CERIFy: Using Business Process Mapping to Engage with Research Information Management Processes and the CERIF Data Model

Mahendra Mahey[a] and Niamh Brennan[b]

[a] UKOLN, University of Bath, Bath, UK

[b] Research Information Systems & Services, Trinity College Library Dublin, Ireland

Summary

This paper outlines the approach of the CERIFy project, which involved engagement with four UK universities and one commercial company, and which can be usefully applied to institutions / organisations / companies who are working in the area of Research Information Management (RIM) and Current Research Information Systems (CRISs) to demonstrate the use of CERIF in ways which are immediately meaningful to the institutional research information manager or end-user. The CERIFy methodology puts the 'user' at the heart of system development and data exchange by using Business Process Analysis. The results can be used to facilitate data exchange / interoperation with systems and to assist with the procurement / development / modification / improvement of a CERIF-based CRIS system. We believe that the CERIFy approach offers an accessible, transferrable and scalable way for organisations to develop local knowledge and experience of the interoperation between their own systems and data, and the CERIF data model. CERIFy provides concrete examples of how the approach facilitated the data modelling of improved (standardised and integrated) RIM processes such as Indicators of Esteem and a bi-directional data exchange process of CRIS-based institutional data with an external commercial system (Thomson Reuters InCites) using CERIF-XML.

1 Background to CERIF in the UK

The Joint Information Systems Committee (JISC) has commissioned and published a number of reports highlighting the importance of CERIF to Research Information Management (RIM), notably the "Exchanging Research Information in the UK" (EXRI-UK) report which identified CERIF as the best option for improving the interoperability and exchange of research information (Rogers et al., 2010) and an analysis of the potential savings to the UK higher education sector of using CERIF, as evidenced by Bolton (2010).

The CERIFy project[1] is part of a raft of UK projects (now totalling over £2 million pounds in funding) supported by JISC[2] and in close co-operation with EuroCRIS which have, together with the work of other funders/bodies, rapidly increased the number of UK institutions who are now

[1] CERIFy project website, http://cerify.ukoln.ac.uk [Accessed 20/04/2012]

[2] JISC RIM page, http://www.jisc.ac.uk/whatwedo/themes/informationenvironment/researchinfomgt.aspx [Accessed 20/04/2012]

actively engaged with CERIF. The UK is currently at the forefront of European engagement with the CERIF standard (Russell, 2012). Over the past three years, this has contributed to an increased number of institutions procuring CERIF-compliant CRIS systems in the UK and a growing number of experts and practitioners within the UK Higher Education (HE) sector who are engaging with CERIF.

2 The CERIFy Project

The project was a partnership between UKOLN at the University of Bath, England, UK, and Research Information Systems and Services, Trinity College Library Dublin (TCD), Ireland.

The project's main aim was to increase engagement of the UK HE and commercial sector with CERIF and CRISs, and to support an emerging community of practice in Research Information Management (RIM). The project was designed to evaluate, test and demonstrate CERIF in partnership with the four pilot UK HE institutions (Aberystwyth University (AU), University of Bath (UoB), University of Huddersfield (UoH) and Queen's University Belfast (QUB)) and Thomson Reuters (TR) (based in Philadelphia, USA and a member of EuroCRIS). None of the project pilot partners had a working CRIS system based on the CERIF data model. UoH had previously managed a CERIF-based JISC project[3] to explore the management of their research information data, QUB had in-depth experience of a CRIS system (which was built in-house but not CERIF compliant) and none of TR's products were CERIF compliant. However, by the end of the project TR had 'CERIFy-ed' their bibliometric analysis product and service, InCites[4], to enable two-way, interoperable, CERIF-based data exchange with the project's CRIS instances based at TCD, to add value to validated institution-specific data and to provide a valuable new view on current institutional research outputs. TR subsequently CERIFy-ed their CRIS system, Research in View[5].

