# Standardising Research Contexts towards System Interoperability – and More

Brigitte Jörg a,b

aJeiBee Ltd., United Kingdom

beuroCRIS, The Netherlands

Summary

Scientific activities are increasingly data-intensive and thus more and more reliant on access to data and consequently the information related to managing the data namely metadata. Data and metadata are found in information systems to support specific functions in defined contexts. Despite the proliferation of formal metadata standards and a similarity in functions across scientific information systems their interoperability remains a challenge. We anticipate that needs or requirements are prerequisite to define harmonized contexts from where the formal boundaries emerge towards guidance with formal standards selection, and the formal contextual representation and aggregation of meaningful applicable elements. We present recent CERIF work in this respect and thus approach system interoperability from different contextual angles, before we conclude by introducing related activities.

## Introduction

Scientific activities are increasingly data-intensive and thus more and more reliant on access to data and consequently the information related to managing the data namely metadata. Data and metadata are found in information systems to support specific functions in defined contexts [Greenberg 2003], [Jeffery et al. 2013], [Jeffery et al. 2014]. Despite the proliferation of formal metadata standards and a similarity in functions across scientific information systems their interoperability remains a challenge [Tenopir 2011]. A variety of stakeholders are engaged with the production and consumption of research information and data throughout the research lifecycle. These require a reflection in standardized formal interfaces as well as in underlying formal information and data descriptions. We anticipate that needs or requirements are prerequisite to define harmonized contexts from where the formal boundaries emerge towards guidance with formal standards selection, and the formal contextual representation and aggregation of meaningful applicable elements. Funders, institutions, managers, researchers, curators, information technology, policy makers, the media, or the public have their own needs for accessing research information and underlying data[[1]](#footnote-2).

The entity in the center of interest within the Research domain emerged from a long tradition with a well-understood concept – the scholarly publication record. Standardization approaches are reflected in scientific repositories and through metadata formats such as Dublin Core[[2]](#footnote-3), MODS[[3]](#footnote-4), METS[[4]](#footnote-5). More recently, repositories are additionally employed for the collection of datasets. Increasingly, the research community recognizes a wider scope of stakeholders and requirements [Whyte & Allard 2014], and acknowledges need for wider stakeholder coverage through CRIS systems in the academic domain [van Godtsenhoven et al. 2008]. The development of CRIS systems is strongly tied to CERIF – a formal data model[[5]](#footnote-6). The Common European Research Information Format CERIF[[6]](#footnote-7) emerges as a standard format [Rogers et al. 2009], [German Science Council 2013]. CERIF is a recommendation to Member States by the European Commission (EC). The responsibility for continued development and maintenance has been handed over to euroCRIS[[7]](#footnote-8) ([www.eurocris.org](http://www.eurocris.org)) a non-profit organization registered in the Netherlands. The office is hosted at DANS, the Data Archiving and Networked Services ([www.dans.knaw.nl](http://www.dans.knaw.nl)).

Beyond the mere record type approach for information storage and exchange, CRISs are based on a domain model to supply a formal syntax for the description of domain entities and a declared semantics for the formal incorporation of multiple contextual vocabularies [Joerg et al. 2007], [Joerg et al. 2010][[8]](#footnote-9). CERIF thus provides formal constructs to represent contexts or profiles that reflect stakeholder needs. *Note:* CERIF does not initially supply the definitions or requirements i.e. contextual boundary specifications but allows to incorporate and formalize them. It is thus open with respect to contexts or directions for aggregations and application of vocabularies according to identified requirements.

The following section two briefly introduces CERIF. Section three explains how available profiles or contexts represent stakeholder views and how CERIF preserves meaning through a model-guided formal representation. Section four refers to ongoing related activities while section five concludes.

## Standardising Research Contexts through CERIF

We understand, a context is a harmonization of agreed requirements through engaged stakeholders; i.e. the definition of a maximum valid range of employable entities and their relationships for their contextual formal aggregation and thus implementation; the specification for standardized formal boundaries.

|  |  |
| --- | --- |
|  | *Figure 1: CERIF entities and their relationships without anticipated contexts.*CERIF allows for the implementation of contexts according to requirements. The requirements guide the selection of entities, their relationships and the applicable contextual vocabularies. The examples in sections three explain identified contexts and demonstrate their implementation through CERIF. |

## Identified Contexts

This section introduces identified requirements or research contexts, and demonstrates how these contexts or profiles represent stakeholder views and how CERIF enables for the preservation of their meaning through its model-guided approach and by supplying formal syntactic and semantic constructs. The UK Jisc-funded UKRISS project developed a core information reporting profile in CERIF [Joerg et al. 2014]. The CASRAI Abridged CV was mapped to CERIF [Joerg et al. 2014]. The Snowball Metrics project chose CERIF for the formal description of the Metrics [Clements et al. 2014]. The EC-funded OpenAIRE Open Access pilot employs a CERIF profile for CRIS interoperation [Houssos et al. 2014].

