



UNIVERSITÀ DEGLI STUDI DI MILANO
UNIRES – ITALIAN CENTRE FOR RESEARCH
ON UNIVERSITIES AND HIGHER EDUCATION SYSTEMS

**ASSESSING THE RELIABILITY AND VALIDITY
OF GOOGLE SCHOLAR INDICATORS.
THE CASE OF SOCIAL SCIENCES IN ITALY**

ex CAVIB Scholar
Progetto ANVUR Concorso Pubblico di Idee di Ricerca 2014
Ferruccio Biolcati-Rinaldi, Francesco Molteni, Silvia Salini



Italian Centre for Research on
Universities & HE Systems

Associazione di Valutazione - AIV
XX Congresso nazionale
PROMUOVERE E VALUTARE
IL BENESSERE DELLE COMUNITÀ
Padova, 21-23 aprile 2017

Indice

1. Il disegno della ricerca
2. I dataset analizzati e la copertura di Scholar
3. L'attendibilità e la validità degli indicatori bibliometrici
4. Conclusioni



Il disegno di ricerca



UNIVERSITÀ DEGLI STUDI DI MILANO
UNIRIS - ITALIAN CENTRE FOR RESEARCH
ON UNIVERSITIES AND HIGHER EDUCATION SYSTEMS

Dimensioni della produzione scientifica

- Con **realizzazione (o output)** si intende la capacità del ricercatore di trasformare il proprio lavoro scientifico nel principale dei prodotti di ricerca, ossia in pubblicazioni
 - Esempi di indicatori di realizzazione sono il numero di pubblicazioni complessive in un determinato periodo nelle diverse fonti di dati, il numero di libri, capitoli, articoli, ecc.
- Con la seconda dimensione si accoglie l'interpretazione delle citazioni come misura di **riconoscimento scientifico**
 - Esempi di indicatori di riconoscimento sono il numero di citazioni per pubblicazione e l'indice h degli autori
- Infine, la **rilevanza** fa riferimento all'importanza che una comunità scientifica attribuisce ai diversi tipi di pubblicazione (libro, articolo, atti di convegno, mappa, brevetti, ecc.) o alle diverse sedi di pubblicazione (riviste, case editrici, collane editoriali)
 - Un esempio di indicatore di rilevanza è la collocazione degli articoli nelle riviste di Fascia A delle liste ANVUR ASN



Obiettivo: valutazione degli indicatori bibliometrici tratti da Google Scholar

- **Copertura**
 - Capacità di includere la produzione scientifica
 - Confronti incrociati tra le diverse fonti di dati per valutare la **copertura** della produzione scientifica nelle aree cosiddette non bibliometriche
 - Analisi sui dataset al livello di base delle **singole pubblicazioni**
- **Attendibilità**
 - Riproducibilità del risultato
 - L'attendibilità viene declinata come **coerenza interna** e valutata sulla base delle analisi delle correlazioni tra i diversi indicatori
 - Analisi sui dataset al livello aggregato degli **autori**
- **Validità:**
 - Rilevare effettivamente il concetto
 - Si propone di applicare la validazione **per criterio**. Questa procedura è applicabile per quelle dimensioni legate alla qualità della produzione scientifica (riconoscimento e rilevanza)
 - Analisi sui dataset al livello aggregato degli **autori**. **Criteri:**
 - esiti individuali Vqr 2004-2010 (con elaborazioni micro in capo ad ANVUR)
 - esiti individuali Asn 2012

