Harmonising and Formalising Research Administration Profiles – CASRAI / CERIF

CRIS 2014, May 13, 2014

Brigitte Jörg – Director, Jei Bee Ltd., Board Member, euroCRIS  
Thorsten Höllrigl – Head of Product Development, CRIS, Thomson Reuters  
David Baker – Founder & Executive Director, CASRAI
Outline

- CASRAI Mission & Approach
- euroCRIS & CERIF XML
- What do we want to do?
  - Example Use Case: Abridged CV
  - Why do we need a mapping or crosswalk?
- CASRAI / CERIF XML Crosswalk
- Example Use Case Implementation
- Lessons Learned
- Conclusion
- Q&A
Outline

• CASRAI Mission & Approach
• euroCRIS & CERIF XML
• What do we want to do?
  – Example Use Case: Abridged CV
  – Why do we need a mapping or crosswalk?
• CASRAI / CERIF XML Crosswalk
• Example Use Case Implementation
• Lessons Learned
• Conclusion
• Q&A
CASRAI Mission & Approach

- Non-profit consortia of funders, institutions, libraries, vendors and other stakeholders
- Mission: improve the flow & management of admin data within and between stakeholders
- Approach: common dictionary of terminology & catalogue of exchangeable business objects
- Partners: euroCRIS, VIVO, ORCID, RDA, others
CASRAI Status

• Networks:
  – **Existing**: Canada & UK
  – **Pilot**: Sweden & Portugal
  – **Exploration**: Netherlands & US

• Projects
  – **CA**: Classification Taxonomy, Impacts; Funding Results; Admin KPI
  – **UK**: Data Management Planning; Org IDs
  – **International**: Peer Review Services; Contributor Roles

• European Presence
  – Annual ReConnect Eu events (new)
## CASRAI Profile Building Blocks

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

• CASRAI Mission & Approach
• euroCRIS & CERIF XML
• What do we want to do?
  – Example Use Case: Abridged CV
  – Why do we need a mapping?
• CASRAI / CERIF XML Crosswalk
• Example Use Case Implementation

• Q&A
euroCRIS & CERIF XML

- non-profit, entrusted by EC to develop and maintain the CERIF model towards interoperability of CRISs
- supported by large international membership of funders, research-involved organisations including global solution providers
- Partner in multiple EC and nationally funded projects
- Strategic Partnerships with Jisc, CASRAI, VIVO, ORCID, COAR, CODATA, ESF, ALLEA, APA, EUNIS + EARMA (in preparation)
- Events: Bi-annual Membership Meetings; Annual Strategic Seminar; every 2-years CRIS Conferences
CERIF XML

- Formal, scalable domain ontology (relational)
- CERIF XML for data exchange
- Generic
  - usable for heterogeneous environments
  - extensible without prejudicing the core data model
- Guaranteed interoperability at least at core level
- Open towards ingestion of any vocabulary
- Able to represent Data Profiles
- Needs building blocks for boundary definitions
# CERIF XML Building Blocks

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

- CASRAI Mission & Approach
- euroCRIS & CERIF XML
- What do we want to do?
  - Example Use Case: Abridged CV
  - Why do we need a mapping or crosswalk?
- CASRAI / CERIF XML Crosswalk
- Example Use Case Implementation
- Lessons Learned
- Conclusion
- Q&A
What do we want to do (requirements)?

Example Use Case: Abridged CV

„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.“
Outline

- CASRAI Mission & Approach
- euroCRIS & CERIF XML

**What do we want to do?**
- Example Use Case: Abridged CV
- Why do we need a mapping or crosswalk?

- CASRAI / CERIF XML Mapping
- Example Use Case Implementation
- Lessons Learned
- Conclusion
- Q&A
Why do we need a mapping or crosswalk?

• General goal:
  – Better understanding of the areas where the two approaches can complement each other
  – Combining the strong points of each approach

• Exemplary results:
  – Widely adopted CERIF XML can be used to exchange information
  – CASRAI when it comes to the vocabulary/dictionary; i.e. understanding
  – Benefit of concepts from CASRAI in CERIF, for example, when it comes to definition of compliance
  – Concept of profiles allows a very clear definition of compliance
Why do we need a mapping?
Supplier perspective

• Current research management solutions need to implement those standards in a complementary way

• Challenges
  – Semantics
    • What is a research project?
    • What is a review activity?
  – Syntax
    • Different date formats
    • Naming

