# GL Transparency: Through a Glass[1] Clearly[2]

Keith G Jeffery STFC, Anne Asserson UiB

## Abstract

GL (Grey literature, interpreted here as grey objects) is very heterogeneous in content, form and quality. Most GL objects evolve through a workflow. Some of these phases involve some form of evaluation or peer review, commonly internal within the management structure of an organisation and possibly involving external advice, including from 'friendly peers' via an e-preprint mechanism. Unlike white literature the evaluation process commonly is unrecorded and undocumented. This leads to accusations that grey literature lacks quality and transparency. This paper proposes how the GL community can overcome this – generally unfounded – accusation, building on our previous work.

A GL repository records the intellectual property of that organisation (2004). We have demonstrated that effective use of this resource requires that the metadata is

---

[1] Grey Literature Architecture for Sustainable Systems

[2] "For now we see through a glass, darkly". The Bible: 1 Corinthians xiii, 12

formalised (1999, 2004) – more precisely in a CERIF-CRIS (Common European Research Information Format – Current Research Information System) (2005). The GL is then available in the context of the work of the organisation and/or its stakeholders managing strategy, evaluation, funding and cost-accounting, innovation and knowledge transfer and public information (2005). This provides user-evaluated assurance on the quality and relevance of the grey object. CERIF provides temporally-based relationships between grey objects (and white objects) thus recording evolution of the object during the workflow – hence provenance. This concept was further refined as 'Greyscape' (2007) and the technologies for interoperation – in order to provide the underpinning homogeneous access to the heterogeneous repositories - surveyed (2008). Efficiency of using CERIF was outlined in (2009). Using advanced hyperactive objects (2006) is postponed until the requirement is realised by the community.

CERIF-CRIS provides the capability for greater quality and transparency through novel methods of evaluating quality, provenance and review including Web2.0 recommender-type systems as well as conventional review mechanisms. CERIF-CRIS provides the way to overcome criticism of GL.

The key messages are:

1. formal metadata associated with grey literature repositories improves relevance and quality;

2. transparency requires recording the workflow phases of a grey object within the context of a research information system;
3. a solution – CERIF – exists already which covers these requirements.

# 1 BACKGROUND

## 1.1 Previous Work

For more than two decades, the authors have worked on research information in the widest sense comprising information not only about grey literature (grey objects) but also all the outputs of research (products, patents, publications) and the context within which the research was done including projects, organizations, funding, persons, facilities, equipment, events. Within the GL community we have highlighted the issues as we see them:

1. the need for formal metadata to allow machine understanding and therefore scalable operations (Jeffery 1999);
2. the enhancement of repositories of grey (and other) e-publications by linking with CRIS (Current Research Information Systems) (Jeffery and Asserson 2004);
3. 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);

4. an architectural model for scaleable, highly distributed, workflowed repositories of grey literature based on hyperactive 'intelligent' documents (Jeffery and Asserson 2006).
5. A 'from 10,000 metres altitude' view of the grey information landscape 'Greyscape' based on the hypothesis that grey literature is the foundation for the knowledge economy (Jeffery and Asserson 2007).
6. An analysis of interoperation architectures among research information systems 'INTEREST' (Jeffery and Asserson 2008).
7. A proposal that Grey Literature should be seen within the context of e-Science supported by a CERIF-CRIS (Jeffery and Asserson 2009).

## 1.2 The Requirement

Our work has convinced us of the need in the Grey literature community for two key technologies:

a) Metadata with formal syntax and declared semantics to allow reliable, scalable management and interoperation of grey resources;

b) Workflow within a research process context to minimize effort for the researcher, research manager, librarian or other knowledge worker, to record the provenance of a grey object and thus to increase accuracy, relevance and contextual awareness of the research information;

These technologies are required for many purposes in GL including – but not limited to – transparency. 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.

We contend that the currently widely-accepted metadata standards for GL – namely Dublin Core (DC) and (MARC) – are insufficient for the purposes of:

a) Discovery (relying on multilingual semantics over multicharactersets);

b) Management (utilising especially the formal syntax);

c) Utilisation (including security and privacy);

d) Understanding (relying on semantics);

e) Re-purposing (relying on both syntax and semantics);

f) Contextualising (utilising contextual metadata such as project, organisation);

g) Provenance (the stages through which the material has been);

h) Preservation/curation (for later re-use by future researchers);

i) Quality assessment (utilising the recorded workflow steps (provenance));

Without appropriate metadata transparency is lost (and the impact of the work recorded in the GL object is much reduced).

