



University  
of Glasgow

Ginty, K., and McCutcheon, V. (2013) *Cerif for Datasets Final Report*. Project Report. University of Glasgow

Copyright © 2013 The Authors.

<http://eprints.gla.ac.uk/86350/>

Deposited on: 10 October 2013

Enlighten – Research publications by members of the University of Glasgow  
<http://eprints.gla.ac.uk>



## JISC Final Report

Project Information			
<b>Project Identifier</b>	<i>To be completed by JISC</i>		
<b>Project Title</b>	CERIF for Datasets (C4D)		
<b>Project Hashtag</b>	#cerif4data		
<b>Start Date</b>	1 <sup>st</sup> October 2011	<b>End Date</b>	30 <sup>th</sup> September 2013
<b>Lead Institution</b>	University of Sunderland (University of Glasgow from 1 <sup>st</sup> April 2013)		
<b>Project Director</b>	Kevin Ginty, Sunderland and Valerie McCutcheon, Glasgow		
<b>Project Manager</b>	Kevin Ginty, Sunderland and Valerie McCutcheon, Glasgow		
<b>Contact email</b>	<a href="mailto:gintyk@gmail.com">gintyk@gmail.com</a> & <a href="mailto:valerie.mccutcheon@glasgow.ac.uk">valerie.mccutcheon@glasgow.ac.uk</a>		
<b>Partner Institutions</b>	<ul style="list-style-type: none"> <li>• University of Sunderland</li> <li>• University of Glasgow</li> <li>• University of St Andrews</li> <li>• EPSRC</li> <li>• NERC</li> <li>• EuroCRIS</li> </ul>		
<b>Project Web URL</b>	<a href="http://cerif4datasets.wordpress.com/">http://cerif4datasets.wordpress.com/</a>		
<b>Programme Name</b>	<i>Managing Research Data Programme 2011-2013</i>		
<b>Programme Manager</b>	Simon Hodson		

Document Information			
<b>Author(s)</b>	Valerie McCutcheon		
<b>Project Role(s)</b>	Project Manager		
<b>Date</b>	31 <sup>st</sup> July 2013	<b>Filename</b>	C4D_Final_Report.doc
<b>URL</b>	<a href="http://cerif4datasets.wordpress.com/">http://cerif4datasets.wordpress.com/</a>		
<b>Access</b>	This report is for general dissemination		

Document History		
Version	Date	Comments
0.5	31/07/13	Draft updated from pre extension version
0.6	19/08/13	Draft incorporating AC and JD comments
0.7	28/08/13	Final draft after project team window to review and check versus benefits and evidence document.
1.0	30/09/13	Submitted

## Table of Contents

NB: This table of contents 'auto-populates' - to update the table of contents – place cursor in the table of contents, right-click your mouse, click 'update field', select appropriate option

<b>1</b>	<b>ACKNOWLEDGEMENTS .....</b>	<b>3</b>
<b>2</b>	<b>PROJECT SUMMARY .....</b>	<b>3</b>
<b>3</b>	<b>THE C4D PROJECT – OUTCOMES &amp; IMPACT .....</b>	<b>3</b>
3.1	PROJECT OUTPUTS AND OUTCOMES.....	3
3.2	HOW DID YOU GO ABOUT ACHIEVING YOUR OUTPUTS / OUTCOMES?.....	6
3.2.1	<i>Key Aims of C4D</i> .....	7
3.2.2	<i>Outcomes</i> .....	7
3.2.3	<i>Methodology</i> .....	7
3.3	WHAT DID YOU LEARN? .....	7
3.4	IMMEDIATE IMPACT .....	8
3.5	FUTURE IMPACT .....	9
<b>4</b>	<b>CONCLUSIONS.....</b>	<b>9</b>
<b>5</b>	<b>RECOMMENDATIONS.....</b>	<b>10</b>
<b>6</b>	<b>IMPLICATIONS FOR THE FUTURE .....</b>	<b>12</b>
<b>7</b>	<b>APPENDIX.....</b>	<b>12</b>

## 1 Acknowledgements

The C4D project (CERIF for Datasets) is funded under the Managing Research Data programme from JISC. This report summarises the project's activities and achievements during its duration. The partners in the project are the University of Sunderland (lead), the University of Glasgow, the University of St Andrews, EPSRC, NERC and EuroCRIS. This final report reflects input received from all partners.

