Hyperactive Grey Objects

Keith G Jeffery

Director, IT &
International Strategy

CCLRC

keith.g.jeffery@rl.ac.uk

Anne G S Asserson

Research Department

University of Bergen

anne.asserson@fa.uib.no

GL8 2006 New Orleans
20061204-05
1

Background

- Previous papers on Grey literature by the authors (in the GL Conference Series) have described
  - the need for formal metadata to allow machine understanding and therefore scalable operations;
  - the enhancement of repositories of grey (and other) e-publications by linking with CRIS (Current Research Information Systems);
  - 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.
- This paper takes the development one step further and proposes 'intelligent' grey objects.
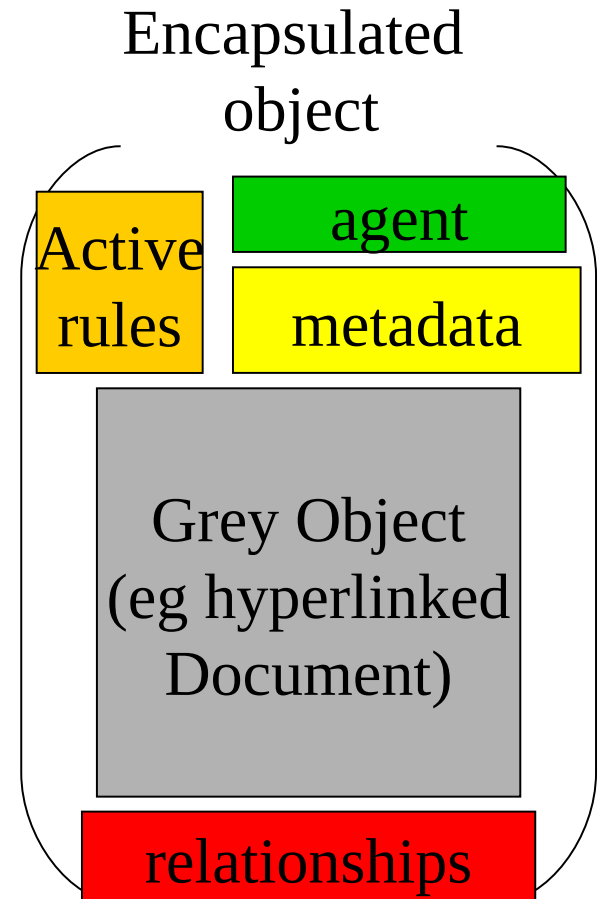
Hypothesis

- The hypothesis is in 2 parts:
  - that the use of passive catalogs of metadata does not scale
    - in a highly distributed environment with millions of nodes
    - with vastly increased volumes of R&D output grey publications with associated metadata;
  - that a new paradigm is required that
    - integrates grey with white literature and other R&D outputs such as software, data, products and patents
    - in a self-managing, self-optimising way and that this paradigm manages automatically curation, provenance digital rights, trust, security and privacy.

State of the Art

- Concerning the problem of scaling
  - Existing repositories catalogs - cost of input / update;
  - Harvesting takes increasing time - ensuring non-currency
  - To obtain and utilise results the end-user expends much manual effort / intelligence
  - The elapsed time of
    - the network
    - the centralised (or centrally controlled distributed) catalog server searches
    - end-user intervention

    becomes unacceptable.

- Concerning the proposed solution
  - there is no paradigm currently known to the authors that satisfies the requirement - hence our proposal is developed.

The Notion

- Hyperactive combines both
  - hyperlinking
  - active properties of a (grey) object.
- Hyperlinking implies multimedia components linked to form the object and also external links to other resources.
- The term active implies that objects do not (only) lie passively in a repository to be retrieved by end–users. They 'get a life' and the object moves through the network knowing where it is going.

Encapsulated object

Active rules

agent

metadata

Grey Object
(eg hyperlinked
Document)

relationships

GL8 2006 New Orleans

20061204-05

5

How it Works

- A hyperactive grey object is wrapped by formal metadata, active rules, relationships, associated (software) agent.

- It moves through process steps

- The workflow is based on the rules and information in the CDR (Corporate Data Repository) and the active rules within the encapsulated grey object

- Once the object is deposited, the agent associated with it actively checks and builds the relationships

- On publication: the agent pushes the object to the end-users (or systems)

- The agents check the object and user (or system) restrictive metadata

- Alternatively the object can be found passively by end-user or system agents using its descriptive metadata. (again the agents check the restrictive metadata)

Wider Context

- The object can also associate itself with other objects forming relationships utilising metadata or content.

  – Declared relationships include references and citations;

  – Workflowed relationships include versions and also links to corporate information and research datasets and software;

  – Inferenced relationships are discovered relationships such as between documents by different authors developed from an earlier idea of a third author.

CDR System

Workflow authorise

Push system

Relation System

publication
review
deposit
authorise
author

push
publication
review
deposit
authorise
author

review
deposit
authorise
author

deposit
authorise
author

authorise
author

author

create

Review/
approve

Deposit

Peer
Review

Public
ation

Publication
ation

Institutional Repository
system

Other Institutional
Repository
systems

Institutional research data
and software repository
system

Other Institutional research
Data and software repository
systems

action

relationlink

metadata

active rules

© Keith G Jeffery, Anne G S Asserson   **GL8 2006 New Orleans**   20061204-05   8

Current State

- Components of this paradigm have been implemented to some extent
  - Formalised metadata (and links to CRIS);
  - Active rules
  - Workflow using CDR;
  - Hyperlinking of Objects;
  - Inferencing (usually for plagiarism);
  - Agents
  - Self-* properties
- Much of it in a GRIDs environment

Conclusion

1. Minimum effort by the end-user

2. Maximum input by system

3. Maximum use of information by system

    a) To manage workflow

    b) To manage relationships

    c) To generate relationships

    d) To assist retrieval

4. Scalable

This surely is

    harnessing the power of grey.

Anne G S Asserson
Research Department
University of Bergen

anne.asserson@fa.uib.no

Keith G Jeffery
Director, IT &
International Strategy
CCLRC
keith.g.jeffery@rl.ac.uk