

# Archivi istituzionali e valutazione della ricerca: l'importanza dei dati e della loro validazione

**Paola Galimberti**

*Università degli Studi di Milano*

## *Abstract*

La valutazione della ricerca sta assumendo un ruolo sempre più importante nella governance degli atenei. Ci sono problemi riguardanti la duplicazione e l'incompletezza dei dati, che non vengono risolti dal database ministeriale, il Sito Docente del MIUR. Le istituzioni si trovano dunque di fronte al problema di organizzare le diverse fonti informative a loro disposizione in modo che possano colloquiare fra di loro, che contengano tutti i dati anagrafici e legati alle affiliazioni e che descrivano la produzione scientifica dell'istituzione sia quantitativamente che qualitativamente rispetto a tutte le tipologie di output, possibilmente secondo standard internazionali. Alcuni Atenei italiani (fra cui l'Università degli Studi di Milano) hanno avviato il processo di collegamento dei dati tratti dagli archivi istituzionali alla valutazione della ricerca attraverso l'uso di SURplus, il CRIS proposto da CILEA, con buoni risultati, poiché l'utilizzo dell'archivio istituzionale come anagrafe della ricerca ma anche come database bibliometrico minimizza le differenze tra le diverse aree e le diverse tipologie di lavori di ricerca, consentendo elaborazioni non possibili in altri casi e con gli strumenti tradizionali.

Research assessment is gaining a relevant role in University governance. The main problems concern data duplication and incompleteness, that are not fixed by the "Sito Docente" database by Italian Ministry of Education and Research. Institutions use different information software to exchange data and fully describing scientific output, according to international standards. Some Italian universities (included the University of Milan) have started a process for linking data from institutional repositories to the research assessment, using SURplus, the CRIS by CILEA. They have good results, since using the institutional archive as the research output database as well as source of bibliometric data, minimizes differences among different areas and kinds of scientific output, allowing statistic results that other tools can not provide.

*Keywords:* Ricerca, Valutazione, CRIS, SURplus.

## **La valutazione della ricerca: in primo luogo una problema di disponibilità dei dati**

Anche in Italia la valutazione della ricerca sta assumendo un ruolo sempre più importante nella governance degli atenei e da esercizio confinato ai percorsi di carriera è stata estesa ai gruppi, alle strutture di ricerca e alle istituzioni ed include, almeno potenzialmente, l'attesa dell'impatto dei risultati di ricerca sul contesto sociale ed economico.

Perché i gruppi, le strutture, i singoli possano essere valutati occorrono dati affidabili, completi ed esaustivi. Raccogliere dati ai fini della valutazione è un processo lungo e costoso che necessita di validazione sia rispetto alla forma che rispetto al contenuto.

Le istituzioni di ricerca hanno da tempo accettato di affidare questo compito a operatori commerciali che hanno costruito un monopolio e

che rivendono a caro prezzo dati sulla cui elaborazione le istituzioni non hanno alcun controllo. Inoltre Database come WoS o Scopus forniscono una rappresentazione solo parziale della produzione scientifica delle istituzioni di ricerca e da ciò deriva che le analisi che si possono fare su questi dati sono parziali: sia per quanto riguarda le tipologie prese in esame che per quanto riguarda le discipline coinvolte.

Spesse volte la confrontabilità è limitata dal linguaggio usato, dalle modalità di divulgazione, dalle politiche editoriali. Si spende molto per avere dati parziali che sono poi difficili da comparare (soprattutto per quanto riguarda le aree delle scienze umane e sociali).

Ci sono problemi di deduplicazione e di affiliazione per quanto riguarda i nomi degli autori e le attribuzioni che solo gli autori possono risolvere, lavorando alla pulizia di dati che poi verranno rivenduti dal gestore della

base dati alla loro istituzione. Esistono alternative ai database commerciali (Google Scholar ad es., un altro monopolio) ma forniscono dati troppo "sporchi" e dunque inutilizzabili.

I dati che servono per la valutazione sono di vario genere: dati bibliografici, dati amministrativi, dati anagrafici, economici. Essi sono disseminati in una molteplicità di banche dati che spesso non colloquiano fra di loro. Raccolgerli e metterli insieme per un esercizio di valutazione è oneroso in termini di tempo e di costi.

Il database del Ministero, il sito docente del MIUR, risulta poco completo e i dati esposti sono piuttosto inaffidabili perché caricati direttamente dai docenti senza particolari regole e non validati.

Le istituzioni si trovano dunque di fronte al problema di organizzare le diverse fonti informative a loro disposizione in modo che possano colloquiare fra di loro, che contengano tutti i dati anagrafici e legati alle affiliazioni e che descrivano la produzione scientifica dell'istituzione sia quantitativamente che qualitativamente rispetto a tutte le tipologie di output, possibilmente secondo standard internazionali.

