isi | Institute for Scientific Information

Global Research Report

# Data categorization: understanding choices and outcomes

Martin Szomszor, Jonathan Adams,
David A. Pendlebury and Gordon Rogers

Clarivate™

# Author biographies

**Dr Martin Szomszor** is Director at the Institute for Scientific Information (ISI). He has a background in Computer Science, with expertise in knowledge engineering, machine learning, and natural language processing. He was named a 2015 top-50 UK Information Age data leader for his work in creating the REF2014 impact case studies database for the Higher Education Funding Council for England (HEFCE). ORCiD: https://orcid.org/0000-0003-0347-3527.

**Jonathan Adams** is Chief Scientist at the Institute for Scientific Information. He is also a Visiting Professor at King's College London, Policy Institute. In 2017 he was awarded an Honorary D.Sc. by the University of Exeter, for his work in higher education and research policy. ORCiD: https://orcid.org/0000-0002-0325-4431.

**David Pendlebury** is Head of Research Analysis at ISI. Since 1983 he has used Web of Science data to study the structure and dynamics of research. He worked for many years with ISI founder Eugene Garfield. With Henry Small, David developed the Web of Science Essential Science Indicators™. ORCiD: https://orcid.org/0000-0001-5074-1593.

**Gordon Rogers** is a Senior Data Scientist at the Institute for Scientific Information. He has worked in the fields of bibliometrics and data analysis for the past 10 years, supporting clients around the world in evaluating their research portfolio and strategy. ORCiD: https://orcid.org/0000-0002-9971-2731.

# Foundational past, visionary future

## About the Institute for Scientific Information

The Institute for Scientific Information™ at Clarivate has pioneered the organization of the world's research information for more than half a century. Today it remains committed to promoting integrity in research whilst improving the retrieval, interpretation and utility of scientific information. It maintains the knowledge corpus upon which the Web of Science™ index and related information and analytical content and services are built. It disseminates that knowledge externally through events, conferences and publications whilst conducting primary research to sustain, extend and improve the knowledge base. For more information, please visit www.clarivate.com/webofsciencegroup/solutions/isi-institute-for-scientific-information/.

ISBN 978-1-8382799-1-2

Cover image: Apples and oranges, Kate Andrews

# Introduction

This Global Research Report is about the way we recognize natural divisions of knowledge and research, or, more specifically, the way we categorize publications for discovery, analysis, management and policy purposes. We describe a history of classification that has generally been 'top down', the characteristics of exemplar systems from around the world that feature in Clarivate products, and the introduction of new bottom-up approaches that draw on research data itself. Then we review the analytical consequences of applying them to national and institutional data and draw attention to the effect of different classification schemes on document counts and citation impact calculations.

This report is not only descriptive. It also promotes the need for good practice in data management as part of the responsible use of research metrics. Choosing a categorical system for research data is not a value-free decision: "Words are the bugles of social change," wrote Charles Handy in *The Age of Unreason* (Handy, 1989, p. 17). "When our language changes, behavior will not be far behind." The labeling of topics and disciplines reflects this in research. Being aware of the characteristics and limitations of the ways we categorize research publications is important to research management because it influences the way we think about established and innovative research topics, the way we analyze research activity and performance, and even the way we set up organizations to do research. There are many ways of organizing this information, most of which are sensibly and coherently related but may have developed for specific purposes. Choosing a categorization scheme by happenstance rather than informed choice, or for a purpose other than that for which it was intended, can lead to equally uninformed outcomes.

**The chief organizing unit of research is the specialty, an "invisible college" that represents "some type of natural order in science … Our method of indexing papers by descriptors or other terms is almost certainly at variance with this natural order … If we can successfully define the natural order, we will have created a sort of giant atlas of the corpus of scientific papers that can be maintained in real time for classifying and monitoring developments as they occur."**

**– Derek J. de Solla Price (1980)**

# Categorizing research

There is no universal template onto which knowledge can be pasted, as historian of science Derek J. de Solla Price recognized. Knowledge is a continuous spectrum but it has long been found convenient, and increasingly necessary, to organize information by inferring boundaries between things. Plato, in the 4th century BCE, described a division of knowledge into arithmetic, geometry, music and astronomy.

Ethnobiologists suggest that the human inclination to categorize nature is innate. There seems to be a universal rule that identifies non-human animals, first in the categories of bird/fish/snake and then by progressively adding worm-bug/mammal (Atran, 1990).

As the knowledge base grew, the message that humans need to organize and prioritize information about the world around them was reinforced. It may have been possible for 16th century natural philosophers to know most of what was being studied but specialisms appeared and divisions grew. The 19th century saw the innovation of formal, relational library indexing, because it was no longer possible to file and recover the proliferating mass of documents if they were shelved without some ordering.

## The 19th century: Library of Congress and Dewey

The Dewey Decimal Classification (1876) was one version and the Library of Congress Classification (1897) was another. Paul Otlet and S. R. Ranganathan advanced other systems, Universal Decimal Classification (1904) and Colon Classification (1933) respectively, and there are many others, including that of E. Wyndham Hulme (see sidebar). The Dewey and Library of Congress classification systems established the idea of hierarchical information structures. The Library of Congress' Class Q refers to Science and within that Subclass QK refers to Botany. Dewey follows a similar system but numerically rather than alphabetically. Some classes became anachronistic: knowledge classification systems are a product of their time! The popularity of these classifications stemmed from the ability they gave to users rapidly to store and retrieve both existing and new documents based on their primary contents.

## The 20th century: the Science Citation Index

The growth of research knowledge continued. Journals proliferate, established fields evolve and fission, new fields appear, and the spread became too great to be readily appreciated by an individual researcher. The need for information about information emerged.

In the 1950s Eugene Garfield, founder of the Institute for Scientific Information (ISI), recognized that the ability to work at the leading edge depended on researchers' current awareness of discovery in their fields. Traditional, labor-intensive indexing fell far behind the growth of the scientific literature after World War II. Existing indexes, such as *Chemical Abstracts*, offered information years out-of-date. Garfield decided to collate contents pages of the latest issues of important journals and publish these as *Current Contents*, a weekly discipline-specific bulletin that allowed researchers to review recently published journals in their areas of interest.

But Garfield also saw defects in discipline-specific classification, especially as handled by traditional indexing approaches using controlled vocabularies and subject headings. His advocacy of citation indexing (Garfield, 1955) was an attempt at, as he said, "breaking the subject index barrier" (Garfield, 1957). When he produced the first commercially available Science Citation Index (SCI)™ in 1964, he showed that related papers could be identified through citations – a navigation method embedding the expert knowledge that authors added to papers in the form of cited references (Garfield, 1964).

The Science Citation Index did not dispense with classification. Each journal was assigned to one or more field categories, to aid information retrieval where a user searched by field classification. As a unified index of science, spanning all domains, Garfield began by identifying the most influential titles in each field and then extended the corpus through consultation with experts and analysis of the journals cited most in each field (Pudovkin & Garfield, 2002).

The SCI was not designed to be encyclopedic but selective, covering the (initially about 600) internationally influential journals then published.

The journals indexed in the SCI became digitized in an accessible format as the Web of Science™ (1997), classified into disciplines, fields and subfields (Web of Science journal categories). More than 20,000 journals are now indexed in the Web of Science, currently grouped into 254 categories. Every item in each journal in a category is indexed and reported, cover-to-cover and coverage is constantly reviewed as knowledge flows and evolves. Clarivate also uses the Essential Science Indicators (ESI)™ database which has 21 broad subject categories, again journal-defined, plus a multidisciplinary category; it does not cover the arts or humanities. The difference between these systems is not just their coverage and granularity but also that the Web of Science categories are inclusive, so a journal may appear in more than one category, whereas ESI categories are exclusive. The Web of Science also has 'Research Areas' which are intermediate between these other two but exclusive, like ESI.

**E. Wyndham Hulme**
Librarian of the Patent Office of Great Britain 1894 to 1919

E. Wyndham Hulme proposed a system for the classification of books according to both content and publication attributes (Dousa, 2017). He deserves attention not only for his theory of classification but also for suggesting that counting works in each category through time could reveal trends in knowledge accumulation across fields and in emerging specialty areas. "Book classification is shelf classification, and shelf classification carried to its furthest limits leads necessarily to uniformity in the extension and definition of its classes," he wrote. "Add to this a chronological order of books in their classes and your scheme of classification acquires a new value: for it presents for each period a bibliographical counterpart of the corresponding growth of the activities of the human mind" (Hulme, 1923). This quantitative approach anticipated scientometric studies by half a century and distinguishes Hulme as a pioneer and visionary.

## Research management and assessment

Since the 1990s, research management has become more active at national and institutional level and new systems for categorizing research activity have emerged across the globe. As signals of change, universities reordered their departmental structures: for example, departments of Botany and Zoology disappeared, Genetics and Biochemistry emerged, and these in turn were subsumed into Schools of Biological Sciences. Business schools spawned novel and complex departmental portfolios. Medical schools lost ancient specialties and gained new technologies such as Nuclear Medicine.

The Organisation for Economic Co-operation and Development (OECD), taking a global perspective, developed the Frascati classification (1963, revised 2007). This contrasts with university structures because OECD categories focus on function rather than content. Research is undertaken for different reasons and Frascati recognizes: *basic*, which aims to acquire new knowledge not directed toward any particular use; *applied*, which is investigation to acquire knowledge directed to a specific practical objective; or *experimental development*, which is a systematic based on existing knowledge. There are six OECD subject-based Fields of Science (FoS) to which activity under the three research types may be assigned: Natural sciences; Engineering and technology; Medical and health sciences; Agricultural sciences; Social sciences; and Humanities. The FoS classification is one of several available for users of InCites™ – a Clarivate platform for research performance analysis.

The OECD FoS look like conventional academic faculties, but they are not. Agriculture signals that this is in fact economic categorisation. The OECD is interested in measuring expenditure, people, activity and outcomes devoted to or created by research and development performed in particular industrial sectors. So, whereas academia recognizes Microbiology as a category in its own right, the OECD is interested in whether it is Microbial physiology (Natural science), Microbial disease (Medicine and health) or Microbial soil ecology (Agriculture). There is a similar matching issue between Chemistry as a science department and Chemical industries: industry uses engineering, mathematics, and so on. Such functional distinctions are entirely reasonable but must be properly understood by users.

# National categorical systems

There are many other well-established categorical classifications in addition to the OECD, the Web of Science and ESI. The examples described below are included in InCites, so we note how they map to the Web of Science.