## 2 Project Summary

C4D has developed a service to facilitate **discovery of research datasets**. Although research datasets are extremely heterogeneous, they have some commonality of research context which can be recorded in CERIF. Working closely with EuroCRIS and the Research Councils (RCs) we have been able to produce a Metadata-CERIF mapping which has advanced the CERIF **standard** to better handle research data.

In the JISC funded project, IRIOS2 (Integrated Research Input and Output System 2)<sup>1</sup> we extended an established platform (IRIOS) to better handle research outputs linked to awards. We demonstrated import of metadata about awards from RCUK systems and tested an efficient way to **share research information** between the organisations in CERIF format. Since all partners in the IRIOS2 project need to store metadata about research datasets, developing the IRIOS2 functionality to support datasets was seen as the next logical step in working together.

By 'CERIFYing' existing research dataset metadata conventions we were able to provide access to research data in an environment which also holds information on research projects and research outputs. By adapting IRIOS2 to manage research data, we took advantage of existing robust and secure platforms already in place and familiar to some users.

The C4D platform at Sunderland was used as a basis for defining metadata requirements and exploring data exchange between the three partners.

Whilst Sunderland plan to use the C4D system as their data registry. Glasgow has developed a live ePrints system, and St Andrews have produced a PURE prototype subject to third party planning release schedules. Each is based on a completely different configuration and this was part of the logic for the partnership in the first place as it allowed us to explore **standards that would work regardless of the underlying software used**. The strength of CERIF being that it allows commonality to be applied to different manifestations of dataset metadata.

Each Higher Education Institution (HEI) partner also developed their institutional research data management policy. This included an analysis of existing research data environments, and funder requirements.

## 3 The C4D project – Outcomes & Impact

The C4D project has been completed over a relatively short timescale of 24 months. This was possibly largely because of the existence of certain elements already in place allowing us to focus on customisation rather than a full development cycle. Project deliverables are published on the C4D website (<http://cerif4datasets.wordpress.com>)

### 3.1 Project Outputs and Outcomes

The deliverables are reported via the blog page:

<http://cerif4datasets.wordpress.com/c4d-deliverables/>

---

<sup>1</sup> <http://irios2.wordpress.com/>

<b>Output / Outcome</b>	<b>Brief Description and URLs (where applicable)</b>
Website	<a href="http://cerif4datasets.wordpress.com">Project website and blog (http://cerif4datasets.wordpress.com)</a>
Report	<b>Metadata Ontology</b> <a href="http://cerif4datasets.files.wordpress.com/2013/01/d2-1_metadata_ontology_v2-1_final.docx">http://cerif4datasets.files.wordpress.com/2013/01/d2-1_metadata_ontology_v2-1_final.docx</a>
Report	<b>Taxonomy Definition</b> <a href="http://cerif4datasets.files.wordpress.com/2013/01/d2-2_taxonomy_definition_v2-3.docx">http://cerif4datasets.files.wordpress.com/2013/01/d2-2_taxonomy_definition_v2-3.docx</a>
Report	<b>Metadata ontology upload requirements and process definition</b> <a href="http://cerif4datasets.files.wordpress.com/2013/01/d3-1_ontology_upload_requirements_v2-2a.docx">http://cerif4datasets.files.wordpress.com/2013/01/d3-1_ontology_upload_requirements_v2-2a.docx</a>
Report	<b>Metadata upload requirements and process definition</b> <a href="http://cerif4datasets.files.wordpress.com/2013/01/d3-2_metadata_upload_requirements_final.docx">http://cerif4datasets.files.wordpress.com/2013/01/d3-2_metadata_upload_requirements_final.docx</a>
Service	<p><b>Interim C4D infrastructure system across three platforms</b></p> <p>Sunderland <a href="http://c4dbeta.unis4ne-system.net/login.cfm">http://c4dbeta.unis4ne-system.net/login.cfm</a>  Accounts for the system are provided for registered users. For more information, please contact Kevin Ginty (<a href="mailto:kevin.ginty@sunderland.ac.uk">kevin.ginty@sunderland.ac.uk</a>).</p> <p>Glasgow <a href="http://researchdata.gla.ac.uk/">http://researchdata.gla.ac.uk/</a>  Glasgow produced a live ePrints registry (January 2013) and continue to work with other ePrints sites to compare and rationalise metadata definitions, roll out their system and consider linkage to the forthcoming national registry.</p> <p>Any user can enter a dataset which is then queued for administrative review. Contact <a href="mailto:Valerie.mccutcheon@glasgow.ac.uk">Valerie.mccutcheon@glasgow.ac.uk</a> for further details.</p> <p>St Andrews worked with the PURE UK User Group to extend the 'out of the box' datasets metadata profile to produce a beta version of a data registry for which is scheduled for release in version 4.17, Oct 2013.  Contact Anna Clements for further details <a href="mailto:akc@st-andrews.ac.uk">akc@st-andrews.ac.uk</a></p>

