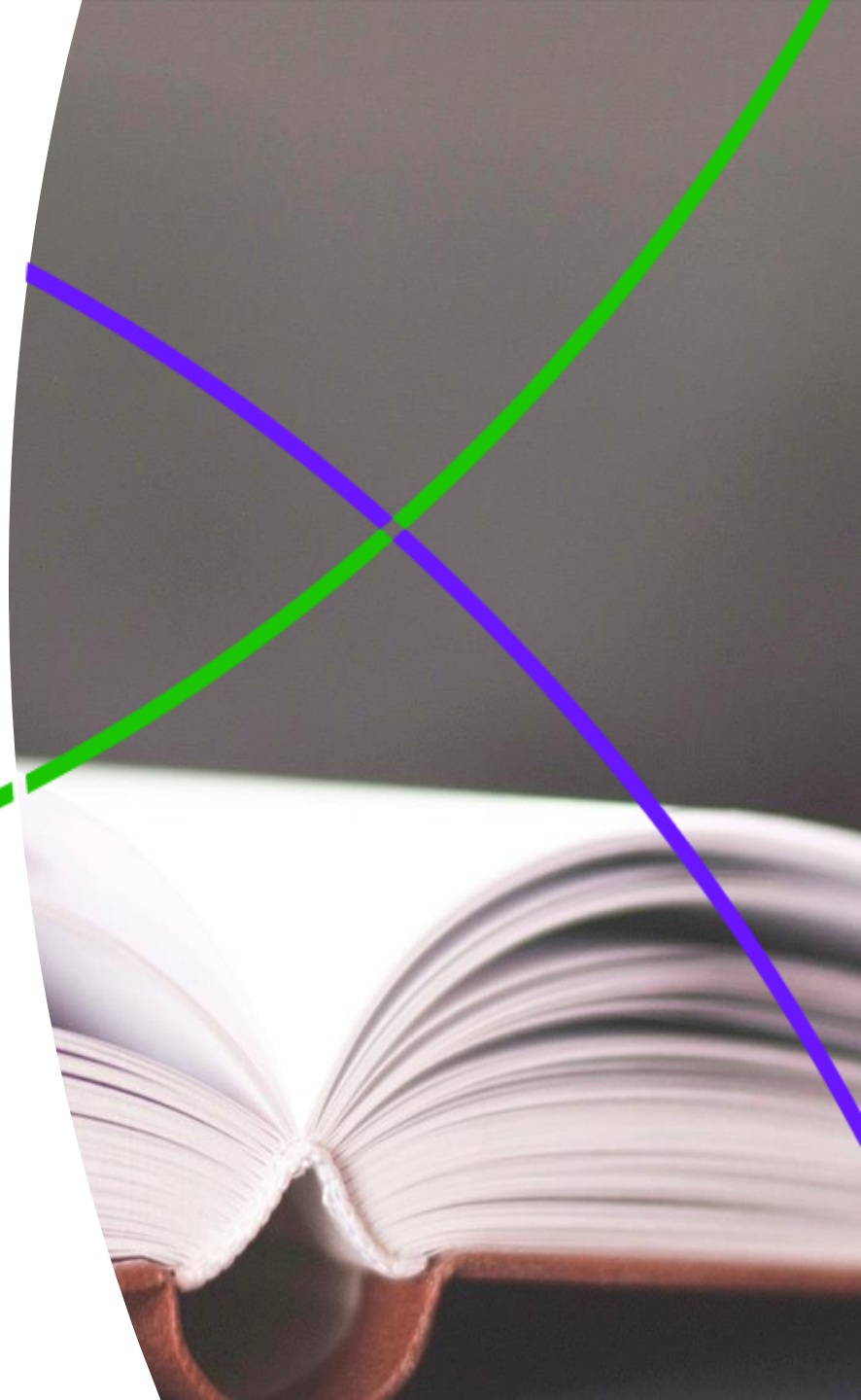
Bibliometrics in a Nutshell

Dr. Evangelia Lipitakis
Pre-Sale Solutions Consultant
Evangelia.lipitakis@clarivate.com

Adriana Filip
Solution Specialist
Adriana.filip@clarivate.com

October 2017

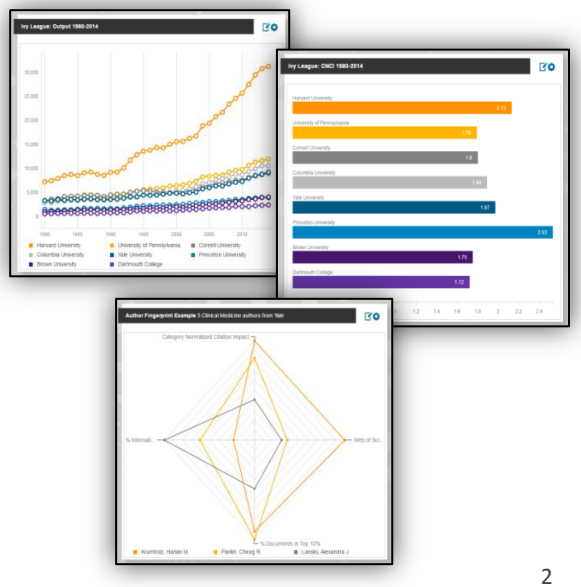
Learn more about bibliometrics and research analytics at our LibGuides pages: http://clarivate.libguides.com



Web of Science
Trust the difference

AGENDA

- Examples of bibliometric indicators, what they are used to measure, their limitations and appropriate use
- Data & metadata considerations
- Practical hands-on session



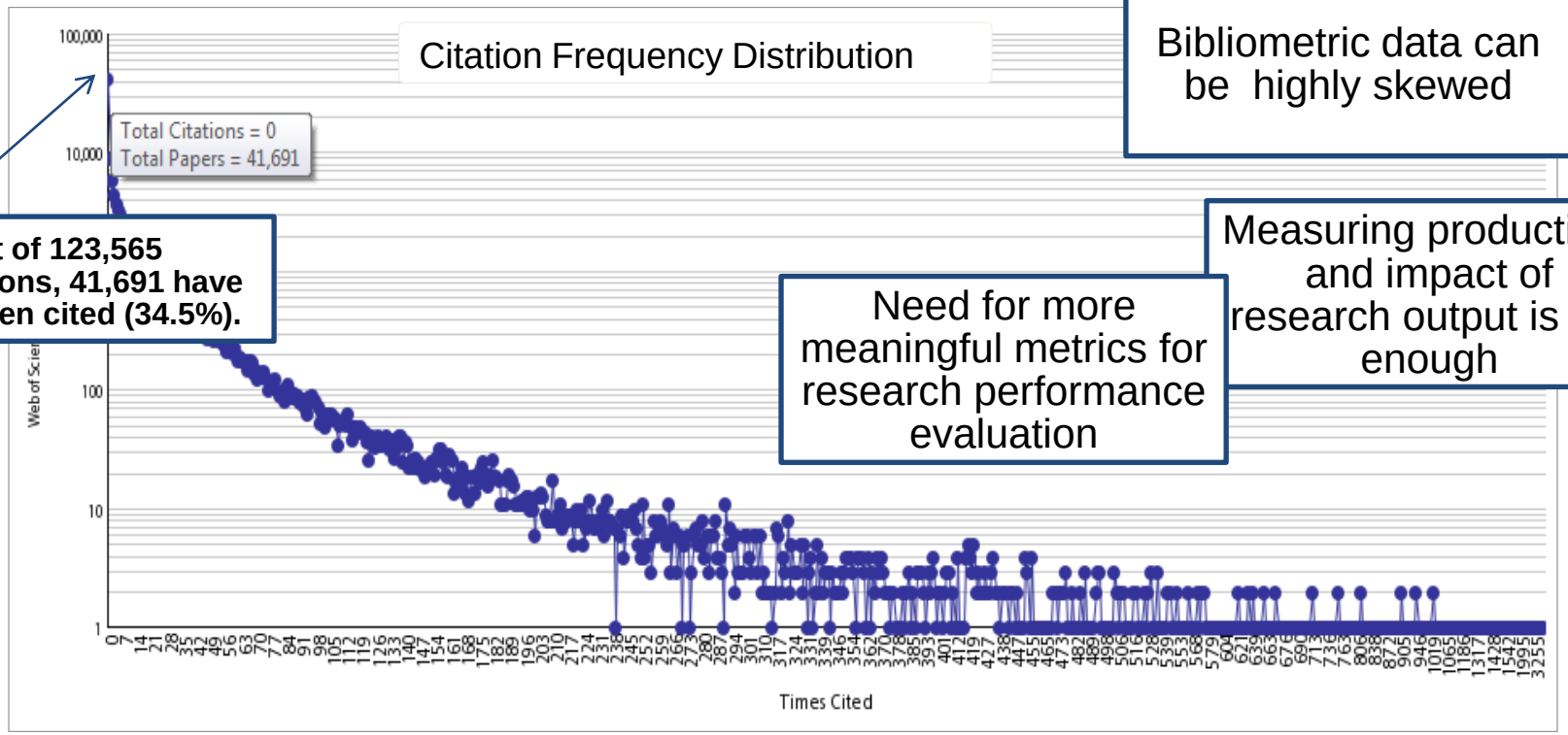
PART I:

BIBLIOMETRIC INDICATORS FOR MEASURING RESEACH IMPACT

% DOCUMENTS CITED

The percentage of documents that have received at least one citation in a set of publications

% Documents Cited = 65.5%



Bibliometric data can be highly skewed

Out of 123,565 publications, 41,691 have never been cited (34.5%).

Measuring productivity and impact of research output is not enough

Need for more meaningful metrics for research performance evaluation

CITATION IMPACT

The total number of citation divided by the total number of publications in a set

- Also known as 'Average Citation Rate' or 'Citations per Publication'

Examples	Total Publications	Total Citations	*Citation Impact*
Researcher A	1	50	50
Researcher B	10	200	20

Researcher A: *Citation Impact* = 50
Researcher B: *Citation Impact* = 20

Even though Researcher B has published more documents and received more citations overall.

Does not account for differences in the fields

H-INDEX

A researcher has an *h-index*, if he/she has at least h publications for which he/she has received at least h citations

- Introduced by physicist J. Hirsch in 2005

+ combines productivity (number of documents) and impact (number of citations)
+ can be applied to any level of aggregation
+ encourages large amounts of impactful research work

- highly time-dependent measure
- ignores the researcher's age
- does not account for field differences

Example	Total Publications	Total Citations	*Citation Impact*	*h-index*
Researcher A	1	50	50	1
Researcher B	10	200	20	10
Researcher C	10	200	20	5

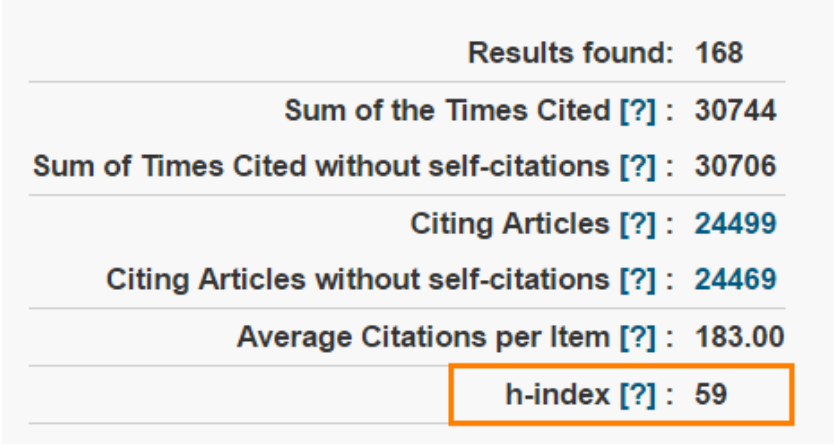
H-INDEX

To calculate h-index, sort number of publications in descending order according to number of citations

Rank	Author A	Author B	
1	24	1,020	
2	20	220	
3	18	110	
4	12	11	h-index: 4
5	6	4	
6	3	2	
7	2	1	
8	1	0	
9	1	0	
10	0	0	

h-index: 5

You can use Web of Science to find the h-index of an author



DIFFERENCES IN AVERAGE CITATION RATES

InCites™ Essential Science Indicators℠ THOMSON REUTERS

Indicators | Field Baselines | Citation Thresholds

Field Baselines

Baselines are annualized expected citation rates for papers in a research field.

