

The Snowball Project

Agreeing metrics for research
information management

John Green, Imperial College London

ESF Member Organisation Forum

Evaluation of publicly funded research

4th Workshop, Bern, Switzerland: 7-8 November 2011

Imperial College
London

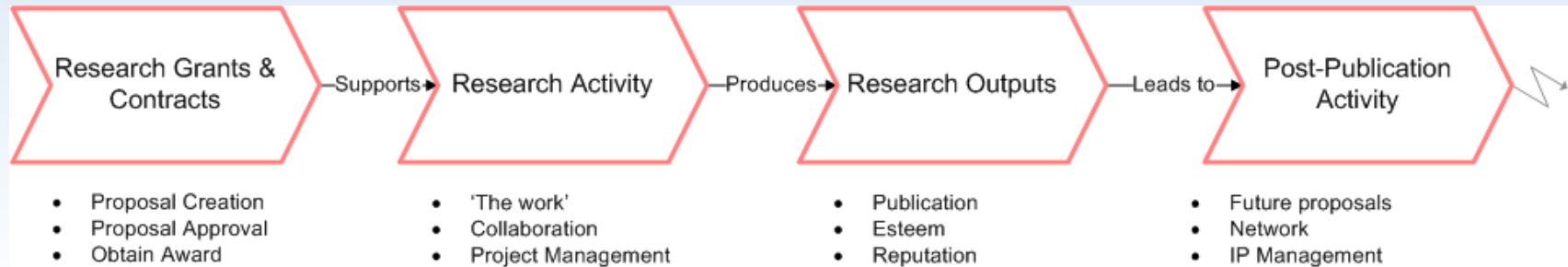


UNIVERSITY OF
CAMBRIDGE



Underlying systems for managing research

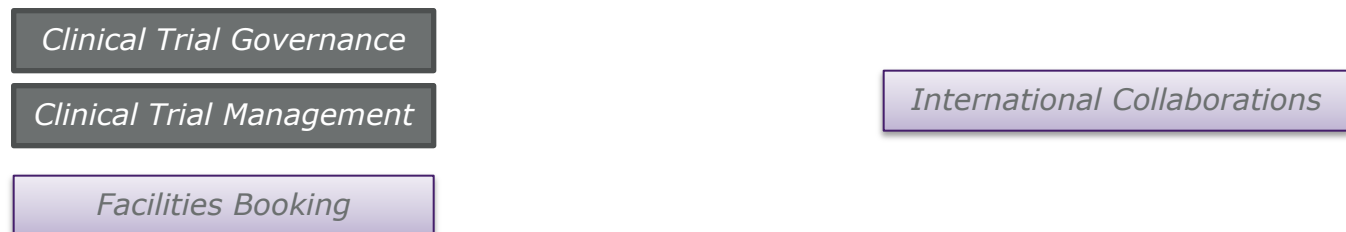
