Exchanging Research Information in the UK

EXRI-UK: A study funded by JISC

Nikki Rogers, Lesly Huxley, University of Bristol
Nicky Ferguson, CLAX Ltd

December 2009
CONTENTS:

1 INTRODUCTION .............................................................................................................. 3
  1.1 BACKGROUND .................................................................................................................. 3
  1.2 EXRI-UK PROJECT ......................................................................................................... 3

2 SCENARIOS .................................................................................................................... 4
  2.1 HIGHER EDUCATION INSTITUTIONS ................................................................................ 4
  2.2 INDIVIDUAL PRINCIPAL INVESTIGATOR (PI) ................................................................. 5
  2.3 RESEARCH OFFICE – EARLY STAGE (PRE-APPROVAL) ............................................... 6
  2.4 RESEARCH OFFICE – POST-APPROVAL ....................................................................... 7
  2.5 WEB PROMOTION OF AN INSTITUTION'S RESEARCH .................................................... 9
  2.6 RESEARCH COUNCILS .................................................................................................... 9
  2.7 HIGHER EDUCATION STATISTICS AGENCY ..................................................................11
  2.8 HEFCE, SFC, HEFCW AND DENI ..................................................................................12

3 CRITERIA, BENEFITS AND BARRIERS ...................................................................... 14
  3.1 INTRODUCTION ............................................................................................................... 14
  3.2 BENEFITS ....................................................................................................................... 14
  3.3 BARRIERS ....................................................................................................................... 15
  3.4 EXTRAPOLATED CRITERIA ............................................................................................. 15

4 TECHNICAL APPRAISAL ............................................................................................. 16
  4.1 CRITERIA FOR A RESEARCH INFORMATION EXCHANGE STANDARD .................. 16
  4.2 DATA MODEL ON WHICH THE STANDARD IS BASED .................................................. 16
  4.3 SYNTAX FOR EXCHANGING DATA ............................................................................... 20
  4.4 POLICY ......................................................................................................................... 22

5 RECOMMENDATIONS, ROADMAP AND RISK ASSESSMENT ................................. 23
  5.1 RECOMMENDATIONS .................................................................................................... 23
  5.2 ROADMAP ..................................................................................................................... 26
  5.3 RISK ASSESSMENT ........................................................................................................ 27

APPENDIX A: EXRI – UK PROJECT BACKGROUND ....................................................... 28
APPENDIX B: REFERENCE GROUP MEMBERSHIP AND OTHER INTERVIEWEES ..... 30
APPENDIX C: STAKEHOLDER GROUPS .......................................................................... 31
APPENDIX D: QUESTIONNAIRE ....................................................................................... 32
APPENDIX E: GLOSSARY ................................................................................................. 35
APPENDIX F: FURTHER REFERENCES ................................................................................ 37
APPENDIX G: FURTHER REFERENCES ................................................................................ 38

The project was led by the University of Bristol’s Institute for Learning and Research Technology (ILRT¹). The team comprised:

Nikki Rogers (Project Manager, ILRT)
Dr Lesly Huxley (Project consultant, ILRT)
Nicky Ferguson (Project consultant, CLAX Ltd²)

¹ ILRT. <http://www.ilrt.bristol.ac.uk/>
² CLAX Limited. <http://clax.co.uk/>
1 INTRODUCTION

1.1 BACKGROUND

Universities need to manage information about the research they host, in order to inform strategic decisions about that research, to ease reporting to external stakeholders such as funding councils and research funders, and to offer useful services to those within and beyond the institution's boundaries.

There is a lot of work at the moment in this area in the UK, complementing that in other countries. In both the Netherlands and Denmark, for example, universities use a common system to document core information about research (METIS\(^3\) and PURE\(^4\) respectively). Both of these systems are based around the CERIF\(^5\) data model. In the UK, JISC, HEFCE, the Research Councils and others are funding a range of work to help the sector better manage and exchange information about research, covering institutional infrastructure (joining up institutional systems), national infrastructure (building services and interoperability to share research information), as well as providing guidance, support and opportunities to share experiences and work together.

By 'research information' in this context, we mean information about research activity and projects, researchers, funding organisations, research organisations, funding streams, research outputs and impact.

During 2009-2011, JISC is funding a significant strand of work covering these areas and entitled Research Information Management.

1.2 EXRI-UK PROJECT

The JISC-funded Exchanging Research Information UK (EXRI-UK) project involves an exploration of current and future scenarios for the exchange of research information in the UK. An outline specification of work packages for the study was developed during discussion at a workshop to review CERIF on 8 June 2009. The study is intended to describe scenarios for research information exchange, outline key requirements and benefits and to appraise options, with related risks and 'roadmap' of actions. The project also aims to explore specifically whether any particular format (e.g. CERIF) would be a suitable format to adopt as a standard for research information exchange in the UK.

We were asked to work in a consultative way with relevant stakeholders and established a Reference Group as a core element of that consultation. Project deliverables were released in phases to the Reference Group for feedback prior to incorporation in our final project report, delivered to JISC on 23 December 2009.

The project was asked:

i) To identify and document scenarios, requirements and criteria for exchanging research information in the UK.

To produce the scenarios, we were asked to work in a consultative way with all relevant stakeholders throughout the project's lifetime. The scenarios (Section 2) in the main report aim to identify and describe situations, and related high level requirements, where exchanging research information does or will offer benefits.

ii) To appraise the options and, specifically, whether any particular format for exchanging research information (e.g. CERIF) would be suitable.

The Technical Appraisal (Section 4) in the main report fulfils this requirement.

---

\(^3\) METIS <http://metis.hosting.kun.nl/metas/>

\(^4\) PURE <http://atira.dk/en/pure/>

\(^5\) CERIF <http://www.eurocris.org/cerif/introduction/>
The report also includes final recommendations, a Roadmap and an outline risk assessment, fulfilling the other requirements of our brief:

a. To work with key stakeholders to develop an initial UK Roadmap of actions for successful implementation of the candidate approach.

b. To develop an outline risk assessment relating to the Roadmap.

2 SCENARIOS

This section of our final report forms Deliverable A in line with our project plan and aims to "identify and document scenarios for exchanging research information in the UK". The scenarios identified below are based on the initial stakeholder mapping and interviews undertaken in the project. Stakeholders are detailed in appendices to the final report.

2.1 HIGHER EDUCATION INSTITUTIONS

Higher Education Institutions incorporate many role-based scenarios covering the entire research lifecycle. We had intended to draw a distinction in our analysis between research intensive institutions and those with a less intensive research profile, but our discussions with stakeholders convinced us that, although the administrative structure and resources may vary, there is no difference in the essential challenges and needs for the exchange of research activity information. Figure 1 below reflects the many and varied interactions between the institution and external agencies and the institution's own internal systems and process interactions.

![Figure 1: Research Information Exchange – Higher Education Institutions](image-url)

* Institutional systems to process staff, student, financial and publications data and so on are implied conceptually within the large light blue central box in this diagram. Solutions for how these tasks are actually organised and integrated are decided on by institutions (e.g. through repositories, data warehouse layers, IRMAS, CRIS systems etc.). However, exchange standards may benefit internal data exchange too, though conformance would be optional.
2.2 **INDIVIDUAL PRINCIPAL INVESTIGATOR (PI)**

2.2.1 **Research information exchanges**
- Submits a CV when joining an institution, often containing publications information that the institution will need
- Submits staff information forms to their institution, with details of gender and ethnicity and so on often then used by the institution in HESA returns
- Submits proposals to research councils by filling in Je-S form
- Validates and adds to their institutional publications information for the RAE/REF
- Validates and submits outputs information (including publications information) to research councils
- May submit proposals and outputs to other funders, such as Charities or Industrial partners
- Tries to find out about the research being conducted by a researcher (or group) with whom they would like to collaborate, and/or that researcher’s contact details

2.2.2 **In summary**
The scenario for an individual PI is based around the typical research lifecycle.

2.2.3 **Commentary**
An Individual PI may receive research opportunities information from their institution's research office or from external sources (such as via email subscription to ResearchResearch). Simultaneously and/or consequently they may be networking with or seeking potential partners with whom to collaborate in pursuing research opportunities.

Such networking may well take place in forums that don't require the exchange of digital information (for example at conferences). However PIs may typically also seek information about other researchers via the web. If they visit the personal page of a researcher at another institution, or the research forum of a group at another institution, they would need to have accurate research information. (Note: this activity varies widely across subject disciplines - in some narrow scientific disciplines for example global researchers in one research area are all known to each other. However interdisciplinary research information may also be relevant).

Where information about UK research and researchers is not easily available we could suggest there is a consequential loss to UK research activity - see for example Les Carr's recent blog on difficulties in trying to retrieve information about research at another institution ('Taking Communication Seriously', Blogpost September 2009)\(^6\). In addition to publications, the researchers themselves may well also want to display items like 'esteem indicators' and potentially 'evidence of impact' (depending on decisions about how impact will be measured by different funders).

Information about previous supervision (and examination) of post-graduate research (PGR) students is often required by HEIs at some stage. Of less importance to the HEI but crucial for the PI is previously won research income (although this is required on some proposals for research funding).

The next stages in the research lifecycle for a PI may typically involve the development and submission of an application to a funding body, with a necessary internal approval process. They may well enter research data (about their proposal) into the Research Councils’ Je-S system or into some other system for the electronic submission of a proposal for funding. If an award of funds is later made that information will be captured - most likely electronically within an institution's central or departmental system and may or may not ever be linked to the original submission data. We note that while a Research Management and Administration System (RMAS)\(^7\) may be a key part of this process, the data generated may well contain information useful for business intelligence or for a later REF submission exercise, and

---

\(^6\) [http://repositoryman.blogspot.com/]
\(^7\) [http://as.exeter.ac.uk/rmas/report.htm]
therefore there is linkage here with a CRIS-like system\(^8\). Undertaking and managing the research itself may involve collaboration with other institutions or bodies.

Reporting on research carried out will take several forms: progress reports internally and to funders, budget outputs internally and to funders, final reports internally and to funders, summarising outputs, annual submissions to funding bodies (ultimately to be through a single Research Outcomes system for the Research Councils, but through several such systems in the near future), responding to institutional calls for making annual HESA returns and perhaps annual or perhaps irregular calls from institutions for updates to repository or CRIS with a view also towards approximately 5-yearly compilation of information for future REFs.

### 2.3 RESEARCH OFFICE – EARLY STAGE (PRE-APPROVAL)

#### 2.3.1 Research information exchanges

- Collates and disseminates information about research opportunities to PIs
- Collates information about potential partners for grant bids, consortia and partnership developments for PIs
- Gathers and where necessary synthesises/interprets information on regulations and frameworks relating to different funders or partners, including government (ethics, animal experimentation, etc) and makes available to PIs
- Negotiates agreements with external bodies
- Records and monitors applications/partnerships in progress
- Collates information from own internal systems and data gathered from external bodies (e.g. research councils) on success rates
- Disseminates reporting information on success rates to senior management, faculties and departments (taking into account sensitivity issues about this information being widely distributed)

#### 2.3.2 In summary

The Research Office role may be largely centralised in an HEI, although at least some activity listed above may also take place at local (department, school, faculty, college) level. The pre-approval phase of research information exchange involves many interfaces with individuals within and outside the HEI, and with a variety of internal and external systems holding research information. Some processes are ad hoc rather than systematic and regular. Processes include gathering, synthesising and disseminating information about research opportunities, potential partners and any relevant legal or other frameworks; providing resources to support application development, monitoring application progress and reporting on success rates.

#### 2.3.3 Commentary

Research opportunity information may be provided via a third party (for example the UK Research Office - UKRO\(^9\)) and/or through regular harvesting or manual gathering of information from Research Councils’ websites for example. Delivery of opportunity information may be via dedicated web interfaces, email or other alerts and may be created automatically or manually. The Research Office may match opportunities to PIs based on their own knowledge and/or preferences specified by PIs and/or on the basis of subject, discipline, field etc information held in research management or corporate systems.

There is a general sense of ‘direction of travel’ from the government towards more collaborative research, regional partnerships, consortia etc. The Research Office will rely to some extent on PIs’ knowledge of their own field, as the opportunities for gathering transparent information systematically about potential partners are currently limited. Much of the information (from HESA, from the Research Councils and HEIs’ own websites) about other HEIs can only be aggregated at a high level, making comparisons of like-for-like difficult and assessment of potential partners’ strengths something of a ‘leap of faith’. Partnerships with industry or the increasing development of HE-NHS partnerships for research increase the level of complexity and highlight the difficulties of research information exchange between sectors.