Further, we contend that unless GL material is collected in the context of a research workflow of services acting on the grey objects:

a) the threshold barrier to collection is high and discourages those producing the GL from providing the metadata (or even the source material);

b) 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;

We propose that both of these problems are overcome by use of a CERIF-CRIS.

## 2 THE HYPOTHESIS

The hypothesis is in three assertions:

1. formal metadata associated with grey literature repositories improves relevance and quality;
2. recording the workflow phases of a grey object within the context of a research information system provides provenance;
3. a solution – CERIF – exists already which covers these requirements.

Utilisation of this technology provides a GLASS (Grey Literature Architecture for Sustainable Systems) enabling GL users previously "seeing through a glass darkly" to see clearly.

# 3 PROPOSED ARCHITECTURE

## 3.1 Introduction

The proposed GLASS – to achieve all the required aspects of a GL environment including transparency - consists of grey objects, metadata and services operating over a virtualised e-infrastructure based on GRIDs (as proposed in (Jeffery and Asserson 2009) or CLOUD technology (for a survey and analysis see (Schubert, Jeffery, Neidecker-Lutz 2010)) thus in the same domain as that in which researchers do their other work. In this way activities associated with GL are not divorced from observation, experimentation, simulation or project management.

## 3.2 Grey Objects

It is expected that the grey objects will be heterogeneous (either in the local collection or the virtual collection obtained by accessing across heterogeneous distributed GL repositories) and of various (multi)media types. The only architectural problems concerning the objects are to ensure that appropriate services are available to utilise them (see list of functions in section 1.2). This implies rich metadata related to the objects to characterise the way in which they are utilised.

### 3.3 Metadata for Grey Objects

There are several classifications of metadata and multiple standards across many domains of scholarly research. The classification of metadata should relate to the purposes for which it will be utilised (through services available to the user) related to the object. For example schema metadata is used to assure integrity whereas descriptive metadata is used – among other purposes – for discovery.

**Table 1: Metadata Kinds for Grey Objects Related to Services**

| SERVICE | METADATA | COMMENTS |
|---|---|---|
| Discovery | Descriptive | Multicharacterset, multilingual |
| Management | Schema<br>Descriptive<br>Restrictive<br>Navigational<br>Provenance<br>Curation/Preservation | Depending on the management process different kinds of metadata are utilised |
| Utilisation | Schema<br>Descriptive<br>Restrictive<br>Navigational | The schema metadata connects the object to the service assuring integrity, the descriptive metadata |

| | | |
|---|---|---|
| | | assures relevance and the restrictive metadata assures rights, security, privacy compliance |
| Understanding | Descriptive<br>Provenance | The descriptive metadata assures relevance. The provenance metadata illuminates the evolution of the object |
| Re-Purposing | Schema<br>Descriptive<br>Restrictive<br>Navigational<br>Contextual<br>Provenance<br>Curation/Preservation | In order to re-use a grey object as much information about it as possible is required to assure that the re-use is valid. |
| Contextualising | Contextual | For example placing the object in the context of a research project, or related to a research facility |
| Provenance | Provenance | One aspect of quality |

| | | |
|---|---|---|
| Preserving / Curating | Schema<br>Descriptive<br>Restrictive<br>Navigational<br>Contextual<br>Provenance<br>Curation/Preservation | All metadata is needed to allow re-purposing at a later time when the grey object creator may be unavailable |
| Quality Assessment | Schema<br>Contextual<br>Provenance | The schema metadata provides integrity, provenance metadata describes the object evolution and contextual metadata covers the research context |

Clearly there are advantages if the metadata for grey objects is stored within one standard structural environment. CERIF provides such an environment covering all the kinds of metadata outlined above, except schema which – by definition – relates to the conceptual, logical and physical representation of the object within the hosting environment. The availability of this rich metadata for grey data objects assures transparency.

### 3.4 Services

Services are executed to fulfil the requirements of the end-user. Services, themselves, require metadata in the same way as grey objects. Services need to be discovered, managed etc. Services can be :

(a) object-independent i.e. generic processes that act on any data or

(b) object-dependent i.e. including and enclosing the object(s) together with the processes.

Examples of (a) are the relational algebra operators (select, project, difference, union, join) which then act on any relational table(s) whatever the data stored in those tables. Examples of (b) include currency conversion services where the current exchange rate table is incorporated within the service and its associated atomic processes. There is an argument that composed services should themselves be described by metadata and contain both processes and objects, each of which is also described by metadata.

