



CRIS 2014

## Harmonising and Formalising Research Administration Profiles CASRAI / CERIF

Brigitte Jörg<sup>a,b\*</sup>, Thorsten Höllrigl<sup>c</sup>, David Baker<sup>d</sup>

<sup>a</sup>euroCRIS, The Netherlands <sup>b</sup>JeiBee, United Kingdom

<sup>c</sup>Thomson Reuters, Germany

<sup>d</sup>CASRAI, Canada

---

### Abstract

CASRAI and CERIF are international standardisation initiatives in the domain of Research Information Management. CASRAI develops and maintains a standard extensible vocabulary and exchangeable data profiles that reflect the business requirements of involved stakeholders. A data profile specifies the maximal ideal space of its application with compliant data records. CERIF is a data model supplying standard formal syntax and declared semantics to preserve the meaning inherent in identified requirements. It enables the transformation of conceptual descriptions into formal representation thereof and thus their meaningful re-use as well as a semantically compliant and syntactically valid data interchange. With this paper we share the experience, and the lessons learned from the transformation of CASRAI profiles into CERIF XML through the example of an Abridged CV.

© 2014 Published by Elsevier B.V. Open access under [CC BY-NC-ND license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Peer-review under responsibility of euroCRIS

*Keywords:* Research Information; Research Administration; Harmonisation; Data Profile; CV; CASRAI; CERIF.

---

### 1. Introduction

Despite continuous advancements in information technologies interoperability between information systems is still in its infancy. Likewise in the research publishing domain, where the most common interchange format is OAI-PMH<sup>1</sup> combined with Dublin Core – a metadata element set to describe *resources* intended for usage with compatible vocabularies in the context of application profiles on the basis of DCAM, the DCMI Abstract Model<sup>2</sup>. OAI-PMH and Dublin Core are widely used for harvesting of scholarly repositories. Recently developed application

---

\* Corresponding author. Tel.: +44(0)2082798026.

E-mail address: [brigitte.joerg@gmail.com](mailto:brigitte.joerg@gmail.com)

profiles in this sense are for example RIOXX<sup>3</sup> for aggregation of metadata from UK Open Access repositories, or OpenAIRE<sup>4</sup> in support of defining and implementing local data management policies in compliance with the Open Access demands of the European Commission (EC). Historically, scholarly repositories have been dedicated to open access for scholarly publications and increasingly they are employed for storage of datasets. Repositories manage a tiny bit of the research ecosystem granting access to scholarly outputs by enabling the deposit of files. They have been little concerned with the precision and consistency in output related metadata<sup>5</sup>. Metadata quality however has been a major concern with CRIS systems since their early inception through the CERIF model in 1991 and 2000 in the format of an EC recommendation to Member States<sup>6</sup>. In 2002, the EC entrusted the responsibility over CERIF activities to euroCRIS, an international not-for-profit organisation dedicated to the development of Research Information Systems and their interoperability through CERIF. euroCRIS is supported by a large international membership of funders and research-involved organisations including global solution providers. Since the 2000 version, the CERIF releases have been developed continuously through contributions from a very active community. The latest updates and specifications are available from the euroCRIS website<sup>7</sup>.

Current Research Information Systems (CRIS) have been recognised as to being in the center of the scholarly information interoperability framework<sup>5</sup>. The need for a broader contextual metadata coverage beyond mere output deposit is reflected in the increasing uptake of CRIS systems and through availability of guidelines such as the OpenAIRE<sup>8,9</sup> Guidelines for CRIS Managers. The history of CRIS systems is strongly tied to CERIF – as a formal research domain model it supplies standard constructs for representing any context though not initially defining it. Any CERIF implementation follows from a requirement or scope analysis.

Since 2011 euroCRIS maintains a strategic partnership with CASRAI – the Consortia Advancing Standards in Research Administration Information<sup>10</sup> – an international non-profit standards initiative in the domain of Research Information Management. CASRAI develops and maintains a standard extensible vocabulary and exchangeable data profiles that reflect business requirements of involved stakeholders. CASRAI data standards take the form of common data profile specifications defined as harmonized standards that specify a subset of information required by users of an inter-organisational work process<sup>11</sup>. A CASRAI data profile thus specifies the maximal ideal space of its business application through compliant data records. A data profile is therefore different from an application profile, which is seen as a formal metadata schema consisting of data elements drawn from one or more namespaces optimized for a particular local application<sup>12</sup>.

