

(META)VALUTARE LA RICERCA

di Anna Giunta, Luca Salvatici
03 febbraio 2012

Una leggenda metropolitana (si spera) narra di un commissario che si vantava di non aver letto i titoli dei candidati a un concorso "per evitare di farsi influenzare". Un simile comportamento potrebbe, però, rivelarsi quello più efficiente per quanto riguarda la valutazione dei prodotti di ricerca coinvolti nella prossima VQR 2004-2010.

In vista della ormai imminente Valutazione della Qualità della Ricerca (VQR) condotta dall'Anvur (Agenzia Nazionale per la Valutazione del Sistema Universitario e della Ricerca), Nel Merito ha già ospitato alcuni interventi (Menegatti, 18 Novembre 2011; Barucci, 25 Novembre 2011) interessanti. Ci sembra che ormai stia emergendo all'interno della comunità scientifica italiana una cultura condivisa della responsabilità per i risultati conseguiti. Tuttavia, se nessuno sembra mettere in discussione l'opportunità della valutazione, il dibattito è assai vivace riguardo alle modalità della valutazione stessa.

Tuttavia, prima di discutere sulle metodologie di valutazione, occorre decidere cosa valutare. Come ricorda Barucci, la valutazione della ricerca può avere due obiettivi principali: la valutazione delle strutture, con conseguenze sull'assegnazione delle risorse, e la valutazione dei singoli ricercatori, con conseguenze sul reclutamento e sulla carriera. In entrambi i casi, l'unità di osservazione sono i prodotti di ricerca individuali (nella maggior parte dei casi articoli o monografie), ma poiché il VQR 2004-2010 non intende valutare i singoli ricercatori, riteniamo che sarebbe opportuno adottare, dove possibile, come oggetto della valutazione la qualità e della sede editoriale di pubblicazione piuttosto che la pubblicazione in sé. Si tratterebbe di una 'metavalutazione', ovvero di una valutazione delle sedi di pubblicazione piuttosto che dei prodotti in sé, dove l'indicazione di qualità attribuita alla rivista/collana editoriale verrebbe trasferita in modo automatico ai prodotti che vi sono contenuti.

L'adozione di un simile approccio sarebbe difficilmente difendibile qualora si trattasse di valutare la produzione scientifica dei singoli ricercatori. Infatti, pur ipotizzando di aver misurato esattamente la qualità media di una rivista, è evidente che vi sarà inevitabilmente una significativa variabilità nella qualità dei singoli articoli pubblicati. Tale variabilità rappresenta l'errore insito in qualsiasi metavalutazione, ma tale errore è destinato a ridursi all'aumentare del campione oggetto di valutazione. La metavalutazione, quindi, non rappresenta la soluzione più accurata, ma potrebbe essere una soluzione più efficiente quando si tratta di valutare enti e strutture di ricerca.

Se si guarda alla passata esperienza della Valutazione Triennale della Ricerca (VTR) 2001-2003, si rileva che: ci sono voluti tre anni per espletare l'intera procedura; il primo Comitato di Indirizzo per la Valutazione della Ricerca - CIVR - ha consegnato la relazione finale nel 2006; il costo è stato di 3,5 milioni di euro; ha coinvolto 150 esperti e 6000 valutatori per valutare 18500 prodotti. Il VQR 2004-2010 si prospetta come un esercizio ancor più ambizioso: circa 210000 prodotti che dovranno essere gestiti dai 450 componenti del Gruppo di Esperti della Valutazione (GEV). Ciascun componente, quindi, avrà in media un "carico" di quasi 500 prodotti e, se si rispetterà l'obiettivo di valutarne tramite peer review almeno la metà (più uno), inviando a ciascun revisore un numero di prodotti compreso tra 10 e 20 (valori ragionevoli ma tutt'altro che piccoli), ci sarebbe bisogno di un "esercito" di valutatori oscillante fra le 10000 e le 20000 unità.

La nostra tesi, quindi, è che una metavalutazione ridurrebbe significativamente i costi e i tempi della valutazione: nelle attuali condizioni di bilancio e con la necessità di abbreviare sensibilmente i tempi (il MIUR nella distribuzione della quota premiale del Fondo di Finanziamento Ordinario utilizza le informazioni della VTR che risalgono ormai a 10 anni fa!), oculatezza e tempestività ci sembrano virtù di non poco conto. E' meglio ricevere in tempi ragionevolmente rapidi una fotografia abbastanza accurata della qualità attuale della ricerca svolta nelle diverse sedi, piuttosto che un'immagine dai contorni più definiti (ma assai più costosa) di quello che le strutture erano nel passato.