These systems were developed with diverse objectives in mind, though usually related to the general principle that they should make sense to both government, as the research funder, and researcher, who delivers the outcome, in order to organize a mutually equitable assessment process. Each is the product of significant policy work, review with stakeholders, pilot projects and experience. None are casual creations, all have consensus support from users, all are seen to function well in delivery, and yet they do not all do the same thing. Different classifications may produce different outcomes, as we show later, but none are right or wrong: they were created for different objectives.

## United Kingdom – research assessment

The first Research Selectivity Exercise was in 1986. Universities submitted portfolios of their research activity, which were assessed by peer review via 150 discipline-based panels, a number that was reduced to 72 Units of Assessment (UOAs) for the 1992 Research Assessment Exercise (RAE), to 67 UOAs for RAE 2008, and to 36 UOAs under four main panels for the 2014 Research Excellence Framework (REF). The RAE/REF UOA structure mimics university departments. It was established with assessment in mind; universities' need to make comparable submissions to panels; and assessment grades becoming weighting factors in a funding formula that would inform the institutions. The progressive reduction in category count, from 150 to 36 over 30 years, shows that unduly granular systems may be unhelpful: categories may overlap in content and journal use when outputs submitted in 2014 (36 UOAs) are analyzed but panels can become highly self-referential in a fine-grained UOA structure.

The UOAs are mapped to the Web of Science database by reference to the frequency with which articles from particular journals have been submitted by academics for assessment. This builds up an inclusive core coverage for each UOA and a small but minimal overlap at the interdisciplinary margins.

## Australia-New Zealand – assessment and policy

The Australia-New Zealand Standard Research Classification (ANZSRC), originally developed by the Australian Bureau of Statistics (1993) and recently updated, crosses economic and academic boundaries. ANZSRC is used by policy bodies and is popular in research analysis. It has three dimensions: Types of Activity (TOA), derived from the OECD; 17 Socio-Economic Objectives (SEO); and 22 Fields of Research (FOR). The SEO and FOR systems are hierarchical to six digits and combine the Dewey-type concept with a contemporary range of subject groups. Thus SEO Division 86 (Manufacturing) has Group 8607 (Agricultural chemicals) and Objective 860702 (Chemical fertilizers), while FOR Division 09 (Engineering) has Group 0901 (Aerospace engineering) and Field 090101 (Aerodynamics). (In our notation we abbreviate Divisions as L1 and Groups as L2). The ANZSRC is employed by the Australian Research Council (ARC) in the cyclical assessment process of Excellence in Research for Australia (ERA). Submissions are made to each FOR by reference to expert-assigned journal lists. Each FOR has specified and exclusive content.

The FOR journal lists are mapped directly to journals indexed in the Web of Science. The lists are purposed for Australia and do not capture some extra-regional work, which may result in a reduced tally for a country or institution compared with direct analysis of the Web of Science data.

## Brazil: CAPES – evaluation of research staff

The Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) classification schema was created by the Foundation CAPES, linked to the Ministry of Education, to support the evaluation and skills improvement of higher education staff in Brazil. It is a hierarchical classification structured in three levels: nine broad areas; 49 evaluation areas; and 121 more granular sub-areas. Data on research performance in higher-level fields incorporate data on research performance in subordinate fields. For example, research in the Health Sciences (code 6) will include data on research in all nine subordinate evaluation areas (6.1 to 6.9). A report on research in the evaluation area Physical Education (6.8) will include data on research in all three subordinate sub-areas (6.8.1 to 6.8.3).

To map to the Web of Science, category elements within the CAPES category schema are compared to scope notes for the Web of Science journal categories. Each Web of Science category is assigned to one (or occasionally more than one) of the 121 CAPES sub-area subject codes and each CAPES category absorbs multiple Web of Science journal categories.

## Brazil: FAPESP – research investment

The Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) classification scheme was created to evaluate the scientific and technological development of the Brazilian state of São Paulo, which has prioritized investment in research. The classification is hierarchical and is structured into two levels: nine high-level categories and 72 more detailed categories. Research performance data in high-level fields incorporates data on research performance in subordinate fields. For example, a report on research in Health Sciences (4) will include data on research in all seven subordinate subject fields (4.1 to 4.7).

The FAPESP classification includes all Web of Science categories. Most journal categories are mapped to only one FAPESP category and multiple categories can be mapped to the same FAPESP category.

## Italy – university administration and review

The Agenzia Nazionale di Valutazione del Sistema Universitario e della Ricerca (ANVUR) classification scheme is based on an official academic fields and disciplines list for Italian universities' research and teaching, organized around 17 broad categories. ANVUR's approach links performance evaluation with primary knowledge-based missions: scientific research, teaching (only for universities) and socio-economic impact. This includes coordination of institutional Independent Evaluation Units (Nuclei di Valutazione); general analysis, statistical reports and benchmarks; sharing best practice; and on-site visits. In 2009, the Italian Parliament introduced new outcome-oriented performance evaluation for all public institutions. An Independent Commission for the Evaluation, Transparency and Integrity of Government (CIVIT) was appointed to monitor and evaluate strategic planning, performance and accountability. In 2013, ANVUR took over CIVIT's role regarding public universities and research institutes controlled by the Ministry of Education, Universities and Research (MIUR).

The ANVUR classification has been mapped to the Web of Science categories in a joint project between ANVUR and Clarivate to establish a foundation for bibliometric analyses carried out by ANVUR in 2013. The study developed indicators of international research standing where ANVUR assesses university research quality in an ANVUR Evaluation of Research Quality framework.

## Japan – grant funding analysis

The KAKEN category definitions are based on the Grant-in-Aid for Scientific Research (KAKENHI) Database. The objective is to develop researcher-led proposals, from basic to applied research in the humanities, social sciences and natural sciences. The grants provide financial support for creative and pioneering research projects that are expected to become the foundation of social development. The research projects are selected using a peer-review process, screened by multiple researchers in a field of specialization close to that of the applicant. Submissions to KAKENHI are made in approximately 300 categories, which are organized hierarchically into four levels.

The 2007 version has 66 categories in level 3 and 10 categories in level 2. Categories at each of these two levels were mapped to Web of Science categories via a study that collated articles linked to KAKENHI grants via the database operated by Japan's National Institute of Informatics (NII).

## China – university teaching and research

The China State Council Academic Degree Committee (SCADC) classification is based on the degree-granting and academic training directory published in 2018 by SCADC and the Ministry of Education of the People's Republic of China. It thus has a strong orientation towards the academic curricular structure, although it is now also used for research evaluation. The SCADC classification is hierarchical and has two levels. There are 13 broad-level categories represented by two-digit codes and 96 more granular categories numbered according to the broader category in which they fall. For example, Biology (0710) will roll up into Natural Science (07).

Clarivate has worked with SCADC to develop a journal mapping from the Web of Science to the SCADC classification. Some categories defy satisfactory translation into bibliometric analysis because they are under-represented in publications or because of subject area overlap. Some 33 subordinate SCADC categories would not translate into international mapping and one broad category (Military Science) is not recognized in international literature.

# Research on classification

Scientometric researchers take a special interest in classification since it is essential for calculating performance measures (Sjögårde 2019; Waltman 2016). Citations are collected as evidence of research impact but, for fair comparisons of papers across publication fields and years, the raw count must be normalized. Each field has a characteristic citation pattern and older papers have had more time to accrue citations. So, to calculate a normalized or relativized citation count for each paper, a reference set is needed drawn from papers with a similar citation pattern, given their age and focus. The average citation count for the reference set is then the baseline for citations to each particular paper (Moed, 2005; Waltman & van Eck, 2019).

There is no ground truth or gold standard to validate a reference system. Different approaches to their creation have strengths and weaknesses, conveniences and inconveniences. There are three main ways of defining such reference sets: **journal-defined fields**, in which a paper published in a journal belongs to the journal-associated field; **supervised information retrieval**, in which similar papers are aggregated using keywords, author names, journals and other attributes, as well as citation linkages; and article-level, **algorithmically constructed classification** (Zitt et al, 2019).

**Journal-defined fields,** such as the Web of Science categories, use a combination of citation analysis and informed judgment to define categories that are reasonable approximations of fields and subfields. They are understandable, accessible and convenient, and replication in data analysis is simple. Scientometrics has traditionally favored such systems, which allows for comparisons across studies (Glänzel & Schubert, 2003; Glänzel et al, 2009; Leydesdorff & Rafols, 2009). Among several journal-defined classification systems, the Web of Science categories have been, and likely still are, the most frequently used. However, the contents of multidisciplinary journals, such as *Science* and *Nature*, cannot be assigned to one or a few categories. A solution is to assign each paper in these serials to a category based on their cited references and the frequency with which cites to specific fields appear (Glänzel et al, 1999). In ESI, papers in multidisciplinary journals are treated in this way and assigned to ESI categories. This is not a simple undertaking for the end-user analyst in other contexts but generally available only to scientometricians, who build systems for such reassignments.

More importantly, journal-defined fields are incomplete representations of research fields since papers belonging to them are often published in journals assigned elsewhere. In fact, highly cited papers often appear in high-impact multidisciplinary journals. Furthermore, while the journals within a category are well-defined, they may also be heterogeneous in their focus and citation density. In such a case, comparing citations for a paper to the baseline of the category may seriously disadvantage (or advantage) the paper (van Eck et al, 2013). A recent study based on data from the Chinese Science Citation Database™ reports a roughly 50% accuracy rate in assigning papers to journal-defined categories (Shu et al, 2019). Finally, because journal-defined fields are typically, like the Web of Science categories, designed to aid information retrieval, the assignments may have consequences when deployed for a different purpose, a phenomenon called "indexer effects" (Rafols & Leydesdorff, 2009). In conclusion, journal-defined fields are robust but not refined representations of field and citation characteristics (Leydesdorff & Bornman, 2016; Milojevic, 2020; Wang & Waltman, 2016).

**Supervised information retrieval,** the second approach, requires expertise in subject matter, is tedious and little-used (Haunschild et al., 2018; Lewison, 1996). It increases the likelihood of finding highly similar content with the probability that a homogeneous set will provide a sound reference set for citation normalization. This comes at a cost: there are few people with search and subject expertise to create these custom collections; and the work is a one-off exercise with poor replicability.

**Algorithmic classification** uses relationships between papers, not groupings of journals, to define structure and fields, usually with hierarchical organization. Field delineation is most often based on analysis of a citation network, but lexical features and other attributes can be used and hybrid systems have been proposed (Boyack & Klavans, 2020; Janssens et al, 2009; Yu et al, 2017; Zitt & Bassecoulard, 2006). This bottom-up approach relies on the "association of ideas" concept described in proposing the first citation index (Garfield, 1955).