CASRAI is a standards development organisation representing an international community of leading research funders and institutions. These collaborate to ensure seamless interoperability of research information. CASRAI develops a dictionary for relevant community profiles, such as: Research Activity Profile; Research Personnel Profile, Academic Funding CV, Non-academic Funding CV, Student CV, Abridged CV. Increasingly suppliers are involved in the initiative to incorporate the developed dictionaries into their products. CASRAI standards projects can be advanced either globally from the start, or nationally first before broader global adoption – depending on the objectives of the initiating stakeholders. As one example of ‘national-first’ standards projects, the Canadian CASRAI network activities currently focus on a development of standards for Research Classification, Funding Results Announcements, Administrative Key Performance Indicators, and Health Research Impacts.

As another example of ‘national-first’ projects, the United Kingdom CASRAI network (pilot project<sup>13</sup>) focuses on Data Management Plans, Authority Files for Organisations, Research Contributions, and Open Access Financial Reporting.

The research management *ecosystem* requires interoperability at various levels, within and beyond organisational, discipline and national boundaries, where a multitude of over-lapping use cases or requirements apply. We want to share our experience with the crosswalk from CASRAI business requirements into formal CERIF XML through the example of an Abridged CV.

The remainder of the paper is organised as follows. Section 2 introduces the CASRAI and CERIF constructs that are relevant in this context. Section 3 presents the business case of an Abridged CV and describes the implementation or transformation process from a CASRAI data profile into a CERIF XML representation thereof. Section 4 informs about the lessons learned and section 5 about the conclusions.

## 2. CASRAI and CERIF

CASRAI data profiles reflect business requirements of involved stakeholders. Examples of current profiles are available through the public CASRAI dictionary <http://dictionary.casrai.org/>. Before we explain the implementation or transformation process from a CASRAI profile into a formal CERIF XML description we want to introduce the major building blocks that constitute a CASRAI profile and a formal CERIF XML representation thereof.

### 2.1. CASRAI Building Blocks

CASRAI Profiles are open specifications for portable ‘business documents’ that can be produced and consumed by any two or more participants in a shared work process. They can be thought of as technology-neutral *objects* that contain all data needed for a specific exchange between parties. They represent the agreement shared by the parties on what is needed and they therefore remove the need for a costly and high-friction ‘point-to-point’ negotiation or mapping. The contents of any CASRAI Profile all come from a collaboratively developed and maintained vocabulary of terms (with labels and meanings) extensible for variations by discipline, domain or nation. These terms are structured through conceptual building blocks such as profile, grouping, record type, field or list as defined in table 1.

Table 1. CASRAI Profile Building Blocks.

Concept	Description	Examples
Profile	A pre-defined subset of data that satisfies the information requirements of a specific work process.	Abridged CV
Grouping	A collection of related records to aid human readability.	Identification, Education, Employment, Contact, Funding, Grants (Sub-Grouping)
Record Type	A single and complete set of related fields.	Person Info, Research Classification, Degrees, Supervisors, Professional Designations, E-mail Addresses, Mailing Addresses, Phone Numbers, Multi-year Details, Grant Participants,
Field	A single piece of information.	First Name, Gender, Keywords, Degree Name,
List	An authoritative collection of coded terms that constrain a field.	Salutation, Topic, Field of Application, Discipline,

CASRAI Profiles are intended to serve as a *canonical messaging model* (CMM) ensuring business agreement among the thousands of disparate information management tools in place across the global research enterprise. This business-layer CMM is intended to be *carried* in any commonly adopted technology-layer like CERIF XML.

## 2.2. CERIF XML Building blocks

CERIF XML is inspired by relational CERIF in that it preserves the entities and time-stamped relationship constructs and in that it complies with the formal syntax of the domain model entities including those of the semantic layer<sup>13</sup>. CERIF XML allows for a representation of various contexts or profiles – following from identified requirements – through the constructs explained in table 2 and inline with the CERIF structure<sup>14, 15</sup>.

Table 2. CERIF XML Building Blocks.

