JISC cetis

centre for educational technology
& interoperability standards

Analytics Series

Vol.1, No. 4 Analytics for Understanding Research

By Mark van Harmelen
Hedtek Ltd

CETIS Analytics Series ISSN 2051-9214

Produced by CETIS for JISC

Analytics for Understanding Research

Mark van Harmelen
Hedtek Ltd

Table of Contents

1. Executive Summary

Analytics seeks to expose meaningful patterns in data. In this paper, we are concerned with analytics as applied to the process and outputs of research. The general aim is to help optimise research processes and deliver improved research results.

Analytics is the use of mathematical and algorithmic methods to describe part of the real world, reducing real-world complexity to a more easily understandable form. The users of analytics seek to use the outputs of analytics to better understand that part of the world; often to inform planning and decision-making processes. Applied to research, the aim of analytics is to aid in understanding research in order to better undertake processes of planning, development, support, enactment, assessment and management of research.

Analytics has had a relatively a long history in relation to research: the landmark development of citation-based analytics was approximately fifty years ago. Since then the field has developed considerably, both as a result of the development of new forms of analytics, and, recently, in response to new opportunities for analytics offered by the Web.

Exciting new forms of analytics are in development. These include methods to visualise research for comparison and planning purposes, new methods – altmetrics – that exploit information about the dissemination of research that may be extracted from the Web, and social network and semantic analysis. These methods offer to markedly broaden the application areas of analytics.

The view here is that the use of analytics to understand research is a given part of contemporaneous research, at researcher, research group, institution, national and international levels. Given the fundamental importance of assessment of research and the role that analytics may play, it is of paramount importance for the future of research to construct institutional and national assessment frameworks that use analytics appropriately.

Evidence-based impact agendas are increasingly permeating research, and adding extra impetus to the development and adoption of analytics. Analytics that are used for the assessment of impact are of concern to individual researchers, research groups, universities (and other institutions), cross-institutional groups, funding bodies and governments. UK universities are likely to increase their adoption of Current Research Information Systems (CRIS) that track and summarise data describing research within a university. At the same time, there is also discussion of increased 'professionalisation' of research management at an institutional level, which in part refers to increasing standardisation of the profession and its practices across institutions.

The impetus to assess research is, for these and other social, economic and organisational reasons, inevitable. In such a situation, reduction of research to 'easily understandable' numbers is attractive, and there is a consequent danger of over-reliance on analytic results without seeing the larger picture.

With an increased impetus to assess research, it seems likely that individual researchers, research groups, departments and universities will start to adopt practices of research reputation management.

However, the use of analytics to understand research is an area fraught with difficulties that include questions about the adequacy of proxies, validity of statistical methods, understanding of indicators and metrics obtained by analytics, and the practical use of those indicators and metrics in helping to develop, support, assess and manage research.

To use analytics effectively, one must at least understand some of these aspects of analytics, and certainly understand the limitations of different analytic approaches. Researchers, research managers and senior staff might benefit from analytics awareness and training events.

Various opportunities and attendant risks are discussed in section 5. The busy reader might care to read that section before (or instead of) any others.

2. Introduction

CETIS commissioned this paper to investigate and report on analytics within research and research management. The aim is to provide insight and knowledge for a general audience, including those in UK Higher Education

The paper is structured as follows. A general introduction to analytics is provided in this section. Section 3 describes four examples to impart a flavour of the uses of analytics. Section 4 contains a discussion of a four major ways of performing analytics. Section 5 contains concluding observations with an opportunity and risk analysis.

2.1 Analytics in the research domain

Analytics allows industry and academia to seek meaningful patterns in data, in ways that are pervasive, ubiquitous, automated and cost effective, and in forms that are easily digestible.

> Organizations such as Amazon, Harrah's, Capital One, and the Boston Red Sox have dominated their fields by deploying industrial-strength analytics across a wide variety of activities. [Davenport 2006]

A wide variety of analytic methods are already in use in research. These include bibliometrics (concerned with the analysis of citations), scientometrics ("concerned with the quantitative features and characteristics of science and scientific research" [Scientometrics 2012]), social network analysis (concerned with who works with whom), and research, to some extent, semantic approaches (concerned with domain knowledge).

Analytics is certainly important for UK research, and a national success story:

> The strength of UK universities and the wider knowledge base is a national asset. Our knowledge base is the most productive in the G8, with a depth and breadth of expertise across over 400 areas of distinctive research strength. The UK produces 14% of the most highly cited papers and our Higher Education Institutions generate over £3 billion in external income each year. [BIS 2011]

Notably, there is a place for analytics in UK research to help maintain and increase this success, for example through the identification of collaboration opportunities:

> The UK is among the world's top research nations, but its research base can only thrive if it engages with the best minds, organisations and facilities wherever they are placed in the world. A thriving research base is essential to maintain competitiveness and to bring benefit to the society and economy of the UK. [RCUK 2012a]

Looking forward, the pace of contemporary cultural, technological and environmental change seems certain to depend on research capacity and infrastructure. Consequently it is essential to seek greater effectiveness in the research sector. Recognising and exploiting the wealth of tacit knowledge and data in the sector through the use of analytics is one major hope for the future.

However, there are risks, and due care must be exercised: evidence from the research about analytics in other contexts combined with the research into academic research suggests that analytics-driven change offers significant opportunities but also substantial risks.

Research is a complex human activity, and analytics data – though often interesting – are hard to interpret and contextualise for maximal effect. There appear to be risks for the long-term future if current qualitative management practices are replaced by purely quantitative target-based management techniques.

2.2 The growth in research and the need for analytics

Research is growing rapidly, and with it, the need for analytics to help make sense of ever increasing volumes of data.

Using data from Elsevier's Scopus, The Royal Society [2011a] estimated that in 1999–2003 there were 5,493,483 publications globally and in 2004–2008 there were 7,330,334. Citations are increasing at a faster rate than publications; between 1999 and 2008 citations grew by 55%, and publications by 33%.

International research collaboration has increased significantly. For example Adams *et al* [2007] report on increases in collaboration across main disciplines in Australia, Canada, China, France, Germany, Japan, the UK and the USA. Between 1996-2000 and 2001-2005 increases by country varied from 30% for France to over 100% for China.

The World Intellectual Property Organisation [WIPO 2012] records that numbers of patents are increasing, in part because of the global growth in intellectual property, and in part because of strategic patenting activities; see figure 1.

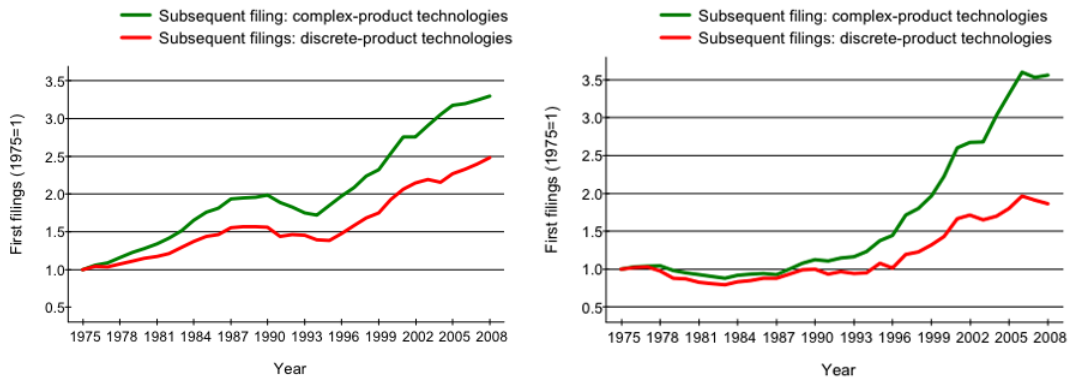

Figure 1: Growth in patent filings, on the left to initially protect intellectual property, and on the right, as part of strategic approaches to protection.

Meanwhile the impact agenda is becomingly increasingly important at levels varying from the impact of individual papers and individual researchers, though institutional impact, to impact at a national or international level. Responses include use of existing indicators and a search for new indicators: for example, the Global Innovation Index [GII 2012] and, in Europe, the development of a new innovation indicator by the Innovation Union Information and Intelligence System [IUIIS 2012].

The impact of the Web on research has been immense, enabling raw data, computational systems, research outputs and data about research to be globally distributed and readily available; albeit sometimes at financial cost. By making communication, data, data handling and analytic facilities readily available, the Web has been an enabler for the enactment of science. With this has come a vast increase in the availability of information about research. In turn, information about research and its enactment leads to further advances as it is analysed and exploited in diverse ways.

Yet despite the growing need for analytics to help make sense of research, we are still coming to terms with the validity (or not) of certain kinds of analytics and their use. Existing research provides a pool of potentially useful analytic techniques and metrics, each with different strengths and weaknesses. Applicability and interpretation of metrics may vary between fields even within the same organisational unit, and generalisation of results may not be possible across fields. It is widely acknowledged that different metrics have different advantages and disadvantages, and a former 'gold standard' of analytically derived impact, the Journal Impact Factor, is now debunked, at least for individual researcher evaluation. Further, the literature contains statistical critiques of some established metrics, and some newer metrics are still of unknown worth.

Inevitably, with future increases in the volume of research, analytics will play an increasing role in making sense of the research landscape and its finer-grained research activities. With new analytic techniques the areas of applicability of analytics will increase. However, there is a need to take great care in using analytics, not only to ensure that appropriate metrics are used, but also to ensure that metrics are used in sensible ways: for example, as only one part of an assessment for career progression, or as a carefully triangulated approach in developing national research programmes.

2.3 Quantitative study and the components of analytic solutions

Domains of interest that use analytics for quantitative study may be described thus:

- *Informetrics* – the quantitative study of all information. Informetrics includes
  - *Scientometrics* – the quantitative study of science and technology,
  - *Bibliometrics* – the quantitative study of scholarly information,
  - *Cybermetrics* – the quantitative study of electronic information, including
    - *Webometrics* – the quantitative study of the Web.
  - *Mathematical sociology* – the use of mathematics to model social phenomena.
  - *Social Network Analysis (SNA)* – the analysis of connections or social ties between researchers. Often seen as part of webometrics and mathematical sociology.
  - *Altmetrics* – a 'movement' concerned with "the creation and study of new metrics based on the Social Web for analyzing and informing scholarship" [Laloup 2011].

In fact, while this description is reasonable for the purposes of this paper, it is only a partial description of a complex field that has many interpretations: different disciplines and different funders tend to use different names for the same thing, and different researchers may structure the 'sub-disciplines' of informetrics differently. For example SNA may be considered part of mathematical sociology while elsewhere it may be viewed as part of cybermetrics or webometrics. Generalising further, there are four major components to an analytic solution. These are shown in figure 2.

Examining the layers in figure 2, we see:

- Applications of analytics in the real world. As examples, assessment of the impact of a funding programme, use of an evidence base to set science policy, discovery of potential collaborators.
- Visualisation of the results of analysis, allowing users of analytic results to perceive and understand analytic results in order to act on them.
- Methods, the algorithmic means of analysis of raw data and the approaches, science, statistics and mathematics behind those algorithms. In this paper there is a gross classification of methods into four sometimes overlapping sub-categories:
  - Metrics, which are computational methods of diverse kinds, for example, acting over bibliometric data.
  - Methods based on the analysis of statistics about the use of resources – this is a sufficiently homogeneous and distinct set of methods so as to be described separately from metrics.
  - Social Network Analysis, the analysis of links between people, in this case, researchers.
  - Semantic methods, a growing set of methods that concentrate, inter alia, on the assignment of meaning to data.
- Data: The raw materials for analytics, for example, data about publications, data about funders, grants and grant holders, data that is the output of research activities, and so on.
- Technological infrastructure: The computational infrastructure needed to realise an analytic approach.

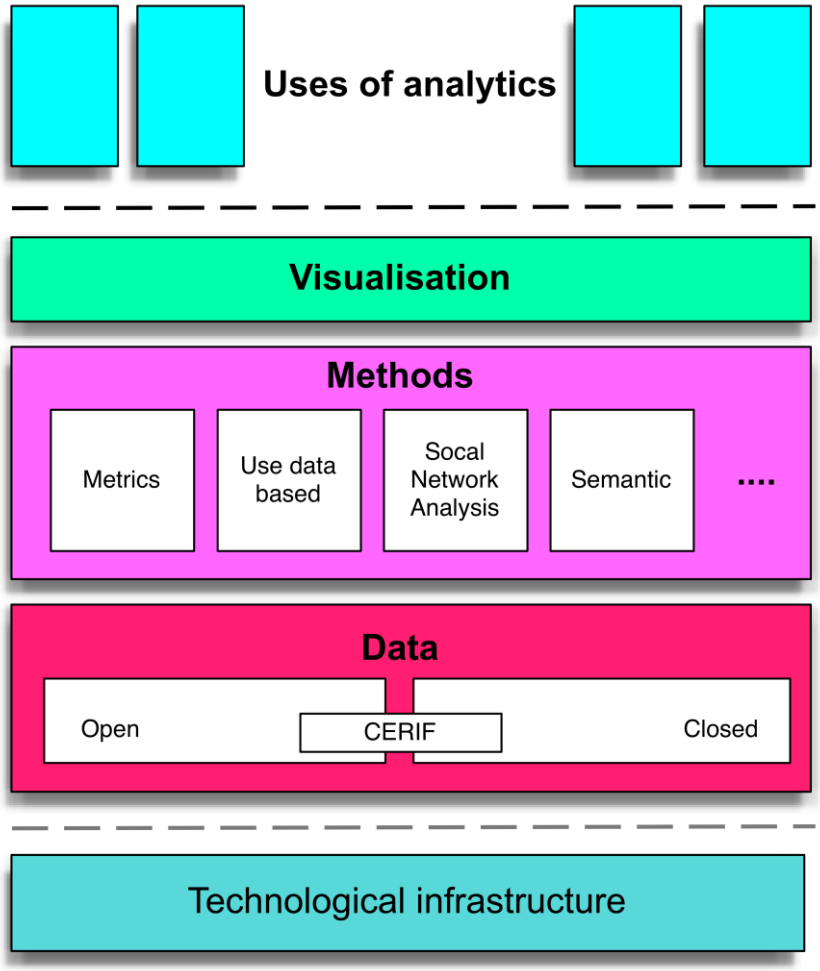

Figure 2: Analytics solutions

The focus of this paper is largely on applications and methods, though, in passing, attention is paid to visualisation and data.

This paper does not consider the form of technological solutions, except to say here that the bottom four layers (or 'the technology stack') shown in figure 2 may have considerable variation, and may be implemented on desktop machines and/or servers, and have user interfaces provided by desktop applications, Web browsers and/or mobile apps.

2.4 Problems in the use of analytics

It is well to note impediments to use of analytics: difficulty in interpretation, inability to generalise results across fields and privacy issues. In dealing analytics is it wise to consider the kinds of impact of research, since impact has many meanings that in turn affect the use of analytics.

While researchers have been interested in the process of research, and have used quantitative techniques to gather evidence about research for many years, interpretation remains challenging even for those with an academic interest in analytics. The literature is rich, complex and highly technical.

Lay users of analytics include (i) researchers in fields other than analytics research and (ii) research managers, including university senior staff with a responsibility for research. With lack of technical knowledge of analytics by lay users several important kinds of questions emerge:

- Are the lay users using suitable metrics for their task in hand?
- Are lay users interpreting suitable metrics appropriately?
- Are suitable metrics being used appropriately within the institution?
- Are metrics that are being used for assessment part of a broader approach that incorporates other non-metric-based indicators?

Lay users may therefore need tools that hide the 'nuts and bolts' of an analytic approach and ease interpretation given that the tools are optimal for the task in hand. Research maps (section 2.1) are interesting in this respect given their robustness in respect of the source of data used to construct the maps and statistical methods used in construction of the maps. Current Research Information Systems (section 2.6) may need to be approached with an initial exploration of the metrics they provide from the points of view of suitability for task, interpretation and effects on the institution.