3 The CERIFy Approach

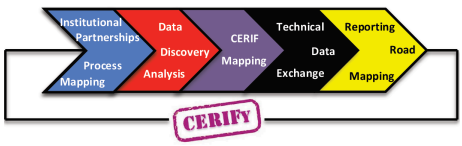

Figure 1: The CERIFy approach summarised

The approach attempts to analyse the context of Research Information Management (RIM), namely the people, business processes, systems, data and policies, in order to develop application profiles of the CERIF standard which are immediately meaningful to those working with RIM at the institutional / organisational level.

It follows the pattern of identifying and moving from the institution's RIM processes to the actual data behind them. It employed a very practical and accessible, 'end-to-end' methodology of re-

[3] Using Business Process Management Tools and Methods for Building Research Information Management (BRIM) Project, http://www.jisc.ac.uk/media/documents/programmes/RIM/RIMBRIM_FinalReport.pdf [Accessed 20/04/2012]

[4] InCites, http://researchanalytics.thomsonreuters.com/incites/ [Accessed 20/04/2012]

[5] Research in View, http://researchanalytics.thomsonreuters.com/researchinview/ [Accessed 20/04/2012]

quirements elicitation, people-centred organizational / business modelling, data discovery, data analysis, data mapping, data exchange, and software demonstration.

3.1 Process Analysis

3.1.1 The importance of users in process analysis

CERIFy focuses on the users / stakeholders involved in the RIM processes. CERIFy adopted the classic approach adopted by Business Process Mapping[6], a change management tool, where users articulate the processes they employ ('as is') and how they would like improve them ('to be'). The approach mirrors other user-centred paradigms such as the Trust DR Model (Casey et al., 2009) which provides an organisational framework / tool necessary for the analysis of institutional readiness for change through technology. Although that tool was developed for the teaching and learning environment it has been adapted for guidance in the research context (see Figure 2). There are similar parallels with socio-cognitive engineering approaches to system development[7].

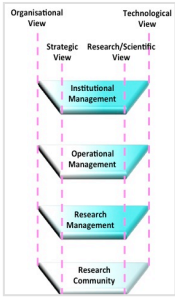

Figure 2: The Trust DR Model (Adapted for Research)

3.1.2 Mapping Current RIM Research to RIM Processes

The first part of the CERIFy approach was to provide a view of existing RIM research mapped to RIM processes to clearly see where research had already been carried out and where the gaps might still be. This resource informed important decisions to be made later in the CERIFy project, especially where RIM processes that merited further investigation were being considered as candidates for engaging with the CERIF data model. The results of this research are available as a publicly available and updatable RIM matrix / mapping spread sheet resource[8].

3.1.3 Institutional RIM Investigation and Analysis

A number of questionnaires were developed to elicit more detailed information about the stakeholders, processes and data within the institution. All are available on the CERIFy project website, under the 'Resources' tab[9]. These questionnaires were distributed to the project partners prior to the site visits and formed the basis for the first process maps for each institution, as well as providing the preliminary knowledge about the institution's readiness to engage with CERIF. The preparatory investigation also identified the institutional stakeholders involved in RIM processes.

---

[6] JISC InfoNet Process Review Data Collection Template: http://www.jiscinfonet.ac.uk/InfoKits/infokit-related-files/process-doc-template [Accessed 20/04/2012]

[7] Socio-cognitive engineering: a methodology for the design of human-centred technology, http://www.eee.bham.ac.uk/sharplem/Papers/mediate%20ejor.pdf, [Accessed, 20/04/2012]

[8] RIM matrix of current research in the area mapped by RIM processes, http://goo.gl/lC5V3 [Accessed 20/04/2012]

[9] CERIFy project resources, http://cerify.ukoln.ac.uk/resources [Accessed 20/04/2012]

User Requirements Elicitation

User Requirements Elicitation with a full range of stakeholders (administrative and academic) took place during the two-day site visits and consisted of interviews using an online form (completed face to face or remotely) with subsequent synthesis and anonymisation of responses (where requested).

On-site RIM Business Process Mapping

RIM processes in relation to institutional and stakeholder priorities were obtained through the 'RIM Process Review' prior to each site visit. As part of this, key research processes were identified by the institutional Research Manager or equivalent within each of the four partner institutions.