Henry Small, for many years ISI's Chief Scientist, pioneered co-citation in the 1970s as a method to define the specialty structure of science. Co-citation links publications related by frequent pairwise citation (Small, 1973). In 1974, Small and Belver Griffith of Drexel University, Philadelphia, used co-citation clusters to create maps of research specialties and employed multidimensional scaling to ordinate clusters according to their calculated similarity. Clustering algorithms and software for visualization are now a significant research activity in what has been termed algorithmically constructed classification systems using citation relationships among papers, including bibliographic coupling, co-citation and direct citation, and other techniques (Ahlgren & Colliander, 2009; Ahlgren et al 2020; Sjögårde & Ahlgren, 2018, 2020; Traag et al 2019; Waltman et al 2020; Waltman & van Eck, 2012). Certain characteristics are deemed more desirable than others, such as the range of cluster numbers and sizes (Perianes-Rodriguez & Ruiz-Castillo, 2017; Ruiz-Castillo & Waltman, 2015; Šubelj et al 2016). Nonetheless, it is difficult to assert that these methods are always advantageous or to be preferred since, at higher levels of aggregation, we can show that there is much agreement between article-level and journal-defined classification and performance measures. An evident disadvantage is the black-box nature of algorithms: different research groups are unable to recreate categorization schemes; and variation in the output of different algorithms using the same data and mappings raises a question as to whether the results are essentially algorithmic artefacts (Gläser et al, 2017).

## The effect of category granularity

Granularity in construction matters with any classification scheme and influences analysis. Zitt et al (2005) noted the possibility that Category Normalized Citation Impact (CNCI) would change according to the level (described as the 'zoom') at which any normalization occurs. A similar issue had been noted by Hirst (1978) in relation to 'Discipline Impact Factors'; comparison of bibliometric indicators across fields had been reviewed by Schubert & Braun (1993, 1996); and Glänzel & Moed (2002) also commented on the effect of different levels of aggregation.

Adams et al. (2008) tested the effect of the 'Zitt zoom' on research performance indicators by analyzing the relative impact of articles submitted for assessment in the U.K. RAE 2001 at different levels of normalization (Table 1). The data for university departments at the three highest grades (4, 5 and 5*) awarded in three Units of Assessment (UOA13 Psychology, UOA14 Biological Sciences and UOA19 Physics) showed a positive relationship between peer judgements and citation impact at some, but not all, levels of data aggregation. When citation counts were normalized at journal level there was little difference between impact metrics at any grade but normalization relative to Web of Science category or the entire UOA produced statistically significant higher relative impact for higher graded units, supporting Zitt et al.'s (2005) analysis.

The risk of fine-grained assessment in evaluation is that a category becomes self-referential. Material submitted by lower ranked units is implicitly sourced from journals of lower average impact than that submitted by leading units. Relative to the journal, the papers are of similar impact to the medium in which they are published. Only when we zoom out, to e.g. the Web of Science level, is the higher absolute citation count for papers from more highly graded units apparent.

**Table 1:** The average Category Normalized Citation Impact (CNCI) of articles and reviews published during 1996 to 2000 by research staff at U.K. universities for units graded 4, 5 or 5* in the Research Assessment Exercise 2001 (RAE 2001). Data are shown for three Units of Assessment (UOA) with the numbers of units at each grade and the CNCI for their publications with citation counts normalized at three levels of granularity: the journal of publication; the Web of Science journal category; and the data set for the entire UOA. (Adams et al., 2008)

| | **UOA13 Psychology** | | | | **UOA14 Biological sciences** | | | | **UOA19 Physics** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Average CNCI | | | | Average CNCI | | | | Average CNCI | | |
| **Grade at RAE 2001** | Number of units | Journal | Web of Science | UOA | Number of units | Journal | Web of Science | UOA | Number of units | Journal | Web of Science | UOA |
| **Grade 4** | 17 | 1.22 | 1.40 | 0.80 | 17 | 1.29 | 2.35 | 1.89 | 15 | 1.28 | 1.84 | 1.98 |
| **Grade 5** | 17 | 1.18 | 1.80 | 1.05 | 30 | 1.11 | 2.33 | 2.33 | 23 | 1.47 | 2.51 | 2.96 |
| **Grade 5*** | 12 | 1.32 | 2.38 | 1.63 | 11 | 1.18 | 2.53 | 2.93 | 5 | 1.82 | 3.32 | 3.75 |

# A new approach: bottom-up classification

The citation-based classification of articles and reviews is a bottom-up categorical system, in which individual elements are progressively linked into larger units with shared characteristics based on features in the underlying data.

Clarivate has now introduced a citation-based classification into InCites, developed collaboratively with the leading academic scientometrics team at the Centre for Science and Technology Studies (CWTS), Leiden University (Netherlands). It is intended to exploit the advantages claimed for article-level algorithmic classification:

- Greater accuracy in representing microclusters, or specialties;
- Increased homogeneity of content; and
- Improved citation normalization.

This innovation contrasts with the historical journal classification approach.  A data-driven approach produces categorization informed by article metadata rather than human concepts. It can be tuned, for example to produce more or fewer clusters (fine-grained or coarse classifications). And the underlying mechanics of how the categories emerge are derived from the data model.

To evaluate the feasibility and potential use of this approach, Clarivate collaborated with CWTS to build an algorithm that creates a categorical structure based on the citation network in the Web of Science. The algorithm supports a hierarchical system with a series of discrete levels progressively aggregating the smallest clusters. Levels were defined as micro (the most granular clusters created), meso (the first level of aggregation that groups similar micro clusters together), and macro (the largest aggregations that group meso clusters).

Challenges were jointly explored to inform decisions on how final data should be produced and updated.

**Timescale** – the analysis was applied to content post-1980 to align with data contained in our analytics product InCites and to reflect the timeframe most often used in longitudinal bibliometric analysis.

**Document types** – although the majority of bibliometric analyses are performed using only articles and reviews, all document types were included in the analysis. Note that only documents that have a citation link to others (i.e. citing or cited) can be incorporated into the classification, so some 'front matter' from journals, book chapters and proceedings may not be included.

**Cluster sizes** – it is possible to tune an algorithm to produce any number of clusters, but it is important to consider the volume of each cluster and its effect on the normalization of citation counts. InCites makes possible Category Normalized Citation Impact (CNCI) indices and Percentile metrics. A suitable minimum volume must exist in each category/year combination or normalization baselines would become unstable. Analysis showed that ~2,000 micro clusters could be created under this constraint, but more would introduce undue sparsity.

**Similarity measure** – the measures that can be used to determine similarity via a citation network are:

- **bibliographic coupling** (similarity based on shared cited references);
- **co-citation** (similarity based on citation of documents pairwise); and
- **direct citation** (any linkage).

Co-citation excluded too many documents from classification. The trade-off between bibliographic coupling and direct citation relates to dynamism. If similarity is measured via bibliographic coupling, the position of a document is static (the list of cited references never changes). However, a direct citation solution includes citing documents, which increase over time as more documents and citations are added, leading to the possibility that a document's cluster assignment could evolve. This is desirable, so existing documents may be reassigned as new topics emerge.

**Updating** – classification of static data is straightforward but a real-world solution must cope with regular updates so we had to consider how a bottom-up system should evolve over time. For practical reasons, two forms of updating are implemented:

**1. Monthly** – as new data are ingested, documents are assigned to an existing micro cluster based on their cited references.

**2. Yearly** – the clustering algorithm is rerun annually. At this point new micro clusters may emerge, or some documents may drift between different clusters (drift will likely affect recent documents, cited most in the years soon after publication). The algorithm has been tuned to moderate creation of new meso or macro level clusters during the yearly update.

Usability is the driving motivation for this solution. The resulting schema should be relatively static so users do not need to completely relearn the landscape with each annual update. At the same time, it should incorporate some agility to evolve so emerging topics can be established. To this end, CWTS incorporated a damping criterion that ensures no more than 5% of articles would change clusters during a yearly update.

**Labeling** – the outcome of the algorithm is assignment of documents to micro clusters; a hierarchy to group micro into meso clusters; and then to group meso into macro clusters. Clusters are labeled for preliminary user identification of contents. Labels were created for macro and meso level clusters by expert curation informed by domain experts and by summary information for a cluster, such as frequent author keywords and Web of Science categories. Micro clusters are allocated a label automatically based on the most significant author keywords associated with documents in the cluster.