Extrapolating from and generalising results in one area may be difficult. De Bellis [2009] notes:

> it would be fundamentally wrong to draw general conclusions from experiments performed in such a wide variety of disciplinary and institutional settings; empirical evidence, once and for all, is not additive.

That is not to say that generalising is impossible in all cases. Rather, generalising should be approached with caution. Even the use of metrics for comparison may be difficult. For example, relatively well-accepted and popular (albeit sometimes contentious) bibliographic metrics such as Hirsch's h-index [Hirsch 2005] are sensitive to different citation rates for individual papers in different disciplines. Scaling has been proposed as a solution [Iglesias and Pecharromán 2007].

Use of technology for analytics may lead to privacy issues. Other papers in the CETIS Analytics Series address these issues and these are not considered further here.

Inevitably, a discussion of analytics leads to a discussion of quality and impact.

In the interests of brevity, questions of what constitutes research quality are totally omitted from this paper. However, since analytics is so often directed at questions of impact, impact is considered in this report. Analytically derived metrics are often considered to be able to measure impact, but as we will see in section 4.1, available measures have some flaws in this respect, and are only recommended as part of a broader approach.

The UK's Research Excellence Framework (REF) defines impact (for the purposes of the REF) as

> an effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life ….
>
> Impact includes, but is not limited to, an effect on, change or benefit to:
>
> - the activity, attitude, awareness, behaviour, capacity, opportunity, performance, policy, practice, process or understanding
> - of an audience, beneficiary, community, constituency, organisation or individuals
> - in any geographic location whether locally, regionally, nationally or internationally.
>
> Impact includes the reduction or prevention of harm, risk, cost or other negative effects. [REF 2011, Annex C, paragraphs 4-6]

Definitions of some kinds of impact are offered by Research Councils UK:

> Academic impact: The demonstrable contribution that excellent research makes to academic advances, across and within disciplines, including significant advances in understanding, methods, theory and application. …
>
> Economic and societal impacts: The demonstrable contribution that excellent research makes to society and the economy. Economic and societal impacts embrace all the extremely diverse ways in which research-related knowledge and skills benefit individuals, organisations and nations by:
>
> - fostering global economic performance, and specifically the economic competitiveness of the United Kingdom,
> - increasing the effectiveness of public services and policy,
> - enhancing quality of life, health and creative output.
>
> [RCUK 2012b]

More generally "a research impact is a recorded or otherwise auditable occasion of influence from academic research on another actor or organization" [LSE 2011].

Besides the REF's characterisation of impact in "economy, society, culture, public policy or services, health, the environment or quality of life", and RCUK's definition of academic, economic and social impact, impact may have other meanings, including scholarly impact, educational impact and epistemological impact. Given this range it is always important to be specific about the kind of impact being discussed or 'measured'.

With increased use of new channels for dissemination of research and scientific information, there may be attendant difficulties in measurement and interpretation of impact that leverage available statistics. For example, issues around the measurement of impact of a YouTube chemistry channel, *The Periodic Table of Videos*, are discussed in [Haran and Poliakoff 2011].

Impact assessment may be applied to different actors: individual researchers, research groups, departments, universities and other research-active organisations. Impact assessment may also be applied to other entities, for example, individual journal papers, and journals themselves.

Difficulties may arise in what is chosen to indicate particular types of impact. For example, some proxies for different kinds of impact are article level citation counts, co-citations, patents granted, research grants obtained, and download counts. These may or may not be useful in helping to indicate a particular kind of impact.

The methods used to derive metrics of impact are themselves subject to considerable discussion that is often of a detailed statistical nature. As just one example, Leydesdorff and Opthof [2010] critique Scopus's original version of the Source Normal Impact per Paper (SNIP) metric.[1]

The choice of proxies is as important as the analytic methods in use to produce indicators. These, their appropriateness, their use, and the institutional or national impact of their application need to be carefully considered by experts and their advantages and disadvantages need to be understood by lay users of analytics.

[1] A second version of SNIP, SNIP 2, is now available via Scopus [Scopus 2012].

3. Examples

By way of providing an introduction to and flavour of analytics as applied in research and research management, various examples are provided as illustration. Here we describe six diverse examples:

- The use of overlay science maps to enable comparisons of research activities.
- Assessment of the impact of research on a national scale.
- Identification of collaboration opportunities.
- Research planning and management facilities.
- Ways in which researchers may improve their own impact.

3.1 Science maps

Maps are appealing visualisation aids: they provide a visual representation that draws on spatiality and meshes well with our abilities to interpret spatial data.

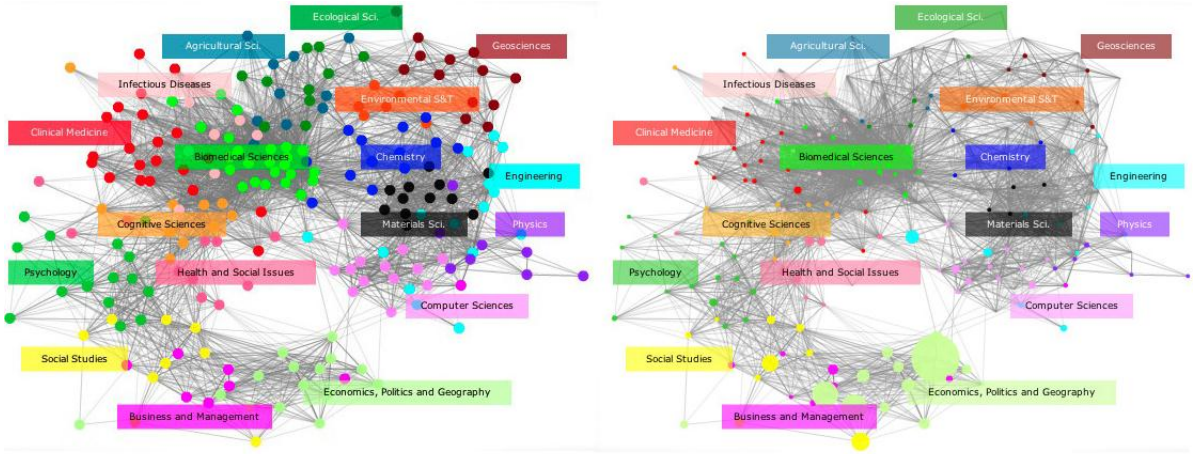

Figure 3: Left: Map of science, labelled with major disciplines, dots represent sub-disciplines. Right: Overlaid with research at LSE, where dots represent areas of research activity as revealed by publication activity. Size of dots is representative of numbers of publications in that discipline. Maps are screenshots from the interactive mapping facilities at http://idr.gatech.edu/maps

Science maps were developed in the 70s. However, maps covering all of science are a more recent development. A global science map provides a spatial representation of research across science; areas on the map represent disciplines. Overlay research maps [Rafols *et al* 2010] incorporate a mapping of data about research activities onto a science map; see figures 3-5. One advantage to the overlay approach here is that the underlying analysis is hidden from the casual user, and allows for easy interpretation by non-specialists.

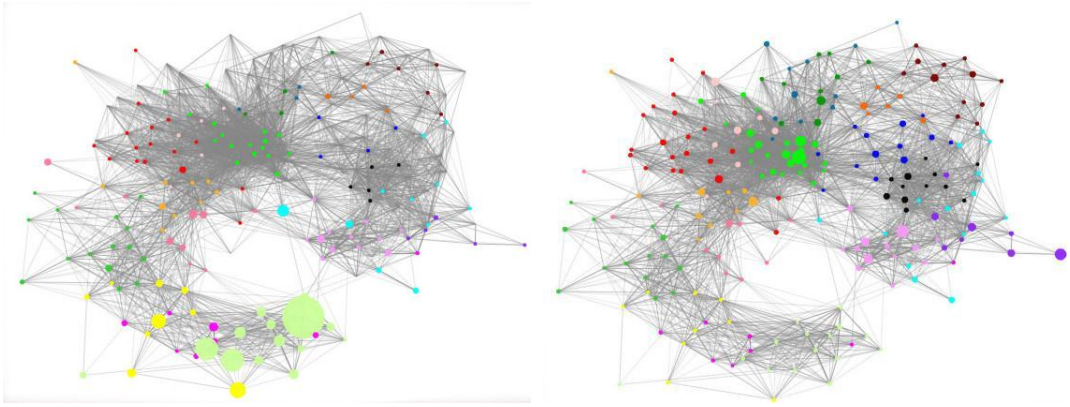

Figure 4: Comparison of research activities.
Left: Research at the LSE (without labels for comparative purposes).
Right: Research at the University of Edinburgh.
Maps are screenshots from the interactive mapping facilities at http://idr.gatech.edu/maps

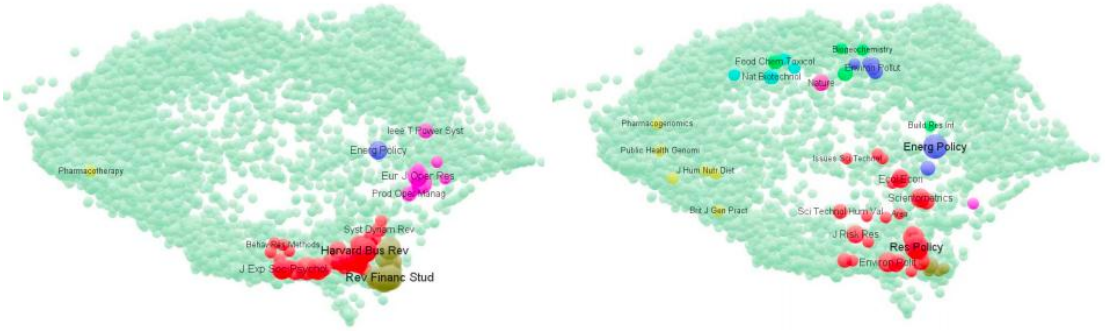

Figure 5: Comparison of 2006-2009 journal publication portfolios
Left: The London Business School
Right: The Science and Technology Policy Research Unit SPRU at the University of Sussex [Leydesdorff and Rafols 2011].

Research maps may be constructed from different sources, most often (as here) from bibliographic databases (PubMed, Web of Science, Scopus), or from other sources, for example, hybrid text/citation clustering or click-streams generated by users of journal websites. Maps tend to consistency in structure. Advantageously, this consistency and a standardised map layout allow for the presentation of easily comparable data for non-specialist research managers and policy makers.

Rafols *et al* [2010] describe uses in comparing the publishing profiles of research organisations, disciplinary differences between nations, publication outcomes between funding agencies, and degrees of interdisciplinarity at the laboratory level. Further successes include mapping the diffusion of topics

across disciplines and the exploration of emerging technologies: for example maps depicting the development of nanotechnologies[2], see also Leydesdorff's interactive version[3].

In fact, Rafols *et al* [2010] claim that research maps provide a richer basis for analytics than other unidimensional bibliometric techniques such as rankings:

> In our opinion, scientometric tools remain error-prone representations and fair use can only be defined reflexively. Maps, however, allow for more interpretative flexibility than rankings. By specifying the basis, limits, opportunities and pitfalls of the global and overlay maps of science we try to avoid the widespread problems that have beset the policy and management (mis-)use of bibliometric indicators such as the impact factor.

They conclude:

> In our opinion, overlay maps provide significant advantages in the readability and contextualisation of disciplinary data and in the interpretation of cognitive diversity. As it is the case with maps in general, overlays are more helpful than indicators to accommodate reflexive scrutiny and plural perspectives. Given the potential benefits of using overlay maps for research policy, we provide the reader with an interactive webpage to explore overlays (http://idr.gatech.edu/maps) and a freeware-based toolkit (available at http://www.leydesdorff.net/overlaytoolkit).

To summarise, key insights are that academic advances are making available new quantitative techniques whose results are more accessible to a wider audience with fewer risks than previously.

3.2 Assessment of impact at national level

Assessment of economic impact at a national level is being considered by Science and Technology for America's Reinvestment: Measuring the Effect of Research on Innovation, Competitiveness and Science (STAR METRICS) programme [OSTP 2012a].

STAR METRICS was established in 2010 to provide evidence about the effects of the 2009 American Recovery and Reinvestment Act as an economic stimulus:

> It is essential to document with solid evidence the returns our Nation is obtaining from its investment in research and development. STAR METRICS is an important element of doing just that.
>
> *John P. Holdren, Assistant to the President for Science and Technology and Director of the White House Office of Science and Technology Policy, June 1, 2010*
> [NIH 2010]

[2] http://tinyurl.com/nanomaps

[3] http://tinyurl.com/nanointeract

This vision has expanded, the programme is now viewed as key in contributing a better empirical basis for US science policy decisions. The 'theoretical' underpinnings of the service include the idea that there can and should be a Science of Science Policy [OSTP 2012b] and that a data infrastructure should be built to enable that to be pursued [Largent and Lane 2012]. As such, STAR METRICS aims to "monitor the impact of US government science investments on employment, knowledge generation, and health outcomes" [NIH 2010].

The programme has two stages of engagement with universities and other institutions that perform research:

- Phase 1: Measuring jobs saved or created by Federal investment in the US economy. This is analytically trivial, but more challenging computationally, since the intent is to relieve institutions of manual reporting, by interfacing directly to financial systems to obtain automated returns of staff names against research budget codes.
- Phase 2: Expanding tracking to all research awards, measuring the impact in terms of publications, patents, and spinoff companies, with the intent of increasing "the evidence base … on the cumulative impact of science investments on the nation's R&D work force and the role of science in [global] competitiveness." [Holdren 2009]
- In Phase 2 four metrics are of particular interest:
  - Economic growth, measured through indicators such as patents and business start-ups.
  - Workforce outcomes, measured by student mobility into the workforce and other employment data.
  - Scientific knowledge, measured through publications and citations.
  - Social outcomes, measured by long-term health and environmental impacts.

What is exciting about STAR MERICS is not the level or sophistication of the analytics being employed, but rather that STAR METRICS is providing a national platform for the collection of evidence for research analytics to inform national fiscal policy and the development of Science of Science Policy.

In considering platforms of this kind, a careful approach should be adopted in at least three respects: use of appropriate analytics that reliably inform lay users; adoption and promulgation of a strong awareness of the limitations in approach and analytics chosen, for example, that analytics may under-represent the future impact and benefit of blue skies research that proves successful in the long-term.

A key insight is that creating a similar platform for UK research at a national level, populated with openly available data, may bring broad benefits including improved research efficiency, an evidence base to inform investment into research and, possibly, a richer public understanding of research and the role it plays in economic and societal development.

3.3 Identifying collaboration opportunities

With some discipline-specific exceptions, collaborative research is increasingly important in a world of multidisciplinary research where many scientists work on problems that cross discipline boundaries.

> Science in general, and biomedical research in particular, is becoming more collaborative. As a result, collaboration with the right individuals, teams and institutions is increasingly crucial for scientific progress. [Schleyer *et al* 2012]

An interesting animation[4] of the growth in international research is provided by the Royal Society [2011b]. As an aside, there are some problems with the data pointed out on the referenced page, providing an example of the value of informed and contextualised interpretation of the output of analytics.

Katz and Martin [1997] provide a survey of research collaboration, pointing to different kinds of collaboration: between individuals, institutions and countries. Adams *et al* [2007] provides some of the drivers for research collaboration:

> International research collaboration is a rapidly growing component of core research activity for all countries. It is driven by a consonance between top-down and bottom-up objectives. Collaboration is encouraged at a policy level because it provides access to a wider range of facilities and resources. It enables researchers to participate in networks of cutting-edge and innovative activity. For researchers, collaboration provides opportunities to move further and faster by working with other leading people in their field. It is therefore unsurprising that collaborative research is also identified as contributing to some of the highest impact activity.