Building on this information, during the two-day site visits, a minimum of two research processes per institution were mapped in detail on-site, in partnership with a range of individuals who engaged with these processes and had a detailed knowledge of their operation within the institution. The process maps were created rapidly and the participants were encouraged to view, correct, refine and add detail to each process map on Day 2 of each site visit. To inform the 'to be' process i.e. an improved process based on standardised research information/CRIS environment, institutional stakeholders were asked to identify the issues associated with the current process and to suggest ways in which it might be improved.

In addition to the processes mapped during the site visits, institutional participants were invited to suggest other RIM processes which they felt should be standardised as a priority. While there were variants within the complete list of processes by each partner institution as priorities for standardisation and integration, the following processes were identified by all four partners: Benchmarking, Esteem Indicators, InCites Data Exchange (i.e. Data Exchange with TR), Awards Management (specifically Pre-Awards Management).

Criteria for Process Selection

The criteria for deciding which processes were to be examined in further detail were largely based on:

- Whether the process was already comprehensively researched and whether further research would result in wider benefit to the community
- The strategic importance to the institution
- The consensus with other project partners, time and the limited resources of the project

RIM Process Map Template

The RIM Manager / Officer worked with the CERIFy team to draw up the 'as is' and 'to be' processes using the 'RIM process Map' template provided. These maps, which originated before and during the site visits, were typically refined and completed after the visit through correspondence and feedback. They included a written description of the process with any issues noted.

3.2 Data Discovery and Data Analysis

The CERIFy Data Surgery[10] was attended by representatives of all partner institutions. The function of the Data Surgery was to finalise the process maps with the partners and to use the resulting

[10] CERIFy Data Surgery, http://cerify.ukoln.ac.uk/node/233 [Accessed 20/04/2012]

generic views of each process to share information and agree on the data elements and related matters associated with each process. The CERIFy Data Surgery included moving from 'RIM Process Mapping' to 'RIM Data Discovery and Data Analysis'. The 'Process Based Investigation Sheet' recorded detailed information behind each selected 'as is' and 'to be' process, e.g. Process, Step, System used, whether the system was auto accessed, whether the data element was to be exported for external reporting, Data sources, Data standard and Data field names. Agreement across institutions was made as to the generic data elements that would be required for the four processes identified. The results of the CERIFy Data Surgery are available via the Project's webpages and include; multimedia interviews with the participants, the process maps for each of the four processes analysed, the results of the user requirements elicitation work which had been undertaken during and after the site visits, and the results of desktop research by the Project Team in relation to each of the four processes.

3.3 CERIF Mapping / Modelling

Agreement amongst partners was made as to the defined 'generic data elements' for each of the 'to be' processes in the CERIFy Data Surgery. Two of these processes were chosen, using a further application of the criteria for process selection (detailed above): *InCites* and *Esteem*. The data elements, systems and reporting requirements associated with each of these processes were detailed thoroughly in close liaison with those project partners with the most experience in the specific processes and with TR. Those data elements were refined, agreed, modelled and then mapped to CERIF in a CERIF Mapping / Modelling Workshop with Brigitte Joerg (EuroCRIS), the project team, expert partners and TR. We recommend this type of Mapping / Modelling workshop as a template for future applications which may help institutions to map and model their processes. The CERIF data model for the InCites process was subsequently tested in a two-way data exchange between the CERIFy CRIS and TR using CERIF XML and FTP. In order to illustrate the CERIFy approach to the two selected processes, the following sections detail: 1) InCites Data Exchange and 2) Esteem Indicators.

4 InCites Data Exchange

InCites from TR is a customized, citation-based research evaluation tool on the Web that lets users analyse their productivity and benchmark their output against peers worldwide. All of CERIFy's institutional partners recognised the value of exchanging institutional data on people and publications along with their institutional affiliations (departments, research centres etc.) with an authoritative source of bibliometric data, such as InCites, in order to improve the accuracy of the external bibliometric resource and achieve the closest possible match of its data with the institution's authorship. In addition, the possibility of taking that matched data back into the institution's CRIS, along with the citations metrics and analyses was seen as offering great potential for improved research reporting and research evaluation exercises.

