

BIBLIOMETRICS AS A RESEARCH MANAGEMENT TOOL WORKSHOP

Marcin Kapczynski,
Strategic Business Manager
Scientific & Scholarly Research

Clarivate
Analytics

Formerly the IP & Science
business of Thomson Reuters

AGENDA

- Origins of bibliometrics
- Overview of main bibliometric indicators; application and limitations
- Data quality considerations
- How and where are these metrics used?
- Live demonstration and Hands-on Practice



— **ORIGIN OF BIBLIOMETRICS**



WHAT IS BIBLIOMETRICS?

Bibliometrics is “the statistical analysis of bibliographic data, mainly in scientific and technical literatures. It measures the amount of scientific activity in a subject category, journal, country, topic or other area of interest.”¹

Citation analysis is a commonly used bibliometric method



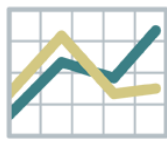
WHY USE BIBLIOMETRICS?



Funding pressures



Efforts at more objective approaches to allocation of credit



Various reporting & demonstration of achievement



Understand ones position in the global Science competition



Elaborating a publication strategy



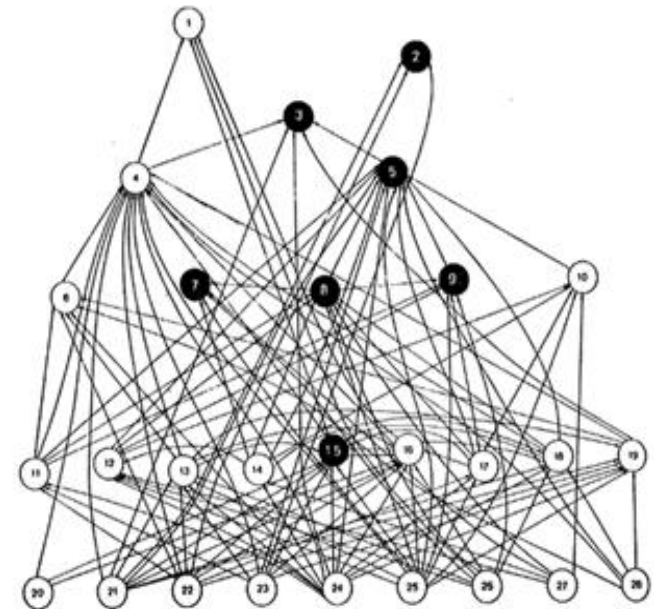
Manage and assess the visibility of an institution

A black and white photograph of a middle-aged man with a mustache and receding hair, smiling warmly. He is wearing a light-colored checkered button-down shirt under a dark jacket. He is seated at a desk, with his hands resting on a stack of papers or books. The background is slightly out of focus, showing what appears to be a window with horizontal blinds.

Science 122 (3159), p.108-11, July 1955

Clarivate
Analytics

Formerly the IP & Science
business of Thomson Reuters



THE WEB OF SCIENCE: BIRTH AND DEVELOPMENT

Journal Citation Reports

Science Citation Index



SciSearch

Custom citation projects & national indicators (mainframe)



Web of Science



Century of Science

InCites

Book Citation Index

1960 | 1965 | 1970 | 1975 | 1980 | 1985 | 1990 | 1995 | 2000 | 2005 | 2010 | 2015



Arts & Humanities Citation Index

Social Sciences Citation Index

Essential Science Indicators

PC-based citation sets and indicators for journals, nations, and institutions

Century of Social Science



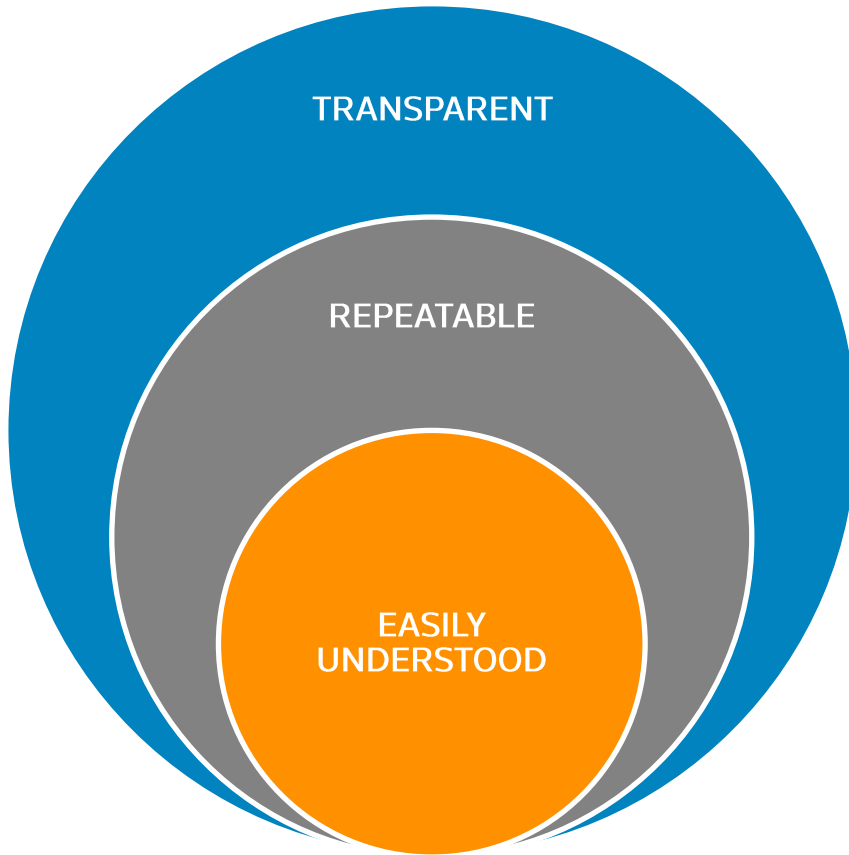
Clarivate Analytics

Formerly the IP & Science business of Thomson Reuters

WHY CITATION METRICS?

Citation metrics are:

- **Transparent**
Uses data on peer reviewed journals and citations received by those articles.
- **Repeatable**
Quantitatively track and measure research outcomes.
- **Meaningful**
Citations are an indicator of an article's impact and usefulness to the worldwide research community; they are the mode by which peers acknowledge each other's research.



CITATION ANALYSIS: STRENGTHS AND WEAKNESSES



- **Objective**, easy and low cost procedure
- **Repeatable**
- It can take into account massive amounts of **data**
- Baselines and thresholds can be created for **true benchmarking**
- Positive **correlation** with peer review (less true for Social Sciences)

- No qualitative differentiation between citations
- Citations measure scientific impact/ utility/ merit, not always quality
- Citations vary across different subject fields and time
- Citation coverage depends on their sources

BEST PRACTICE: INFORMED PEER REVIEW

“The literature does support the idea of supplementing peer review by bibliometrics (**informed** peer review).”¹



“Metrics should **support**, not supplant, expert judgment.”¹

CITATION METRICS ARE **ONE PIECE** OF THE RESEARCH PERFORMANCE PUZZLE.