\(^9\) UK Research Office <http://www.ukro.ac.uk/>
Pre-submission tools, particularly for costing proposals, may be provided by the Research Office or, for example, the Finance Office. Financial information may be integrated from corporate systems via the Research Office and any tools (web-based or spreadsheet templates) they supply for costing and pricing, or this may be drawn directly by the PI from finance systems. The Research Office monitors progress of applications and gathers information on success rates. This may be via internal tracking and reporting systems, and/or on the basis of information supplied directly by Research Councils and other funders or indeed internally from PIs or local organisation units. This may be in spreadsheet format and may or may not be able to be imported into internal systems, where these exist. There is still a level of immaturity in many HEIs’ research management systems and processes so that the tracking and reporting of success rates can be difficult; however it is done, it is a sensitive issue that requires various permissions to be built in to any views on the data made available to deans, heads of department/schools etc.

2.4 RESEARCH OFFICE – POST-APPROVAL

2.4.1 Research information exchanges

- Coordinate, collate, post-process and synthesise information required for RAE/REF, drawing on local operational (e.g. HR, Finance, Student, Repository) systems and 3rd parties (publishers, citation indices, applications such as Symplectic as half-way houses)
- May specify and maintain bespoke system for collation, modelling and submission of RAE/REF data (from spreadsheet to relational system with web interface)
- Have sight of and some 'late stage' input at strategic / analysis levels of operational returns e.g. for HESA finance, students, staff
- Coordinate, collate, post-process and synthesise information required for Research Activity Survey (soon to be incorporated in HESA) (in Scotland, 'Other Activity Indicators' instead of 'Research Activity Survey')
- Develop mappings between local/central systems containing research information and data requirements for RAE/REF and other statutory returns, where necessary, trying to address variations in interpretation/definition of data fields for different purposes (for example, FTEs, student ‘year’, disability status, unit of assessment, subject and cost centre)
- Collate and exchange data on postgraduate research, research seminars and research equipment
- Knowledge transfer activities
- Consultation with partners, funders and clients to inform the strategic development of research activity
- Undertake analysis of data from HESA and other statutory returns for internal purposes
- Analyse available data on composition of RAE/REF panels and requirements to maximize submission
- Work with departments/faculties/PIs to ensure that non-funded research can be included as necessary (not necessarily captured in institutional systems re: grants etc)

2.4.2 In summary

Post-approval, the Research Office role will involve monitoring research progress, supporting dissemination and exploitation (including knowledge transfer) and coordinating all information from local, central and 3rd party 'systems' - however basic or sophisticated - for the RAE and, in future, the REF. The latter is a strategic activity, involving selection and choice in submission of staff, of their submission against which Units of Assessment, and which outputs and impact to include.

Beyond the Research Office role, a number of related strategic or operational roles may be undertaken - again, centrally or locally - which involve research accounting and the gathering of information on staff and students for internal, operational, organisational purposes. These will feed into the strategic processes relating to REF/RAE, to the internal strategic business intelligence relating to the monitoring and enhancement of an institution's research impact, research income and research capacity and to the exploitation of research through knowledge transfer activity.
2.4.3 Commentary

The Research Office role-holders interact at human and systems levels with many staff and systems across the university and with external bodies to collate, post-process and submit the appropriate data. Some have specified and implemented bespoke systems that provide a coordinating and modelling layer over corporate and external systems and data. Others make use of spreadsheets and internal reporting. They often encounter tensions between the need to allow local views and manipulation of data and reporting, and the strategic institutional needs, particularly where local systems prevail.

Whilst they may have some input during final stages of operational returns e.g. to HESA, their primary role is to coordinate the processes involved in the RAE/REF submission. This is seen as an important strategic activity with clear funding implications. A wide range of people, functions and systems from across the institution is involved in contributing data as part of the normal operation of the institution, for other statutory returns (e.g. HESA) or specifically for the RAE/REF return. Submission to the RAE/REF is based on requirements specified by HEFCE; internal systems will need to be mapped or adapted as requirements change.

The RMAS report has documented the difference between the support from administrative staff provided to PIs up to the point at which a grant is awarded, and that provided post-approval, when often a different range of systems and processes apply, often moving from local systems and spreadsheets to corporate systems for the management of research and/or specific systems for collating REF data. Although much of this takes place within the institutional ‘black box’ (or in the light blue box in figure 1), much of these data and processes are also of use for institutional business intelligence, planning and benchmarking – and it is at the latter point that research information exchange standards may be of particular use.

Figure 2 below (reproduced with kind permission from St Andrews University) shows in more detail some of the ‘black box’ interactions as well as exchange of information between an institution and external bodies; section 2.5 below also refers to that element of the diagram relating to web-based promotion of research. Other green boxes on the left might include ‘Estates’ and ‘Facilities’ data (for space and equipment usage).
Note that the label "complex and inflexible" in the diagram above is not a value judgement on central facilities, more a comment that the centralised systems tend to be designed for internal purposes and functions very different from those performed by a CRIS or required for collation and submission of research data in a number of different forms to a number of different funders and data gatherers.

### 2.5 WEB PROMOTION OF AN INSTITUTION’S RESEARCH

#### 2.5.1 Research information exchange

- Identify research information for promotion in the public domain and publish it via personal researcher web pages and other institutional website features.

#### 2.5.2 Commentary

A research intensive institution may have a dedicated research web group and be increasingly seeking how best to promote the research data exemplifying their institution via the web. For example, publishing international collaborations via sophisticated visualization interfaces (for example geographical views) on web pages may be a priority. Institutional websites can be a useful vehicle through which to attract researchers at other institutions or potential industry partners interested in research collaborations, or simply to disseminate research outcomes to the local community and the general public. We may anticipate that as the current government Linked Data focus gathers momentum future scenarios valuable to the UK will require the ability for research outputs to be cross-referenced with government reports and initiatives for example.

The inter institutional sharing of research data, such as with “pools” in Scotland and other regional (or other) collaborations elsewhere across the UK is another point at which comprehensive research information can be exchanged. St Andrews and Aberdeen Universities, for example, anticipate they will be sharing research data that will benefit them in research collaborations, such as potentially exchanging facilities-sharing data as well as recording proportions of researcher FTEs dedicated to research work and so on. The ResearchRevealed and BRII projects funded by the JISC at the Universities of Bristol and Oxford respectively are planning to demonstrate how to share research information across their institutions by the end of 2010 and using Semantic Web technology.

### 2.6 RESEARCH COUNCILS

Figure 3 below shows the main areas where research activity information is exchanged between the Research Councils and outside bodies/individuals. It should be noted that many other levels of complexity could be added – showing data exchange between the Research Councils themselves, for example. And also that what happens in the light blue box is under the direct control of the council and therefore considered an internal process. This may include exchanges with institutes or programs or centres (some of which may be at HEIs and some at Research Council premises). The exact details of where and who should not concern us, suffice it to say that a standard for the exchange of research information, if agreed, would be essential for use in exchanges which go outside the light blue box and optional but potentially helpful for some interactions inside the light blue box i.e. within the council's internal processes (e.g. with "internal" institutes etc).

---

10 This September 2009 blog post by Les Carr at Southampton University illustrates an increasing need to make research information available to others on demand via the web: <http://repositoryman.blogspot.com/2009/09/taking-communication-seriously.html>  

2.6.1 Research information exchanges

- Exchange (collect and feedback) activity data from over 150 HEIs
- Allow ongoing data input from organisations and individual PIs, identified by grant/project\(^{12}\)
- Exchange data on postgraduate research and researchers
- Capture activity and facility data from intra mural research centres which may not be HEI-based
- Collect, collate, package and present output data
- Present activity data for government, public and press
- Collect and archive/store integral research data/datasets, allow and control access
- Exchange data with a variety of peer reviewers for funding and assessment
- Collect data to inform strategic decisions on areas (themes) of funding
- Collect outputs for open access / repository purposes Collect data from postgrad students, exchange with HEIs and research centres
- Exchange with other parts of government e.g. ONS
- Use HESA data for destination analysis and context setting
- Exchange activity data with international partner agencies/councils/programmes/ and collaborators
  - Exchange data/metadata with subject or institutional repositories
  - Exchange data/metadata with publishers

2.6.2 In summary

The Research Council scenario is based on communicating with a large number of other parties in an environment where the specifics of the information exchanged are liable to frequent change. Any standard would therefore need to be flexible and adaptable.

\(^{12}\) Note that in contrast to HESA (annual retrospective from institution) and REF (every 5 years from institution), Research Councils have allowed ongoing data input from grant holders; one possibility is a future move to annual collection from HEI, but they will allow ongoing input of publications etc.
2.6.3 Commentary
Historically the Councils have collected data from individual PIs and the Councils’ own centres or programmes, but they are moving towards encouraging institutions to collect data from the PIs and submit that data to the Councils. The similarities between this process and the REF would then be more obvious and visible. There are also significant moves towards harmonisation of systems between Research Councils (both outward-facing and back end); and also towards harmonisation of vocabularies and classification schemes.

The Councils recognise that now is the time to be seriously considering adopting standards to facilitate data exchange - with HEIs, with other government agencies, with press and publishers with their own institutes, councils and programmes and with each other. In the past, research activity data has been made available on an ad hoc basis but it is recognised that in the future it will need to be made more generally available, even if one disadvantage of this is that data may be misinterpreted or used for purposes for which it was not intended.

2.7 HIGHER EDUCATION STATISTICS AGENCY
The Higher Education Statistics Agency (HESA) is the official agency for the collection, analysis and dissemination of quantitative information about higher education. It provides a system of data collection, analysis, and dissemination in relation to higher education in the whole of the United Kingdom that aims to deliver, on a charitable basis, the information required by Higher Education Institutions for their planning and development. It also operates on behalf of the higher education sector to support and enhance the ability of government and its agencies to determine higher education policy and allocate funding and to meet the public interest in the availability of quantitative information about higher education. HESA data covers research students (in the individualised HESA Student Record), academic staff (the individualised HESA Staff Record) and some research-specific financial information in the HESA Finance Record.

2.7.1 Research information exchanges
- Annual retrospective data collection from all publicly funded UK HEIs
- From multiple points within an institution (e.g. HR for staff, registry for students)
- Mostly statutory requirements but some data (e.g. Research income by cost centre in the Finance Record) is there because the sector has said that it wants this information collected/published centrally
- Analyse and publish through Heidi –<http://www.heidi.ac.uk/>
- Supplies full datasets to Statutory Customers
- Statutory Customers include HEFCs for UK, Research Councils, TDA plus government departments, for example, BIS
- Ad hoc reports for commercial clients
- Advisory service on how to use the data, IP rights and responsibilities
- A quality assessment process takes place with many iterations of contact and clarification with institutions and multiple redrafting (typically 20+)
- Some tracking of postgraduates after they leave as impact measure

2.7.2 In summary
HESA has a mainly statutory responsibility for data collection but staff also emphasise that government reports talk of "collecting data once and using it many times" and indeed the printed version of the REF guide specifically mentions and recommends HESA's likely role in the REF and other future data collection exercises. HESA collect data from central points within institutions and do not currently deal with departments or individual members of staff. They will in the future be collecting data on research facilities and buildings. There are significant differences between the student data and the staff data which they collect, the most obvious is that staff names are not currently collected.
2.7.3 Commentary

Examples of the use of data include:

- By HEFCE for monitoring funding allocations
- By BIS for monitoring and developing policy
- By HESA itself for producing and publishing performance indicators

HESA have developed tight data protection terms to allow them to safely deal with non-statutory customers at the individual data level. They already use standards such as ONS country codes (ISO standard) and a standard character set, which is a subset of Unicode. HESA do not comment on the internal processes the institutions use to create that data as long as it is in the correct format. So HESA collects statistical data that impacts on institutions from the institutions. They do perform validation checks, credibility checks (e.g. "is it right that 80% of your students are female?") and a final verification from the head of institution.