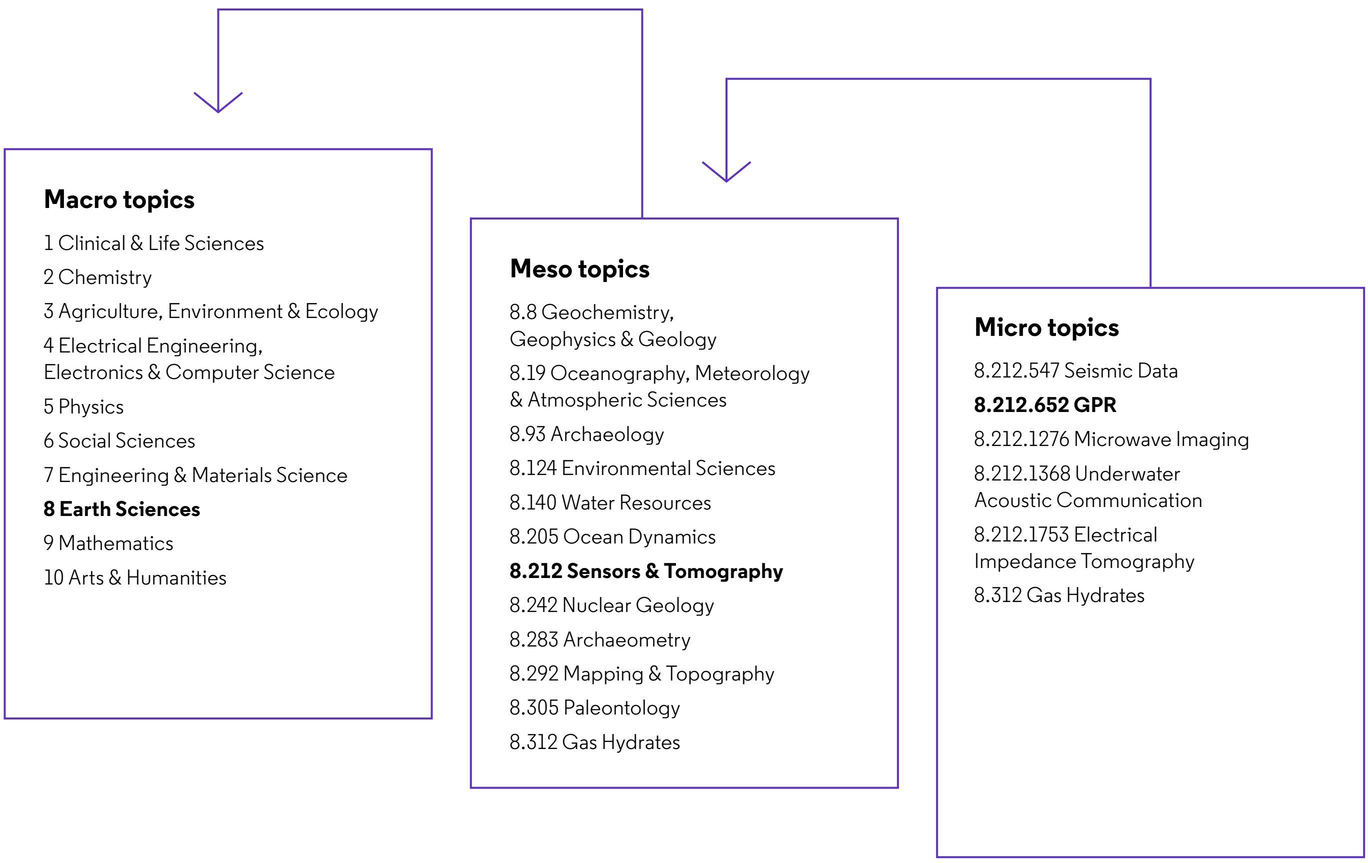
## Citation Topics

The new categorization scheme 'Citation Topics', based on the CWTS methodology, was added to InCites in December 2020. The name 'Citation Topics' is used because the classification system is built from the citation network and the resulting clusters vary in nature. Some align well with disciplinary labels (e.g. ophthalmology) while others are focused on specific diseases, materials or analytical techniques. The term 'topic' seemed more suitable than category or class, which imply a formal structure.

The current implementation (December 2020) is composed of 10 macro topics, 326 meso topics and 2,444 micro topics. More than 60 million documents were analyzed and more than 50 million were assigned a Citation Topic. For the substantive research document types (articles and reviews), a high percentage of documents could be assigned to a topic for the full data from 1980 onwards (92% of articles and 96% of reviews) and this improved further for the most recent five years (95% of articles and 99% of reviews).

In Figure 1, we list the 10 broad macro topics and illustrate how 'Earth Sciences', as an example, is aggregated from 12 meso topics, one of which ('Sensors & Tomography') is aggregated from five micro topics, of which one is highlighted for GPR (ground penetrating radar).

**Figure 1: Three levels of the Citation Topics hierarchy described for Earth Sciences**



## Citation Topics compared to other classifications

To compare the categorical structure in Citation Topics (article-level clustering) with established journal-based schemes, we created a series of research maps to visualize the similarities and differences (Figure 2. A-C). The maps focus on the ESI category "Geosciences" and include only the 48,000 articles and reviews published in 2015. In these diagrams, documents are located on a common mapping landscape, i.e. the same layout is used for each figure, with each document positioned according to its cited references and their similarity to all other documents (bibliographic coupling). We utilize Uniform Manifold Approximation (UMAP) (McInnes & Healy, 2018) to project the feature space onto two dimensions so the document space can be easily visualized. Then, in each plot, a different color scheme is applied to show how the same set of 'ESI-Geosciences' documents would be categorized in that classification system. The three exemplar systems are: Macro Citation Topics; Meso Citation Topics; and Web of Science categories.

The maps show that there are differences in how the same content is grouped or clustered by a particular classification scheme. Although the methodological differences between bottom-up (citation clustering) versus top-down (journal categories) are substantial, it is clear that these classification groups align across the landscape. This suggests that a 'natural order' is underpinning the guiding principles inherent in both expert judgment and citation linkage. Differences within this can be explained by interpretation of the way one discipline or field relates to other, cognate areas. An evident benefit of Citation Topics is that the more granular level of categorization (Micro Topics) gives opportunities for new groups to appear, which was not possible with older journal-based schemes.

**Figure 2A. Citation Topics, Macro.** The first picture shows how 2015 documents in the ESI field of Geosciences are arranged and colored according to Macro Topics. The majority of the ESI content has been clustered by citation links in the Macro Topic – Earth Sciences (green). On the right of the plot, a significant component has been allocated to Engineering (orange) and, on the left, another cluster is assigned to Physics (purple).

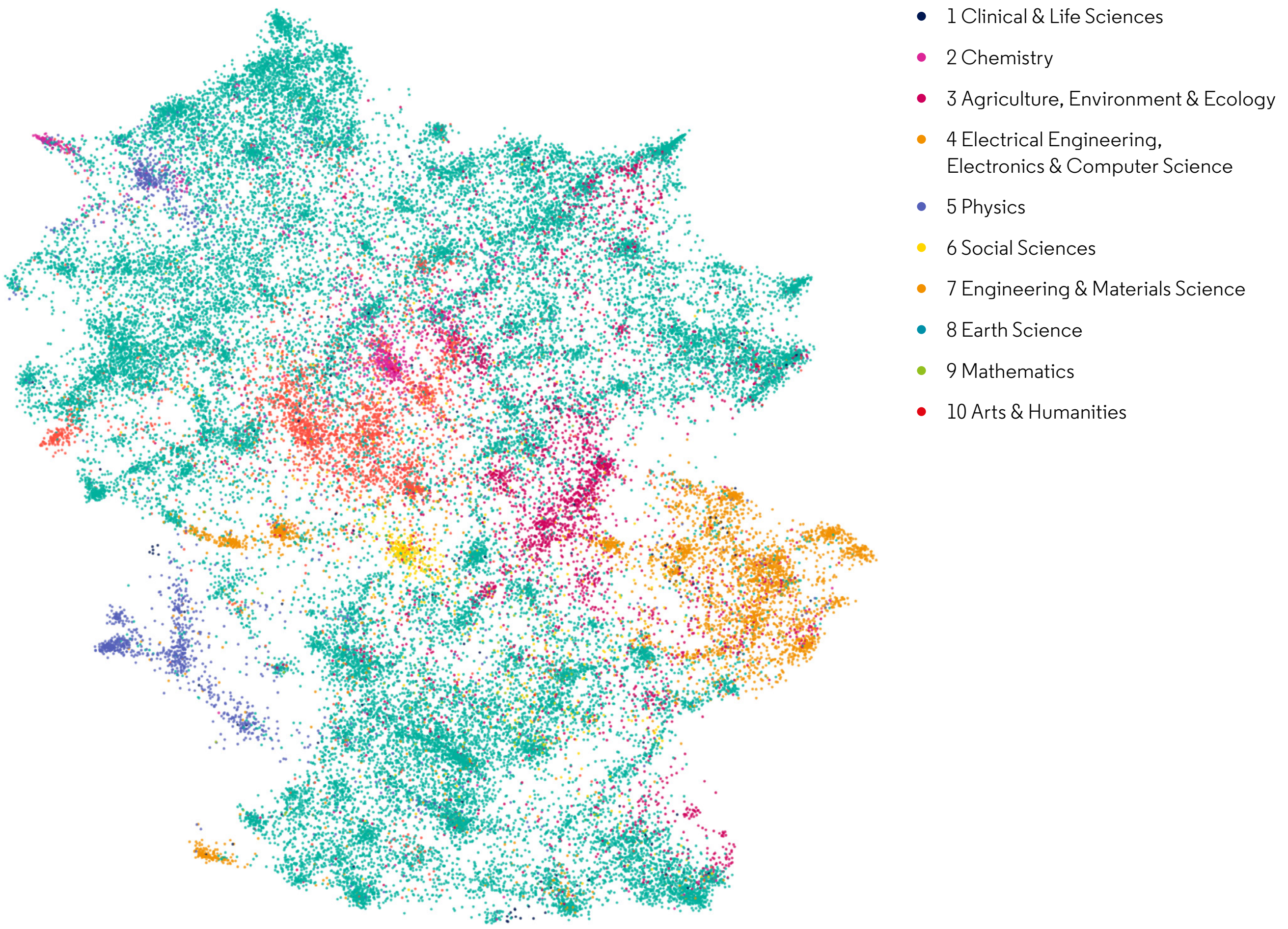


**Figure 2B. Citation Topics, Meso.** The second map shows Meso Topics so this follows the same citation-based article-level categorization but at a finer level of detail. Here it is apparent that the major groupings of 'Geochemistry, Geophysics & Geology' (8.8, top, green) and 'Oceanography, Meteorology & Atmospheric Sciences' (8.19, bottom, orange) segment the content.