- They complement other types of assessment.
- Combination of multiple methodologies and data sources
- Examples of data sources / methodologies:
 - # of and value of Grants awarded
 - # of awards (e.g. Nobel Prizes)
 - Peer evaluation
 - **Publication counts**
 - **Citation counts/citation metrics**
- Combination of factors
 - None of these measures works perfectly on its own, there are always anomalies.
 - Human judgment is required to interpret the results.



— **OVERVIEW OF MAIN METRICS**



MAIN BIBLIOMETRIC INDICATORS



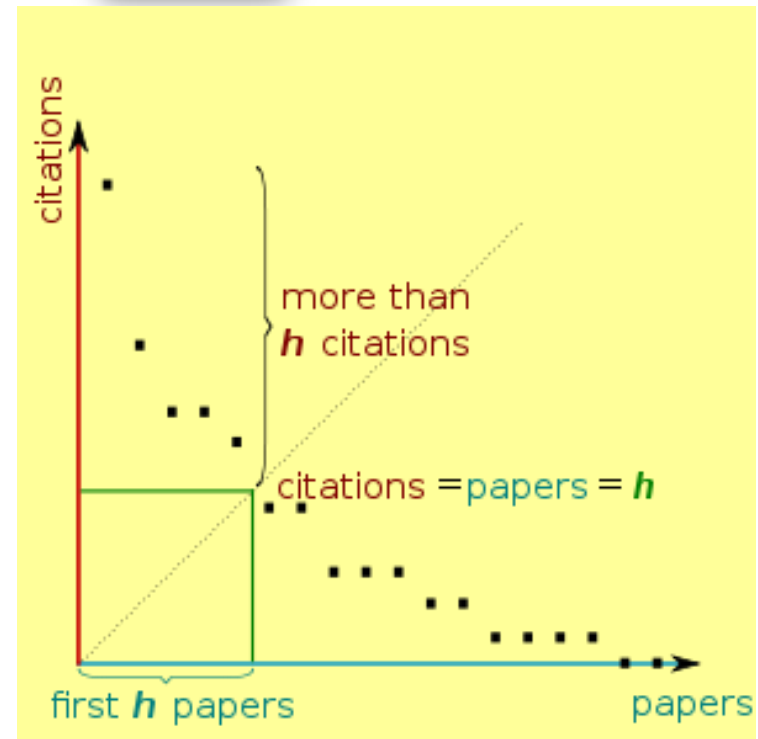
- Productivity (number of papers produced)
- Citation counts (number of citation received)
- **H-Index**
- **Citation average**
- **Journal Impact Factor**
- **Normalized Citation Impact**

THE H-INDEX

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



CITATION IMPACT (AVERAGE CITATIONS)

DR. DOE



10 Documents

60 Total Cites

6 h-index

Average
Citation
per paper:
6

DR. BROWN



10 Documents

60 Total Cites

2 h-index

Average
Citation
per paper:
6

H-index as a Tie breaker

CITATION IMPACT (AVERAGE CITATIONS)

DR. SMITH



7 Documents 273 Total Cites 4 h-index

Average
Citation
per paper:

39

DR. JONES



4 Documents 92 Total Cites 4 h-index

Average
Citation
per paper:

23

DR. ZHANG



10 Documents 40 Total Cites 4 h-index

Average
Citation
per paper:

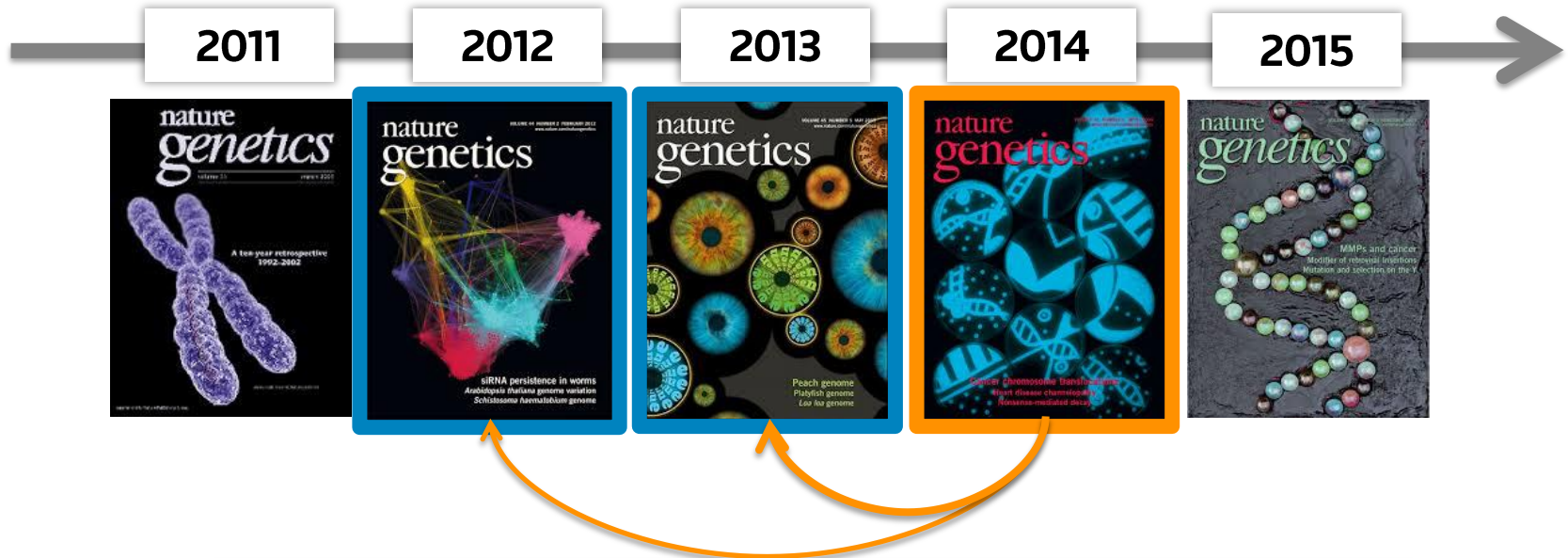
4

A **pillar** of best practices:

Do **NOT** rely on a single indicator, use a combination of them

JOURNAL IMPACT FACTOR REFLECTS A **JOURNAL'S** OVERALL PERFORMANCE

2014 Impact Factor: Ratio of **citations from 2014** to papers published in **2013 and 2012** to **papers published in 2012-2013**



Cites in 2014 to items published in:	2013 =6182	Number of items published in:	2013 =212
	2012 =6645		2012 =225
	Sum: 12827		Sum: 437