Citation Rates *are yearly averages of citations per paper.*

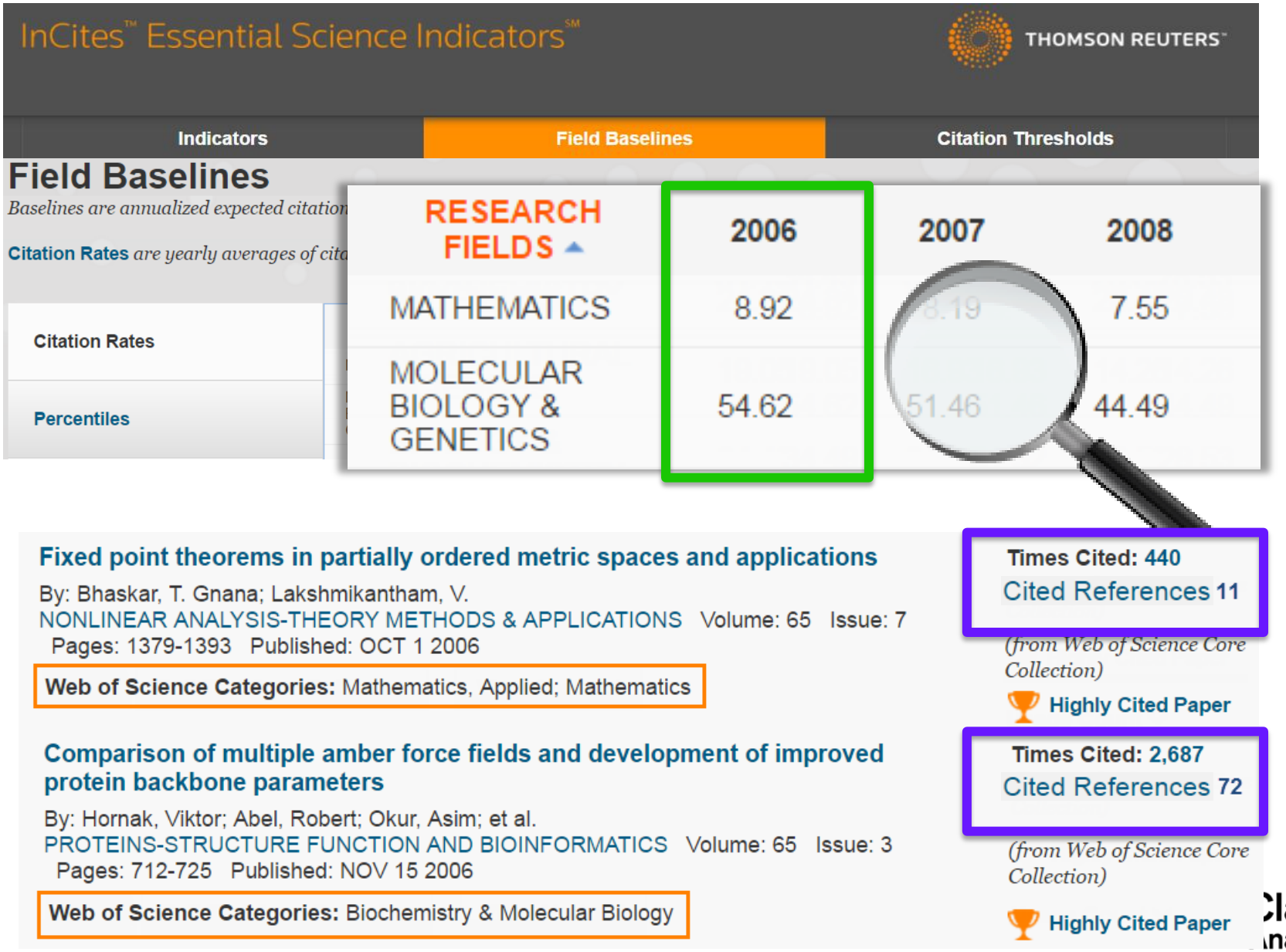
RESEARCH FIELDS	2007	2008	2009	2010	2011	2012	20	
ALL FIELDS	23.80	21.64	19.81	17.77	15.13	12.61		
AGRICULTURAL SCIENCES	18.02	15.28	13.81	12.49	10.50	8.77		
BIOLOGY & BIOCHEMISTRY	32.79	30.05	27.72	24.00	20.16	16.68		
CHEMISTRY	23.96	23.54	21.70	20.41	18.18	16.12		
CLINICAL MEDICINE	25.73	23.20	21.21	18.62	15.78	13.07		
COMPUTER SCIENCE	13.18	12.29	12.15	10.57	9.10	7.17		
ECONOMICS & BUSINESS	18.74	15.60	13.87	11.97	9.78	7.36		
ENGINEERING	13.05	12.03	12.04	10.98	9.60	8.00	6.62	4.93
ENVIRONMENT/ECOLOGY	28.88	26.32	22.66	20.40	16.94	14.20	10.69	7.53
GEOSCIENCES	23.92	22.58	21.07	18.11	15.99	12.87	10.04	6.97
IMMUNOLOGY	38.78	35.41	32.07	27.75	23.39	18.55	14.84	10.56
MATERIALS SCIENCE	20.92	19.12	18.36	17.90	15.90	13.88	11.33	9.01
MATHEMATICS	8.68	8.07	7.25	6.49	5.29	4.16	3.13	2.22
MICROBIOLOGY	31.44	28.12	25.89	23.12	18.17	14.83	12.04	8.70
MOLECULAR BIOLOGY & GENETICS	53.86	46.85	42.59	37.09	31.15	24.29	19.01	12.96

Citation Impact can vary significantly across different disciplines and time periods.

Cannot be compared without some form of normalization to allow for the differences in fields and time

Clarivate Analytics

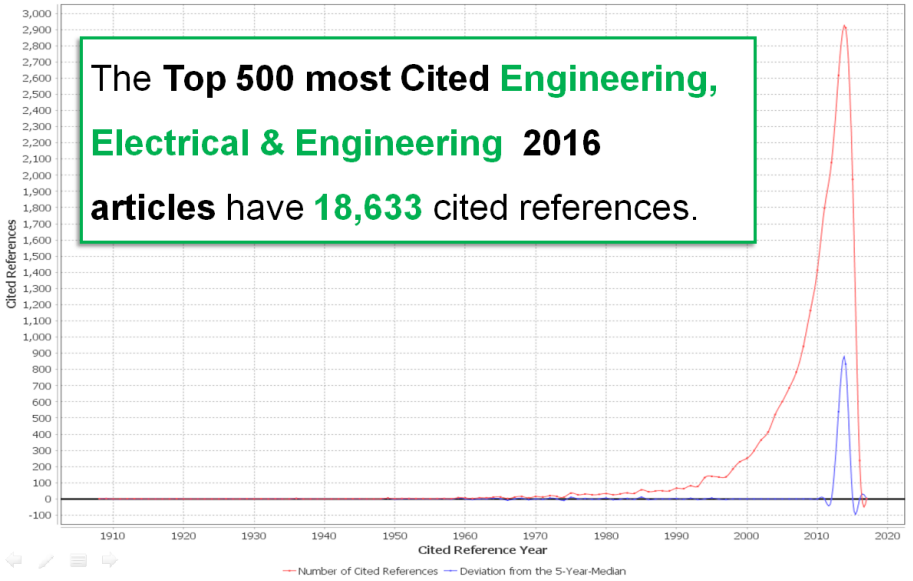
DIFFERENCES IN AVERAGE CITATION RATES



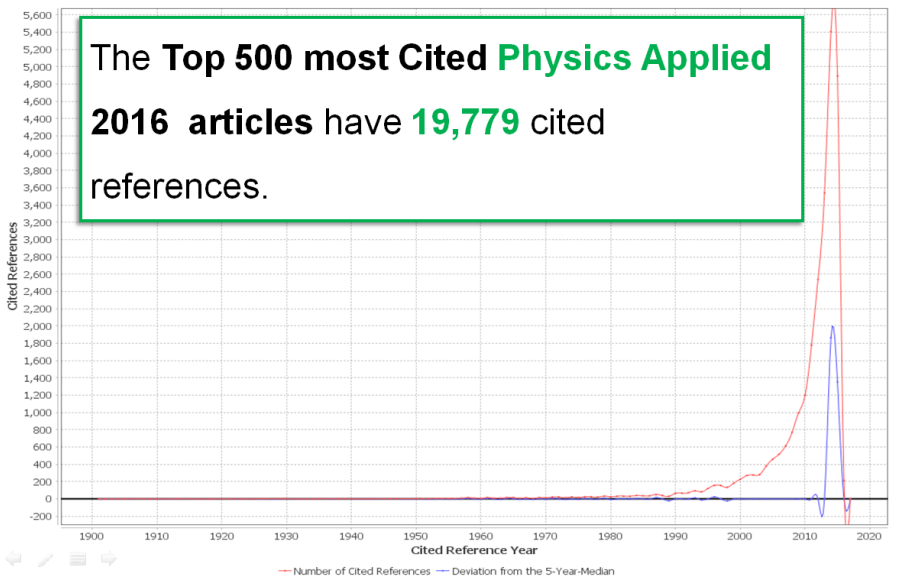
Clarivate Analytics

EXAMPLES OF DIFFERENCES IN CITING BEHAVIOUR PER CATEGORIES

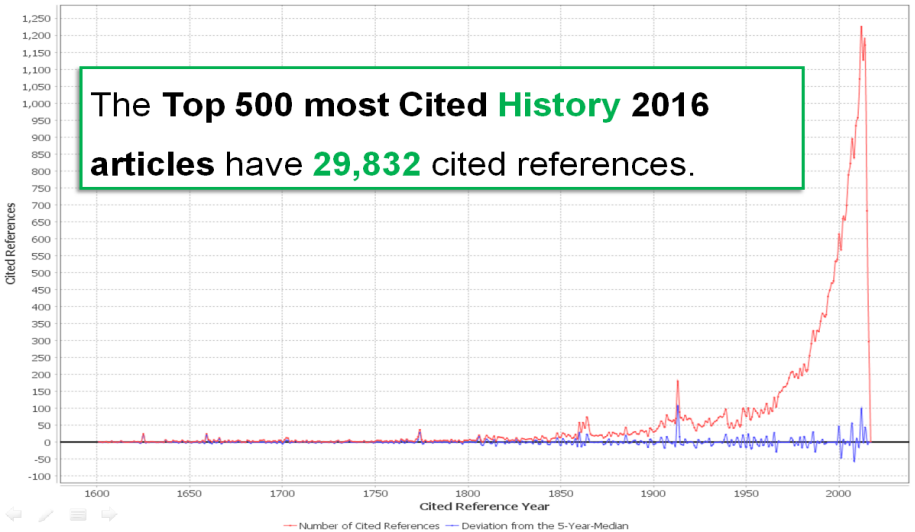
Engineering, Electrical & Engineering Top 500 Most Cited 2016 Articles: Cited References Distribution