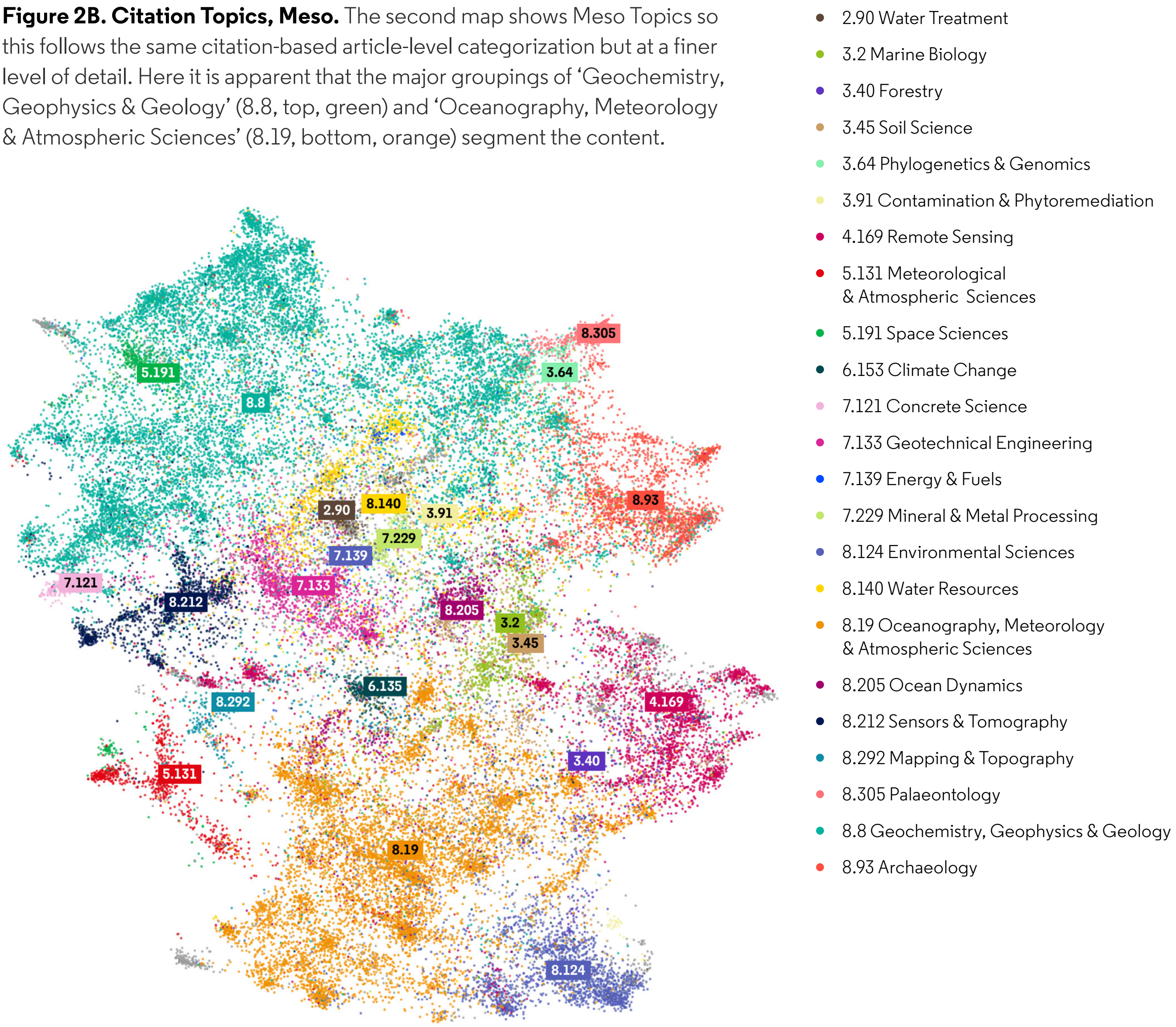
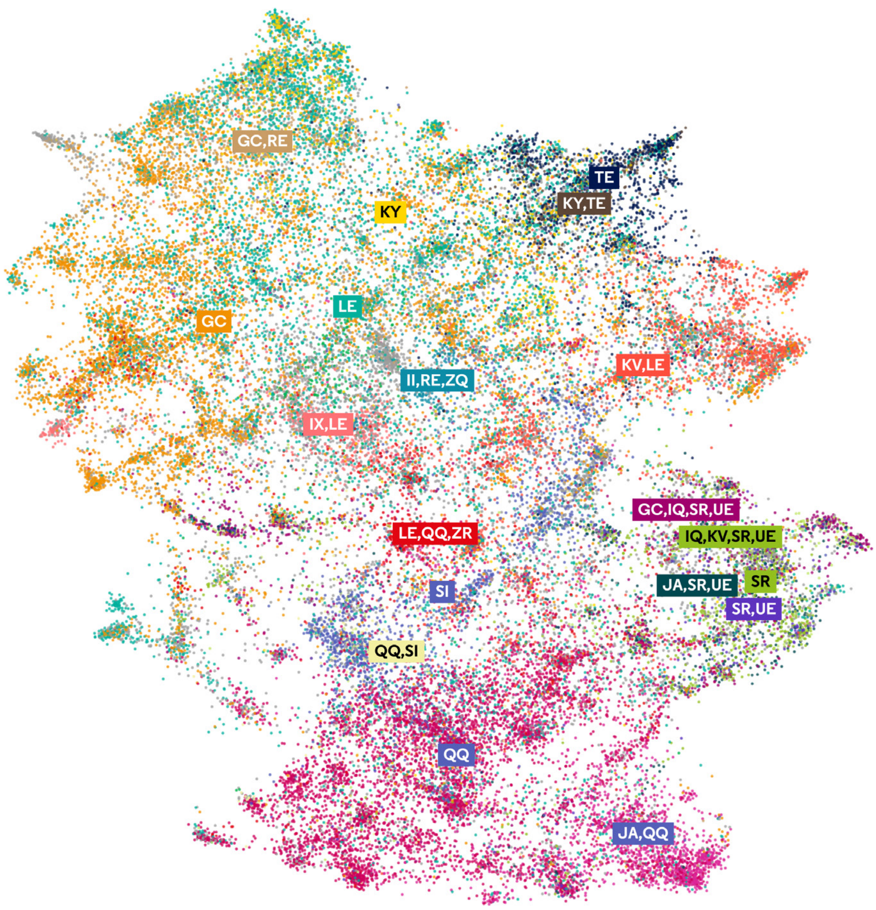


**Figure 2C. Web of Science Categories.** The final picture shows the Web of Science categories to which the set of papers have been assigned. These are mapped at the journal level and many journals are assigned to multiple categories. Hence, a color may correspond to a category-combination and similar colors make the map slightly harder to decipher, especially on the right side where a number of engineering categories are located. The main groupings at the top of the map are Geosciences, multidisciplinary (green, LE), Geochemistry & Geophysics (orange, GC), and Geology (yellow, KY) which overlaps broadly with the Meso Topic 8.8 Geochemistry, Geophysics & Geology. Similarly, the major group at the bottom is composed of Meteorology & Atmospheric Sciences (purple, QQ) and QQ with Environmental Sciences (pink, QQ+JA). Astute readers will notice a portion of the map (circled) in the lower group that is not included with the Meteorological content (purple and pink) but is grouped together in the corresponding Meso cluster 8.19 Oceanography, Meteorology & Atmospheric Sciences. This content is specifically related to Oceanography (SI) and highlights the slight but real differences in granularity that are reflected in the assigned labels.




- [LE] Geosciences, Multidisciplinary
- [GC] Geochemistry & Geophysics
- [QQ] Meteorology & Atmospheric Sciences
- [KV, LE] Geography, Physical Geosciences, Multidisciplinary
- [SI] Oceanography
- [JA, QQ] Environmental Sciences, Meteorology & Atmospheric Sciences
- [KY] Geology
- [TE] Palaeontology
- [SR] Remote Sensing
- [LE, QQ, ZR] Geosciences, Multidisciplinary, Meteorology & Atmospheric Sciences, Water Resources
- [GC, IQ, SR, UE] Geochemistry & Geophysics
- [GC, RE] Geochemistry & Geophysics, Mineralogy
- [ID, IP] Energy & Fuels, Engineering, Petroleum
- [IX, LE] Engineering, Geological, Geosciences, Multidisciplinary
- [JA, SR, UE] Environmental Sciences, Remote Sensing, Imaging Science & Photographic Technology
- [IQ, KV, SR, UE] Engineering, Electrical & Electronic
- [KY, TE] Geology, Palaeontology
- [II, RE, ZQ] Engineering, Chemical, Mineralogy, Mining & Mineral Processing
- [SR, UE] Remote Sensing, Image Science & Photographic Technology
- [QQ, SI] Meteorology & Atmospheric Sciences, Oceanography

# The impact of classification

InCites, Clarivate's analytical bibliometric package, provides the user with multiple choices of top-down data classifications that now includes the bottom-up Citation Topic classification based on our work with CWTS. How much does the choice of data classification affect the range of data used, the way it is grouped, the degree of granularity, and how does it change baselines as well as sample content?

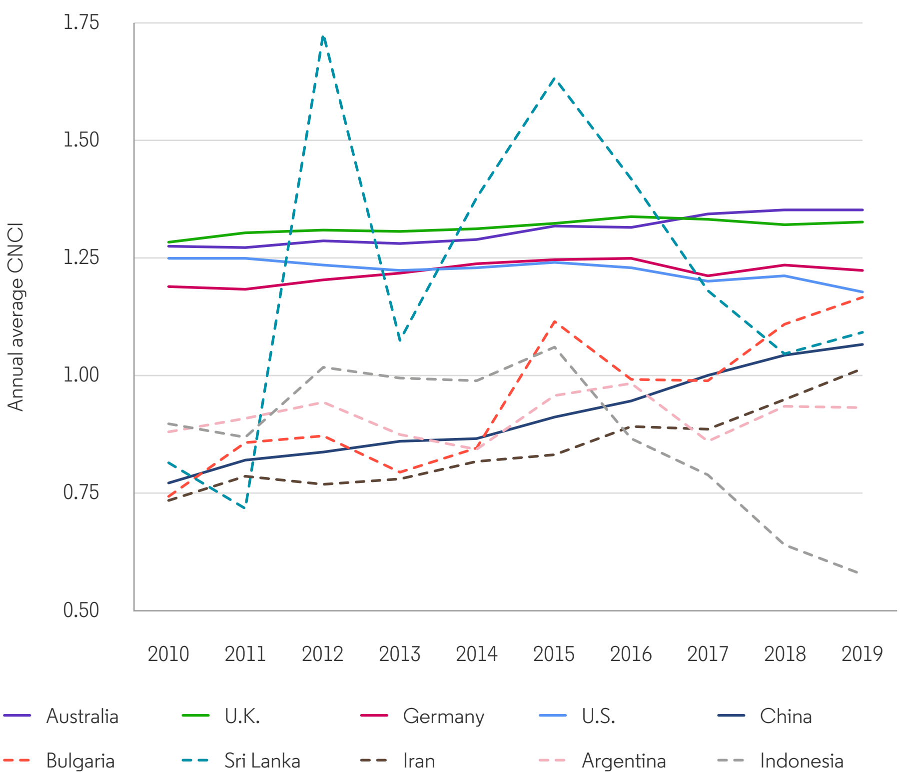
To explore these questions, we compared the outcome of analyses based on the 254 journal-based categories of the Web of Science (covering all subject domains) with those based on alternative classifications. Our summary metrics (Table 2) describe the publication volume (count of articles and reviews) and citation impact (CNCI) of ten countries. Five of these have both large and well-funded research economies (U.S., China, U.K., Germany and Australia) and the other five, while improving, presently have both relatively weaker funding and smaller research output.

Does Sri Lanka really have an average CNCI equal to the U.S. when the latter produces more than 500 times as many publications? What does it mean if Iran has the highest rate of cited papers when it is the second lowest in average CNCI? How, in other words, are these point metrics compiled and calculated? The data on CNCI trends (Figure 3) suggest that average index for at least 2 of the 10 nations may be unreliable: the average for Sri Lanka is not only high but volatile. Does that introduce doubt about the more stable values? It certainly suggests a need to know more about the mass of publications that feed the indicators for each economy. Can a single indicator stand for millions of publications and tens of millions of citations? More information is evidently required to properly interpret Table 2 and Figure 3, and to explore how such results might be influenced by choosing a particular classification system.