In 2011 the Royal Society reports a contemporary researcher-driven impetus to collaborate:

> Collaboration is increasing for a variety of reasons. Enabling factors such as advances in communication technology and cheaper travel have played a part, but the primary driver of most collaboration is individual scientists. In seeking to work with the best of their peers and to gain access to complementary resources, equipment and knowledge, researchers fundamentally enhance the quality and improve the efficiency of their work.
> [Royal Society 2011a].

Katz and Martin [1997] discuss difficulties in measuring collaboration, particularly through the most common mechanism, co-authorship of scientific papers. Here they point to the many different reasons as to why co-authors appear as authors of papers, for example because of significant and meaningful collaborative effort, because of minor contribution, because they may have secured funding, but not have performed research, or for various "social" reasons.

Nonetheless, failing other more accurate metrics, co-authored paper counts are widely used as an indicator of collaborative research activity.

With an increase in collaboration, there comes an attendant need for assistance in finding collaborators. It may be that researcher match-making proves to be its greatest worth for multi-disciplinary research. In

[4] http://tiny.url /collabgrowth

that setting researchers seeking collaborators in a related but unknown field will most appreciate help in finding potential collaborators and well-connected 'hub' individuals.

One approach to finding collaborators is by means of Research Networking Systems, which

> …are systems which support individual researchers' efforts to form and maintain optimal collaborative relationships for conducting productive research within a specific context. [Schleyer *et al* 2012]

As a simple solution, database technologies may be used to build Research Network Systems to supply data on researcher interests. However one UK university known to the author is having difficulties with a simple database approach, mostly because of mismatches between descriptors of similar or overlapping research interests.

Because of the standardised nature of Medical Subject Headings (MeSH), these mismatches were eliminated in the Faculty Research Interests Project (FRIP) at the University of Pittsburgh, and its more recent development into Digital Vita [Schleyer *et al* 2012] [Schleyer private communication 2012]. Entries for 1,800+ researchers in the university's Health Sciences Center are described using MeSH, researchers' MeSH descriptors having been automatically extracted from researcher publications whose details are available in PubMed. PubMed contains "more than 22 million citations for biomedical literature from MEDLINE, life science journals, and online books" [PubMed 2012].

There are many other approaches to analytics for the identification of collaboration opportunities. For example, Stephens [2007] discusses using commercial products utilising semantic technologies and graph matching to perform network analysis to identify well-connected 'hub' individuals. This could be used in research to identify who to approach in the process of finding research collaborators. As an example from within UK HE, Liu *et al* [2005] describe the use of semantic technology to integrate expertise indications from multiple institutional sources to find collaborators and expertise within the University of Leeds.

A future where researchers and their support staff are assisted by smart software agents capable of highlighting a short list of good matches discovered from a massive quantity of prospects sounds attractive.

3.4 Research planning and management

Research planning can be carried out at many levels, from individual projects to national and international scales.

In the process of research planning, the end game must be to pick successful areas for research. Unfortunately, ultimate success cannot, in general, be predicted, and analytics may be of little leverage in such endeavours. For example Rafols *et al* [2010] state that

> bibliometrics cannot provide definite, 'closed' answers to science policy questions (such as 'picking the winners')

However at a less ambitious level than picking winners, analytics may be able to help in aspects of research planning, management and administration.

Discussion now turns to the use of analytics in helping manage current research within an institution. Bittner and Müller [2011] provide references to approaches in various European countries. This paper focuses on institutional Current Research Information Systems (CRIS), described by Russell [2012] as

> a system which supports the collection and management of a range of research information derived from various sources, together with reporting and analysis functions

Bittner and Müller [2011] provide summary information about CRIS stakeholders and users:

> Research information systems, also referred to as Current Research Information Systems (CRIS) ... are software tools used by the various actors in the research process. Their uses are manifold, ranging from the documentation of research projects and their results over easing the management of research to simplifying research assessment. In doing so, various parties are involved in the application of research information systems, including funding agencies, assessment bodies, policy makers, research institutions, and researchers themselves.

In the UK there is particular emphasis on CRIS to support the Common European Research Information Format (CERIF) standard data model developed by the European Organisation For International Research Information.

Russell [2012, and private communication 2012] reports a trend towards adoption of CERIF-based CRIS amongst UK institutions, and that adoption is actively encouraged by JISC. CERIF was not used before 2009, uptake started to increase in 2010 and expanded rapidly in 2011. Incremental improvements continue to be made to the standard, for example [MICE 2011]. As of March 2011, 51 institutions in the UK (30.7% of UK HEIs) had adopted a CERIF-based CRIS system, no doubt reflecting the view that CRIS

> is becoming a crucial tool for providing management and strategic data and reporting. [Russell 2012]

The trends reported by Russell [2012] suggest that the UK CRIS ecosystem will be dominated by a small number of commercial vendors; possibly even a single vendor. This may or may not be a negative influence, depending on the extent that a single dominant CRIS determines a particular kind of interpretation of the data captured by the system.

It may therefore be valuable to contrast the properties of closed ecosystems with more open approaches, where the research and management community can build their own analytics systems to exploit and interpret data as they desire, rather than potentially being given 'a view' by a closed proprietary system.

3.5 Research reputation management

The final example is rather different to the above, concentrating not on analytics *per se*, but rather on an approach to maximising (measures of) individual researcher impact. This is research reputation management.

We concentrate on the LSE's Impact of Social Sciences Blog [LSE 2011] and the LSE's Public Policy Group's handbook *Maximising the Impacts of your Social Research: A Handbook for Social Scientists* [LSE PPG 2011a].

As the blog post accompanying the handbook points out, there has previously been little in the way of advice to academics to maximise their impact:

> For the past year a team of academics based at the London School of Economics, the University of Leeds and Imperial College have been working on a 'Research Impacts' project aimed at developing precise methods for measuring and evaluating the impact of research in the public sphere. We believe our data will be of interest to all UK universities to better capture and track the impacts of their social science research and applications work.
>
> Part of our task is to develop guidance for colleagues interested in this field. In the past, there has been no one source of systematic advice on how to maximize the academic impacts of your research in terms of citations and other measures of influence. And almost no sources at all have helped researchers to achieve greater visibility and impacts with audiences outside the university. Instead researchers have had to rely on informal knowledge and picking up random hints and tips here and there from colleagues, and from their own personal experience. [LSE PPG, 2011b]

The handbook discusses citations, becoming better cited, external research impact, and achieving greater impact. It forms an invaluable resource, particularly when coupled with free to use tools like Google Scholar and Harzing's Publish or Perish.

Two conclusions spring out from the blog. Firstly, that measurement and assessment of impact should be performed in a domain-specific context; for example, there are much higher citation counts prevalent as the norm in science subjects compared to humanities subjects. Secondly, that scholarly impact as measured by citation counts is only one form of impact and that there may be many more forms of impact to measure.

Thus, in a guest post on the blog, Puustinen and Edwards [2012] provide an example of academics measuring their own impact in a variety of different ways:

> Who gives a tweet? After 24 hours and 860 downloads, we think quite a few actually do

Puustinen and Edwards go on to explain aspects of their approach to impact:

> With the impact agenda in everyone's mind but with no consensus on how best to demonstrate the impact of research, at NCRM [the National Centre for Research Methods at the University of Southampton] we have set Key Performance Indicators for the website,

> in addition to monitoring the performance of some of the print materials via print-specific website addresses and QR codes. By making sure that not only do we publicise NCRM research, but also are able to track the effectiveness of those publicity activities, we trust that we will be in a good position to demonstrate the short and long-term impacts of our research.

One key insight here is that researchers (and particularly beginning researchers) may be under-informed as to how to maximise the impact of their work and manage their research reputation.

A second key insight is that indicators of impact may be multifarious and changing as new Web technologies arise; see altmetrics in section 4.1 for newer web-enabled metrics.

The final insights are that researchers need tools to measure impact, and that researchers' interests may be best served if the researchers collect all potential indicators.

4. Methods

We define methods as groups of similar kinds of computational techniques, and we divide them into four broad classes of techniques:

- *Metrics*, where we confine ourselves to techniques that are specifically concerned with the impact of journals, articles and authors.
- *Analysis of use data*, where traces of the use of resources by end users are subsequently used to enhance discovery, recommendation and resource management.
- *Social Network Analysis*, concerned with links between people and/or institutions and the uses to which this data may be put.
- *Semantic methods*, concerned with the assignment of and inference of meaning.

Three notes are in order:

- In places, there is overlap between these methods.
- Further, there is no claim to completeness here; the field is vast.
- As indicated in section 2.3, methods do not stand alone from technology, data and visualisation; all the methods discussed need these complementary components to be usable.

4.1 Metrics

We examine metrics that aim to measure impact:

- *Bibliometrics and citation analysis*, to assess the impact of journals, articles and authors.
- *Webometrics*, which uses statistics derived from the web and web-based hyperlinks to provide reputation analyses of web sites and institutions which have a presence on the web.
- *Altmetrics*, which uses, largely, Web-based social media to assess the impact of published works in a more immediate way than bibliometrics and citation analysis.

Where scholarly impact is concerned, three kinds of impact are of interest:

- *Journal level impact* is used as a proxy for a journal's importance, influence and impact in its field relative to other journals in the field.
- *Article level impact* provides scholarly impact for individual articles, regardless of where they may be published. Article level impact is sometimes referred to as Article Level Metrics.
- *Individual researcher impact* is often interpreted as scholarly impact. Individual researcher impact may be based on raw citation counts, or may use metrics such as the h-index [Hirsch, 2005] or the g-index [Egghe 2006a].

Bibliometrics and citation analysis

Bibliometric analysis based on citation data is of central interest in understanding research. De Bellis [2009] points to the importance of citations in analysing research activity:

> bibliographic citations are quite unique … , because the connections they establish between documents are the operation of the scientists themselves in the process of exposing (and propagandizing) their findings to the community of their peers.

Often, bibliographic citation analysis is concerned with the assessment of scholarly impact, but there are other uses, as indicated by Kuhn, in his *Structure of Scientific Revolutions* [1962], namely, citations as a potential indicator of revolution in scientific paradigms:

> ...if I am right that each scientific revolution alters the historical perspective of the community that experiences it, then that change of perspective should affect the structure of post-revolutionary textbooks and research publications. One such effect – a shift in the distribution of the technical literature cited in the footnotes to research reports – ought to be studied as a possible index to the occurrence of revolutions. [Kuhn 1962]

While the author is not aware of specific predictive use, citation based analysis has been used for *post hoc* analysis. For example, Garfield *et al* [1964] graphed the network of paper and citation linkages that lead to the discovery of the structure of DNA (see also later graphing using HistCite [Scimaps 2012a] [TR 2012c]) and mapping the development of nanotechnologies[5] [Leydesdorff and Schank 2008] [Scimaps 2012b].

However, in this section we are more interested in more traditional use of citations in bibliometrics to assess scholarly impact. Pioneered by Garfield in 1955 [Garfield 1955], subsequent advances included bibliographic coupling [Kessler 1963], document co-citation [Small 1973], and author co-citation analysis [White and Griffith 1981].

As indicated above, broad categories of citation-based bibliometric analysis have emerged, notably journal impact factors and, in response to criticism of journal impact factors, article level metrics. Some indication of the large variety of metrics is supplied by, for example, [Okubo 1997], [Rehn *et al* 2007] and [Rehn and Kronman 2008].

There are several citation analysis systems available, both freely and on a commercial basis. Free systems include CiteSeer[6], Google Scholar[7] and Publish or Perish[8]. Commercial offerings include

[5] See also the animation at http://tiyurl.com/nanoani

[6] http://citeseer.ist.psu.edu/

[7] http://scholar.google.com/

[8] http://www.harzing.com/pop.htm

THOMSON REUTERS Journal Citation Reports[9], InCites[10], Web of Science[11], and Elsevier's SciVerse tools[12], including Scopus[13].

However, depending on use, bibliographic metrics can be controversial. Eugene Garfield, to whom we can attribute the genesis of modern citation-based bibliometric metrics, notes:

> I first mentioned the idea of an impact factor in 1955. At that time it did not occur to me that it would one day become the subject of widespread controversy. Like nuclear energy, the impact factor has become a mixed blessing. I expected that it would be used constructively while recognizing that in the wrong hands it might be abused.
> …
> The use of the term "impact factor" has gradually evolved, especially in Europe, to include both journal and author impact. This ambiguity often causes problems. It is one thing to use impact factors to compare journals and quite another to use them to compare authors. [Garfield 1999]

Much attention has been paid to the shortcomings of journal impact factors, be those problems associated with the reliability of the metrics, or with the use of the metrics, particularly to measure the scholarly impact of individual researchers. Given the importance of metrics measuring scholarly impact, it is worth investigating journal impact factors and the THOMSON REUTERS Journal Impact Factor (JIF) in a little more depth.

Generally, journal impact factors similar to the JIF measure the current years' citation count for an 'average' paper published in the journal during the preceding *n* years. Normalisation is applied by dividing this count by the number of citable items published in the preceding *n* years. For the JIF, *n* is two. See [TR 2012a] for some further discussion of algorithms to derive journal impact factors of this nature.

The JIF [TR 2012a], which is published in *Journal Citation Reports* [TR 2012b], is widely used but also widely critiqued, for example:

> The JIF has achieved a dominant position among metrics of scientific impact for two reasons. First, it is published as part of a well-known, commonly available citation database (Thomson Scientific's JCR). Second, it has a simple and intuitive definition. The JIF is now commonly used to measure the impact of journals and by extension the impact of the articles they have published, and by even further extension the authors of these articles, their departments, their universities and even entire countries. However, the JIF has a

[9] http://thomsonreuters.com/products_services/science/science_products/a-z/journal_citation_reports/

[10] http://researchanalytics.thomsonreuters.com/incites/

[11] http://apps.webofknowledge.com/

[12] http://www.info.sciverse.com/

[13] http://www.info.sciverse.com/scopus

> number of undesirable properties which have been extensively discussed in the literature [references]. This had led to a situation in which most experts agree that the JIF is a far from perfect measure of scientific impact but it is still generally used because of the lack of accepted alternatives. [Bollen *et al* 2009]

Criticism of the JIF includes, for example, variation of the impact of individual articles in the same journal (leading to criticism of the use of journal impact factors for individual researcher impact assessment) and documented cases of editors and/or publishers manipulating journal impact factors by insisting that authors reference publications in a journal before acceptance of their papers in that journal. Interested readers are referred, in the first instance, to the overview of criticisms that appears in Wikipedia [2012a]. The shortcomings are also well documented in the literature: for example [Lozano et al 2012] provides a summary of criticism of journal impact factors as a measure of the impact of researchers.

Measures of the impact of single papers (rather than journal impact factors) are now generally viewed as more indicative of the impact of individual publications. Article Level Metrics (ALM) are being increasingly adopted: for example, The Public Library of Science (PLOS), an influential collection of seven peer-reviewed and web-published Open Access journals, initiated its ALM programme in 2009 [PLOS 2012]. PLOS provides article-level citation data extracted from four third-party citation extraction services (Scopus, Web of Science, PubMed Central, and CrossRef) for each article published in a PLOS journal.

Even despite widespread knowledge of the deficiencies of journal impact factors, the Times Higher Education [Jump 2012] reports what appears to be inappropriate use in the UK earlier this year. According to the report, Queen Mary, University of London (QMUL) used a journal impact factor, together with research income data, to decide on redundancies. This led to reiteration of criticism of journal impact factors in the context of the redundancies, for example collected at [QMUL UCU 2012] and in the comments to [Gaskell 2012]. In response, University and College Union members planned strike action at the use of “crude measures” in deciding redundancies. Despite the criticism of QMUL, it appeared that the University would institute a “new performance assessment regime for academics across the Faculty of Science and Engineering that is based on similar metrics to the redundancy programme” [Jump 2012].

However, assessment of individual researcher impact has generally moved on from assessment by journal impact factor of the journals that researchers publish in.