The Research Process



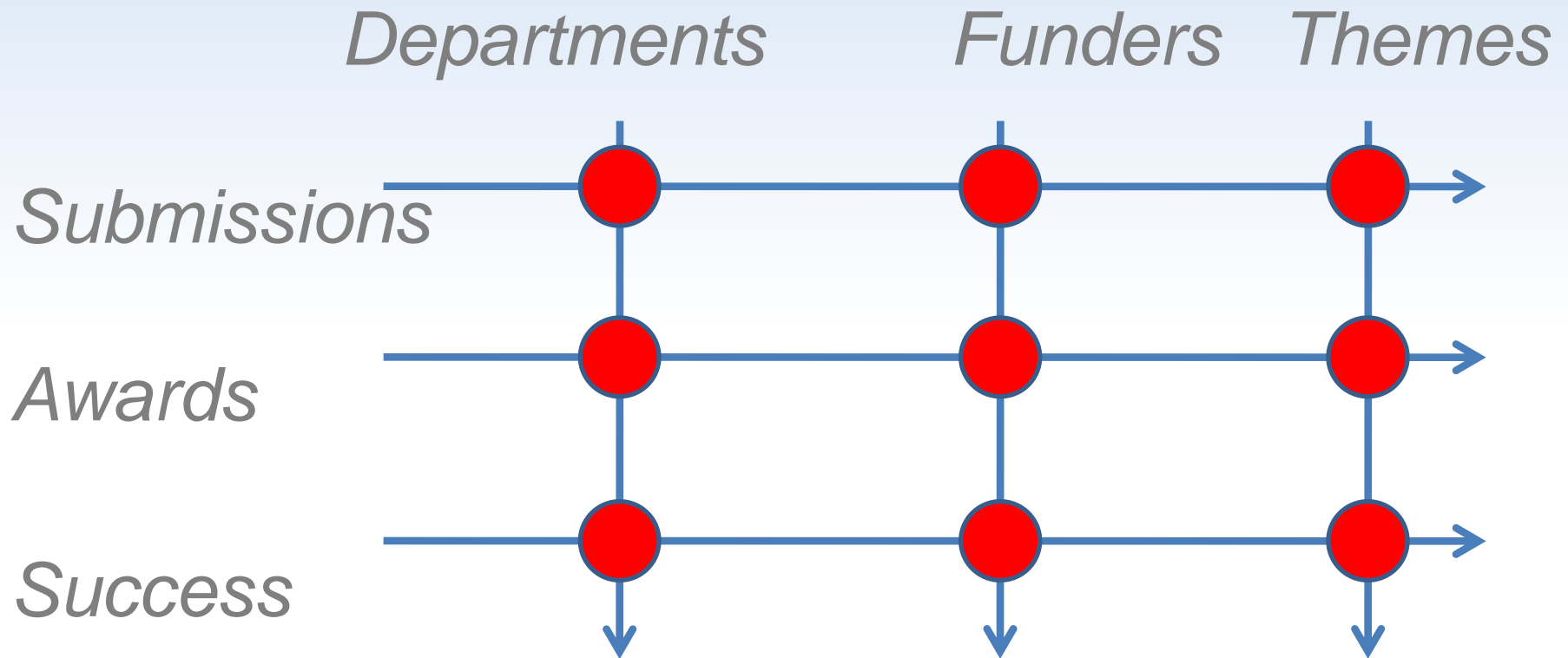
Imperial's Core Research Systems




Optional / Specialist Systems

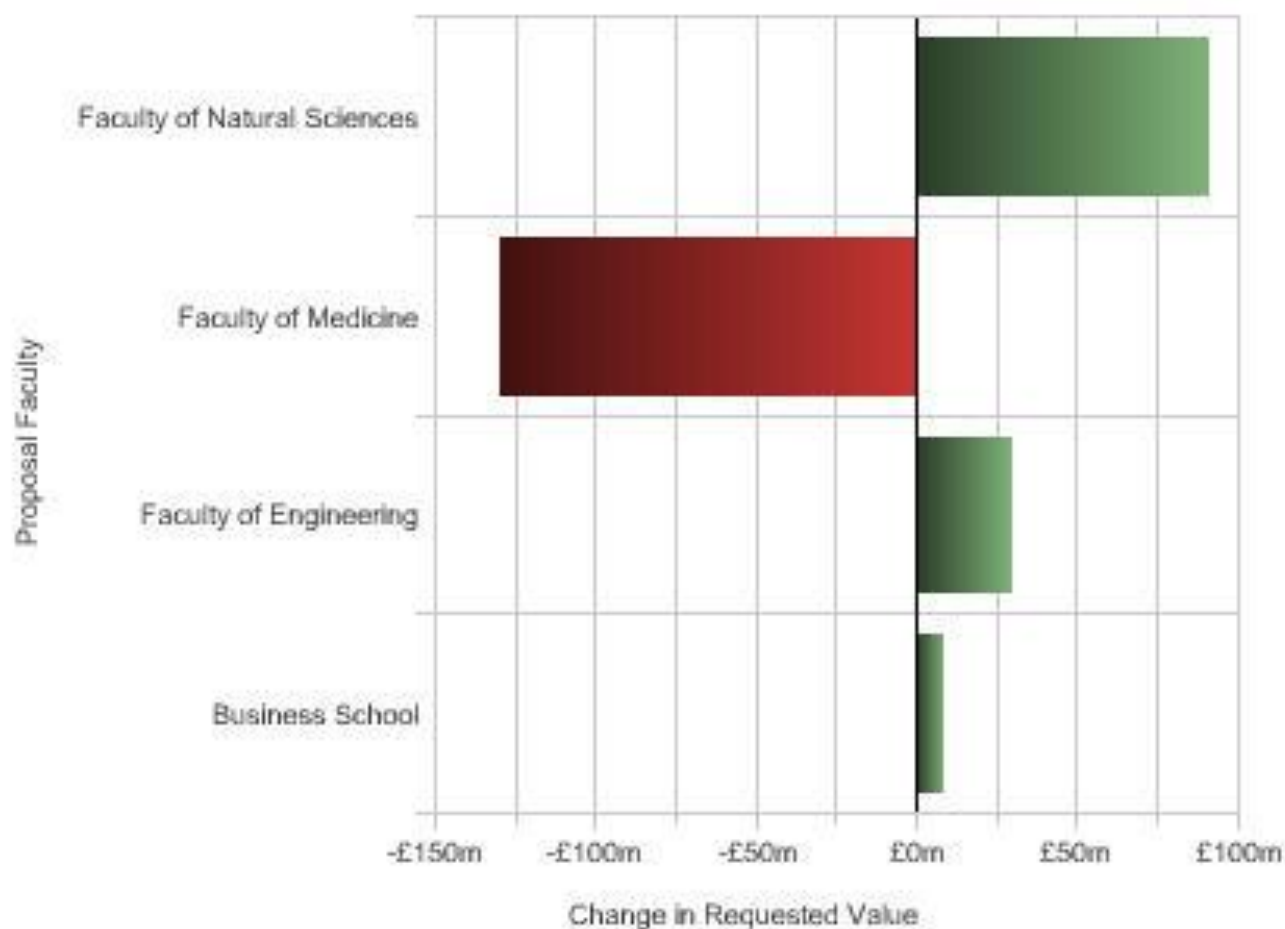


Dashboard layout



Faculty Submission Volumes

Select Analysis Type 

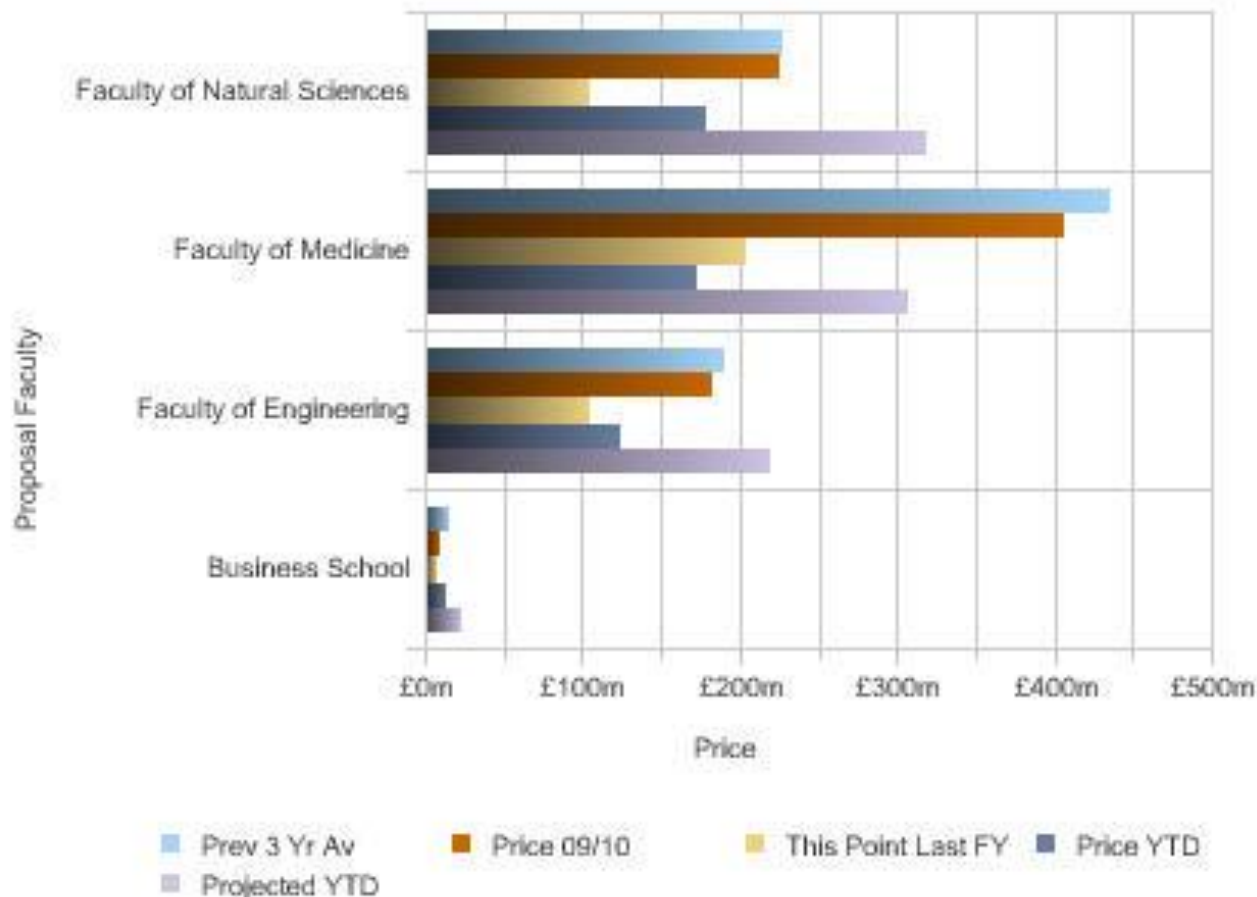


The chart displays the difference in total requested value between the projected current year total and the previous 3-year average.

