

Roars e l'arte del Fake checking

E' di qualche giorno fa un roboante articolo di Roars che grida al fake per una espressione utilizzata in un articolo de IlSole24 Ore secondo cui Roberto Cingolani, Direttore Scientifico di IIT sia "tra i primi 10 material scientist" al mondo, ricavando questa informazione da Google Scholar, uno dei data base più diffusi nella comunità scientifica.

<http://www.ilsole24ore.com/art/commenti-e-idee/2017-08-05/come-scienza-cerco-l-essenza-cose-181456.shtml?uuid=AE4ABs7B&fromSearch>

<https://www.roars.it/online/vero-o-falso-per-il-sole-24-ore-cingolani-iit-e-uno-dei-dieci-scientiati-più-citati-del-mondo-per-le-scienze-dei-materiali/>

La ragionevole puntualizzazione sul riferimento di Google Scholar fatta da Roars è stata subito accettata da Luca De Biase con tanto di scuse e ringraziamenti per la segnalazione. Ma questo non è bastato e ROARS ha voluto ricostruire un profilo che secondo il blog sarebbe più veritiero.

Il lavoro realizzato da Roars, ben lungi dall'essere esente da critiche nel metodo e nel merito, è una vera e propria manifestazione di utilizzo spregiudicato di dati e informazioni. E a questo punto la giusta segnalazione, passa la mano ad una operazione di vero e proprio fake checking, aggravato dal sospetto che il blog Roars – non essendo una testata registrata (operazione da pochi minuti in tribunale) – si garantisca "parole in libertà".

Le principali "lacune" del check riguardano il data base utilizzato (solo anni 2001, 2014, 2015, 2016 che analizzano i 10 anni precedenti), il periodo di tempo analizzato (limitato prevalentemente ai 10 anni di direzione scientifica), gli ambiti di ricerca (l'attività di RC è multidisciplinare), l'attendibilità stessa della fonte: secondo il britannico The Guardian Clarivate Analytics è fonte "semplicistica, controversa, con una capacità di influenza indebita".

Da non dimenticare che le fonti di Roars sono prevalentemente su abbonamento (non accessibili a chi non è nel mondo della ricerca) mentre Google Scholar è libero.

Roberto Cingolani è uno dei migliori scienziati italiani nei suoi campi di indagine a livello nazionale e internazionale.

Roberto Cingolani è un ricercatore la cui attività scientifica vanta 950 pubblicazioni su WOS con oltre 20000 citazioni e h-index 72. Alla produzione scientifica aggiunge la realizzazione di alcuni dei principali laboratori italiani multidisciplinari quali NNL e IIT. (<https://www.iit.it/executive-committee/48-roberto-cingolani-cv-eng/file>)

Il riferimento circa H-index e citazioni su Google Scholar è sicuramente accurato. E in più è una fonte più completa per chi voglia rappresentare l'intera carriera di uno scienziato e non soltanto 10 anni.

E i Tag ? Utilizzando *Materials Science* Roberto Cingolani sarebbe tra i primi 70, fra i primi 55 utilizzando *Nanotechnology*.

L'utilizzo di Google Scholar come fonte può essere opinabile. Rimane il fatto che è utilizzato da gran parte del mondo scientifico. Google Scholar è fonte indiscutibile? No. Ovvio. E Clarivate Analytics ?

semplicistico, controverso, con una capacità di influenza indebita almeno a sentire ***the Guardian*** (Figura 1)

<https://www.theguardian.com/science/2017/jun/01/peer-review-is-essential-to-good-science-its-time-to-credit-expert-reviewers>

in job or promotion applications - of their contributions to peer review.

The system benefits journals because it can be used to identify researchers who can be trusted to do a good job if approached with a peer review request. It should also help to weed out those [authors who fraudulently pose as reviewers](#) to get their papers published.

In the past five years Publons has accumulated a user-base of over 150,000 researchers, who have collectively performed over 800,000 reviews. Following the acquisition by Clarivate Analytics for an undisclosed sum, Preston and Johnston, who remain in post along with the rest of the Publons staff, are looking to expand.

Clarivate Analytics is best known in academic circles as the company that acquired the citation database Web of Science from Thomson Reuters in 2016. That gave the company the means to calculate journal impact factors, a metric that is often linked simplistically - and controversially - to the quality of research outputs, and has acquired undue influence in research assessment.