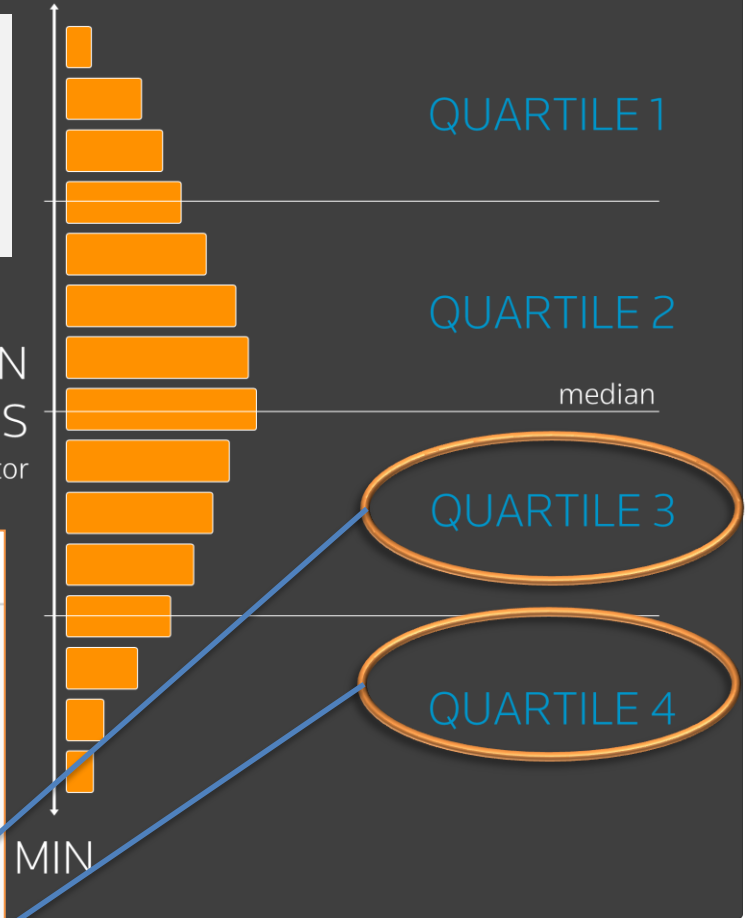
Calculation= $\frac{\text{Cites to recent items}}{\text{Number of recent items}} = \frac{12827}{437} = 29.352$

JOURNAL IMPACT FACTOR REFLECTS A **JOURNAL'S** OVERALL PERFORMANCE

To contextualise the Journal Impact Factor, we calculate normalised metrics, based on percentiles distribution of Impact Factors of journals **in the same category**

DISTRIBUTION
OF VALUES
e.g Impact Factor

MAX



JOURNAL OF CONTINUING EDUCATION IN THE HEALTH PROFESSIONS

Impact Factor

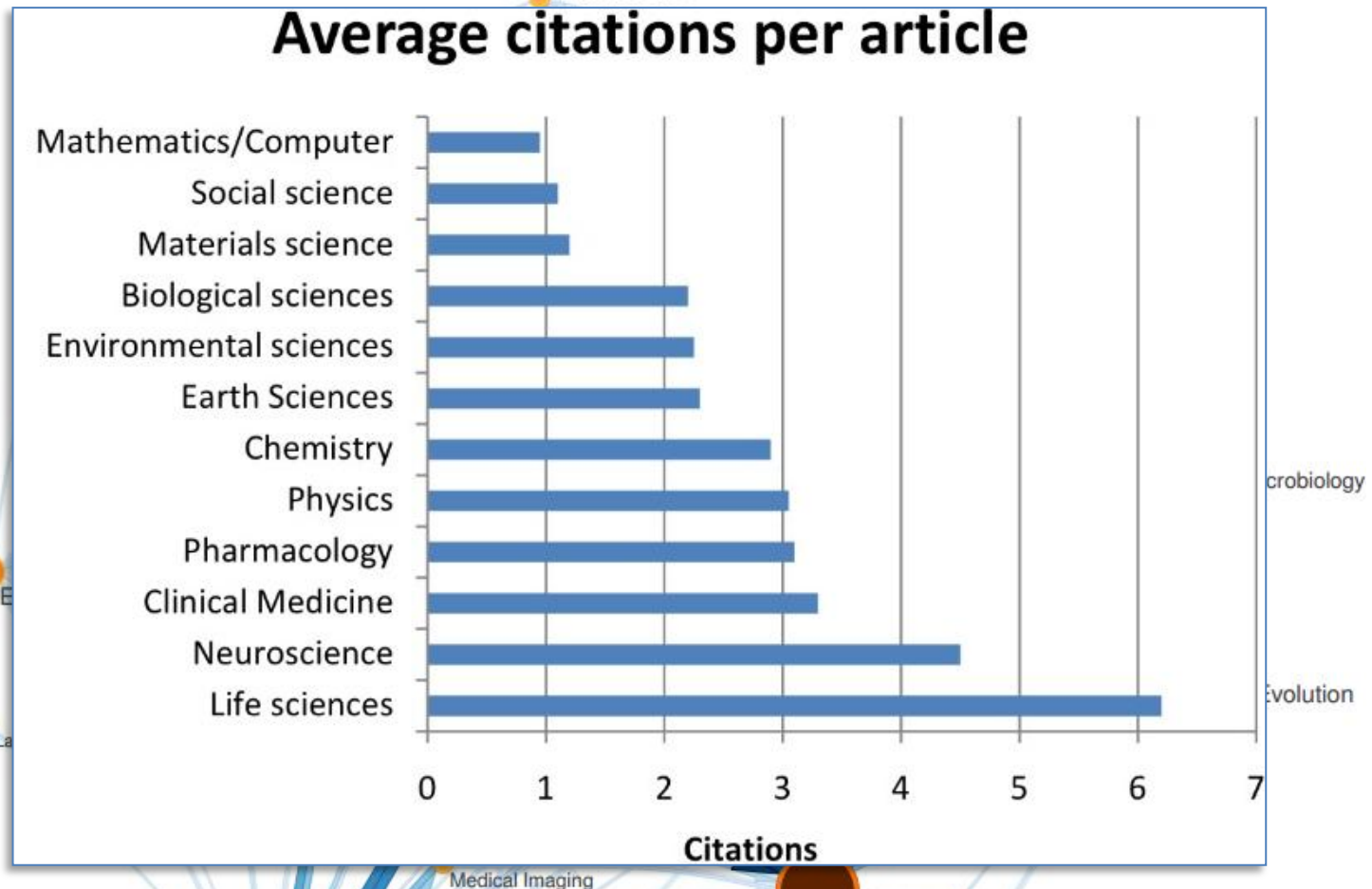
1.112 **1.406**

2015

5 year

JCR® Category	Rank in Category	Quartile in Category
EDUCATION, SCIENTIFIC DISCIPLINES	24 of 40	Q3
HEALTH CARE SCIENCES & SERVICES	70 of 87	Q4

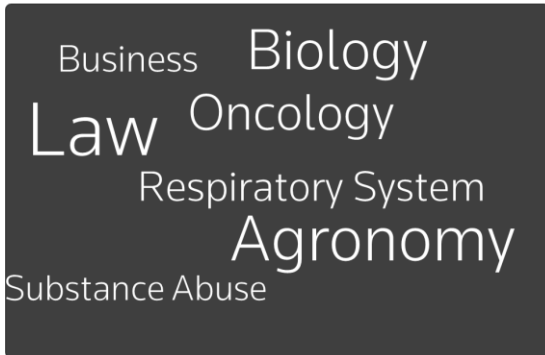
THE INFLUENCE OF CONTEXT ON METRICS



Research communities have radically **different** “Citation behaviours” and patterns.