Service	<p><b>Final C4D infrastructure across three platforms</b></p> <p>Status of each system is reported at:  <a href="http://cerif4datasets.wordpress.com/system-status/">http://cerif4datasets.wordpress.com/system-status/</a></p> <p>See appendix A for screenshots of systems</p>
Report**	<p><b>Deployment plan across three platforms</b></p> <p><b>Incorporated in final report and blog</b></p>
Report**	<p><b>Sustainability analysis report across three platforms</b></p> <p><b>Incorporated in final report and blog</b></p>
Report	<p><b>CERIF-Metadata mapping</b>  <a href="http://cerif4datasets.files.wordpress.com/2013/03/c4d-cerif-metadata-implementation.pdf">http://cerif4datasets.files.wordpress.com/2013/03/c4d-cerif-metadata-implementation.pdf</a></p>
Report**	<p><b>Institutional research data management policy proposals</b></p> <p>Link to an overview document:  <a href="http://cerif4datasets.files.wordpress.com/2013/01/d5-3_institutional_research_data_management_policy_proposals-final.docx">http://cerif4datasets.files.wordpress.com/2013/01/d5-3_institutional_research_data_management_policy_proposals-final.docx</a></p> <p>Sunderland</p> <p>Have a draft policy awaiting internal approval.</p> <p>Glasgow</p> <p><a href="http://www.gla.ac.uk/services/datamanagement/rdm-at-gu/">http://www.gla.ac.uk/services/datamanagement/rdm-at-gu/</a></p> <p>These documents will be updated as processes and support is refined.</p> <p>St Andrews</p> <p>Have a draft policy awaiting internal approval.</p>
*Report	<p><b>Guidance notes for researchers</b>  <a href="http://cerif4datasets.files.wordpress.com/2013/01/d5-4_guidance_notes_for_researchers-final.docx">http://cerif4datasets.files.wordpress.com/2013/01/d5-4_guidance_notes_for_researchers-final.docx</a></p>

Report	<b>Framework Research Data Management Policy for Use by HEIs</b> <a href="http://cerif4datasets.files.wordpress.com/2013/01/d5-5_research_data_management_framework_final.docx">http://cerif4datasets.files.wordpress.com/2013/01/d5-5_research_data_management_framework_final.docx</a>
Report	<b>Project Final Report (This document)</b>
Report	<b>Dissemination Report</b> Included in Final Report
Report	<b>Workshop Report Workshop outputs available at:</b> <a href="http://cerif4datasets.wordpress.com/2013/08/02/cerif-4-datasets-workshops-outputs-july-2013/">http://cerif4datasets.wordpress.com/2013/08/02/cerif-4-datasets-workshops-outputs-july-2013/</a>

The project outputs were disseminated via several channels including:

- Poster displays at several conferences and events including ARMA and OIA8
- A Workshop on 'Contextualizing Publications and Datasets' and a paper at the Theory and Practice of Digital Libraries 2013 conference
- Two workshops that, in addition to sharing the findings of this project, resulted in a report and a number of potential actions for sustainability (see section 5 Recommendations). Workshop outputs available at: <http://cerif4datasets.wordpress.com/2013/08/02/cerif-4-datasets-workshops-outputs-july-2013/>
- On-going sharing via formal and informal meetings, our blog and professional networks

### ***3.2 How did you go about achieving your outputs/outcomes?***

C4D produced a framework for incorporating metadata into CERIF by building a demonstrator allowing for the import and management of metadata related to the selected research datasets.

