



McCutcheon, V., Barr, M., and McHugh, A. (2012) Engage - Using Data About Research Clusters to Enhance Collaboration. Project Report. University of Glasgow, Glasgow, UK.

<http://eprints.gla.ac.uk/69724/>

Deposited on: 19th September 2012

Engage - Using Data About Research Clusters to Enhance Collaboration

Final Report

September 2012

Authors: Valerie McCutcheon, Matthew Barr, Andrew McHugh

Table of Contents

Content	Page
1. Acknowledgements	3
2. Overview	3
3. Background	4
4. Aims and Objectives	5
5. Key Findings	11
6. Project Management Comments	12
7. Outputs and Outcomes	12
8. Recommendations	12

1. Acknowledgements

The Engage project was part-funded by JISC under the Business Intelligence programme.

2. Overview

The Engage project was an 18 month programme of work at the University of Glasgow. The project aimed to explore ideas for improved evidence-based strategic decision making through better access to existing data held within core systems at the University.

It focussed on the classification of research activity by theme.

The University of Glasgow Strategy 'Glasgow 2020: A Global Vision' has a number of aims in the area of delivering excellent research that we wanted to support with this work.

The strategy includes the following statements:

"Aligned with the priorities of our research funders we will, over the course of the next five years, provide resource to establish further thematic, self-sustaining, world class Research Institutes and Centres....based on combinations of our current disciplinary strengths'

'We will also encourage Research Networks that bring together researchers (academic staff, postdoctoral and graduate students) from more than one College, to develop their common interests in a strategically significant research cluster. Networks enable the building of research communities, the establishment of an internal and external presence, and the coordination of bids for research funding"

It outlines plans for:

- Fit-for-purpose infrastructure (including maximising the value and number of uses of the information held on corporate systems in support of the Universities strategic aims)
- Improving support for collaboration to help researchers reach the forefront of their field and develop research strengths and multidisciplinary activities
- Extending knowledge exchange, and business and community engagement
- Enhancing networks of staff (which will facilitate value for money for funders)

The Senior Management Group wanted to have research clusters represented in core systems and the custodians of the systems were therefore both directed and supported in exploring this functionality.

We wanted to provide a better method than the existing tools for people to find potential collaborators. For example, the research map <http://www.gla.ac.uk/services/researchmap/> is labour intensive (and therefore not currently well maintained) and it is constrained to more rigid subject classifications. The A-Z subject classification on the University website is based on the organisational structure stored in the HR System <http://www.gla.ac.uk/subjects/>

The project focused on manifesting a visualisation of research clusters for the College of Arts and making recommendations for the future data and system requirements for research cluster management.

Research themes could be used to group research information already available in core systems and present it for different purposes - from marketing and identifying collaboration opportunities, to management information and workload modelling.

Our latest project Encapsulate follows on from, and utilises some of the ideas developed in, Engage. It will use a digital repository of research outputs (from our repository 'Enlighten') to explore a better method of automatically identifying relevant academic expertise within HEI's to help businesses. For more information on the Encapsulate project visit

<http://academicexpertise.wordpress.com/>

This report is the Engage project management summary and we welcome any suggestions or feedback which will help fulfil the University of Glasgow strategic aims.

3. Background

Within the University of Glasgow we have a well-defined formal organisational structure. We have a long history of re-using research related data across different systems. Interfaces have existed between the core systems for many years. In recent years our JISC projects 'Enrich' and 'Enquire' have succeeded in linking outputs in the repository with people and projects in the Research Support System.

It is often difficult both within the organisation, and with external bodies to find out who is interested in different themes and therefore it can be time consuming or failure may result when trying to identify potential collaborators or partners for projects. A scan of the current environment makes it clear that we need to understand and identify research themes to:

- Deliver research that meets funding priorities
- Foster interdisciplinary research
- Align strategic priorities and areas of excellence within the University
- Support relevant research clusters as per our strategy
- Market expertise of the research clusters
- Facilitate business and community interaction
- Improve information sharing
- Minimise duplication of effort
- Share resources

Traditional methods of finding collaborators include networking in the research community, subject searches or serendipity. The Engage project seeks to offer another way of identifying potential collaborators as well as identifying previously unexplored collaboration opportunities.