It is therefore essential to **contextualize** citation metrics

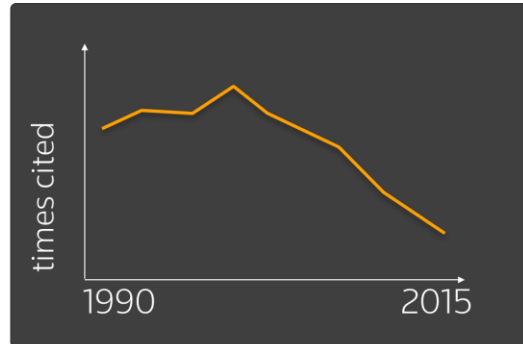
NORMALIZATION



CATEGORY

citation patterns differ by subject category

e.g. nanotechnology vs law



TIME

citations accumulate over time and at different rates depending on article age and category

e.g. new articles may accumulate citations quickly, older ones more slowly or not at all



DOCUMENT TYPE

citations differ by document type within a journal

e.g. reviews are generally more heavily cited than articles, and editorials, book reviews etc. may go uncited

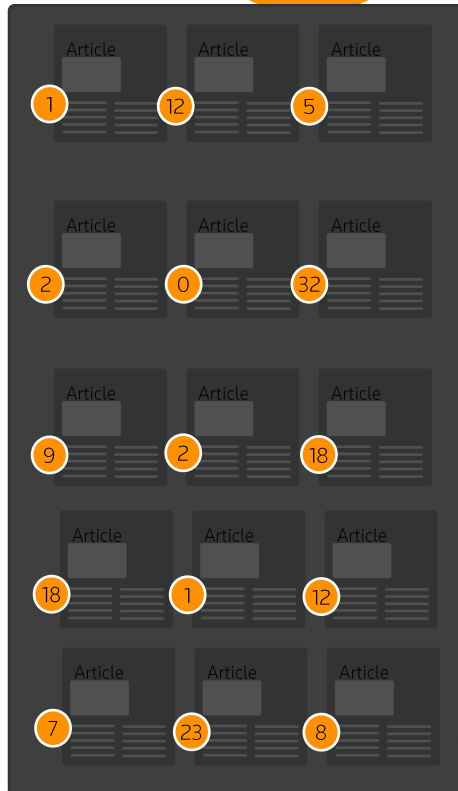
NORMALIZATION PUTS DATA INTO CONTEXT— IS AN ENTITY DOING BETTER OR WORSE THAN WOULD BE EXPECTED IN A CATEGORY?

NORMALIZATION

IS

20

"good" or "bad?"



EXPECTED NUMBER OF CITATIONS FOR ARTICLE SET*

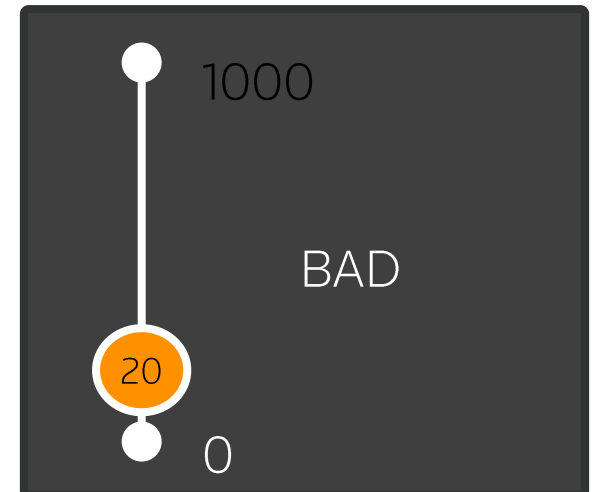
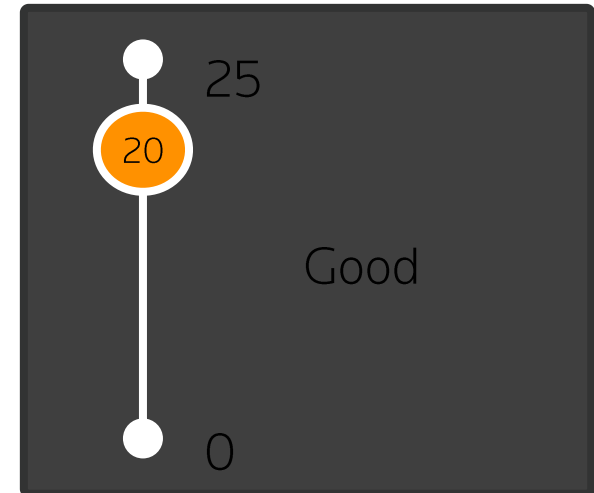
AVERAGE CALCULATED
FROM SUM OF ALL
DOCUMENTS IN THE
SELECTED GROUP