**Table 2.** Summary metrics for the research output (numbers of articles and reviews indexed in the Web of Science) and performance (average category normalized citation impact, CNCI world average = 1.0) of 10 regions during a recent 10-year period (2010 to 2019). Regions are ranked on CNCI.

| | Papers | CNCI | Citations | % cited |
|---|---|---|---|---|
| U.K. | 1,981,903 | 1.41 | 26,932,154 | 65.6 |
| Australia | 888,127 | 1.41 | 12,626,406 | 72.4 |
| U.S. | 6,838,175 | 1.31 | 90,031,964 | 63.9 |
| Sri Lanka | 13,068 | 1.31 | 170,284 | 63.6 |
| Germany | 1,615,968 | 1.30 | 23,029,125 | 71.1 |
| Bulgaria | 38,366 | 1.01 | 360,385 | 60.2 |
| China | 3,743,888 | 0.99 | 39,306,476 | 71.5 |
| Argentina | 121,077 | 0.96 | 1,321,844 | 71.4 |
| Iran | 362,748 | 0.91 | 3,428,680 | 77.9 |
| Indonesia | 85,885 | 0.81 | 342,576 | 39.1 |

**Figure 3.** Annual trends in Category Normalized Citation Impact (CNCI) for 10 regions (five established and five growing research economies) during a recent 10-year period (2010 to 2019).



## Classification and volume

The total available Web of Science publication dataset is often greater than the number of papers actually assigned to each country if any of six other classifications available in InCites are chosen (Table 3). Some, especially the journal lists for the ANZSRC FORs, reduce the available data for countries such as Indonesia by as much as half. Even for the U.S. and the U.K. the publication set is down by 20% (the L1 Divisional categories) or 35% (the more specific L2 Group categories). By contrast, the schema for the U.K. REF and those used in Brazil by CAPES and FAPESP essentially draw on the full source material.

**Table 3.** The ratio between numbers of papers assigned to the 10 regions listed in Table 3 via the Web of Science journal-based disciplinary category scheme and six other classifications used in InCites. Variation in the scope of the literature that is covered will affect both the numerator and denominator citation counts in any subsequent normalization calculation of citation impact.

| | ESI | FOR L1 | FOR L2 | REF2014 | CAPES49 | FAPESP |
|---|---|---|---|---|---|---|
| U.S. | 0.85 | 0.80 | 0.64 | 1.00 | 1.00 | 1.00 |
| China | 0.78 | 0.71 | 0.54 | 1.00 | 1.00 | 1.00 |
| U.K. | 0.81 | 0.80 | 0.65 | 1.00 | 1.00 | 1.00 |
| Germany | 0.84 | 0.77 | 0.61 | 1.00 | 1.00 | 1.00 |
| Australia | 0.86 | 0.84 | 0.67 | 1.00 | 1.00 | 1.00 |
| Iran | 0.90 | 0.80 | 0.64 | 1.00 | 1.00 | 1.00 |
| Argentina | 0.90 | 0.82 | 0.66 | 1.00 | 1.00 | 1.00 |
| Indonesia | 0.34 | 0.35 | 0.26 | 0.98 | 1.00 | 1.00 |
| Bulgaria | 0.75 | 0.60 | 0.48 | 0.98 | 1.00 | 1.00 |
| Sri Lanka | 0.76 | 0.72 | 0.59 | 0.99 | 1.00 | 1.00 |

## Classification and impact

The schema also affect CNCI. Reassuringly, there is a high degree of correlation between CNCI values obtained from citation counts normalized under different categorical systems but the correlation is not perfect. There can be differences both in the y-intercept, which would move all values up or down, and in the slope, which would differentially affect organizations with lower and higher average impact.

Matching data categorization to the objectives of the assessment is essential if equity is to be maintained across all parties under assessment. The average CNCI for 86 U.K. universities (2015 to 2019), taken across all discipline categories in each of several different categorical systems, is shown in Figure 4. The effect of moving from the Web of Science journal categories to the FOR 2-digit Level 1 is to depress most institutional CNCIs but this is most marked below world average CNCI and almost negligible at the upper end of the distribution. There are some evident outliers, so the effect is far from uniform. There is a much closer correlation between the CNCI values for the Web of Science categories and the categories created by the CWTS Citation Topic clustering. Comparison of CNCI using CWTS meso categories and the FOR1 categories shows again that the FOR system depresses the CNCI values. A shift to a finer-grained level, using the CWTS micro and the ANZ FOR Level 2 categories, produces a similar effect and depression in the low CNCI part of the distribution is relatively greater. (Figure 4)

**Figure 4.A-D.** Correlations between the average Category Normalized Citation Impact (CNCI) of Web of Science indexed publications for 86 U.K. universities (2015 to 2019) when different schema are used to categorize the institutional data and create global benchmarks. All the correlations are highly significant but the variance about the regression differs for specific institutions. To track this, four universities with distinct research histories and portfolios are highlighted with a constant color point. (Web of Science categories map journals to 254 fields; ANZSRC Fields of Research (FOR) L1 = journals mapped to 24 broad Divisions and L2 = 212 specific Groups nested within L1; CWTS MESO and MICRO refer to coarse and fine Citation Topic categories developed by CWTS, University of Leiden).

**Figure 4A. FOR1 vs Web of Science, correlation = 0.97**

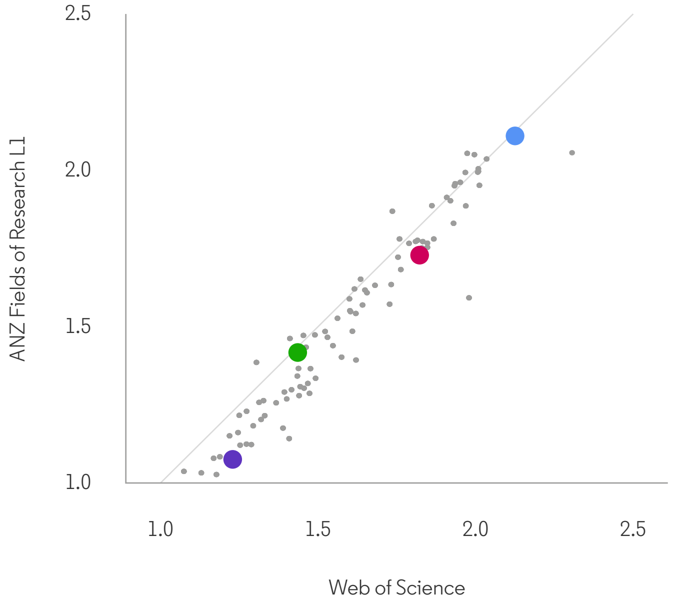



**Figure 4B. CWTS Meso vs FOR1, corr = 0.95**

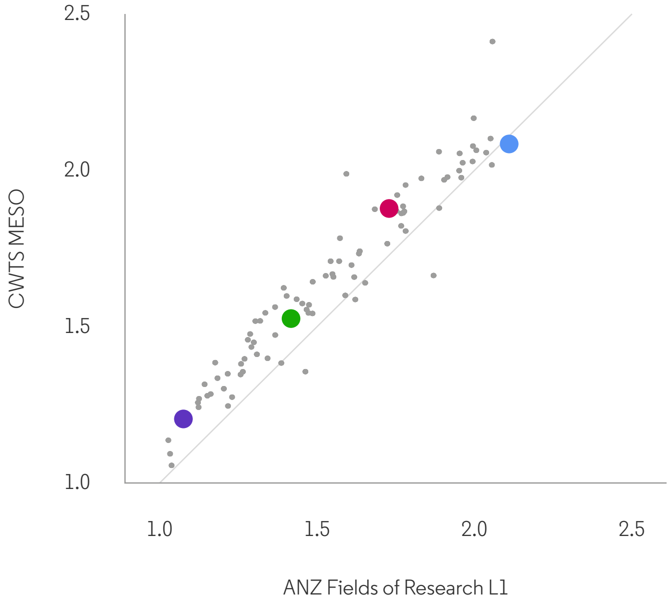



**Figure 4C. CWTS Meso vs Web of Science, corr = 0.99**

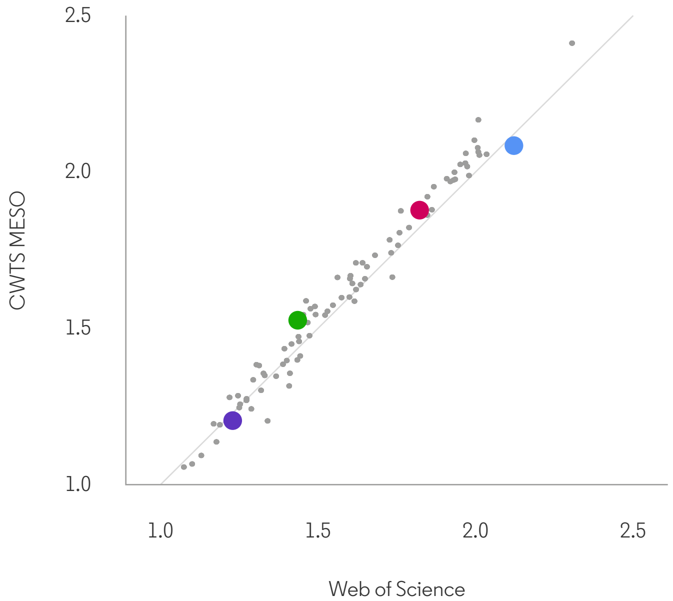
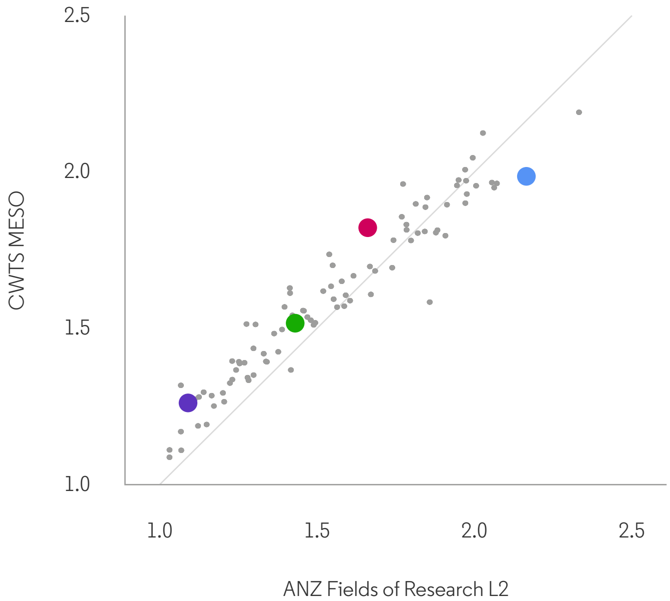