Source: InfoEd

Faculty Submission Volumes

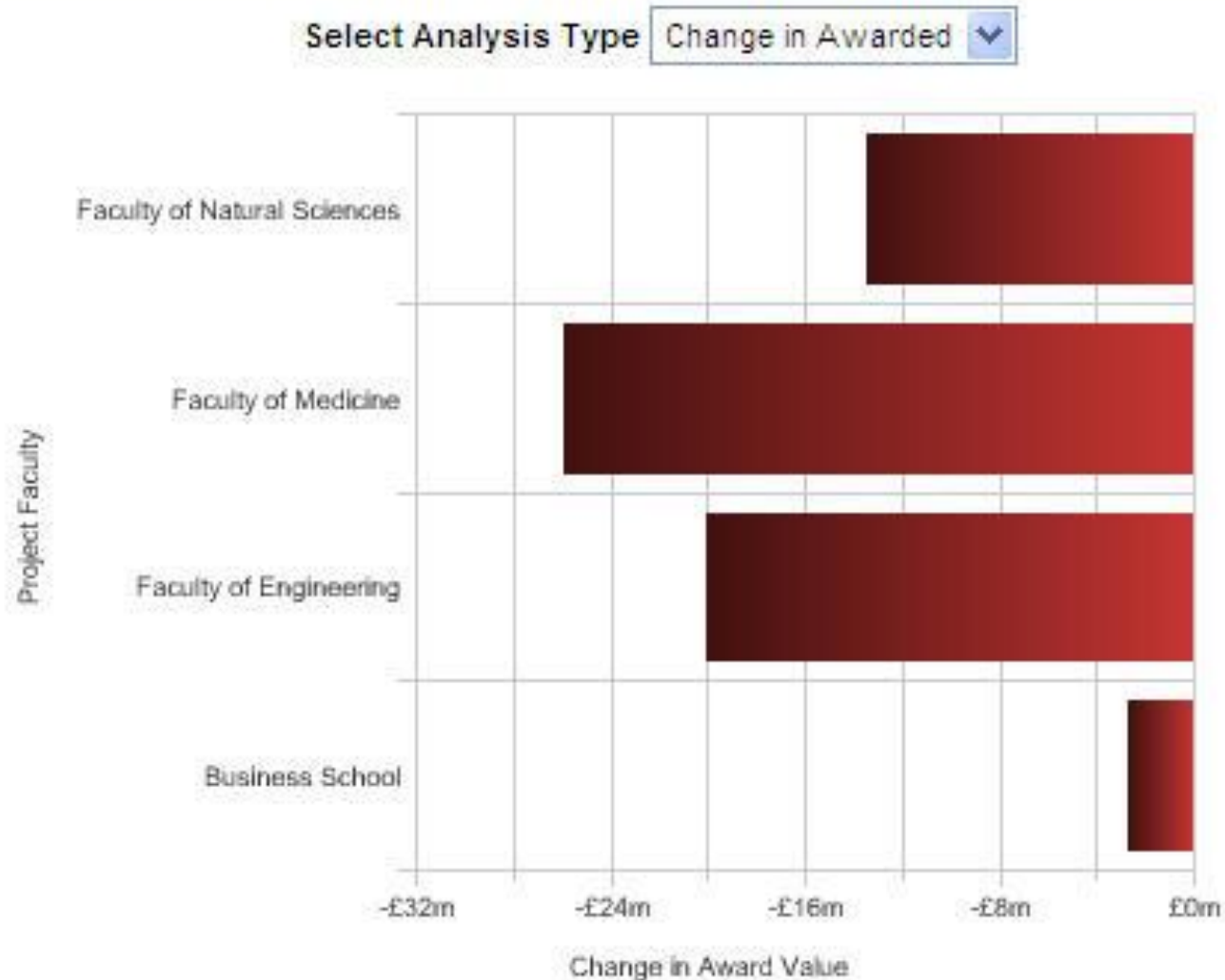
Select Analysis Type



The graph displays the requested amount for the previous three-year average, the previous financial year total, the amount at this point in the previous financial year, the current financial year to date and the linear projected total for the current year.

Source: InfoEd

Faculty Award Volumes

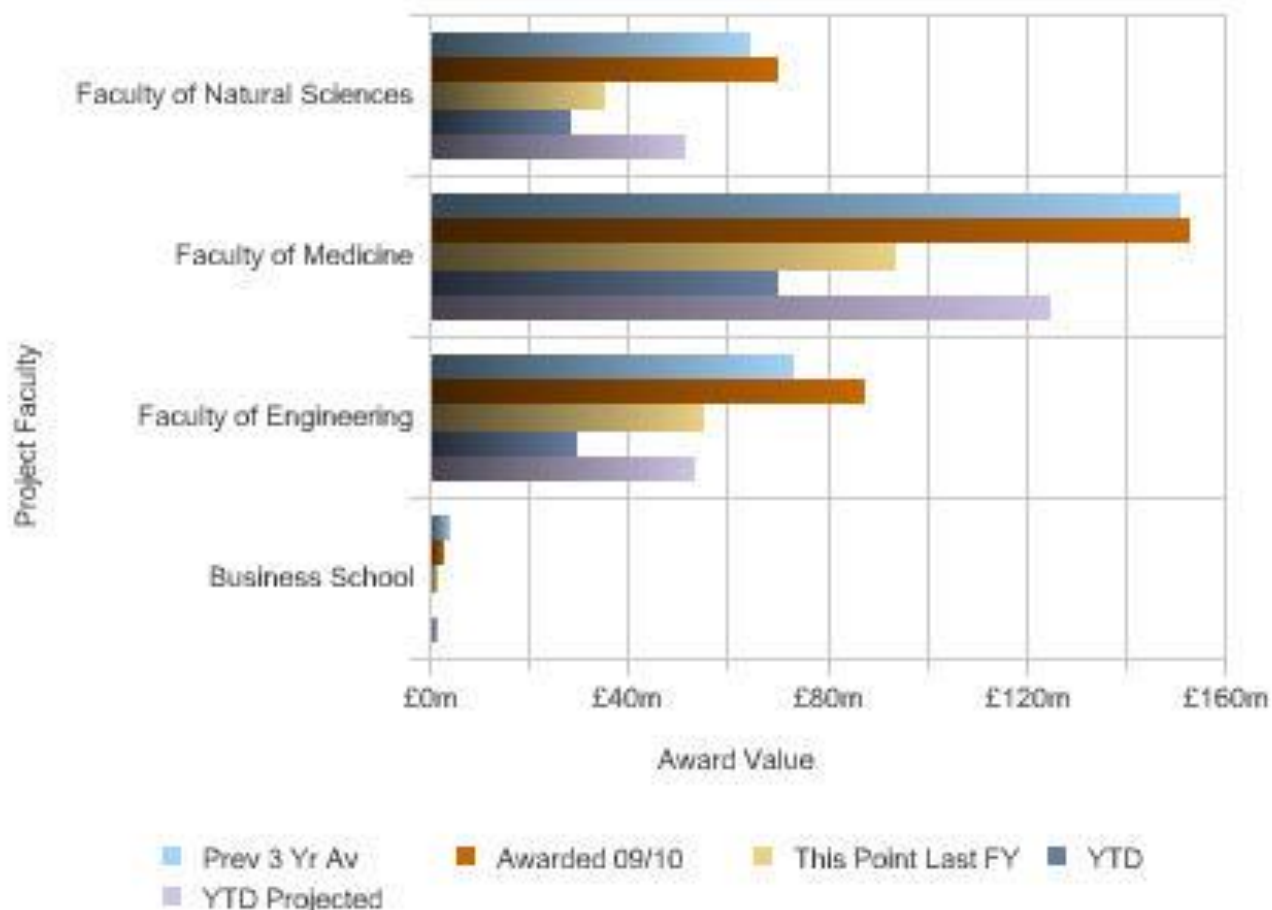


The chart shows the difference in total awarded amount between the projected current year total and the previous 3-year average.