A survey of current datasets and metadata was undertaken as part of the project, with a prediction that the area chosen would be one which all partner institutions were active. The outcomes of the study showed that marine science was the subject area with the common activity by all three university partners. The approach of C4D was to focus on marine sciences whilst also ensuring the developed service has a generic nature.

We researched the MEDIN metadata standard and mapped this to CERIF v1.5. 24 of the 30 elements mapped directly to CERIF. Recommendations for the remaining elements were put forward to the euroCRIS CERIF Task Group.

We proactively engaged with EuroCRIS via participation in CERIF Task Group meetings and discussions to ensure that the suggestions on metadata for CERIF for Datasets were considered for a new version of CERIF.

Each partner also pursued a roadmap to research data management support within their institution.

### 3.2.1 Key Aims of C4D

- Evaluate the feasibility of using CERIF as a standard protocol to represent research data metadata
- Demonstrate how this data metadata can be used alongside other CERIF entities for projects, people, institutions and outputs to increase accessibility to dataset

### 3.2.2 Outcomes

A proposed common CERIF metadata ontology for datasets was produced and delivered to EuroCRIS for amalgamation into the standard. A new version of CERIF incorporating key suggestions from this project was released for testing in July 2013. Details at <http://www.eurocris.org/index.php?page=CERIF-1.6&t=1>  
Delivery of an ontology and ontology driven interface provided the ability to import, export, discover, explore and link metadata.

The registries at each partner have progressed and will facilitate the process of publishing metadata to enable research datasets to be discoverable.

### 3.2.3 Methodology

A lot of time was consumed agreeing a list of fields to consider and formulating and refining the CERIF-metadata mapping to ensure its approval by EuroCRIS and the ultimate extension of the CERIF standard to incorporate research data metadata.

Each organisation worked on their own registry specification but bearing in mind the generic specification that was being developed for proposal to EuroCRIS.

Some testing was done to identify errors in code and any areas where systems were not sufficiently friendly. Amendments were made and then further cycles of testing took place. Test schedules were retained.

Regular meetings ensured that all parties knew what was expected of them and were able to input relevant expertise.

For the ePrints and PURE developments there was frequent discussion with the support services for these products.

## 3.3 What did you learn?

CERIF is able to record rich metadata about datasets and link to other entities within the CERIF model; CERIF also has a well governed maintenance and enhancement structure for improving the model.

There is similar work ongoing in the EU Engage project. This has developed a three-layer metadata model for research datasets. Paper presented by the C4D team at CRIS2012  
[http://www.eurocris.org/Uploads/Web%20pages/CRIS%202012%20-%20Prague/CRIS2012\\_2\\_full\\_paper.pdf](http://www.eurocris.org/Uploads/Web%20pages/CRIS%202012%20-%20Prague/CRIS2012_2_full_paper.pdf)

There is a lot of current activity in RDM and we need to work smarter at sharing ideas and reducing the cost if re-inventing wheels.

On the metadata side, the consensus for the minimum requirement appears to be the DataCite mandatory and recommended fields. <http://schema.datacite.org/meta/kernel-3/index.html>



If partners had not been involved in this project we would have been much less aware of this wide range of RDM activities and further back in our developments.

There is still a lot of uncertainty about the requirements for RDM and we should not worry too much about this as long as we are heading in the right direction. As institutions we aim to provide tools and support for RDM should staff require it but without pushing too many demands onto researchers who have many conflicting demands on their time already.

There is much more data stored in inappropriate or inaccessible places than we thought.

There is a considerable culture change needed in the way researchers view their data so that they understand and accept the desirability of making data open and reusable; just using the stick of funder policy is not a good solution.

### **3.4 Immediate Impact**

C4D has prompted EuroCRIS to advance the next version of CERIF in order to accommodate additional requirements for research data metadata. As of July 2013 a new version of CERIF – version 1.6 has been made available for testing. This incorporates many suggestions from the C4D CERIF mapping and work and notes that C4D was a major influence. We had hoped that the new version would have been available earlier to allow us to test the official EuroCRIS version within the lifetime of the C4D project however testing and iteration of the standard will continue beyond the project.