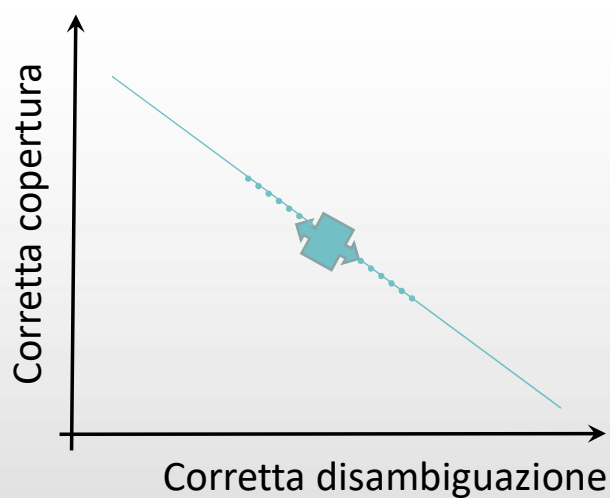
I dataset utilizzati

Fonte	Note	Fornitore
Scholar	<i>Versione grezza</i>	EVA
Scholar	<i>Versione disambiguata per autore</i>	EVA
Scholar	<i>Versione disambiguata per SSD</i>	EVA
WOS	Interrogazione controllata	CAVIB
SCOPUS	Interrogazione controllata	CAVIB
SUA-RD	Anni 2011-2013	ANVUR
ASN	Anno 2012	ANVUR
VQR 2004-2010	Non dataset ma risultati analisi micro	ANVUR

EVA

- Il dataset Scholar è stato elaborato all'interno del progetto *EVA - Estrazione, validazione e analisi dei dati di Google Scholar per i settori non bibliometrici* sviluppato da Alfio Ferrara, Stefano Montanelli e Stefano Verzillo dell'Università di Milano.
- L'acquisizione e la gestione dei dati di Google Scholar si è articolata in quattro fasi distinte:
 - la costruzione di un'anagrafica dei docenti,
 - l'implementazione del software di acquisizione,
 - l'eliminazione delle pubblicazioni spurie (problema dell'interrogazione «cognome nome»),
 - la disambiguazione dei nominativi per una corretta attribuzione delle pubblicazioni (dizionario costruito a partire dai CV ASN 2012).

Trade off tra disambiguazione e copertura



I dataset analizzati e la copertura di Scholar



UNIVERSITÀ DEGLI STUDI DI MILANO
UNIRIS - ITALIAN CENTRE FOR RESEARCH
ON UNIVERSITIES AND HIGHER EDUCATION SYSTEMS

La popolazione di riferimento (31.12.2014)

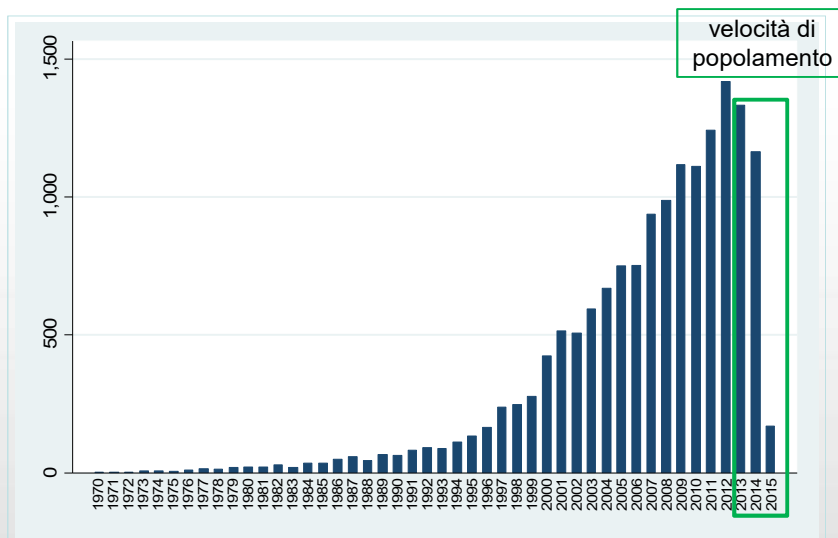
	Perc.	(N)
<i>Genere</i>		
Maschio	60,9	(1034)
Femmina	39,1	(663)
<i>Coorte di nascita</i>		
1940-1949	13,1	(218)
1950-1959	23,4	(391)
1960-1969	33,7	(562)
1970-1983	29,8	(498)
<i>Ruolo</i>		
Professore ordinario	22,2	(376)
Professore associato	29,9	(507)
Ricercatore a tempo indeterminato	38,5	(655)
Ricercatore a tempo determinato	9,4	(159)
<i>Settore concorsuale</i>		
14/A1 - Filosofia politica	6,2	(105)
14/A2 - Scienza politica	12,9	(218)
14/B1 - Storia delle dottrine e delle istituzioni politiche	11,1	(189)
14/B2 - Storia delle relazioni internazionali, delle società e delle istituzioni extraeuropee	8,8	(150)
14/C1 - Sociologia generale	23,7	(401)
14/C2 - Sociologia dei processi culturali e comunicativi	18,3	(311)
14/C3 - Sociologia dei fenomeni politici e giuridici	6,9	(117)
14/D1 - Sociologia dei processi economici, del lavoro, dell'ambiente e del territorio	12,1	(206)
<i>Dimensione dell'università</i>		
Grande	55,1	(922)
Media	25,5	(427)
Piccola	16,0	(267)
Molto piccola	3,4	(57)
<i>Zona dove ha sede l'università</i>		
Nord-Ovest	23,8	(404)
Nord-Est	19,2	(325)
Centro	24,3	(412)
Sud	19,7	(335)
Isole	10,4	(177)
Telematica	2,6	(44)