We would like to record 'research themes' in addition to the formal organisational unit codes. Research themes could be described as work areas. A person might be involved in several themes at any one time. They are less static than organisational units. Research activities in different disciplinary areas may have intellectual cross-connections which are not immediately evident to managers or individual researchers but can nevertheless be identified by research mapping. This requires modelling tools which are flexible enough to

accommodate different and shifting interdisciplinary configurations and that are not constrained by conventional disciplinary boundaries.

At the University of Glasgow we have a number of formal and informal research theme classifications. A myriad of research themes can be found in many organisations but often these are very informal and not included in core systems. We wanted to identify something that could be applied across the University and across organisational boundaries and be stored in core systems so that information was more readily discoverable by research theme.

4. Aims and Objectives

- As this project was funded by JISC one of the key requirements was to **Field Test the JISC InfoNet BI resource** and provide the JISC InfoNet BI team with feedback.
- To **explore definitions** of what a research cluster, theme, or group is and the data requirements of research clusters.
- To **create a method for visualising and presenting** data associated with research clusters.
- To **share our story** including any issues encountered with the sector.

4. 1 Field Testing the JISC InfoNet BI Resource

The JISC Infonet BI Resource is intended as a practical support guide to help organisations explore how to use information they already have in their core systems in support of strategic aims.

The resource website is: <http://www.jiscinfonet.ac.uk/bi>

We found the JISC BI resource useful in a number of ways.

The maturity model provided is intended to measure the level of development of business intelligence at an organisation. Details can be found at:

<http://www.jiscinfonet.ac.uk/infokits/strategy/environment-scanning/maturity-model>

In terms of the 'Engage' project we utilised this model specifically to look at the maturity of 'research themes' at the University of Glasgow.

It helped us articulate where we were at the beginning of the project and what action might be appropriate to improve the methods of managing research themes at the University.

The descriptions used on the BI site were generally very clear and precise for example from the descriptions of BI implementation issues it was easy for us to consider ways in which we had addressed these:

- We aimed to drive process improvement by reducing the burden of identifying common research interests in a more intelligence and automated way

- We used some innovative technology and are pursuing these in partnership with suppliers (e.g. ePrints) and data suppliers (e.g. other HEI's involved in the Encapsulate project, and internal systems custodians)
- We have considered change management and have suggested that re-visiting the business case and explaining the value add is essential to Senior Management if they wish to make a robust case to implement our ideas live in core systems in terms of making the financial case and getting buy-in from users.
- The project was based on re-using existing internal data in a more sophisticated way to help address requirements stated in our institutional strategy. There was on-going engagement with data suppliers from a range of core systems.
- There were data quality and definition aspects to the project. For example we considered how data could be maintained with relevant topics attached to entities and where data management responsibilities might be placed.
- Within the specific area of research themes we could be considered a maturity exemplar. We are not aware of many other organisations that are considering how to incorporate detailed research themes with staff, project, and output data and utilise this for improved discovery of research related information.

As well as using the InfoKit to help us ensure we kept within the scope of the programme it was (and will be) very useful for describing the potential benefits to internal stakeholders.

4.2 Exploring Definitions

For our prototype we were keen to identify a subject classification that was comprehensive and comprehensible to cover the research at the University. We also wanted to utilise a standard set of definitions if something suitable existed.

The University of Glasgow organisational units cannot be used as robust means of information categorisation, since they were not really designed for this purpose.

Considerable research was done and a focus group was held with researchers and research administrators in the College of Arts and the input to this was used to inform the prototype.

We looked at a number of options before finally settling on a unified classification scheme used by the UK Research Councils. Some of these considered but not adopted were:

NAME	PROS	CONS
Library of Congress		<ul style="list-style-type: none"> • Insufficient granularity
HESA Cost Centres		<ul style="list-style-type: none"> • Insufficient granularity
Eurostat's Classifications metadata	<ul style="list-style-type: none"> • Hugely detailed 	<ul style="list-style-type: none"> • Unwieldy metadata • Economic focus
The EU's Nomenclature for the analysis and Comparison of Scientific Programmes and Budgets (NABS) classification	<ul style="list-style-type: none"> • Strong research focus 	<ul style="list-style-type: none"> • Good for science but not so for 'non-science' topics
The Universal Decimal	<ul style="list-style-type: none"> • Good fit to needs 	<ul style="list-style-type: none"> • Not closely related to