150 Total Cites in group

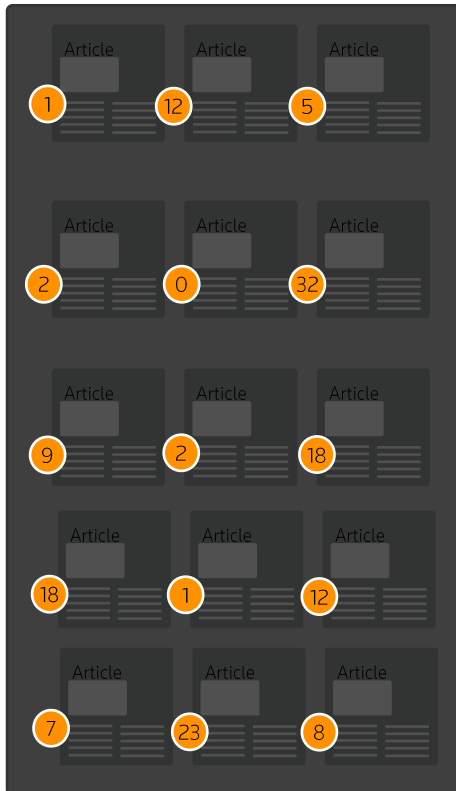
15 DOCUMENTS IN
GROUP

10 Expected
cites

*Based upon category or journal,
publication year, and document type



NORMALIZATION



EXPECTED NUMBER OF CITATIONS FOR ARTICLE SET*

AVERAGE CALCULATED
FROM SUM OF ALL
DOCUMENTS IN THE
SELECTED GROUP

150 Total Cites in group

15 DOCUMENTS IN
GROUP

10 Expected
cites

*Based upon category or journal,
publication year, and document type

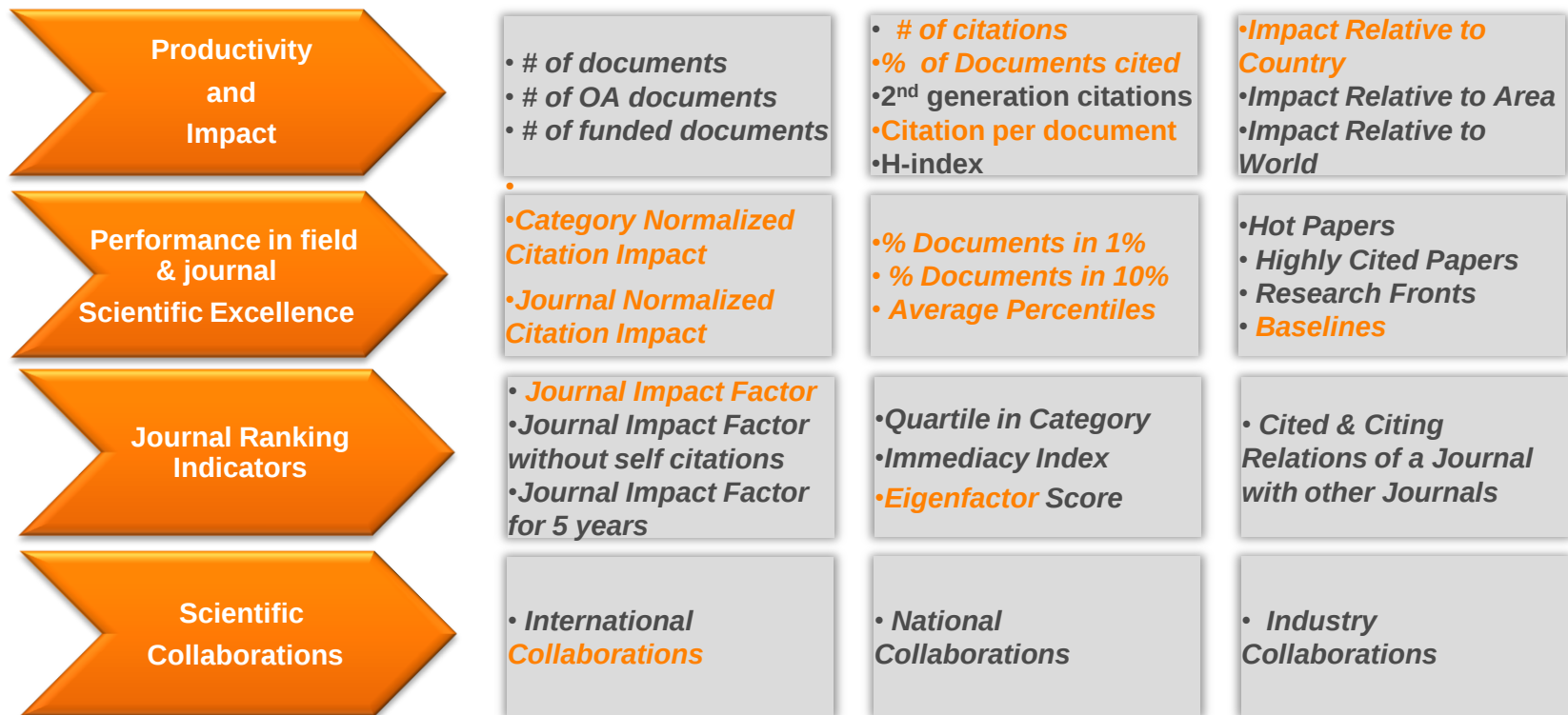
NORMALIZED RATIO OF ACTUAL / EXPECTED CITATIONS



** Also applies to ORGANIZATIONS, COUNTRIES, RESEARCH
AREAS, AND ANY OTHER CLUSTER OF PAPERS

BIBLIOMETRIC INDICATORS

- What *can* and *what* should be measured?
- What are appropriate measures for the purpose?
- Each indicator answers 1 question, not ALL questions.