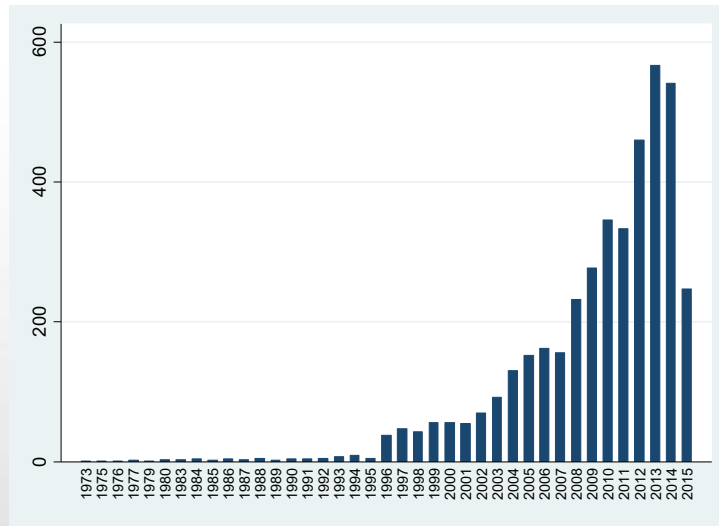
L'andamento nel tempo

- L'andamento nel tempo della produzione scientifica, così come rilevata nei tre dataset bibliometrici, mescola almeno due fenomeni:
 - la crescente capacità di copertura dei dataset bibliometrici della produzione scientifica
 - l'effettivo andamento della produzione scientifica che possiamo ipotizzare che risenta di due tendenze contrastanti:
 - aumento della produttività individuale (anche per l'irrompere delle logiche valutative)
 - riduzione del corpo accademico
- Implicazioni per qualsiasi confronto longitudinale, necessità di separare i due fenomeni