TR already offered a mechanism for the exchange of institutional data with InCites, however this process used a non-standardised data schema and a relatively cumbersome exchange process. QUB had previously tested this process and was extremely well positioned to offer expertise and suggest improvements. TR was receptive to testing the new approach and deeply engaged with the data mapping to CERIF and subsequent data exchange.

4.1 InCites Data Exchange - “To Be” Process

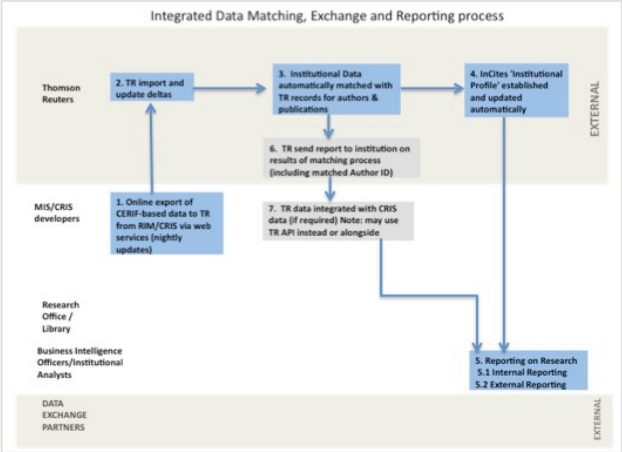

Figure 3: InCites Data Exchange ‘To Be’ Process

In the idealised ‘to be’ process for InCites data exchange it was agreed that data elements would be CERIF-based and only changes should be automatically ‘pulled’ by TR from the institutional CRIS each night. InCites would be updated more frequently by TR (weekly was suggested). The institutional data would be automatically matched to TR records and the ‘Institutional Profile’ would be automatically updated within the web-based InCites User Interface. A bi-directional data exchange process would also allow the ‘matched data’ in InCites to be sent back to the institutional CRIS, so that additional data on metrics and other information derived from InCites would be automatically incorporated into the institutional CRIS. While this ‘to be’ process could not be fully realised within the timeframe and scope of the project, it was used to inform the next stage of the project: data analysis, CERIF mapping and data exchange.

4.2 InCites Data Analysis, CERIF Mapping and Data Exchange

The data exchange process with InCites required a CERIF representation of the researcher with relationships to publications and organisations. This was strongly informed by the experience of QUB and the deep engagement of TR. The overview of the results and the proposed mappings and related entities and roles were formalized at the CERIF

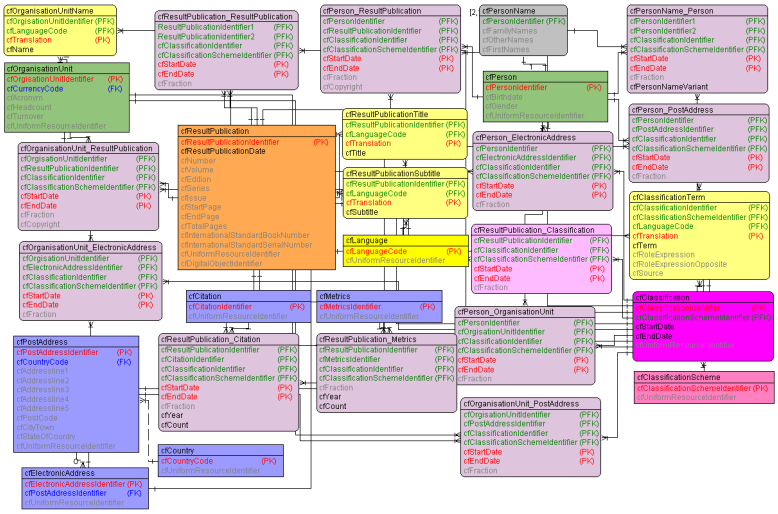

Figure 4: CERIFy InCites data model

Mapping / Modelling Workshop detailed above. Standardisation of data elements prior to the mapping with CERIF included the creation of a definitive list of publication ‘types’ based on a