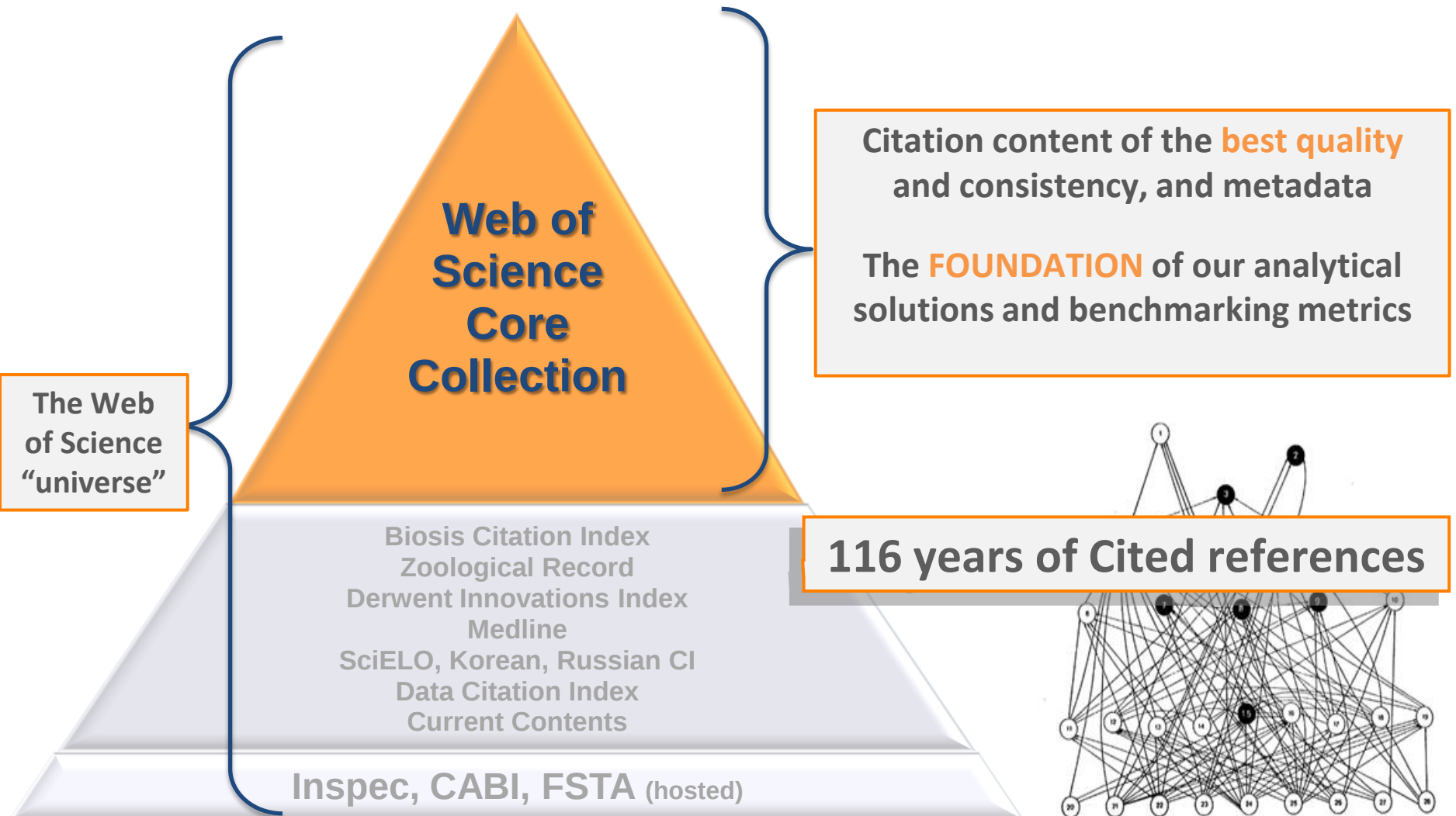


— DATA QUALITY CONSIDERATIONS

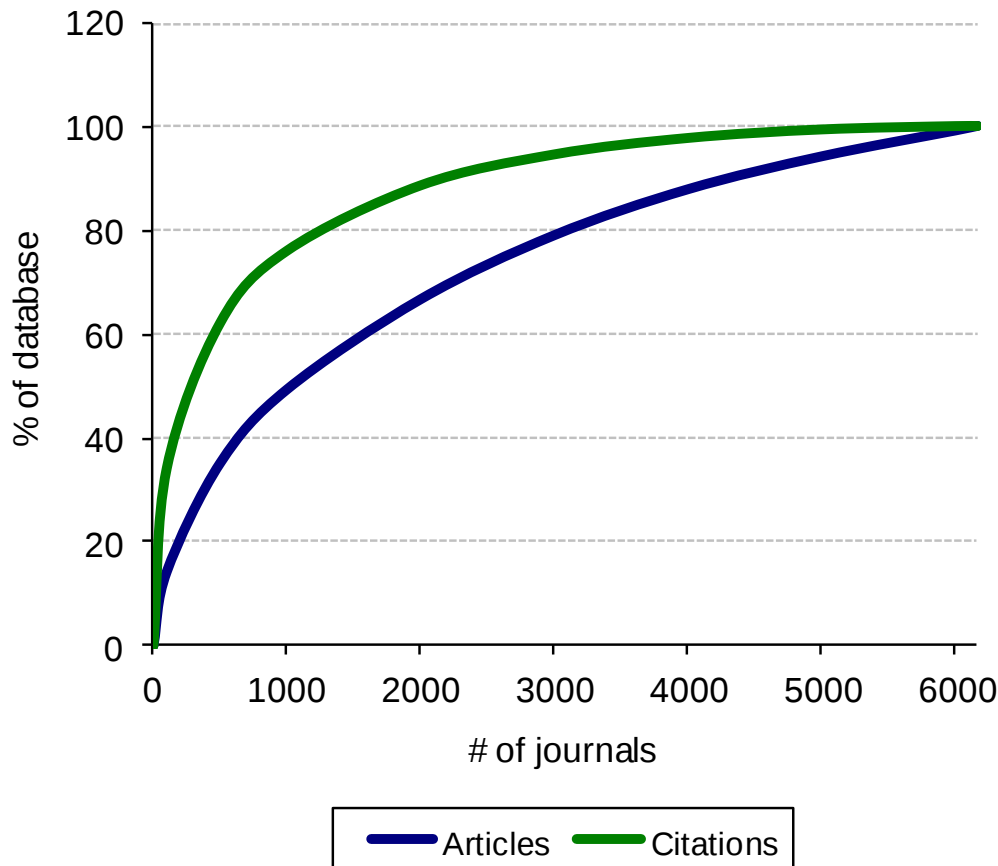
**CITATION METRICS ARE ONLY AS
GOOD AS THEIR SOURCE**

- The **necessary quality** of the foundational data behind the metrics

CITATION METRICS ARE ONLY AS GOOD AS THEIR SOURCE



SELECTION: WHERE IS THE RELEVANT CONTENT?



40% of the journals represent:

- 80% of the publications
- 92% of cited papers

4% of the journals represent:

- 30% of the publications
- 51% of cited papers

Also: Who selects the data?
(Garfield's idea was to keep this process neutral)

THE WEB OF SCIENCE CORE COLLECTION DEVELOPMENT

JOURNAL SELECTION PROCESS

- The philosophy of the database is to offer content of the **highest quality**
- Selection is based on quality criteria established and developed for **over 50 years**
- Today, 18 **FULL TIME** editors **fully dedicated** are in charge of content selection and maintaining the quality and coherence of the database (**150 Years of experience**)
- **12% ACCEPTANCE RATE** in 2014

INFORMATION PROVIDER, NOT PRIMARY PUBLISHER

- Metrics need to reflect reality, and metrics are as objective as their data source
- None of our editors are involved in journal edition or research publishing
- **NEUTRALITY + OBJECTIVITY = SELECTION OF THE HIGHEST QUALITY**

THE WEB OF SCIENCE CORE COLLECTION



63M publications: 12,760 Journals, 12,000+ annual conferences (8.5 M records), 70,645 Books

Consistency and reliability: ideal for research evaluation, e.g. field averages

Depth: 100+ years, including cited references

Unmatched expertise: 50+ years of citation analysis and research evaluation

High quality metadata: Funding acknowledgments from 2008

1 billion cited references, as far back as 1898

PUBLICATION METADATA

The IMGT/HLA database

By: Robinson, J (Robinson, James)^[1]; Halliwell, JA (Halliwell, Jason A.)^[1]; McWilliam, H (McWilliam, Hamish)^[2]; Lopez, R (Lopez, Rodrigo)^[2]; Parham, P (Parham, Peter)^[3]; Marsh, SGE (Marsh, Steven G. E.)^[1,4]

NUCLEIC ACIDS RESEARCH

Volume: 41

DOI: 10.1093/

Published: JA

[View Journal](#)

Abstract

It is 14 years s

The HLA comp

encode proteins of the im

Author Information

Reprint Address: Marsh, SGE (reprint author)

+ Royal Free Hosp, A

Addresses:

+ [1] Royal Free Hos

Categories / Classification

Research Areas: Biochemi

Web of Science Categor

Document Information

Document Type: Article

Language: English

Conexio

Abbott Molecular Laboratories Inc.

The more data points you can use,
the more accurate and flexible your
analysis can be

Journals are indexed Cover-to-Cover

All authors are indexed

All affiliations are indexed

All funding sources are indexed (2008)

quences.

genes

ol is the responsibility

Citation network

204 Times Cited

ences

[View Related Records](#)

ce Core Collection

192 in BIOSIS Citation Index
4 in Chinese Science Citation
Database

BIBLIOMETRIC CHALLENGE: DOES THE DATA REQUIRE EDITING TO REMOVE ARTIFACTS?

• INSTITUTION NAMES

ST ISTVAN UNIV	CENT EUROPE UNIV	ACAD SCI
ST ISTVAN UNIV GODOLLO	CENT EUROPEAN UNIV	ACAD SCI BUDAPEST
SVENT ISTVAN UNIV	CENT EUROPEAN UNIV BUDAPEST	ACAD SCI HONGRIE
SYENT ISTVAN UNIV	CENT EUROPEAN UNIV BUSINESS SCH	ACAD SCI HUNGARIAN PEOPLES REPUB
SZEBT ISTVAN UNIV	CENT EUROPEN UNIV	ACAD SCI HUNGARY
SZECHENYI ISTVAN UNIV	CENTRAL EUROPEAN UNIV	
SZECHENYI ISTVAN UNIV APPL SCI	CEU	
SZEINT ISTVAN UNIV	UNIV CENT EUROPE	
SZENT ISTAVAN UNIV		
SZENT ISTVAIN UNIV		
SZENT ISTVAN EGYET		
SZENT ISTVAN EGYETEM		
SZENT ISTVAN UNIV		

• AUTHOR NAMES

- Is ROTHWELL N different from ROTHWELL NJ?
- LEE S returns **14,022** publications; SUZUKI T (**21,462**)

DOES THE DATA REQUIRE EDITING TO REMOVE ARTEFACTS?