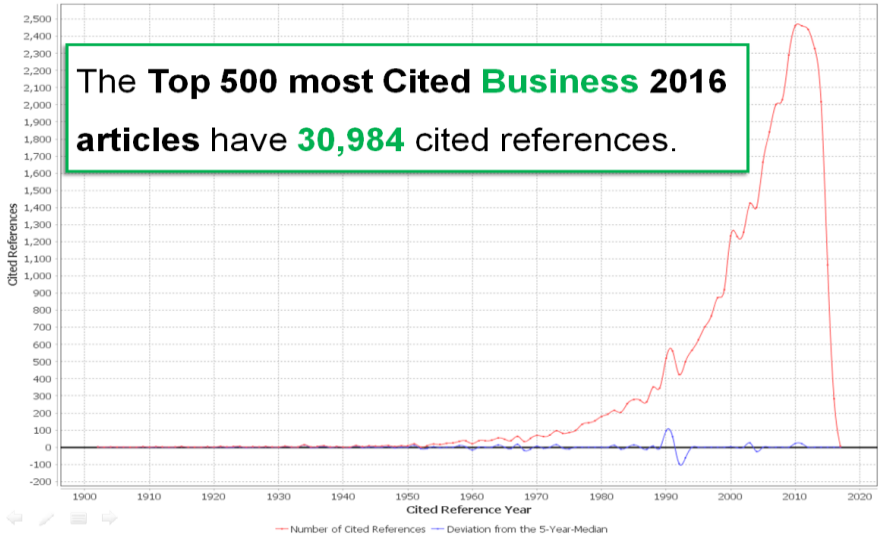
Physics Applied Top 500 Most Cited 2016 Articles: Cited References Distribution



History Top 500 Most Cited 2016 Articles: Cited References Distribution



Business Top 500 Most Cited 2016 Articles: Cited References Distribution



WHY NORMALIZATION?

“Normalization is a prerequisite for reliable benchmarks and hence for evaluative science policy” (Debackere, 2015)

The average number of citations varies significantly across disciplines and journals

NECESSITY: **FIELD** AND **JOURNAL** NORMALIZATION

Citations are dynamic; they grow over time and cannot be compared across different time periods.
Also the “citation maturity” rate differs between fields

NECESSITY: **TIME** NORMALIZATION

Different publication types have different citation behaviour, an article does not statistically receive as much citations as a review

NECESSITY**: DOCUMENT TYPE** NORMALIZATION

NORMALIZATION AT PAPER LEVEL (Category)

How many citations should I expect from my papers?
How do my papers perform in my field?
How do other researchers perform in my field?

Average of citations received by an article published in 2012 in the Management category

Indicator of performance in the Management category for this Article published in 2006:
If>1, performs higher than average
If<1, performs lower than average

Article Title	Authors	Source	Research Area	Volume	Issue	Pages	Publication Date	Times Cited	Journal Expected Citations	Category Expected Citations	Journal Normalized Citation Impact	Category Normalized Citation Impact	Percentile in Subject Area	Journal Impact Factor
Toward a Theory of Coordinating: Creating Coordinating Mechanisms in Practice	Jarzabkowski, Paula A.; Le, Jane K.; Feldman, Martha S.	ORGANIZATION SCIENCE	MANAGEMENT	23	4	907-927	2012	43	21.88	7.34	1.97	5.86	1.97	3.36

Times Cited/Category Expected Citations = 43/7.34 = 5.86

NORMALIZATION AT PAPER LEVEL (Journal)

How do my papers perform in the journals I publish?
How is my research perceived by the journals I publish in?
Is there a journal article level metric to help me go beyond the Journal Impact Factor ?

Indicator of performance of this Article in the Organization Science journal:
If>1, performs higher than average
If<1, performs lower than average

Average of citations received by an article published in 2012 in the Organization Science journal

Article Title	Authors	Source	Research Area	Volume	Issue	Pages	Publication Date	Times Cited	Journal Expected Citations	Category Expected Citations	Journal Normalized Citation Impact	Category Normalized Citation Impact	Percentile in Subject Area	Journal Impact Factor
Toward a Theory of Coordinating: Creating Coordinating Mechanisms in Practice	Jarzabkowski, Paula A.; Le, Jane K.; Feldman, Martha S.	ORGANIZATION SCIENCE	MANAGEMENT	23	4	907-927	2012	43	21.88	7.34	1.97	5.86	1.97	3.36

Times Cited/Journal Expected Citations = 43/21.88 = 1.97

PERCENTILES: DOCUMENTS IN TOP 1% & 10%

Knowing I am better than average is not enough.
Where do my research papers stand in competition to other papers?
Do I have highly cited papers amongst my publications?

Article Title	Authors	Source	Research Area	Volume	Issue	Pages	Publication Date	Times Cited	Journal Expected Citations	Category Expected Citations	Journal Normalized Citation Impact	Category Normalized Citation Impact	Percentile in Subject Area	Journal Impact Factor
Toward a Theory of Coordinating: Creating Coordinating Mechanisms in Practice	Jarzabkowski, Paula A.; Le, Jane K.; Feldman, Martha S.	ORGANIZATION SCIENCE	MANAGEMENT	23	4	907-927	2012	43	21.88	7.34	1.97	5.86	1.97	3.36

Percentile is a value above which a certain proportion of the observations fall

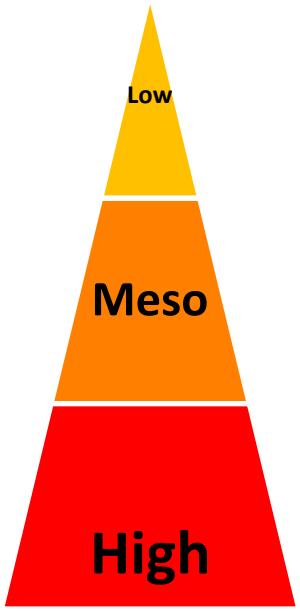
Percentiles allow the classification of publications into meaningful citation impact classes

The smaller the percentile number, the higher the number of citations (in a scale of 0-100)

HIGHLY CITED PAPERS & HOT PAPERS (ESI)

A class of selected indicators measuring scientific excellence and top performance which can be used to benchmark research performance against field baselines worldwide

Level of Aggregation



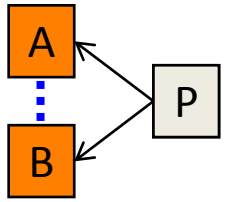
	Citation Percentile	Data years examined
Highly Cited Papers	1%	10
Hot Papers	0.1%	2
Researchers	1%	10
Institutions	1%	10
Journals	50%	10
Countries	50%	10

CO-CITATION ANALYSIS

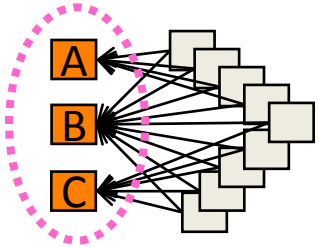
Co-Citation Analysis and Clustering: How Does It Work?

Counting the number of times that a given pair of documents (or authors or journals) are co-cited. The more papers that co-cite the pair, the stronger the relationship. This relationship is *dynamic (new papers may be published which cite the pair) and forward looking.*

1. When paper A and B are "co-cited" by paper P, A and B are likely to have topical similarity.



2. When co-citation is frequent, it forms a group of papers that are topically associated to one another.



Henry Small, "Co-Citation in the Scientific Literature: A New Measure of the Relationship Between Two Documents," *Journal of the American Society for Information Science,* 24(4): 265-69, July/August 1973

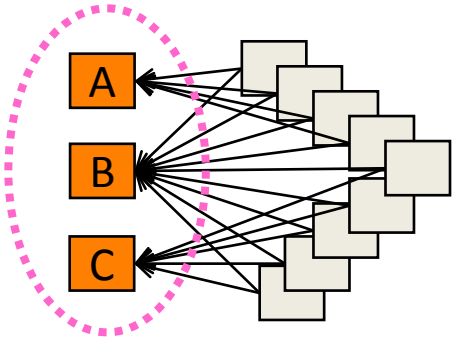
RESEARCH FRONTS (ESSENTIAL SCIENCE INDICATORS)

Clusters of papers belonging to the 1% most highly cited papers that are frequently cited together; A Research Front is formed.

Research Fronts consist of a group of highly cited Core Papers and a set of Citing Papers that frequently cite the Core Papers

A
B
C

are highly cited and influential papers that have left a mark in their field



Co-citing papers reveal the uptake of data, techniques and concepts revealed in the Core Papers

Top Three Research Fronts in Chemistry

Map View by Top / Hot / Highly Cited Papers — Show Visualization +

Report View by Selection — Customize

	Research Fronts	Highly Cited Papers	Mean Year
1	NIR HYPERSPECTRAL IMAGING SYSTEM;INFRARED HYPERSPECTRAL IMAGING;NEAR-INFRARED HYPERSPECTRAL IMAGING;NIR HYPERSPECTRAL IMAGING;VISIBLE/NEAR-INFRARED HYPERSPECTRAL IMAGING PREDICTION	48	2012.4
2	ENANTIOSELECTIVE ELECTROPHILIC TRIFLUOROMETHYLTHIOLATION;LEWIS ACID-CATALYZED ELECTROPHILIC TRIFLUOROMETHYLTHIOLATION;BROSTED ACID-CATALYZED ELECTROPHILIC TRIFLUOROMETHYLTHIOLATION;BASE-CATALYZED ELECTROPHILIC TRIFLUOROMETHYLTHIOLATION;ELECTROPHILIC AROMATIC TRIFLUOROMETHYLTHIOLATION	46	2013.6
2	EMISSION DYE BASED RED FLUORESCENT ORGANIC NANOPARTICLES;TETRAPHENYLETHENE-BASED AGGREGATION-INDUCED EMISSION FLUORESCENT ORGANIC NANOPARTICLES;POLYMERIZABLE AGGREGATION-INDUCED EMISSION DYE-BASED FLUORESCENT NANOPARTICLES;EMISSION DYE-BASED FLUORESCENT ORGANIC NANOPARTICLES;AIE BASED FLUORESCENT ORGANIC NANOPARTICLES	46	2012.7

Research fronts are drivers of innovation and scientific discovery in their fields

The name of the Research Fronts comes from a summarization of the titles of the cited papers

