

# The Snowball Project

## Agreeing metrics for research information management

Dr Lisa Colledge, [l.colledge@elsevier.com](mailto:l.colledge@elsevier.com)

Imperial College  
London



UNIVERSITY OF  
CAMBRIDGE



# Background

- Growing recognition of need for research intelligence and performance management frameworks and metrics
- Frustration over tools available
  - Duplication of effort, manually intensive systems, and overall inefficiency
  - Dissatisfaction with data and tools available to integrate information
  - An appetite for more detailed research intelligence and for more sophisticated data tools and systems
- Frustration over availability of metrics to make sensible measurements
  - Different stakeholders demand similar information differently
  - Limitations of external benchmarking through inconsistent definitions and calculations

# Research information management UK study, 2010

- [www.researchdatatools.com/downloads/2010-research-information-management-2.pdf](http://www.researchdatatools.com/downloads/2010-research-information-management-2.pdf)
- Joint Imperial-Elsevier JISC-funded study of research information management
  - Institutions should work more collaboratively with each other
  - Institutions and funders should work together to identify commonality in systems and processes
  - Institutions should develop stronger relationships with suppliers
  - An agreed national framework for data and metric standards is needed
  - Suppliers should participate in the development of data and metric standards
  - Institutions should be encouraged to develop long-term system strategies focussed on core research management processes and information needs

- **Institutions and funders** should work more collaboratively to identify commonality in systems and processes, so they may share data in more cost-effective and less resource-intensive ways
- **Institutions and funders** should be encouraged to develop long-term system strategies focussed upon core research management processes and information needs

**FUNDERS**

**INSTITUTIONS**

**A national framework  
for data standards should  
be developed with  
stakeholders, and used  
across the sector**

**SUPPLIERS**

- **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** should develop stronger relationships with **suppliers** and work with them to define their needs more clearly

- **Suppliers** should participate in the development of data standards with the sector in an effort to drive consistency in research systems

**Figure 1:** Summary of recommendations from *Research information management: Developing tools to inform the management of research and translating existing good practice* (2010)

# Benchmarking

- With no holistic approach, it is not unusual for institutions to submit different information for the same data point in various external data-gathering exercises
  - Impossible to benchmark meaningfully without clear and shared data elements and metrics driven by institutions
- Institutions have allowed the demands of external stakeholders to determine the data and the data-definitions they collect and measure
- Benchmarking requires the combination of data from institutions, suppliers (proprietary) and third parties

# Second phase project: Snowball

- **Overall goal:** to facilitate external benchmarking by ensuring that institutions can confidently benchmark against their peers
- Self-funded, voluntary project
- Aims of second phase
  - Institutions define metrics and consensus method of calculation
  - Define all possible sources of the data elements for the metrics
  - Establish a three-year roadmap for ‘snowball effect’ to ensure adoption in sector

**Imperial College  
London**



**UNIVERSITY OF  
CAMBRIDGE**



# Define the landscape of research activities




	1. RESEARCH INPUTS	2. RESEARCH PROCESS	3. RESEARCH OUTPUTS/OUTCOMES
• Research Grants	<p>1a) Research applications</p> <p>1b) Research awards</p> <ul style="list-style-type: none"> <li>• Price / Overhead recovery</li> <li>• Philanthropy</li> </ul>	<p>2a) Research spend (income)</p> <ul style="list-style-type: none"> <li>• Space utilisation</li> <li>• Staff recruitment</li> <li>• Start / end date slippage</li> </ul>	<p>3a) Publications and citations</p> <p>3b) Esteem measures</p> <p>3c) Collaboration (co-authorship)</p> <ul style="list-style-type: none"> <li>• Socio-economic impact</li> </ul>
• Post Graduate Education	<p>1c) PGR volumes</p> <ul style="list-style-type: none"> <li>• PGT volumes</li> <li>• International PGT volumes</li> <li>• UG to PG conversion rates</li> </ul>	<ul style="list-style-type: none"> <li>• PG Experience – contact time</li> <li>• PG Experience – facilities</li> </ul>	<ul style="list-style-type: none"> <li>• Alumni / destination of leavers</li> </ul> <p>3d) Completion rates</p> <ul style="list-style-type: none"> <li>• Skills development (impact)</li> </ul>
• Enterprise activities	<p>1d) Industrial income</p> <ul style="list-style-type: none"> <li>• Industry engagement</li> </ul>	<ul style="list-style-type: none"> <li>• Contract turnaround times</li> <li>• Industry research spend (income)</li> <li>• UG to PG conversion rates</li> </ul>	<p>3e) Patenting</p> <p>3f) Licensing income</p> <p>3g) Spin-out generation / income</p> <ul style="list-style-type: none"> <li>• KTPs numbers</li> <li>• Consultancy income</li> </ul>
4) DENOMINATORS	<p>4A. (NUMBER OF) PEOPLE</p> <ul style="list-style-type: none"> <li>• Researcher, authors</li> <li>• Principal / Co-investigators</li> <li>• Academic staff by category</li> <li>• Research assistants</li> <li>• PGR Students</li> <li>• UG / PGT Students</li> <li>• Post doctoral staff</li> <li>• Support staff</li> </ul>	<p>4B. ORGANISATIONS</p> <ul style="list-style-type: none"> <li>• Institution</li> <li>• Faculty</li> <li>• Department / School</li> <li>• Unit of Assessment (UoA)</li> <li>• HESA cost centre</li> <li>• Groups / clusters</li> <li>• Funders by type: RC etc.</li> <li>• Centres / Institutes</li> </ul>	<p>4C. THEMES/ SCHEMES</p> <ul style="list-style-type: none"> <li>• Standard grants</li> <li>• Strategic initiatives (Calls)</li> <li>• Grand challenges</li> <li>• Subject areas</li> <li>• Keywords</li> </ul>