The h-index [Hirsch 2005] is a very popular metric to assess individual researcher impact. A researcher’s h-index *h* is the (highest) number of papers that a researcher has written that have each been cited at least *h* times. Hirsch claims that the index is representative, in one number, of value of a researcher’s contribution, the diversity of that contribution, and how sustained the contribution has been over time.

The h-index is undoubtedly an easy-to-understand metric, but it does hide information. Imagine researcher A has twenty papers each cited at twenty times, and five more papers each cited five times, whereas researcher B has twenty papers cited twenty times, and fifty papers each cited nineteen times. One might think that B has had a greater scholarly impact than A, but if assessed using the h-index alone, each researcher has an equal h-index, of twenty. Other criticisms, including the h-index favouring

longer-established researchers with greater publication counts, are usefully summarised in Wikipedia [Wikipedia 2012b].

Hirsch [2005] provides some caveats to the use of the h-index. Two of these is are notable in the context of the recommendation at the end of this sub-section:

> Obviously, a single number can never give more than a rough approximation to an individual's multifaceted profile, and many other factors should be considered in combination in evaluating an individual. Furthermore, the fact that there can always be exceptions to rules should be kept in mind, especially in life-changing decisions such as the granting or denying of tenure. [Hirsch 2005].

The g-index [Egghe 2006a] gives a higher weight to highly cited articles, addressing one criticism of the h-index, that it does not take account of highly cited papers well. A researcher's g-index *g* is the (highest) number of the researcher's papers that each received g^2 or more citations. Egghe provides an comparison between himself and Small (an important earlier researcher in bibliometrics) where with similar h-indices, the g-index reveals a difference based on more highly cited papers, where Egghe's g-index is 19 and Small's is 39 [Egghe 2006b]

Other metrics include the e-index, to compare researchers in different fields with different citation rates, and metrics that account for the age of a paper, in effect giving higher weightings to more recent papers' citation counts. Simpler metrics are sometimes used, number of publications, number of citations, average number of publications per year, and average number of citations per year.

Recent work by Wagner and Leydesdorff [2012] which proposes a new "second generation" Integrated Impact Indicator (I3) that addresses normalisation of citation counts and other common concerns.

A comprehensive selection of metrics is supplied by Harzing's Publish or Perish software, and ReaderMeter.org provides statistics built on Open Data supplied by Mendeley via its API[14] [Henning 2012].

[14] As a testament to how Open Data about research is being reused widely: "Imagine the rich ecosystem of third-party Facebook and Twitter apps, now emerging in the domain of science. More than 240 applications for research collaboration, measurement, visualization, semantic markup, and discovery – all of which have been developed in the past year – receive a constant flow of data from Mendeley. Today, Mendeley announced that the number of queries to its database (termed "API calls") from those external applications had surpassed 100 million per month." [Henning 2012]. This is all the more remarkable, sine the API has only existed for seventeen months.

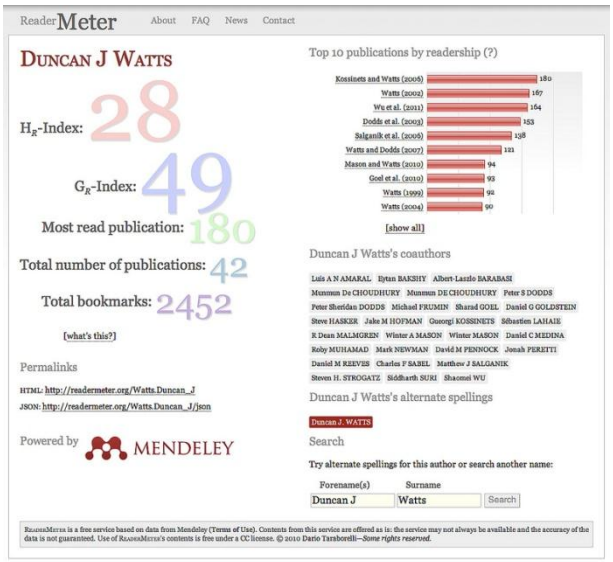

Figure 6: ReaderMeter.org display

http://www.flickr.com/photos/mendeley/7839392988/in/set-72157631195319638/

Three international mathematics associations (the International Mathematical Union in cooperation with the International Council on Industrial and Applied Mathematics and the Institute of Mathematical Statistics) provide a report on the use of quantitative assessment of research. In a press release about the report, the associations warn against use of citation-based bibliometrics as the sole indicator of research quality:

> The report is written from a mathematical perspective and strongly cautions against the over-reliance on citation statistics such as the impact factor and h-index. These are often promoted because of the belief in their accuracy, objectivity, and simplicity, but these beliefs are unfounded.
>
> Among the report's key findings:
>
> - Statistics are not more accurate when they are improperly used; statistics can mislead when they are misused or misunderstood.
> - The objectivity of citations is illusory because the meaning of citations is not well-understood. A citation's meaning can be very far from "impact".
> - While having a single number to judge quality is indeed simple, it can lead to a shallow understanding of something as complicated as research. Numbers are not inherently superior to sound judgments.
>
> The report promotes the sensible use of citation statistics in evaluating research and points out several common misuses. It is written by mathematical scientists about a widespread application of mathematics. While the authors of the report recognize that assessment must be practical and that easily-derived citation statistics will be part of the process, they caution that citations provide only a limited and incomplete view of research quality. Research is too important, they say, to measure its value with only a single coarse tool. [ICIAM 2008]

The view taken here is that while individual article level metrics can provide useful information to researchers and to research management, there are caveats to the use of citation based bibliometric metrics in general:

- Immediacy of results is, in general, not good. Results depend on the publication cycle time, from submission to publication, and on some inevitable delay in uptake of publications as citations. In a discipline where the time from submission to publication is short, results will be more immediately available. But in disciplines with long publication cycles results will lag.
- Some disciplines, notably mathematics and physics, are starting to use publication media other than peer-reviewed journals, the traditional source of citation-based bibliometric data. Particularly, Web-hosted publication mechanisms fall outside the remit of traditional bibliometric analysis and are more amenable to altmetric methods.
- Citation analysis is sensitive to the journals and other sources selected for inclusion in the analysis. In amelioration, very large data sets of article-level publication and citation data are available, and should serve for most purposes.
- Problems with names and identity are troublesome for citation analysis:

  > The academic reward and reputational system, in fact, rests on the postulate that it is always possible to identify and assign the individual intellectual responsibility of a piece of scientific or technical work. [De Bellis 2009]

  However there remain considerable difficulties in determining a unique identity for each researcher. There is as yet no common researcher ID in the UK, although the JISC Research Identifiers Task and Finish Group has recently recommended the use of the Open Researcher and Contributor ID (ORCID) initiative [ORHID 2012], based on THOMSON's ResearcherID, as being most likely to meet UK needs [JISC/TFG undated]. Problems with names and identity are also being addressed outside the UK [Rotenberga and Kushmericka 2011].
- Citation analysis falls short of typical managerial concerns and should be used with care:

  > Citation analysis is not a substitute or shortcut for critical thinking; it is, instead, a point of departure for those willing to explore the avenues to thorough evaluation... citations tell us nothing about a researcher's teaching ability, administrative talent, or other non-scholarly contributions. And they do not necessarily reflect the usefulness of research for curing disease, finding new drugs, and so on. [Garfield 1985]

In one approach of interest, the Higher Education Funding Council for England (HEFCE) preferred expert review over bibliometrics in the context of the UK Research Evaluation Framework (REF):

> Bibliometrics are not sufficiently robust at this stage to be used formulaically or to replace expert review in the REF. However there is considerable scope for citation information to be used to inform expert review. [HEFCE 2009]

In summary, there is a very valid and widely held view that conventional citation-based bibliometric metrics are not suitable for use as the sole indicator of impact.

Thus the key recommendation here is that bibliographic metrics for the assessment of the impact of individual researchers or groups of researchers should not be used in isolation. Instead, a better

approach is to use a combination of various kinds of assessment (each of which have advantages and disadvantages) to provide a system of checks and counterbalances.

Such a system might usefully include:

- Peer assessment, to describe, contextualise and explain aspects of research and research performance.
- Two bibliographic indicators: one to indicate the size of the core productive output and one to indicate its scholarly impact.
- Other indicators deemed relevant and useful, for example economic impact or funding performance.
- Self assessment, to describe, contextualise and explain aspects of research and research performance.

Webometrics

The emergence of the World Wide Web has resulted in an analogous field to bibliometrics, but applied to the Web rather than serials per se. First proposed by Almind and Ingwersen [1997] webometrics is

> the study of the quantitative aspects of the construction and use of information resources, structures and technologies on the Web drawing on bibliometric and informetric approaches. [Björneborn and Ingwersen 2004]

The field is also defined by Thelwall [2009] as:

> the study of web-based content with primarily quantitative methods for social science research goals using techniques that are not specific to one field of study

The discussion of webometrics that appears below is largely limited to Internet use and reputation analysis. For other aspects of webometrics please see, for instance, [Thelwall 2010] and [Thelwall 2012].

Sampled visits to a site provide an indicative approximation of number of visitors, and are typically generated, on a global scale, using monitoring software that a large number of users install in their browsers.[15] To illustrate the state of the art, Alexa[16] provides for no cost rich metrics for the world's more popular web sites. Table 1 below shows three sites, ordered in decreasing order of popularity and reputation. Here visits to the site are used as a proxy for popularity and counts of links pointing to the site are used as a proxy for reputation[17].

---

15 This is distinct from use of other web analytics software, for example Google Analytics, which is deployed on a website-by-website basis to record the sources, numbers and basic behaviour of visitors to particular Web sites.

16 http://www.alexa.com/ the self-styled "leading provider of free, global web metrics."

17 All rankings in table 1 were generated on 1 December 2012.

Organisation	Organisation type	URL	Popularity low => good	Reputation high => good
CERN	european laboratory	cern.ch	20,567	19,873
Fraunhofer	national laboratory	fraunhofer.de	23,592	15,278
INRIA	national laboratory	inria.fr	34,832	9,787

Table 1: Popularity (by traffic rank) and reputation (by inbound hyperlink count) for three national and international research laboratories.

Google Trends, which monitors search term use, provides another form of web analytics. Google Trends' data has proven useful over a range of topics such as tracking disease [Pelat *et al* 2009] [Valdivia and Monge-Corella 2010] and economic forecasting [Schmidt and Vosen 2011].

There are other, more specialist rankings. For example CrunchBase[18] tracks technology companies, people and ideas, and illustrates how trend analytics might be used by academia to drive interest and highlight popular recent research.

While metrics for popularity and reputation exist as above, these are not correlated with institutional research capabilities or research reputation, in part because the Web hosts consumer sites as well as research and research-oriented sites. (For example, Facebook is ranked second for traffic by Alexa.)

De Bellis [2009] is clear about this:

> Although the reputation and visibility of an academic institution are partially reflected in the 'situation impact' of its website, no evidence exists, so far, that link rates might be determined by (or used as an indicator of) research performance. Web visibility and academic performance are, once and for all, different affairs.

Some researchers argue [De Bellis 2009] that general measures may need to be supplemented by more specialised measures, and have developed alternatives.

One such example, at webometrics.info, is a ranking of world universities supplied by the Cybermetrics Lab, part of the Spanish National Research Council. This presents sophisticated ranking data backed by academic research, for example [Aguillo *et al* 2008], for over 20,000 institutions in a variety of forms including graphical and aggregated forms suitable for various audiences. Interested readers can peruse the site to see the rankings and associated methodology used to generate the rankings.

18 http://www.crunchbase.com/

Similar services, also from webometrics.info, supply rankings for research centres, open access repositories, business schools, and hospitals. Here, the Ranking Web of Repositories ranks repositories that "have their own web domain or subdomain and include at least peer-reviewed papers to be considered (services that contain only archives, databanks or learning objects are not ranked)" [RWR 2012a]. The metrics used for this purpose are described by Aguillo *et al* [2010].

The site does come with a suitable caveat:

> We intend to motivate both institutions and scholars to have a web presence that reflect accurately their activities. If the web performance of an institution is below the expected position according to their academic excellence, institution authorities should reconsider their web policy, promoting substantial increases of the volume and quality of their electronic publications. [RWR 2012b]

The UK's top ten UK repositories according to the ranking metrics used by the site are shown in table 2.

United Kingdom

ranking	World Rank	Instituto	Size	Visibility	Files Rich	scholar
1	6	UK PubMed Central	3	11	39	6
2	28	University of Southampton ePrints	153	36	79	61
3	72	Open Research Online	371	87	372	110
4	76	LSE Research Online London School of Economics and Political Science	495	102	243	123
5	78	Cogprints Cognitive Sciences ePrint Archive	303	26	442	491
6	85	Natural Environmental Research Council Open Research Archive	646	85	260	180
7	97	University of Southampton: Department of Electronics and Computer Science	207	43	754	371
8	108	University of Glasgow ePrints	388	84	366	314
9	127	University College London ePrints	451	115	358	299
10	146	UCL Discovery University College London	133	321	323	114

Table 2: Top ten UK open access repositories for scientific papers with world ranking from http://repositories.webometrics.info/en/Europe/United%20Kingdom. Columns on the right are described at http://repositories.webometrics.info/en/Methodology.

Altmetrics

> Altmetrics is the creation and study of new metrics based on the Social Web for analyzing, and informing scholarship [Altmetrics 2012].

Altmetrics is generally concerned with a movement away from bibliometrics and scientometrics as the sole (or even the valid) measures of impact of scientific publications. This is entirely congruent with new publication and use opportunities that have been made possible by the Web, and the promulgation and discussion of these resources that has been made possible by the Social Web [Priem *et al* 2010].

In an empirical study Priem *et al* [2012] examined the sources of altmetric data "in the wild" by gathering altmetric events relating to 21,096 research articles that were published in the seven PLOS journals between 18 August 2003 and 23 December 2010. Figure 7 shows the distribution of approximately 1.8 million altmetric events partitioned according to event types and the article provides further detailed analysis of the dataset.

Conclusions include that there is no shortage of altmetric data, that the altmetric data and citation data cluster differently, meaning that neither on its own adequately portrays impact, and that there are different "flavours" to use, for example, some articles are heavily read, but not cited much.

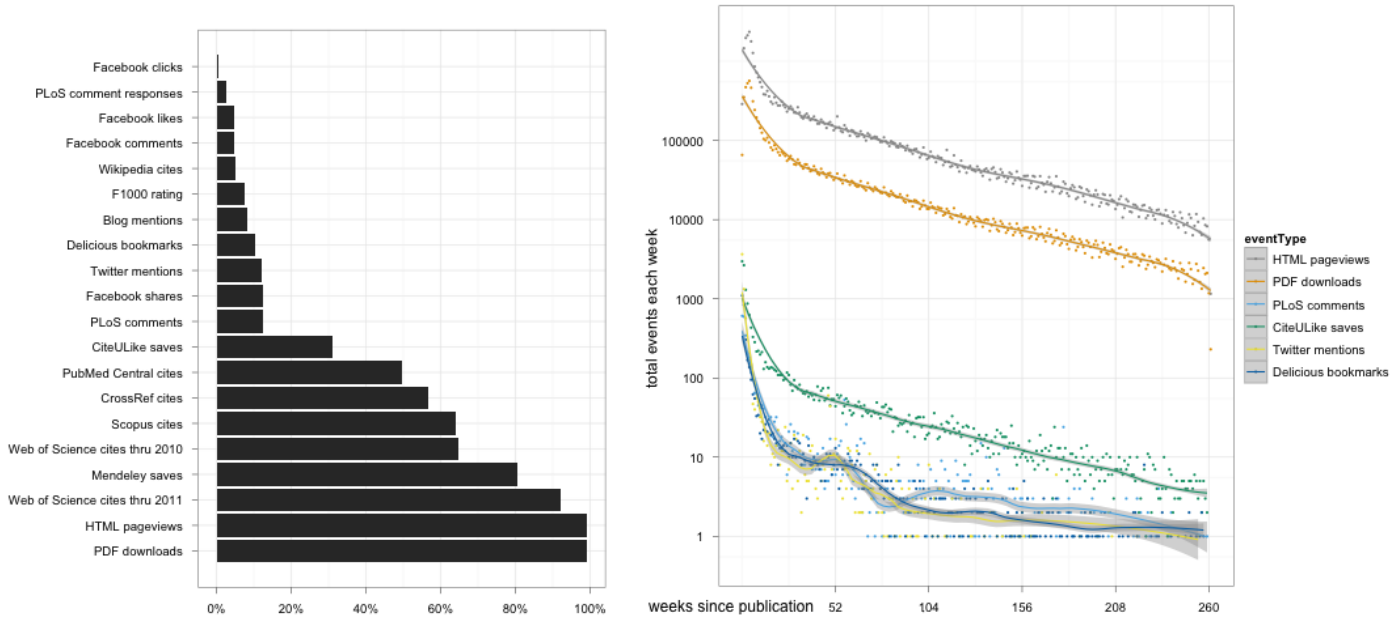

Figure 7: Left: Proportion of 21,096 research articles (in the seven PLOS journals) with at least one altmetric, by metrics.
Right: Altmetric event frequency in (up to) the five years following publication.
[Priem et al 2012].