The Guardian view on the end of the peer review
[Editorial: Nature and the BMJ](#)

Annette Thomas, a member of the board of Clarivate Analytics, insists that its investment in Publons will boost the profile and protect the independence of the platform, thereby enhancing its value to the research community.



Figura 1: articolo di *The Guardian*

“Clarivate Analytics è meglio conosciuto nei circoli accademici come azienda che ha acquisito il database di citazione Web of Science da Thomson Reuters nel 2016. Questo ha dato all'azienda i mezzi per calcolare gli impact factors, una metrica che spesso è legata semplicisticamente - e in modo controverso - alla qualità dei risultati della ricerca e ha acquisito un'influenza indebita nella valutazione della ricerca”

In sintesi: Google Scholar è un database pubblico che raccoglie pubblicazioni lungo l'intero arco della carriera di uno scienziato. Clarivate WOS è anch'esso su base volontaria e ha un limite di tempo di 10 anni. E' disponibile solo su abbonamento.

Ma la verifica va oltre.

1. “Nelle liste Clarivate degli Highly Cited, Cingolani non c’è”

Negli archivi ufficiali disponibili pubblicamente sul sito di HCR ci sono esclusivamente gli anni 2001, 2014, 2015, 2016 <http://hcr.stateofinnovation.com/page/archives>.

Il rilascio 2016 (preso a riferimento da Roars) si riferisce al decennio 2004-2014 anni in cui Roberto Cingolani si dedicava principalmente a ricoprire il ruolo di Direttore Scientifico di IIT, e quindi con ovvia minore produzione scientifica.

Inoltre il pacchetto di dati utilizzato per stilare questa lista si riferisce a un decennio di pubblicazioni precedente all'anno rilasciato e non a tutta la produzione scientifica (Google Scholar tiene in considerazione tutta la carriera scientifica).

Le pubblicazioni vengono categorizzate [in base alla subject area Web of Science](#) in cui ricadono le riviste (<http://hcr.stateofinnovation.com/page/methodology>).

E' noto che l'attività di Cingolani, pur mantenendo un unico focus generale - come da lui indicato nel profilo di Google Scholar e sul curriculum vitae (<https://www.iit.it/executive-committee/48-roberto-cingolani-cv-eng/file>) - se viene categorizzata con le aree "rigide" usate dalle piattaforme come Web of Science – Clarivate Analytics o altre, può portare a conclusioni non puntuali. La sua attività ricadrebbe anche, ad esempio, in Physics and Chemistry (Figura 2). Questo significa che la classifica utilizzata da Roars è parziale e considera soltanto una parte dei lavori.

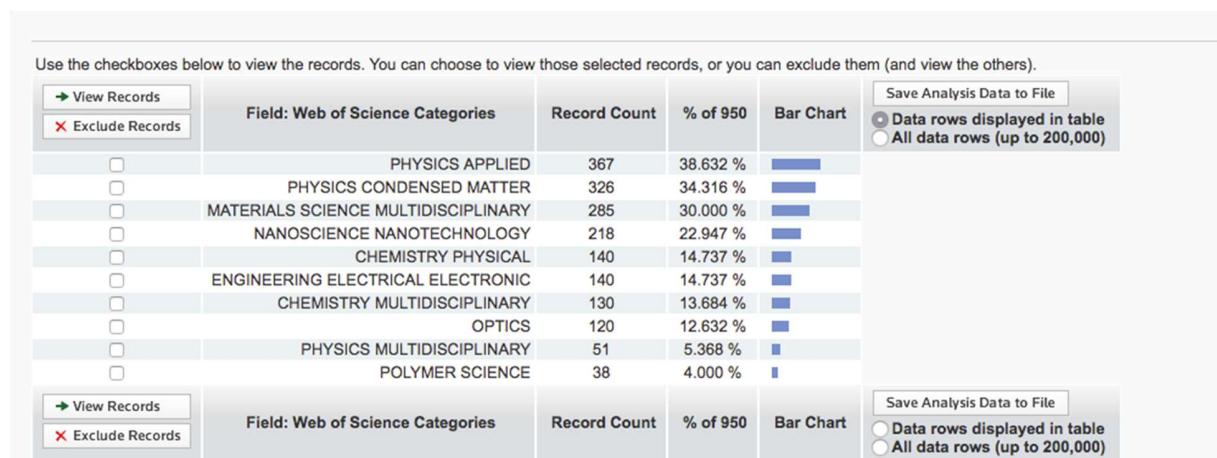


Figura 2: Aree di ricerca di Roberto Cingolani

Sotto il profilo metodologico non premettere questa nota alle analisi fatte, dimostra una certa superficialità e lascia insinuare il dubbio che si voglia fare un utilizzo strumentale dei dati (che poi è il vero problema dei *big data*: si trova sempre un riferimento quantitativo a supporto delle proprie tesi).