Peraltro, sulla base dei numeri sopra citati e della capacità di carico dei valutatori, è lecito attendersi che molti di loro saranno in qualche modo costretti ad utilizzare tutte le informazioni in loro possesso per portare a termine il proprio compito in tempi ragionevoli. Poiché la sede della pubblicazione non è "blind", è ragionevole aspettarsi che lo status della rivista (eccellente, mediocre, e così via) finisca per influenzare il

giudizio dei valutatori sui prodotti in esame. Al termine di un lungo e costoso processo di peer review, quindi, potremmo scoprire che le valutazioni soggettive sono altamente correlate con il ranking bibliometrico delle sedi di pubblicazione, e questo renderebbe palese l'inefficienza della procedura di valutazione in corso.

A scanso di equivoci, non riteniamo che la metavalutazione sia una panacea universalmente. Evidentemente, essa è difficilmente applicabile a quei settori dove i prodotti di ricerca sono rappresentati da monogra e in lingua nazionale, edizioni critiche, risultati di scavi, lessici, cataloghi di mostre o curatele, ovvero tutti quei prodotti per cui è difficile ipotizzare una tassonomia chiusa (prima ancora che un ordinamento secondo criteri di qualità) delle sedi di pubblicazione. Tuttavia andrebbe adottata là dove possibile, e certamente nel caso dell'Area 13 (Scienze economiche e statistiche) dove la produzione scientifica è prevalentemente su base pubblicistica e può essere valutata sulla base della rilevanza e qualità della sede editoriale di pubblicazione. A questo punto, però, si pone comunque la questione di come valutare tali sedi.

Rispetto alla tradizionale contrapposizione tra peer review (soggettiva) e valutazione bibliometrica (oggettiva), la metavalutazione è, almeno in linea di principio, compatibile con entrambi gli approcci o loro combinazioni. Da una parte, infatti, si potrebbe immaginare di chiedere ai revisori di fornire un'opinione su di una rivista nel suo insieme (invece che su singoli articoli in essa contenuti). Dall'altra, si può scegliere di fare riferimento a indici bibliometrici che, in effetti, sono stati pensati proprio per valutare l'impatto (che non sempre e non necessariamente coincide con la qualità) delle riviste piuttosto che di specifici articoli.

La nostra proposta è di adottare un paradigma del tipo struttura-condotta-performance. La valutazione della “struttura” si baserebbe sulla qualità dei componenti degli organi della rivista (redazione, comitato scientifico, referee). La valutazione della “condotta” prenderebbe in considerazione elementi come la modalità con cui gli articoli sono sottoposti a referaggio; i tempi di risposta agli autori; la puntualità nella pubblicazione dei numeri; la predisposizione di un archivio dei referaggi. Infine la “performance” potrebbe essere basata, oltre che su parametri di diffusione (numero di abbonamenti eventualmente pesati sulla base della tipologia più o meno istituzionale) sui tradizionali indicatori bibliometrici. Su questi indici esiste un'ampia letteratura che ha evidenziato pregi e difetti dell'impact factor e delle sue possibili alternative, ma è bene ricordare il monito di Barucci che l'adozione “di un puro criterio meccanico può portare a risultati non sempre soddisfacenti ed è soprattutto illusorio che esista in modo oggettivo”.