Source: GMS

Faculty Award Volumes

Select Analysis Type

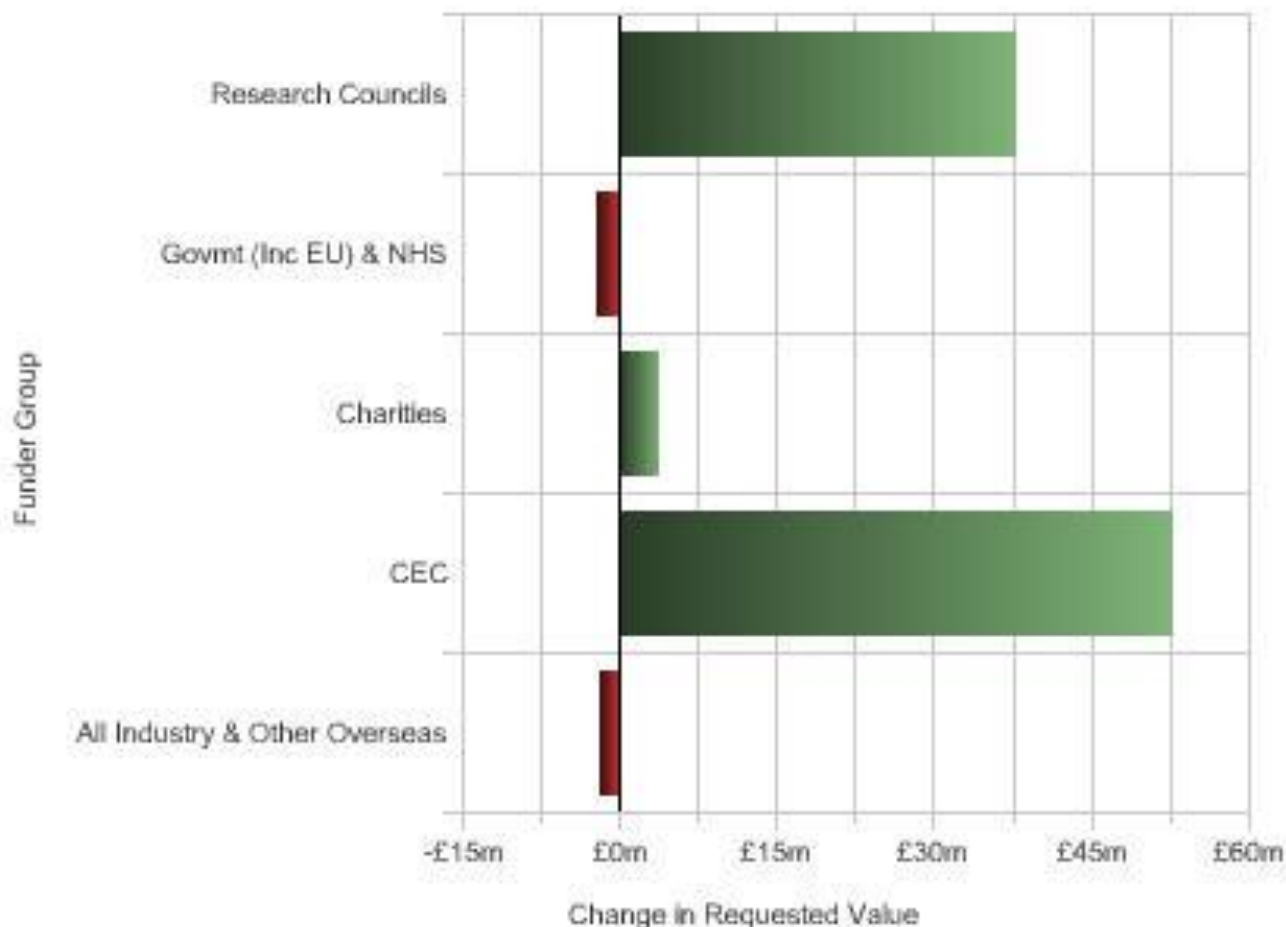


The graph displays the awarded amount for the previous three-year average, the previous financial year total, the amount at this point in the previous financial year, the current financial year to date and the linear projected total for the current year.

Source: GMS

Sponsor Submission Volumes

Select Analysis Type

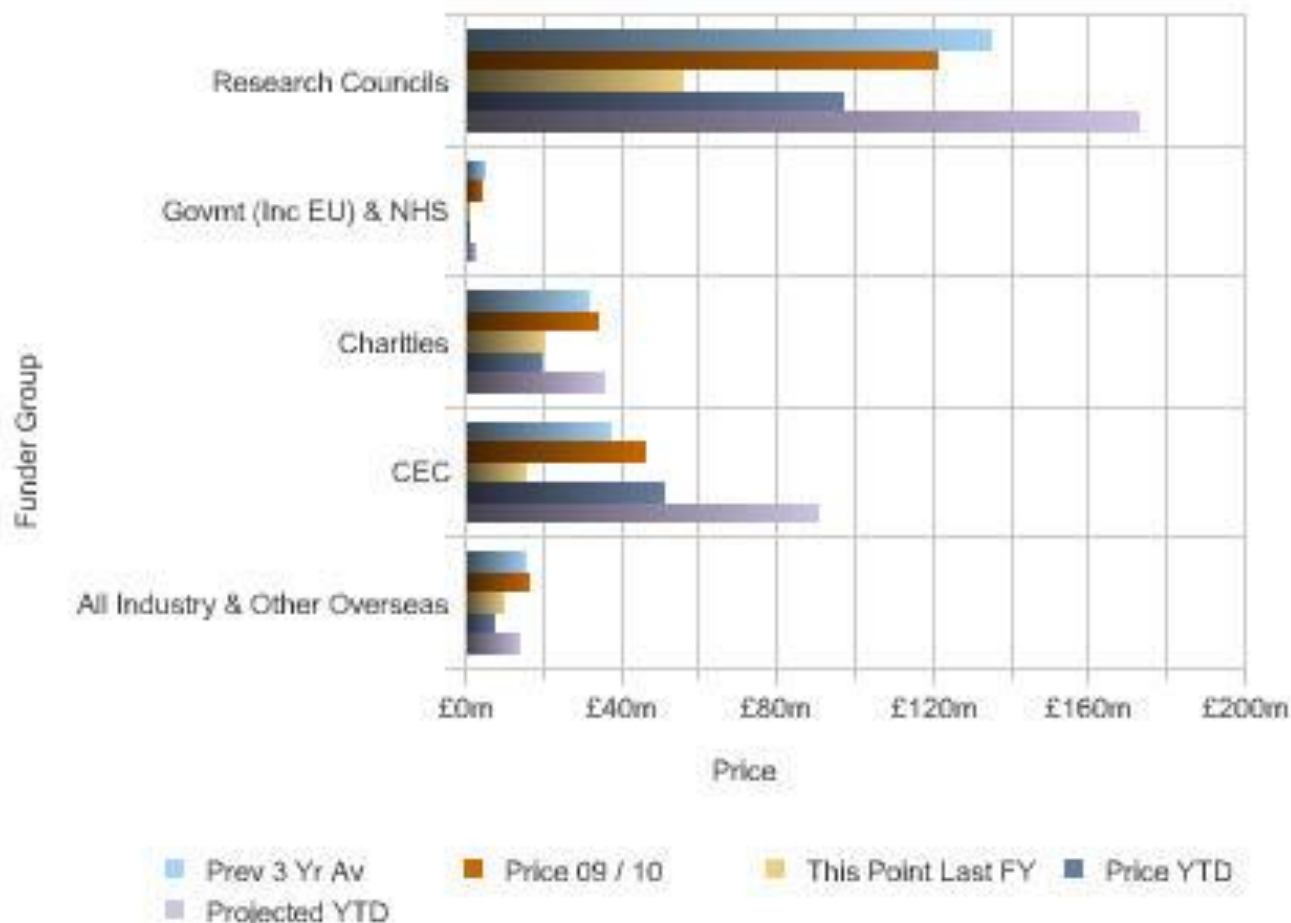


The graph displays the difference in price by sponsor group between the current year's projected total and the previous three-year average.

