GL Transparency:

Through a GLASS Clearly

Keith G Jeffery
Director, IT &
International Strategy, STFC

keith.jeffery@stfc.ac.uk

Anne G S Asserson
University Library
University of Bergen

anne.asserson@fa.uib.no

Authors

Keith G Jeffery
STFC-RAL

Anne Asserson
UiB

Structure

- The Challenge
- Background
- The Hypothesis
- Proposed Architecture
  - Metadata
  - Workflow
  - Architectural Solution
- Conclusion

The Challenge

- THE ACCUSATION
  - GL documents (objects) lack transparency
  - (and implicitly quality)
- THE SOLUTION
  - Formal metadata
  - Workflow
  - Recording provenance
  - Recording context

Transparency is defined in physics as the property of allowing light to pass through a material while more generally it implies openness, communication, and accountability. The latter meaning is used in this paper.

Contention: Metadata

Dublin Core (DC) and (MARC) – are insufficient for the purposes of:

- Discovery
- Management
- Utilisation
- Understanding
- Re-purposing
- Contextualising
- Provenance
- Preservation/curation
- Quality assessment

Explanation w.r.t. GOs (Grey Objects)

- Existence of relevant GOs
- Organising optimally GOs
- Using GOs for research
- The content of the GO
- A different use of the GO
- GO related to e.g. Project, group
- The processing steps of the GO
- GO available and understandable
- Compare the GO

Contention: Workflow

- unless GL material is collected in the context of a research workflow of services acting on the GL:
- the threshold barrier to collection is high and discourages those producing the GL from providing the metadata (or even the source material);
- associated contextual information is lost including any quality controls or peer review, or information allowing reputational judgement – thus transparency, so essential for confidence and trust in the information, is also lost;

Background: Previous Work

- the need for formal metadata to allow machine understanding and therefore scalable operations (Jeffery 1999)
- the enhancement of repositories of grey (and other) e-publications by linking with CRIS (Jeffery and Asserson 2004)
- the use of the research process to collect metadata incrementally reducing the threshold barrier for end-users and improving quality in an ambient GRIDs environment (Jeffery and Asserson 2005)
- an architectural model for scaleable, highly distributed, workflowed repositories of GLbased on hyperactive 'intelligent' documents (Jeffery and Asserson 2006)
- 'Greyscape' based on the hypothesis that grey literature is the foundation for the knowledge economy (Jeffery and Asserson 2007)
- An analysis of interoperation architectures among research information systems 'INTEREST' (Jeffery and Asserson 2008).
- A proposal that Grey Literature should be seen within the context of e-Science supported by a CERIF-CRIS (Jeffery and Asserson 2009)

Hypothesis

The ‘transparency problem’ can be solved easily by two technologies for which implemented examples are used widely:

- formal metadata associated with grey literature repositories improves relevance and quality;
- recording the workflow phases of a grey object within the context of a research information system records provenance;

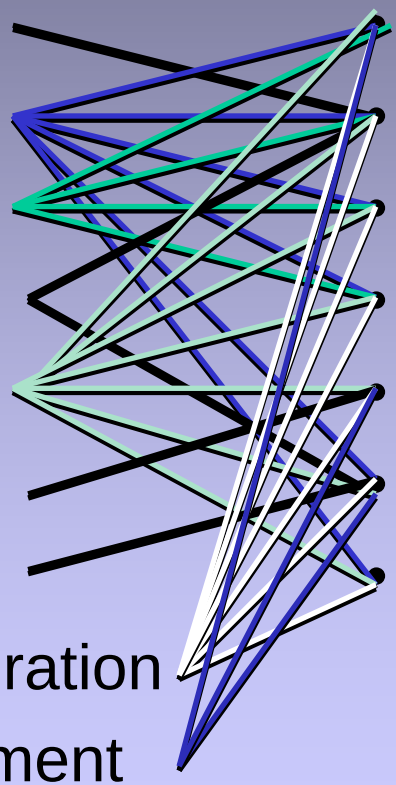
The solution – CERIF – exists already which covers these requirements.

