Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Common Specification on Research Information and Data & CERIF

Possibilities and limits

R.Arovelius, K. Pettersson

EuroCris, Espoo May 2019

Outline

- Introduction to common specifications (CS).
- CS and research records & data.
- Implementing CERIF into CS.
- Possibilities and limitations.

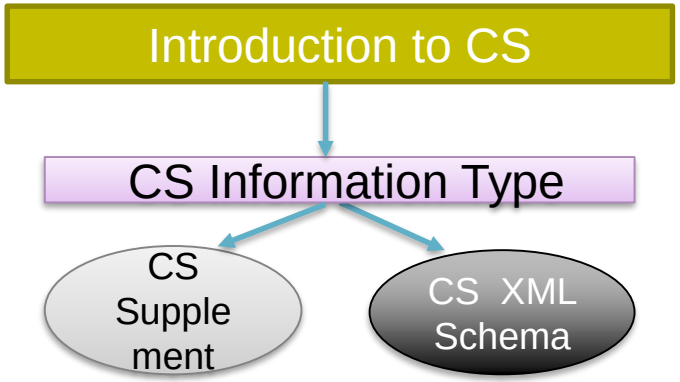
What is a CS?

- Format for exchange of information between different systems.
- Well defined and based on an existing international standard and its exchange format.
- Defines and describes metadata for information and packaging in informations packages (IP).
- Can be descriptive or/and normative.
- Administrated by the National Archives in Sweden.

https://riksarkivet.se/intro-fgs

What is the structure of a CS?

- Introduction to the management and use on the top level.
- Technical description and metadata fields on the middle level.
- Supplement as value lists, explanations and XML- schema.



- Specifications: http://riksarkivet.se/fgs-earkiv
- XML-schema: http://xml.ra.se/e-arkiv/

CS and information package (IP)

- CS is always content speciific and defines only one type of the content.
- CS Research connects to other specification, e.g. personel and archival description.
- May combine and embrace different standards at the package level.

CS Research records & data

- Standardised way for management of research information.
- Metadata
- PID
- Possibility for efficient and quality-assured transfer of information from different system during the entire life cycle.
- Semantic and organisational interoperability
- Reuse of research information and longevity

Content related common specification for all types of records from research process.

Technical requirements

- A CS should consist in a national adaption to international standards.
- XML should be used as data exchange format.
- XSD (or some other type of XML schema) have to be used in order to validate XML documents according to the CS.
- CERIF XML is an XML exchange format with a defined XSD.

Adapting CERIF XML to CS

- Metadata elements related to the different phases of the research process have to be mapped to the CERIF entities.
- After that, we can define an XML Schema which selects a subset of documents valid according to the main CERIF XML schema.

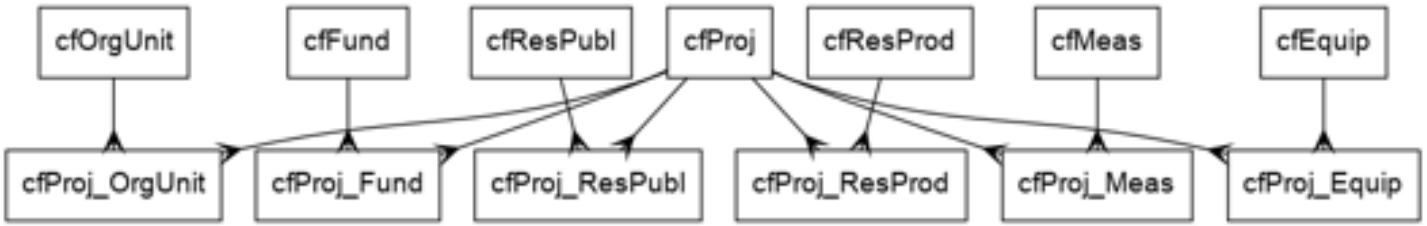


Figure 1: Possible subset of CERIF to be included in CS.

Relations between entities in CERIF

- Entities typically have few attributes except for key (ID).
- The information model is largely based on relating entities to other entities.
- These relations are defined by special link entitites, e.g.:
  - Between projects and persons (cfProj_Pers).
  - Between projects (cfProj_Proj).
  - Between persons and publications (cfPers_ResPubl).
  - If cfProj is used for processes, cfClass may be connected both to cfProj and cfResPubl/cfResProd, and used for record types.

Possibilities and limitations

- Standard approach for management and reuse of research information and data, longevity and trust.

- Interoperability of all kinds and possibility for dissemination of information from different sources.

Limitations:

- CS is just a recommendation.
- Need to extend data model beyond CERIF (preservation metadata) – related standards as e.g. PREMIS at the package level.
- CS must be implemented as only XML-schema - other formats as e.g. the JSON format might be easier to use.