The Statutory Customers will change data requirements but HESA aim to give institutions one year's notice of major changes (two years before collection). There will be an overhead in using a data standard both when changes are made to the standard and also when customers require a change that is not easily accommodated within the standard, so a standard must be flexible and reasonably responsive. HESA staff emphasise the desirability for the whole sector of a common lexicon (i.e. what is a "course", what is an "organisation unit", what is "a project"? They see the main benefits of a standard approach accruing to the institutions – whether data exchange occurs within or between institutions, a standard would be helpful and lead to efficiency savings.

The lack of staff identifiers collected could be a hindrance to further analysis – they record "primary employment function" which identifies cost centre and source of salary (identifying Research Councils individually) but this would clearly not be enough to properly assess research activity and funding, particularly as some staff have multiple and changing sources of funding.

HESA currently collect information about the institution as a whole, its staff, students and finances. Although some of the information they collect is research-specific, it would be an expansion of its role to collect the kind of detailed information required by the Research Councils. Similarly, collecting all the information the REF requires would be a wider remit. There is no obvious reason why this should not happen in the future and there are obvious efficiency gains in having all the HE data collection exercises under one roof, providing the responsible agencies were confident that their needs were being met. If each institution were only dealing with one point of collection for this data rather than three, then presumably this would also lead to efficiency gains at HEIs. It might be argued that in a world of linked data with all HEIs using a standard data format for research activity data it becomes irrelevant where the collection exercises are based, but it is likely that a single point of collection would push forward any progress towards standardisation and help to remove obstacles and barriers.

2.8 HEFCE, SFC, HEFCW and DENI

Out of the UK higher education funding bodies, we have interviewed only HEFCE and we use HEFCE as an example below. We recognise that there are some differences in terminology and collection activities across the UK but we regard the HEFCE example as sufficiently representative.

2.8.1 Research information exchanges

- Collect data from some 150 HEIs approximately every five years for a large scale assessment. The Research Assessment Exercise (RAE) as this was known has now been replaced by the Research Excellence Framework (REF), whose first collection is intended to take place in 2013 - <http://www.hefce.ac.uk/Research/ref/>)
- Segment all collected research data along subject lines and distribute to databases for REF evaluation panels, the data to include links to uniquely identified research outputs either in a digital form (for example publications in open access repositories) or in a material form (for example books temporarily housed in warehouses for access by REF evaluation panel members)
- Complement research data collected by retrieving further data possibly from Research Councils, from HESA – to be confirmed - (regarding institutional research income and postgraduate student details) and Citation Suppliers
• Publish data on REF outcomes (in the form of institutional profiles) via the HEFCE, SFC, HEFCW and DENI websites and through their official paper publications, with an accompanying media launch
• Eventually publish all REF submissions made (exactly how this is to be done is to be confirmed)

2.8.2 Summary
HEFCE collects data approximately every five years for a large scale Research Excellence Framework (REF) exercise, the first of which is intended to fully take place in 2013. REF is designed to be a comprehensive assessment of research excellence to enable funding allocations to HEIs within a highly selective funding system (QR).

2.8.3 Commentary
The data about research in HEIs is expected to come from four major sources (the first of which will be the most significant):

1. HEIs – Research publications, impact and environment qualitative information (narrative descriptions will be classified against impact indicators that are still being developed).
2. HESA – it is hoped HESA will provide a) research income b) post-graduate student data. This part will play a comparatively small role in the assessment.
3. Research Councils – it is envisaged that HEFCE may submit a query to Research Councils for research facilities data relating to particular individuals.
4. Citation indices - citation information should come directly from a citation supplier, with some level of view or edit for HEIs in relation to their submission.

HEFCE recognize that to acquire REF data via system-to-system uploads from institutions would be ideal (institutions having the opportunity to choose beforehand which case studies, citations and so on form part of their submissions per the Units of Assessments contained within their research faculties), but that they - HEFCE - will also have to continue to offer (as with the RAE) the option for submissions to be made via an online manual upload system. The manual option is for institutions simply unable to comply technically/organisationally with centralised systems that operate in a machine-to-machine online environment.

The pool of collected UK REF data is to be segmented according to subject disciplines and distributed to the appropriate specialist panels, each of which are made up of assessors (35 people or more). These assessors will look at a considerable amount of digital information, including the publications themselves, the metadata ideally linking to online full-text available from open access repositories or CRISes in the case of publications, for example, but in many cases they may follow information links that refer them to review hard copy books, for example.

HEFCE intend to use the REF data as evidence with which to assess quality of research being undertaken within institutions relating to key elements – research outputs, impact and environment. HEFCE will publish the results as profiles which for each institution show the proportion of their research activities falling into star rating categories. Overall institutional profiles will be used by HEFCE for government funding allocation decisions. HEFCE will use some of the REF data for their equality and diversity monitoring as well. How they will archive data (previously, for the RAE, via HERO - <http://www.hero.ac.uk/>) is yet to be decided.
3 CRITERIA, BENEFITS AND BARRIERS

3.1 INTRODUCTION

The scenarios outline the various processes in which research information is exchanged, for the range of stakeholders identified for the EXRI-UK project.

Deliverable B was to present the criteria that any recommended approach would need to meet to support UK higher education’s common research information exchange needs, derived from development of scenarios in consultation with stakeholders.

In some respects it has proved difficult to pull out criteria in isolation, as research information exchange in the UK is in its infancy with respect to any standardisation. Some criteria emerge in the scenarios; this section aims to summarise criteria based on the benefits and barriers to the adoption of a common research information exchange format specifically identified by stakeholders through interviews, meetings and consultation events. These criteria, or requirements, are explored in relevant sections of the Technical Appraisal (section 4) and inform, where appropriate, key points of the Roadmap included in section 5.

3.2 BENEFITS

Stakeholders identified the following anticipated benefits to the adoption of a common research information exchange format:

**Efficiency gains**
- Reduction of burden on HEIs and researchers in reporting terms – “collect data once, re-use many times”
- More efficient HEI internal systems – clearer reporting requirements leading to a better understanding of information dependencies and organisational workflows
- Reduction in the costs of information access and exchange across stakeholder groups
- Greater integration of statutory and other sector-wide collections of data
- Greater understanding of what questions can be asked usefully of the data, to the benefit of all stakeholder groups

**Making the most of data for UK research**
- Information exchange across the research lifecycle
- More open access by third parties, the public and researchers
- Increased promotion opportunities for UK research online, enhancing the profile of research projects, HEIs, Funding Councils and the UK government on an international scale
- Improved understanding of research by the general public
- More consistent provision of research and funding opportunities to researchers
- Improved evidence-based decision-making at HEI, funder and national levels
- Wider opportunities for like-for-like data comparison at greater levels of granularity, improving benchmarking in the UK, EU and more widely internationally
- Enhanced Virtual Research Environments (VREs) via standardised exchanges of information to support research collaborations and knowledge transfer in a more consistent way - within the UK and potentially in Europe and beyond

**Flexibility and interoperability**
- Opportunity to output data in a number of different ways for different purposes, e.g. HEIs would be able to interoperate more easily with funder systems, statutory reporting systems, other HEIs
- Enhanced ability to cope with changing requirements for statutory returns
- Easier to extend a common standard to accommodate changes over time with regular reviews and an agreed updating process
- Opportunity to arrive at common data definitions including for bibliographic data, a common lexicon, across various statutory requirements
3.3 BARRIERS

Stakeholders identified the converse of all of the above as barriers – now or in the future – to the adoption of a common standard, but the following were particularly highlighted:

Poor Use and Roll-out of a Standard
- The risk of a lack of common data definitions that would then lead to misinterpretation of a standard, thus hampering potential efficiency and comparability gains
- The risk that a lack of harmonisation of requirements across data collection ‘agents’ would counter gains made by the use of a common data standard and continue to force HEIs to undertake costly reporting exercises
- The risk that the standard does not provide sufficient coverage of the research domain to satisfy multi-stakeholder requirements
- The risk that a lack of documentation for a standard, such as a lack of appropriate mappings and conceptual models (e.g. Unit of Assessment, cost centre, subject, FTE etc) would make it difficult to apply
- Fears that the adoption of a standard that isn't flexible enough to respond to changes or accommodate new needs on a regular basis would cause significant problems for all stakeholders
- The risk that despite a standard, a lack of clear agreement on what research data is being collected and why would cause tension for HEIs developing or buying in systems to support research information integration and reporting (e.g. Research Councils’ focus on grants/projects; REF’s focus on Units of Assessment)
- The risk for a range of stakeholders that a lack of a common standard for DOIs would reduce the potential for interoperability gains

Lack of Resources
- Insufficient cost-benefit analysis to convince of the need for initial investment of time/effort/funds to adopt a new standard and anticipated benefits
- Requirement for specialist or additional technical resource in order to implement a common standard and input/output formats
- Few or under-developed commercially available systems
- With fuller integration/interoperability and richer data, the burden of information requests may become much greater, thereby negating efficiency savings

Fears over Security/transparency
- Potential anxiety over the re-use of data: what will happen to it next and how could it be used in different contexts?
- Lack of clarity about confidentiality, access and authentication for sensitive data

3.4 EXTRAPOLATED CRITERIA

The above summary of stakeholders’ perceptions of current and potential benefits and barriers suggest that the following will need to be taken into consideration in the technical appraisal, recommendations and roadmap to implementation of the adoption of any common research information exchange standard for the UK:

- Support for single collection, multiple re-use of data
- Comprehensive coverage of the research domain
- Compatibility with the wider (UK/international) research environments
- Interoperability with range of existing internal systems
- Extensibility and flexibility
- Regular review and update process
- Comprehensive documentation
- Agreed common data definitions and lexicon across various statutory requirements
- Agreed common lexicon
- National mappings for different requirements
- Open standard
- Clarification of purposes for which data are being collected
- Relatively low adoption costs proportionate to benefits
- Timescales for initial adoption and subsequent development align with existing initiatives
4 TECHNICAL APPRAISAL

4.1 CRITERIA FOR A RESEARCH INFORMATION EXCHANGE STANDARD

We describe here the key requirements for a UK standard for the exchange of research information. These are based on stakeholder consultation undertaken throughout the EXRI-UK project and a technical appraisal of contender options. For each requirement a summary evaluation of CERIF\(^\text{13}\) against other options is given. In other words, we consider in this "criteria" section the "fitness for purpose" of technical standards considered by the project for the exchange of research information between key agencies and organisations in the UK, given the differing and common requirements of those agencies.

We also note that the exchange of research information in standardised formats, such as compliant data with well-defined XML schemas, is intended for machine-to-machine exchanges of data, for example in institutional system uploads to funding authorities. Where institutions cannot implement or integrate systems to support a standard, we suggest that people such as researchers will need to make data entries by hand through a user interface effectively hiding – for usability purposes - the research exchange technical format. In other words, “hand crafting” XML-encoded data, for example, will not be necessary.

4.2 DATA MODEL ON WHICH THE STANDARD IS BASED

Data models formally define the conceptual model of entities/data elements and the relationships between them for a particular domain. In this case, we deal with the research domain and the data model should describe such concepts as Researchers, Projects, Organisations, Publications etc, and the relationships between them as well as the context of those relationships where necessary. This could include, for example, where we might want to state that Researcher X is linked to an Organisational Unit Y by a relationship such as “affiliation” and that that relationship may only be true for the time period Z – thus giving context to that relationship.

For this technical review we consider known Semantic Web\(^\text{14}\) approaches to modelling the research domain contrasted with the CERIF data model. The range of Semantic Web vocabularies considered includes:

- RES: A vocabulary for describing academic researchers: <http://www.medsci.ox.ac.uk/vocab/researcher/0.1>
- FOAF: The Friend of a Friend (FOAF) project vocabulary, describing people, the links between them and the things they create and do: <http://www.foaf-project.org/>
- DCQ/DCMI terms from the Dublin Core Metadata Initiative - <http://dublincore.org/>  
- OWL TIME for describing temporal content: <http://www.w3.org/TR/owl-time/>  
- ARPFO ioAcademic Research Project Funding Ontology: <http://vocab.ouls.ox.ac.uk/projectfunding/projectfunding-0.1.html>  
- Academic Institution Internal Structure Ontology (AIISO) <http://vocab.org/aiiso/schema>  
- Academic Institution Internal Structure Ontology Roles (AIISO Roles) <http://vocab.org/aiiso-roles/schema>

The vocabularies described above may be reused in part or in whole and must be used in combination if they are to offer a full enough specification of an extensible common schema for representing research information. Several JISC-funded projects have been using combinations of the above schemas independently and recently began to collaborate to develop alignment in their approaches (the DoTAC

\(^{13}\) <http://www.eurocris.org/cerif/cerif-releases/cerif-2008/>  
\(^{14}\) Semantic Web Advanced Development <http://www.w3.org/2000/01/sw/>
project at the University of Southampton, the BRII project at the University of Oxford and the ResearchRevealed project at the University of Bristol). We now consider CERIF and Semantic Web data modelling approaches in terms of a number of requirements arising from scenarios considered by the EXRI-UK project.