Classification Summary (udcS) (based on the larger UDC Master Reference File)	<ul style="list-style-type: none"> Licensed under Creative Commons Attribution Share Alike licence International 	familiar UK nomenclature
Standard Industry Codes		<ul style="list-style-type: none"> Economic focus less suited to research activity

The rationale for choosing the Research Councils' classification scheme included the facts that it is:

- Cross-disciplinary - the Arts are as well represented as the Sciences
- Hierarchical - allowing us to make more intuitive associations between related research
- Already related to a large percentage of funding (Research Council funding accounts for 1/3 of our annual research income)
- More detailed than the organisational structure codes available at the University
- Widely understood within the University and across the sector

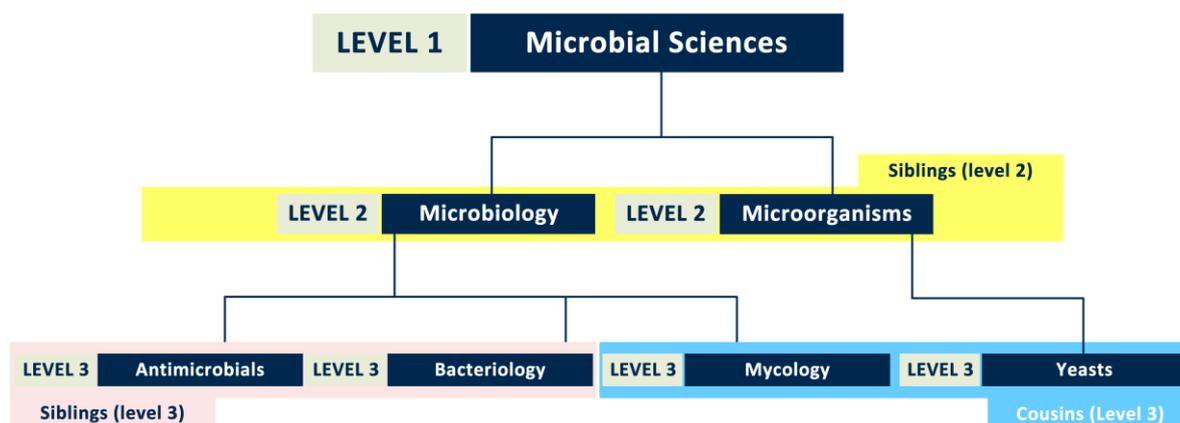


Diagram 1 – Extract from the RCUK classification system

For the 'Engage' project we imported a flat data list of RCUK research classifications to our demonstrator, creating a relational database version of the classifications that preserved their implicit hierarchical structure.

Another JISC project in which we are involved 'IRIOS2' illustrates that we can obtain the Research Council classification and award data as an export from the Research Council systems and potentially use them as a basis for our 'theme' classification albeit we 'only' get classifications for some Research Council's at present and not all Research Council's or all funders. This is a good potential base that could remove a portion of the data entry and be developed to facilitate assignation of standard research themes to all of our projects (and their outputs).

The current IRIOS2 demonstrator includes a number of other classifications which might be of use - e.g. FRASCATI and Eurostat 2007 NABS.

Further details can be found at <http://irios2.wordpress.com/>

The University has an Oracle-driven HR System and we explored the process of adding research themes to a test version of the system. Themes could potentially be controlled centrally – where a rigid structure was needed for consistency or reporting, or administered via the self-service portal where the classification was to be more fluid. The current version of the HR System allows for 12 themes per person so some modification would be required if it were decided to implement in the live system.

4.3 Creating Visualisations

We used a range of technologies a summary of which can be found on the technology table in the post:

<http://researchclusters.wordpress.com/2012/06/28/a-poster-on-engage/>

The repository – '*Enlighten*' – provides a central location for our research outputs, such as papers, journal articles and book chapters. What we aimed to do was to use the publications deposited by researchers to inform their research profile, by analysing the text of their papers' abstracts or full texts (where copyright and licencing allows) to pull out keywords.

We did not use the data entered for our mini-REF exercises in this prototype as we had specified what purposes that would be used for. If we wanted to use REF data in future data protection statements would need to be re-examined.