Review of particle physics

By: Amsler, C (Amsler, C.)^[21]; Doser, M (Doser, M.)^[3]; Antonelli, M (Antonelli, M.)^[4]; Asner, DA (Asner, D. M.)^[5]; Babu, KS (Babu, K. S.)^[6]; Baer, H (Baer, H.)^[7]; Band, HR (Band, H. R.)^[8]; Barnett, RM (Barnett, R. M.)^[9]; Bergren, E (Bergren, E.); Beringer, J (Beringer, J.)^[9]; Bernardi, G (Bernardi, G.)^[10,11,12]; Bertl, V (Bertl, W.)^[13]; Bichsel, H (Bichsel, H.)^[14]; Biebel, O (Biebel, O.)^[15]; Bloch, P (Bloch, P.)^[3]; Blucher, E (Blucher, E.)^[16]; Blusk, S (Blusk, S.)^[17]; Cahn, RN (Cahn, R. N.)^[9]; Carena, M (Carena, M.)^[16,18,19]; Caso, C (Caso, C.)^[1,20]; Ceccucci, A (Ceccucci, A.)^[3]; Chakraborty, D (Chakraborty, D.)^[21]; Chen, MC (Chen, M. C.)^[22]; Chivukula, R (Chivukula, R.)^[23]; Cowan, C (Cowan, C.)^[24]; Dahl, O (Dahl, O.)^[9]; D'Ambrosio, G (D'Ambrosio, G.)^[1]; Dobrescu, B (Dobrescu, B.)^[18]; Erler, J (Erler, J.)^[32]; Ezhela, B (Ezhela, B.)^[36]; Gaiser, TK (Gaiser, TK.)^[36]; Gherghetta, T (Gherghetta, T.)^[39]; Giudice, V (Giudice, V.)^[42,43]; Groom, D (Groom, D.)^[42,43]

[82] Univ Alabama, Dept Phys & Astron, Tuscaloosa, AL 35487 USA

[83] Univ Pisa, INFN, I-56127 Pisa, Italy

[84] Univ Pisa, Dipartimento Fis, I-56127 Pisa, Italy

[85] Univ Gottingen, Inst Phys 2, D-37077 Gottingen, Germany

Organization-Enhanced Name(s)
University of Gottingen

[86] Ohio State Univ, Dept Phys, Columbus, OH 43210 USA

[87] Max Planck Inst Phys & Astrophys, Werner Heisenberg Inst, D-80805 Mu

Organization-Enhanced Name(s)
Max Planck Society

[88] Johannes Gutenberg Univ Mainz, Inst Phys, D-55099 Mainz, Germany

Organization-Enhanced Name(s)
Johannes Gutenberg University of Mainz

[89] Tufts Univ, Medford, MA 02155 USA

Citation Network

4,214 Times Cited

4,762 Cited References

[View Related Records](#)

[View Citation Map](#)

[Create Citation Alert](#)

(data from Web of Science™ Core Collection)

Author	ResearcherID	ORCID Number
Sarkar, Subir	G-5978-2011	http://orcid.org/0000-0002-3542-858X
Patrignani, Claudia	C-5223-2009	http://orcid.org/0000-0002-5882-1747
Stahl, Achim	E-8846-2011	http://orcid.org/0000-0002-8369-7506
Chivukula, R. Sekhar	C-3367-2012	http://orcid.org/0000-0002-4142-1077
Punzi, Giovanni	J-4947-2012	http://orcid.org/0000-0002-8346-9052
Murayama, Hitoshi	A-4286-2011	

6000+ institutions manually unified today.
CNRS alone uses 18K+ rules

Direct Synchronisation with ResearcherID
Regular downloads of ORCID data

[93] Ecole Polytech Fed Lausanne, CH-1015 Lausanne, Switzerland		
[94] Univ British Columbia, Dept Phys & Astron, Vancouver, BC V6T 1Z1, Can		http://orcid.org/0000-0001-9250-246X
Woody, CL (Woody, C. L.) ^[65,111] ; Workman, RL (Workman, R. L.); Yamamoto, A (Yamamoto, A.) ^[48]		http://orcid.org/0000-0002-7630-8605

ORGANISATION/SEGMENTATION OF DATA

▼ Research Area Schemas

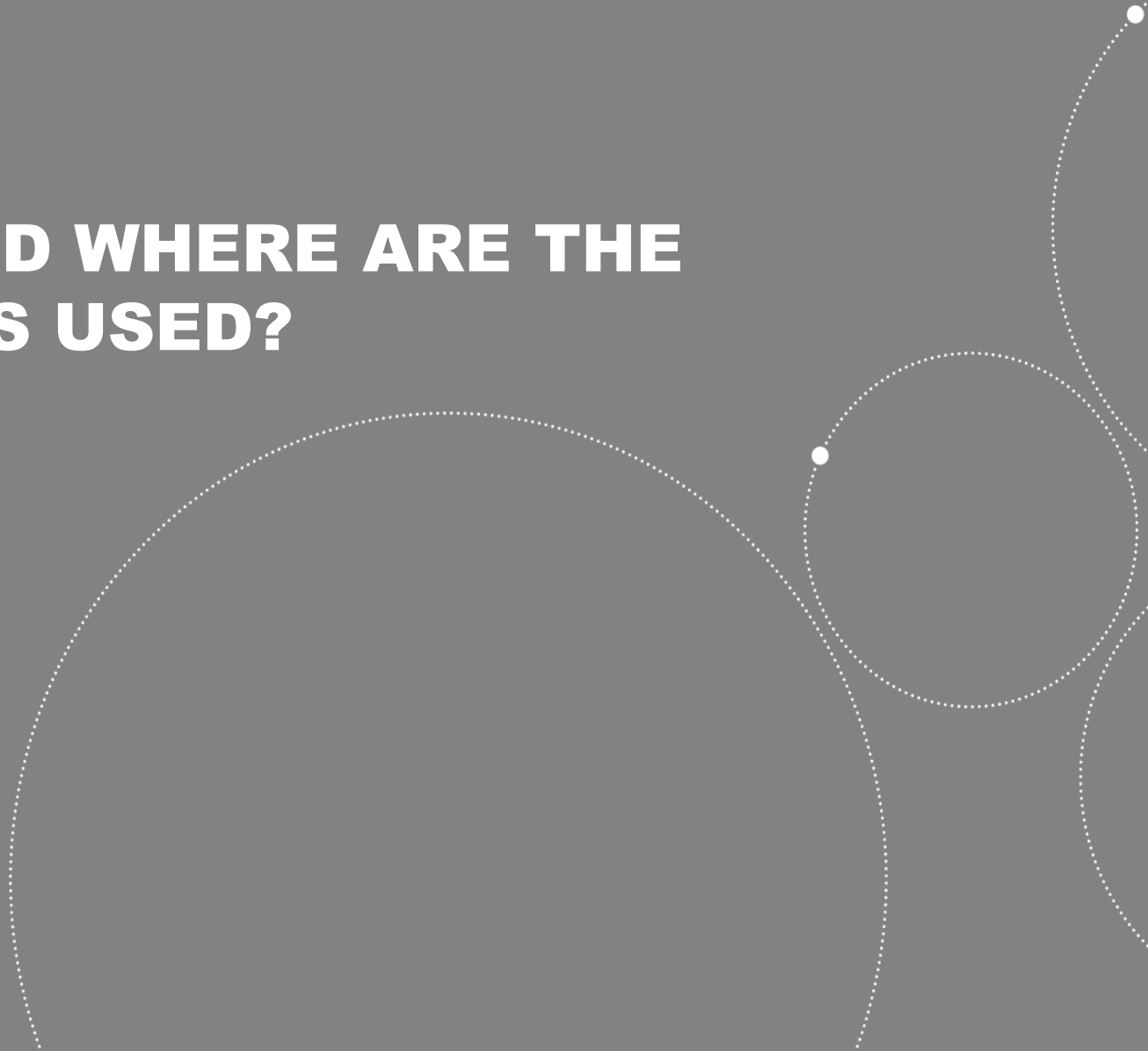
- ▶ ANVUR Category Scheme
- ▶ Australia ERA 2012 FOR
- ▶ China SCADC Subject Categories
- ▶ Essential Science Indicators
- ▶ FAPESP (Brazil)
- ▶ GIPP
- ▶ OECD Category Scheme
- ▶ UK 2008 RAE Units of Assessment
- ▶ UK 2014 REF Units of Assessment
- ▶ Web of Science Research Areas
- ▶ KAKEN Category Scheme