Source: InfoEd

Sponsor Submission Volumes

Select Analysis Type

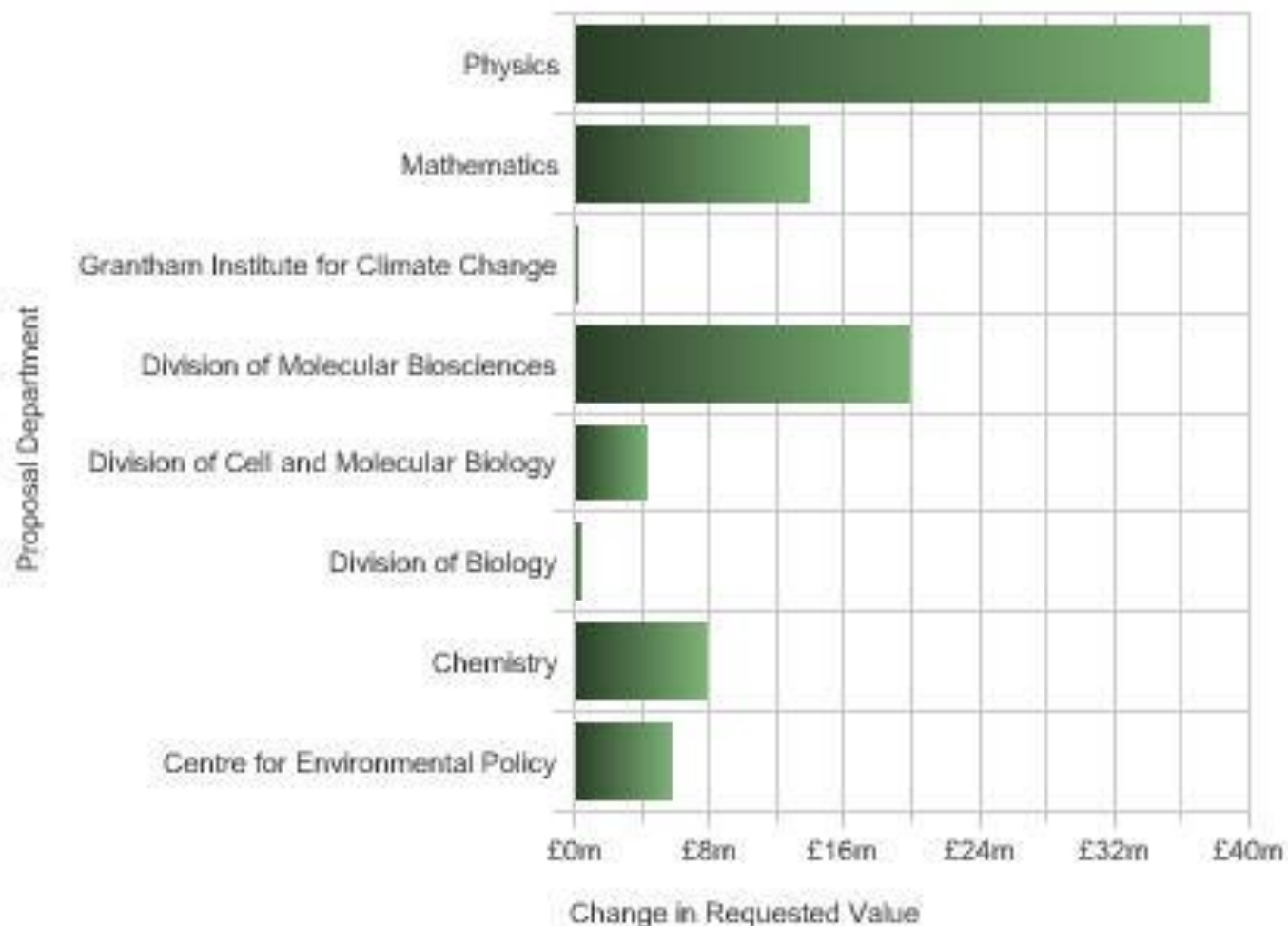


The graph displays the requested amount for the previous three-year average, the previous financial year total, the amount at this point in the previous financial year, the current financial year to date and the linear projected total for the current year.

Source: InfoEd

Departmental Submission Volumes

Select Analysis Type



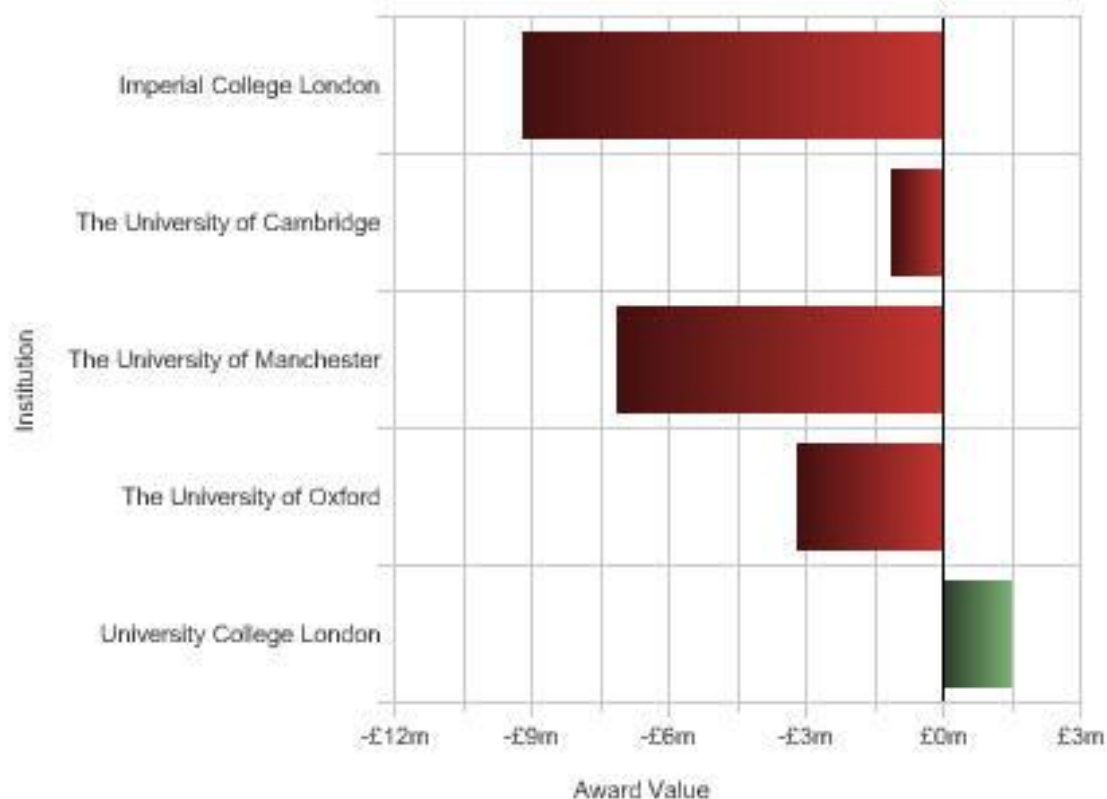
The chart displays the difference in total requested value between the projected current year total and the previous 3-year average.

Source: InfoEd

Award Volumes by Competitor

Select Analysis Type

Funder



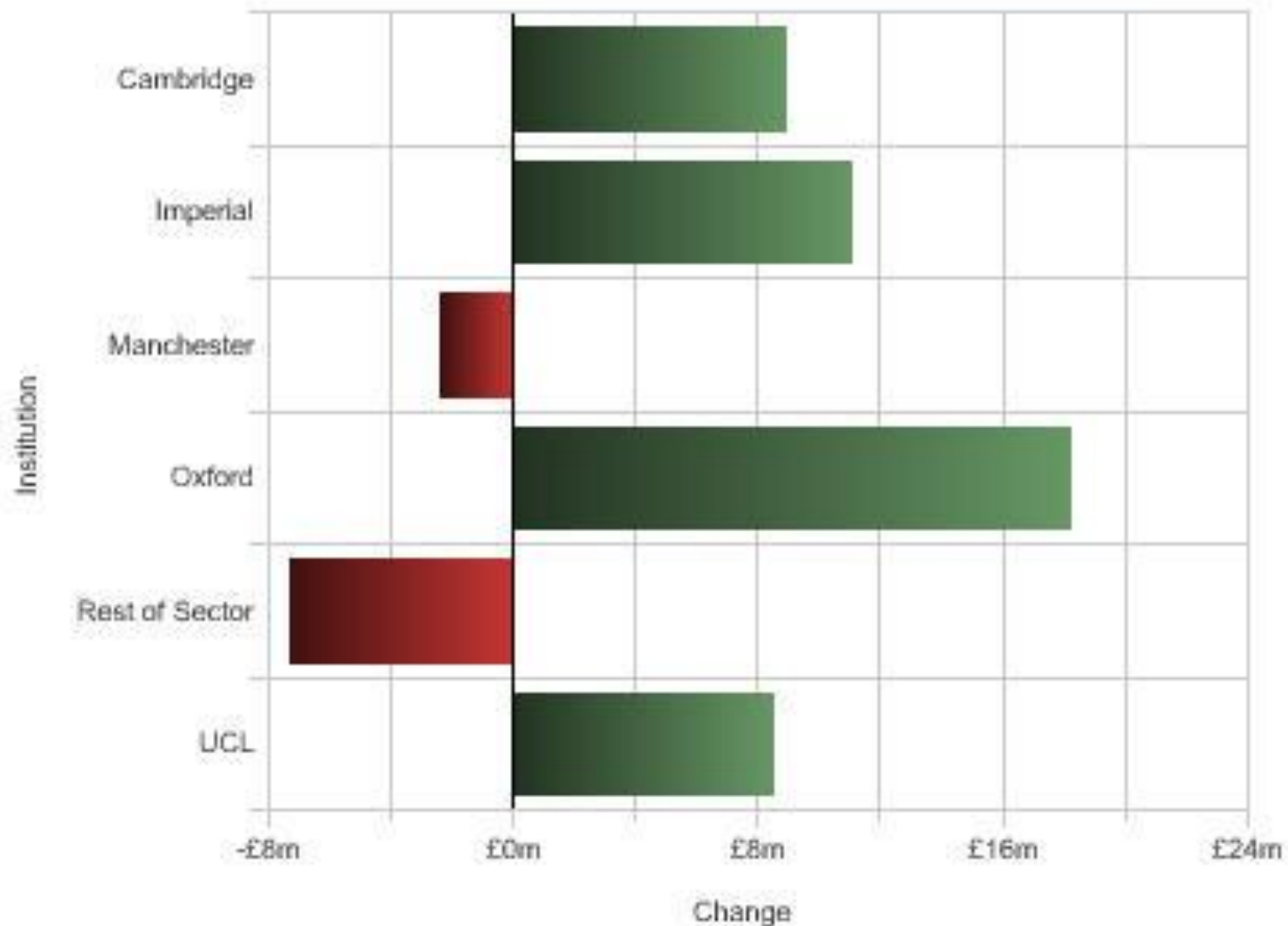
The graph displays the difference in awarded budgets between the financial year 2008 / 09 and the previous three-year average. 2009 / 10 figures will be used instead once all funders have published complete information for that period.