The web team are already harvesting lists of publication data from the repository for researcher profiles. The 'Engage' prototype made use of similar data harvested from the repository to identify themes.

We Used Google Analytics to review the search terms a visitor used to find a researcher. We also considered creating a version of the "Amazon Recommends" system, whereby you might be presented with other researchers that match your interests, based on those you had already viewed. This appears in the prototype in the form of the 'Related Researchers' part of the research profiles.

One of the central aims of the *Engage* project was to ensure that academics and academic administrators will not be asked to input data that already exists elsewhere. We looked at utilising existing data from a number of the University's core systems, including:

- Data such as name, organisational unit and staff number already automatically flows from our **HR System** to our repository.

- Similarly information about projects, awards and applications from our **research support system** is made available in the repository where it is linked to outputs from the projects.

We added a 'research methodology' field to our research profiles to allow us to make cross-disciplinary connections between researchers based on the methodologies employed in their research. Researchers might be related, for example, by their use of a method such as prosopography, but working in otherwise disconnected disciplines such as history and geography, or sociology and art history.

Feedback from those we demonstrated the system to was good. A summary of the demonstration can be found at:

<http://researchclusters.wordpress.com/2012/07/31/feedback-on-the-engage-prototype/>

A range of Visualisations was developed including:

- Simple tag clouds
- Venn diagrams showing potential overlap of interests with colleagues
- Bubble charts that might help identify potential collaborators
- Motion charts showing how interests of a person have changed over time
- Pie charts
- Connections

The 'extended search' option allows users to search a wider variety of sources, including the full text of PDF articles held in the institutional repository. These searches are more resource-intensive, and therefore take slightly longer to perform, but can provide richer results.

Likewise, the option to allow for stemming – reducing keywords to their root or stem form – means that search results can be conflated to include variations on the entered search terms.

Visualisations

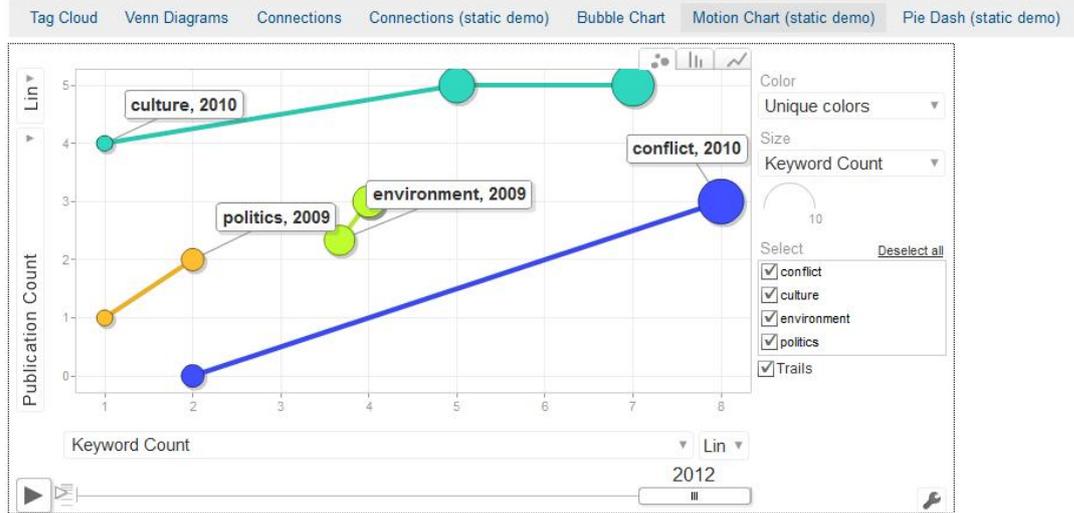


Diagram 2 – Example visualisation produced during Engage project

To further illustrate the re-use of data and the ideas produced during the project it might be of interest to readers that the programmer also developed an international activity map as a separate task outwith the project. It uses similar technology and there is a potential to re-use themes and data from core systems. Senior Management will be discussing whether to allocate resource to further development and integration of the map during Autumn 2012. This might include addition of fields to core systems, guidance on what data should be entered into which core system so that it can feed the application, and methods for geocoding of data.