### Per un uso non convenzionale dell'archivio istituzionale

Gli archivi istituzionali sono collezioni digitali dei prodotti della ricerca liberamente accessibili a tutti in cui le informazioni relative alla produzione scientifica dei ricercatori vengono raccolte in maniera centralizzata, secondo standard internazionali<sup>1</sup> ed esposti ai motori di ricerca attraverso protocolli che permettono l'interoperabilità e la comparabilità sia a livello nazionale che a livello internazionale (OAI-PMH<sup>2</sup>). Lo scopo principale per cui sono creati è quello di documentare in maniera permanente, in un unico punto, l'attività di ricerca svolta all'interno dell'ateneo. Essi fungono da indicatori tangibili della qualità della ricerca di una istituzione aumentandone la visibilità, il prestigio e il valore pubblico<sup>3</sup>. Sono più di 1400

gli archivi istituzionali nel mondo e in Italia una cinquantina di atenei ne posseggono uno<sup>4</sup>. Un archivio istituzionale dialoga con gli altri archivi istituzionali e con i principali motori di ricerca sia specialistici che generalisti<sup>5</sup> che indicizzano i metadati e li rendono disponibili per la metaricerca.

Ultimamente si è cominciato a vedere gli archivi istituzionali come contenitori di una grande quantità di dati non solo bibliografici, e ad utilizzarli come nucleo centrale nei sistemi di CRIS (Current Research Information System). Questi sistemi integrano dati provenienti da fonti diverse (progetti, finanziamenti, anagrafiche, repertori, competenze) e costruiscono statistiche, report e modelli per la valutazione dell'attività scientifica di una struttura basandosi appunto sull'output (vale a dire le pubblicazioni scientifiche).

Ovviamente l'esistenza di un archivio è necessaria ma non sufficiente perché le informazioni contenute possano essere usate come fonte di dati per la valutazione. L'archivio infatti restituisce, nel bene e nel male, ciò che contiene ovvero ciò che è stato inserito, potendo quindi presentare tutti i difetti e i limiti segnalati per i data base: copertura, errori, ambiguità, ecc. C'è necessità di una policy chiara e condivisa rispetto agli oggetti da depositare e ai metadati necessari a descriverli, ai soggetti titolati a depositare i propri lavori, alla tempistica entro cui depositare, all'obbligo di deposito almeno dei metadati e preferibilmente anche dei full-text, alla definizione di un workflow che preveda la validazione dei metadati prima della loro pubblicazione.

La presenza dei full-text negli archivi (compatibilmente con i vincoli posti dalla legge sul diritto d'autore ed in particolare dai contratti di edizione sottoscritti dagli autori) permette l'applicazione di metriche del web (Web impact factor, Usage factor, numero dei download, ecc) che tracciano il reale uso di un lavoro e il suo livello di penetrazione nella comunità scientifica (non solo disciplinare, ma in maniera interdisciplinare) e che, essendo svincolate da modelli prestabiliti, possono essere riaggreate ad hoc a seconda delle

<sup>1</sup> Ad esempio Dublin Core <http://dublincore.org/>

<sup>2</sup> <http://www.openarchives.org/OAI/openarchivesprotocol.html>  
Il protocollo OAI PMH è uno standard mediante il quale i service provider interrogano i data provider per prelevarne i metadati contenuti e creando su questa base servizi a valore aggiunto.

<sup>3</sup> C. Lynch, (2003) sottolinea come i contenuti digitali in un sistema non coordinato siano a rischio. Nelle istituzioni vi è una grande massa di informazioni digitali decentrate che devono essere in qualche modo raccolte condivise e preservate dall'organizzazione. L'archivio istituzionale nasce dunque anche con questo fine.

<sup>4</sup> Dati OpenDoar dicembre 2010 [www.opendoar.org](http://www.opendoar.org) Gli archivi italiani non hanno tutti il medesimo livello di copertura temporale e disciplinare. Ci sono atenei in cui l'archivio funge anche da anagrafe della ricerca e in cui almeno il deposito dei metadati è obbligatorio, ci sono atenei che hanno reso obbligatorio il deposito solo di alcune tipologie di pubblicazioni (ad esempio le tesi di dottorato), ci sono atenei in cui il deposito è lasciato all'iniziativa dei ricercatori.

<sup>5</sup> Google, Yahoo, ma anche Google Scholar, Scientific Commons, Base, Oaister.

diverse esigenze valutative, cancellando (in parte) tutta una serie di comportamenti adattivi e opportunistici tipici di chi è abituato ad essere valutato sulla base degli indicatori tradizionali (IF, ecc.).