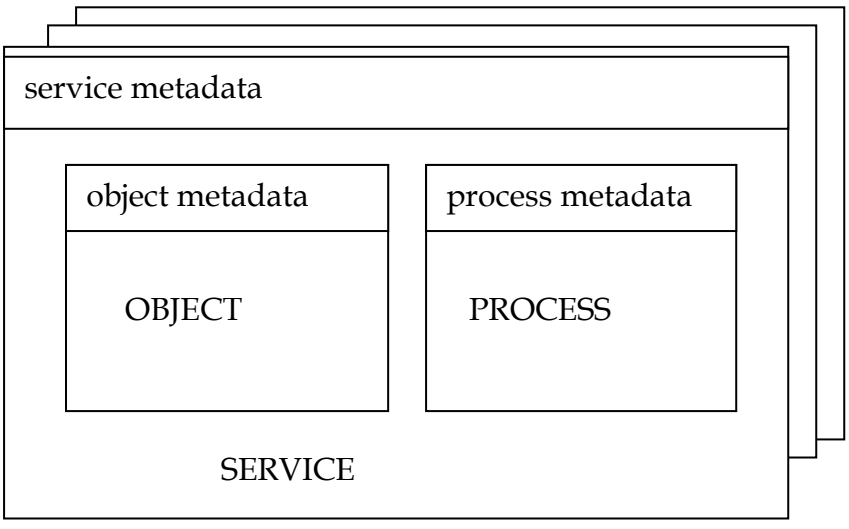


**Figure 1: Service, Object, Process Metadata**

However, whereas objects can be collected together in collections (usually as a set of similar objects based on some parameters e.g. all grey objects generated in 2010 relating to global warming and climate change) services also may be composed – that is groups of atomic services are linked together in some ordered fashion to execute the requirements of the end-user. A simple sequence of services (e.g. select, count) is a simple workflow. A more complex workflow has decision points and branches. However, for reasons of performance and resilience – especially in a heterogeneous distributed environment – the composition can include multiple parallel tracks of workflow with replicated services. It is necessary that each service can be executed anywhere – depending on requirements of performance, security etc – which demands that the services are mobile – that is the program code can be moved to the locus of execution. This leads to the requirement for self-organising (composing, managing, optimizing) services because the complexity of matching the requirement (including non-functional aspects such as performance, security, adherence to a service level agreement or quality of service) to the execution environment (distributed, heterogenous, parallel, multi-tenanted) is too great and too dynamic for human management.

Within the architecture task group of euroCRIS, a set of services for any CERIF-CRIS is being discussed.

### 3.5 Metadata for Services

The services provided require metadata in order for them to be utilised and – more importantly – for them to be utilised correctly.

1. Schema metadata: analogous to the schema controlling integrity in a data object is used to assure integrity in the service particularly in the parameters and input/output declarations;
2. Navigational metadata: analogous to that for data objects is used to locate the service:
3. Descriptive metadata: analogous to that for data objects is used to discover the service and then (together with the schema and restrictive metadata) to assure fitness for purpose;
4. Restrictive metadata: analogous to that for data objects assures enforcement fo non-functional properties of the service such as performance, security, privacy, rights management, (micro-)payment for usage;
5. Provenance metadata: analogous to that for data objects records the transition states of the service as it evolves;
6. Curation/Preservation metadata: analogous to that for data objects records the additional information require to assure (a) preservation of the service code and documentation (specification) over time e.g. through media conversion and evolution; (b) curation such that the purpose and characteristics may be understood in future time;

7. Contextual metadata: analogous to that for data objects describing how the service fits within a context of other research information such as projects, organizations, facilities, equipment etc. CERIF is the preferred standard for use here.

The availability of this rich metadata for services in the GL domain assures transparency.

# 4 CONCLUSION

The proposed GLASS architecture achieves transparency through several mechanisms:

1. 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;
2. encouraging the provision of full metadata using CERIF to cover all aspects of services thus maximizing the potential utilization (including in composed services) and providing information relating to integrity and quality;
3. through CERIF defining metadata with formal syntax (for reliable computer processing) and declared semantics (for computer or human understanding);
4. 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;

# 5 REFERENCES

(Asserson and Jeffery 2004) Asserson, A; Jeffery, K.G.; 'Research Output Publications and CRIS' in A Nase, G van Grootel (Eds) Proceedings CRIS2004 Conference, Leuven University Press ISBN 90 5867 3839 May 2004 pp 29-40 (available under www.eurocris.org )

(Asserson and Jeffery 2005) Asserson, A; Jeffery, K.G.; 'Research Output Publications and CRIS' The Grey Journal volume 1 number 1: Spring 2005 TextRelease/Greynet ISSN 1574-1796 pp5-8