The authors point to various limitations, including the challenge of data quality within altmetrics, difficulties in generalising from altmetric results, and the relative ease of gaming altmetrics (although experience suggests that careful data mining may be a solution to the latter).

The authors point to some intriguing possibilities:

> In the future, tools like these may allow researchers to keep a "live CV" showing up-to-date indicators of their works' impact next to each product they have produced. Funding organizations could keep timely portfolios of their grants impact across different populations and communities. Tenure, promotion, and hiring committees could assemble research teams whose members have different sets of "impact specialties," just as sport managers might assemble teams of players with different physical and mental skill sets. Altmetrics also have great potential for supporting personalized recommender systems; citation-based versions of these have been created before [*reference*], but suffer from the long lag between an idea's insemination and citation; altmetrics could solve this problem.

Various altmetrics tools exist, for example, Crowdometer[19], PaperCritic[20], PLoS Impact Explorer[21], ReaderMeter[22], ScienceCard[23], and Total-impact[24].

In addition, at least two commercial specialists offer altmetric services. Plum Analytics provides research analytic reports, interpretation and consultancy based on metrics that include several altmetrics [Plum 2012]. Altmetric.com computes an altmetric score, and offers open data via an API, an altmetric explorer and a bookmarklet (see figure 8) that displays basic altmetric data [Altmetric.com 2012].

However, because the altmetrics movement and its methods are an extremely recent advance there is generally a lack of objective information about altmetrics:

> Despite the growing speculation and early exploratory investigation into the value of altmetrics, there remains little concrete, objective research into the properties of these metrics: their validity, their potential value and flaws, and their relationship to established measures. Nor has there been any large umbrella to bring these multiple approaches together. [Laloup 2011]

19 http://crowdometer.org/

20 http://www.papercritic.com/

21 http://altmetric.com/interface/plos.html

22 http://readermeter.org/

23 http://sciencecard.org/

24 http://total-impact.org/

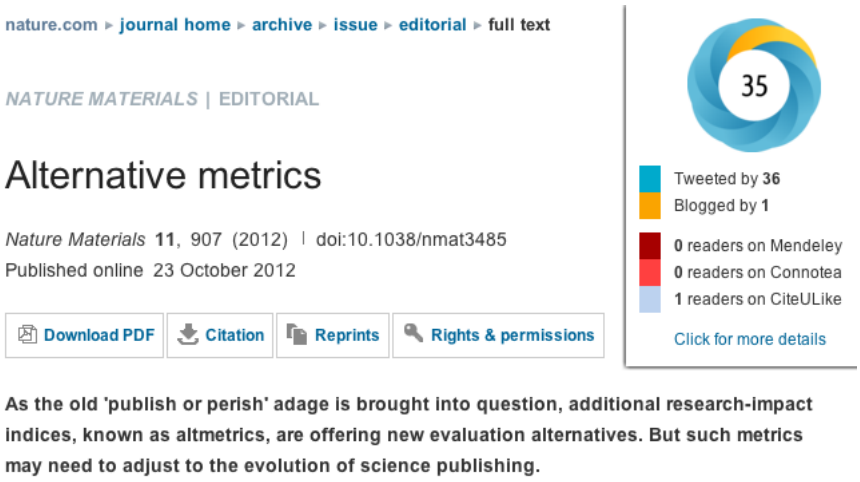

Figure 8: altmetrics.com bookmarklet (upper right) displaying basic altmetric data about an editorial in NATURE MATERIALS *including altmetrics.com's altmetric score (in the circle).* [NATURE MATERIALS 2012].

PLOS, the collection of seven peer-reviewed and web-delivered Open Access journals mentioned above (in the context of Article Level Metrics), also supplies some altmetrics: blog mentions are tracked using third party blog aggregators, research blogging.org and Nature Blogs. PLOS point out that its altmetrics are not currently comprehensive due to coverage of only part of the blogosphere, and difficulties the third party aggregation services have in identifying references to articles in PLOS journals. PLOS also tracks and publishes statistics as to how many times each article is cited in Wikipedia.

Figure 9: PLOS altmetrics for the article Eurekometrics: Analysing the Nature of Discovery *[Arbesman and Christakis 2011], showing a pop-up with more detailed Facebook mentions.*

It seems likely that altmetrics will likely yield major new measures of impact that are concerned with the 'buzz' around recently published work, rather than the work's more formal use as a citation in a later papers.

4.2 Analysis of use data

Information about the use of resources provides a proxy for interest in resources and their content.

Data about the use of resources (*use data*) can be collected from a variety of online services and connected physical devices. Common examples of use data include data that records web page accesses, views and downloads of documents including, importantly, accesses or downloads of journal articles.

There are a variety of applications for use data in analytics, including in the areas of altmetric webometrics (see section 4.1 above).

There are a variety of names for use data including *activity data*, *attention data*, and *the data exhaust* (that records user activity). Sources include click streams, log files, records of accesses and downloads, and records from physical devices.

PLOS, the collection of seven peer-reviewed and web-delivered Open Access journals mentioned above (in relation to both Article Level Metrics and altmetrics), tracks use and publishes aggregated use statistics. Article views for HTML, PDF and XML format articles are published in graphical format as an aggregate metric and as month-by-month totals. PLOS also retrieves and publishes similar usage data from PubMed Central. Currently PLOS provides some caveats in relation to interpreting its article level usage data.

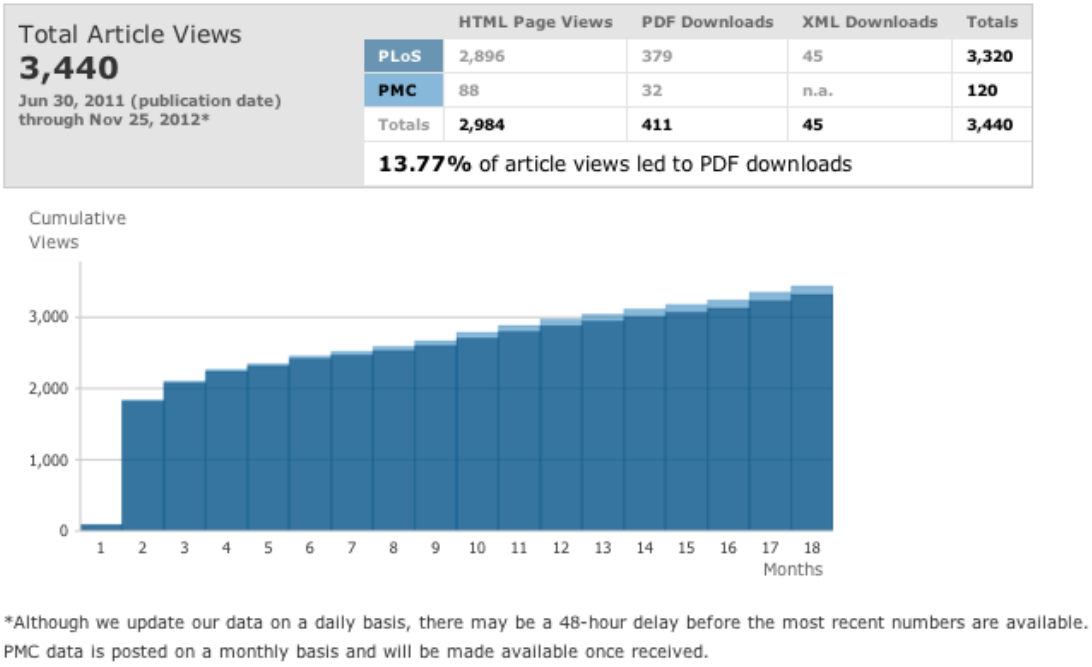

Figure 10: PLOS use data for the article Eurekometrics: Analysing the Nature of Discovery *[Arbesman and Christakis 2011].*

In the last few years activity data has risen to prominence in UK HE. Kay and van Harmelen [2012] describe advantages in using HE activity data:

> The HE sector is potentially in an advantageous position. A number of systems and services collect activity data, ranging from the VLE and the library to help desks and security. However, to make sense of human activity, context is king. Thanks to detailed knowledge of each user's context held in such as registration and learning systems (for example, level of study, course or specialism, course module choices and even performance), institutions have a powerful basis to segment, analyse and exploit low level human activity data. Consequently, activity data can enable an institution or a service to understand and support users more effectively and to manage resources more efficiently.

Various application areas for the analytic treatment of activity data are being exploited within HE. One application area, concerned with student retention and success, falls outside the remit of this paper.[25] A second area is the use of analytics to enhance experience, notably the learner experience, and to some extent also the researcher experience. An example is the use of recommendation techniques for library resources to create recommendations for researchers; see, for example, [Manista 2012][26] A second area, resource management, is very much applicable to research; a stand out application is management of journal subscriptions.

To illustrate use of activity data we describe some recent JISC projects:

- RISE is an experimental recommendation system implemented at the Open University. RISE uses data about the activities of library users and the courses they are on to provide recommendations to users who are using the Open University's Library's search facilities. Three kinds of recommendations are offered:
  - Search related: "People using similar search terms often viewed"
  - Relationship based: "These resources may be related to others you viewed recently" "People who looked at this resource also looked at…"
  - Course based: "People on your course(s) viewed…"

  In a survey 61% of users rated RISE's recommendations as "very useful" or "quite useful", and on a separate question 50% of users rated the recommendations as "very relevant" or "quite relevant" [RISE 2011].
- The Copac Activity Data Project (CopacAD) [COPAC 2012] has been undertaken by the team responsible for Copac, a union catalogue funded by JISC that aggregates UK HE library catalogues. With its work on CopacAD, the Copac team aims to (i) further understanding of the potential of aggregating and sharing library circulation data to support recommender functionality at local and national levels, and to (ii) scope the feasibility of a shared national service to support

[25] Interested readers are referred to [van Harmelen and Workman 2012] in this series.

[26] However, this is only one way of deriving recommendations, others might be on the basis of social network analysis, or by use of citation patterns.

circulation data aggregation, normalisation and distribution via an open API. The project is an extension of the Surfacing the Academic Long Tail (SALT) project, which focussed on how recommender systems can help to expose the long tail of underutilised but useful items for the humanities. SALT utilised ten years' circulation data from the University of Manchester and, from analysis of this data, provided recommendations through an experimental user interface to Copac and via an API [COPAC 2011]. During the past year (2012) a follow-up project, CopacAD, has added circulation data from the Universities of Huddersfield, Cambridge, Sussex and Lincoln, and updated the API.[27] JISC is now exploring how to take this work forward more formally as a service.

- The Open University project UCAID [d'Aquin et al 2011] addresses the problem of obtaining user data that meaningfully captures and represents user activity across a number of web sites. One could imagine application of UCAID's techniques in the research domain; generating exploitable use data to assist researchers who use multiple Web sites for related purposes. UCAID uses semantic techniques to aggregate, integrate and analyse these sources of activity data and might appear here under semantic methods (section 4.4). However, because of its close relationship to the sources of use data, UCAID appears in this section.
- JISC Collections' KnowledgeBase+ (KB+) falls in the resource management class of activity data applications. KB+ is directed at ensuring improvements in the supply of information to the research community through its journal management facilities. KB+ will utilise journal usage statistics (aggregated activity data) provided by another JISC Collections' service, JUSP. JUSP is particularly concerned with the aggregation of journal usage statistics from publisher web sites.

Significant use data projects exist outside the JISC ecosystem. Two are mentioned here.

- Metridoc [Zucca 2012a, 2012b], from the University of Pennsylvania, is an extensible framework that supports assessment and analytics in the context of higher education library management. Metridoc provides business intelligence to give library managers greater insight into use and effectiveness of their services. Metridoc collects use (and other) data from heterogeneous library and institutional sources including data on research consultation and library instruction, gate counts, discovery tool use, counter statistics, resource sharing data, and accounting data. Metridoc provides advances in the aggregation, normalisation, unification, visualisation and exploitation of these data for library oversight and management purposes.
- There is also a focus on much larger data sets of usage information, as, for example, by MESUR which studies science through the medium of large-scale usage and citation data. The MESUR database contains over 1B article-level use events during 2002-2007. MESUR [2012] states:

  > The collected use data spans more than 100,000 serials (including newspapers, magazines, etc.) and is related to journal citation data that spans more than 100,000 serials (including scientific journals) and nearly 10 years (1996-2006). In addition we have obtained significant publisher-provided COUNTER usage reports that span nearly 2,000 institutions worldwide.

[27] Updated API at http://vm-salt.mimas.ac.uk/copacAD/getSuggestions.api

4.3 Social network analysis

We turn now to techniques that focus on the shape and influence of research communities on the Web, and discuss social network analysis (SNA) as an illustrative technique.

Like bibliometrics, SNA is concerned with network analysis, but this time analysis of networks of people and/or institutions, rather than networks of citations. While SNA techniques predate the Web, they are a natural fit to analyse the interconnections between researchers who establish connections in online social networks. As such, SNA can properly be considered a technique in the general area of webometrics.

Again a pithy and general definition is hard to find – see for example [Scott 1987], a standard text in the area, or [Gretzel 2001] for definitions – but it is reasonable to state that SNA seeks to identify patterns and strengths of relations, connections or interactions between actors, where the actors may be individuals or institutions.

One might assume, given the rise of online social networks of researchers, widespread application of SNA techniques to analyse connections between researchers. However, the author has failed to find significant application of SNA techniques in the analysis of the online networks of researchers. This is to be expected, online networks tend to keep their social graphs (of who is connected to whom) private.

However, there is some readily apparent work that leads one author (De Bellis) to believe in promise for network analysis and SNA techniques. These

> have been promisingly applied in a series of case studies dealing with topics as diverse as e-commerce; social movements; and interpersonal, interorganisational, and international communication, altogether confirming the feasibility of web sociometric mining
> [De Bellis 2009]

Other analysis of networks might be properly termed webometrics: for example, a rich body of literature exists which applies classic network analysis to the hypertextual structure of the web, see [De Bellis 2009, ch8]. To take an example, Ortega and Aguillo [2009] use social network analysis techniques to visualise the relationships between 1,000 institutions, leading to some of the information appearing at webometrics.info, discussed in section 3.2.

Looking forward, Big Data technology [McKinsey 2011] promises to allow similar analytics to be run frequently, allowing the academic ecosystem to be explored visually to discover emergent trends.

Flink [Mika 2005], also discussed under semantic methods, extracts network information on linkages between semantic web researchers from their Friend of A Friend (FOAF) profiles and other sources (for example bibliographic citation data), and displays these in various ways (see figure 11).

