Colledge, Lisa; Clements, Anna; Aisati, M'hamed el; Rutherford, Scott: *Project Snowball – sharing data for cross-institutional benchmarking.* In: Jeffery, Keith G; Dvořák, Jan (eds.): E-Infrastructures for Research and Innovation: Linking Information Systems to Improve Scientific Knowledge Production: Proceedings of the 11th International Conference on Current Research Information Systems (June 6-9, 2012, Prague, Czech Republic). Pp. 253-260. ISBN 978-80-86742-33-5. Available from: www.eurocris.org.

Project Snowball – sharing data for cross-institutional benchmarking

Lisa Colledge[a], Anna Clements[b],
M'hamed el Aisati[a], Scott Rutherford[c]
[a] Elsevier, Amsterdam, The Netherlands
[b] University of St Andrews, UK
[c] Queens University Belfast, UK

Summary

Project Snowball (www.projectsnowball.info) is a public service project that aims to help universities benchmark their performance across a broad range of research activities. The objectives of the project are to determine a standard set of common metrics and to share the methodology behind those metrics publically. With a set of common metrics defined by and adopted across UK institutions, it will be possible for them to make meaningful comparisons with each other. There is global interest in the opportunities that Project Snowball represents, and it remains to be seen how generally applicable the method that is being worked out for the UK will be.

1 The starting point: a study on managing research-related data

In 2010, Elsevier and Imperial College London participated in a joint JISC-funded study within the English higher education sector. The resulting report (Imperial College London, Elsevier and JISC, 2010) reviewed the sector's efforts and experiences of implementing research management systems. It evaluated and compared the tools that universities currently use to manage data related to research. It also aimed to identify problems with the current approaches used, and to publicise elements of good practice. The key recommendations were:

- Institutions should work more collaboratively to harmonise their approach to research management processes and to minimise wasteful duplication of investment in research management systems
- Institutions and funders should work more collaboratively to identify commonality in systems and processes so they might share data in more cost effective and less resource-intensive ways
- Institutions should develop stronger relationships with suppliers and work with them to define their needs more clearly
- A national framework for data and metric standards should be developed with stakeholders and used across the sector

- Suppliers should participate in the development of data standards with the sector in an effort to drive consistency in research systems
- Institutions, supported by funding organisations, should be encouraged to develop long-term system strategies focussed upon core research

The report highlighted considerable dissatisfaction with the data and tools available to integrate information from disparate systems, and a frustration that different stakeholders, including funders, demand similar information in differing formats with differing definitions. Institutions and funders have tended to implement their own bespoke solutions and examples of success from collaborative approaches are few and far between.

The study revealed that institutions and funders recognise data as an essential element in strategic management and decision making, but that there is a lack of consensus among stakeholders on the metrics that should be used for measurement and evaluation. The study showed that, without clear and shared metrics, institutions find that it is almost impossible to benchmark meaningfully and that as a result they are hampered in their ability to establish strategic direction. The lack of a shared definition of metrics makes it difficult for institutions to measure performance against peer institutions and plan for the long term. It mitigates against suppliers developing data systems that could be used across the sector as each institution wants different data structures and metrics.

The focus on and pressures of external data requests have meant that institutions have allowed the demands of other stakeholders to determine the data and the data-definitions they collect and measure, rather than considering what would be best suited to their own purposes.

The results of the study were presented at a sector-wide workshop held at Imperial College London in August 2010. Stakeholders at this workshop agreed that it was essential that the issues and opportunities for improvement identified by the study be addressed, and that the recommendations made in the study should form the basis of further work. Many who participated in either the study or the results workshop voiced their belief that work needed to be done to harmonise research information systems, and that the study should not gather dust on the shelf and have no practical application. Therefore, following the completion and publication of the report, the Imperial College/Elsevier team initiated a second phase of work that would address and build on the recommendations.

2 Second phase: Snowball – agreeing on benchmarking metrics for UK research performance

The Snowball Project (www.projectsnowball.info) has a tightly defined overall goal: to facilitate external benchmarking by ensuring that institutions can confidently compare research data in a like-with-like manner. It aspires to make the practice of research information management easier and more efficient, by facilitating and driving the storage and provision of meaningful data for use by higher education institutions and other stakeholders. Achieving a consistent understanding of metrics will simplify interaction between institutions, suppliers and funders, enable more effective benchmarking, and improve the efficiency of data management.