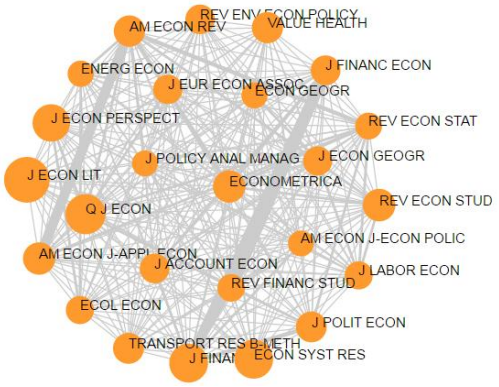
Web of Science
Trust the difference

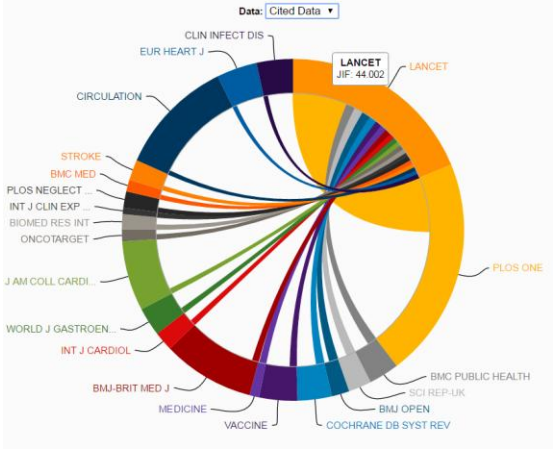
Clarivate Analytics

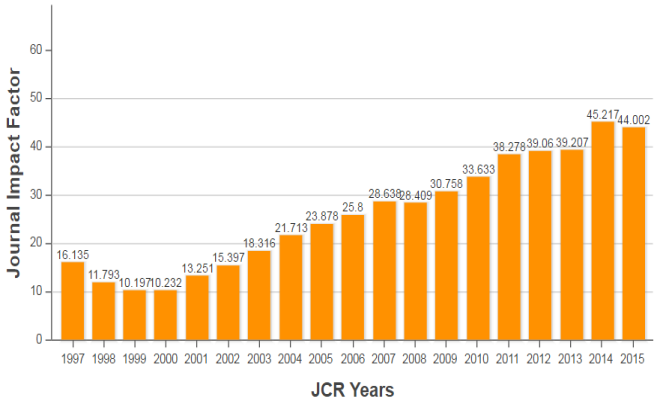
JOURNAL RANKING INDICATORS

Some commonly used journal ranking indicators available via the Journal Citation Reports

- Journal Impact Factor
- 5 year Journal Impact Factor
- Immediacy index
- Eigenfactor score







THE JOURNAL IMPACT FACTOR

- The journal impact factor is a measure of the frequency with which the "average article" in a journal has been cited in a particular year.
- The impact factor will help you evaluate a journal's relative importance, especially when you compare it to others in *the same field*

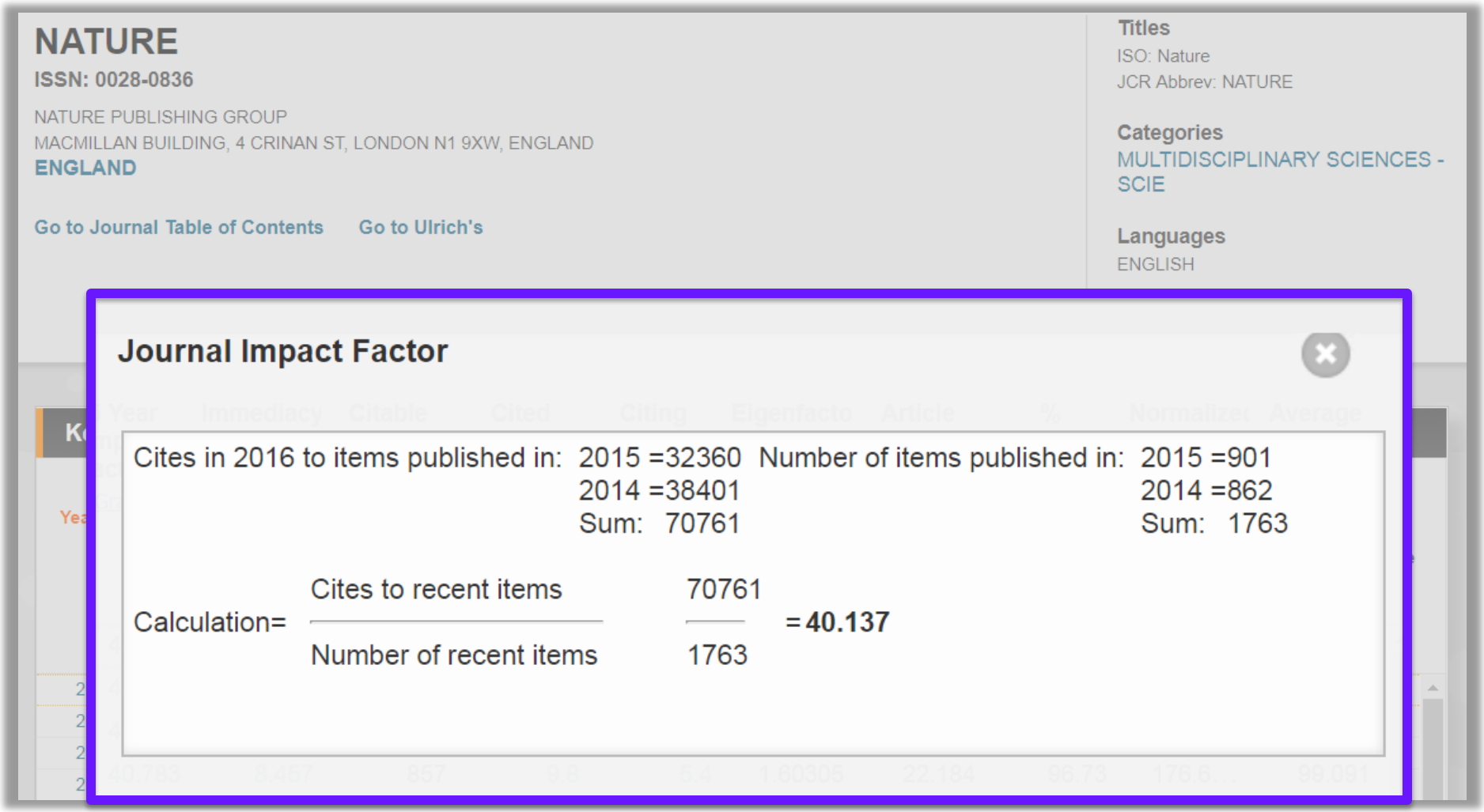
Ranking journals within the same field can help:

- To spot new journals increasing their impact
- To learn evolving contents of existing journals

One common misuse of the IF is to evaluate papers, or *people*

Responsible use of Journal Ranking Indicators*: Use Journal Impact Factor at Journal Level*

CALCULATION OF THE IMPACT FACTOR



JIF: DISPARITIES IN CATEGORIES

APPLIED SURFACE SCIENCE

Impact Factor

3.387 *2016*

3.184 *5 year*

JCR® Category	*Rank in Category*	*Quartile in Category*
CHEMISTRY, PHYSICAL	**46 of 146**	**Q2**
MATERIALS SCIENCE, COATINGS & FILMS	**1 of 19**	**Q1**
PHYSICS, APPLIED	**32 of 148**	**Q1**
PHYSICS, CONDENSED MATTER	**19 of 67**	**Q2**

Data from the 2016 edition of Journal Citation Reports

Publisher
ELSEVIER SCIENCE BV, PO BOX 211, 1000 AE AMSTERDAM, NETHERLANDS

ISSN: 0169-4332
eISSN: 1873-5584

Research Domain
Chemistry
Materials Science
Physics

JIF 2 YEARS: DISPARITIES IN CATEGORIES

The picture is different for Sciences disciplines such as ‘Clinical Medicine’ category where we can see a shorter citation lag

	Full Journal Title	Total Cites	Journal Impact Factor	5 Year Impact Factor	Imm Inc
1	NEW ENGLAND JOURNAL OF MEDICINE	283,525	59.558	56.170	
2	JAMA-JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION	129,909	37.684	33.569	
3	JOURNAL OF CLINICAL ONCOLOGY	141,362	20.982	18.021	
4	BMJ-British Medical Journal	93,118	19.697	17.687	
5	Nature Reviews Clinical Oncology	5,447	18.786	16.376	
6	GASTROENTEROLOGY	67,726	18.187	15.417	
7	JOURNAL OF THE AMERICAN COLLEGE OF CARDIOLOGY	84,932	17.759	15.776	
8	CIRCULATION	155,582	17.202	16.252	
9	Nature Reviews Endocrinology	4,758	15.432	14.125	

JIF 5 YEARS: DISPARITIES IN CATEGORIES

Citations accumulate slower for Social Sciences journals across time thus the 5 Year Impact Factor is often higher than its 2-Year counterpart

	Full Journal Title	Total Cites	Journal Impact Factor	5 Year Impact Factor
1	JOURNAL OF ECONOMIC LITERATURE	6,360	6.614	11.168
2	QUARTERLY JOURNAL OF ECONOMICS	19,063	5.538	9.607
3	JOURNAL OF FINANCE	24,013	5.290	7.728
4	JOURNAL OF ECONOMIC PERSPECTIVES	8,096	5.012	6.658
5	JOURNAL OF ACCOUNTING & ECONOMICS	5,373	3.535	5.973
5	JOURNAL OF POLITICAL ECONOMY	18,039	3.750	5.973
7	JOURNAL OF FINANCIAL ECONOMICS	18,347	3.541	5.833
8	American Economic Journal-Applied Economics	1,412	4.038	5.428
9	ECONOMETRICA	24,957	4.053	5.399
10	REVIEW OF FINANCIAL STUDIES	9,405	3.119	5.265
11	AMERICAN ECONOMIC REVIEW	35,805	3.833	5.140

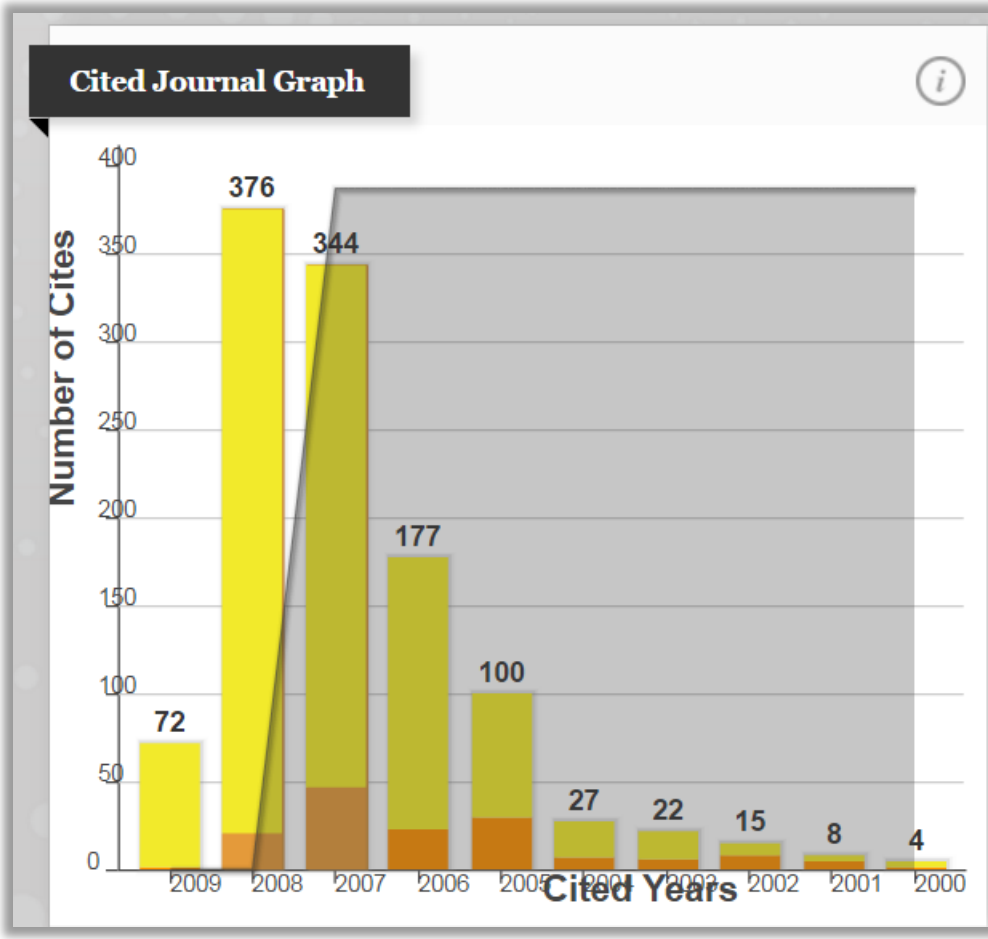
SELF CITATIONS AT JOURNAL LEVEL

- REV BRAS FARMACOGN: Regional coverage Expansion
- First Journal Impact Factor in 2009 was 3.462