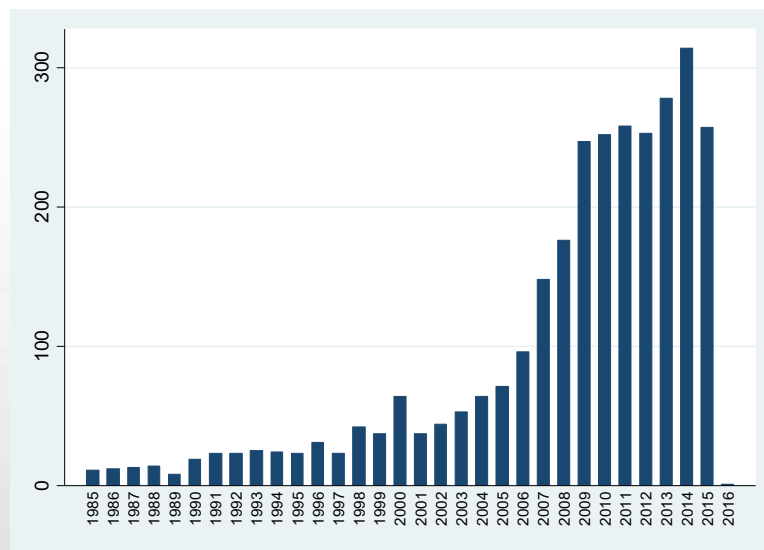
Il dataset SCHOLAR/EVA



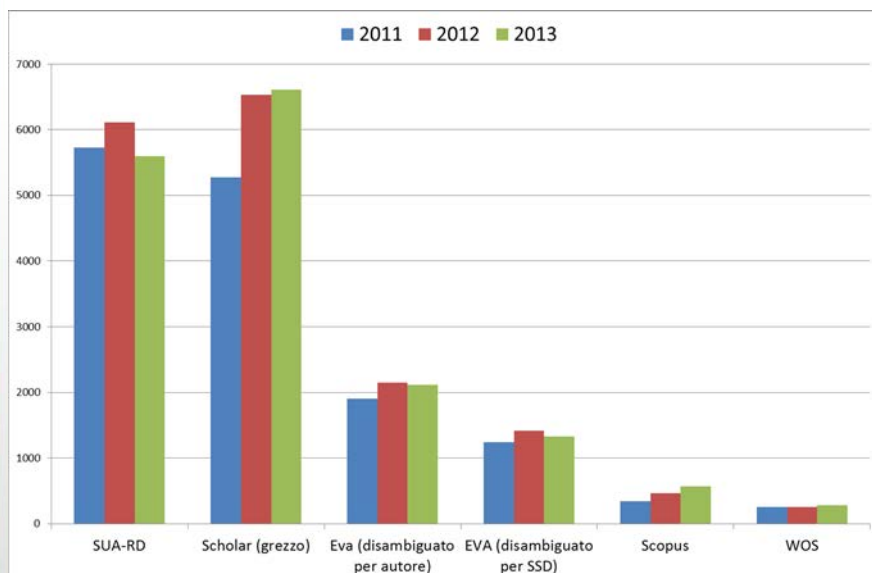
Il dataset Scopus



Il dataset WOS



Il dataset SUA-RD a confronto con gli altri dataset



La copertura di *Scholar/Eva* della produzione scientifica

	SUA-RD (%)	(N)		Scopus (%)	(N)
Scholar/Eva	14,7	(17513)	SUA-RD	61,6	(1356)
Scopus	4,7	(17608)	WOS (%)		(N)
WOS	2,9	(17662)	SUA-RD	60,7	(780)

Una chiave di lettura dei dati

- SUA-RD potrebbe essere parzialmente inflazionato
- EVA potrebbe avere privilegiato la corretta disambiguazione a detrimento della corretta copertura
- Gli algoritmi di copertura sembrano problematici
- Ma in ogni caso Scholar non è esaustivo come ce lo immaginiamo



L'attendibilità e la validità
degli indicatori bibliometrici
tratti da Google Scholar



Tab. 7.1. Statistiche descrittive degli indicatori bibliometrici

	Mediana	Media	Campo di variazione	Dev. std.	Coeff. di variazione	Asim- metria	Curtosi	(N)
<i>Contributi</i>								
Scholar	4	8,2	147	11,7	1,4	3,7	26,3	(1697)
Scopus	1	2,1	41	3,9	1,8	3,3	19,0	(1697)
WOS	0	1,4	53	3,1	2,1	5,5	60,8	(1697)
<i>Citazioni per contributo</i>								
Scholar	2,0	5,5	216,6	13,5	2,4	8,5	102,5	(1372)
Scopus	0,6	2,5	72,6	5,8	2,2	5,9	52,4	(924)
WOS	1,0	2,5	87,0	6,0	2,3	7,0	74,9	(692)
<i>Indice H</i>								
Scholar	2	2,6	22	2,9	1,1	2,3	10,1	(1372)
Scopus	1	1,1	12	1,5	1,3	2,4	11,3	(924)
WOS	1	1,0	10	1,2	1,2	2,0	9,3	(692)

Tab. 7.4. Correlazioni degli indicatori di realizzazione e riconoscimento scientifico in Scholar, Scopus e WOS (correlazione di Spearman delle versioni originali)

	Contributi	Citazioni per contributo	Indice H
Scholar-Scopus	0,48	0,64	0,65
Scholar-WOS	0,42	0,59	0,57
Scopus-WOS	0,65	0,74	0,72
(N)	(1697)	(534)	(534)

Tab. 7.5. Correlazioni tra il punteggio medio nella VQR 2004-2010 e le versioni originali degli indicatori di riconoscimento scientifico calcolati in Scholar, Scopus e WOS per due diversi periodi temporali