(CERIF) www.eurocris.org/cerif

(DC) http://www.oclc.org:5046/research/dublin_core/

(DSpace) DSpace repository product homepage http://www.dspace.org/

(ePubs) ePubs repository product homepage http://epubs.cclrc.ac.uk/

(eprints) ePrints repository product homepage http://www.eprints.org/

(Fedora) Fedora repository product homepage http://www.fedora.info/

(FRIDA) http://frida.uio.no

(Jeffery 1999) Jeffery, K G: 'An Architecture for Grey Literature in a R and D Context' Proceedings GL'99 (Grey Literature) Conference Washington DC October 1999 http://www.konbib.nl/greynet/frame4.htm

(Jeffery, 2000). Jeffery, K G: 'Metadata': in Brinkkemper,J; Lindencrona,E; Solvberg,A (Eds): 'Information Systems Engineering' Springer Verlag, London 2000. ISBN 1-85233-317-0.

(Jeffery 2004a) Jeffery, K.G.; 'GRIDs, Databases and Information Systems Engineering Research' in Bertino,E; Christodoulakis,S; Plexousakis,D; Christophies,V; Koubarakis,M; Bohm,K; Ferrari,E (Eds) Advances in Database Technology - EDBT 2004 Springer LNCS2992 pp3-16 ISBN 3-540-21200-0 March 2004

(Jeffery 2004b) Jeffery, K.G.; 'The New Technologies: can CRISs Benefit' in A Nase, G van Grootel (Eds) Proceedings CRIS2004 Conference, Leuven University Press ISBN 90 5867 3839 May 2004 pp 77-88 (available under www.eurocris.org )

(Jeffery 2005) K G Jeffery CRISs, Architectures and CERIF CCLRC-RAL Technical Report RAL-TR-2005-003 (2005)

(Jeffery and Asserson 2004) K G Jeffery, A G S Asserson; Relating Intellectual Property Products to the Corporate Context; Proceedings Grey Literature 6 Conference, New York, December 2004; TextRelease; ISBN 90-77484-03-5

(Jeffery and Asserson 2005) K G Jeffery, A G S Asserson 'Grey in the R&D Process'; Proceedings Grey Literature 7 Conference, Nancy, December 2005; TextRelease; ISBN 90-77484-06-X ISSN 1386-2316

(Jeffery and Asserson 2006a) Keith G Jeffery, Anne Asserson: 'CRIS Central Relating Information System' in Anne Gams Steine Asserson, Eduard J Simons (Eds) 'Enabling Interaction and Quality: Beyond he Hanseatic League'; Proceedings 8th International Conference on Current Research Information Systems CRIS2006 Conference, Bergen, May 2006 pp109-120 Leuven University Press ISBN 978 90 5867 536 1

(Jeffery and Asserson 2006b) Keith G Jeffery, Anne Asserson: 'Supporting the Research Process with a CRIS' in Anne Gams Steine Asserson, Eduard J Simons (Eds) 'Enabling Interaction and Quality: Beyond the Hanseatic League'; Proceedings 8th International Conference on Current Research Information Systems CRIS2006

Conference, Bergen, May 2006 pp 121-130 Leuven University Press ISBN 978 90 5867 536 1

(Jeffery and Asserson 2006c) Keith G Jeffery, Anne Asserson: 'Hyperactive Grey Objects' Proceedings Grey Literature 8 Conference (GL8), New Orleans, December 2006; TextRelease; ISBN 90-77484-08-6. ISSN 1386-2316 ; No. 8-06-X

(Jeffery and Asserson 2007) Keith G Jeffery, Anne Asserson: 'Greyscape' Opening Paper in Proceedings Grey Literature 9 Conference Antwerp (GL9) 10-11 December 2007 pp9-14; Textrelease, Amsterdam; ISSN 1386-2316

(Jeffery and Asserson 2008) Keith G Jeffery, Anne Asserson: 'INTEREST' Proceedings Grey Literature Conference Amsterdam 8-9 December 2008 in Tenth International Conference on Grey Literature : Designing the Grey Grid for Information Society, 8-9 December 2008, Science Park Amsterdam, The Netherlands ed. by Dominic J. Farace and Jerry Frantzen ; GreyNet, Grey Literature Network Service. - Amsterdam : TextRelease, February 2009. GL-Conference series, ISSN 1386-2316; No. 10. - ISBN 978-90-77484-11-1.

(MARC) http:// http://www.loc.gov/marc/

(PURE) http:// http://www.atira.dk/en/pure/

(RDF) http://www.w3.org/RDF/

Schubert, Jeffery, Neidecker-Lutz 2010 'The Future of Cloud Computing' http://cordis.europa.eu/fp7/ict/ssai/docs/cloud-report-final.pdf .

(XML) http://www.w3.org/XML/

(Z39.50) http:// http://www.loc.gov/z3950/agency/