The scope of the second phase project was expanded beyond England to ensure that consensus was as broadly relevant as possible. Eight universities were approached and agreed to collaborate on the project. These are: Imperial College London, Queen's University Belfast, University Col-

lege London, University of Bristol, University of Cambridge, University of Leeds, University of Oxford, and University of St Andrews. The Snowball team is comprised of the core Elsevier/Imperial College team that conducted the initial JISC-funded study and representatives from each of the Snowball project partners, typically both directors of research offices or senior staff responsible for research management and strategy, as well as technical specialists with expertise in data systems, structure and use.

2.1 The landscape of metrics desired by institutions for the establishment of their strategic direction

Previous research (Green and Langley, 2009) has shown that university research management offices are asked to perform a multitude of functions. The first task of the Snowball team was therefore to define the landscape of research activities that would be included in the project's scope. This included consideration of *inputs, throughputs, and outputs* of the research process, and identified broad sets of measures that were associated with each. Inputs, throughputs, and outputs were further distinguished in three ways: those connected to *research grants, postgraduate education, and enterprise activities*. The Snowball team also identified a common set of denominators that could be applied to metrics at various levels of aggregation, such as principal investigator, unit of assessment, or funding body categories or themes. The metric landscape is shown in *Figure 1*.

	Research Inputs	Research Process	Research Outputs/ Outcomes
Research Grants	1a) Research applications 1b) Research awards • Price/ Overhead recovery • Philanthropy	2a) Research spend (income) • Space utilisation • Staff recruitment • Start/ end date slippage	3a) Publications & citations 3b) Esteem measures 3c) Collaboration (co-authorship) • Socio-economic Impact
Post Graduate Education	1c) PGR volumes • PGT volumes • International PGT volumes • UG to PG conversion rates	• PG Experience – contact time • PG Experience - facilities	• Alumni/ destination of leavers 3d) Completion rates • Skills development (impact)
Enterprise activities	1d) Industrial income • Industry engagement	• Contract turnaround times • Industry research spend (income) • UG to PG conversion rates	3e) Patenting 3f) Licensing income 3g) Spin-out generation / income • KTPs numbers • Consultancy income

Denominators	(Number of) People	Organisations	Themes/ Schemes
	• Researcher, authors • Principal / Co-investigators • Academic staff by category • Research assistants • PGR Students • UG / PGT Students • Post doctoral staff • Support staff	• Institution • Faculty • Department / School • Unit of Assessment (UoA) • Groups / clusters • Funders by type: RC etc • Centres / Institutes	• Standard grants • Strategic initiatives (Calls) • Grand challenges • Subject areas • Keywords

Figure 1: Definition of the institutional landscape of research activities

Colledge, Lisa; Clements, Anna; Aisati, M'hamed el; Rutherford, Scott: *Project Snowball – sharing data for cross-institutional benchmarking.* In: Jeffery, Keith G; Dvořák, Jan (eds.): E-Infrastructures for Research and Innovation: Linking Information Systems to Improve Scientific Knowledge Production: Proceedings of the 11th International Conference on Current Research Information Systems (June 6-9, 2012, Prague, Czech Republic). Pp. 253-260. ISBN 978-80-86742-33-5. Available from: www.eurocris.org.

2.2 Data collection experiment and lessons learnt from Snowball prototype

Having established the landscape, the Snowball team decided to try to gather the data and calculate and deliver metrics to the Snowball project partners. It was agreed that each participating institution would collect and contribute data on ten anonymised researchers in chemistry, and that Elsevier would contribute Scopus data on each identified researcher to enable the completion of publication, citation, and collaboration metrics. Snowball project partners were asked to complete an Excel-based template of data components for each metric, which was fed to developers at Elsevier, who constructed a simple prototype analytical benchmarking tool in the form of a basic dashboard (*Figure 2*).