# Data collection experiment

- Each of the Snowball partner institutions would collect and contribute data on ten anonymised researchers
- Key challenges institutions faced:
  - Data were not readily available
  - The request had to be completed manually
  - Data were spread across multiple departments and/or systems with different ownerships within the institution and therefore permissions were needed to access it
  - The time period to gather the data was too short
  - Some concerns about confidentiality, especially in relation to third stream (i.e. commercial) activity; spin-out, patenting, and licensing information was viewed as commercially confidential
  - Engagement with industry was difficult to report as some of such activities were not mapped to the researchers involved

**Figure 5:** Results of data collection experiment

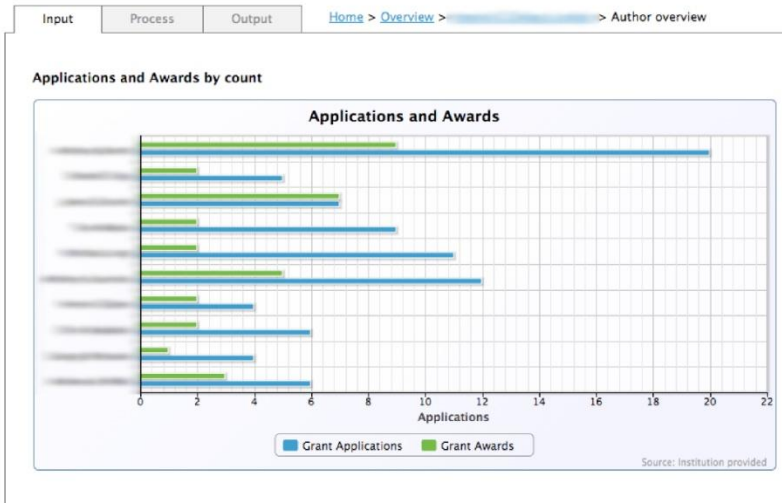
Data request			HEI								
Section	Corresponding metrics	Measure category	1	2	3	4	5	6	7		
1		Identifying researchers	Provided in a consistent way	Provided in a consistent way	Provided, some work needed	Provided in a consistent way	Provided in a consistent way	Provided in a consistent way		*	
2	1.a.i-1.b.iv	Research applications and awards	Provided, some work needed	Provided, some work needed	Provided, some work needed	Provided, some work needed	Provided, some work needed	Provided, some work needed		*	
ad 2		Metrics per funder	Provided, some work needed	Not provided	Not provided	Not provided	Not provided	Not provided		*	
3	1.c.1-1.c.iv	PGR volumes	Provided in a consistent way	Provided, some work needed	Not provided	Provided, some work needed	Not provided	Provided, some work needed		*	
4	1.d.i-1.d.iii	Industrial income	(Included in other categories)								*
5	1.e.i	Price / Overhead recovery	Provided in a consistent way	Not provided	Provided, some work needed	Provided in a consistent way	Not provided	Not provided		*	
6	1.f.i	Industry engagement	Provided in a consistent way	Not provided	Provided in a consistent way	Provided, some work needed	Provided in a consistent way	Not provided		*	
7	2.a.i-2.a.iv	Research spend (income)	Provided, some work needed	Provided in a consistent way	Not provided	Not provided	Not provided	Not provided	Provided in a consistent way		*
	(provided by Elsevier)	Publication and citation data	Provided in a consistent way	Provided in a consistent way	Provided in a consistent way	Provided in a consistent way	Provided in a consistent way	Provided in a consistent way			
8	3.b.i-3.b.ix	Esteem measures	Not provided	Provided in a consistent way	Not provided	Provided, some work needed	Provided, some work needed	Not provided			
9	3.d.1-3.d.iii	Completion rates	Provided in a consistent way	Provided in a consistent way	Not provided	Not provided	Not provided	Not provided			
10	3.e.i	Patenting	Not provided	Not provided	Not provided	Not provided	Not provided	Not provided		**	
11	3.f.i-3.f.ii	Licencing	Not provided	Not provided	Not provided	Not provided	Not provided	Not provided		**	
12	3.g.i-3.g.iii	Spin-out generation / income	Not provided	Not provided	Not provided	Not provided	Not provided	Not provided		**	