Link to CERIF version 1.6:

<http://cerif4datasets.wordpress.com/2013/07/25/new-version-of-cerif-available-for-testing/>

The development of an institutional Research Data Management (RDM) policy proposal has been beneficial at each of the partner HEIs. The development of an RDM policy has given researchers easier access to guidelines needed when seeking funding. In addition to an RDM policy we have produced a series of guidance notes which have been made available to researchers.

The partners have seen a number of impacts. We have been able to illustrate progress in setting up a registry and data repository to senior management and this has resulted in:

- Increased confidence in our abilities to deliver the RDM agenda and further momentum towards investment and support within the organisation.
- Increased awareness raising of data management requirements and support. For example, at Glasgow, the Vice-Principal for Research and Enterprise gave a very clear overview of open access and research data management requirements at the annual research staff conference in April 2013. This event is heavily attended by early career researchers.
- Increased ability to make a business case and secure resources for RDM
- Increased knowledge and awareness in support staff
- Ability to inform policy and guidance increased significantly from being involved in this project

The C4D project has raised our awareness of data management requirements and we continue to integrate data management requirements into the research support workflows. For example, at Glasgow, we now have a more definitive check of whether datasets were produced as part of the research in our open access workflow for publications.

In addition feedback shows that the wider community has truly benefitted from sharing ideas and advice via our blog and our workshops. These stakeholders are keen to maintain this sharing as it keeps them informed and saves them time.

### **3.5 Future Impact**

HEI's should be better able to meet RCUK ROS requirements in reporting instances of datasets and exchange information in a standard format.

Each of the partners now has a clear roadmap to supporting datasets via their registries. We will be able to collect metadata about datasets, ensure that storage is robust and offer improved research data access (more freely illustrating research excellence and making information available to wider society) by publishing the metadata.

By having clearer plans and standard systems we will become more efficient and able to apply economies of scale instead of ad hoc support.

We will continue to discuss API's and data synchronisation methods with other registries such as UKDA, ADS and NERC with whom we have already had discussions. Some marketing of data extraction methods for such services would be useful. Perhaps this will be linked in to the development of the national registry. Discussions are ongoing with the DCC on ensuring that the national register is CERIF compliant and the DCC Director, Kevin Ashley, will be participating in the euroCRIS strategic seminar in Brussels in September. It is important that stakeholders are involved in planning to avoid duplicate data transfers as far as possible and have clarity on the authority data sources.

The efficiency gains of the combined functionality of IRIOS2 and C4D are potentially very great. By using the system researchers are able to more fully manage research activities between multiple Research Organisations (ROs). Reception of the service has been positive to date.

The workshops have produced some excellent ideas for future improvements in communications and information sharing within stakeholders including:

- Encourage stakeholder groups (including Jisc) to invest time and money in saving tax payer money by clarifying specifications and supporting re-use of ideas across the sector
- Inform senior managers of sector wide issues and facilitate comparison with local issues

We also asked attendees 'What areas of Data Management do you feel need further communication to the research community?' The comments may be useful to Jisc in planning future support for RDM.

Details can be found on the document 'Attendees Feedback' at:

<http://cerif4datasets.wordpress.com/2013/08/02/cerif-4-datasets-workshops-outputs-july-2013/>

Further support and enhanced systems for research data management at each partner will be facilitated by the kick start that has arisen from C4D allowing us to clarify metadata requirements and design and deliver initial data registries.

The C4D service itself forms part of an integrated service used by the University of Sunderland and other HEI's and this could be made available to other organisations who wish to utilise this for research output management.

Similarly developments in ePrints and PURE will continue and will be shared by the user groups. C4D had been a driver for focussed user group activity as evidenced by ePrints and PURE meetings and forward plans.

## **4 Conclusions**

The general conclusion from the C4D project is that the CERIF standard can be used successfully as a standard protocol to represent research data metadata, and by continuing the evolution of CERIF we can ensure it becomes the standard for research activities.

The continued promotion and evolution of the CERIF standard would be greatly beneficial to the wider research community. Awareness of the standard has increased significantly over the previous two years with many researchers opting to use one or multiple CERIF based system and many vendors changing their systems to be more CERIF compliant. Continued evolution of the service could see CERIF become the standard for all research activities.