La compilazione di classifiche delle riviste e degli editori per valutare i prodotti è una responsabilità della comunità scientifica nazionale (nelle sue possibili articolazioni istituzionali: CRUI, CUN, società scientifiche, ANVUR, nuclei di valutazione…) eventualmente integrata dall'esperienza internazionale (utilizzando, ad esempio, la lista del CNRS francese, come suggerito da Baccini: <http://www.roars.it/online/?p=2949>) per mettersi al riparo dai comportamenti opportunistici delle ‘baronie' che hanno ingessato l'università e la ricerca italiana votandole al declino. In nessun caso, però, andrebbe demandata ad aziende private che adottano criteri poco trasparenti e non sono responsabili nei confronti di coloro che subiscono le conseguenze delle loro scelte. A questo proposito è emblematico il caso delle due banche dati più importanti per quanto riguarda la classificazione delle riviste, prescelte dell'Anvur: Web of Science (WoS) e Scopus. Se queste ultime verranno adottate come fonti esclusive, come sembra si sia deciso nell'ambito della VQR, le scelte di inclusione o meno nelle proprie banche dati operate da Thomson, nel caso di WoS, e Elsevier nel caso di Scopus incideranno sull'affidabilità degli indici bibliometrici e sulla stessa consistenza e distribuzione geografica dell'editoria scientifica: questo, almeno nel caso di Elsevier, configura tra l'altro un macroscopico conflitto di interessi.

Da questo punto di vista, riteniamo che qualche insegnamento possa essere tratto dal recente dibattito sulle conseguenze dell'operato delle agenzie di rating sui mercati finanziari. L'analogia è evidente se si pensa che, in entrambi i casi, abbiamo a che fare con società private che hanno come obiettivo la valutazione (di titoli finanziari o di prodotti di ricerca). Nel caso delle agenzie di rating, è stata la regolamentazione pubblica a dare legittimazione al loro operato, rendendole così centrali per alcuni mercati finanziari: l'obiettivo era quello di imporre comportamenti prudenziali (divieto di acquisto di titoli che non hanno un certo rating) ad alcuni investitori istituzionali (ad esempio i fondi pensione), ma vicende come quella della Lehman Brothers hanno dimostrato che i comportamenti rischiosi da contrastare possono caratterizzare le stesse agenzie di rating.

Allo stesso modo avvertiamo forte il rischio che il VQR 2004-2010, con l'obiettivo di vincolare il comportamento dei valutatori, finisca per dare legittimazione a soggetti che perseguono finalità perfettamente comprensibili per aziende private ma non necessariamente coerenti con una politica della ricerca nazionale

(comunque essa sia definita). A questo proposito va ricordato che si tratterebbe di una legittimazione largamente immotivata in quanto i processi di selezione per l'inclusione in WoS e Scopus non si avvicinano neanche lontanamente agli standard ben descritti da Menegatti con riferimento alle “grandi riviste scientifiche internazionali”. Quale opinione avremmo di una rivista che vantasse di far valutare meticolosamente gli articoli da revisori eccellenti ma li rifiutasse senza fornire alcuna spiegazione (questo è il caso di WoS)? Oppure come giudicheremmo una rivista che respingesse un articolo dopo due anni solamente perchè manca l'abstract e non fosse disposta a riconsiderarlo prima di altri due anni (questo è il caso di Scopus)? Si tratta di pratiche discutibili, oltre che opache, che rischiano di essere (ingiustificatamente) avallate dalle scelte fatte dall'Italia, come da altri paesi, in sede di valutazione della ricerca.

Allargare l'insieme dei data base di riferimento, includendo, per esempio, anche RePec, avrebbe il vantaggio di fornire un indicatore bibliometrico sull'impatto esercitato da un maggior numero di riviste. L'indicatore prodotto da RePec non è esente da difetti e dunque la proposta non è di assumerlo in blocco, quanto di mettere a punto un algoritmo di impatto che abbia natura composita, che tenga conto di data base diversi. In ogni caso, bisognerebbe tenere presente che la valutazione non ha solamente un obiettivo di astratta conoscenza di ciò che è successo nel passato, quanto e soprattutto quello di indirizzare (sperabilmente in una direzione condivisibile) il futuro della ricerca svolta nelle strutture italiane.

Da questo punto di vista è bene ricordare che vi sono solamente 6 riviste italiane di economia in Scopus e, addirittura, 3 in WOS. Se l'Anvur ritiene che nelle discipline economiche, così come nel campo delle scienze “dure”, non ci sia alcun motivo per mantenere un dibattito nazionale (dove l'aggettivo riguarda tanto la lingua quanto l'oggetto del dibattito), il messaggio che non ha alcun valore scientifico pubblicare sulla maggior parte delle riviste italiane di economia è del tutto coerente. Se così non fosse, però, bisognerebbe fare attenzione che un simile messaggio non passi attraverso le scelte operate, in modo più o meno consapevole, da altri soggetti fuori dal suo controllo.