 Provided in a consistent way  
 Provided, some work needed  
 Not provided

\* Data available to capture, but a lot of labour involved

\*\* Difficult to capture due to unaligned systems and commercial confidentiality

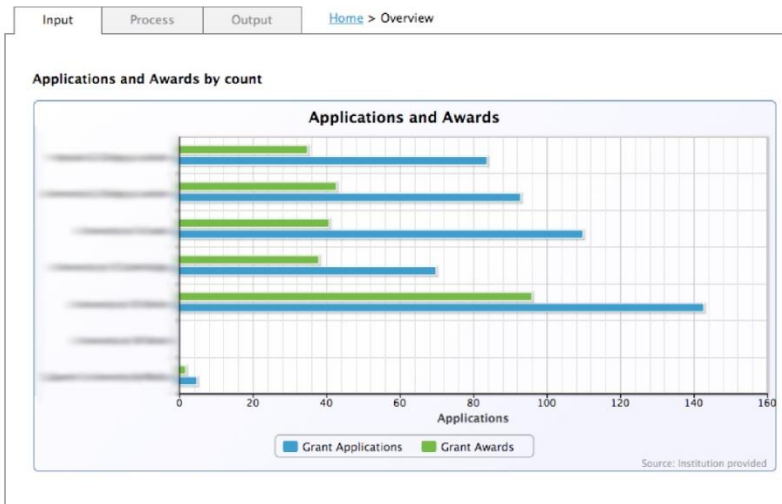
Figure 6: Prototype benchmarking tool



a) Application count by year (of request) by institution



c) Publication count by institution



b) Application count by year (of request) by researcher



d) Publication count by researcher

# Key lessons

- Key lessons taken from the data collection experiment included:
  - The availability of data
  - Manual labour in data collection
  - Definitions
  - Confidentiality
- Affirmation of Snowball vision:
  - "very worthwhile"
  - "the right thing to be doing"
  - "hugely valuable"

# Key conclusions

- Strong support for consistently defined, standardised metrics to enable cross-institutional benchmarking from common data sources, with analytical tools on top
- There is a strong need to integrate data from different sources to increase the scope of the metrics that can be generated
- The 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 between institutions and across denominators

# Current activity

- Expert group has reached consensus on definitions of data elements and method of calculation of priority metrics and denominators
- Working on a scalable way of building priority metrics for all researchers in partner institutions
  - Initially with three of the eight institutions with experts advising to ensure approach is scalable
- Developing relations with funders to source data
- Using data from all three types of data sources: institutional, proprietary and third party

# Sample metric: success rate

**Description:** proportion of all applications on which a decision has been recorded in HEI system that were successful

**Exclude:**

- Applications without a decision
- Applications of collaborative partners
- Outlines, expressions of interest, extensions, supplements, studentships, Doctoral Training Centres / Grants / Awards
- Industry funding

**Read out:**

- HEI financial year (August-July) and rolling 12 month period
- Absolute value, number and %
- Normalised by researcher

**Denominators:**

- HESA cost centre and institution
- Funder type and individual funding bodies (RCs aggregated as well as individually)

# Medium term aims

- Establish a UK sector-wide standard for institutional performance metrics
  - Come to shared, agreed understanding of what institutions, funders, and other stakeholders wish to measure and benchmark against
- Enable all UK HEIs and other key stakeholders to develop the capability to deploy standardised metrics for benchmarking and reporting
  - Snowball methodology for calculating metrics and the framework that they sit in will be made freely available across the sector
- Work with entities that hold data on multiple institutions
- Easier collaborations with suppliers and more effective and efficient data management
- Regular updates published – first output in press now

[www.projectsnowball.info](http://www.projectsnowball.info) (coming soon...)

Lisa Colledge - [l.colledge@elsevier.com](mailto:l.colledge@elsevier.com)

John Green - [j.green@imperial.ac.uk](mailto:j.green@imperial.ac.uk)

Nick Fowler - [n.fowler@elsevier.com](mailto:n.fowler@elsevier.com)

Imperial College  
London



UNIVERSITY OF  
CAMBRIDGE



University of  
St Andrews