• Avoid inconsistencies
Outline

• CASRAI Mission & Approach
• euroCRIS & CERIF XML
• What do we want to do?
  – Example Use Case: Abridged CV
  – Why do we need a mapping or crosswalk?
• CASRAI / CERIF Crosswalk
• Example Use Case Implementation
• Lessons Learned
• Conclusion
• Q&A
<table>
<thead>
<tr>
<th>CASRAI Concept</th>
<th>CERIF Constructs</th>
<th>Notes with regard to implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Profile</td>
<td>–</td>
<td>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.</td>
</tr>
<tr>
<td>Grouping</td>
<td>Semantic Entity, Relations, Vocabulary Scheme, Vocabulary Term, Multilinguality</td>
<td>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)</td>
</tr>
<tr>
<td>Record Type</td>
<td>Entity, Relations, Attributes, Semantic Entity, Vocabulary Schemes, Vocabulary Terms, Time</td>
<td>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.</td>
</tr>
<tr>
<td>Field</td>
<td>Entity, Relations, Attributes, Vocabulary Terms, Vocabulary Schemes, Semantic Entity, Multilinguality</td>
<td>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.</td>
</tr>
<tr>
<td>List</td>
<td>Semantic Entity, Relations, Vocabulary Scheme, Vocabulary Term, Multilinguality</td>
<td>A list in CERIF is handled similar as a grouping. Only, it maintains a different inherent structure as compared to the grouping.</td>
</tr>
</tbody>
</table>
Outline

• CASRAI Mission & Approach
• euroCRIS & CERIF XML
• What do we want to do?
  – Example Use Case: Abridged CV
  – Why do we need a mapping or crosswalk?
• CASRAI / CERIF XML Crosswalk
• Example Use Case Implementation
• Lessons Learned
• Conclusion
• Q&A
## Example Use Case Implementation

<table>
<thead>
<tr>
<th>CASRAI Abridged CV Profile</th>
<th>CERIF XML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Profile</td>
<td>--</td>
</tr>
<tr>
<td>Identification</td>
<td>cfPers.cfPersId</td>
</tr>
<tr>
<td></td>
<td>cfPers.cfFedId.cfFedId=unique-person-identifier</td>
</tr>
<tr>
<td></td>
<td>cfPers.cfPers_Class.cfClassId=identification-vocabulary-term-uuid</td>
</tr>
<tr>
<td>Person Info</td>
<td>cfPers.cfPers_Class.cfClassId=person-info-vocabulary-term-uuid</td>
</tr>
<tr>
<td>Salutation</td>
<td>cfPers.cfPersName_Pers.cfClassId=salutation-vocabulary-term-uuid</td>
</tr>
<tr>
<td>FirstName</td>
<td>cfPers.cfPersName_Pers.cfFirstNames</td>
</tr>
<tr>
<td>Middle Name</td>
<td>cfPers.cfPersName_Pers.cfFirstNames</td>
</tr>
<tr>
<td>Family Name</td>
<td>cfPers.cfPersName_Pers.cfFamilyNames</td>
</tr>
<tr>
<td>Presented Name</td>
<td>cfPers.cfPersName_Pers.cfClassId=presented-name-vocabulary-term-uuid</td>
</tr>
<tr>
<td>Previous Family Name</td>
<td>cfPers.cfPersName_Pers.cfClassId=previous-family-name-vocabulary-term-uuid</td>
</tr>
<tr>
<td>Correspondance Language</td>
<td>cfPers.cfPers_Class.cfClassId=correspondence-language-vocabulary-term-uuid</td>
</tr>
<tr>
<td>Gender</td>
<td>cfPers.cfGender</td>
</tr>
<tr>
<td>Date of Birth</td>
<td>cfPers.cfBirthdate</td>
</tr>
<tr>
<td>Designated Group</td>
<td>cfPers.cfPers_Class.cfClassId=designated-group-vocabulary-term-uuid</td>
</tr>
<tr>
<td>Research Classification</td>
<td>cfPers.cfPers_Class.cfClassId=research-classification-vocabulary-term-uuid</td>
</tr>
<tr>
<td>Topic</td>
<td>cfPers.cfPers_Class.cfClassId=topic-vocabulary-term-uuid</td>
</tr>
<tr>
<td>Field of Application</td>
<td>cfPers.cfPers_Class.cfClassId=field-of-application-vocabulary-term-uuid</td>
</tr>
<tr>
<td>Discipline</td>
<td>cfPers.cfPers_Class.cfClassId=discipline-vocabulary-term-uuid</td>
</tr>
<tr>
<td>Keywords</td>
<td>cfPers.cfKeyw</td>
</tr>
<tr>
<td>Education</td>
<td>cfPers.cfPers_Class.cfClassId=education-vocabulary-term-uuid</td>
</tr>
<tr>
<td>Degrees</td>
<td>cfPers.cfPers_Qual.cfQualId=degree-reference-vocabulary-term-uuid</td>
</tr>
<tr>
<td>Supervisors</td>
<td>cfPers.cfPers_Pers.cfPersId=supervisor-identifier</td>
</tr>
<tr>
<td>Supervisor Role</td>
<td>cfPers.cfPers_Pers.cfClassId=supervisor-role-vocabulary-term-uuid</td>
</tr>
<tr>
<td>Supervisor First Name</td>
<td>cfPers.cfPersName_Pers.cfFirstNames</td>
</tr>
</tbody>
</table>
Selection of CERIF Entities
Employed CERIF Building Blocks for CASRAI Abridged CV