In sintesi: Clarivate è un ranking migliore di Google Scholar? No. E la presenza o meno del Prof. Cingolani nelle liste Clarivate degli Highly Cited non si può escludere. Dire che sia assente (così come che sia presente) è una affermazione non verificabile rispetto agli anni non consultabili.

Gli stessi fruitori di Clarivate sollevano delle perplessità: (<http://hcr.stateofinnovation.com/page/frequently-asked-questions>)

Why does a junior member of our faculty appear on the Highly Cited Researchers 2016 list, but a more senior member does not? Again, the specific methodology used in generating the new list (see above) can turn up researchers – even so-called junior researchers – who have contributed multiple Highly Cited Papers during 2004-2014, whereas more senior and even more cited scientists may not have been identified because they did not publish as many Highly Cited Papers in this field (as we defined it, see below) during this period.

We think the result conforms to what we were trying to achieve – the identification of researchers with substantial contemporary impact as measured by the number of Highly Cited Papers produced, even if those papers, in terms of total citations, do not sum to more than that of other researchers who have longer publication and higher citation records over their entire careers.

You say you selected top researchers according to specific fields as defined in Essential Science Indicators. But what about researchers who have Highly Cited Papers across several fields, such as Molecular Biology and Genetics, Clinical Medicine, and Psychiatry/Psychology? How did you account for such cross-disciplinary impact? We analyzed each *Essential Science Indicators* field separately. If we had attempted to identify researchers across fields we would have faced the problem of determining a cross-field baseline for inclusion in an all-field list. In other words, the percentage of Highly Cited Papers required for inclusion differs across fields. In some fields, the number of Highly Cited Papers required for selection was low compared to others – as much as five to one – and each individual would present a different mix of fields in which they published their Highly Cited Papers and in different quantities, thereby creating a seemingly unmanageable number of combinations by which to calculate a baseline for inclusion. This is a challenging problem that we are working to solve.

2. Come si classifica Cingolani? E' 2746-esimo:

Il lasso temporale preso in esame è temporalmente limitato e riguarda prevalentemente il periodo relativo all'incarico da direttore scientifico, in cui l'attività scientifica cala (si rammenta che Google Scholar riguarda l'intero arco della vita scientifica. Da questo punto di vista una fonte più rappresentativa).

Cingolani compare al 2746-esimo posto compilando la lista degli scienziati maggiormente citati utilizzando gli Essential Science Indicators (accessibili solo su abbonamento), che per altro non è una riproduzione più estesa della lista degli Highly Cited Researchers.

I due riferimenti non sono correlabili. Il non dirlo rappresenta una mancanza di rigore metodologico rilevante.

<http://ipscience-help.thomsonreuters.com/incitesLiveESI/dataSubscriptionGroup/dataAndSubs.html>

Se si applicano gli stessi filtri della lista del punto 1. non si ottengono gli stessi 153 ricercatori dell'altra. In allegato la lista dei primi 13 nomi per citazioni e per Highly Cited Papers in Materials Science. Già da questi pochi nomi si può capire che c'è qualcosa che non torna: circa metà dei primi 13 non sono tra i 153. Un altro aspetto di criticità è anche legato al nome del ricercatore, soprattutto se con cognomi di origine asiatica. Si possono infatti contare 150 CHEN, un centinaio di LEE, 200 LI e ZHANG, 230 WANG. (Figura 3)

Results List		Map View by Top / Hot / Highly Cited Papers					Show Visualization +	
Authors		Report View by Selection					Customize	
Total:	Authors	Web of Science Documents	Cites	Cites/Paper	Highly Cited Papers			
5545								
564	CAO, WQ	79	1,483	18.77		8		
564	CHEN, YH	431	4,972	11.54		8		
564	CHEN, YL	341	4,354	12.77		8		
564	CHOI, SH	244	3,575	14.65		8		
564	CHOY, WCH	45	1,822	40.49		8		
564	CHU, YH	116	4,223	36.41		8		
564	DING, B	120	2,886	24.05		8		
564	DOMEN, K	55	5,115	93.00		8		
564	FENG, J	333	6,324	18.99		8		
564	FUSIL, S	14	1,788	127.71		8		
564	GAO, C	210	4,296	20.46		8		
564	GAO, W	292	3,837	13.14		8		
564	GRAETZEL, M	32	2,081	65.03		8		

Figura 3: Classifica Highly Cited Papers

Questo non confuta ovviamente la validità della classifica. Semplicemente suggerisce che bisognerebbe verificare caso per caso.