It was clear from this data collection experiment and the development of the prototype that there was strong support for the concept of consistently defined, standardised metrics to enable cross-institutional benchmarking from common data sources, with analytical tools on top. This experiment reinforced the need to integrate data from different sources to increase the scope of the metrics that can be generated, but this method of data collection employed was a struggle and not scalable. Despite the significant challenges identified, all involved strongly endorsed the concept of an analytical tool that enables comparison and benchmarking, and expressed a desire to continue participation in Project Snowball while benefitting from the following key lessons:

- *The availability of data:* Snowball partner institutions were mostly able to provide application and award data, postgraduate research volumes, and completion rates, but data on industrial consultancy, patenting and licensing, and esteem measures were either not readily available within their research management systems, or institutions were not prepared to share this potentially sensitive information. The Snowball team agreed that alternative methods and sources to capture and reflect patenting and licensing, and esteem measures needed to be sought.
- *Manual labour in data collection:* Where data were available, it often took considerable effort to provide it in the requested format. Most Snowball project partners found it difficult and labour-intensive to gather all funding data for just ten researchers. The Snowball team agreed that on a larger scale an automated or decentralized approach would be necessary and that where possible the group should consider working directly with potential central data providers.
- *Definitions:* The Snowball partner institutions have in-depth knowledge of research administration and engaging with industry through patenting, licensing, and spin-outs, but mechanisms need to be identified so that data can be reported in a suitable way. The Snowball team agreed to form a working group of experts to refine and agree definitions, and to eliminate inconsistencies.
- *Confidentiality:* Some Snowball partner institutions chose not to link funding data to researchers despite a confidentiality agreement that information on individual researchers would only be available to the institution with which that individual was connected. Furthermore, commercial confidentiality prevented universities from reporting on patenting and licensing metrics. The Snowball team agreed that they should consider strengthening legal assurances, while continuing to gather metrics on an opt-in basis.

Figure 2: Screenshots from the simple prototype benchmarking dashboard

Colledge, Lisa; Clements, Anna; Aisati, M'hamed el; Rutherford, Scott: *Project Snowball – sharing data for cross-institutional benchmarking.* In: Jeffery, Keith G; Dvořák, Jan (eds.): E-Infrastructures for Research and Innovation: Linking Information Systems to Improve Scientific Knowledge Production: Proceedings of the 11th International Conference on Current Research Information Systems (June 6-9, 2012, Prague, Czech Republic). Pp. 253-260. ISBN 978-80-86742-33-5. Available from: www.eurocris.org.

2.3 Current focus

The Snowball team has felt that the partnership approach is working well and has already been valuable and productive. The team is now looking to identify and implement scalable and less labour-intensive methods to integrate data from different sources and increase the scope of metrics that can be generated, for entire institutions rather than for a handful of researchers.

The Snowball project partners recognised that in order to proceed, a team with more detailed knowledge of the data would be needed. A *subgroup of data and systems experts* familiar with day-to-day questions about data and systems requirements, with members from each of the partner universities, was therefore created, with the aim of establishing a consensus on the methods by which each metric should be calculated and by which denominators it should be displayed. Over a series of meetings between May and August 2011, this group consolidated and refined the list of metrics, and defined with complete clarity how they should be calculated and sourced. As a result, by the end of August 2011, agreement had been reached on almost all of the metrics identified as priorities by the Snowball team.

Dashboards that enable external benchmarking against identified and defined metrics are being created for Snowball partner institutions. While collecting data for entire institutions, calculating priority metrics, and generating benchmarks, this group of data and systems experts continues to advise the Snowball team and to revisit the metrics definitions as necessary. The group also aims to bridge inconsistencies across the sector and to ensure that existing national standards, such HESA cost centres and CERIF, are adopted within the Snowball project wherever possible and sensible.

2.4 Next steps

The Snowball Project aims to establish a UK sector-wide standard for metrics, with consensus on definitions and institutional performance metrics. It aims, ultimately, to enable all UK higher education institutions and other key stakeholders to develop the capability to deploy a set of standardised metrics for benchmarking and reporting purposes. A central repository that is open to everyone will ensure that project documentation and reports are readily available to everyone facing the challenges of, and/or interested in, external benchmarking. This will be accompanied by the consensus definitions, identified data sources, and a roadmap to enable the adoption of the project's vision across the sector with guidance on how others can use this roadmap to achieve the Snowball vision themselves.