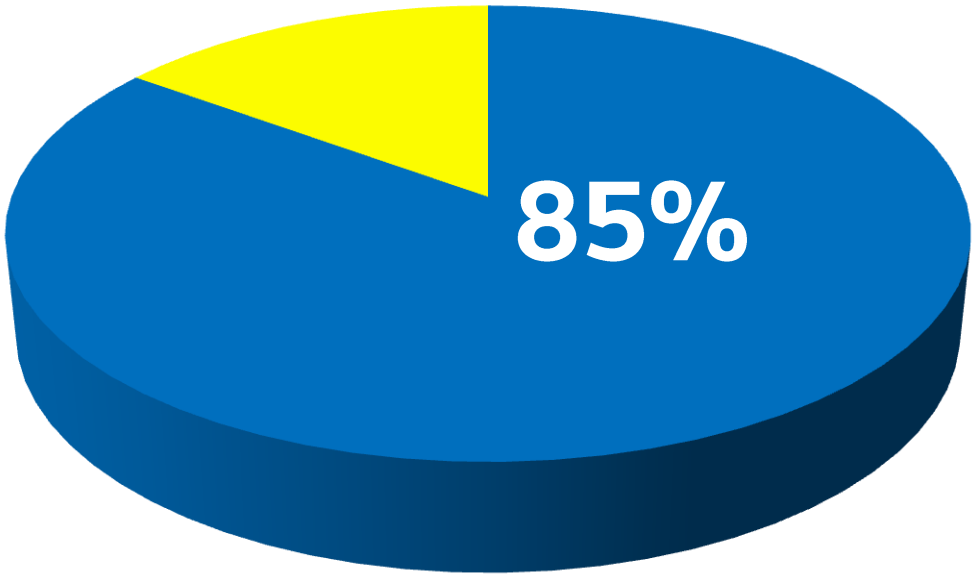
Key Indicators

Year	Total Cites Graph	Journal Impact Factor Graph	Impact Factor Without Journal Self Cites Graph	5 Year Impact Factor Graph	Immediacy Index Graph
2016	1,415	1.059	0.957	1.197	0.270
2015	1,134	0.956	0.762	0.914	0.049
2014	1,057	0.834	0.806	0.807	0.029
2013	966	0.796	0.689	0.865	0.041
2012	822	0.676	0.509	0.886	0.098
2009	1,163	3.462	0.326	Not A…	0.576

Journal was suppressed from 2010 & 2011 JCR



SELF CITATIONS AT JOURNAL LEVEL



Most journals have self-citation rates of less than or equal to 15%

Source: *JCR* Science Edition (2010)

- Excessive self-citation weakens the integrity of the journal's Impact Factor
- Journals with excessive self-citation may be suppressed from Journal Citation Reports until the problem is corrected

More information on journal suppression is available at: http://wokinfo.com/media/pdf/jcr-suppression.pdf

SPECIAL CASE: MUTUAL CITATIONS

CELL TRANSPLANTATION
ISSN: 0963-6897
COGNIZANT COMMUNICATION CORP
3 HARTSDALE ROAD, ELMSFORD, NY 10523-3701
USA

Journal self-citations are concentrated in Journal Impact Factor years

High-value citation partners show extreme concentration

Cited Journal Data

	Impact	Citing Journal	All Yrs	2010	2009	2008	2007	2008	2007	2006	2005	2004	2003	2002	2001	Rest
1		ALL Journals	3,711	95	777	805	331	805	331	277	172	208	186	153	95	612
2		ALL OTHERS (360)	360	5	32	52	38	52	38	40	24	24	27	15	15	88
3	6.204	CELL TRANSPLANT	718	49	250	234	58	234	58	33	16	19	13	8	8	30
4	1.699	MED SCI MONITOR	445	0	226	219	0	219	0	0	0	0	0	0	0	0
5	1.524	THESCIENTIFICWORLDJO	96	0	96	0	0	0	0	0	0	0	0	0	0	0
6	4.636	TISSUE ENG	96	1	7	7	7	7	7	13	3	10	8	9	4	27
7	7.883	BIOMATERIALS	77	0	9	15	7	15	7	3	6	12	1	3	1	20
8	0.993	TRANSPL P	67	4	15	13	4	13	4	7	3	1	6	5	2	7
9	3.676	TRANSPLANTATION	57	2	4	7	1	7	1	4	3	6	3	6	1	20
10	1.379	ADV EXP MED BIOL	46	0	1	8	2	8	2	6	1	3	3	2	3	17
11	4.791	STEM CELLS DEV	42	0	6	8	7	8	7	4	1	5	5	0	1	5
12	3.044	J BIOMED MATER RES A	41	0	0	6	1	6	1	2	3	4	1	6	3	15
13		METHOD MOL BIOL	34	0	0	2	3	2	3	7	3	5	8	1	1	4
14	13.577	ADV DRUG DELIVER REV	31	1	3	4	4	4	4	2	4	3	0	4	0	6
15	2.925	CYTOTHERAPY	30	0	2	6	7	6	7	3	1	2	2	4	0	3

SPECIAL CASE: MUTUAL CITATIONS

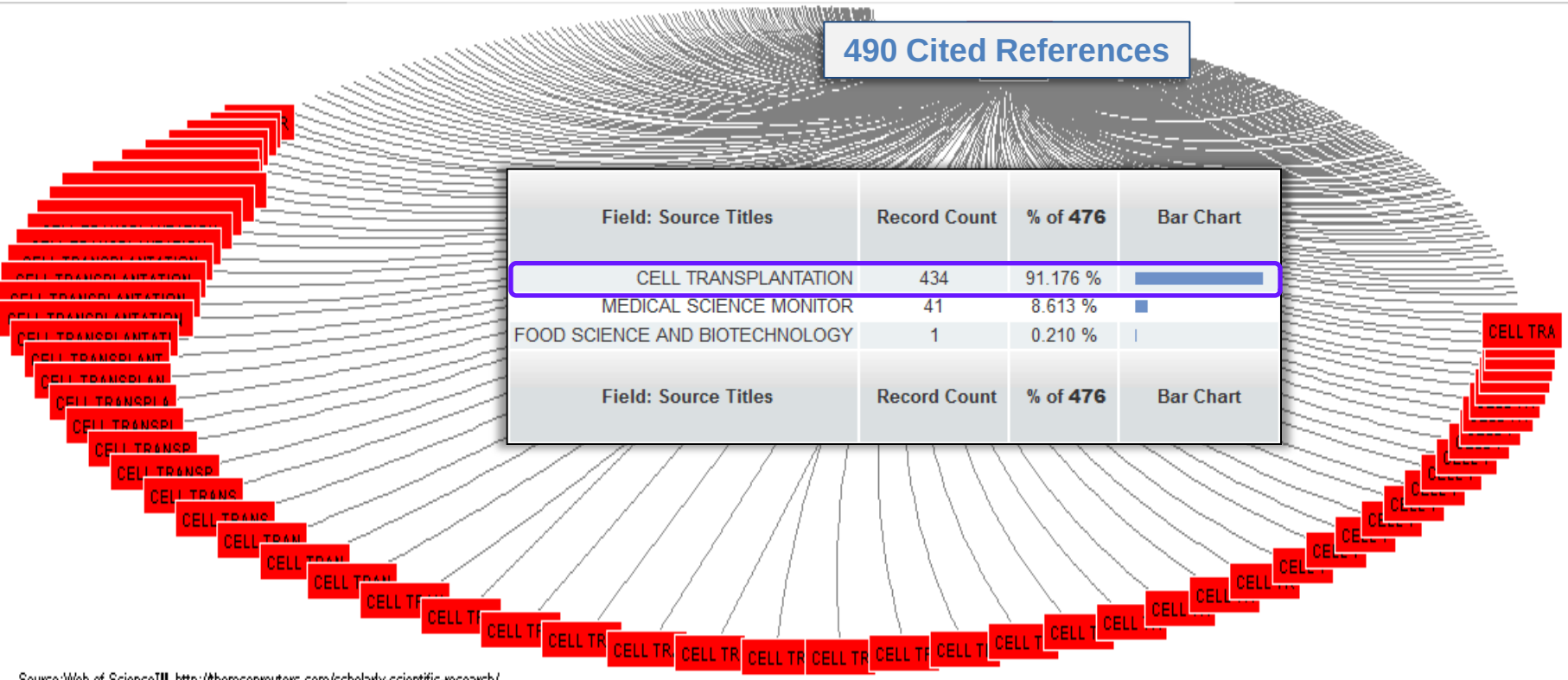
Stem cells have the potential to rejuvenate regenerative medicine research

By: Eve, DJ (Eve, David J.)[1]; Fillmore, RW (Fillmore, Randolph W.)[1]; Borlongan, CV (Borlongan, Cesar V.)[1]; Sanberg, PR (Sanberg, Paul R.)[1]

MEDICAL SCIENCE MONITOR
Volume: 16 **Issue:** 10 **Pages:** RA197-RA217
Published: 2010
View Journal Information

Citation Network

9 Times Cited
490 Cited References
View Related Records
View Citation Map
Create Citation Alert
(data from Web of Science ™ Core Collection)



Field: Source Titles	Record Count	% of 476	Bar Chart
CELL TRANSPLANTATION	434	91.176 %	
MEDICAL SCIENCE MONITOR	41	8.613 %	
FOOD SCIENCE AND BIOTECHNOLOGY	1	0.210 %	
Field: Source Titles	**Record Count**	**% of 476**	**Bar Chart**