### 4.2.1 Coverage

A data model to encapsulate research information for the UK must be fully specified and detailed enough to provide suitable coverage for as wide a range of scenarios discussed in this report as possible.

**Discussion:** Early examples of institutional “mappings” to CERIF (that is, conceptual mappings of their internal system data models to CERIF 2008) are promising – the Universities of Aberdeen and St Andrews state that this exercise has proved much easier than expected.

The ResearchRevealed project at Bristol is developing a Semantic Web schema in partnership with the Oxford BRII and DoTAC, Southampton projects and with other interested parties. Early work builds on the DoTAC project’s initial attempt to map CERIF schemas to metadata vocabularies (see http://www.dotac.info/mapping/) and reveals that CERIF 2008 has good coverage of the research domain in many places, but with a small number of perceived gaps for example a way to model research collaborations between people or organisational units using a standardised semantic classification scheme. The mechanism is certainly there to explicitly express collaboration roles between people but it is not clear whether there is a standardised classification scheme to describe roles such as ‘manager’ or ‘collaborator’ or ‘co-author’ such as might be required with some degree of granularity for UK purposes. Eurocris do however state “additional classification schemes will be developed and formalized as part of the CERIF Semantics in upcoming CERIF releases.”

Regarding the CERIF 2008 data model, for the REF there are some potential omissions such as modelling “Impact” and classifications that will emerge as “Impact Indicators”. The finalised REF specification is not yet published however. The JISC-funded Readiness4REF project says it has successfully managed to map the RAE2008 specification in its entirety (since the specific inclusion of two new features in CERIF by agreement with Eurocris). For HESA records the data model might need to be extended to model the full coverage required of Student and Income data – the REF will also require information on these. EuroCRIS has indicated that this can be discussed.

For Research Councils the data model appears at first glance to be sufficient but a comprehensive mapping exercise and statement of data modelling requirements has not yet been attempted for all the UK Research Councils. Requirements to model industry collaborations and for institutions to create returns to funders such as Charities and Industry partners are also not likely to be fully met by the data model for CERIF 2008. However data model extensibility can cater for shortcomings as discussed next.

### 4.2.2 Extensibility

It is important to be able to change and extend the data model for new information needs or greater levels of granularity because requirements for collection and use of UK research information change and will continue to change over time. For example there may be further iterations of the REF, or changing government statistical requirements or changing business models in the institutional management of information in varying economic climates.

**Discussion:** We note that workshops held by Eurocris at St Andrew’s University in Autumn 2009 were able to establish the need for additions to CERIF 2008 (in order to record personal costs linked to people’s names – previously not done due to EU law – and percentage FTEs for researchers working on...
projects). This sort of extensibility appears possible on a continuing basis for the CERIF standard, without compromising its core data model and thus not hindering interoperability gains in a wider context than just the UK. Membership of Eurocris is offered at a fee and full membership allows institutions and individuals to cast votes for changes to the CERIF standard. It may be possible for the UK to adopt a subset of CERIF, providing additional contextual data (outside of the minimal requirement) useful for understanding (by machine or human). Should additional entities, attributes or classification scheme terms need to be added then this is possible with the voting and CERIF extensions mechanisms.

We note that Semantic Web schemas are similarly extensible. Extending a standard over time, of course, also implies the need for version control and this is discussed below.

4.2.3 Flexibility

By flexibility we mean the ability for data based on the standard data model to be repurposed for a variety of different uses. One use might be the population of web pages both with human readable data and with embedded metadata to be consumed by “intelligent” agents on the web (such as HTML pages embedded with RDFa20 in a Linked Data21 context). Another use would be the transference of data between reporting systems such as between institutional and Research Council systems (systems typically implemented as relational databases, possibly supported by a Web Services architectural layer).

Discussion:

It would be possible for the CERIF abstract data model to be customised for a variety of different purposes, retaining a UK “super data model” of which all the others would be subsets. For example in the UK we could specify a CERIF4REF, a CERIF4ResearchOutputs and so on. We could also map the model to a Semantic Web compliant vocabulary, reusing suitable vocabularies wherever possible, and call this CERIF4LinkedDataPublishing, say. These subsets of the main model could be defined as profiles, together with relevant documentation describing the research information exchange scenarios they support and the related syntax and architectures intended for their use.

4.2.4 Ability to classify data

Given that some research data is potentially sensitive (such as collaboration information between educational institutions and partners) the ability to flag certain data with privacy flags or confidentiality classifications may be important. The Australian Research Council for example publishes business rules to accompany its submission instructions, one of which is BR00822 and specifies sensitivity codes as follows,

1 = Commercially sensitive
2 = Culturally sensitive
3 = Non public
4 = Commercially sensitive and culturally sensitive
5 = Commercially sensitive and non public
6 = Culturally sensitive and non public
7 = Commercially sensitive, culturally sensitive and non public

CERIF 2008 is not defined with sensitivity data but could be extended (see 1.2 Extensibility above) for example to express sensitivity via Context metadata (see 1.7 below).

4.2.5 Unique resource identifiers

URIs23 are used to uniquely identify resources described in metadata – for example each person can be assigned a URI, each organisational unit, each research output and so on. They are not the same as URLs which are network locations, although URLs can be used for URIs. URIs are especially important where data is to be aggregated from several sources, for example where an individual's forename and surname

---

20 <http://www.w3.org/TR/xhtml-rdfa-primer/>
21 <http://www.w3.org/DesignIssues/LinkedData.html>. For a good example of where Linked Data is being used see http://www.websci10.org/home.html, see also Appendix G
23 <http://www.w3.org/TR/webarch/#identification>
combinations are not enough to uniquely identify a researcher in an institution (i.e. several researchers may have the same name even within the same institution).

**Discussion:** URI schemes often have to be supported by centralised systems as there has to be multi-party agreement on who gets to generate unique identifiers when organisations have their own systems. The JISC-funded NAMES2 project runs until March 2011 and is developing work on an earlier authority prototype and investigating requirements for a name authority service for UK repositories; it may offer useful outcomes towards a shared service in the more general area of exchanging research information in the future.

We note that in the CERIF specification Core and Result entities (person, project, organisation, result publication and so on) have URIs but that link entities do not. In the Semantic Web languages such as RDF, even properties – which link resources – have URIs.

**4.2.6 Comprehensibility**

This is the question of how easy a proposed data model is to use and interpret, given that a standard is of no real use unless it becomes popular and is widely adopted! When uptake of a standard is successful we realise “the network effect”.

**Discussion:** For the EXRI-UK project we interviewed several parties who have attempted to map their data models to the CERIF data model. For example the Readiness4REF project reported that it took several weeks to refine a complete mapping of the RAE 2008 model to CERIF 2008. St Andrew’s and Aberdeen Universities took no longer than this to map their internal data models to the CERIF data model and were pleased about how intuitive they found it to create that mapping.

The first “chunk” of data model mapping often proves easily accessible (for example mapping obvious concepts of “Researcher” and “Organisational Unit”, or mapping “Internet publication” to “research output”). The more granular and specific parts of a data model are more taxing and require experts “on both sides” (for example someone who knows the Eurocris data model well and someone who knows their institutional data model very well) to ease the process. Without experts relating the two data models being mapped there is the danger of poorly generated data even though the standard is apparently being used. In the world of libraries, something similar can be seen when different cataloguers generate very different looking Marc records – often leaving several record fields blank or repeatedly supplying the same value to several different fields, purely because the technical user is not sure of the nuances differentiating parts of the underlying data model, or simply misunderstands it and makes some rather arbitrary decisions. On this key point we note that the initial work by the DoTAC project to map CERIF 2008 to existing research vocabularies (http://www.dotac.info/mapping/) illustrates the difficulties encountered in interpreting a data modeller’s intentions (see the number of question marks against terms to be mapped). It also brings up the notion of use-case driven modelling – if the need to include concepts in a data model is clear it should be well documented and preferably derived from genuine requirements.

Accessible data model mapping for the research information domain depicted by our scenarios in this document is very important because the domain is large and complex and many institutions have developed bespoke internal data models which, while often similar at the core, have specifics that are hard to cross-map. For example Oxford University has a collegiate system to model in their structure of organisational units. Some institutions may have no single coherent data model to start with.

The CERIF 2008 documentation was criticised as “poor” by several parties we spoke to and this would need considerable attention if CERIF were to be built on as a standard for the UK to use. Eurocris as a supporting agency was however described as very helpful.

---


25 See the examples section listed here [http://130.88.120.172:8080/help.html#quickAPIService](http://130.88.120.172:8080/help.html#quickAPIService) for the kinds of lookups such a service might be able to support.

4.2.7 Expression of Context and Controlled Vocabularies

This is the ability to support time or provenance data related to portions of metadata. Data that can be used in relation to parts of the data model when creating conformant data “records” may also need to be constrained with controlled vocabularies, again to support integration by well harmonised implementations across the UK. Without harmonisation it is more problematic to integrate data from multiple sources. Without contextual time and provenance data it is harder to guarantee the accuracy or relevancy of information shared.

Discussion: CERIF 2008 has a mechanism both for expressing time periods and for linking in controlled vocabularies. There are also well known ways to express time and controlled vocabularies using Semantic Web standards. For example the University of Bristol is using SKOS27 – a Semantic Web standard recommended by the W3C - for vocabularies (for subjects, organisational units etc.) linked in to its use of a research ontology. At the University of Oxford the BRII project is trialling one way to associate fragments of data that are true over particular time frames (using “manifests”). Researchers from the BRII, ResearchRevealed and DoTAC projects intend to align their approaches to expressing temporal information by using the OWL Time ontology28 in a consistent way in order to state where data is “true” over some time period as well as at some particular point in time.

We note that context information may state that information contained in a fragment of metadata is only valid in a certain context, or that it has particular attributes: for example it might be necessary to describe that the fragment is only true for a particular time period (the time period for which a researcher is affiliated to an institution say), or that a fragment describing the relationship between a person and a publication has attributes attached to that relationship (perhaps indicating the percentage of the person’s authorial contribution to that publication, for example). RDF has no specific notion of such contexts or relationship attributes built in, while CERIF provides a way to add metadata to “links” between resources, thus providing context – it calls this Semantics. With RDF, this is typically achieved by introducing anonymous resources (intermediary “nodes” in effect) in property assignments (i.e. to represent relationships between resources by an intermediary node to which context information can be attached). This is sufficient, if a little awkward.

4.2.8 Standardised, Open and Used

Ideally a standard should have an ISO or W3C recommendation, the latter perhaps being most useful given the web scenario requirements discussed here. Open standards are preferable to encourage freedom of reuse.

Discussion: Semantic Web languages such as RDF are W3C recommendations. CERIF is EC-recommended. The Semantic Web languages are fully open. RDF for example offers a standardised approach to expressing the data model for any knowledge domain (and no particular domain). In this document we listed (in the Data Model section 1) some existing RDF schemas suitable for the research information domain. All of these Schemas are open and freely reusable. CERIF is declared open as well and the only word of caution we might introduce here is around IPR – Intellectual Property Rights. It could be important to check there is no statement of IPR for CERIF that would prevent someone from taking the extensive data modelling work produced by Eurocris and reusing it, say, with a different syntax and format than the XML usage specified for CERIF 2008. During this project we have uncovered no restrictions that could create “tie-in” to the way knowledge encapsulated in CERIF 2008 might be reused in future, but a formal statement of rights from Eurocris would be highly desirable.

Evidence of usage of a data model is also an important consideration in evaluating whether its take up should be recommended. We return to this in section 2 on Syntax.

4.3 SYNTAX FOR EXCHANGING DATA

4.3.1 Usable for exchanging data over the Internet

Both CERIF in XML and RDF in XML (or its other syntactic variants such as N3) are usable for exchange over the internet.