4.4 Information Sharing

It was very important to us to engage with stakeholders during the project. Not just because the funder advocates this but because besides disseminating what we have done we genuinely expected, and have received, very useful information exchange. The informal network of interested parties helps to pave the way for better information management in the future reducing the need for individuals to struggle with issues already visited by others.

We will be reporting back to Senior Management at the University with a view to revisiting the business case for research themes in core systems and exploring the roadmap for implementing these more fully should Senior Management wish to do so.

As well as a specific workshop for staff in the College of Arts we have had interest from administrators in other colleges and University Services who can see benefits in exposing research activity by theme. They are interested in more automated methods to do so. They also want to ensure that research theme data is accessible via core systems.

We have communicated with the wider HEI community:

- Via our blog: <http://researchclusters.wordpress.com/>
- Via posters and presentations at:
 - ARMA
 - The Digital Curation Centre Roadshow in Belfast
 - OR2012
 - At our Research Staff Conference
 - At our Research Network meetings
- In July 2012 we hosted a workshop to demonstrate the software and discuss ideas and business requirements.

The final report from this workshop can be found at <http://researchclusters.wordpress.com/2012/09/10/workshop-report/>

We have also found the JISC programme meetings both for this 'Business Intelligence' programme but also for related projects on 'Research Information Management' and 'Business and Community Engagement' very useful forums for sharing ideas as many of the initiatives overlap.

It was clear from the BI cluster meeting that there was a desire for better evidence sharing across the sector to support staff in making the business case for these sorts of BI projects as well as marketing benefits to stakeholders.

The challenges of the 'Business and Community Interaction' programme included tackling 'the language barrier' (research outputs need to be intelligible to the stakeholder), and 'Re-use' of information for different needs – both issues that we were keen to understand and offer solutions to in Engage.

There was talk about the need to not just throw information at the world but to target to business and community requirements – exactly what Engage and Encapsulate are doing.

5. Key Findings

5.1 There is an appetite for cross-institutional collaboration to identify and implement methods of modelling research themes and embedding them into core systems. There was interest in an open source offering of the Engage software and it potentially being used across institutional boundaries as well as within specific organisations.

5.2 There is a lack of agreed terminology for research classification.

5.3 There is a lot of interest in coming up with some standard classifications.

5.4 The benefits for researchers and organisations need to be further evidenced to inform a robust business case.

5.5 There needs to be careful management of content so that information is not used inappropriately. E.g.

- Sources such as Web of Science do not allow abstracts to be downloaded and then made public.

- Some activities such as licencing of inventions and contracts may have intellectual property requirements and any process to harvest information must allow for certain information to be excluded from presentation in the public domain.

6. Project Management Comments

For JISC projects we find the wiki's provided an excellent way of sharing documents and information across the internal and external representatives of the project team. We can make information visible to the entire programme and we can also see what other projects are doing and learn from and link up with them. The wiki also acts as a project log and is invaluable when writing up findings such as this report.

At times we found that user expectations were outwith the scope of the project. In future we should definitely ensure we communicate the goals of any projects more effectively and re-iterate these to all audiences.

7. Outputs and Outcomes

7.1 A working prototype where research themes could be:

Auto generated using the RCUK list

OR

Bespoke according to researcher preference

7.2 Clever re-use of keywords to enhance discovery

7.3 Some excellent visualisations which received great feedback at our workshop

7.4 Workshop and final reports that may assist others

7.5 We have sustainability in the form of an extension to the development work. Our new 'Encapsulate' project hopes to improve the success of 3rd parties (e.g. businesses) in finding relevant academic expertise.

7.6 Improved internal discussion around research themes and enhanced linkages of information professionals.

7.7 An informal community of organisations exchanging ideas and views on research themes.

8. Recommendations

8.1 Implement research themes in HR System or other core system. Resource will be required to classify specific awards and outputs by topic even if we semi-automate this.

8.2 Senior Management needs to decide if there is a robust business case. We could provide a short business case from our findings and utilise the advice on the JISC InfoNet kit. <http://www.jiscinfonet.ac.uk/infokits/strategy/environment-scanning/business-case> In addition <http://www.jiscinfonet.ac.uk/infokits/research> may also be worthy of referral.

8.3 Define requirements and select functionality to create visualization and to and expose the raw data for others to create their own views.