number of sources, including CERIF, TR Web of Science and Research in View, Atira's Pure[11], DRIVER[12]/OpenAIRE[13], DCMI[14] etc. The resulting CERIF InCites Data Model was tested in a two-way data exchange process. Information on researchers and their publications from the UoB (and subsequently from Aberystwyth University) were entered into the CERIFy CRIS developed by TCD. The institutional structure (faculties, schools/departments, research centres etc.) for each partner institution had already been modelled within the CRIS by TCD. Data exchange between TR and the CERIFy CRIS was affected through CERIF-XML files extracted from the CERIFy CRIS delivered via FTP to TR and CERIF-XML files from TR sent back to the CRIS. A web services-based process to replace FTP is preferable and is currently under development by TR.

TR matched the data received from the institution with their own publications and bibliometric data and returned the results to the CRIS in CERIF-XML format for inclusion in the CERIFy CRIS. In addition, QUB completed initial testing of the generation of CERIF-XML files for In-Cites data exchange out of their local CRIS. A number of refinements to the CERIF InCites Data Model were made based on the results of the data exchange and the Data Model is now available for use by the wider CRIS/RIM/CERIF community.

5 Esteem Indicators

Ill-defined, unstandardized and partly field-specific, Research Esteem is often partially included with Impact measures, although Esteem usually focuses more narrowly on indicators of scholarly or disciplinary standing, particularly amongst peers. Commonly recognised Esteem indicators include Awards and Honours, specific memberships and elective positions, invited talks/keynote addresses, specific types of editorial roles, etc. Reporting on Esteem is a regular requirement for research evaluation analyses, yet the current processes for capturing and reporting this information were described by the CERIFy partners as opaque and 'woolly'. The result is a universal experience of a time-consuming, largely manual process which is ultimately unsatisfactory. The process was identified by all partners as one which would benefit from standardisation via CERIF and automation through a CRIS.

5.1 Esteem Indicators - "To Be" Process

In the ideal process, the researcher's profile in the CRIS is automatically populated with elements of Esteem indicators, see Figure 5 below. Suggested data sources for editorships and collaborations include InCites / Web of Science Premium API[15], Scopus API[16], PubMed[17] or other external

[11] Atira, http://www.atira.dk/ [Accessed 20/04/2012]

[12] DRIVER, http://www.driver-repository.eu/ [Accessed 20/04/2012]

[13] OpenAIRE, http://www.openaire.eu/ [Accessed 20/04/2012]

[14] Dublin Core Metadata Initiative, DCMI, http://dublincore.org/ [Accessed 20/04/2012]

[15] Web of Science Premium API, http://researchanalytics.thomsonreuters.com/solutions/webservices/ [Accessed 20/04/2012]

[16] Scopus, http://www.scopus.com/ [Accessed 20/04/2012]

[17] PubMed, http://www.ncbi.nlm.nih.gov/pubmed/ [Accessed 20/04/2012]

resources: in addition to bibliographic and bibliometric data these sources provide information on collaborations (national/international; by institution; by country). The development of standardised lists of awards and honours, high quality conferences etc. would assist the automation of this process. Alongside the automatic, ongoing population of the CRIS with data from available authoritative sources, researchers should update the CRIS with additional details of Esteem indicators. A fully populated CRIS with standardised Esteem data can be queried and reports produced as required. Reporting can be filtered and combined with other data before being made available to external agencies and funders. The information (for example, on Awards and Honours) can be used to populate School and researchers' personal webpages.