---

27 Simple Knowledge Organisation System - <http://www.w3.org/2004/02/skos/>
28 <http://www.w3.org/TR/owl-time/>
Discussion: RDF was developed specifically for this purpose, in particular to offer a way to represent the meaning of encoded data – types of resources and relationships between them (for machine-interpretation) and not merely data structures. RDF may be exchanged using several known, standardised and well-used syntaxes. However, RDF/XML, for example, becomes more complex and challenging (to query\(^{29}\)) with the introduction of anonymous nodes in the data model, as would be needed to model complex relationships between researchers and publications, say, where many attributes need to be given for the relationship (expressing time context, type of academic role, percentage proportion of a contribution to a publication and so on). Nonetheless, the major benefits of using RDF and its related Semantic Web languages include its standardised use of namespace referencing and URIs that then permit data from disparate sources to be aggregated and for corresponding concepts and entities to be automatically linked.

The EXRI-UK project encountered some concerns over the suitability of CERIF’s exchange format – mainly the belief that while it is geared up for transferring data between relational databases it is nonetheless not used with an XML format that is “web friendly” or suitable for data integration (for example the dotAC project could not find clear documentation on how to map keys from one dataset to another). Eurocris recommend the creation of several XML files, distinguished with semantics expressed in the filenames, for the exchange of data sets. This introduces quite a significant level of complexity. One of the projects interviewed noted that the BBC’s work on a technical specification for TV Anytime\(^{30}\) resulted recently in a similar style of XML to CERIF’s as an exchange format, so it does not appear completely untypical for complex data modelling to be encoded using the XML syntax in this way.

4.3.2 Usable for serialising from/to relational databases
Both CERIF or a suitable Research Vocabulary in RDF are usable for exchanging data between relational databases. CERIF was developed more specifically for this purpose, its data model being based fairly closely on a relational database entity-relationship model.

4.3.3 Standardised syntax
The XML syntax (used by CERIF and optionally with Semantic Web schemas) is a W3C standard.

4.3.4 Clearly defined for the data model
CERIF XML is defined for use in relation to the data model. Similarly with Semantic Web research ontologies (such as the RES ontology for researchers\(^{31}\), FOAF, DCMI terms and so on).

4.3.5 Well Documented, Supported with clear Evidence of Use
Use of a data model and syntax should be well documented and supported via such things as sample data encoded according to the model and syntax, case studies, documentation describing best practice, workshops and exemplar implementations.

Discussion: This project encountered JISC-funded projects with concerns about the lack of documentation for the use of CERIF and the lack of enough real-world examples of sample data to clearly illustrate mark-up according to the standard. However, Eurocris are willing to run workshops to support uptake of CERIF and the documentation appears to be gradually improving.

Use of the Semantic Web schemas cited earlier in the Data Model section is not well documented although there is collaborative activity in this area to develop and document them. Underlying Semantic Web specifications and implementations of technology are readily found, however there is no formal Semantic Web group in existence to support the sharing of research information.

---

29 By query here we mean to ask questions of the data set using a query language such as SPARQL for the Semantic Web or SQL for relational databases.
30 [http://backstage.bbc.co.uk/feeds/tvradio/doc.html](http://backstage.bbc.co.uk/feeds/tvradio/doc.html)
31 [http://www.medsci.ox.ac.uk/vocab/researcher/0.1](http://www.medsci.ox.ac.uk/vocab/researcher/0.1)
4.4 POLICY

4.4.1 Version control
Required with any standard that will change and/or be extended over time. For example, The standard’s model could be represented using a standard itself - such as the Unified Modelling Language (UML) and a software package such as SubVersion or Concurrent Versions System (CVS) could be used for model version control.

4.4.2 Efficient mechanism for agreeing and disseminating extensions
Approval by committee is important for any changes to a standard so that they can be qualified, reviewed, agreed and disseminated appropriately.

4.4.3 Review process
For currency and continued fitness for purpose of the standard (given that requirements, syntax standards and web infrastructures change and evolve over time) a review process should be planned for the adoption of any standard.

4.4.4 Technical Appraisal - Conclusion
We conclude that CERIF 2008 offers very good coverage of the research domain and extensibility such that its data model could be purposed effectively for the UK research domain. For this, more work would have to be done in developing CERIF’s coverage in relation to the full range of scenarios discussed in this report for research information exchange (in particular to complete data model mappings for the UK Research Councils, HESA the REF and inter-institutional exchanges/promotion of UK research via the web).

However, documentation and evidence of uptake of CERIF 2008 in the UK is limited. While the two Scottish Universities of Aberdeen and St Andrews have successfully mapped their internal system data models to CERIF, they have not yet exchanged CERIF-encoded data between their institutions nor any other institutions and thus have not as yet demonstrated the practical benefit of using CERIF as an exchange format. Nonetheless, there is evidence of take up of various versions of CERIF in other parts of Europe—see Appendix F for more information.

Over 2010 and 2011 we expect very pertinent experience to be gained within the UK with respect to the question of where benefit could be realised from a common exchange format (and indeed if so which one) as well as the costs and issues in implementing one. We recommend reviewing progress when more about outcomes from at least the following activities is known:

- The Readiness4REF project will be demonstrating how sufficient CERIF is for REF reporting purposes with demonstrators at the University of Southampton and Kings College London (due to finish Spring 2011)
- The ResearchRevealed project will be demonstrating inter-institutional research information exchange via the Semantic Web approach, and whether it is easy to integrate this approach with statutory/funder reporting requirements (due to finish Spring 2011)
- Over the next year or so the Scottish universities will be sharing research information using CERIF and be able to report on the increased effectiveness of their research pools

In the meantime it would be possible build on the considerable body of CERIF work and to continue to develop a core view of CERIF 2008 for the UK. Rather than enforcing a “one size fits all” data model in the first instance it would be necessary to develop and publish mappings of a core CERIF for the UK to profiles required for different reporting purposes, with clear explanations of the requirements for use, examples of use, policy and architectural recommendations. It could be that further down the line it would be more efficient/appropriate for the UK to use a Semantic Web-based specification and syntax for the UK view of the CERIF data model than the current CERIF-specified XML usage. It could be that Eurocris itself reviews available technologies and scenarios in future and comes to that conclusion.

Our final note in this regard is that Eurocris is an established supporting (and, according to our interviewees, supportive) organisation – if it were not engaged to support the adoption by the UK of a standard to facilitate research exchange scenarios then it is likely that an alternative organisation would have to be created “from scratch”.

JISC Research Information Management Programme
5 RECOMMENDATIONS, ROADMAP AND RISK ASSESSMENT

We believe the potential benefits of a standard format for exchanging research information are great. It will allow cooperation between researchers, between institutions and research organisations and between research funders, within and beyond the UK. It will allow the rationalisation of data about what research is being funded and is taking place across the UK and the querying of that data by individuals, stakeholders and government. To an outsider who perhaps does not appreciate the complexities and organisation of research funding in the UK, it might appear extraordinary that this is not already possible. Finally it will allow HEIs and research organisations to plan in an ongoing way for the collection of data about research activity and to store that data in one format ready for collection by whichever government agency requires it, saving money, time and allowing easier planning and collection procedures.

In an Australian study in this area - Towards the Australian Data Commons, the authors recommend “the early development of an outline framework that allows progress to be made while the framework is further refined, and the refinement of the framework over time with broader contextual interests”.32 This is the broad approach we recommend in an area where there is a clearly expressed need for standards and harmonisation, yet also fast moving developments and many research projects underway whose results will be relevant and may affect future implementation decisions.

5.1 RECOMMENDATIONS

Recommendation 1

To: all stakeholders By: dates as shown in Roadmap

We recommend the use of CERIF 2008 as a basis for the exchange of research information in the UK, providing that this is read in the context of the other recommendations made below. We believe that this will bring benefits and efficiency gains to institutions, funders and researchers. It is possible that in the future a novel or UK-based standard will be preferred, but we are confident that this will be aided by work that can be done immediately to harmonise and standardise research information exchange by using CERIF 2008 as an exchange format.

Recommendation 2

To: JISC By: February 2010

JISC should seek and receive confirmation that CERIF is an open standard and Eurocris will not assert ownership or restrict use. JISC should also discuss with Eurocris the need to continue to be flexible enough to accommodate regular changes, such as those required by government, and the need for a good version control system and improved documentation. All indications we have received from CERIF users are that Eurocris will meet these requirements, while noting that it is a membership organisation and users play a large part in developing CERIF. We would reiterate that standardising research information exchange in the UK will require support and that if a custom-built standard were recommended then an organisation such as Eurocris would need to be invented.

Recommendation 3

The following harmonisation activities should be carried out:

3A

To: Research Councils, HEFCE, HESA By: June 2010

A harmonisation exercise should be conducted by the Research Councils, HEFCE and HESA to agree meaning of terms within a CERIF context or, where not possible, to agree defined variants. This should be done in liaison with Eurocris as the CERIF 2008 standard will likely need to be extended to accommodate UK requirements in full.

Mapping exercises to CERIF should be conducted by the Research Councils and HESA, to complement the work already done for HEFCE by the Readiness 4 REF project. Readiness 4 REFs experience is that the mapping was not a major undertaking (perhaps two person-weeks for the initial effort), so their experience would be valuable and EuroCRIS should be requested to participate in the process, perhaps through running workshops. Close consultation between all stakeholders mentioned here would be highly advantageous during this work to facilitate the next stage:

To: Research Councils, HEFCE, HESA and preferably other funders  
By: Q2, 2011

Complete implementation of harmonised terms and shared lexicon between research funders and agencies (for collecting new information required from research organisations).

**Recommendation 4**

To: JISC, ARMA, UCISA and HESA  
By: July 2010

JISC and ARMA, in collaboration with UCISA and HESA, should conduct a preliminary exercise to inform all stakeholders, HEIs, research organisations and other UK research funders, including charities and trusts etc, of likely developments with regard to CERIF, with a draft Roadmap.

**Recommendation 5**

To: HEFCE, Research Councils  
By: July 2010

While we believe the cost benefit case is overwhelming for activity to address the fact that different government agencies store, make available and request research activity information in different formats, it may be useful for JISC in their future communications with other stakeholders to consider commissioning a brief, formal economic cost benefit analysis to examine the business case for the changes we recommend, to examine experience of business cases in other countries (available from EuroCRIS) and to quantify the benefits for the UK sector as a whole.

To: Research Councils, HEFCE and HESA  
By: September 2010

Advance notice should then be given by September 2010 to all HEIs and significant research organisations that the Research Councils, HEFCE and HESA will request any new information to be supplied to them in CERIF compatible format by a deadline to be agreed before the announcement. Plans should be made to cater for individual researchers or very small research groups by allowing form submission which will create CERIF compatible information.

**Recommendation 6**

To: HEIs and research organisations  
By: deadline to be agreed

HEIs and other significant research organisations should plan and aim to complete the work necessary to be able to submit, in CERIF compatible format, all research activity data and statistics to collecting organisations.

**Recommendation 7**

To: JISC  
By: July 2010

Fund pilots to look at real exchange of research activity data between HEIs using CERIF.

**Recommendation 8**

The following projects should be reviewed:

8A

To: HEFCE, Research Councils  
By: Q4, 2010

Review results of REF Impact pilots (and to what extent metadata for Impact can be encapsulated by / is appropriate to build in to CERIF).
To: JISC, HEIs  
By: Q4, 2010 
Review results of JISC-funded Research Information Management pilots, including those funded under recommendation 7, plus results of R4REF (data exchange between Kings College, London and University of Southampton) and ResearchRevealed (data exchange between Oxford University and University of Bristol).

8C  
To: JISC initially, then other stakeholders  
By: Q1, 2011  
Review BRII, Research Revealed and DoTAC projects and their collaborative attempt to develop a Semantic Web Research schema mappable to CERIF.

8D  
To: JISC initially, then other stakeholders  
By: Q1, 2011  
Review of open gov data initiative, NAMES2 project, Open Researcher Contributor Identification Initiative (ORCID) and other relevant projects re: the likely need to be assigning common URIs to research entities and sharing UK research data in a wider Linked Data environment.