Central Funds, Infrastructure, Equipment and Knowledge Transfer awards are excluded.

Award Volumes by Research Theme

Select Analysis Type

Theme Top Group

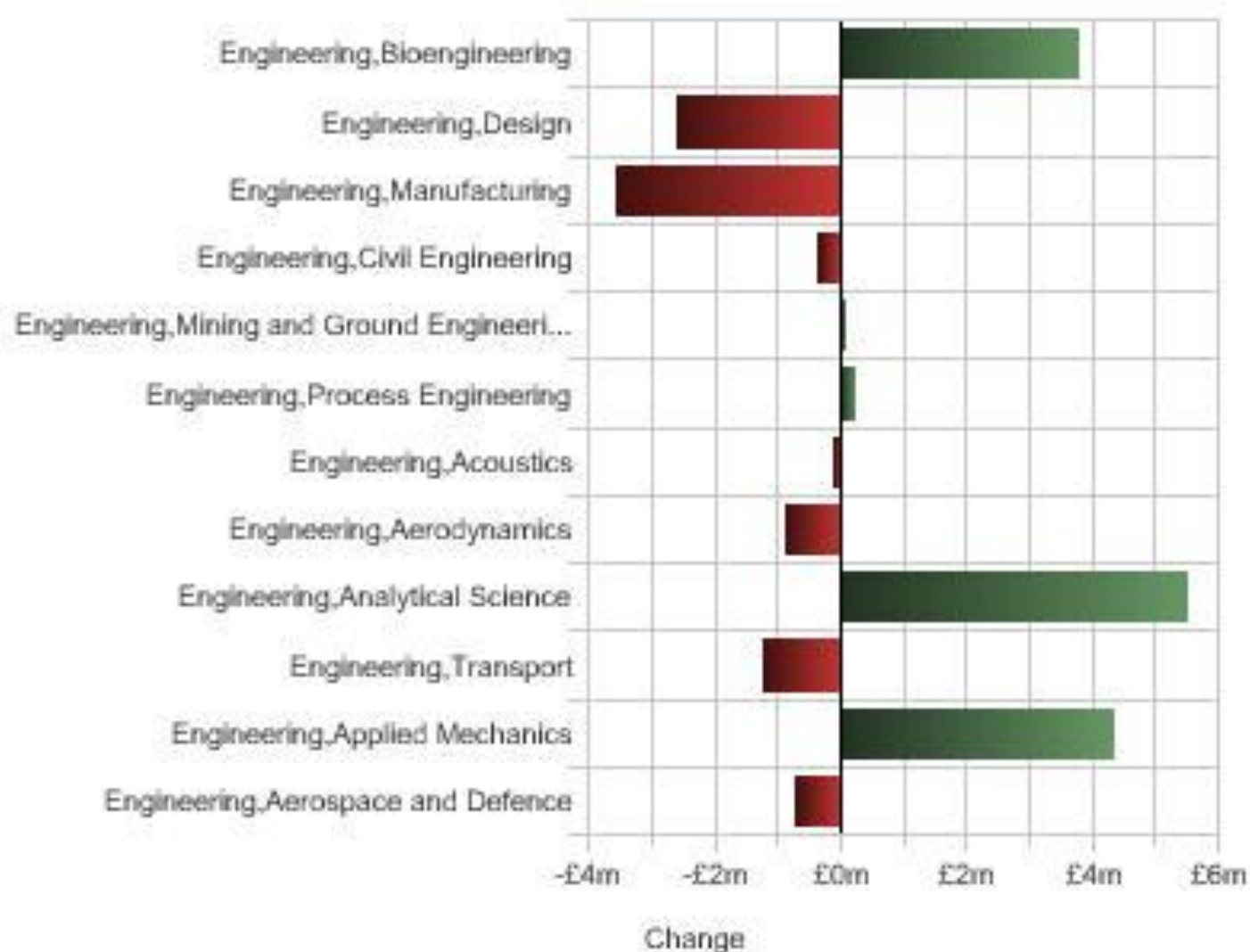


Award Volumes by Research Theme

Select Analysis Type

Institution

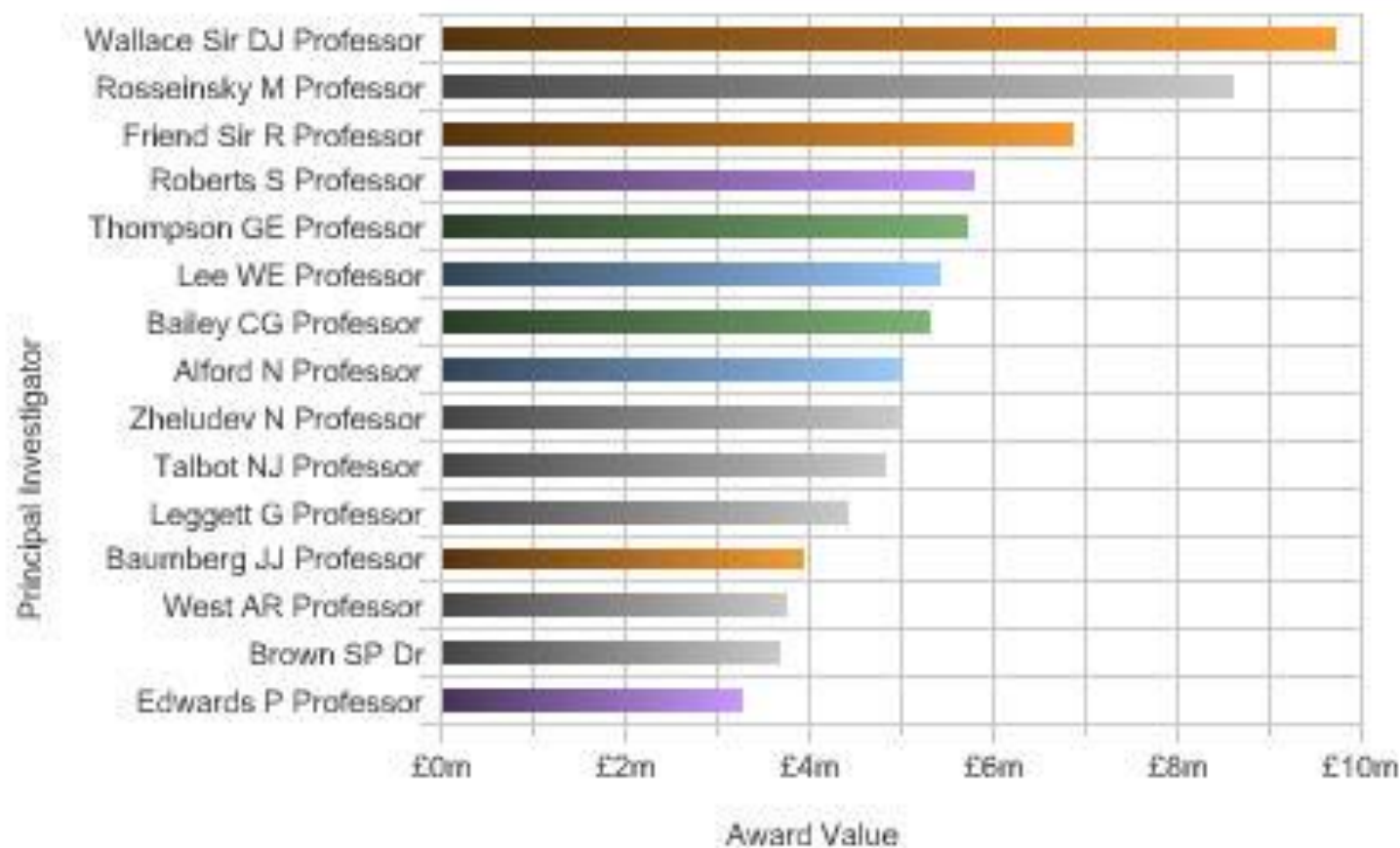
Theme Top Group, Theme Group



Top Academics by Theme Hierarchy

Theme Top Group - Theme Group

Materials - Characterisation



University College London

The University of Cambridge

The University of Manchester

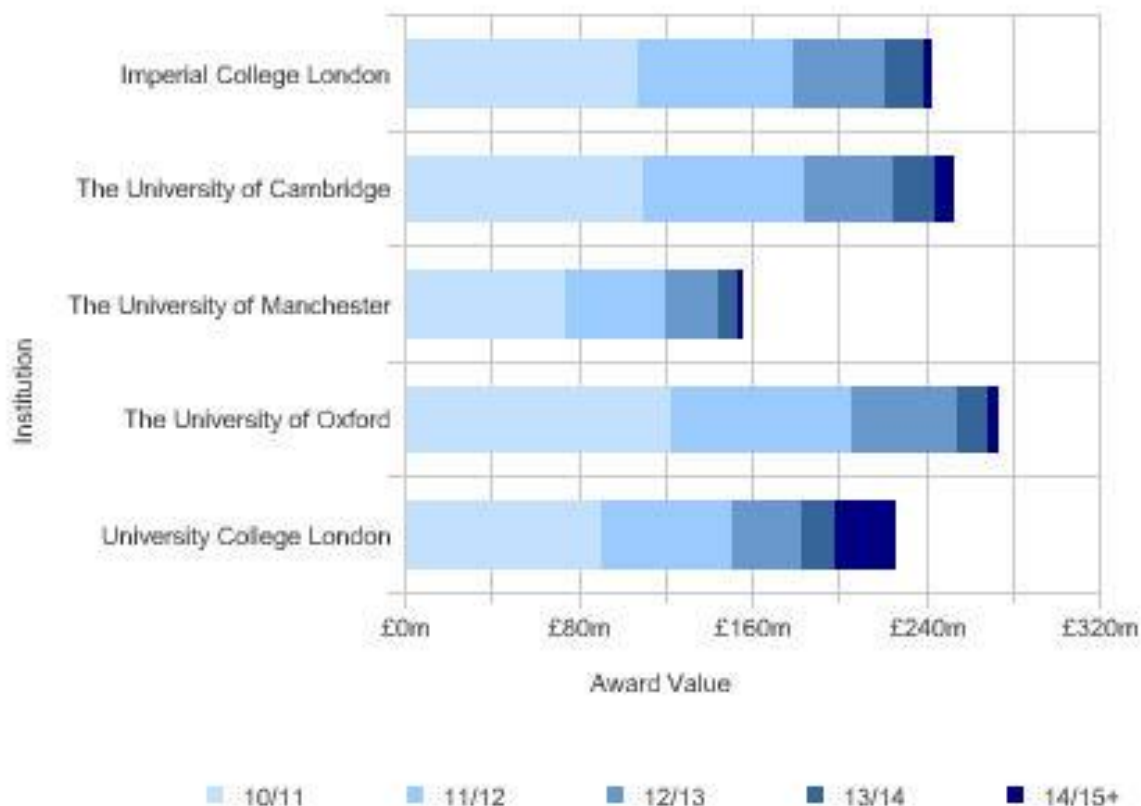
Imperial College London

The University of Oxford

Other Institutions

Award Volumes by Remaining Duration

Select Analysis Type



The graph displays the estimated proportion of awarded budgets that is projected to be spent in each year, based on the assumption of linear spending throughout an award.

All types of awards including Central Funds, Infrastructure, Equipment and Knowledge Transfer awards are included.

Background

- Growing recognition of need for research intelligence and performance management frameworks and metrics
- Dissatisfaction with data and tools available to integrate information from disparate systems
- Frustration that different stakeholders demand similar information in differing formats with differing definitions
- Recognition of duplication of effort, manually intensive systems, and overall inefficiency
- Limitations of external benchmarking through inconsistent definitions of data elements and calculations
- An appetite for more detailed research intelligence and for more sophisticated data tools and systems
 - e.g. recently launched STAR METRICS project in US

Research information management UK study, 2010

- 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
- www.researchdatatools.com/downloads/2010-research-information-management-2.pdf

- **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

- Without clear and shared data elements and metrics driven by institutions, they find that it is almost impossible to benchmark meaningfully
- Lack of a shared definition of metrics makes it difficult for institutions to measure performance against their peers (compare apples with apples)
- 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
- Institutions have allowed the demands of external stakeholders to determine the data and the data-definitions they collect and measure
- Benchmarking requires an institution's own data, proprietary data (e.g. held by funders) and data held by third parties

Second phase project: Snowball

- Self-funded, voluntary project
 - participating institutions are getting perpetual free access to prototype and pilot
 - public service aspect brings value to the sector
- Aims of second phase
 - Define a set of metrics
 - Define all possible sources of the data elements of the metrics calculations
 - Establish a three-year roadmap for adoption in sector
- Address these issues by enabling institutions to benchmark against key research performance and activity metrics on a like-with-like basis
- **Overall goal:** to facilitate external benchmarking by ensuring that institutions can confidently compare research data in a like-with-like manner