The RDM programme has been invaluable to the community in kick starting improved support for research data management.

There is a lot yet to be done and some thought should be given on how best to share ideas across this broad community to minimise cost to the tax payer. This is perhaps an area where Jisc could provide some support.

Options might include consideration of:

- Pooling resources across Institutions and Funders to develop network or sustainable data centres.
- Funding bodies and HEIs collaborating on developing a virtual, discipline-specific layer for the national research data registry being developed by Jisc and DCC that might add value to the data held in individual HEI data repositories (e.g., quality checking metadata, adding more metadata, linking between relevant datasets). This option was discussed during the Jisc MRD programme workshop in March 2013. NERC and ESRC already run data centres for their funded research.

## 5 Recommendations

These recommendations emerged from our work on the project and from the project workshops <http://cerif4datasets.wordpress.com/2013/08/02/cerif-4-datasets-workshops-outputs-july-2013/>

### General Recommendations

1. The Sunderland C4D service can be offered to UK ROs as a combined platform with IRIOS2 which is already known to many researchers.

Cloud computing provides a multitude of options for further development of C4D, deployment of the final system in a cloud computing environment has been planned, and costs outlined for developing and harvesting metadata.

2. A standard format for citing datasets. At present we are leaning towards DataCite's offering but there is no formal standard.

### Recommendations for the wider community

3. The community should join up, share their interpretations and 'specifications' and send some suggestions on our joint interpretation ('Manifesto') to the Department of Business Innovation and Skills and RCUK to ask for agreement or clarification.

There may be some difficulty of ensuring all relevant parties could get involved. Initial suggestions for key stakeholder groups included ARMA, Russell Group, system user groups (such as ePrints, PURE, Symplectic), Higher Education Information Directors Scotland (HEIDS), funders (e.g. HEFCE/SFC/Wellcome Trust) key user groups such as Life Sciences group, Senior Management responsible for research.

As an action we will raise this with ARMA. Perhaps Digital Curation Centre or Jisc could help with this. DCC are in the process of having their MRD module for research administrators evaluated by ARMA members and are keen to keep communications active with this group.

REF as used as an analogy – there is a clear specification and some consultation and working together took place.

4. Beyond metadata harmonisation, there is an argument for funders to work with institutions to provide cost-effective data centres. NERC and ESRC already provide data centres for their funded research. The risk at the moment is that much time and effort will be wasted by each Institution aiming to provide their own siloed infrastructure, workflows and resources to sustain long-term preservation of data. JISC could have a role to play here; as could the DCC.
5. Stakeholders should discuss synchronisation of data exchange so that effort is not duplicated and there is a clear understanding of which source is authoritative. DCC could perhaps incorporate this in their national registry work.
6. Continued liaison between institutions and EuroCRIS to ensure the evolution of CERIF.
7. ePrints and PURE communities should continue to collaborate on dataset specifications and support. The C4D project was one of the drivers in the set up of the ePrints UK user group which first met in June 2013 and plans to meet again regularly. Previously collaboration was often ad hoc. In addition we are scheduling some dataset specific workshops. The PURE user group has a data sub-group that meets up to discuss dataset specification and this group will be reviewing the beta system produced as part of the C4D project and taking this initiative forward to live delivery.

### Recommendations for Jisc

8. Continued promotion of the CERIF standard so that it becomes the standard for research information. It is clear that having a standard can facilitate efficient information exchange and crucially it can represent multiple models.
9. It is very encouraging to see all of the good work done on training and guidance however there is a lot for individual organisations to review. Jisc, DCC or another agency could synthesise all of the best practice on guidance and training into one summary of key points/templates. One good example given was getting groups to share data but not speak so that they could identify any gaps in metadata provision.  
<http://blog.okfn.org/2012/03/30/introducing-our-panton-fellows/>
10. Some investment of time in support for several findings from the workshops should be made. Perhaps this is a role for Jisc or one of the key funders.