Construct	Description *	Examples
Entity	A thing which is recognized as being capable of an independent existence and which can be uniquely identified.	cfPerson, cfOrganisation, cfCurriculumVitae, cfPostalAddress, cfElectronicAddress, cfFunding, cfProject
Relation	A set of tuples for a defined domain, where each element is an attribute value.	cfPerson.cfPersonName_Person, cfPerson.cfPerson_OrganisationUnit, cfPerson.cfPerson_Project
Attribute	Entities may have various attributes to characterise them.	cfPerson.cfGender, cfPerson.cfPersonName_Person.cfFirstNames, cfPerson.cfPersonName_Person.cfFamilyNames,
Timestamp	The date or time at which the record is true in the modelled world, also known as valid time.	cfPerson.cfPerson_OrganisationUnit.cfStartDate cfPerson.cfPerson_OrganisationUnit.cfEndDate
Multilinguality	Applied with multilingual attributes.	cfCurriculumVitae.cfName, cfPerson.cfResearchInterest, cfClassificationTerm
Semantic Entity	An entity employed for declaring CERIF semantics.	cfClassification, cfClassification.cfClassification_cfClassification, cfClassificationScheme, cfClassificationScheme.cfClassificationScheme_cfClassificationScheme
Vocabulary Term	An attribute of the CERIF Classification entity.	cfClassification.cfTerm
Vocabulary Scheme	An entity employed with definitions of CERIF vocabularies.	cfClassificationScheme.cfName

- Inspired by Wikipedia

### 2.3. CASRAI – CERIF

The *crosswalk* between common CASRAI and CERIF building blocks in table 3 supports the understanding of each and shows how and where the two can complement each other.

Table 3. CASRAI – CERIF Crosswalk.

CASRAI Concept	CERIF Constructs	Notes with regard to implementation
Profile	--	CERIF misses a formal construct to define a data profile. It could e.g. be a set of rules for aggregating relevant entities based on defined vocabularies.
Grouping	Semantic Entity, Relations, Vocabulary Scheme, Vocabulary Term, Multilinguality	A grouping in CERIF is enabled by reference to the semantic layer, where a grouping (e.g. Identification) is considered a vocabulary term within a vocabulary scheme, (e.g. CASRAI Abridged CV Grouping)
Record Type	Entity, Relations, Attributes, Semantic Entity, Vocabulary Schemes, Vocabulary Terms, Time	A record type in CERIF follows from the selection of the corresponding basic entity (e.g. cfPerson), its relations (e.g. cfPersName_Pers) and attributes (e.g. cfPers.cfBirthdate) and the employed vocabulary schemes (e.g. Salutations) and terms (e.g. Ms, Dr). A record type label (e.g. Person Info) is enabled by reference to the semantic layer. Timestamps are employed within relations.
Field	Entity, Relations, Attributes, Vocabulary Terms, Vocabulary Schemes, Semantic Entity, Multilinguality	A field is either an attribute or a relation in CERIF XML. A field always belongs to an entity (e.g. corresponding language belongs to a person). Depending on its function (e.g. list or text), it is transformed either into an attribute or into a relation and a labeling vocabulary term (e.g. cfClass.cfTerm) assigned to a vocabulary scheme, thus employing semantic entities.
List	Semantic Entity, Relations, Vocabulary Scheme, Vocabulary Term, Multilinguality	A list in CERIF is handled similar as a grouping. Only, it maintains a different inherent structure as compared to the grouping.

### 3. CASRAI Abridged CV Implementation in CERIF XML

In table 3 we presented a crosswalk between CASRAI and CERIF building blocks. It started with the concept of a profile guiding the transformation process. With table 4 we demonstrate the implementation of a CASRAI Abridged CV in CERIF XML. The Abridged CV was chosen as a good test case. It is a smaller subset of the larger Academic Funding CV Profile for a researcher but it contains all the main structural elements (records, fields, lists, IDs).

The requirements behind the Abridged CV are the following: During the business process of applying for funding to multiple sources the Principal Investigator (PI) will need to include a full CV as part of the information submitted. But in the case where multiple collaborators are also a part of the application they may only be required to submit a smaller subset of data about themselves - hence the Abridged CV.

Table 4. CASRAI Abridged CV Profile in formal CERIF XML.