Recommendation 9  
To: all stakeholders, jointly  
By: Q2 2011  
Review progress and developments in the exchange of CERIF based research activity information in the UK. In particular, examine the possible linked data approach of a Semantic Web / RDF-in-XML syntax for data exchange, using CERIF as the agreed data model standard.

Recommendation 10  
To: all stakeholders, jointly  
Maintain a strategic commitment to the exchange of research activity information in a standard format between institutions, government, researchers, research users and other stakeholders; use the existing Research Information Management Group as a joint advisory and policy group which can task an individual or small steering group to monitor progress, act as an animateur for the national rollout of the standard and set appropriate timescales with the objective that:

- All HEIs and all significant research organisations make CERIF compliant data available (with appropriate access restrictions adjusted to data) for:
  - Collection by funders and other agencies
  - Sharing with collaborators and potential partners
  - Public consumption for web interfaces, by other agents for analysis etc
### 5.2 ROADMAP

#### 2010

<table>
<thead>
<tr>
<th>February '10</th>
<th>February '10</th>
<th>June '10</th>
<th>June '10</th>
<th>July '10</th>
<th>July '10</th>
<th>Q3 2010</th>
<th>Q4 2010</th>
<th>Q4 2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXRI report approved / qualified by Research Information Management Group (represents major stakeholders)</td>
<td>JISC on behalf of UK to receive confirmation that CERIF is open standard and Eurocris will not assert ownership or restrict use.</td>
<td>Mapping exercise for HESA to CERIF and Research Councils to CERIF.</td>
<td>Preliminary exercise with ARMA, UCISA &amp; HESA to inform all stakeholders, other UK research funders, including charities and trusts etc</td>
<td>Fund pilots to look at examples of exchange of research activity data between HEIs using CERIF</td>
<td>Advance notice that RCs, HESA and HEFCE (with others where appropriate) will request information to be supplied to them in a CERIF compatible format. HEIs to begin work.</td>
<td>Review of results of REF Impact pilots (and to what extent metadata for Impact can be encapsulated by is appropriate to build in to CERIF, Fund pilots to look at real examples of exchange of research activity data between HEIs using CERIF</td>
<td>Review of results of June-funded pilots plus results of R4R (exchange b/w Kings and Southampton) and Research Revealed (exchange b/w Oxford and Bristol).</td>
<td></td>
</tr>
</tbody>
</table>

#### Recommendation 2

- 3A
- 3B
- 4
- 7
- 5
- 8A
- 8B

#### 2011

<table>
<thead>
<tr>
<th>Q1 2011</th>
<th>Q1 2011</th>
<th>Q2 2011</th>
<th>Q2 2011</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Review BRII, Research Revealed and DoTAC projects and their collaborative attempt to develop a Semantic Web Research schema mappable to CERIF.</td>
<td>Review of open gov data initiative, ORCID initiative &amp; NAMES2 project re: likely need to be assigning common URIs to research entities and sharing UK research data in a wider Linked Data environment.</td>
<td>Review progress and developments in the exchange of CERIF based research activity information in the UK. In particular, examine the possible linked data approach of a Semantic Web / RDF-in-XML syntax for data exchange, using CERIF as the agreed data model standard.</td>
<td>Complete implementation of harmonised terms and shared lexicon between research funders and agencies. (For information required from research organisations).</td>
<td>HEIs and other significant research organisations to aim to complete work necessary to submit research information to funders in CERIF compatible format.</td>
</tr>
</tbody>
</table>

#### Recommendation 8C

- 8D
- 9
- 3C
- 6

---

33 Supporting Institutional submissions of HESA returns

34 Supporting Institutional submissions of data via what we anticipate will be a Research Outputs system that the majority/all of the Research Councils will adopt
Managing and pushing forward UK implementation

<table>
<thead>
<tr>
<th>All stakeholders</th>
<th>Aim by 2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>All research activity data held by funding agencies and host institutions in the UK is available in a consistent format and can be merged and queried, with appropriate authority restrictions.</td>
<td>All stakeholders maintain a strategic commitment to the exchange of research activity information in a standard format between institutions, government, researchers, research users and other stakeholders; use the existing Research Information Management Group as a joint advisory and policy group which can task an individual or small steering group to monitor progress, act as an animateur for the national rollout of the standard and set appropriate timescales with the objective that:</td>
</tr>
<tr>
<td>• All HEIs and all significant research organisations make CERIF compliant data available (with appropriate access restrictions adjusted to data) for:</td>
<td>Collection of research activity data by different agencies rationalised well in time for REF 2018.</td>
</tr>
<tr>
<td>o Collection by funders and other agencies</td>
<td></td>
</tr>
<tr>
<td>o Sharing with collaborators and potential partners</td>
<td></td>
</tr>
<tr>
<td>o Public consumption for web interfaces, by other agents for analysis etc</td>
<td></td>
</tr>
</tbody>
</table>

5.3 RISK ASSESSMENT

<table>
<thead>
<tr>
<th>Risk</th>
<th>Impact</th>
<th>Likelihood</th>
<th>Countermeasures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Future ownership and IP issues over CERIF</td>
<td>High</td>
<td>Low</td>
<td>Recommendation 2</td>
</tr>
<tr>
<td>EXRI report rejected by major stakeholders</td>
<td>High</td>
<td>Low</td>
<td>RIMG group to take report back to their organisations for ratification</td>
</tr>
<tr>
<td>Differences and disputes over meaning and use of elements and vocabularies</td>
<td>Medium</td>
<td>High</td>
<td>Recommendation 3</td>
</tr>
<tr>
<td>Other significant funders of UK research are (or feel) not included in process</td>
<td>Medium</td>
<td>Medium</td>
<td>Recommendations 3, 4, 5</td>
</tr>
<tr>
<td>HEIs unable or reluctant to modify existing workflows and practices</td>
<td>High</td>
<td>Medium</td>
<td>Recommendations 5, 6, 7, 8</td>
</tr>
<tr>
<td>Serious differences in opinion on future implementation paths</td>
<td>Medium</td>
<td>Medium</td>
<td>Recommendations 8, 9, 10</td>
</tr>
<tr>
<td>HEIs agree to minimal conformance necessary but do not take action for efficiency gains</td>
<td>High</td>
<td>Medium</td>
<td>Recommendations 7, 8</td>
</tr>
<tr>
<td>Changes do not impact on economy and wider environment</td>
<td>Medium</td>
<td>Low</td>
<td>8A, 9, 10</td>
</tr>
</tbody>
</table>
APPENDIX A: EXRI – UK PROJECT BACKGROUND

A1 EXRI-UK PROJECT

As part of this strand, the JISC-funded Exchanging Research Information UK (EXRI-UK) project involves an exploration of current and future scenarios for the exchange of research information in the UK. The study was intended to outline key requirements and benefits and to appraise options for the exchange of research information, with related risks and a ‘roadmap’ of actions. The project also aimed to explore specifically whether any particular format (e.g. CERIF) would be suitable.

The project was based at the Institute for Learning and Research Technology (ILRT) at the University of Bristol and ran from September to December 2009.

The project’s findings will inform JISC’s broader programme of work on research information management (RIM).35

A2 AIMS AND OBJECTIVES

The project had two main aims:

i) To identify and document scenarios, requirements and criteria for exchanging research information in the UK.

ii) To appraise the options for doing so and, specifically, whether any particular format for exchanging research information (e.g. CERIF) would be suitable.

The project had the following objectives relating to the first aim:

i) To work in a consultative way with all relevant stakeholders throughout the project's lifetime.

ii) To identify and describe scenarios and related requirements within the scope noted above, wherein exchanging research information would offer benefits.

iii) To note in each scenario, where possible, who benefits and some assessment of how much.

iv) To prioritise the scenarios by reference to their potential benefit, the feasibility of addressing them and the needs of the key stakeholders (bearing in mind their constraints in terms of timing, resources, etc).

The project had the following objectives related to the second aim:

i) To develop a set of criteria (beyond direct requirements) that should be met by any approach to exchanging research information in the UK.

ii) To identify a range of such approaches, including data models with an exchange format such as CERIF and practical examples of data exchange.

iii) To appraise each identified approach against the scenarios (iv) and other criteria (v), including technical and organisational impacts and risks.

iv) Without undertaking a systematic cost-benefit analysis, to come to a view, in consultation with stakeholders, on where and how the most benefit is likely to be delivered.

v) To work with key stakeholders to develop an initial UK Roadmap of actions for successful implementation of the candidate approach.

vi) To develop an outline risk assessment relating to the Roadmap.

vii) To make recommendations to named bodies, for example JISC, HEFCE, and Research Councils, on next steps.

EXRI-UK was a short project, running between September 2009 and December 2009 inclusive and was not designed to develop either a software system or an “EXRI” standard. As described above, its purpose was to consult widely and to review existing standards for the exchange of research data. We might think of choosing a standard for data exchange as analogous to selecting and agreeing on a highway code for our national road system. In this analogy we agree road markings, the semantics of signs and also speed limits etc. But we do not mandate any particular type of car to drive on this road system – different cars (just as different research software systems) are capable of using it. For maintenance and continued development of our highway code we require a governing body, such as the Department of Transport or the DVLA.

A3 APPROACH
The project team (Nikki Rogers, Dr Lesly Huxley and Nicky Ferguson) aimed to work in a consultative way with all relevant stakeholders throughout the project’s lifetime, via stakeholder interviews and workshops.

A number of stakeholder groups were identified at the start of the project. This informed establishment of a Reference Group for the project in consultation with the JISC Programme Manager.

We engaged with members of the Reference Group and other stakeholders by email, face-to-face or by telephone at key points in the project lifetime to inform the iterative development of project deliverables. Members of the Reference Group and other interviewees are listed in Appendix B. A list of stakeholder groups is provided at Appendix C.

In addition, we presented a progress report to the JISC-convened Research Information Management group which will also be reviewing the final report early in 2010. Major stakeholders are also represented on this group.

There are a number of other projects and activities in hand in the sector on which we drew to avoid duplication and ensure complementarity as appropriate, to inform and broker selection of stakeholders and to help frame and inform desktop research.

The questionnaire used in interviews and as a prompt for email responses where interviews were not possible is included at Appendix D.

A4 DELIVERABLES
Deliverable A: Report to identify, describe and prioritise scenarios for research information exchange (see Section 2 on this report).

Deliverable B: Report to describe Criteria for a standard for research data exchange. We used the outcomes of consultation and interviews to develop criteria that any approach will need to meet to support UK higher education’s common research information exchange scenarios (Section 3).

Deliverable C: Technical appraisal of potential technical approaches to research information exchange, including CERIF (Section 4).

Deliverable D: Outline roadmap of actions that will be needed for successful implementation of any adopted approach, including an outline risk assessment (Section 5).

Deliverable E: A summary set of recommendations (also included in Section 5) on next steps for JISC, HEFCE, Research Councils and other stakeholders as appropriate.

A glossary of terms is provided as Appendix E.

A5 LIMITATIONS AND BOUNDARIES
Although our brief did not include interviewing individual researchers we drew on available insights from the Readiness for REF (R4R) project, the UK Council of Research Repositories, relevant available studies such as the recent RIN report Communicating knowledge: how and why UK researchers publish and disseminate their findings36, the BRII Exploratory Study into the requirements and uses for research activity data at the University of Oxford37 and discussions in institutions. ILRT’s JISC-funded project ResearchRevealed38 conducted two technical workshops during the lifetime of EXRI-UK and these helped inform development of the technical section in this report.

37 http://brii.ouls.ox.ac.uk/documentation/project-milestones/BRIIStakeholderAnalysisReport.pdf
38 http://researchrevealed.ilrt.bris.ac.uk/
APPENDIX B: REFERENCE GROUP MEMBERSHIP AND OTHER INTERVIEWEES

Reference Group members were drawn from the stakeholder groups identified during project planning. All those listed below were interviewed face-to-face, by telephone, provided responses by email, or participated in discussions around the questionnaire themes in group meetings.