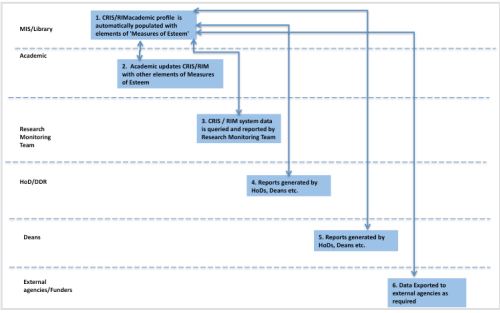

Figure 5: Indicators of Esteem 'To Be' process

5.2 Esteem Indicators Data Analysis and CERIF Mapping

Standardisation of the 'Esteem' process required a conceptual model for "Indicators of Esteem" which could be modelled in CERIF. Data elements required for the Esteem process were identified at the CERIFy Mapping / Modelling Workshop. As with the InCites process, the expertise of Brigitte Joerg, the project team and expert partners shaped the data model.

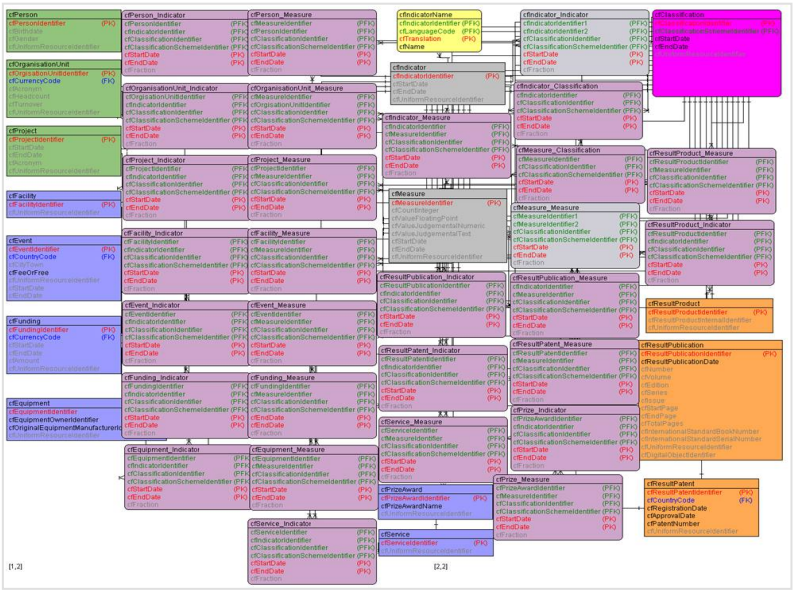

Figure 6: Employed CERIF entities for Esteem Indicators

Similarities with the data model developed for the JISC-funded MICE Project[18] (on Measures of Impact) were recognised and the MICE Model was adapted for use with Esteem. This Model is now available for use by the CRIS/RIM/CERIF community (please see the resources section of the CERIFy website).

[18] Measuring Impact under CERIF, http://mice.cerch.kcl.ac.uk/ [Accessed 20/04/2012]

Refinement of the data elements associated with Indicators of Esteem subsequent to the CERIF Mapping / Modelling Workshop resulted in the development of a 'definitive' mapping of Esteem indicators used in CRISs such as Pure[19], Research in View, TCD's Research Support System (RSS)[20] and in research evaluation exercises and analyses (including Australia's Excellence in Research Initiative (ERA)[21], UK Research Excellence Framework (REF)[22] and others). It was noted that data from the InCites data exchange process could contribute to Esteem reporting namely in the area of 'contribution to discipline' and collaborations (required for REF). The resulting selection of data elements representing indicators of esteem was mapped to CERIF, so that they could be included in a CERIF XML extract and provided to the research information community as an XML Schema for reference and comparison among data suppliers at every level.

6 Conclusions and Recommendations

The CERIFy approach demonstrated to research information stakeholders that creating and exchanging data using CERIF is considerably less daunting than it might appear. Two elements contributed to this, firstly, the processes CERIFy worked with arose naturally from the requirements of the RIM community as represented by the CERIFy partners. A key part of the CERIFy approach is to encourage RIM stakeholders (and members of the academic/research community, in particular) to identify and define the processes of greatest benefit to themselves. Harnessing community engagement from the outset to shape the development of the CRIS and its associated processes offers the best chance for the successful implementation and acceptance of these systems within an institution. Secondly, the availability of the data models associated with specific processes is of key importance. New CERIF data models have been developed for two RIM processes (InCites and Esteem) and a major commercial supplier of data to the HEI sector, TR, has now been 'CERIFy-ed': this company is now prepared to exchange data with universities using CERIF. TR has already improved its InCites data exchange process directly as a result of working on CERIFy. The data model developed by CERIFy can now be applied generally and availed of by other companies.