### UKRISS Research Reporting Profile

The Jisc-funded UK Research Information Shared Service (UKRISS) project developed a core reporting profile in CERIF enabling harmonized reporting on RCUK-funded research. The core profile development followed the requirement that “institution submits final report to funder”.



*Figure 2: UKRISS Core Information Reporting Profile.*

To maintain consistency with the formal reporting objects’ structure an upper reporting level has been introduced. It guided not only the object structure but in addition reflected the funder’s view according to which the reporting objects have finally been classified. A publication or event is thus defined as “Research Output”, while Spinout or Collaboration are considered “Research Outcome”. More detailed information about the formal description of each reporting object, its generic elements, its applicable aggregations and vocabularies based on CERIF are available in the final report and publicly accessible through the UKRISS blog (http://ukriss.cerch.kcl.ac.uk).

### CASRAI Activity Profile

CASRAI is a standards development organisation supported by an international community of research funders and institutions. CASRAI maintains and develops profiles for research administration information, such as: Research Activity; Research Personnel, Academic Funding CV, Non-academic Funding CV, Student CV, Abridged CV. A first draft of a formal CASRAI in CERIF description for the CASRAI Activity profile has been introduced at the CASRAI UK Summit[[9]](#footnote-10) convened by CASRAI and Jisc. Figure 3 reflects the profile elements through formal CERIF constructs and presents the CASRAI dictionary terms within their contextual range of the formal CERIF constructs.



*Figure 3: CASRAI Activity Profile in CERIF.*

A formal crosswalk or mapping of the CASRAI Abridged CV profile in CERIF will be available after the CRIS 2014 conference in Rome [Joerg et al. 2014].

### Snowball Metrics

The Snowball Metrics initiative aims at sharing the knowledge and experiences for a best practice in evidence-based institutional strategic planning. The ‘recipes’ or methodologies are available with the Snowball Metrics Recipe Book ([www.snowballmetrics.com/metrics](http://www.snowballmetrics.com/metrics)). CERIF has been applied as a formal description format for the Snowball Metrics [Clements et al. 2014].



*Figure 4: Generic Snowball Metrics structure in CERIF[[10]](#footnote-11).*

The general structure of a Snowball Measurement has been introduced at euroCRIS Member Meetings in Bonn and in Porto. Figure 4 is an extract of one of the slides presented in Porto where the CERIF task group approved the *cerification*. It shows the formal CERIF constructs behind the generic Snowball layout, which is then applied consistently with each individual Metric. CERIF not only allows for the formal description of each Metric, but also for the formal description of the entire set of Snowball Indicators and their structure.

### OpenAIRE Guidelines for CRIS Managers

OpenAIRE gathers research output related to European funding streams supporting open science and tracking research impact. OpenAIRE Guidelines[[11]](#footnote-12) have been developed for CRIS Managers to expose their metadata in a way that is compatible with the OpenAIRE infrastructure [Houssos et al. 2014]. Formal CERIF XML representations have been provided to ensure the consistent implementation and thus quality with data interchange of defined entities. The range of entities in the OpenAIRE context is indicated in figure 5. For more details we refer to the public website.



*Figure 5: Range of employed OpenAIRE entities in CERIF.*

## Related Work

The identification of stakeholder requirements has been recognized as a prerequisite for their harmonization and thus standardization. Stakeholder views have been investigated through life-cycle model approaches [Whyte & Allard 2014] especially related to Research Data Management activities[[12]](#footnote-13). We therefore want to mention the ongoing related work, activities and initiatives such as with the UK’s Digital Curation Centre (DCC) ([www.dcc.ac.uk/](http://www.dcc.ac.uk/)), and the Research Data Alliance (RDA) ([www.rd-alliance.org](http://www.rd-alliance.org)) in particular, with ongoing and continued collaboration between the RDA Metadata Working groups:

* Data in Context Interest Group
<https://www.rd-alliance.org/group/data-context-ig.html>
* Metadata Standards Directory Group:
<https://www.rd-alliance.org/group/metadata-standards-directory-working-group.html>
* Metadata Interest Group:
<https://www.rd-alliance.org/group/metadata-ig.html>

Relevant further ongoing activities are certainly VIVO ([www.vivoweb.org](http://www.vivoweb.org)), ORCID ([www.orcid.org](http://www.orcid.org)), COAR ([www.coar-repositories.org](http://www.coar-repositories.org)), and also ALLEA ([www.allea.org/](http://www.allea.org/)), EARMA ([www.earma.org/](http://www.earma.org/)), EUNIS ([www.eunis.org/](http://www.eunis.org/)) and APA (<http://www.alliancepermanentaccess.org/>), to mention just a few of the strategic partnerships maintained by euroCRIS.

## Conclusion

The investigation of the introduced approaches and profiles and active engagements in the above mentioned initiatives and activities revealed the need for contextual clarity through identification of requirements for guidance with harmonized formal representations of contexts or profiles. The profiles or specifications then become much more understandable and thus wider re-usable and therefore sharable and interchangeable.

### References

[German Science Council 2013] *Empfehlungen zu einem Kerndatensatz Forschung*: <http://www.wissenschaftsrat.de/download/archiv/2855-13.pdf> (Wissenschaftsrat 2013: see also project website Research Core Dataset: <http://www.research-information.de/Projekte/Research_Core_Dataset/projekte_research_core_dataset.asp>)

[Greenberg 2002] Greenberg J. *Metadata and the World Wide Web.* The Encyclopedia of Library and Information Science, Vol.72, 224-261, Marcel Dekker, New York, 2002.

[Joerg et al. 2014] Jörg B; Höllrigl T: Baker D. *Harmonising and Formalising Research Administration Profiles - CASRAI/CERIF.* In Proceedings CRIS 2014. To appear.

[Joerg et al. 2014] Jörg B; Waddington S; Jones R; Trowell S. *Harmonising Research Reporting – Experiences and Output from UKRISS.* InProceedings CRIS2014. To appear.

[Clements et al. 2014] Clements A.; Jörg B.; C. Lingjærde C.; Chudlarský T.; College L. *The application of the CERIF data format to Snowball Metrics.* In Proceedings CRIS 2014. To appear.

[Joerg 2010] Jörg B. *CERIF: The Common European Reseach Information Format Model.* Data Science Journal. Volume 9, Special Issue: CRISs for the European e-Infrastructure (Jul. 2010), CRIS24-31.

[Jeffery et al. 2014] K. Jeffery; N. Houssos; B. Jörg; A. Asserson. *Research Information Management: the CERIF approach*. International Journal of Metadata, Semantics and Ontologies, Inderscience Publishers, pp. 5-14, Computing and Mathematics, 2014.

Jeffery et al 2013. *A 3-Layer Model for Metadata*. Keith G Jeffery, Anne Asserson, Nikos Houssos, Brigitte Jörg. InProceedings: International Conference on Dublin Core and Metadata Applications 2013.

[Houssos et al. 2014] N. Houssos, B. Jörg, J. Dvořák, P. Príncipe, E. Rodrigues, P. Manghi, M.-K. Elbæk. *OpenAIRE Guidelines for CRIS Managers: Supporting Interoperability of Open Research Information through established standards.* CRIS2014 to appear.

[Joerg et al. 2007] B. Jörg; K. Jeffery; A. Asserson; G. van Grootel. *CERIF 2006 – 1.1 Full Data Model (FDM) – Introduction and Specification*. euroCRIS October 2007.

<http://www.dfki.de/~brigitte/CERIF/CERIF2006_1.1FDM/CERIF2006_FDM_1.1.pdf>

[Joerg et al. 2010] B. Jörg; K. Jeffery; A. Asserson; G. van Grootel; H. Rasmussen; J. Dvorak; D. Zendlukova; A. Price; T. Vestdam; M.K. Elbæk; N. Houssos; R. Voigt; E.J. Simons; *CERIF 2008 – 1.2 Semantics*. euroCRIS, November 2010

http://www.eurocris.org/Uploads/Web%20pages/CERIF2008/Release\_1.2/CERIF2008\_1.2\_Semantics.pdf

[Moreau & Missier 2013] Moreau, L. & Missier, P. (Eds.) (2013) *PROV-DM: The PROV Data Model. W3C Recommendation*. Retrieved Nov 11, 2013 from the World Wide Web: <http://www.w3.org/TR/2013/REC-prov-dm-20130430/>

[Mylopooulos 1992] J. Mylopoulos: *Conceptual Modeling and Telos.* P. Loucopoulos and R. Zicari (Eds.) Conceptual Modeling, Databases and Case. pp. 49-68, Wiley, 1992.