Nel 2009 il gruppo Open Access della Commissione Biblioteche della CRUI ha pubblicato le raccomandazioni su OA e la valutazione dei prodotti della ricerca scientifica. Le raccomandazioni invitano alla creazione di una rete di archivi istituzionali liberamente accessibili, che colloqui con le anagrafi della ricerca locali e con quella centrale secondo standard che rendano i dati confrontabili a livello nazionale e internazionale. Esse, pur offrendo un modello che ha come riferimento la prassi in ambito europeo, si inseriscono male nella frammentata realtà italiana, perché non rispondono con soluzioni chiare alle esigenze di validazione e utilizzo dei dati. Aprono tuttavia la strada per una serie di percorsi possibili fra cui interventi sugli oggetti utilizzabili per l'analisi in termini di universo coperto, elaborazioni sui contenuti testuali, aggiornamento e tracciabilità, una maggior confrontabilità tra aree scientifiche e la sperimentazione di nuovi indicatori (metriche del web), una esplorazione più sistematica dei contesti web-based, la possibilità di riaggregare i dati rispetto alle esigenze contingenti.

Alcuni atenei italiani (fra cui l'Università degli Studi di Milano) hanno avviato il processo di collegamento dei dati tratti dagli archivi istituzionali alla valutazione della ricerca (si tratta degli atenei che hanno adottato la suite Surplus, il CRIS proposto da CILEA<sup>6</sup>). Da una prima analisi delle esperienze in atto è possibile ravvisare un percorso comune:


- 1) completezza e esaustività dei dati da verificarsi attraverso la validazione di personale esperto anche nel caso di inserimento da parte dei ricercatori;
  - 2) utilizzo dei dati ai fini della valutazione sperimentando (e condividendo) nuove procedure soprattutto per quanto riguarda le discipline "problematiche";
  - 3) aumento delle competenze e della consapevolezza di valutatori e committenti;
- ed è possibile trarre alcune conclusioni provvisorie:
- l'utilizzo dell'archivio istituzionale come anagrafe della ricerca ma anche come database bibliometrico minimizza le differenze tra le diverse aree e le diverse tipologie di lavori di ricerca, consentendo

elaborazioni non possibili in altri casi e con gli strumenti tradizionali;

- la presenza di dati "buoni" e "oggettivi" variamente aggregabili a seconda delle diverse esigenze valutative neutralizza parte dei comportamenti opportunistici legati all'uso di alcuni indici bibliometrici.

Siamo solo agli inizi, tuttavia una maggiore attenzione da parte degli atenei verso un uso generalizzato degli archivi (come del resto già viene fatto in altri paesi europei) e una armonizzazione nelle modalità e nella forma della raccolta dei dati potrebbe rappresentare in prospettiva una reale alternativa al monopolio degli operatori commerciali, riportando il controllo sui dati della ricerca nelle mani di chi la produce e permettendo un confronto in ambito nazionale e internazionale anche per quelle aree per cui attualmente la comparabilità non è garantita.

<sup>6</sup> <http://www.cilea.it/index.php?id=surplus>



Univ. Studi di Milano  
Report dei Prodotti della Ricerca con la relativa valutazione filtrati per area di appartenenza e anno  
Pagina 1

Pubblicazioni per l'area AREA 01 - SCIENZE MATEMATICHE nell'anno 2007

Report dei Prodotti della Ricerca con la relativa valutazione filtrati per area di appartenenza e anno

AREA	PUNTEGGIO
Stratigraphy across the Cimmerian unconformity in Eastern Alborz (Neka Valley, Iran): Late Cretaceous glauconitic facies as indicator of a geodynamic event / F. Berra, A Zanchi, M. Mattei, N. Marinoni, A. Nawab. EGU General Assembly 2007, Vienna.	3.20
Stratigraphy across the Cimmerian unconformity in Eastern Alborz (Neka Valley, Iran): Late Cretaceous glauconitic facies as indicator of a geodynamic event / F. Berra, A Zanchi, M. Mattei, N. Marinoni, A. Nawab. EGU General Assembly 2007, Vienna.	1.60
On the evolution equations of mean geometric densities for a class of space and time inhomogeneous stochastic birth-and-growth processes / Vincenzo Capasso, Elena Villa. - (Lecture notes in mathematics - CIME subseries ; 1892). - In: Stochastic Geometry / A. Baddeley, I. Barany, R. Schneider, W. Weil ; [a cura di] W. Weil. - Berlin Heidelberg : Springer Verlag, 2007. - ISBN 978-3-540-38174-7. - p. 267-281.	5.60
On the evolution equations of mean geometric densities for a class of space and time inhomogeneous stochastic birth-and-growth processes / Vincenzo Capasso, Elena Villa. - (Lecture notes in mathematics - CIME subseries ; 1892). - In: Stochastic Geometry / A. Baddeley, I. Barany, R. Schneider, W. Weil ; [a cura di] W. Weil. - Berlin Heidelberg : Springer Verlag, 2007. - ISBN 978-3-540-38174-7. - p. 267-281.	5.60
Multiplier methods for mixed type equations / Kevin R. Payne. - In: International Journal of Applied Mathematics & Statistics. - ISSN 0973-1377. - 8:M07(2007). - p. 58-75.	10.00

Fig. 1 - Servizi di business intelligence per la valutazione della ricerca tramite SURplus di CILEA