➔ GLASS: Grey Literature Architecture for Sustainable Systems

Architectural Solution

SERVICES

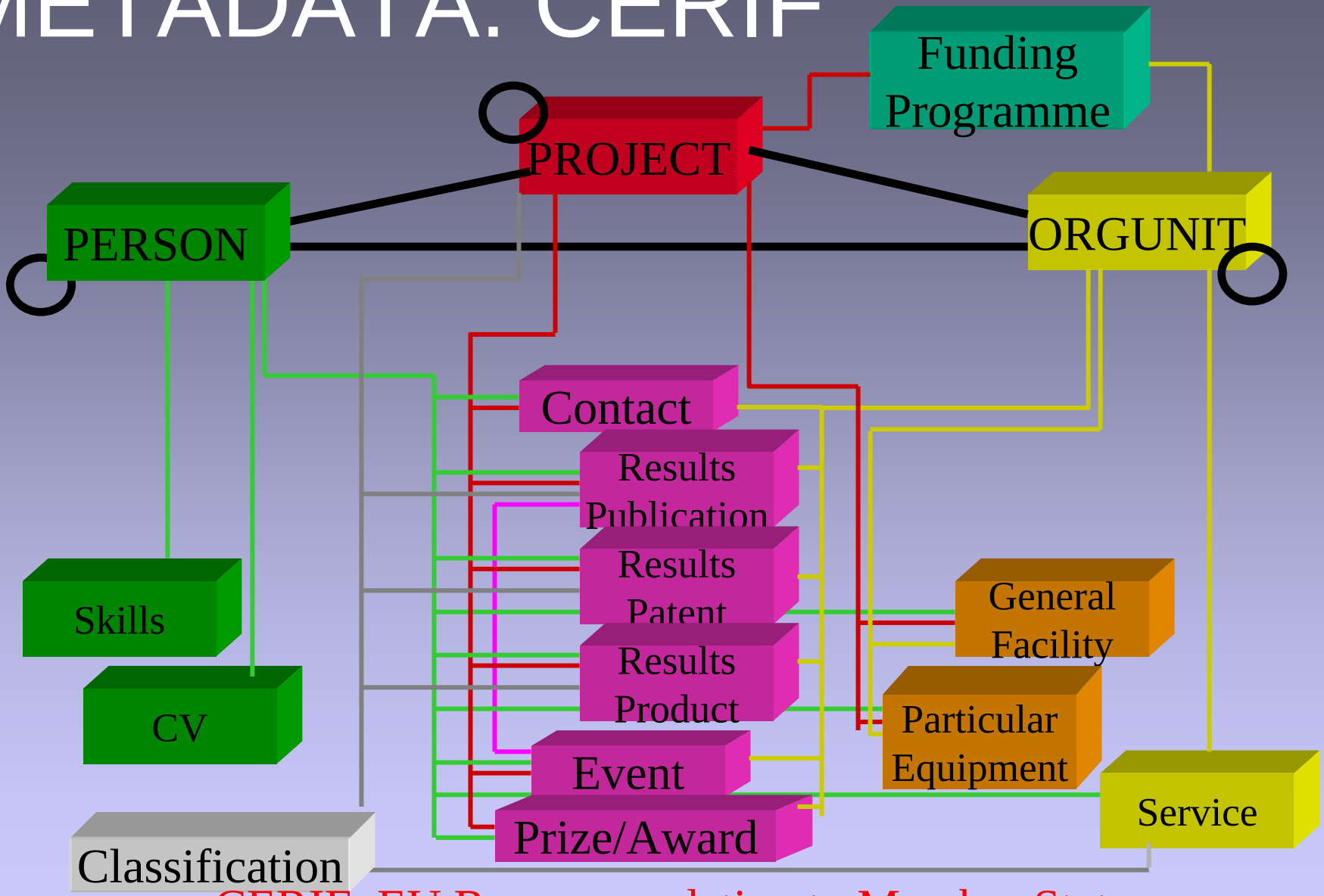
- Discovery
- Management
- Utilisation
- Understanding
- Re-purposing
- Contextualising
- Provenance
- Preservation/curation
- Quality Assessment



METADATA

- Schema
- Descriptive
- Restrictive
- Navigational
- Contextual
- Provenance
- Curation/Preservation