8.4 If it is decided to take the geographical mapping modules forward investigate methods for improved geo-coding. Some interesting examples include:

- The BatchGeo service was mentioned at our workshop as a possible means of geo-coding large quantities of place data for visualisation on a map. Currently, we use the Google Maps API for geo-coding but it would be worth looking into using BatchGeo to efficiently generate latitude and longitude values for existing data derived from our core systems.
- Thompson Reuters work on the movement of researchers across the globe
<http://academicexecutives.elsevier.com/sites/default/files/aeb0201.pdf>

8.5 Investigate methods for extracting data from core systems. IT Services are already exploring data warehouse and analytics options for a range of business requirements. Some interesting examples include:

- Business intelligence tools such as the Oracle Business Intelligence Discover that works with our HR System and LOGIX4 that is being explored for our finance and research support system.
- Mining and visualising research activity using 'Viewshare' as presented by the Library of Congress speaker at OR2012
https://www.conftool.net/or2012/index.php?page=browseSessions&form_session=2
- Thomson Reuter's themescape maps of patent portfolios which had helped to identify emerging technologies. Thompson Reuters also have some citation mapping tools available that we could evaluate.
- From the JISC BI programme it may be worth exchanging views with Open University and Bolton who have committed to a data warehouse approach and may be able to share their business cases.

8.6 Explore visualisation tools. For example:

- Microsoft Academic search has some nice visualisations. It might be useful to check how data is synchronised with institutional systems as apparently the user can just login and change their data in MS Academic if they don't think it is right.
<http://academic.research.microsoft.com>
- Natural language visualisations such as Nutrina BI that make it easier for user to get their questions answered.
- QlikView is currently getting penetration in the HEI market.

8.7 The technology developed and classification could be used more generically to allow better discovery of research outputs such as publication and datasets. This is very much related to a suite of information activities including the Cerif for Datasets, Cerif in Action and IRIOS2 projects that the University is working on with other HEI's. We will liaise with these initiatives accordingly.

8.8 Evaluate usefulness of 'research method' addition to the dataset.

8.9 Ensure advertising and joining up initiatives within the University of Glasgow to avoid duplication of effort and ensure most efficient re-use of data from core systems.

8.10 Explore linkage of the Customer Relationship Management (CRM) system data with other core system data to minimise re-keying and facilitate information sharing. Events on the CRM system are often linked to award on the Research Support System, or outputs on the repository which in turn relate to specific themes.

8.11 Continue to liaise with other HEI's as to how and why they are using research themes. There is a growing interest in recording these in core systems. For example the PURE user group have discussed this and PURE has a 'graph of relations' function. Smartsimple Software has research theme included in the standard fields. As there are a number of different purposes mapping may be required e.g. from the standard theme mapping to Unit of Assessment, or to Web of Science topics.

8.12 Outwith the direct scope of Research Themes but related we should further consider how data outwith the University systems can be compiled for researchers. This is likely to involve a portfolio of tools. Other projects such as Cerif in Action <http://cerifinaction.wordpress.com/> are considering exchange of data between HEI's whilst <http://researchrevealed.ilt.bris.ac.uk/> includes an impact recording tool for gathering researcher specific information from the public domain and integrating it with data already stored on internal systems. The RCUK Gateway to Research is another initiative to expose research activity though this is initially focussed only on work funded by Research Council's. Google Scholar pulls together connections across institutional boundaries. We could consider research agents that look up any 'new' content posted on the web or social media. Attendees at the workshop also suggested a managed bookmark facility where staff could indicate pages that they thought should be harvested. At the JISC programme meeting we had a discussion re possible ways that analysis of social media might benefit HEI's e.g. incorporation of subjects actually tweeted at conferences.

8.13 Consider relationships to other mappings and liaise with other stakeholders to identify key requirements e.g.

- Subject classifications used by the big aggregation's like Web of Science and Scopus e.g. <http://www.info.sciverse.com/scopus/scopus-in-detail/content-coverage-guide/journalclassification>
This could improve linkage with some of the citation benchmarking services offered.
- The IRIOS demonstrator includes a number of other subject classifications such as OECD 2002 FRASCATI classification, Eurostat 2007 NABS Classification.
<http://irios2.wordpress.com/2012/08/21/research-classification/>