CASRAI Abridged CV Profile	CERIF XML
Profile	--
Identification	cfPers.cfPersId cfPers.cfFedId.cfFedId=unique-person-identifier cfPers.cfPers_Class.cfClassId=identification-vocabulary-term-uuid
<i>Person Info</i>	cfPers.cfPers_Class.cfClassId=person-info-vocabulary-term-uuid
Salutation	cfPers.cfPersName_Pers.cfClassId=salutation-vocabulary-term-uuid
First Name	cfPers.cfPersName_Pers.cfFirstNames
Middle Name	cfPers.cfPersName_Pers.cfFirstNames
Family Name	cfPers.cfPersName_Pers.cfFamilyNames
Presented Name	cfPers.cfPersName_Pers.cfClassId=presented-name-vocabulary-term-uuid
Previous Family Name	cfPers.cfPersName_Pers.cfClassId=previous-family-name-vocabulary-term-uuid
Correspondance Language	cfPers.cfPers_Class.cfClassId=correspondence-language-vocabulary-term-uuid
Gender	cfPers.cfGender
Date of Birth	cfPers.cfBirthdate
Designated Group	cfPers.cfPers_Class.cfClassId=designated-group-vocabulary-term-uuid
Research Classification	cfPers.cPers_Class.cfClassId=research-classification-vocabulary-term-uuid
Topic	cfPers.cfPers_Class.cfClassId=topic-vocabulary-term-uuid
Field of Application	cfPers.cfPers_Class.cfClassId=field-of-application-vocabulary-term-uuid
Discipline	cfPers.cfPers_Class.cfClassId=discipline-vocabulary-term-uuid
Keywords	cfPers.cfKeyw
<i>Education</i>	cfPers.cfPers_Class.cfClassId=education-vocabulary-term-uuid
Degrees	cfPers.cfPers_Qual.cfQualId=degree-reference-vocabulary-term-uuid
Supervisors	cfPers.cfPers_Pers.cfPersId=supervisor-identifier
Supervisor Role	cfPers.cfPers_Pers.cfClassId=supervisor-role-vocabulary-term-uuid
Supervisor First Name	cfPers.cfPersName_Pers.cfFirstNames
Supervisor Last Name	cfPers.cfPersName_Pers.cfFamilyNames
Degree Name	cfPers.cfPers_Qual.cfQualId=person-degree-identifier cfQual.cfName
Institution	cfPers.cfPers_OrgUnit.cfOrgUnitId=institution-identifier cfOrgUnit.cfName
Country	cfPers.cfPers_OrgUnit.cfOrgUnitId=institution-identifier cfOrgUnit.cfOrgUnit_Class.cfClassId=country-vocabulary-term-uuid
Degree Status	cfPers.cfPers_Qual.cfQualId=person-degree-identifier cfQual.cfQual_Class.cfClassId=degree-status-vocabulary-term-uuid
Start Date	cfPers.cfPers_OrgUnit.cfOrgUnitId=institution-identifier cfPers.cfPers_OrgUnit.cfClassId=begin-of-study-vocabulary-term-uuid cfPers.cfPers_OrgUnit.cfStartDate=date-value
End Date	cfPers.cfPers_OrgUnit.cfOrgUnitId=institution-identifier cfPers.cfPers_OrgUnit.cfClassId=end-of-study-vocabulary-term-uuid cfPers.cfPers_OrgUnit.cfEndDate=date-value
Expected Completion Date	cfPers.cfPers_Qual.cfQualId=person-degree-identifier cfQual.cfQual_Class.cfClassId=expected-completion-date-vocabulary-term-uuid cfQual.cfQual_Class.cfEndDate=date-value
...	...

The Abridged CV profile is available from the public CASRAI dictionary<sup>16</sup>. The employed CASRAI concepts such as groupings, record types, fields and lists have been introduced in table 1 including some examples. Table 4 provides the formal implementation of some selected CV elements in CERIF XML down to the level of attributes and values.

The implementation of CASRAI elements through CERIF in table 4 demonstrates the capacity of CERIF with respect to formal profile descriptions. Alignments of terms with vocabulary schemes are not considered – these require further investigation with respect to naming and structuring and as to how they relate to formal profile rules not currently supplied with CERIF.

In table 4 only some of the Abridged CV elements have been implemented. A more comprehensive view over the employed profile elements is given with figure 1.

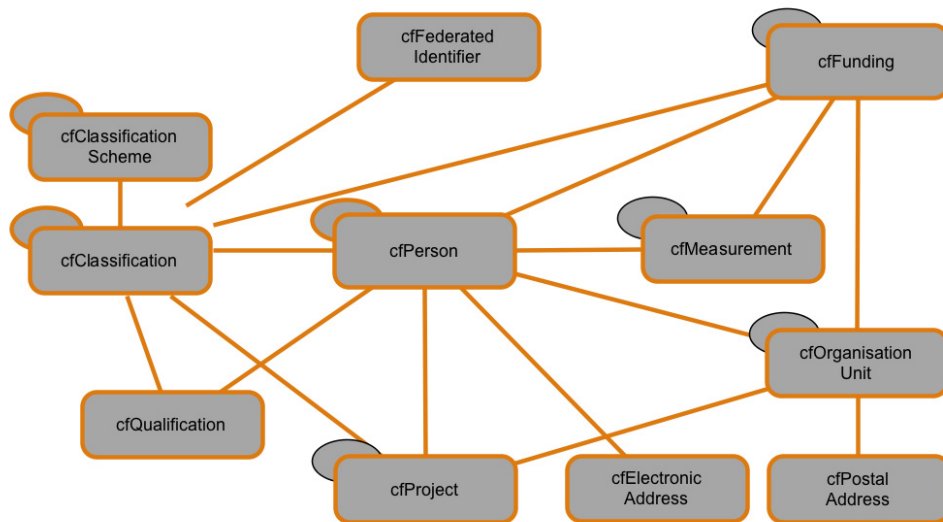


Fig. 1: CASRAI Abridged CV Profile in CERIF.