Prendendo ad esempio uno dei ricercatori con nome WANG Zhong Lin (l'unico WANG presente nei 153 della prima classifica), ordinando la classifica per HCP lui è effettivamente nei primi posti ma davanti a lui ci sono altri due WANG non presenti nella prima classifica. Ordinando per citazioni se ne trovano prima di lui 6 (nessuno presente nella prima classifica). O ancora se si cerca CHUN, YH nome presente nella schermata usata da ROARS nel punto 1, usando questa seconda classificazione scende al 744 posto per citazioni e al 564 per Hcp (Figura 4). Stessa cosa accade per altri che “escono” dai primi 153 researchers più citati secondo la 1

Results List		Map View by Top / Hot / Highly Cited Papers					Show Visualization +	
Authors		Report View by Selection					Customize	
Total:	Authors	Web of Science Documents	Cites	Cites/Paper	Highly Cited Papers			
5545								
1	LIU, J	2,450	41,616	16.99		80		
2	ZHANG, H	2,428	35,168	14.48		69		
3	ZHANG, Q	1,631	26,721	16.38		63		
4	ZHANG, Y	4,077	52,617	12.91		62		
5	LIU, Y	4,260	50,830	11.93		61		
6	WANG, Y	4,316	55,113	12.77		60		
7	WANG, J	3,903	49,009	12.56		57		
7	YANG, Y	2,058	39,596	19.24		57		
9	WANG, ZL	699	22,211	31.78		53		
10	CHEN, J	2,037	29,018	14.25		49		
10	LI, J	3,600	42,387	11.77		49		
10	LIU, Z	913	19,216	21.05		49		
10	ZHANG, J	3,413	38,044	11.15		49		

Figura 4: Classifica Highly Cited Papers per autore

In sintesi: la classifica riportata da Roars non è per nulla esente da dubbi di attendibilità. La bravura e la numerosità dei ricercatori asiatici lascia estasiati.

Anche in questo caso non mancano i dubbi dei fruitori della piattaforma:
(<http://hcr.stateofinnovation.com/page/frequently-asked-questions>)

I have a very common name, and some other people with the same name form (surname and initials) actually work in the same field that I do. How did you make sure not to confuse me and my papers with others and their papers in your analysis? Ensuring correct attribution of papers to authors involved a manual inspection of each Highly Cited Paper. The Highly Cited Papers in an *Essential Science Indicators* field for a specific name and its variants (such as, for example, surname plus one initial or two) were ordered in chronological sequence, the subject of the papers examined as well as the journals in which they were published, the institutional addresses reviewed, and the co-authorships inspected. Often this was sufficient to resolve questions of authorship for a unique individual. Original papers were sometimes consulted to obtain a full name not present in the *Web of Science* bibliographic record (papers indexed before 2008). Reference was made to websites of researchers themselves and their curricula vitae if questions remained, which sometimes arose when a researcher changed institutional affiliations several times during the period surveyed. We would like to think our efforts to resolve authorship questions resulted in 100% clean data, but with any such effort, and more than 3,000 researchers, we likely fell short in some few specific instances and will make adjustments where required.

Conclusioni

Sicuramente Roars ha ragione nel sottolineare che l'espressione Roberto Cingolani è fra "i primi 10 *material scientist* più citati al mondo" ricavata dal Tag di Google Scholar non sia appropriata. Il suo contributo per la ricerca nazionale e internazionale è ben più ampio e non si può limitare alle sole citazioni.

Per la morale di questa storia rimando al monito di Umberto Eco sull'attendibilità delle fonti on line non certificate (o in alternativa alla parabola della pagliuzza e della trave).

Sul checking di Roars avrei la tentazione di utilizzare una delle espressioni più note del "professore" che guidò il grifone a fine anni '80. Mi limito a rimanere nel vocabolario tanto caro al blog: fake.