- cfExpertiseAndSkills
- cfFunding
- cfPerson
- cfQualification
- cfMeasurement
- cfCurriculumVitae
- cfCountry
- cfCurrency
- cfLanguage

- Education
  - Degree
    - Degree Type
    - Degree Status
- Correspondence Language
- Educational Institution
- Department
- Supervisor
- Programme
- cfOrganisationUnit
- Employment
- cFProject
- Contact
- cfPostalAddress
- cfElectronicAddress
- Grant
- Participant
- Institution
- Institution Type
- Degrees
  - Degree Major
  - Degree Type
  - Degree Status
  - Degree

- Grant
- Participant
- Institution
- Institution Type
- Degrees
  - Degree Major
  - Degree Type
  - Degree Status
  - Degree

- Grant
- Participant
- Institution
- Institution Type
- Degrees
  - Degree Major
  - Degree Type
  - Degree Status
  - Degree

- Grant
- Participant
- Institution
- Institution Type
- Degrees
  - Degree Major
  - Degree Type
  - Degree Status
  - Degree

- Grant
- Participant
- Institution
- Institution Type
- Degrees
  - Degree Major
  - Degree Type
  - Degree Status
  - Degree
Employed CERIF Entities without Vocabularies (simple view)
Outline

• CASRAI Mission & Approach
• euroCRIS & CERIF XML
• What do we want to do?
  – Example Use Case: Abridged CV
  – Why do we need a mapping or crosswalk?
• CASRAI / CERIF XML Crosswalk
• Example Use Case Implementation
• Lessons Learned
• Conclusion
• Q&A
Lessons Learned

• Formal crosswalk between CASRAI concepts and CERIF constructs is straightforward within the Abridged CV profile starting from the requirements description.

• Needs further investigation of other requirements or profiles.

• Cardinality, field lengths, time-stamps, 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.
Outline

• CASRAI Mission & Approach
• euroCRIS & CERIF XML
• What do we want to do?
  – Example Use Case: Abridged CV
  – Why do we need a mapping or crosswalk?
• CASRAI / CERIF XML Crosswalk
• Example Use Case Implementation
• Lessons Learned
• Conclusion
• Q&A
Conclusion

• Lessons learned will be forwarded to the CERIF TG encouraging further developments into this direction.

• Crosswalk from other CASRAI profiles into CERIF is on the agenda.

• Authors see a strong benefit from combining the complementary elements of the two approaches in the future.
Outline

• CASRAI Mission & Approach
• euroCRIS & CERIF XML
• What do we want to do?
  – Example Use Case: Abridged CV
  – Why do we need a mapping or crosswalk?
• CASRAI / CERIF XML Crosswalk
• Example Use Case Implementation
• Lessons Learned
• Conclusion
• Q&A
Questions?

Contact

David Baker
Executive Director
Mobile: 613-291-7635
Phone: 613-832-0304
Mail: dbaker@casrai.org
Skype: dbakerskype

Dr. Brigitte Joerg
Director, Jei Bee Ltd.
Board Member, euroCRIS
Mail: contact@jeibee.com
Skype: brigitte.joerg

Dr. Thorsten Hoellrigl
Phone: +49 (0)721 354 90 24
Mobile: +49 (0)152 5310 47 01
Mail: thorsten.hoellrigl@thomsonreuters.com
Skype: t.hoellrigl