#### 4. Lessons Learned

A formal crosswalk between CASRAI concepts and CERIF constructs is straightforward within the Abridged CV profile starting from the requirements description. However, it needs further investigation of other requirements or profiles. Cardinality, field lengths, time-stamps or other integrity constraints have so far not been investigated and need further thought. A formal implementation of business rules for aggregation and for specification of profile boundaries is currently missing in CERIF.

## 5. Conclusion

The lessons learned will be forwarded to the CERIF TG encouraging further developments into this direction. A crosswalk from other CASRAI profiles into CERIF is on the agenda. The authors see a strong benefit from combining the complementary elements of the two approaches in the future.

## References

1. The Open Archives Initiative Protocol for Metadata Harvesting. <http://www.openarchives.org/pmh/>
2. Dublin Core Metadata Element Set, Version 1.1: <http://dublincore.org/documents/dces/>
3. RIOXX – Application Profile Version 1.0: <http://www.rioxn.net/v1-0/>
4. OpenAIRE Guidelines 2.0 for Repository Managers: [http://www.openaire.eu/about-openaire/publications-presentations/public-project-documents/doc\\_download/431-openaire-guidelinesv2-0en.pdf](http://www.openaire.eu/about-openaire/publications-presentations/public-project-documents/doc_download/431-openaire-guidelinesv2-0en.pdf)
5. van Godtsenhoven K, Elbaek M, Schmeltz Pedersen G, Sierman B, Bijsterbosch M, Hochstenbach P, Russel R, Vanderfeesten M. The European Repository Landscape 2008 – Survey on Technology. M. Vernooy-Gerritsen (Ed.). Amsterdam University Press, Amsterdam 2008.
6. CERIF: The Common European Research Information Format. An EU Recommendation to Member States. <http://cordis.europa.eu/cerif/home.html>
7. CERIF 1.6 Release by euroCRIS, August 2013. <http://www.eurocris.org/Index.php?page=CERIF-1.6&t=1>
8. OpenAIRE Guidelines for CRIS Managers: [https://guidelines.openaire.eu/wiki/OpenAIRE\\_Guidelines\\_For\\_CRIS](https://guidelines.openaire.eu/wiki/OpenAIRE_Guidelines_For_CRIS)
9. Houssos N, Jörg B, Dvorak J, Principe P, Rodrigues E, Manghi P, Karstensen Elbaek M. OpenAIRE Guidelines for CRIS Managers: Supporting Interoperability of Open Research Information through established standards. *Procedia Computer Science*. CRIS 2014. May 13-15, 2014
10. Strategic Partnership euroCRIS and CASRAI (November 2011): <https://cordis.europa.eu/wire/index.cfm?fuseaction=article.Detail&rcn=28533>
11. Baker D, Solving data disconnect in global research administration: the CASRAI approach. *Research Global*, November 2013
12. Ball A. Scientific Data Application Profile – Scoping Study Report. June 3<sup>rd</sup>, 2009.
13. Jisc – CASRAI-UK Pilot Project: [http://www.jisc.ac.uk/whatwedo/programmes/di\\_researchmanagement/researchinformation/casraipilot.aspx](http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/researchinformation/casraipilot.aspx)
14. Dvořák J, Jörg B. CERIF 1.5 XML – Data Exchange and Format Specification. euroCRIS 2013. [http://www.eurocris.org/Uploads/Web%20pages/CERIF-1.5/CERIF1.5\\_XML.pdf](http://www.eurocris.org/Uploads/Web%20pages/CERIF-1.5/CERIF1.5_XML.pdf)
15. Jörg B. CERIF: The Common European Research Information Format Model. *Data Science Journal*. Volume 9, Special Issue: CRISs for the European e-Infrastructure (Jul. 2010), CRIS24-31.
16. CASRAI Abridged CV: <http://dictionary.casrai.org/documents/abridged-cv/1.1>