**Imperial College
London**



**UNIVERSITY OF
CAMBRIDGE**

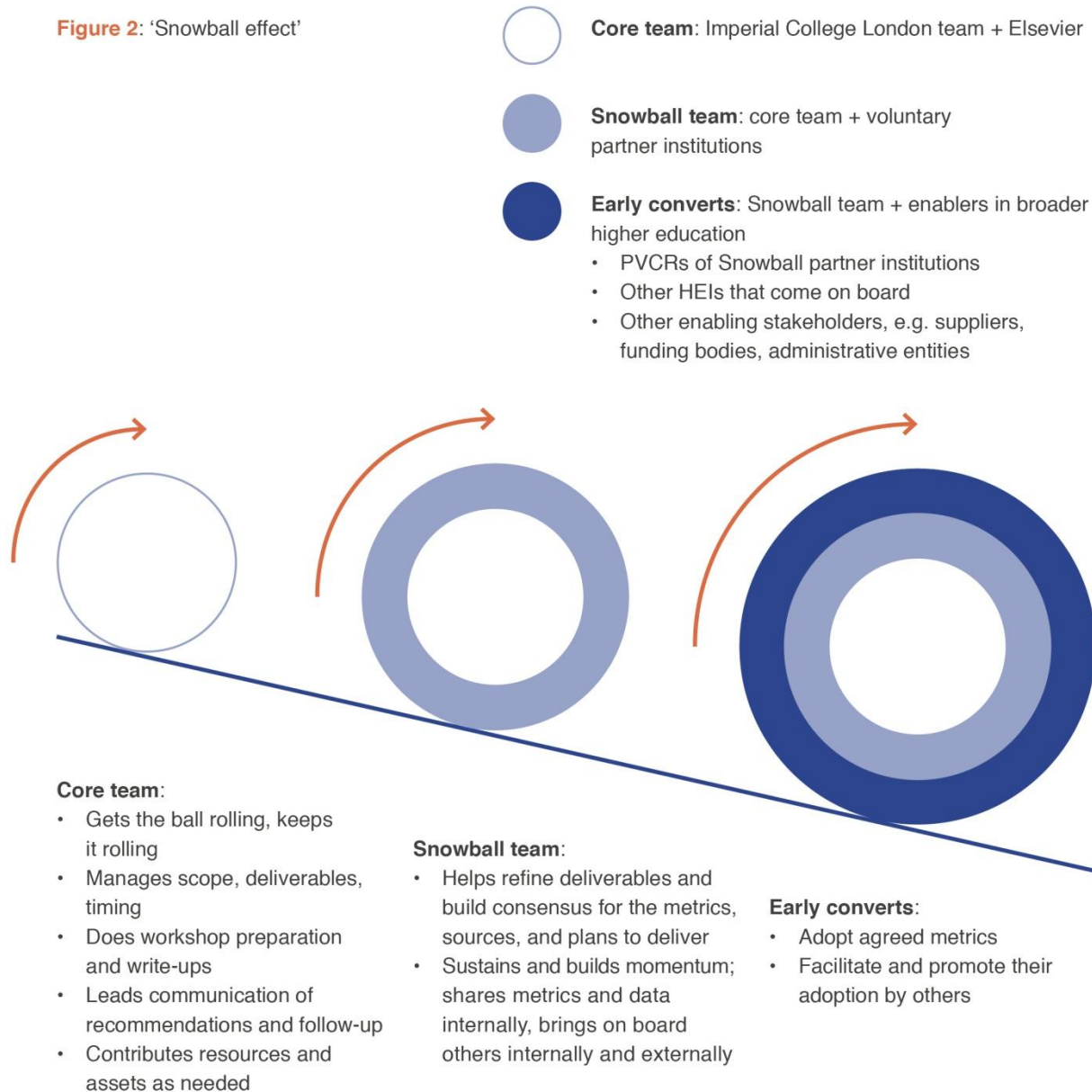


**University of
St Andrews**

Method

- Starting in November 2010, a series of regular workshops organised
- Shape the project's objectives, agree definitions of metrics and sources of data
- Share experiences and knowledge, discuss outcomes
- Aim to create a 'snowball effect' across the sector

Figure 2: 'Snowball effect'



Definition of the landscape of research activities

- Define landscape of activities in the research process
- Identify and agree
 - the data points people wish to measure
 - a common set of denominators
- Prioritise and develop metrics for a subset of core research activities
- Analyse consistent and robust data sources

Figure 3: Definition of the landscape of research activities

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

Measures in scope for Project Snowball (current phase)

Definition and prioritisation of Snowball project metrics

- Sixty-six separate metrics identified
- Definitions used varied considerably, e.g. ‘grant success rates’ and ‘researcher’ calculated and defined in a variety of ways by different institutions
- Pivotal units – the denominators – identified by the Snowball team
- Interrogate metric from a number of perspectives (‘denominators’), including by department, by funder type, by Unit of Assessment, or by a specific research theme

Figure 4: Results of metrics prioritisation

Important, perceived as hard to get

- 1.b.i Percent of grants won vs percent of grants available
- 1.b.ii Amount of grants won vs amount of grants available
- 1.e.i FEC versus amount awarded
- 1.f.i Number of academics involved in enterprise activities
- 3.a.ii Total number of articles authored over lifetime
- 3.a.iv Total citations per article
- 3.a.vi Average citations per article
- 3.a.vii Field Weighted Citation Impact
- 3.a.viii FWCI per article
- 3.a.x Field normalized *h*-index
- 3.a.xi Percentage of articles in top X percentile of citations received
- 3.a.xii Number of books authored in prior year
- 3.a.xiii Total number of books authored over lifetime
- 3.a.xiv Number of book chapters authored in prior year
- 3.a.xv Total number of book chapters authored over lifetime
- 3.a.xvi Number of conference proceedings authored in prior year
- 3.a.xvii Number of [X] authored in prior year
- 3.a.xix Number of [X] authored over lifetime
- 3.b.i Number of specific, high quality (competitive) awards
- 3.b.iii Number of memberships to high quality organisations / learned societies each year
- 3.b.v Number of prestigious awards / medals by year
- 3.b.vi Number of national honours awarded
- 3.b.x Number of outreach activities (public lectures, exhibitions, workshops etc.)
- 3.b.xi Contributions to public policy
- 3.b.xii Number of visits to policy-makers
- 3.b.xiv Journal editorial board membership
- 3.c.i Percent of articles co-authored with a non-home HEI author
- 3.c.ii Percent of articles co-authored with a non-UK author, 2009
- 3.d.i Percentage pass rate for research masters
- 3.d.ii Percentage pass rate for research doctorates
- 3.d.iii Percentage research doctorates completed within three years
- 3.e.i Number of patents granted per year
- 3.e.ii Income from patents per year, before internal distribution
- 3.f.i Number of licenses granted per year
- 3.g.ii Number of spin-outs lasting three years
- 3.g.iii Income from spin-out activity per year

Less important, perceived as hard to get

- 2.a.iv Percentage of actual spend versus budgeted spend by month
- 3.a.ix *h*-index
- 3.a.xvii Number of meeting abstracts authored in prior year
- 3.b.ii Number of prize 'placings'
- 3.b.iv Number of major industrial advisorships each year
- 3.b.vii Number of speaker invitations per year
- 3.b.viii Number of conferences participated in / organised per year
- 3.b.ix Number of media appearances per year
- 3.c.iii Percent of articles co-authored with an author ex-discipline
- 3.c.iv Number of distinctive competencies per HEI
- 3.c.v Percent of books co-authored with a non-home HEI author
- 3.c.vi Percent of books co-authored with a non-UK author
- 3.c.vii Percent of books co-authored with an author ex-discipline

Important, perceived as easy to get

- 1.a.i Percent of grant applications per year that are successful
- 1.b.iii Number of grants won per year, identifying ad hominem
- 1.b.iv Amount of grant funds won per year, identifying ad hominem
- 1.c.ii Number of postgraduates registered for research doctoral degrees by year
- 1.c.iv Number of research doctoral degrees awarded/year
- 1.d.i Volume of industrial income year on year
- 1.d.ii Volume of industrial income per year versus 3 year average
- 1.d.iii Top 10 industry funders over time
- 2.a.i Volume of research spend by month
- 2.a.ii Volume of research spend by year
- 2.a.iii Ratio of research spend from government: vs. non government sources
- 3.a.i Number of articles authored in prior year
- 3.a.iii Total citations per article
- 3.a.v Average citations per article
- 3.b.viii Number of conferences participated in / organised per year
- 3.f.ii Income from licensing activity per year before internal distribution
- 3.g.i Number of spin-outs formed per year

Less important, perceived as easy to get

- 1.c.i Numbers of postgraduates registered for research masters degrees by year
- 1.c.iii Number of research masters degrees awarded/year

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

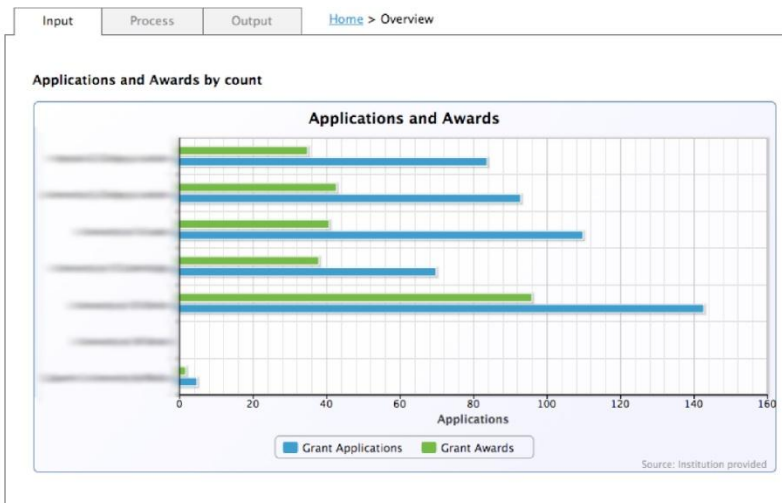
Figure 6: Prototype benchmarking tool



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

Prototype benchmarking tool

- Despite its limitations and restricted nature, regarded as a powerful affirmation of the vision for an external benchmarking tool based on consistently defined and sourced metrics
 - "very worthwhile"
 - "the right thing to be doing"
 - "hugely valuable"

Key conclusions

- There is strong support for the concept of 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

Following the work of our expert group (drawn from the eight institutions), which has reached consensus on the definitions of data fields and metrics:

- Working on a scalable way of implementing, following on from the prototype:
 - including all researchers in institutions
 - using a subset of the metrics defined by our expert group
- Developing relations with funders to source data
- Using data from all three types of data sources: institutional, proprietary and third party
- Working initially with three of the eight institutions (but experts advising to ensure method is scalable)

Medium term aims

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

j.green@imperial.ac.uk
n.fowler@elsevier.com

www.projectsnowball.info

**Imperial College
London**



UNIVERSITY OF
CAMBRIDGE