	Indice H/VQR <i>Correlazione di Pearson</i>			Citazioni per contributo/VQR <i>Correlazione di Spearman</i>		
	<i>r</i>	<i>p</i>	(<i>N</i>)	<i>rho</i>	<i>p</i>	(<i>N</i>)
<i>1998-2014</i>						
Scholar	0,33	0,000	(1127)	0,27	0,000	(1127)
Scopus	0,29	0,000	(750)	0,23	0,000	(750)
WOS	0,26	0,000	(559)	0,21	0,000	(559)
<i>2004-2010</i>						
Scholar	0,33	0,000	(970)	0,27	0,000	(970)
Scopus	0,30	0,000	(463)	0,26	0,000	(463)
WOS	0,25	0,000	(374)	0,19	0,000	(374)

Tab. 7.6. Regressioni logistiche dell'abilitazione per una serie di indicatori bibliometrici normalizzati (indici H e citazioni per contributo in Scholar, Scopus e WOS, indicatori ASN) - ASN 2012

	OR	Std. Err.	z	P> z	Pseudo R ²	Count R ²	Adjusted Count R ²	(N)
<i>Indice H (normalizzato)</i>								
Scholar	1,58	0,119	6,13	0,000	0,05	0,60	0,19	(696)
Scopus	1,60	0,156	4,83	0,000	0,04	0,57	-0,02	(509)
WOS	1,39	0,143	3,23	0,001	0,02	0,64	0,00	(397)
<i>Citazioni per contributo (radice quadrata normalizzata)</i>								
Scholar	1,34	0,112	3,57	0,000	0,01	0,55	0,10	(696)
Scopus	1,22	0,088	2,88	0,004	0,01	0,58	0,00	(509)
WOS	1,22	0,111	2,20	0,028	0,01	0,64	0,00	(397)
<i>Indicatore ASN (normalizzato)</i>								
1	0,98	0,071	-0,39	0,695	0,00	0,48	-0,04	(696)
2	2,09	0,235	6,61	0,000	0,04	0,61	0,21	(509)
3	1,93	0,163	7,84	0,000	0,07	0,64	0,28	(397)

Conclusioni: il livello bibliometrico

- Sostanziale attendibilità e validità degli indicatori bibliometrici tratti da Scholar
- Perimetro dei dataset bibliometrici e conseguenze per attendibilità e validità
- Gli indicatori di riconoscimento scientifico risultano più attendibili degli indicatori di realizzazione
- Difficoltà a estrarre indicatori di rilevanza da Scholar
- L'indice H sembra funzionare meglio del numero medio di citazioni per contributo
- I confronti longitudinali sono inquinati dalla capacità di capacità di copertura dei dataset bibliometrici che cresce nel tempo

Conclusioni: il livello valutativo

- I tre repertori citazionali sembrano essere più simili di quanto si potesse pensare a priori per quanto riguarda la qualità dei dati
- Scholar opera comunque una selezione di quella che è la produzione di una comunità scientifica, è certamente più esteso dei repertori commerciali ma lontano dall'eshaustività
- Il trade-off tra disambiguazione e copertura è ingigantito nel caso di Scholar: qualità della base dati di partenza, algoritmo e grandi numeri
- L'errore è critico per ogni sistema di valutazione
- La soglia di errore consentito varia a seconda dell'oggetto della valutazione (struttura, individuo)
- Il criterio per cui l'errore non deve essere tale da distorcere i risultati della valutazione fatica a reggere a livello individuale perché errore induce crisi della fiducia (nonché più facili risvolti legali)
- Ne deriva una indicazione per la massima cautela nell'utilizzo dei dati Scholar
- Contesto in continuo cambiamento

GRAZIE DELL'ATTENZIONE

PER MAGGIORI INFORMAZIONI:

Ferruccio Biolcati Rinaldi
Dipartimento di Scienze Sociali e Politiche
Università degli Studi di Milano
via Conservatorio, 7 - 20122 Milano (MI) - ITALY
Tel. + 39 02 503 21287 - Fax + 39 02 503 21240 - ferruccio.biolcati@unimi.it



UNIVERSITÀ DEGLI STUDI DI MILANO
UNIRES - ITALIAN CENTRE FOR RESEARCH ON
UNIVERSITIES AND HIGHER EDUCATION SYSTEMS