**Figure 4D. CWTS Micro vs FOR2, corr = 0.93**



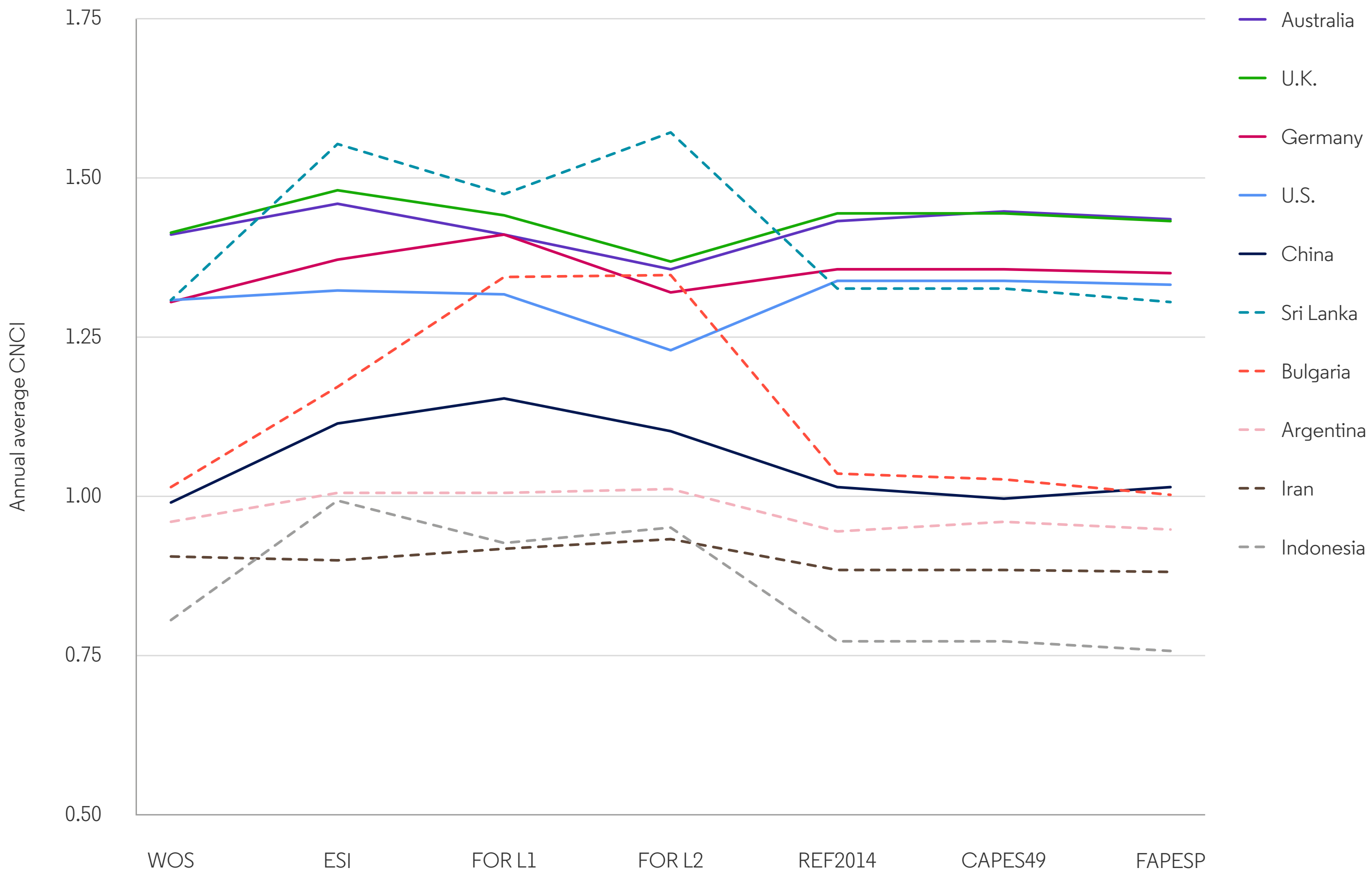
The changes in relative positions for the four tracked universities illustrates the considerable residual variance because the shift from one system to another is never uniform across all four. There are six universities which have an average CNCI of 1.7 when Web of Science journal categories were used for normalization whereas they would display CNCI values ranging between 1.45 and 1.85 if FOR L1 Divisional categories were used (Figure 4.A). Among the four tracked universities, the highest performer gains in the shift from CWTS-MESO to FOR1 (Figure 4.B) and to FOR2 (Figure 4.D), but the other three all suffer a detriment.

Whether these shifts are due to subject mix, because each system assigns journals differently across categories so global baselines change, or another factor, it materially affects the relative institutional outcomes.

We can now compare the outcomes of different classification schemes on national performance indicators. Table 2 (based on Web of Science journal categories) suggested that CNCI for Sri Lanka was similar to that of the U.S. and Germany. Figure 5 shows that the use of either the ESI or the two ANZ FOR schema would have produced outcomes in which Sri Lanka is world-beating. Indonesia's CNCI would also be elevated if these schema are used, but most countries' CNCI is affected much less – although that of the U.S., U.K., Australia and Germany are all slightly depressed under FOR Level 2.

None of these are wrong answers. The lesson here is that choosing a scheme for data selection and aggregation will influence analysis and interpretation, yet none of the alternative schema have been implemented casually or without planning, analysis and prior development and all present reasonable, fact-based outcomes.

**Figure 5.** The average Category Normalized Citation Impact (CNCI) for 10 regions calculated with data normalized under seven different classification schemes. The numbers of publications used to calculate CNCI vary between schema are indicated in Table 3. The graph lines do not imply any connection between distinct schema but are inserted as a visual aid.



# Implications for responsible metrics

The responsible user of bibliometric data needs to be clear whether the data they have are relevant to the evaluation questions they pose; they need to establish an *a priori* understanding of how they will use the data and of the choices of methods to apply; and that choice of methods must include careful understanding of the classification scheme they will use to group and benchmark the data.

It should be clear that evaluators would be incautious if they were to rely solely on the summary information in Table 2 to make judgments about the relative or absolute research strengths, even of whole countries. This should be even more important if they were reviewing a table of institutions from the same countries or a set of their research groups seeking funding, and yet this happens frequently.

Highly granular categorical systems group research papers into small, self-referential pockets that boost the apparent relative citation performance of work that appears poorly cited in familiar topical aggregations. More generally, the effect of a choice of discipline/topic categories for aggregating publications and normalizing citations is two-fold. First, countries with a less developed domestic research base, and less well cited domestic research output, will tend to have smaller publication tallies when more exclusive categorical systems (such as ESI and the ANZSRC FORs) are used. Second, because such categories focus on journals selectively, it is the least well cited part of a country's activity that is omitted, so their average CNCI is raised. So, although publication counts for Sri Lanka, Bulgaria and Indonesia are significantly reduced in an ESI analysis compared with a Web of Science analysis, they nonetheless then have higher average CNCI.

We wholly endorse the views of Professor Henk Moed (Moed, 2020a, 2020b) regarding the need for an evaluation framework in which the context and the purpose of the exercise are overriding considerations. Citations are themselves value-laden constructs with social as well as research weight. Any aggregation of citation counts, subsequent management of the data through normalization and fractionation, and choice of analytical methodology then applied, must introduce further subjective modification that moves from original information towards a stylized indicator.

Users planning a research evaluation should be aware of these summary points:

- **Purpose:** data classification meets user need: some users focus on academic performance while others focus on economic or social benefit; and some align with research fields while others align with a standard curriculum.
  - › USERS need to consider whether their objectives align with the designer of the classification.
- **Categories:** there are many systems for assigning journals and/or individual publications to discipline categories and none is uniquely correct.
  - › USERS should take a researcher's output portfolio into account in choosing a data source.
- **Granularity:** a choice of broad or narrow focus is made when citation counts are normalized against a global benchmark, for comparative purposes or to aggregate data across years and disciplines.
  - › USERS need to be aware of granularity and choose an appropriate level of aggregation.
- **Coverage:** not all topics within an evaluated unit's research are covered equitably by all classification schemes, and the need for equity applies equally to discipline and region.
  - › USERS should determine whether the classification captures data equitably for all stakeholders.

Data categorization is not a trivial consideration in research policy, management and evaluation, as this report has demonstrated. When evaluators are clear about objectives, the questions to be addressed, the relevance of bibliometrics to those questions, the nature of the available data, and the place of the bibliometric analysis within an overall evaluative framework, then they should proceed to work through the issues and determine whether they have fully understood the implications of these and the outcome in the context of their purpose and materials.

# References


Adams, J., Gurney, K., & Jackson, L. (2008). Calibrating the zoom — a test of Zitt's hypothesis. Scientometrics, 75(1), 81-95.

Ahlgren, P., & Colliander, C. (2009). Document–document similarity approaches and science mapping: Experimental comparison of five approaches. Journal of Informetrics, 3(1), 49–63.

Ahlgren, P., Chen, Y., Colliander, C., & van Eck, N. J. (2020). Enhancing direct citations: A comparison of relatedness measures for community detection in a large set of PubMed publications. Quantitative Science Studies, 1(2), 714-729.

Atran, S. (1990). Cognitive foundations of natural history: Towards an anthropology of science. Cambridge, UK: Cambridge University Press. ISBN 978-0521438711

Boyack, K. W., & Klavans, R. (2020). A comparison of large-scale science models based on textual, direct citation and hybrid relatedness. Quantitative Science Studies, 1(4), forthcoming.

Colliander, C., & Ahlgren, P. (2019). Comparison of publication level approaches to ex post citation normalization. Scientometrics, 120(1), 283–300.

Dousa, T. M. (2017). E. Wyndham Hulme's classification of the attributes of books: On an early model of a core bibliographical entity.Knowledge Organization, 44(8), 592-604.

Garfield, E. (1955). Citation indexes for science – a new dimension in documentation through association of ideas. Science, 122(3159), 108-111. http://garfield.library.upenn.edu/papers/science1955.pdf

Garfield, E. (1957). Breaking the subject index barrier – a citation index for chemical patents. Journal of the Patent Office Society, 39(8), 583-595. http://www.garfield.library.upenn.edu/essays/v6p472y1983.pdf

Garfield, E. (1964). Science Citation Index: a new dimension in indexing. Science, 144(3619), 649-654. http://www.garfield.library.upenn.edu/essays/v7p525y1984.pdf

Glänzel, W., Schubert, A., & Czerwon, H. J. (1999). An item-by-item subject classification of papers published in multidisciplinary and general journals using reference analysis. Scientometrics, 44(3), 427-439.

Glänzel, W., & Moed, H. F. (2002). Journal impact measures in bibliometric research. Scientometrics, 53(2), 171 – 193.