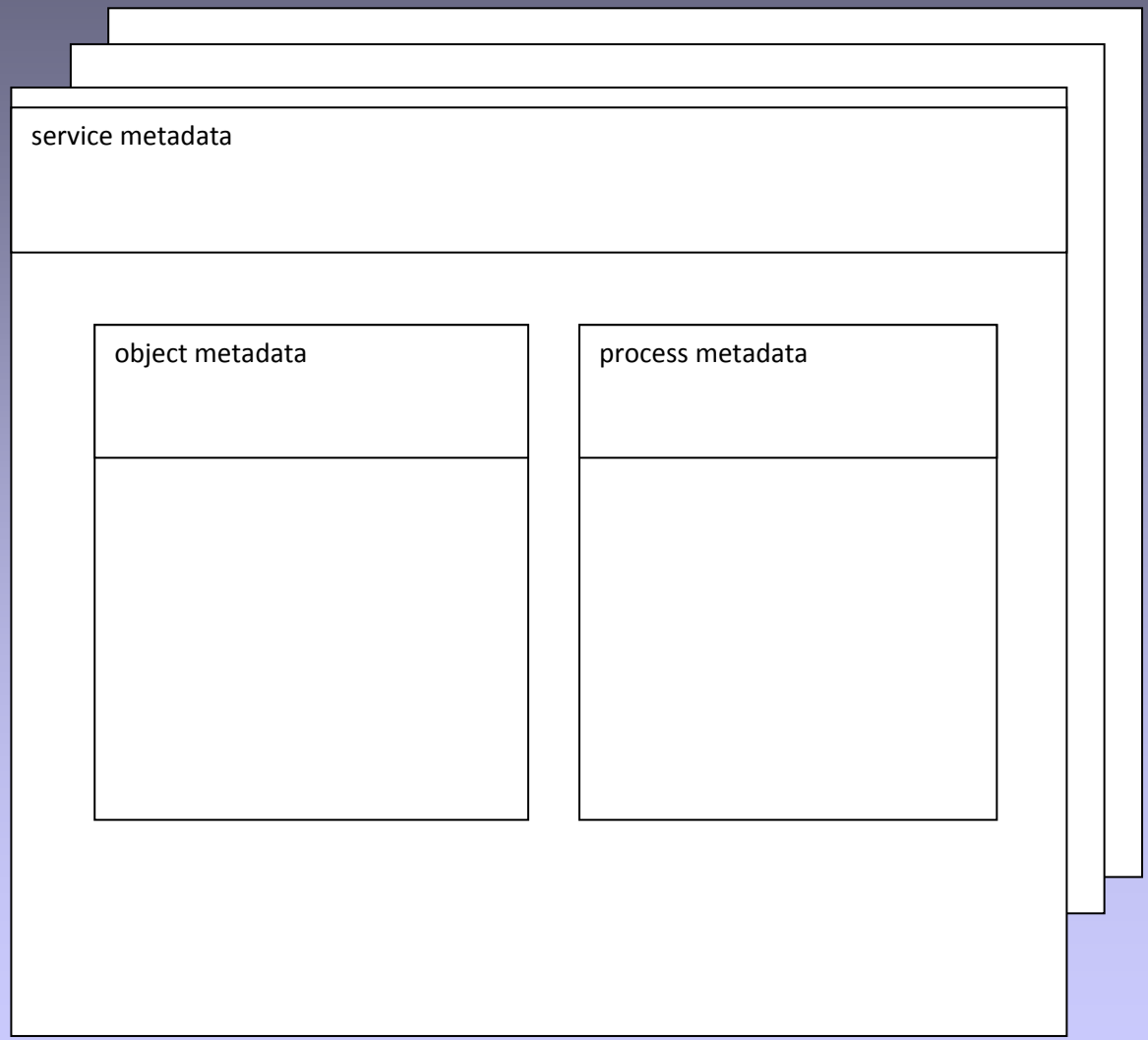
METADATA: CERIF



CERIF: EU Recommendation to Member States

WORKFLOW: SERVICES

- object-independent i.e. generic processes that act on any data
- object-dependent i.e. including and enclosing the object(s) together with the processes.



Conclusion

- The proposed GLASS architecture achieves transparency through several mechanisms:

- encouraging the provision of full metadata using CERIF to cover all aspects of the grey data object thus maximizing the potential utilisation and providing information relating to integrity and quality;
- encouraging the provision of full metadata using CERIF to cover all aspects of services thus maximizing the potential utilisation (including in composed services) and providing information relating to integrity and quality;
- through CERIF defining metadata with formal syntax (for reliable computer processing) and declared semantics (for computer or human understanding);
- through CERIF providing a data model which records the date/time interval associated with any relationship between two base entities. This provides automatically a provenance track and also can be used for non-functional aspects such as security, privacy, rights restrictions;