Incites uses **790** categories
in **14 different schema** to calculate
expected citation rates (baselines)

38 document types
x 35 years of data
x 790 categories
= 1 million baselines
calculated every two months for
better contextualization

**The more data points you can use, the more
accurate and flexible your analysis can be**

— **HOW AND WHERE ARE THE
METRICS USED?**



THE LEIDEN MANIFESTO FOR RESEARCH METRICS



1. Quantitative evaluation should support qualitative assessment by experts
2. Measure performance against the research missions of the institution, group or researcher.
3. Protect excellence in locally relevant research.
4. Keep data collection and analytical processes open, transparent and simple.
5. Allow those evaluated to verify data and analysis.
6. Account for variation by field in publication and citation practices.
7. Base assessment of individual researchers on a qualitative judgment of their portfolio.
8. Avoid misplaced concreteness and false precision.
9. Recognize the systemic effects of assessment and indicators.
10. Scrutinize indicators regularly and update them.

“READY MADE” ANALYTICS PLATFORMS (E.G. INCITES)

InCites includes:

—*Journal Citation Reports*—

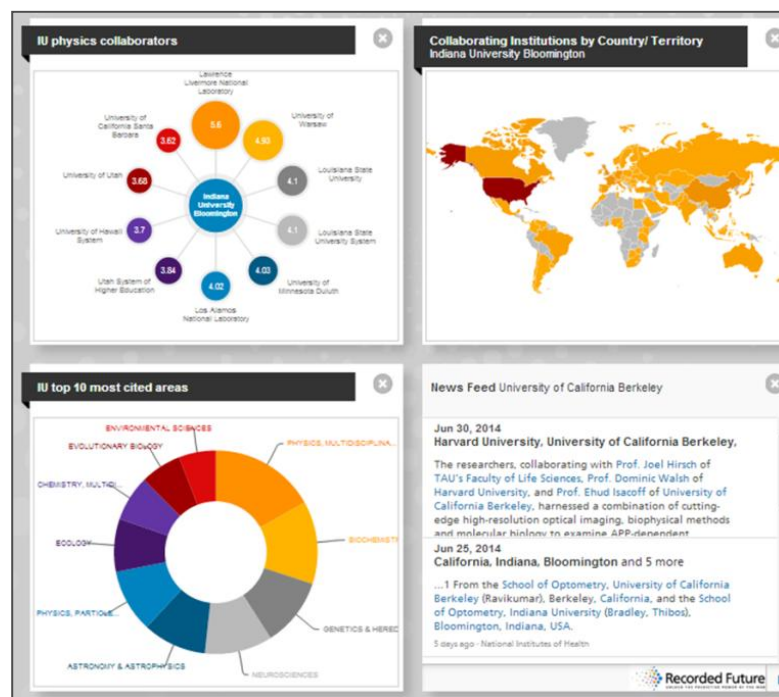
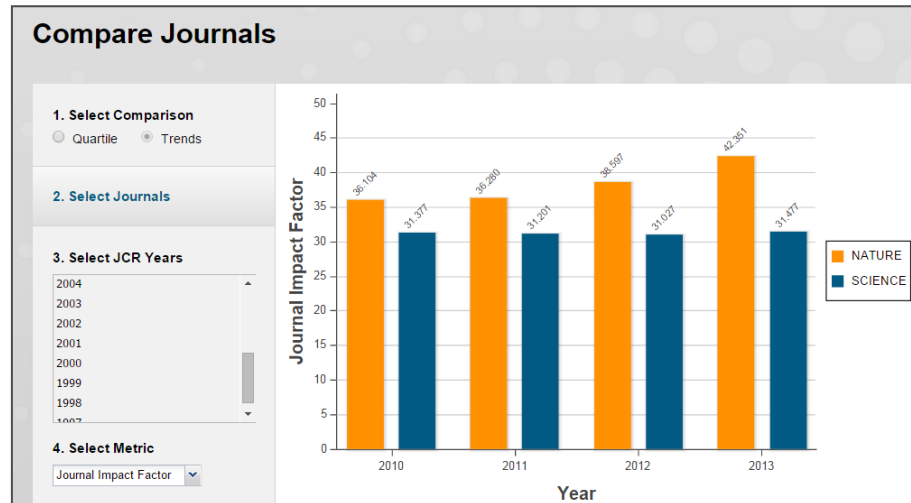
comprehensive journal analysis

—*Essential Science Indicators*—

top cited research from around the world

—*Benchmarking and Analytics (B&A)*—

flexible benchmarking of institutions, journals
authors, funders and collaboration analysis across
35 years of publication activities



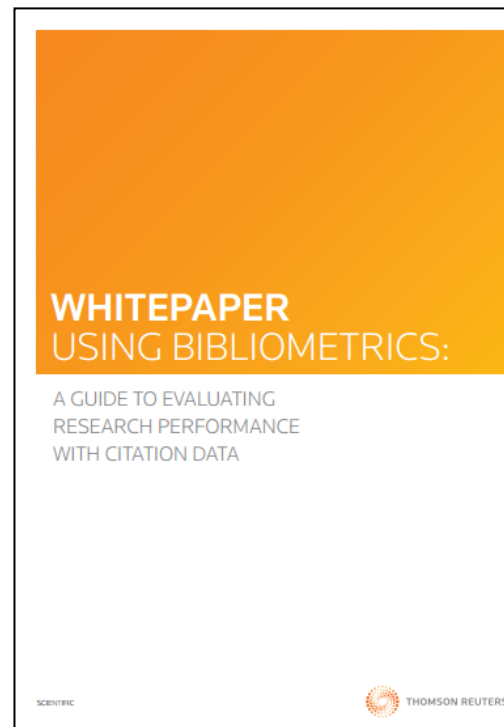
— **LIVE DEMO**



ADDITIONAL REFERENCE MATERIAL



[Read the InCites Indicators Handbook](#)



[Read the bibliometrics whitepaper](#)