Source:Web of Science™, http://thomsonreuters.com/scholarly-scientific-research/

SPECIAL CASE: “NEGATIVE” CITATIONS

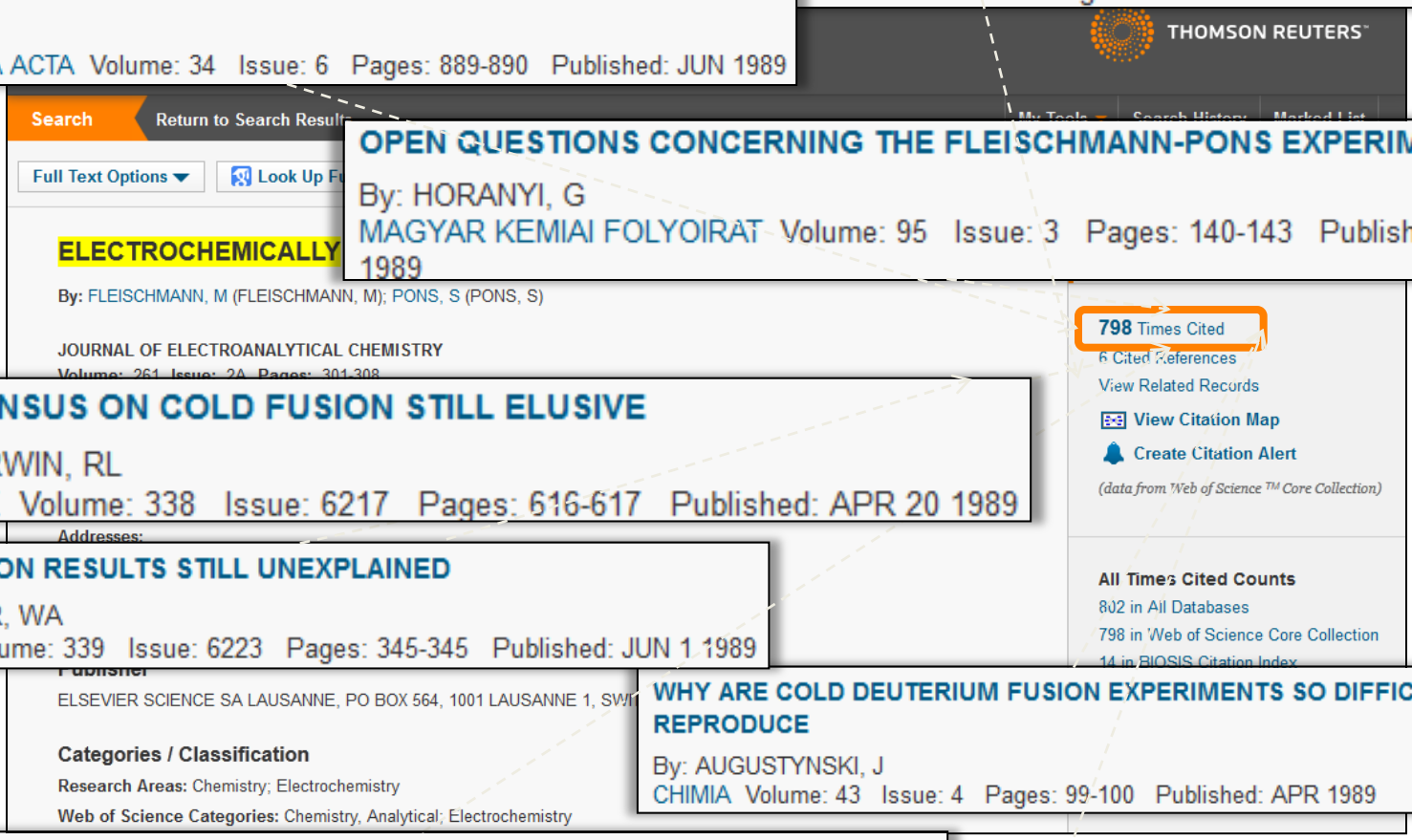
NO NEW FUSION UNDER THE SUN

339 Issue: 6221 Pages: 180-180 Published: MAY 18 1989

SOME DOUBTS ABOUT THE OCCURRENCE OF ELECTROCHEMICALLY INDUCED NUCLEAR-FUSION OF DEUTERIUM

By: HORANYI, G

ELECTROCHIMICA ACTA Volume: 34 Issue: 6 Pages: 889-890 Published: JUN 1989



OPEN QUESTIONS CONCERNING THE FLEISCHMANN-PONS EXPERIMENT

By: HORANYI, G

MAGYAR KEMIAI FOLYOIRAT Volume: 95 Issue: 3 Pages: 140-143 Published: MAR 1989

CONSENSUS ON COLD FUSION STILL ELUSIVE

By: GARWIN, RL

NATURE Volume: 338 Issue: 6217 Pages: 616-617 Published: APR 20 1989

COLD FUSION RESULTS STILL UNEXPLAINED

By: FOWLER, WA

NATURE Volume: 339 Issue: 6223 Pages: 345-345 Published: JUN 1 1989

WHY ARE COLD DEUTERIUM FUSION EXPERIMENTS SO DIFFICULT TO REPRODUCE

By: AUGUSTYNSKI, J

CHIMIA Volume: 43 Issue: 4 Pages: 99-100 Published: APR 1989

COLD FUSION - WHATS GOING ON

By: MCCEVOY, AJ; OSULLIVAN, CTD

NATURE Volume: 338 Issue: 6218 Pages: 711-712 Published: APR 27 1989

PROBLEMS WITH THE GAMMA-RAY SPECTRUM IN THE FLEISCHMANN ET AL EXPERIMENTS

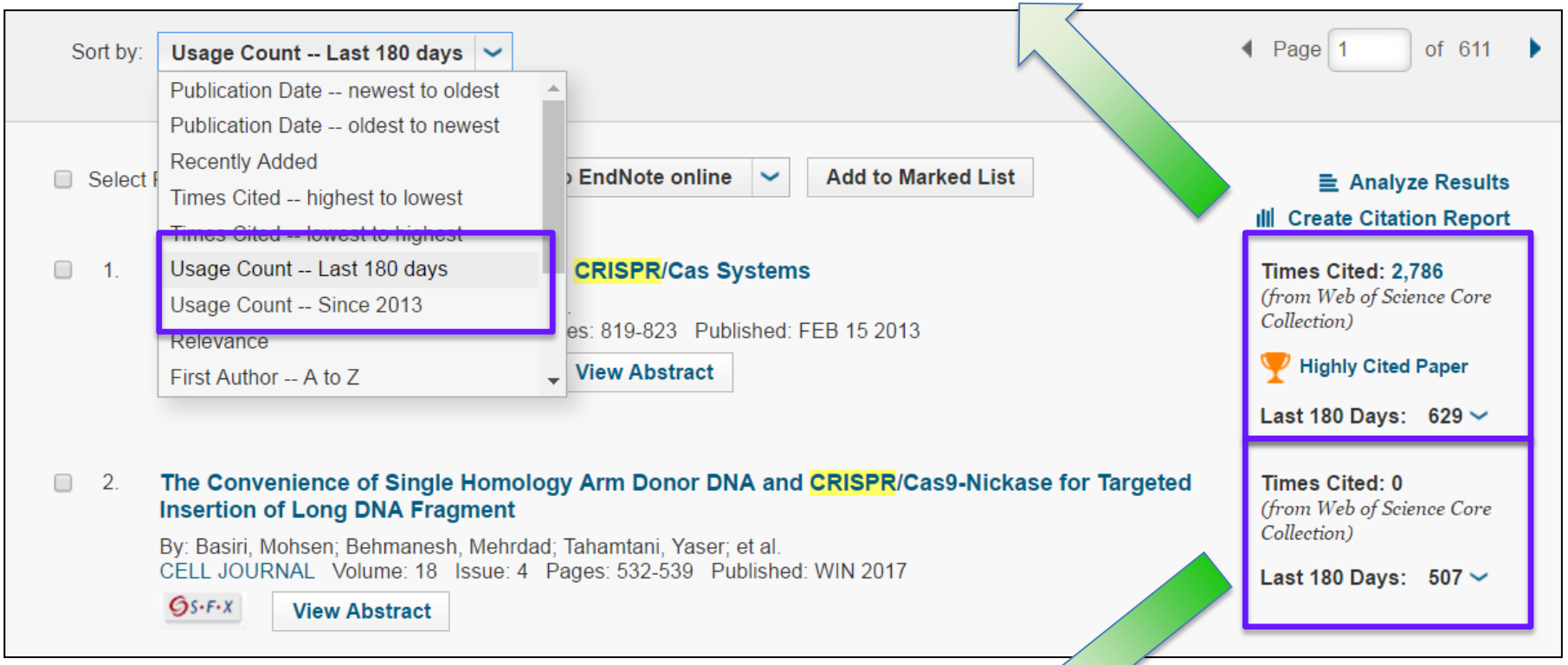
By: PETRASSO, RD; CHEN, X; WENZEL, KW; et al.

NATURE Volume: 339 Issue: 6221 Pages: 183-185 Published: MAY 18 1989

USAGE METRICS: PAPERS THAT ARE RECEIVING A LOT OF ATTENTION FROM THE SCIENTIFIC COMMUNITY, WITHOUT BEING CITED

Usage count is a measure of usage on the platform can show ***INTEREST*** in a publication or a topic prior to, or in the absence of citation activity.

The 1st most used CRISPR paper is a highly cited paper that has received 2,786 citation published in 2013.



The 2nd most used CRISPR publication has received 0 citations being newly published but wos researchers are highly interested in it.

Web of Science
Trust the difference

Clarivate Analytics

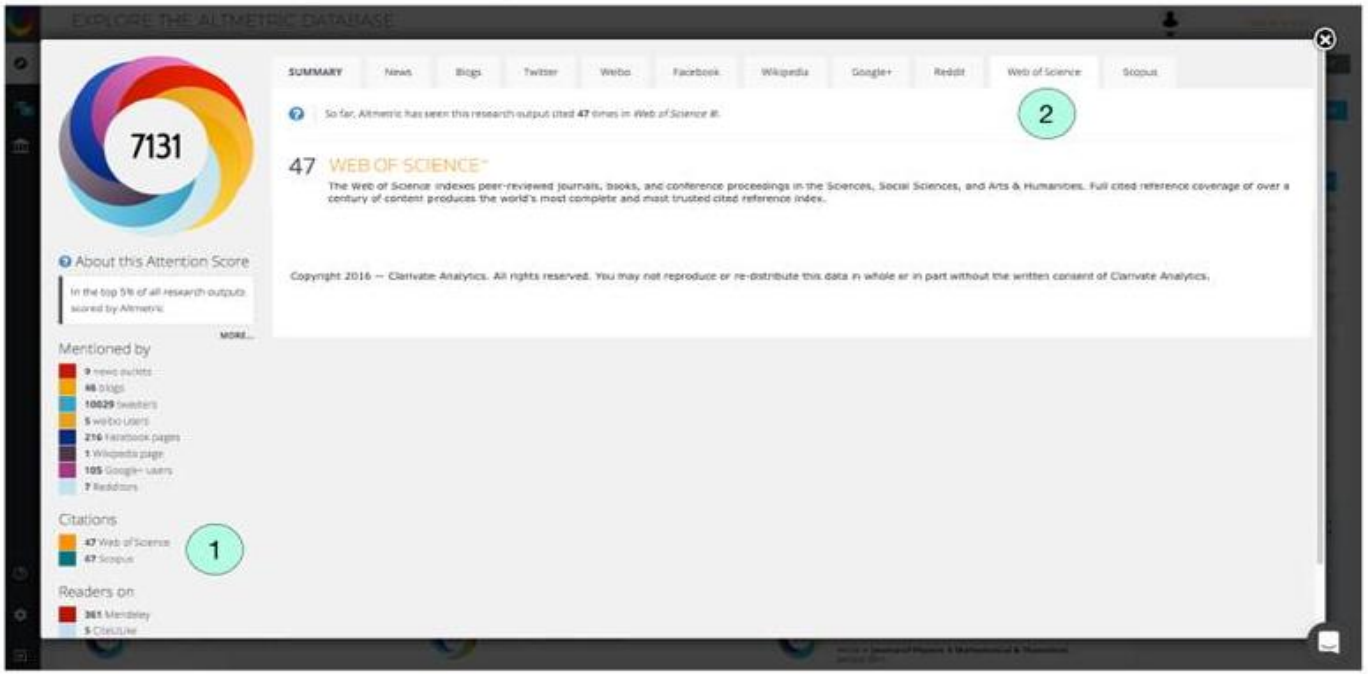
PARTNERSHIP WITH ALTMETRICS

A **new, strategic partnership with Altmetrics**, a Digital Science business since January 2017.