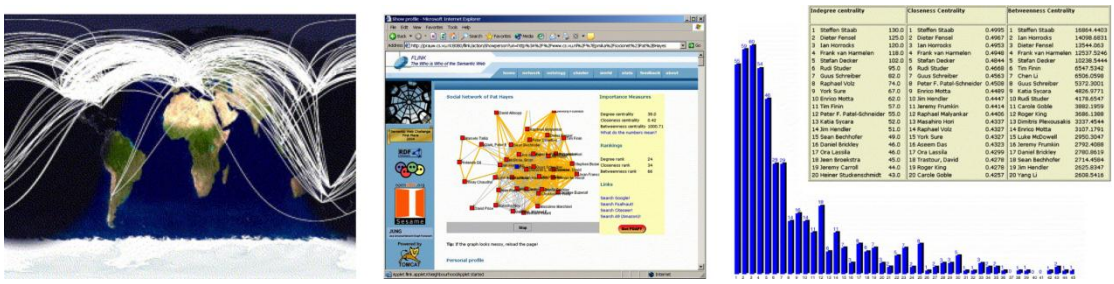

Figure 11: Left and center, depiction of links between semantic web researchers. Right, calculation of simple network statistics from the network [Mika, 2005].

Some work is concerned with classical application of SNA-based sociological techniques to the research and innovation domains. For example: Coulon [2005] summarises work on the application of SNA techniques to innovation. Pham *et al* [2011] apply SNA techniques based on citations to the development of sub-disciplines in computer science.

Other work is 'bubbling under', and points to application in the research domain: For example: Rosen *et al* [2011] *inter alia* provide an overview of SNA analyses of political blogs and bloggers, readily extensible to the research blogosphere. Nair and Dua [2011] propose a novel technique for the identification of communities from tagging behaviour in online social networks. Stieglitz and Dang-Xuan [2012] propose a methodological framework for the analysis of social media (including Twitter, blogs, and public Facebook content) that is applicable, in line with Stieglitz and Dang-Xuan's interests, to political content.

Given these advances, it is worth a detour to examine how researchers are using social media and the growth of online social networks of researchers.

Particularly we are interested in online social networks created using the faculties of social networking sites that are; Web-based services that allow individuals to construct a user profile and form links with other users.

Though specialist online social networks for academia are relatively new [Bullinger 2010], use is growing, according to *The Economist*, mostly amongst researchers in their 20s [Economist 2012].

A more nuanced view of young researchers' attitudes to online social networks is provided by JISC and the British Library's *Researchers of Tomorrow* project. This reports that:

- Take-up of most institutionally-provided and open Web technology tools and applications is low among doctoral students overall.
- Generation Y doctoral students are more likely than older doctoral students to use technology to assist them in their research.
- Generation Y doctoral students tend to use technology applications and social media in their research if they augment, and can be easily absorbed into, existing work practices.

- Levels of use of social media and other applications helpful in retrieving and managing research information are steadily rising among Generation Y doctoral students, but those applications most useful for collaboration and scholarly communications remain among the least used.
- Fellow students and peers are the major influence on whether or not Generation Y doctoral students decide to use a technology application and are their main source of hands-on help. [JISC/BL 2012]

The general view of online social networks for older researchers is that adoption is limited. However, given numbers of enrolments in leading research-oriented online social networks and the advantages offered by, for example, sharing journal article information (down to abstract but not full-text level) within those networks, some doubt must be associated with that statement. For example, the social networks that focus specifically on researchers include the following:

- Academia.edu[28] is a representative *research directory*, hosting over 1.5 million public profiles created by researchers and offering sophisticated search.
- Mendeley[29], with over 1.9 million members, is a representative *research information management site*,
- ResearchGATE[30] is a *research awareness* example with over 1.7 million members, allowing current interests to be tracked and communicated through a profile.

The terms research directory, research information management site and research awareness site are defined in [Bullinger et al 2010], which provides a taxonomy of online social networks for research.

Mendeley typifies a site which offers diverse advantages to researchers, not only can researchers store and share papers, but Mendeley contains an extremely large corpus of information about journal articles: Mendeley's users have collaboratively created, in three years, "a shared database containing [information about] 65 million unique documents and covering – according to recent studies – 97.2% to 99.5% of all research articles published" [Henning 2012]. This searchable database is extremely good; at times during the writing of this paper it served better than Google, Google Scholar and Microsoft Academic Search (the latter the 'poor' relative, due to its search speed).

More examples of online social networks for researchers appear in [Bittner and Müller 2011]. Schleyer *et al* [2012] lists university-based systems including VIVO (University of Florida), Catalyst Profiles (Harvard), and Loki (University of Iowa), and a variety of commercial products including cos[31], Index Copernicus Scientists[32], Research Crossroads[33], BiomedExperts[34] and Epernicus[35].

---

28 http://www.academia.edu/

29 http://www.mendeley.com/

30 http://www.researchgate.net/

31 http://www.cos.com

32 http://scientists.indexcopernicus.com/

33 http://www.researchcrossroads.com/

With the rise of Web 2.0, a further rich ecosystem of independent social networks has been adopted by researchers, sometimes with institutional encouragement: for example, the LSE Public Policy Group encourages academic use of Twitter [Mollet and Moran 2011] [Mollett *et al* 2011].

Other systems exploit citations and author information to create exploitable network data, for example, Social Science Research Network[36] allows the public to browse from a paper to related papers.

In summary, online social networks offer novel sources and types of data as well as posing new challenges in exploiting that data. This evolving landscape suggests that analytics will also need to evolve, and that there may be challenges to researchers and organisations in exploiting that information.

4.4 Semantic methods

Of the methods discussed in this paper, semantic methods are the least utilised, but show great promise for the study of research and management of knowledge. The advantages of being able to categorise research are manifold, not least to categorise existing research, discern trends in research, manage and exploit knowledge and allocate resources adequately, particularly in response to changing directions in research.

It is postulated by this author that adequate and evolvable categorisations are a necessary feature for the longer-term development of analytics for research and research management. In this context, the following is particularly apt:

> … no clear alternative socio-cognitive structure has yet replaced the "old" disciplinary classification. In this fluid context, in which social structure often no longer matches with the dominant cognitive classification in terms of disciplines, it has become increasingly necessary for institutions to understand and make strategic choices about their positions and directions in moving cognitive spaces. [Rafols *et al* 2010]

While prior work in discovering the structure of science and the humanities (as discussed above) is acknowledged, one might look largely to semantic technologies as being able to make the necessary advances in this area. When looking to apply quantitative evidence about research, semantic meaning is crucial. Individuals and organisations looking to use analytic data about research are likely to face similar issues.

Flink [Mika 2005], mentioned above in section 3.3, illustrates the potential for improved understanding when analytics are combined with semantic web technology. Flink builds striking visuals about the shape of the research community and the position of researchers within it, and the relationships between topics studied by that community from public data already available on the web. The promise of big data

---

[34] http://www.biomedexperts.com/

[35] http://www.epernicus.com

[36] http://papers.ssrn.com/

processing is that visualisations of this sort be generated routinely over time and would be available to inform researchers and decision makers.

> Flink employs semantic technology for reasoning with personal information extracted from a number of electronic information sources including web pages, emails, publication archives and FOAF profiles. The acquired knowledge is used for the purposes of social network analysis and for generating a web-based presentation of the community. We demonstrate our novel method to social science based on electronic data using the example of the Semantic Web research community. [Mika 2005]

There is progress with ontological descriptions of domains in science and technology, currently driven by e-Science applications. However, the use of ontologies in the generation of knowledge may perhaps be taken further. Thus, three examples illustrate the state of the art and possible progressions:

- Ontologies may be applied in easing the production of scientific information. Fox *et al* [2009] describe how anthologies are used to describe equipment, observations and phenomena in order to ease the production of observational artefacts within a virtual solar-terrestrial observatory. This is fairly traditional e-Science where knowledge, captured in ontologies is used to help the production of new knowledge. Fox *et al* suggest that their techniques are replicable and cite evidence from other data-intensive fields (vulcanology, tectonics) to support this conclusion.
- Brodaric and Gahegan [2010] suggest that ontologies which describe theories, hypotheses and models may be used to help test these scientific artefacts. This approach is known as semantic science: "Such direct scientist interaction with the ontology-enabled knowledge, i.e. 'in-silico' semantic science, should then help revitalize online scientific methodology by helping generate richer insights, and improving our ability to repeat, report, and validate scientific findings."
- Poole *et al* [2009] also explore semantic science, citing their own work in modelling and representations of geology, by suggesting that ontologies can be used to make probabilistic predictions that can then be tested, for example, as to the location of minerals.

Three commercially available analytic solutions that might be applicable to research and research management are discussed next. They appear in this section because they use semantic methods that work on large bodies of text and various other kinds of unstructured and structured data, up to data warehouse scale.[37]

Semantica [Semantic Research 2012], largely a tool for government intelligence agencies, explores networks of information extracted from raw data, automatically extracting networks of entities that represent people, places, things, events and ideas, combining data from different input sources to create holistic views of the entities. Relationships analytics focuses on the nature and type of the relationship connecting entities as an aid to understanding the network and discovering new relationships. Extensible ontologies are used to typify entity and relationship types.

[37] However, they could also be used as examples in the section on Social Network Analysis.

Directed at the intelligence and enterprise communities, Synthesys [Digital Reasoning 2012a] is directed at automatic analysis of large volumes of textual data using text recognition and entity classification techniques. Using an Oracle big data appliance, benchmarks show Synthesys can process 1M documents in 24 hours and 233M in 300 hours [Oracle 2012]. This kind of ability leads to interesting applications; one is the construction of dynamic social networks of influence from blog posts. [Digital Reasoning 2011] shows the results of analysis of 16M posts posted in 6M blogs.

Recently the above two tools have been integrated with each other [Arnold 2012]. Technical details appear in [Digital Reasoning 2012b]. Synthesys has also been integrated [Digital Reasoning 2012c] with Tableau Desktop [Tableau Software 2012], a very dynamic and visually-oriented business intelligence tool that uses a desktop direct-manipulation metaphor in the analysis and display of data sources that include data warehouses, databases, files and spreadsheets. [Danielson 2012] and [Nelson 2012] respectively show intelligence and financial data based demonstrations of these integrated systems.

While the direction for adoption is not specifically in the areas of analytics for research or research management, examining the videos cited above – [Danielson 2012] [Digital Reasoning 2011] and [Nelson 2012][38] – gives rise to evocative and intriguing thoughts for the deployment of these kinds of tools in the analysis of research and the enactment of research management.

[38] Respectively, http://tinyurl.com/analysisvid1, http://tinyurl.com/analytisvid2, and http://tinyurl.com/analysisvid3

5. Observations and conclusions

The use of analytics to understand research is an area fraught with difficulties that include questions about the adequacy of proxies, validity of statistical methods, understanding of indicators and metrics obtained by analytics, and the practical use of those indicators and metrics in helping to plan, develop, support, enact, assess and manage research.

However, in an uncertain landscape the drive to understand is often paramount: when questions of funding arise the impetus to assess research is, for a variety of social, economic and organisational reasons, unavoidable. In such a situation, while reduction of research to 'easily understandable' numbers is attractive, there is a danger of over-reliance on analytic results without seeing the larger picture.

The rational consensus is that numeric bibliometrics should not be used in isolation, and one might usefully suggest that all metrics should not be used in isolation. Understanding the context of research is key to an informed interpretation of numeric data. This is particularly important for bibliometric measures of scholarly impact, the most commonly-used indicators of the impact of research.

New forms of analytics are in development and use. These include research overlay maps, varying forms of altmetrics, semantic analysis and social network analysis. These offer to broaden the application of analytics by enabling new kinds of applications, for example understanding the relationship between areas of research, or applying semantic knowledge management techniques.

The view here is that a use of analytics to understand research is, despite the impediments mentioned above, a given part of contemporaneous research, both at a researcher and institutional level, and as an activity within the larger research system. Given the fundamental importance of assessment of research and the role that analytics may play, it is of paramount importance to construct assessment frameworks in institutions and beyond that use analytics appropriately. Adoption of research reputation management and enhancement practices may become a major activity in the UK research landscape.

The key risks involved in the use of analytics to understand research include:

- Use of bibliometric indicators as the sole measure of research impact or over-reliance on metrics without any understanding of the context and nature of the research.
- Lack of understanding of analytics and advantages and disadvantages of different indicators on the part of users of those indicators. Managers and decision makers may lack the background needed to interpret existing analytics sensitively.
- In respect to a move towards target-based management empowered by advances in analytics, the suitability of target-based assessment based on analytics is unproven. A wider assessment approach was tentatively recommended above (in most detail on page 29).
- Developing usable analytics solutions for research and research management is a hard problem.
  - CRIS vendors may lack the skills required to implement rich, high quality analytics.

- There is a danger of one or a few vendors supplying systems that impose a particular view of analytics on research management data.

Bearing in mind the reservations and risks identified above, key opportunities are that:

- Access to high-quality timely analytics may enable professionals to gauge their short-term performance, and use experimentation to discover new and novel ways to boost their impact.
- Progress is being made in respect of Current Research Information Systems, their use of CERIF and improvements to the CERIF standard. These will inevitably bring an increased focus on the use of analytics in research management in UK universities.
- Adoption of CERIF-based CRIS across UK HE institutions and research institutes, with automatic retrieval of public data by UK Research Councils may help motivate increases in public funding of scientific and other scholarly activity; vitally important to the UK economy and national economic growth.
- Training as to the advantages, limitations and applicability of analytics may assist in the effective use of analytics its lay users, including researchers, research managers, and those responsible for policy and direction in institutions and beyond.

Finally, there are also key opportunities that are related to technical advances:

- Although not considered here beyond brief mention (on page 26), Open Data offers exciting opportunities for new and widespread analytic exploitation of research data.
- The use of semantic methods is likely to provide domain descriptions that can be leveraged by tools that assist researchers in knowledge management and discovery activities.
- Big data techniques may become widely employed, particularly so in response to large research data sets being published on the web.
- Analytics empowered sociology of science and science of science policy, and analytics applied to the understanding of the structure of research domains may increasingly bring advances in a variety of areas that include identification of new trends in research, further enabling opportunities for research co-operation, support for the research process, and other activities in the research ecosystem.