Alcuni suggerimenti per orientarsi nelle classifiche

Per classificare i ricercatori non esiste un metodo univoco e universale. Ogni sistema ha le sue problematiche e controindicazioni.

Esistono numerose classifiche oltre quella già discussa degli highly cited researchers, come quella di www.topitalianscientists.org (basata su google scholar) che, ad esempio, inserisce al secondo posto dei top researcher italiani nel mondo nel campo delle scienze dei materiali proprio Roberto Cingolani. Questo comunque non fornisce indicazioni assolute sulla qualità dei ricercatori. Chi è presente nella classifica è sicuramente un ottimo scienziato, ma ciò non esclude che chi non viene citato non sia un ottimo scienziato.

I metodi più utilizzati si basano sull'utilizzo del database Web of Science che analizza tutta la produzione del ricercatore basandosi sul nome e offrendo alcuni filtri (non infallibili) per evitare la confusione tra i

casi di omonimia. L'accesso a tale database è fondamentalmente riservato ad istituzioni e ricercatori (Figura 5).

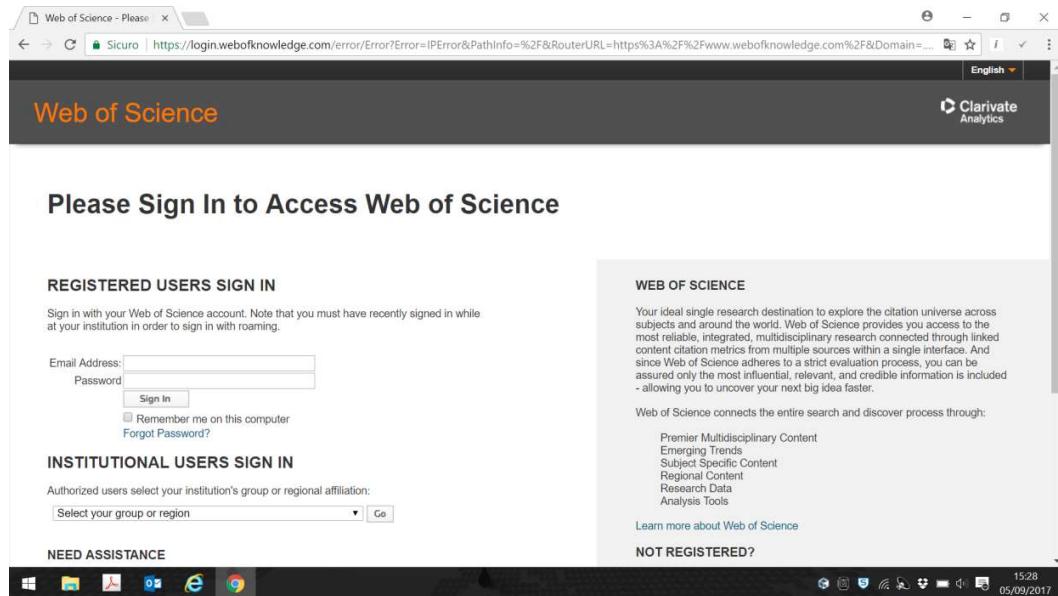


Figura 5: Pagina di Login di Web Of Science, Servizio a pagamento

Google scholar, il motore di ricerca di google dedicato ai ricercatori, è un valido strumento aperto a chiunque per individuare pubblicazioni scientifiche in ogni campo della ricerca. Su base volontaria offre la possibilità di crearsi un proprio profilo. Se accuratamente utilizzato questo strumento può essere più preciso rispetto ad altri database in quanto l'autore può modificare il proprio profilo autonomamente evitando confusioni (quali ad esempio la paternità di pubblicazioni non proprie dovute ad omonimia).

A seconda delle impostazioni, Google scholar suggerisce i lavori che reperisce in rete (compresi report tecnici e partecipazioni a conferenze spesso non indicizzati da Web of Science) e li notifica come in attesa di approvazione o li inserisce direttamente. L'autore che abbia selezionato la modalità "con suggerimento" può rifiutare la paternità della pubblicazione rendendo più veritiero il proprio profilo.

Altro metodo accurato ma non aggiornato automaticamente dal sistema è ResearcherID di Web of Science. In questo caso l'autore ha una pagina di profilo a cui corrisponde un ID univoco e deve procedere personalmente (senza suggerimenti o inserimenti automatici) all'inserimento delle pubblicazioni presenti in nel database.