valutati in modo aggregato, è sufficientemente elevato, le stesse leggi della statistica ci assicurano un grado di attendibilità dei risultati dell'analisi quantitativa che cresce (anche se non linearmente) 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 normalizzazione 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.

Ma la condizione *sine qua non* per l'attribuzione di un peso prevalente ai parametri bibliometrici è la presenza di un numero di soggetti sufficiente a giustificare l'ipotesi che le fluttuazioni di qualità non possano incidere significativamente sull'interpretazione del dato quantitativo.

Restando dunque confermato che una valutazione bibliometrica 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/disincentivanti e premiali operanti 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.

Criteria per la valutazione della ricerca

La valutazione dell'attività di ricerca individuale, a qualunque fine essa sia effettuata, deve comunque fare 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.

L'uso degli indicatori bibliometrici per le valutazioni collettive e aggregate, pur con le restrizioni al dominio di applicazione specificate nella premessa generale, non può comunque ridursi a una mera misurazione quantitativa della 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 in premessa, è 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 importante punto di partenza è fornito dal documento conclusivo prodotto dal gruppo di esperti investito del problema dal Consiglio Scientifico generale del C.N.R., e interessanti spunti sono offerti anche dal gruppo di lavoro costituitosi nell'ambito delle discipline umanistiche.

Un'indicazione che appare emergere in misura largamente condivisa è quella di operare una prima sommaria distinzione riconducibile quasi completamente alla verifica del soddisfacimento dei criteri di scientificità, così come specificati anche nelle proposte del C.U.N.

Per quanto riguarda poi specificamente gli articoli su rivista, mentre è da escludere, per tutti i settori disciplinari, l'ipotesi di un *ranking* assoluto e puntuale di tutte le riviste (quale sarebbe quello offerto dall'IF), deve 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, coordinato dal C.U.N., 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 (eventualmente misurato dall'IF nei settori in cui questo criterio risulti appropriato),
- la presenza nelle principali banche dati internazionali (tra cui la banca dati I.S.I. per i settori in cui la copertura da essa offerta risulti adeguata),
- 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 (inclusi gli Atti di convegni), al di là dei meri criteri di scientificità già contenuti nella proposta del C.U.N., 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.

Se però 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", 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.

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.

IL SEGRETARIO
(firmato Valeo)

IL PRESIDENTE
(firmato Lenzi)