CERIFy revealed that institutional RIM processes vary only to an extent. There is a surprising level of commonality across the UK HE institutions in their experience of the issues, requirements and solutions associated with RIM processes and a concomitant desire to rationalise and standardise many of these processes. CERIFy also learned that all RIM processes, even those which look like standard production processes (e.g. Pre-Award Management) now include a business intelligence function. This adds a range of new reporting requirements to all RIM processes and reinforces the case for integrated/interoperable systems.

The CERIFy project revealed the need for the extension of CERIF in some areas (e.g. InCites includes valuable 'relative indicators' as part of the bibliometric information it supplies). The development of identifiers at a more granular level (for institutional units, for publications, for

[19] Atira, http://www.atira.dk/en/pure/ [Accessed 20/04/2012]

[20] Trinity College Dublin's Research Support System, http://www.tcd.ie/research/rss/ [Accessed 20/04/2012]

[21] The Excellence in Research for Australia (ERA) Initiative, http://www.arc.gov.au/era/ [Accessed 20/04/2012]

[22] Research Excellence Framework, http://www.ref.ac.uk/ [Accessed 20/04/2012]

'provenance' i.e. the source of the data) is vital for effective data exchange using CERIF. Recognition of CERIF by international standards agencies such as ISO was seen as important.
Finally, the experience of CERIFy reinforced our understanding that it is only when RIM is built in to institutional strategies and embedded in institutional policies that the associated systems, datasets and personnel can deliver the promised cost-benefits, efficiencies, improvements in business intelligence and services to the research community.

Acknowledgements

The authors acknowledge the invaluable input of the following people: Josh Brown (JISC), Patricia Brennan and the Thomson Reuters team; the UKOLN team; the TCD team (particularly Kevin Kiely), the Queens University Belfast partners and participants, especially Ricky Rankin, Gavin Mitchell, Craig Wooton; the University of Huddersfield partners and participants, especially Kirsty Taylor; the Aberystwyth University partners and participants, especially Hannah Payne; the University of Bath partners and participants, especially Katy McKen along with Brigitte Joerg (EuroCRIS) and Róisín Croker.

References

Russell, R. (2012): *Adoption of CERIF in Higher Education Institutions in the UK: A Landscape Study:* Report. Available at: http://www.ukoln.ac.uk/isc/reports/cerif-landscape-study-2012/CERIF-UK-landscape-report-v1.0.pdf [Accessed 20/04/2012]

Rogers, N.; Huxley, L.; Ferguson, N. (2010) *Exchanging Research Information in the UK: final report.* Project Report. Available at: http://ie-repository.jisc.ac.uk/448/1/exri_final_v2.pdf [Accessed 20/04/2012]

Bolton, S. (2010): *Business case for the adoption of a UK standard for research information inter-change.* Report to JISC. Available at: http://www.jisc.ac.uk/media/documents/publications/reports/2010/Businesscasefinalreport.pdf [Accessed 20/04/2012]

Casey, J., Proven, J., Dripps, D., "*The TrustDR Framework: a useful conceptual model of IPR and DRM*" http://trustdr.ulster.ac.uk/dev_pak/Readings/Reading_3_TrustDR_Framework.doc [Accessed 20/04/12]

Contact Information

Mahendra Mahey
UKOLN, University of Bath
Claverton Down
Bath
BA2 7AY
UK
m.mahey@bath.ac.uk

Ms Niamh Brennan
Research Information Systems & Services
Trinity College Library Dublin
College Green
Dublin 2
Ireland
niamh.brennan@tcd.ie