[Paskin 2003] N. Paskin. *Components of DRM Systems, Identification and Metadata.* In Becker, E., Buhse, W., Günnewig, D., & Rump, N. (Eds.), Digital Rights Management, Lecture Notes in Computer Science 2770. Retrieved Nov 11, 2013, from the World Wide Web: <http://dx.doi.org/10.1007/10941270_4>

[Rogers et al. 2013] N. Rogers, L. Huxley, N. Ferguson. *Exchanging Research Information in the UK*. *EXRI-UK: A study funded by JISC.* 2009. Web. 6. Sept. 2013. <http://repository.jisc.ac.uk/448>

[Tenopir et al. 2011] Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, et al. (2011) *Data Sharing by Scientists*: Practices and Perceptions. PLoS ONE 6(6): e21101. doi:10.1371/journal.pone.0021101

[Whyte & Allard 2014] Angus Whyte and Suzie Allard. *How to Discover Requirements for Research Data Management Services.* Digital Curation Centre (DCC), March 2014. <http://www.dcc.ac.uk/sites/default/files/documents/publications/DCC_Howto_Discover_Requirements.pdf>

1. The Economic and Social benefits of big data (EUROPA press releases database, Neelie Kroes, May 23/05/2013):
 http://europa.eu/rapid/press-release\_SPEECH-13-450\_en.htm [↑](#footnote-ref-2)
2. Dublin Core Metadata Initiative (DCMI): http://dublincore.org/ [↑](#footnote-ref-3)
3. Metadata Object Description Schema (MODS): http://www.loc.gov/standards/mods/ [↑](#footnote-ref-4)
4. Metadata Enconding and Transmission Standard (METS): http://www.loc.gov/standards/mets/ [↑](#footnote-ref-5)
5. “Conceptual Modeling is the activity of formally describing some aspects of the physical and social world around us for purpose of understanding and communication” [Mylopoulos 1992] [↑](#footnote-ref-6)
6. CERIF: An EU Recommendation to Member States (by CORDIS). http://cordis.europa.eu/cerif/home.html [↑](#footnote-ref-7)
7. euroCRIS Mission: „Advance Interoperability in the Research Community through CERIF”. http://www.eurocris.org [↑](#footnote-ref-8)
8. CERIF allows for the representation of any structure and mappings in between vocabularies and vocabulary terms through its declared semantics or its so-called Semantic Layer [Jörg et al. 2007]. [↑](#footnote-ref-9)
9. Two Peas in a Pod (CASRAI and CERIF): <http://cerifsupport.org/2012/12/17/cerif-casrai-profiling/> [↑](#footnote-ref-10)
10. Snowball Metrics presentation at euroCRIS Members Meeting in Porto (November 2013): [http://www.eurocris.org/Uploads/Web%20pages/members\_meetings/201211%20-%20Madrid,%20Spain/TG%20meetings%20-%20Anna%20Clements-Lisa%20Colledge%20(Snowball%20Metrics).ppt](http://www.eurocris.org/Uploads/Web%20pages/members_meetings/201211%20-%20Madrid%2C%20Spain/TG%20meetings%20-%20Anna%20Clements-Lisa%20Colledge%20%28Snowball%20Metrics%29.ppt) [↑](#footnote-ref-11)
11. OpenAIRE Guidelines for CRIS Managers: [https://guidelines.openaire.eu/wiki/OpenAIRE\_Guidelines:\_For\_CRIS](https://guidelines.openaire.eu/wiki/OpenAIRE_Guidelines%3A_For_CRIS) [↑](#footnote-ref-12)
12. A rich collection of requirements resulted rom a CKAN or RDM workshop: <https://docs.google.com/spreadsheet/ccc?key=0Arh4BnSV2XSIdGxibkhLXzhsUVZzVTl6MWdYOWVMcWc&usp=sharing#gid=0> [↑](#footnote-ref-13)