Designed **to provide mutual subscribers of the Web of Science and Altmetric Explorer with Times Cited counts visible in Altmetric Explorer with outbound links into the Web of Science**, hence with a more complete picture of the interest in their research and its impact. .

A link from Explorer to the Web of Science Times Cited count is one of the most requested enhancements from their customers.

We are committed to inserting our data into varied customers workflows.



Altmetric

Clarivate Analytics

RESPONSIBLE USE OF BIBLOMETRIC INDICATORS: A "BASKET" OF INDICATORS – NO MAGIC RECIPE FITS ALL

- What *can* and *what* should be measured?
- What are appropriate measures for the purpose?

Productivity And Impact	Normalization	Top Performance	Scientific Collaborations	Journal Ranking Indicators
Web of Science Documents	Category Normalized Citation Impact	% Documents in Top 1%	% Industry Collaborations	Journal Impact Factor
Times Cited	Category Expected Citations	% Documents in Top 10%	% International Collaborations	Impact Factor w/o Self Cites
Citation Impact	Journal Normalized Citation Impact	Average percentile	Collaborations with Organizations	5 year Impact Factor
% of documents cited	Journal Expected Citations	Highly Cited Papers	Collaborations with Countries	Immediacy Index
H Index		Hot Papers	Collaborations with Authors	Eigenfactor

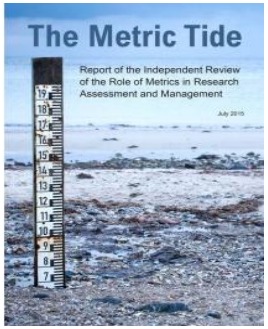
Web of Science
Trust the difference

Clarivate Analytics

RESPONSIBLE USE OF BIBLIOMETRICS - OVERARCHING PRINCIPLES

Ten Rules in Using Publication and Citation Analysis

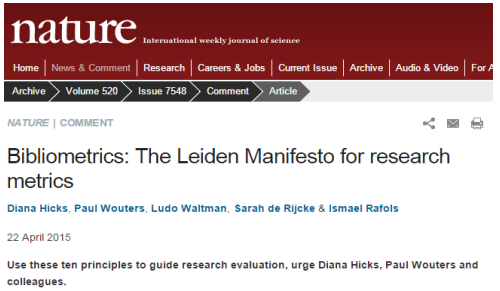
1. Consider whether available data can address the question
2. Confirm that the data collected are relevant to the question
3. Recognize the skewed nature of citation data
4. Judge whether data require editing to remove “artifacts”
5. Use relative measures, not just absolute counts
6. Choose publication types, field definitions, and years of data
7. Compare like with like
8. Obtain multiple measures
9. Decide on whole or fractional counting
10. Ask whether the results are reasonable



San Francisco DORA
Declaration on Research Assessment

And, above all, present the results openly and honestly

David Pendlebury (2008): “Using Bibliometrics in Evaluating Research”


nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For A

Archive > Volume 520 > Issue 7548 > Comment > Article

NATURE | COMMENT

Bibliometrics: The Leiden Manifesto for research metrics

Diana Hicks, Paul Wouters, Ludo Waltman, Sarah de Rijcke & Ismael Rafols

22 April 2015

Use these ten principles to guide research evaluation, urge Diana Hicks, Paul Wouters and colleagues.

PART III:

MEET THE WOS DATA

DATA & METADATA INDEXING
CONSISTENCY IS THE KEY TO VALIDITY

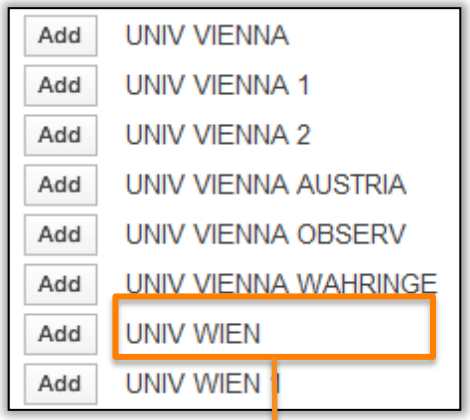
- Consistent indexing for complete analysis
  - Cover-to-cover indexing
  - All author names
  - All author addresses (affiliations)
  - Funding Agencies & Grant Numbers (Funding text)
  - Subject Area Classification
  - Open Access

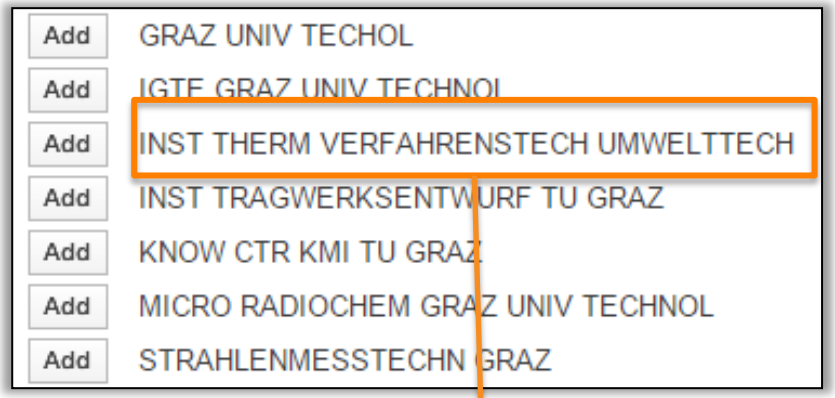
COVER TO COVER INDEXING IS ESSENTIAL FOR PRODUCING RELIABLE JOURNAL RANKING INDICATORS

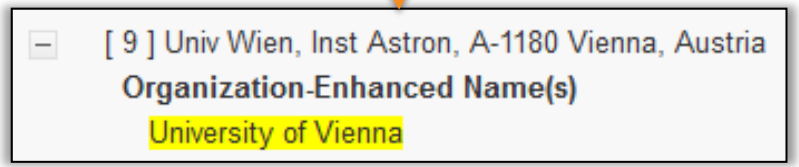
40+ document types curated and properly assigned to the correct type

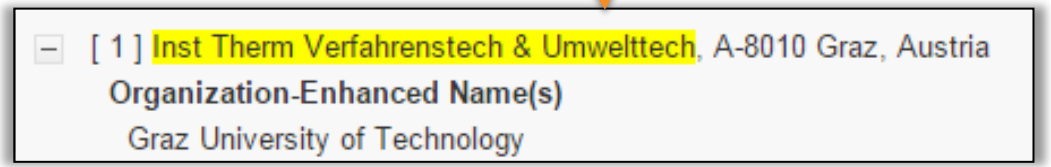
How can you claim that your Journal metrics (Impact Factor, SNIP, etc.) reflect reality if you do not index the entire journal?

DIFFERENT LEVELS OF METADATA QUALITY
ENHANCED ORGANIZATIONS NAMES





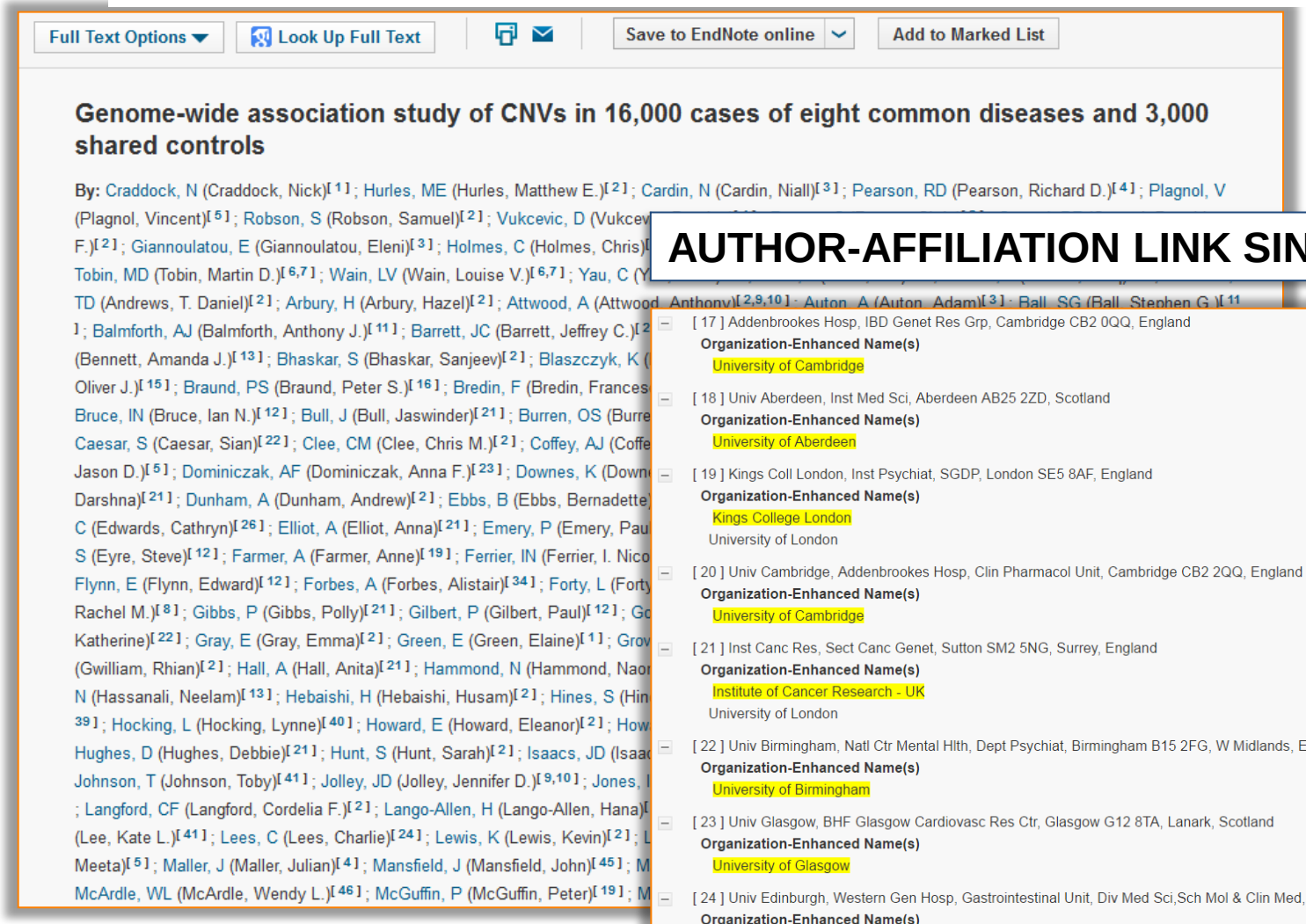




Unification rules sets are built in complete transparency, using internal and external expertise

We communicate rules to institutions → They validate/modify/ complete the rules → Rules are updated and applied to more than a century of publication activity