6. References

[Adams *et al* 2007] Adams J, Gurney K and Marshal S, *Patterns of international collaboration for the UK and leading partners*, Summary report commissioned by the UK Office of Science and Innovation, Evidence Ltd, June 2007. http://image.guardian.co.uk/sys-files/Education/documents/2007/07/13/OSICollaborationSummaryRepo.pdf

[Aguillo *et al* 2008] Aguillo IF, Ortega JL and Fernández M. (2008). Webometric Ranking of World Universities: Introduction, Methodology, and Future Developments. *Higher Education in Europe*, **33**(2/3), 2008. http://digital.csic.es/bitstream/10261/32190/1/Ranking%20of%20Repositories.pdf

[Aguillo *et al* 2010] Aguillo IF, Ortega JL, Fernández M and Utrilla AM, Indicators for a webometric ranking of open access repositories, *Scientometrics*, **82**(3), 2010. http://digital.csic.es/bitstream/10261/32190/1/Ranking%20of%20Repositories.pdf

[Almind and Ingwersen 1997] Tomas C. Almind and Peter Ingwersen (1997). Informetric analyses on the World Wide Web: Methodological approaches to 'webometrics', *Journal of Documentation*, **53**(4), 1997. http://www.emeraldinsight.com/journals.htm?articleid=864059

[Altmetric.com 2012] Altmetric.com, About us, *Altmetric Web site*, 2012. http://altmetric.com/help.php

[Altmetrics 2012] altmetrics.org, About, Web page, *altmetrics.org Web site*, 2012. http://altmetrics.org/about/

[Arbesman and Christakis 2011] Arbesman, S, and Christakis, NA, Eurekometrics: Analysing the Nature of Discovery, *PLOS COMPUTATIONAL BIOLOGY*, **7(**6), June 2011. http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1002072

[Arnold 2012] Arnold, SE, Digital Reasoning and Semantic Research Tie Up, *Beyond Search*, 2012. http://arnoldit.com/wordpress/2012/04/02/digital-reasoning-and-semantic-research-tie-up/

[BIS 2011] Department of Business Innovation and Skills, *Innovation and Research Strategy for Growth*, Parliamentary Report, Department of Business Innovation and Skills, UK Government, December 2011. http://www.bis.gov.uk/assets/biscore/innovation/docs/i/11-1387-innovation-and-research-strategy-for-growth

[Bittner and Müller 2011] Sven Bittner S and Müller A, Social networking tools and research information systems: Do they compete? *Proc ACM WebSci'11*, June 2011. Also published in *Webology*, **8**(1), 2011. http://www.webology.org/2011/v8n1/a82.html

[Björneborn and Ingwersen 2004] Björneborn Land Ingwersen P, Toward a basic framework for webometrics, *Journal of the American Society for Information Science and Technology*, **55**(14), December 2004. http://onlinelibrary.wiley.com/doi/10.1002/asi.20077/abstract

[Bollen *et al* 2009] Bollen, J, Van de Sompel, H, Hagberg, A, and Chute, R, A Principal Component Analysis of 39 Scientific Impact Measures, *PLoS ONE*, **4**(6), June 2009. http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0006022

[Brodaric and Gahegan 2010] Brodaric B, and Gahegan M, Ontology use for semantic e-Science, *Semantic Web*, 10(1-2), IOS Press1, 2012.
http://iospress.metapress.com/content/e4557v7151g0n628/?issue=1&genre=article&spage=149&issn=1570-0844&volume=1 and http://iospress.metapress.com/content/e4557v7151g0n628/fulltext.pdf

[Bullinger 2010] Bullinger A, Hallerstede SH, Renken U, Soeldner JH, Moeslein KM, Towards Research Collaboration – a Taxonomy of Social Research Network Sites, *Proc Sixteenth Americas Conference on Information Systems*, 2010.
http://www.academia.edu/603019/Towards_Research_Collaboration_a_Taxonomy_of_Social_Research_Network_Sites

[COPAC 2011] COPAC SALT Team, SALT Recommender API, *COPAC Activity Data Blog*,
http://copac.ac.uk/innovations/activity-data/?page_id=227

[COPAC 2012] COPAC, *COPAC Activity Data Blog*, 2012. http://copac.ac.uk/innovations/activity-data/

[Coulon 2005] Coulon, F, The use of Social Network Analysis in Innovation Research: A literature review, *DRUID Academy Winter 2005 PhD Conference*, 2005.
http://www.druid.dk/conferences/winter2005/papers/dw2005-305.pdf

[Davenport 2006] Davenport TH, Competing on Analytics, *Harvard Business Review*, **84**(1), January 2006. http://hbr.org/2006/01/competing-on-analytics/ar/1

[Danielson 2012] Danielson D, Synthesys Tableau 3Min S Demo HD, *YouTube*, 2012.
http://www.youtube.com/watch?v=yOiZ2yqH6P0

[d'Aquin et al 2011] d'Aquin, M, Elahi, S, and Motta, E, *Semantic Technologies to Support the User-Centric Analysis of Activity Data*, Knowledge Media Institute, The Open University, 2011.
http://sdow.semanticweb.org/2011/pub/sdow2011_paper_8.pdf

[De Bellis 2009] De Bellis, N, *Bibliometrics and Citation Analysis: From the Science Citation Index to Cybermetrics*, Scarecrow Press Inc, ISBN 978-0-8108-6713-0, 2009.
http://www.amazon.com/Bibliometrics-Citation-Analysis-Science-Cybermetrics/dp/0810867133

[Digital Reasoning 2011] Digital Reasoning, Synthesys Overview, *YouTube*, 2011.
http://www.youtube.com/watch?v=3msW0RvzoJ4

[Digital Reasoning 2012a] *Digital Reasoning, Synthesys Technology Overview*, White Paper, 2012.
http://www.digitalreasoning.com/wp-content/uploads/2011/12/Synthesys_Technology_Overview.pdf

[Digital Reasoning 2012b] *Digital Reasoning, Semantica-Synthesis Integration*, White paper, 2012.
http://www.digitalreasoning.com/wp-content/uploads/2012/03/Semantica-Synthesys-Integration-201203221.pdf

[Digital Reasoning 2012c] Digital Reasoning, Digital Reasoning and Tableau Software Partner to Advance Big Data Analytics, Press Release, *Digital Reasoning Web site*, 2012.

http://www.digitalreasoning.com/2012/company-news/digital-reasoning-and-tableau-software-partner-to-advance-big-data-analytics/

[Fox et al 2009] Fox P, McGuinness DL, Cinquini L, West P, Garcia J, Benedict JL, Middleton D, Ontology-supported scientific data frameworks: The virtual solar-terrestrial observatory experience, *Computers & Geosciences*, 35(4), 2009. http://www.sciencedirect.com/science/article/pii/S0098300409000247

[Economist 2012] The Economist, Professor Facebook: More connective tissue may make academia more efficient, *The Economist*, 2012. http://www.economist.com/node/21547218

[Egghe, 2006a] Egghe, L, Theory and practise of the g-index, *Scientometrics*, **69**(1), 2006. http://www.springerlink.com/content/4119257t25h0852w/

[Egghe, 2006b] Egghe, L, An improvement of the H-index: the G-index. *ISSI Newsletter*, 2006. http://pds4.egloos.com/pds/200703/08/11/g_index.pdf

[Garfield 1955] Garfield E, Citation Indexes for Science A New Dimension in Documentation through Association of Ideas, *Science*, 122, July 1955. http://garfield.library.upenn.edu/papers/science1955.pdf

[Garfield *et al* 1964] Garfield, E, Sher IH, and Torpie, RJ. *The Use of Citation Data in Writing the History of Science*, Report for the Air Force Office of Scientific Research, Institute for Scientific Information. http://www.garfield.library.upenn.edu/papers/useofcitdatawritinghistofsci.pdf

[Garfield 1985] Garfield, E, Uses and abuses of citation frequency, *Current Contents*, **43**, October 1985. http://garfield.library.upenn.edu/essays/v8p403y1985.pdf

[Garfield 1999] Garfield E, Journal impact factor: a brief review, *Canadian Medical Association Journal*, **161**(8), October 1999. http://www.cmaj.ca/content/161/8/979.full?ijkey=82f872a0c0569207b070f68f997e2278f801fd58

[Gaskell 2012] Gaskell, S, From here to posterity, *Times Higher Education*, 9 August 2012. http://www.timeshighereducation.co.uk/story.asp?storycode=420801

[GII 2012] *Global Innovation Index 2012 Edition*, Web site, INSEAD and the World Intellectual Property Organisation, 2012. http://www.globalinnovationindex.org/gii/

[Gretzel 2001] Gretzel U, Social Network Analysis: Introduction and Resources, Web page, 2001. http://lrs.ed.uiuc.edu/tse-portal/analysis/social-network-analysis/

[HEFCE 2009] Higher Education Funding Council for England, *Report on the pilot exercise to develop bibliometric indicators for the Research Excellence Framework*, Issues Paper, HEFCE, September 2009. http://www.hefce.ac.uk/media/hefce1/pubs/hefce/2009/0939/09_39.pdf

[Hirsch 2005] Hirsch JE, An index to quantify an individual's scientific research output, *PNAS*, **102**(46), November 2005. http://www.pnas.org/content/102/46/16569

[Haran and Poliakoff 2011] Haran B and Poliakoff M, How to measure the impact of chemistry on the small screen, *Nature Chemistry*, **3**, February 2011. http://www.nature.com/nchem/journal/v3/n3/full/nchem.990.html

[Henning 2012] Henning V, Mendeley handles 100 million calls for Open Science, per month, *MENDELEYBLOG*, 2012. [http://blog.mendeley.com/open-access/mendeley-handles-100-million-calls-for-open-science-per-month

[Hirsch 2005] Hirsch, JE, An index to quantify an individual's scientific research output, *PNAS*, **102**(46), November 2005. http://www.pnas.org/content/102/46/16569

[Holdren 2009] Holdren J, quoted in Mervis J, When Counting Jobs Isn't Enough, *Science*, 326, November 2009: http://www.ncbi.nlm.nih.gov/pubmed/19892954

[ICIAM 2008] ICIAM, Report of the IMU/IMS/ICIAM Joint Committee on *Quantitative Assessment of Research*, *ICIAM Web Site*, June 2008. http://www.iciam.org/QAR/

[IUIIS 2012] Innovation Union Information and Intelligence System, Development of an Innovation Headline Indicator, Commitment 34-A, *Innovation Union Information and Intelligence System*, May 2012. http://i3s.ec.europa.eu/commitment/40.html

[Iglesias and Pecharromán 2007] Iglesias JE and Pecharromán C, Scaling the h-index for different scientific ISI fields, *Scientometrics*, **73**(3), 2007. http://www.springerlink.com/content/8x4088qp27227818/

[JISC/BL 2012] The Researcher of Tomorrow, Project web site, JISC and The British Library, 2012. http://www.jisc.ac.uk/publications/reports/2012/researchers-of-tomorrow.aspx

[JISC/TFG undated] JISC Task and Finish Group, *Task and Finish Group Recommendations*, JISC, undated. http://www.jisc.ac.uk/media/4/C/E/%7B4CE0831F-62DF-46FA-A95C-D96802E39CA0%7DTask-Finish-Group-Recommendations-and-ORCID-initiative-and-principles.pdf

[Jump 2012] Jump, Queen Mary job losses provoke industrial action, *Times Higher Education*, 4 October 2012. http://www.timeshighereducation.co.uk/story.asp?storyCode=421385&sectioncode=26

[Katz and Martin1997] Katz JS and Martin BR, *What is Research Collaboration*, SPRU, 1995. http://www.sussex.ac.uk/Users/sylvank/pubs/Res_col9.pdf

[Kay and van Harmelen 2012] Kay, D, and van Harmelen, M, *Activity Data: Delivering benefits from the Data Deluge*, JISC Digital Infrastructure Directions Report: Activity Data: Analytics and Metrics, in publication. See http://www.jisc.ac.uk/fundingopportunities/funding_calls/2012/02/did_activity_data.aspx

[Kessler 1963] Kessler, MM, Bibliographic coupling between scientific papers, *American Documentation*, **14**(1), January 1963. http://onlinelibrary.wiley.com/doi/10.1002/asi.5090140103/abstract

[Kuhn 1962] Kuhn, T, *The Structure of Scientific Revolutions*, University of Chicago Press, ISBN 0226458083, 1962.

[Liu *et al* 2005] Liu P, Curson J and Dew P, Use of RDF for expertise matching within academia, *Knowledge and Information Systems*, **8**(1), 2005. http://www.springerlink.com/content/hdcavuex15pkljw4/

[Laloup 2011] Laloup J, Altmetrics: Tracking scholarly impact on the social Web – PLoS ONE Collection, Everyone PLOS ONE Community Blog, *PLOS BLOGS*, November 2011. http://blogs.plos.org/everyone/2011/11/08/altmetrics-tracking-scholarly-impact-on-the-social-web-plos-one-collection/

[Largent and Lane 2012] Largent MA, Lane JI, STAR METRICS and the Science of Science Policy, *Review of Policy Research*, 29(3), May 2012. http://onlinelibrary.wiley.com/doi/10.1111/j.1541-1338.2012.00567.x/abstract

[Leydesdorff and Opthof 2010] Leydesdorff L and Opthof T, Scopus's Source Normalized Impact per Paper (SNIP) versus a Journal Impact Factor based on Fractional Counting of Citations, *Journal of the American Society for Information Science & Technology*, **61**(11), November 2010. http://onlinelibrary.wiley.com/doi/10.1002/asi.21371/abstract and http://arxiv.org/pdf/1004.3580.pdf

[Leydesdorff and Rafols 2011] Leydesdorff and Rafols I, Interactive Overlays: A New Method for Generating Global Journal Maps from Web-of-Science Data. *Journal of Informetrics*, **6**(2), 2011. http://www.leydesdorff.net/journalmaps/journalmaps.pdf

[Leydesdorff and Schank 2008] Leydesdorff L, and Schank, T, Dynamic animations of journal maps: Indicators of structural changes and interdisciplinary developments, *Journal of the American Society for Information Science and Technology*, **59**(11), September 2008. http://arxiv.org/pdf/0911.1437.pdf

[Lozano *et al* 2012] Lozano GA, Larivière V and Gingras Y, The weakening relationship between the impact factor and papers' citations in the digital age. *J Am Soc Inf Sci*, 2012. http://arxiv.org/pdf/1205.4328.pdf

[LSE 2011] LSE, Introduction: Defining Research Impacts, *Impact of Social Sciences Blog*, LSE, 2011. http://blogs.lse.ac.uk/impactofsocialsciences/introduction

[LSE PPG 2011a] LSE Public Policy Group, *Maximizing The Impacts Of Your Research: A Handbook For Social Scientists*, London School of Economics, April 2011 http://www2.lse.ac.uk/government/research/resgroups/LSEPublicPolicy/Docs/LSE_Impact_Handbook_April_2011.pdf

[LSE PPG 2011b] The Handbook, *LSE Impact of Social Sciences blog*, 2011. http://blogs.lse.ac.uk/impactofsocialsciences/the-handbook/

[NIH 2010] National Institute of Health, STAR METRICS: New Way to Measure the Impact of Federally Funded Research, *NIH News*, June 2010. http://www.nih.gov/news/health/jun2010/od-01.htm

[Manista 2012] Manista, F, Working with Academics and the COPAC Recommender, *Copac* blog*, July 2012. http://copac.ac.uk/innovations/activity-data/?author=8

[McKinsey 2011] Manyika, J, Chui, M, Brown, B, Bughin, J, Dobbs, R, Roxburgh, C, Byers, A, *Big data: The next frontier for innovation, competition, and productivity*, McKinsey Global Institute Research Report, 2011.
http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation

[MESUR 2012] MESUR, About, *MESUR*, 2012. http://mesur.informatics.indiana.edu/?page_id=2

[MICE 2011] JISC MICE Project, *MICE Blog*, King's College London, 2011. http://mice.cerch.kcl.ac.uk/

[Mika 2005] Mika P, Flink: Semantic Web technology for the extraction and analysis of social networks, *Web Semantics: Science, Services and Agents on the World Wide Web*, 3(2–3), 2005.
http://www.sciencedirect.com/science/article/pii/S1570826805000089 and
http://www.cs.vu.nl/~pmika/research/papers/JWS-Flink.pdf

[Mollet and Moran 2011] Mollet A and Moran D, The LSE Impact Blog's new guide to using Twitter in university research, teaching, and impact activities, is now available, *British Politics and Policy at LSE blog*, 2011. http://blogs.lse.ac.uk/politicsandpolicy/2011/10/15/twitter-guide/

[Mollett et al 2011] Amy Mollett, Danielle Moran and Patrick Dunleavy, *Using Twitter in university research*, Manual, LSE, 2011.
http://eprints.lse.ac.uk/38489/1/Using_Twitter_in_university_research,_teaching_and_impact_activities_(LSE_RO).pdf

[Nair and Dua 2011] Nair, V, and Dua, S, Folksonomy-based ad hoc community detection in online social networks, *SOCIAL NETWORK ANALYSIS AND MINING*, **2**(4), 2012.
http://www.springerlink.com/content/004ph10763577100/

[NATURE MATERIALS 2012] NATURE MATERIALS, Alternative Metrics, Editorial, *NATURE MATERIALS*, **11**, November 2012. http://www.nature.com/nmat/journal/v11/n11/full/nmat3485.html