These include:

- Provision of good examples of benefits gained that can be used to encourage researchers to engage with the RDM agenda (The Jisc MRD programme Evidence Gatherers should have some outcomes that will help with this in early September)
- Provision of a list of possible counter arguments to 'anti-champion' concerns.
- Provision of some clear benefits of good RDM that can be used to garner Senior Management support
- Provision of a summary of suitable value add storage options that might appeal to researchers and might be more robust than some of the storage in common use at present
- Provision of some useful guidance on how to overcome key RDM issues from an institutional perspective. This could cover some of the key issues raised at the workshops <http://cerif4datasets.wordpress.com/2013/08/02/cerif-4-datasets-workshops-outputs-july-2013/>

- DCC will be working with HEIs to develop workflow-specific training over the next 12 months through the 1-2-1 programme so this should help to embed good practice in institutions.
- Standardise the standards and work together to plan how to deploy. E.g. authority lists of organisations are not a trivial matter to deploy given historic status and system complexities.
- Provision of typical infrastructure models from well established data registries and repositories

## 6 Implications for the future

The C4D project has demonstrated that CERIF is viable as a standard protocol for representing research data metadata. In the broader context CERIF continues to grow into a viable standard for all research administration activities.

Looking ahead the work conducted during C4D on the CERIF-metadata mapping can be used as a reference model for more extensive coverage of metadata outside of the area of marine sciences. The current mapping provides a quite focused view with some areas of generic coverage.

As promotion of the combined C4D/IRIOS2 platform has been successful during the life of the project we will look to offer the service to other research organisation in the near future, doing so will enable the platform to be maintained and may lead to additional development. The positives of this combined service will be not only providing a service for managing research data metadata but also combining this metadata with information already held on projects, people, organisations and publications.

The work in the EU Engage project in building a three-layer metadata model with CERIF acting as the bridge between the simple DC-style discovery metadata and the detailed discipline-specific metadata for the individual datasets is of great interest. This project is benefitting from the enhanced CERIF model for datasets resulting from the C4D work and one aim for the future is to explore implementing a similar three-layer model at both the institutional and aggregator (e.g. national register) level.

## 7 Appendix

Glossary of Acronyms and Terms used in this report

CERIF – Common European Research Information Format  
RC – Research Council  
RO - Research Organisation  
HEI – Higher Education Institution  
RDM – Research Data Management  
NERC – National Environment Research Council  
EPSRC - Engineering and Physical Sciences Research Council  
IRIOS2 – Integrated Research Input and Output System 2  
EuroCRIS - the European Current Research Information Systems

## Glasgow – LIVE system ePrints

The screenshot shows a web form for dataset submission. The form is divided into several sections: Title, Description, Creators, Dataset Details, Restrictions, Divisions, and Subjects. Red annotations are placed to the right of the form, pointing to specific fields and sections.

**Title:** A text input field. Annotation: "Title of dataset being deposited/registered".

**Description:** A large text area. Annotation: "Description of what the dataset contains/is used for".

**Creators:** A table with columns: Family Name, Given Name / Initials, Email, GUID. There are four rows and a "More input rows" button. Annotations: "URL to the dataset" and "Software required to open dataset" (pointing to the Software Required field below).

**Dataset Details:** Includes fields for "URL of dataset (if not uploaded)", "Software Required", "Media of Output", "Related URLs", "Publisher", and "Publication Date". Annotations: "URL to the dataset", "Software required to open dataset", and "Restriction on access to the dataset" (pointing to the Restrictions section).

**Restrictions:** A text input field. Annotation: "Restriction on access to the dataset".

**Divisions:** A search box for "University of Glasgow Division" with a dropdown menu showing "all of" and a "Search" button. Below is a list of divisions: College of Arts, College of Medical/Veterinary and Life Sciences, College of Science and Engineering, College of Social Sciences, University Services. Annotation: "Glasgow University Organisational Units".

**Subjects:** A search box for "Search for subject" with a dropdown menu showing "all of" and "Search" and "Clear" buttons. Below is a list of subjects: Agrienvironmental science, Animal science, Archaeology, Anatomy observation, Astronomy theory, Atmospheric physics and chemistry, Atomic and molecular physics, Bioinformatics. Annotation: "RCUK Classification Scheme" and "Full three levels to improve".

St Andrews – BETA system PURE

Sunderland – BETA system UNIS (bespoke platform used by several HEI's in the North East of England)