Reference Group Membership

- Les Carr, UK Council of Research Repositories, Readiness for REF project and School of Electronics and Computer Science, University of Southampton
- Dr Mary Davies, Deputy Director of ISS, Kings College London, SCONUL’s Performance Management Group and Readiness for REF project
- Scott Rutherford, Research Excellence Framework (REF) Programme Coordinator, Higher Education Funding Council for England (HEFCE)
- Simon Kerridge, ARMA and Head of Graduate Research Support and Assistant Director (Research), Academic Services, University of Sunderland
- Ian McCormick, Association of Research Managers and Administrators (ARMA) and Head of Research Services, University of East Anglia
- Dr Glenn Swafford, Director of Research Services, University of Oxford
- Dave Thomas, Head of Information Systems and Technology, Natural Environment Research Council (NERC)
- Alan Green, Business Analyst, Science and Technology Facilities Council (STFC) and Project Manager, Research Outcomes Project (formerly OOCS)
- Deborah Welland, Project Manager, Research Management and Administration System (RMAS) Group and Assistant Director, Business Improvement and Systems Support, University of Exeter
- Andy Youell, Director of Quality and Development, Higher Education Statistics Agency (HESA)

Other interviewees

- Alison Allden, Chief Executive, HESA
- Steve Bailey, Senior Adviser, JISC infoNet
- Stuart Bolton, Consultant
- Niamh Brennan, Programme Manager, Research Information Systems & Services, Trinity College Library Dublin.
- Emma Caseley, Strategy and Planning Division, Imperial College
- Victoria Cassely, Assistant Director, Information Services, Engineering and Physical Sciences Research Council (EPSRC)
- Anna Clements, University Data Architect, St Andrews University and EuroCRIS member, with Dr. Helen Reddy, Senior Research Policy Officer, St Andrews University.
- Lesley Dinsdale, Head of Research and Enterprise Policy, Research and Enterprise Development, University of Bristol
- Dale Heenan, Web Project Manager, Economic and Social Research Council (ESRC)
- Jill Jones, Personnel Systems Manager, University of Bristol
- Professor Doug Kell, Chief Executive, Biotechnology and Biological Sciences Research Council (BBRSC)
- Dr David Langley, Director, Research Enterprise and Development, University of Bristol
- Gerry Lawson, Research Information Systems (NERC)
- Josie Lewis-Gibbs, Strategy and Planning Division, Imperial College
- Pam McPherson Barrett, Higher Education Policy Adviser (REF development), HEFCE
- Graeme Rosenberg, Senior Higher Education Policy Adviser (REF development), HEFCE
- Wendy White, Manager of Institutional Repository, University of Southampton
APPENDIX C: STAKEHOLDER GROUPS

The following stakeholder groups were identified during initial project specification and planning in consultation with JISC. Reference Group and interviewee selection and contact with related projects were based on this initial analysis.

**Stakeholder group 1:**
Research managers and administrators in higher education institutions, research systems staff, Association of Research Managers and Administrators (ARMA) UK.
**Stake:** internal institutional strategic and operational decision-making; Research Excellence Framework (REF): research performance and impact; research information management; benchmarking; statutory data returns; public engagement with research.

**Stakeholder group 2:**
Higher Education Funding Council for England (HEFCE)
**Stake:** research funding allocation; research impact;

**Stakeholder group 3:**
JISC
**Stake:** supporting research information management in UK higher education; project sign-off and dissemination

**Stakeholder group 4:**
Higher Education Statistics Agency (HESA)
**Stake:** statutory data collection on staff, students and finance; potential future role in REF data collection

**Stakeholder group 5:**
Research Councils and Research Councils UK (RCUK)
**Stake:** Research funding information; research impact; Research management information systems; research performance; promotion of UK research; public understanding of research

**Stakeholder group 6:**
Researchers in higher education institutions **NOTE:** direct engagement with individual researchers is out of scope of the EXRI project; scenarios to be gleaned from engagement with related projects and perspectives of other stakeholder groups.
**Stake:** Research funding information; research impact; research performance; promotion of research; research collaborations; promotion and reward; public engagement with research; research publications

**Stakeholder group 7:**
Research Libraries, Repositories and other research information organisations
**Stake:** Research impact; promotion of research; research publications

**Stakeholder group 8:**
Publishers **NOTE:** publishers are on the periphery of the project’s scope; contact one or two if possible
**Stake:** research publications; research impact; promotion of research;

**Stakeholder group 9:**
EuroCRIS
**Stake:** UK higher education could be a key client

**Stakeholder group 10:**
Related projects
**Stake:** research publications; research impact; promotion of research; REF; CERIF
APPENDIX D: QUESTIONNAIRE

JISC EXRI-UK Project: Interview schedule

NOTE: the questions are designed to be posed in face-to-face or telephone interviews, and therefore include subsidiary prompts that may be used, depending on the response to each main question.

INTRODUCTION

1. Can you tell me something about your role and how this relates to the collection / provision / processing of research information?

CURRENT SCENARIOS and REQUIREMENTS

2. We want to find out about the different scenarios in which research information is currently collected, provided and exchanged. Can you tell me [more] about the research information you collect / provide / process?

Some prompts may be needed following this initial question:

- Why do you collect | provide | process this information?
- Is there a statutory requirement? from whom?
- Where does the information come from? (individuals, organisations - which ones)
- How often is the information collected | provided | processed? (does that vary for different information)?
- How long have you collected | provided this information?
- Has it changed over time? In what way?
- Is the information used for any other (e.g. internal) reasons? If so, what? (how, why)
- Do you exchange or share this information with others? (who, why?)
- Do you take in or use research information from other sources (other collectors | providers)? (who, why?)
- Can you give me a couple of typical use cases or scenarios which indicate the end user(s) of this information and how they might use it?

3. Can you tell me [more] about the detail of the data structure of the research information you collect / provide?

If no:
Go on to ask questions from 4 onwards

If yes, prompt for:

- Do you use any kind of controlled vocabulary or set of values? If so, what? Can we have a copy?
- Can you say what you call the major elements or classes of information (for example, if there is a concept of 'research output', or 'project', or 'publication', what labels do you give them?)
- What is the coverage (e.g. are details of institutions, subjects etc included?)
- Are there any dependencies or hierarchies that link different bits of information?
- How were these data structures developed? By whom? When? Why?
- Is this a standard that is used by anyone else? Who? For what purpose?

4. Can you tell me [more] about the processes and systems you use for collection / provision / exchange of research information?

If no:
Go on to ask questions from 5 onwards

If yes, prompt for:

- Who processes the information?
- What systems and processes are used? Do they differ for different purposes?
- Do you have to apply any interpretation or secondary coding to process the information? What? Why?
• Are you in the process of building or buying in any systems for processing research information?
• Do you extend or add value to any of the information? What? Why?
• Is there any 'authority' control over who can process the data?
• Is the processed data published or shared in any way? What? With whom? Why? (reports, statistics, raw data)

5. We want to get a feel for priorities: is any of the information you collect / provide more important for your purposes than others? Is there any information that you absolutely must have, but other data that you could manage without?
• What (in as much detail as possible - can they provide a ranking?)
• Why is this so important?
• Do you allow/accept any 'nil' returns? For what? Why?

6. Can you tell me if and how often your research information formats and processes change?
Do you ever change, add or delete data elements? How often? Why?
• Have you changed the way you store / process research information over the years? How? Why?
• What was the impact?
• Do you have plans in the next year to make any changes to your current data structures, processes or systems? What? Why? When?

FUTURE SCENARIOS, STANDARDS AND BENEFITS

7. Can you envisage any scenarios involving the exchange of research information that you don't currently operate, but could be of benefit?
[If they give an example, prompt for more. If they can't think of any [more] examples, suggest one or more of the following and get their opinion - and explore further in Q8]:
• Exchanging [more] between departments/faculties/central services in an institution
• Exchanging [more] research information between institutions in a geographical or 'mission' grouping
• Exchanging research information at national level (e.g. as Scottish HEIs currently do)
• Exchanging information between HEIs, research and statutory organisations in a different way?
• Any others (promotion of research, public understanding of science ...)?

8. We also want to explore benefits and barriers relating to research information exchange, and particularly of greater standardisation. What do you think are the benefits of the [current and future] scenarios you've described? What are the barriers to achieving them?
Prompts:
• What are the main benefits of collecting / providing / exchanging the research information we've been discussing?
• For whom? Why?
• Are there any difficulties in collecting / providing / exchanging? What are they? How could they be resolved?
• Are there resource issues that could be barriers to achieving these scenarios?
• Are there any data sensitivity issues (e.g., protected industry data) or concerns relating to sharing information in a standardised way?
• Any other issues?
• What would happen if this information weren't collected / provided / exchanged?
• Are there any other things you'd like to be able to do with this information? What? Why?

9. Are you aware of the CERIF data format? What do you think of it?
Have you evaluated it in relation to your research information processes?
• Is it too limited? In what way? What's missing?
• Is it too exhaustive? In what way? What wouldn't you use?

10. If the UK adopted a single standard (e.g. a particular XML schema, or CERIF, say) for describing, collecting and sharing research information, what impact would that have?
• What benefits might there be? To whom? Why?
• What would be the impact on your inputs and outputs?
• What would be the impact on your systems, processes?
• Would there be any impact on organisational structures, roles?
• Would it require any additional or different resource or skills?
• Overall, how disruptive would greater standardisation be?
• What timescales do you think would be needed?
• Can you foresee any advantages to adoption of a common standard? What? Why?
• And any disadvantages? What? Why?

CLOSURE
11. Is there anything else you would like to say about your perspectives on current or future scenarios for the exchange of research information?

12. Have we missed anything in our questions that you believe is essential to the development of scenarios or requirements? If so, what should we have asked, or asked differently?

13. Within the project timeframe, we expect to be able to interview a small number of additional stakeholder representatives (up to 20). We have already drawn up a list of potential interviewees in your organisation/stakeholder area [to be listed by interviewer at time of interview]. Which of these would you recommend we interview as a priority?