Glänzel, W., & Schubert, A. (2003). A new classification scheme of science fields and subfields designed for scientometric evaluation purposes. Scientometrics, 56(3), 357-367.

Glänzel, W., Thijs, B., Schubert, A., & Debackere, K. (2009). Subfield-specific normalized relative indicators and a new generation of relational charts: Methodological foundations illustrated on the assessment of institutional research performance. Scientometrics, 78(1), 165-188.

Gläser, J., Glänzel, W., & Scharnhorst, A. (2017). Same data – different results? Towards a comparative approach to the identification of thematic structures in science. Scientometrics, 111(2), 981-998.

Handy, C. (1989). The Age of Unreason. Boston, MA: Harvard Business School Press.

Haunschild, R., Schier, H., Marx, W., & Bornmann, L. (2018). Algorithmically generated subject categories based on citation relations: An empirical micro study using papers on overall water splitting. Journal of Informetrics, 12(2), 436-447.

Hirst, G. (1978). Discipline impact factors: A method for determining core journal sets. Journal of the American Society for Information Science, 29(4), 171-172.

Hulme, E.W. (1923). Statistical Bibliography in Relation to the Growth of Modern Civilization: Two Lectures Delivered in the University of Cambridge in May, 1922. London: Butler & Tanner, Grafton & Co.

Janssens, F., Zhang, L., De Moor, B., & Glänzel, W. (2009). Hybrid clustering for validation and improvement of subject-classification schemes. Information Processing & Management, 45(6), 683-702.

Kessler, M. M. (1963). Bibliographic coupling between scientific papers. American Documentation, 14(1), 10-25.

Klavans, R., & Boyack, K. W. (2011). Using global mapping to create more accurate document level maps of research fields. Journal of the American Society for Information Science and Technology, 62(1), 118.

Klavans, R., & Boyack, K. W. (2017). Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge? Journal of the Association of Information Science and Technology, 68(4), 984-998.

Lewison, G. (1996). The definition of biomedical research subfields with title keywords and application to the analysis of research outputs. Research Evaluation, 6(1): 25–36.

Leydesdorff, L., & Rafols, I. (2009). A global map of science based on the ISI subject categories. Journal of the American Society for Information Science and Technology, 60(2), 348-362.

Leydesdorff, L., & Bornmann, L. (2016). The operationalization of "fields" as WoS subject categories in evaluative bibliometrics: The cases of "library and information science" and "science & technology studies," Journal of the Association of Information Science and Technology, 67(3), 707-714.

Leydesdorff, L., Comins, J. A., Sorensen, A. A., Bornmann, L., & Hellsten, I. (2016). Cited references and Medical Subject Headings (MeSH) as two different knowledge representations: Clustering and mappings at the paper level. Scientometrics, 109(3), 2077-2091.

McInnes, L., & Healy, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, ArXiv e-prints 1802.03426.

Milojević, S. (2020). Practical method to reclassify Web of Science articles into unique subject categories and broad disciplines. Quantitative Science Studies, 1(1), 183-206.

Moed, H. F. (2005). Citation analysis in research evaluation. Dordrecht: Springer, Chp. 14.3, 187-189.

Moed, H. F. (2020a). Appropriate use of metrics in research assessment of autonomous academic institutions. Scholarly Assessment Reports, 2(1), 1.

Moed, H. F. (2020b). How evaluative informetrics relates to scientific, socio-historical, political, ethical, and personal values. Scholarly Assessment Reports, 2(1), 9.

Neuhaus, C., & Daniel, H.-D. (2009). A new reference standard for citation analysis in chemistry and related fields based on the sections of Chemical Abstracts. Scientometrics, 78(2), 219-229.

Perianes-Rodriguez, A., & Ruiz-Castillo, J. (2017). A comparison of the Web of Science and publication-level classification systems of science. Journal of Informetrics, 11(1), 32-45.

Price, D. J. de S. (1965). Networks of scientific papers. Science, 149(3683), 510-515.

Price, D. J. de S. (1980). Foreword. In Garfield, E., Essays of an Information Scientist. Volume 3. 1977-1978. Philadelphia, PA: ISI Press, v-ix. http://www.garfield.library.upenn.edu/essays/v3forewordy1977-78.pdf

Pudovkin, A. I., & Garfield, E. (2002). Algorithmic procedure for finding semantically related journals. Journal of the American Society for Information Science and Technology, 53(13), 1113-1119. http://garfield.library.upenn.edu/papers/pudovkinsemanticallyrelatedjournals2002.html

Rafols, I., & Leydesdorff, L. (2009). Content-based and algorithmic classifications of journals: Perspectives on the dynamics of scientific communication and indexer effects. Journal of the American Society for Information Science and Technology, 60(9), 1823-1835.

Ruiz-Castillo, J., & Waltman, L. (2015). Field-normalized citation impact indicators using algorithmically constructed classification systems of science. Journal of Informetrics, 9(1), 102–117.

Schubert, A., & Braun, T. (1993). Reference standards for citation-based assessments. Scientometrics, 26(1), 21-35.

Schubert, A., & Braun, T. (1996). Cross-field normalization of scientometric indicators. Scientometrics,
36(3), 311-324.

Shu, F., Julien, C.A., Zhang, L., Qiu, J., Zhang, J., & Larivière, V. (2019). Comparing journal and paper level classifications of science. Journal of Informetrics, 13(1), 202225.

Shu, F., Ma, J., & Larivière, V. (2020). Classifications of science and their effects on bibliometric evaluations. Scientometrics, 125(3), 2727-2744.

Sjögårde, P. (2019). The challenge of categorizing research. Leiden Madtrics. https://leidenmadtrics.nl/articles/the-challenge-of-categorizing-research

Sjögårde, P., & Ahlgren, P. (2018). Granularity of algorithmically constructed publication-level classifications of research publications: Identification of topics. Journal of Informetrics, 12(1), 133–152.

Sjögårde, P., & Ahlgren, P. (2020). Granularity of algorithmically constructed publication-level classifications of research publications: Identification of specialties. Quantitative Science Studies, 1(1): 207-237.

Small, H. (1973). 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), 265-269.

Small, H., & Griffith, B. C. (1974). The structure of scientific literatures I: Identifying and graphing specialties. Science Studies, 4(1), 1740.

Small, H. (2003). Paradigms, citations, and maps of science: A personal history. Journal of the American Society for Information Science and Technology, 54(5), 394 –399.

Šubelj, L., van Eck, N. J., & Waltman, L. (2016). Clustering scientific publications based on citation relations: A systematic comparison of different methods. PLoS ONE, 11(4), e0154404.

Traag, V. A., Waltman, L., & van Eck, N. J. (2019). From Louvain to Leiden: Guaranteeing well-connected communities. Scientific Reports, 9, article number 5233.

Van Eck, N. J., Waltman, L., Van Raan, A. F., Klautz, R. J., & Peul, W. C. (2013). Citation analysis may severely underestimate the impact of clinical research as compared to basic research. PLoS ONE, 8(4), e62395.

Waltman, L. (2016). A review of the literature of citation impact indicators. Journal of Informetrics, 10(2), 365-391.

Waltman, L., & van Eck, N. J. (2012). A new methodology for constructing a publication-level classification system of science. Journal of the American Society for Information Science and Technology, 63(12), 2378-2392.

Waltman, L., & van Eck, N. J. (2019). Field normalization of scientometric indicators. In Wolfgang Glänzel, Henk F. Moed, Ulrich Schmoch, and Mike Thelwall, Eds., Springer Handbook of Science and Technology Indicators. Cham, Switzerland: Springer Nature, 2019, 281-300. ISBN-13: 978-3030025106

Waltman, L., Boyack, K. W., Colavizza, G., & van Eck, N. J. (2020). A principled methodology for comparing relatedness measures for clustering publications. Quantitative Science Studies, 1(2), 691-713.

Wang, Q., & Waltman, L. (2016). Large-scale analysis of the accuracy of the journal classification systems of the Web of Science and Scopus. Journal of Informetrics, 10(2), 347-364.

Yu, D., Wang, W., Zhang, S., Zhang, W., & Liu, R. (2017). Hybrid self-optimized clustering model based on citation links and textual features to detect research topics. PLoS ONE, 12(10), e0187164.

Zhang, L., Janssens, F., Liang, L., & Glänzel, W. (2010). Journal cross-citation analysis for validation and improvement of journal-based subject classification in bibliometric research. Scientometrics, 82(3), 687-706.

Zitt, M., Ramanana-Rahary, S., & Bassecoulard, E. (2005). Relativity of citation performance and excellence measures: From cross-field to cross-scale effects of field-normalisation. Scientometrics, 63(2), 373-401.

Zitt, M., & Bassecoulard, E. (2006). Delineating complex scientific fields by an hybrid lexical-citation method: An application to nanosciences. Information Processing and Management, 42(6), 1513–1531.

Zitt, M., Lelu, A., Cadot, M., & Cabanac, G. (2019). Bibliometric delineation of scientific fields. In Wolfgang Glänzel, Henk F. Moed, Ulrich Schmoch, and Mike Thelwall, Eds., Springer Handbook of Science and Technology Indicators. Cham, Switzerland: Springer Nature, 2019, 25-68. ISBN-13: 978-3030025106

## About the Global Research Report series from the Institute for Scientific Information (ISI)

**Our Global Research Reports draw on our unique industry insights to offer analysis, ideas and commentary to enlighten and stimulate debate.**

Each one demonstrates the huge potential of research data to inform management issues in research assessment and research policy and to accelerate development of the global research base.

If you would like to find out more about ISI, visit: webofsciencegroup.com/ISI

e: ISI@clarivate.com

**Previous reports include:**

Profiles, not metrics: beyond single point metrics

Research Integrity: Understanding our shared responsibility for a sustainable scholarly ecosystem.

Identifying Research Fronts in the Web of Science: From Metrics to Meaning

The Value of Bibliometric Databases

Multi-Authorship and Research Analytics

The 2020 Annual G20 Scorecard

Navigating the structure of Research on Sustainable Development Goals

## About Clarivate

Clarivate™ is a global leader in providing solutions to accelerate the lifecycle of innovation. Our bold mission is to help customers solve some of the world's most complex problems by providing actionable information and insights that reduce the time from new ideas to life-changing inventions in the areas of science and intellectual property. We help customers discover, protect and commercialize their inventions using our trusted subscription and technology-based solutions coupled with deep domain expertise. For more information, please visit clarivate.com.

**The Web of Science™**, part of Clarivate, organizes the world's research information to enable academia, corporations, publishers and governments to accelerate the pace of research. It is powered by the world's largest publisher-neutral citation index and research intelligence platform.

**webofsciencegroup.com/isi**



WS612750460 / 06