Interest in Project Snowball is widespread, and is not limited to the UK. The situation of institutions and funders tending to work in isolation, leading to duplicated effort, systems that are manually intensive, and inefficiency overall, exists in all research-intensive countries around the world. In the United States, for example, the recently launched STAR METRICS project (www.starmetrics.nih.gov) addresses the need to establish and measure how higher education institutions spend government grants. The project aims to create a repository of data and tools that will be useful in assessing the impact of federal R&D investments, a goal akin to that of the United Kingdom's research councils. Other, similar projects are underway in Europe.

Ongoing research shows that, for institutional strategic planning, the Snowball Metrics are globally considered comprehensive and relevant. Encouragingly, a recent survey of members of Elsevier's Research Administrators *Innovation Explorers Community* found a high level of agreement

that the metrics developed by the UK Snowball project team were relevant outside the UK (*Figure 3*).

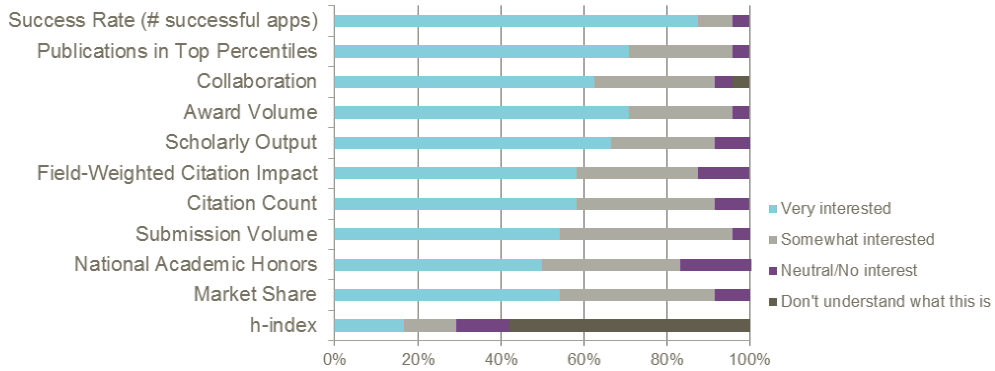

Figure 3: Survey results from Innovation Explorers Research Administrators Community

There is a global desire to be able to benchmark across peer institutions, but data and/or metrics sharing is limited to compulsory returns to national evaluation programmes or funding agencies, and very rarely takes place, if at all, between institutions. It is clear that there is a strong appetite for more detailed research intelligence that applies to diverse universities, and for more sophisticated data tools and systems, that extends beyond the United Kingdom, and that institutions are looking for help to establish metrics that apply to diverse universities, and share data across them. It remains to be seen how generally applicable across the globe the method that is being worked out by the Project Snowball team will be.

References

Green, J. and Langley, D. (2009), *Professionalising Research Management*. Available at www.researchdatatools.com/downloads/2009-professionalising-research-management-2.pdf

Imperial College London, Elsevier and JISC (2010), *Research information management: Developing tools to inform the management of research and translating existing good practice* (freely available at www.projectsnowball.info)

Colledge, Lisa; Clements, Anna; Aisati, M'hamed el; Rutherford, Scott: *Project Snowball – sharing data for cross-institutional benchmarking.* In: Jeffery, Keith G; Dvořák, Jan (eds.): E-Infrastructures for Research and Innovation: Linking Information Systems to Improve Scientific Knowledge Production: Proceedings of the 11th International Conference on Current Research Information Systems (June 6-9, 2012, Prague, Czech Republic). Pp. 253-260. ISBN 978-80-86742-33-5. Available from: www.eurocris.org.

Contact Information

Dr Lisa Colledge
Snowball Project Director
Elsevier
Radarweg 29
1043 NX Amsterdam
The Netherlands
l.colledge@elsevier.com

M'hamed el Aisati
Head of Product Technology and Analytical Services
Elsevier
Radarweg 29
1043 NX Amsterdam
The Netherlands
m.aisati@elsevier.com

Anna Clements
Enterprise Architect
University of St Andrews
Butts Wynd
St Andrews
KY16 9AL
UK
akc@st-andrews.ac.uk

Scott Rutherford
Director of Research and Enterprise
Research & Enterprise
Lanyon North
Queen's University Belfast
Belfast
BT7 1NN
UK
s.rutherford@qub.ac.uk