Thank you for your time
### APPENDIX E: GLOSSARY

<table>
<thead>
<tr>
<th>Acronym or term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARMA</td>
<td>Association of Research Managers and Administrators UK <a href="http://www.arma.ac.uk/">http://www.arma.ac.uk/</a></td>
</tr>
<tr>
<td>BBSRC</td>
<td>Biotechnology and Biological Sciences Research Council <a href="http://www.bbsrc.ac.uk/">http://www.bbsrc.ac.uk/</a></td>
</tr>
<tr>
<td>BIS</td>
<td>Department of Business Innovation and Skills <a href="http://www.bis.gov.uk/">http://www.bis.gov.uk/</a></td>
</tr>
<tr>
<td>BRII</td>
<td>Building the Research Information Infrastructure project, Oxford University <a href="http://brii.ouls.ox.ac.uk/">http://brii.ouls.ox.ac.uk/</a></td>
</tr>
<tr>
<td>DCMI</td>
<td>Dublin Core Metadata Initiative <a href="http://dublincore.org/">http://dublincore.org/</a></td>
</tr>
<tr>
<td>dotAC</td>
<td>Project at University of Southampton ‘exploring the UK Research Landscape’ <a href="http://www.jisc.ac.uk/whatwedo/programmes/inf11/jiscri/dotac">http://www.jisc.ac.uk/whatwedo/programmes/inf11/jiscri/dotac</a></td>
</tr>
<tr>
<td>EPSRC</td>
<td>Engineering and Physical Sciences Research Council <a href="http://www.epsrc.ac.uk/">http://www.epsrc.ac.uk/</a></td>
</tr>
<tr>
<td>ESRC</td>
<td>Economic and Social Research Council <a href="http://www.esrc.ac.uk/">http://www.esrc.ac.uk/</a></td>
</tr>
<tr>
<td>EuroCRIS</td>
<td>A not-for-profit association which aims to be the internationally recognised point of reference for all matters relating to Current Research Information Systems (CRIS) <a href="http://www.eurocris.org/public/home/">http://www.eurocris.org/public/home</a></td>
</tr>
<tr>
<td>EXRI-UK</td>
<td>Exchanging Research Information UK – the short name assigned to this study <a href="http://exri.ilrt.bris.ac.uk/">http://exri.ilrt.bris.ac.uk/</a></td>
</tr>
<tr>
<td>FOAF</td>
<td>Friend of a Friend project - creating a Web of machine-readable pages describing people, the links between them and the things they create and do. <a href="http://www.foaf-project.org/">http://www.foaf-project.org/</a></td>
</tr>
<tr>
<td>FTE</td>
<td>Full-time equivalent</td>
</tr>
<tr>
<td>HEFCE</td>
<td>Higher Education Funding Council for England <a href="http://www.hefce.ac.uk/">http://www.hefce.ac.uk/</a></td>
</tr>
<tr>
<td>HEI</td>
<td>Higher Education Institution</td>
</tr>
<tr>
<td>Heidi</td>
<td>Higher Education Information Database for Institutions <a href="https://heidi.hesa.ac.uk/">https://heidi.hesa.ac.uk/</a></td>
</tr>
<tr>
<td>HERO</td>
<td>Higher Education and Research Opportunities in the UK <a href="http://www.hero.ac.uk/">http://www.hero.ac.uk/</a></td>
</tr>
<tr>
<td>HESA</td>
<td>Higher Education Statistics Agency <a href="http://www.hesa.ac.uk">http://www.hesa.ac.uk</a></td>
</tr>
<tr>
<td>HR</td>
<td>Human Resources</td>
</tr>
<tr>
<td>ILRT</td>
<td>Institute for Learning and Research Technology – host to the EXRI-UK project <a href="http://www.ilrt.bristol.ac.uk/">http://www.ilrt.bristol.ac.uk/</a></td>
</tr>
<tr>
<td>IPR</td>
<td>Intellectual Property Rights</td>
</tr>
<tr>
<td>ISO</td>
<td>International Organization for Standardization <a href="http://www.iso.org/iso/home.htm">http://www.iso.org/iso/home.htm</a></td>
</tr>
<tr>
<td>JACS</td>
<td>Joint Academic Coding System – introduced by HESA and UCAS See JACS2 specification at: <a href="http://tinyurl.com/39tpcy">http://tinyurl.com/39tpcy</a> (HESA)</td>
</tr>
<tr>
<td>J-eS</td>
<td>Joint Electronic Submission system – for submission and monitoring of research grant applications to the Research Councils <a href="https://je-s.rcuk.ac.uk/JeS2WebLoginSite/Login.aspx">https://je-s.rcuk.ac.uk/JeS2WebLoginSite/Login.aspx</a></td>
</tr>
<tr>
<td>JISC</td>
<td>The funder of the EXRI-UK study <a href="http://www.jisc.ac.uk/">http://www.jisc.ac.uk/</a></td>
</tr>
<tr>
<td>Linked Data</td>
<td>See Tim Berners-Lee’s detailed description at: <a href="http://www.w3.org/DesignIssues/LinkedData.html">http://www.w3.org/DesignIssues/LinkedData.html</a></td>
</tr>
<tr>
<td>METIS</td>
<td>Dutch research information database system <a href="http://www.metis.hosting.kun.nl/metis/">http://www.metis.hosting.kun.nl/metis/</a></td>
</tr>
<tr>
<td>N3</td>
<td>Notation 3, an alternative to RDF/XML for the exchange of semantic Web data - <a href="http://www.w3.org/2000/10/swap/Primer.html">http://www.w3.org/2000/10/swap/Primer.html</a></td>
</tr>
<tr>
<td>NERC</td>
<td>Natural Environment Research Council <a href="http://www.nerc.ac.uk/">http://www.nerc.ac.uk/</a></td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>NHS</td>
<td>National Health Service</td>
</tr>
<tr>
<td>OOCS</td>
<td>Outputs and Outcomes System – a Research Councils initiative to develop a system for collection of research outputs and outcomes (now the Research Outcomes Project) <a href="http://www.rcuk.ac.uk/aboutrcuk/efficiency/Researchoutcomes/ropdocs.htm">http://www.rcuk.ac.uk/aboutrcuk/efficiency/Researchoutcomes/ropdocs.htm</a></td>
</tr>
<tr>
<td>ORCID</td>
<td>Open Researcher Contributor Identification Initiative <a href="http://orcid.securesites.net/">http://orcid.securesites.net/</a></td>
</tr>
<tr>
<td>PI</td>
<td>Principal Investigator</td>
</tr>
<tr>
<td>PURE</td>
<td>A modular, standard-software CRIS and full-text repository, which can be configured and customized to meet individual requirements <a href="http://atira.dk/en/pure/">http://atira.dk/en/pure/</a></td>
</tr>
<tr>
<td>QR</td>
<td>Quality-related funding – part of the HEFCE funding formula for HEIs <a href="http://www.hefce.ac.uk/research/funding/resfund/">http://www.hefce.ac.uk/research/funding/resfund/</a></td>
</tr>
<tr>
<td>ResearchRevealed</td>
<td>Project based at the University of Bristol, <a href="http://researchrevealed.ilrt.bris.ac.uk/">http://researchrevealed.ilrt.bris.ac.uk/</a></td>
</tr>
<tr>
<td>R4R</td>
<td>Readiness for REF project led by King’s College London <a href="http://www.kcl.ac.uk/iss/cerch/projects/portfolio/r4r.html">http://www.kcl.ac.uk/iss/cerch/projects/portfolio/r4r.html</a></td>
</tr>
<tr>
<td>RAES</td>
<td>Research Assessment Exercise – see information about RAES2008 at: <a href="http://www.rae.ac.uk/">http://www.rae.ac.uk/</a></td>
</tr>
<tr>
<td>RAS</td>
<td>Research Activity Survey (HEFCE) – see information about RAS2008 at: <a href="http://www.hefce.ac.uk/pubs/hefce/2008/08_29/">http://www.hefce.ac.uk/pubs/hefce/2008/08_29/</a></td>
</tr>
<tr>
<td>RC(s)</td>
<td>Research Council(s)</td>
</tr>
<tr>
<td>RDF</td>
<td>Resource Description Framework <a href="http://www.w3.org/RDF/">http://www.w3.org/RDF/</a></td>
</tr>
<tr>
<td>RDFa</td>
<td>RDF – in – attributes <a href="http://www.w3.org/TR/xhtml-rdfa-primer/">http://www.w3.org/TR/xhtml-rdfa-primer/</a></td>
</tr>
<tr>
<td>REF</td>
<td>Research Excellence Framework, successor to the RAE <a href="http://www.hefce.ac.uk/Research/ref/">http://www.hefce.ac.uk/Research/ref/</a></td>
</tr>
<tr>
<td>RIM</td>
<td>Research Information Management (JISC programme) <a href="http://www.jisc.ac.uk/whatwedothemes/informationenvironment/ResearchInfoMgt.aspx">http://www.jisc.ac.uk/whatwedothemes/informationenvironment/ResearchInfoMgt.aspx</a></td>
</tr>
<tr>
<td>RIMG</td>
<td>Research Information Management Group – A meeting of various stakeholders for the exchange of research information in the UK, facilitated by JISC</td>
</tr>
<tr>
<td>RIN</td>
<td>Research Information Network <a href="http://www.rin.ac.uk/">http://www.rin.ac.uk/</a></td>
</tr>
<tr>
<td>RMAS</td>
<td>Research Management and Administration System – to be specified by the RMAS Group <a href="http://as.exeter.ac.uk/rmas/">http://as.exeter.ac.uk/rmas/</a></td>
</tr>
<tr>
<td>SCONUL</td>
<td>Society of College, National and University Libraries <a href="http://www.sconul.ac.uk/">http://www.sconul.ac.uk/</a></td>
</tr>
<tr>
<td>Semantic Web</td>
<td>a common framework that allows data to be shared and reused across application, enterprise, and community boundaries <a href="http://www.w3.org/2001/sw/">http://www.w3.org/2001/sw/</a></td>
</tr>
<tr>
<td>SKOS</td>
<td>Simple Knowledge Organisation System <a href="http://www.w3.org/2004/02/skos/">http://www.w3.org/2004/02/skos/</a></td>
</tr>
<tr>
<td>SOAP</td>
<td>Simple Object Access Protocol</td>
</tr>
<tr>
<td>SPARQL</td>
<td>SPARQL Protocol and RDF Query Language</td>
</tr>
<tr>
<td>SRW/U</td>
<td>Search Retrieve via the Web / URL</td>
</tr>
<tr>
<td>STFC</td>
<td>Science and Technology Facilities Council <a href="http://www.stfc.ac.uk/">http://www.stfc.ac.uk/</a></td>
</tr>
<tr>
<td>Symplectic</td>
<td>Publications management and research management supplier <a href="http://www.symplectic.co.uk/">http://www.symplectic.co.uk/</a></td>
</tr>
<tr>
<td>TDA</td>
<td>Training and Development Agency for Schools <a href="http://www.tda.gov.uk/">www.tda.gov.uk</a></td>
</tr>
<tr>
<td>UKRO</td>
<td>UK Research Office <a href="http://www.ukro.ac.uk/">http://www.ukro.ac.uk/</a></td>
</tr>
<tr>
<td>URI</td>
<td>Unique Resource Identifier</td>
</tr>
<tr>
<td>URL</td>
<td>Uniform Resource Locator</td>
</tr>
<tr>
<td>W3C</td>
<td>World Wide Web Consortium <a href="http://www.w3.org/">http://www.w3.org/</a></td>
</tr>
<tr>
<td>XML</td>
<td>Extensible Markup Language <a href="http://www.w3.org/XML/">http://www.w3.org/XML/</a></td>
</tr>
<tr>
<td>Z39.50</td>
<td>Search/retrieval standard maintained by Library of Congress <a href="http://www.loc.gov/z3950/agency/">http://www.loc.gov/z3950/agency/</a></td>
</tr>
</tbody>
</table>
APPENDIX F: FURTHER REFERENCES

- SWRC (Semantic Web for Research Communities) [http://ontoware.org/projects/swrc](http://ontoware.org/projects/swrc), University of Melbourne
- CDISC - [http://www.cdisc.org/](http://www.cdisc.org/) - acquisition, exchange, submission and archive of clinical research data and metadata

Sites demonstrating the use of CERIF in the Exchange of Research Information

This is not intended to be a fully comprehensive list, but was usefully provided by a representative from Eurocris:

- BE: FRIS ([http://www.ewi-vlaanderen.be/en](http://www.ewi-vlaanderen.be/en)), National (Flanders) CRIS use CERIF as an interchange / collection format between Flemish universities
- DK: ([http://www.atira.dk/en/pure/](http://www.atira.dk/en/pure/)) All Danish universities using CERIF-CRISs implemented by atira using the PURE system which is fully CERIF driven
- NL: METIS ([http://metis.hosting.kun.nl/metis/](http://metis.hosting.kun.nl/metis/)) All NL universities using (extended) CERIF 1991 standard and are about to update to the current CERIF (2008) standard; linked to repositories (DARE) via the NARCIS project
- UK: Scottish universities (led by St. Andrews and Aberdeen) implementing CERIF-CRIS using PURE from Atira
- UK: CDR at STFC/RAL, CERIF-CRIS linked to institutional repository and used for directories driving workflow
- EU: IST World project ([http://www.ist-world.org/](http://www.ist-world.org/)) Using CERIF as an exchange / collection format in between more than 15 European countries, adding a view layer via advanced visualization techniques
- NO: Frida ([http://frida.usit.uio.no/](http://frida.usit.uio.no/)), National CERIF-CRIS linked to institutional repositories in individual universities
- IC: ICERIS ([http://www.ris.is/index_eng.html](http://www.ris.is/index_eng.html))
APPENDIX G: FURTHER REFERENCES

Example of good practice in using Linked Data in relation to typical website document – here the conference have made some of the data available as RDF (at the URL http://www.websci10.org/WebSci10.rdf) and have given the URI to be used to identify the conference:

The RDF given appears like this:

```
<rdf:RDF>
  <swc:ConferenceEvent rdf:about="http://www.websci10.org/WebSci10">
    <cal:summary>Web Science Conference 2010</cal:summary>
    <cal:dtstart>2010-04-26</cal:dtstart>
    <cal:dtend>2010-04-27</cal:dtend>
    <event:place rdf:resource="http://dbpedia.org/resource/Raleigh,_North_Carolina"/>
    <event:time rdf:resource="http://www.websci10.org/WebSci10#time"/>
    <dc:title>Web Science Conference 2010</dc:title>
    <rdfs:seeAlso rdf:resource="http://www.websci10.org/dates.rdf"/>
  </swc:ConferenceEvent>
  <tl:Interval rdf:about="http://www.websci10.org/WebSci10#time">
    <tl:beginsAt rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2010-04-26</tl:beginsAt>
    <tl:endsAt rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2010-04-27</tl:endsAt>
  </tl:Interval>
</rdf:RDF>
```