DIFFERENT LEVELS OF METADATA QUALITY
ALL AUTHOR NAMES, ALL ADDRESSES



DIFFERENT LEVELS OF METADATA QUALITY
FUNDING ACKNOWLEDGEMENTS SINCE 2008

Funding

Funding Agency	Grant N
International Global Atmospheric Chemistry (IGAC) project	
Climate Program Office of the National Oceanic and Atmospheric Administration (NOAA)	
Radiation Sciences Program of the National Aeronautic and Space Administration (NASA)	
CSC	
Tully Graphics	
IGAC via the Joint Institute for the Study of the Atmosphere and Ocean (JISAO) under NOAA	NA10OAR 2035
U.S. EPA	RD-83503
NASA	RD-83503
NSF	ATM 08-5
DOE	DE-SC00 DE-AC06
Royal Society Wolfson Research Merit award	
Joint DECC/Defra Met Office Hadley Centre Climate Programme	GA01101
U.S. Department of Energy (DOE), Office of Science, Scientific Discovery through Advanced Computing (SciDAC) program	
DOE Decadal and Regional Climate Prediction using Earth System Models (EaSM) program	
Ministry of Education, Culture, Sports, Science, and Technology (MEXT)	
Japan Science and Technology Agency (JST)	
global environment research fund of the Japanese Ministry of the Environment	A-1101
EUCAARI project (EU-FP6)	34684
U.S. National Science Foundation	
European Union Seventh Research Framework Programme (MACC project)	218793
U.S. NSF	ARC-06-
National Basic Research Program of China	2011CB4

View funding text

CURRENTLY WORKING TOWARDS UNIFICATION OF FUNDERS

MORE THAN 1,000 FUNDERS UNIFIED IN INCITES

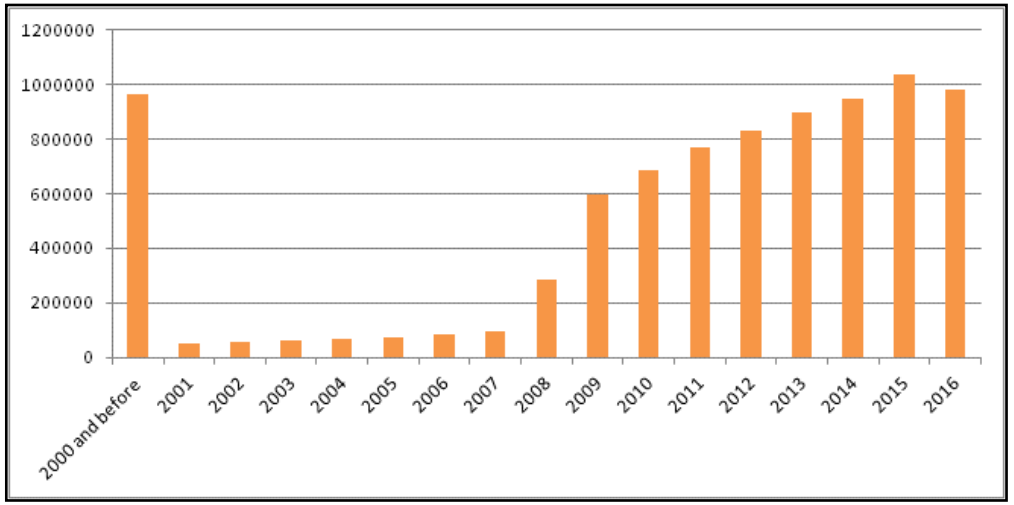
(British Heart Foundation, Medical Research Council, European Commission, NASA, HEFCE, NERC, RCUK, EPSRC, Wellcome Trust, Leverhulme Trust WHO, European Cooperation in Science and Technology (COST), Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT), Deutsche Forschungsgemeinschaft, Research Council of Norway, Dutch Cancer Society, etc)

China (2011CB403405). The views expressed in this paper are those of the authors and do not necessarily reflect the views of policies of the U.S. Environmental Protection Agency.

CAPTURING FUNDING DATA IN WEB OF SCIENCE

- Agency names and grant numbers are captured from an article's Acknowledgments section exactly as named
- Science Citation Index-Expanded (articles and reviews) in August 2008-present
- Social Sciences Citation Index (all document types) and Emerging Sources Citation Index articles in 2015-present
- Funding acknowledgments coming from Medline® and ResearchFish® are also indexed; 1.5 million of new funding acknowledgments (22% growth), resulting in a significant coverage for years before 2008 where ~1 million papers with funding info

Publication year	*Records with at least one funding agency*
2008	14,23%
2009	28,12%
2010	32,53%
2011	34,60%
2012	35,69%
2013	37,06%
2014	38,22%
2015	39,22%
2016	44,65%



DIFFERENT LEVELS OF METADATA QUALITY
OPEN ACCESS JOURNALS

Coming Soon..

Web of Science will provide direct access to additional, legal Open Access content

Clarivate Analytics has invested in technology so that you can soon:

- ✓ **Find Hybrid Gold OA articles when searching the Web of Science**
- ✓ **Find Green OA articles when searching the Web of Science**

Article-level Open Access identification will help you **find legally available Green & Hybrid articles in the Web of Science.**

To develop this capability, we have given a grant to Impactstory.

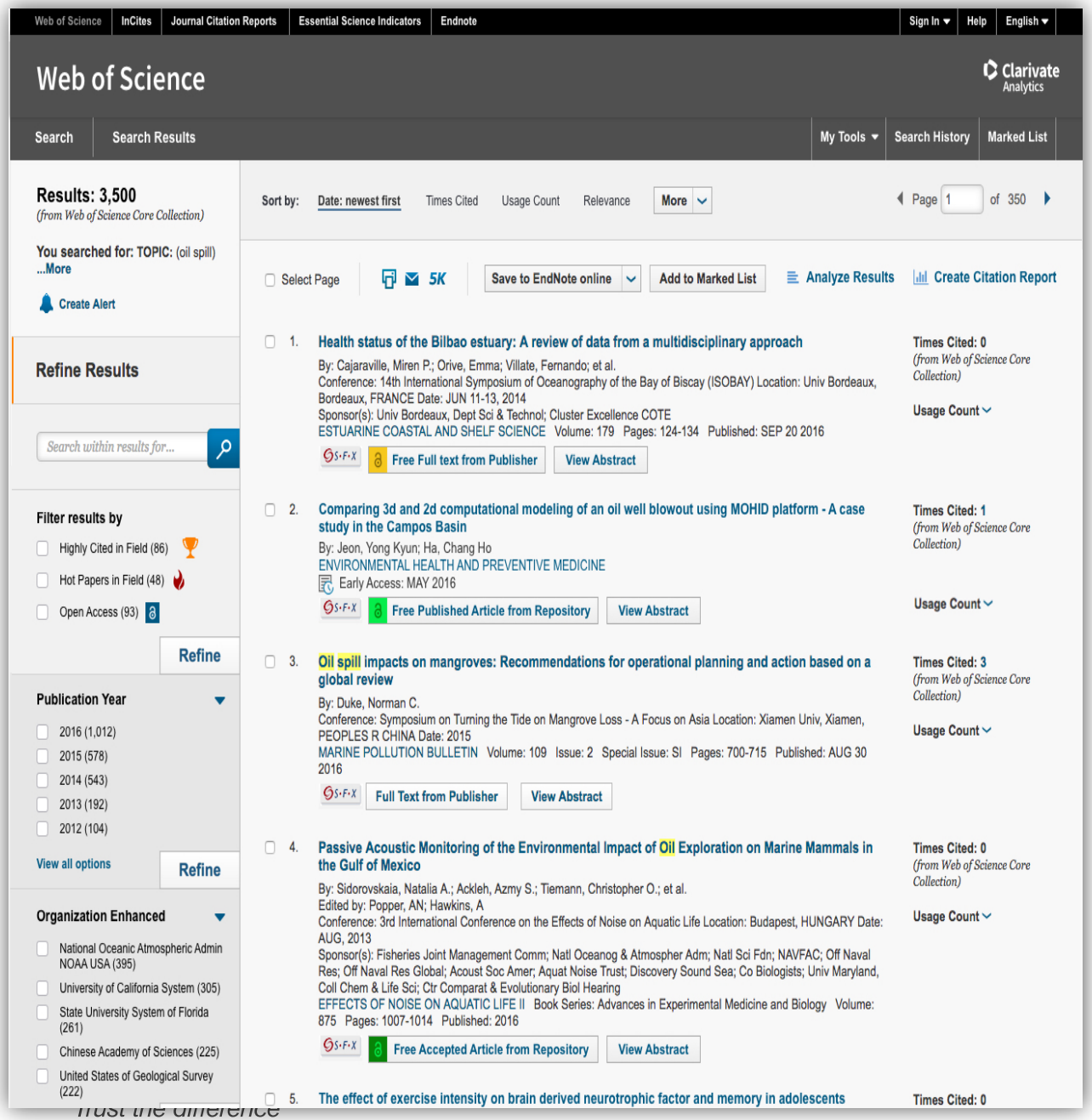
Impactstory

oadoi

The grant funds improvements to Impactstory's oaDOI technology. We are using oaDOI to provide **reliable linking to the best available version of OA content.**

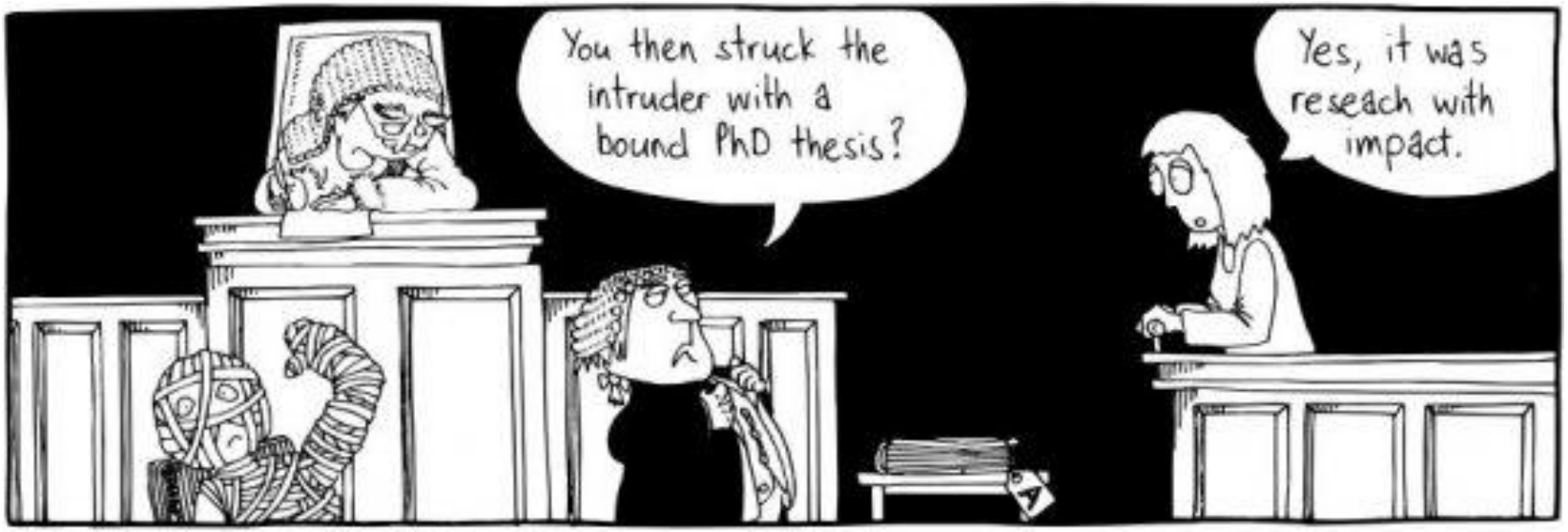
- For Green OA articles, Web of Science will only link to peer-reviewed items from open repositories, NOT "pre-prints." We will identify two types of Green OA articles:
  - Accepted Manuscript
  - Published Version
- For all OA articles, Web of Science will preference links to the publisher's version, when available.

Coming Soon..



Expanded Open Access identification will help you **find legally available Green & Hybrid articles.**

Clarivate Analytics

Trust the difference

THANK YOU! QUESTIONS & DISCUSSION!