[Nelson 2012] Nelson, E, Tableau Synthesys Voice, *Vimeo*, 2012. http://vimeo.com/51924043

[Okubo 1997] Okubo Y, BLIOMETRIC INDICATORS AND ANALYSIS OF RESEARCH SYSTEMS: METHODS AND EXAMPLES, *STI WORKING PAPERS*, 1997/1, OECD, 1997.
http://search.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=OCDE/GD(97)41&docLanguage=En

[Oracle 2012] Oracle, *Big Data and Natural Language: Extracting Insight From Text*, White Paper, Oracle, 2012. http://www.digitalreasoning.com/wp-content/uploads/2012/10/Oracle-BDA-Digital-Reasoning-White-Paper_October-2012.pdf

[ORCID 2012] ORCID, What is ORCHID, *ORCHID Web site*, 2012. http://about.orcid.org/about/what-is-orcid

[Ortega and Aguillo 2009] Ortega, JL, and Aguillo, I, Mapping World-class universities on the Web, *Journal of Information Processing and Management*, **45**(2), March 2009. http://dl.acm.org/citation.cfm?id=1508677 and http://internetlab.cchs.csic.es/cv/11/world_map/map.html

[OSTP 2012a] The White House Office of Science and Technology Policy, the National Science Foundation and the National Institutes of Health, STAR METRICS, *National Institute for Health Web site,* 2012. https://www.starmetrics.nih.gov/

[OSTP 2012b] The White House Office of Science and Technology Policy, About SoSP, *OSTP Science of Science Policy Web site*, 2012. http://scienceofsciencepolicy.net/page/about-sosp

[Pelat *et al* 2009] Pelat C, Turbelin C, Bar-Hen A, Flahault A, Valleron A, More diseases tracked by using Google Trends, *Emerging Infectious Diseases*, **15**(8), August 2009. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2815981/

[Pham *et al* 2011] Pharm, MC, Klamma, R, and Jarke, M, Development of Computer Science Disciplines - A Social Network Analysis Approach, *arXiv*, March 2011. http://arxiv.org/abs/1103.1977

[Plum 2012] Plumb Analytics, About, Web page, *altmetrics.org web site*, 2012.. http://www.plumanalytics.com/about.html

[Poole *et al* 2009] Poole D, Smythe C and Sharma R, Ontology Design for Scientific Theories That Make Probabilistic Predictions, *IEEE Inteligent Systems*, 24(1), 2009. http://www.georeferenceonline.com/OntologyDesignforScientificTheoriesIEEE.pdf

[PLOS 2012] Public Library of Science, ARTICLE-LEVEL METRICS, *PLOS Web site*, 2012. http://article-level-metrics.plos.org/

[Priem *et al* 2010] Priem, J, Taraborelli, D, Groth, P. and Neylon, C. altmetrics: a manifesto, *altmetrics Web site,* 2010. http://altmetrics.org/manifesto/

[Priem *et al* 2012] Priem, J, Piwowar, HA, and Hemminger, BM, Altmetrics in the Wild: Using Social Media to Explore Scholarly Impact, *arXiv*, March 2012. http://arxiv.org/abs/1203.4745

[PubMed 2012] PubMed, PubMed Home Page, US National Library of Medicine and National Institutes of Health, 2012. http://www.ncbi.nlm.nih.gov/pubmed

[Puustinen and Edwards 2012] Puustinen, K, and Edwards, R, Who gives a tweet? After 24 hours and 860 downloads, we think quite a few actually do, *LSE Impact of Social Science blog*, May 2012. http://blogs.lse.ac.uk/impactofsocialsciences/2012/05/18/who-gives-a-tweet-860-downloads/

[QMUL UCU 2012] University and College Union Branch of Queen Mary, University of London, Criticism of Metric Enforced Redundancies Continues, *QMUL UCU Blog*, July 2012. http://qmucu.wordpress.com/2012/07/06/criticism-of-metric-enforced-redundancies-at-queen-mary-continues/

[Rafols *et al* 2010] Rafols I, Porter AL and Leydesdorff L, Science overlay maps: a new tool for research policy and library management, *Journal of the American Society for Information Science and Technology*, **61**(9), September 2010. http://onlinelibrary.wiley.com/doi/10.1002/asi.21368/full

[RCUK 2012a] Research Councils UK, International, Web page*, Research Councils UK Web site*, 2012. http://www.rcuk.ac.uk/international/Pages/International2.aspx

[RCUK 2012b] Research Councils UK, What do Research Councils mean by 'impact'? Web page, *RCUK Web site*, 2012. http://www.rcuk.ac.uk/kei/impacts/Pages/meanbyimpact.aspx

[REF 2011] Research Evaluation Framework, *Assessment framework and guidance on submissions*, REF 02.2011, Research Evaluation Framework 2014, July 2011. http://www.ref.ac.uk/pubs/2011-02/

[Rehn *et al* 2007] Rehn C, Kronman U and Wadskog D, *Bibliometric indicators – definitions and usage at Karolinska Institutet*, Appendix to the Bibliometric handbook for Karolinska Institutet, Version 1.0, Karolinska Institutet University Library, August 2007. http://kib.ki.se/sites/kib.ki.se/files/Bibliometric_indicators_definitions_1.0.pdf

[Rehn *et al* 2008] Rehn C and Kromman U, *Bibliometric handbook for Karolinska Institutet*, Version 1.05, Karolinska Institutet University Library, December 2008. http://ki.se/content/1/c6/01/79/31/bibliometric_handbook_karolinska_institutet_v_1.05.pdf

[RISE 2011] The RISE Project, RISE Measuring Success, *RISE Blog*, July 2011. http://www.open.ac.uk/blogs/RISE/2011/07/13/rise-measuring-success/

[Rotenberga and Kushmericka 2011] Rotenberga, A, and Kushmericka, A, The Author Challenge: Identification of Self in the Scholarly Literature, *Cataloging & Classification Quarterly*, **49**(6), September 2011. http://www.tandfonline.com/doi/full/10.1080/01639374.2011.606405

[Royal Society 2011a] Royal Society, *Knowledge networks and nations: Global scientific collaboration in the 21st Century*, Final Report, Royal Society, March 2011. http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/publications/2011/4294976134.pdf

[Royal Society 2011b] The Royal Society, *International collaboration as a proportion of national output 1996-2008*, Knowledge, Networks and Nations Motion Graph, Royal Society, March 2011. http://royalsociety.org/policy/reports/knowledge-networks-nations/motion-graph/

[Rosen et al 2011] Rosen, D, Barnett, GA, and Kim, JH, Social networks and online environments: when science and practice co-evolve, *SOCIAL NETWORK ANALYSIS AND MINING*, **1**(1), October 2011. http://www.springerlink.com/content/n7557l83h8473335/

[Russell 2012] Russell, R, *Adoption of CERIF in Higher Education Institutions in the UK: A Landscape Study*, UKLON, 2012. http://opus.bath.ac.uk/30979/

[RWR 2012a] Ranking Web of Repositories, English Home Page, *Ranking Web of Repositories web site*, 2012. http://repositories.webometrics.info/en

[RWR 2012b] Ranking Web of Repositories, English Objectives Page, *Ranking Web of Repositories web site*, 2012. http://repositories.webometrics.info/en/Objetives

[Schleyer *et al* 2012] Schleyer T, Butler B, Song M and Spallek H, Conceptualizing and advancing research networking systems, *ACM Transactions on Computer-Human Interaction*, **19**(1), 2012. http://dl.acm.org/citation.cfm?id=2147785

[Scimaps 2012a] HistCite™ Visualization of DNA Development, *Scimaps Web site*, 2012. http://scimaps.org/maps/map/histcite_visualizati_52/

[Scimaps 2012b] The Emergence of Nanoscience & Technology, *Scimaps Web site*, 2012. http://scimaps.org/maps/map/the_emergence_of_nan_121/ and http://scimaps.org/maps/map/the_emergence_of_nan_121/detail/

[Scopus 2012] Scopus, About SNIP, Web page, *Journal Metrics Web site*, Scopus, 2012. http://www.journalmetrics.com/index.php

[Small 1973] Small, HG, 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), 1973. http://onlinelibrary.wiley.com/doi/10.1002/asi.4630240406/abstract

[Stephens 2007] Stephens S, THE ENTEPRISE SEMANTIC WEB Techologies and Applications for the Real World, in Cardoso J*, Hepp M, Lytras MD (Eds.): The Semantic Web: Real-World Applications from Industry, Semantic Web And Beyond, Computing for Human Experience* , Vol. 6 Springer 2007, ISBN 978-0-387-48531-7, 2007. http://tinyurl.com/stephens2007

[Semantic Research 2012] Semantic Research, Products Page, *Semantic Research web site*, 2012. http://www.semanticresearch.com/solutions/products http://www.semanticresearch.com/solutions/semantica-features

[Scientometrics 2012], Scientometrics journal home page, *Scientometrics*, Springer. 2012. http://www.springer.com/computer/database+management+%26+information+retrieval/journal/11192

[Schmidt and Vosen 2011] Schmidt T and Vosen S, A Monthly Consumption Indicator for Germany Based on Internet Search Query Data, *Ruhr Economic Papers #208*, Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI), ISSN 1864-4872, October 2010. http://www.rwi-essen.de/publikationen/ruhr-economic-papers/309/

[Scott 1987] Scott, J, *Social Network Analysis: A Handbook*. Sage Publications, 1987.

[Stieglitz and Dang-Xuan 2012] Stieglitz, S, and Dang-Xuan, L, Social media and political communication: a social media analytics framework, *SOCIAL NETWORK ANALYSIS AND MINING*, July/August 2012. http://www.springerlink.com/content/xhxq7x174280461n/

[Tableau Software 2012] Tableau Software, Tableau Desktop, *Tableau Software web site*, 2012. http://www.tableausoftware.com/products/desktop

[Thewall 2009] Thelwall T, *Introduction to Webometrics: Quantitative Web Research for the Social Sciences*, Morgan & Claypool, ISBN 978-1-59829-993-9, 1990.

[Thelwall 2010] Thelwall, M, Webometrics: emergent or doomed? Proc Seventh International Conference on Conceptions of Library and Information Science, *Information Research*, **15**(4), December 2010. http://informationr.net/ir/15-4/colis713.html

[Thelwall 2012] Thelwall, M, A history of Webometrics, *Bulletin of the American Society for Information Science and Technology*, **38**(6), August/September 2012. http://www.asis.org/Bulletin/Aug-12/AugSep12_Thelwall.html and http://onlinelibrary.wiley.com/doi/10.1002/bult.2012.1720380606/abstract

[TR 2012a] THOMSON REUTERS, THE TOMSON REUTERS IMPACT FACTOR, Web page, *THOMSON REUTERS Web site*, 2012. http://thomsonreuters.com/products_services/science/free/essays/impact_factor/

[TR 2012b] THOMSON REUTERS, JOURNAL CITATION REPORTS, Web page, *THOMSON REUTERS Web site*, 2012. http://thomsonreuters.com/products_services/science/science_products/a-z/journal_citation_reports/

[TR 2012c] THOMSON REUTERS, HISTCITE, Web page, *THOMSON REUTERS Web site*, 2012. http://thomsonreuters.com/products_services/science/science_products/a-z/histcite/

[Valdivia and Monge-Corella 2010] Valdivia A, Monge-Corella S. Diseases tracked by using Google trends, Spain [letter, *Emerging Infectious Diseases*, **16**(1), January 2010 http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2874385/

[van Harmelen and Workman 2012] van Harmelen, M, and Workman, D, Analytics for Learning and Teaching, *CETIS Analytics Series*, **1**(3), 2012. http://publications.cetis.ac.uk/wp-content/uploads/2012/11/Analytics-for-Learning-and-Teaching-Vol1-No3.pdf

[White and Griffith 1981] White, HD, and Griffith, BC, Author cocitation: A literature measure of intellectual structure, *Journal of the American Society for Information Science*, **32**(3), May 1981. http://onlinelibrary.wiley.com/doi/10.1002/asi.4630320302/abstract

[Wikipedia 2012a] Wikipedia contributors, Impact Factor Criticisms, *Wikipedia*, 2012. http://en.wikipedia.org/wiki/Impact_factor#Criticisms

[Wikipedia 2012b] Wikipedia contributors, h-index Criticism, *Wikipedia*, 2012. http://en.wikipedia.org/wiki/H-index#Criticism

[WIPO 2012] World Intellectual Property Organisation Patent Cooperation Treaty Working Group, *THE SURGE IN WORLDWHDE PATENT APPLICATIONS*, Report, World Intellectual Property Organisation, May June 2012. http://www.wipo.int/edocs/mdocs/pct/en/pct_wg_5/pct_wg_5_4.pdf

[Zucca 2012a] Zucca, J, *Penn Libraries DATA FARM Testing What's Possible with Library Data*, Presentation, 2012 http://prezi.com/ntatay4nd0ng/copy-of-metridoc/

[Zucca 2012b] Barker, T, and Zucca, J, Metridoc: Extensible Infrastructure for Metrics and Analytics, *2012 Digital Library Federation Forum*, Digital Library Federation, 2012. http://www.diglib.org/forums/2012forum/metridoc-extensible-infrastructure-for-metrics-and-analytics/

About the Author

Dr Mark van Harmelen is the Director of Hedtek Ltd, where he performs IT strategy consultancy and directs the development of computer systems for clients and Hedtek's own product portfolio. Mark is also an Honorary Research Fellow in the University of Manchester's School of Computer Science. He has previously worked as a Lecturer at the University of Manchester, a Senior Researcher at Matushita Electric Industrial's Tokyo Research Centre, a consultant to industry, and a South African Cabinet appointee tasked with formulating the 2005 establishment of the Meraka Institute in South Africa.

Hedtek Ltd works for a variety of commercial and academic clients, developing IT strategy, performing research, and building highly usable computer systems that leverage the Web, cloud and mobile platforms. Hedtek also provides analytics consultancy to academia.

CETIS Analytics Series

http://publications.cetis.ac.uk/c/analytics

Acknowledgements

The author acknowledges help and assistance from Robert Burrill Donkin, Neil Jacobs, and David Kay. Dan Danielson, Liam Earny, Loet Leydesdorff, Ross MacIntyre, Geraint North, Joy Palmer, Norman Paton, Bijan Parsia, Stephen Pearson, Ismael Rafols, Catharina Rehn, Rosemary Russell, Titus Schleyer, Robert Stevens, Daniel Wadskog and Joe Zucca kindly answered diverse queries. Thanks are extended to all.

The CETIS Analytics Series was commissioned by Myles Danson (JISC programme manager) to give an overview of current thinking around analytics in post-16 education in the UK. In addition to the authors the following people have contributed to the production of the CETIS Analytics Series; Lorna Campbell (CETIS), Adam Cooper (CETIS), Rob Englebright (JISC), Neil Jacobs (JISC), Sheila MacNeill (CETIS) and Christina Smart (CETIS). Design by: http://www.consul4design.com

About this White Paper

Title: CETIS Analytics Series Vol. 1 No. 4: Analytics for Understanding Research

Author: Mark van Harmelen (Hedtek Ltd)

Date: November 2012

URI: http://publications.cetis.ac.uk/2012/518

ISSN 2051-9214

hedtek

Published by The University of Bolton

About CETIS

CETIS are globally recognised as leading experts on interoperability and technology standards in learning, education and training. We work with our clients and partners to develop policy and strategy, providing impartial and independent advice on technology and standards. CETIS are active in the development and implementation of open standards and represent our clients in national, European and global standards bodies and industry consortia, and have been instrumental in developing and promoting the adoption of technology and standards for course advertising, open education resources, assessment, and student data management, opening new markets and creating opportunities for innovation.

For more information visit our website: http://jisc.cetis.ac.uk/

The Analytics Series has been produced by CETIS for JISC: http://www.jisc.ac.uk/