Interoperability best practices and solutions: The DL.org Cookbook

Leonardo Candela
ISTI - CNR

“Digital Library and Open Access. Interoperability strategies” Workshop
London, 4 February 2011

Outline

- What a Technology and Methodology Cookbook is expected to be
- An Interoperability Framework
- What the Cookbook is now

Digital Library Technology and Methodology Cookbook

- Scope
  - a portfolio of best practices and pattern solutions to common issues faced when developing large-scale interoperable Digital Library systems
  - organised in patterns, i.e. standard, well-recognized or proven solutions to DL development problems
  - each pattern is characterized by
    - a context, a situation giving rise to a problem
    - a problem, the recurring problem arising in that context
    - a solution, a proven resolution of the problem
    - an assessment, a qualitative evaluation of the approach
- Release schedule
  - RFC version (October 2010)
  - Final version (February 2011 - ...)

The interoperability monster

- A companion in any system build as “collection” of existing constituents
  - “... integration of heterogeneous data. “A major challenge is still really hard is schema mapping—converting data from one format to another. ”... “It sounds straightforward, but it’s very subtle.”
  - “... the “unsolved problem” of querying geographically distributed databases”
- An ever hot topic, e.g. the Digital Agenda for Europe
- Some thoughts on Interoperability (by courtesy of Y. Ioannidis, DL Foundations 2008)
  - Is it difficult?
    - yes it is, it is (almost) impossible
  - Is it about content/functionality?
    - is it about content, functionality, user, policy, quality and architecture, it is about (almost) everything
  - What kind of job it?
    - dirty but critical
    - broad but partitionable
    - complex but fun
    - will never be solved but must be solved even approximately
Cookbook motivations and challenges

• A lot of solutions exist / are developed
  – lack of systematic approach
  – scarce knowledge of “others” solutions
    • “let’s not to reinvent the wheel”
• Provide an organised, pragmatic, comprehensive and effective description of ways to attack the interoperability monster
• Challenges
  – scope (almost everything)
  – partitioning schema (per domain vs. cross-domain)
  – multidisciplinary nature (beyond technology)

*** an unifying interoperability framework is needed ***

Interoperability Framework

• Three main aspects has to be captured:
  – interoperability scenario, i.e. a setting where interoperability takes place
  – interoperability issue, i.e. a problem hindering an interoperability scenario
  – interoperability solution, i.e. an approach aiming at removing an interoperability issue in the context of an interoperability scenario

Interoperability Framework (cont.)

• A plethora of interoperability definitions

Interoperability Framework (cont.)

• Rephrasing and elaborating
  – at least two entities Provider and Consumer
  – willing to “share” a Resource to perform a Task (has preconditions)
  – through a Communication Channel, involving a protocol and information representation
  – across Organizational, Semantic and Technical boundaries of entities
Interoperability Framework: focusing on the three key aspects

- **Organizational** deals with business goals and processes of an entity (Provider or Consumer)
- **Semantic** deals with the meaning of the exchanged resource and the rest of information
- **Technical** deals with technological solution supporting the operation of the Provider / Consumer as well as the communication among the two

- Dependencies / constraints among the three boundaries, e.g. organisational aspects has to be implemented by the technical aspects
- Implicit / hidden information, e.g. the technical aspects might implement part of the organisational aspects only
- Different approaches for diverse boundaries, e.g. human-centric vs. machine-centric
- Complete solutions involve all of them, e.g. the decision to rely on a certain technology might be useless if it is not complemented by proper organisational aspects

Interoperability approaches: Agreement-based

- Include Standard-based approaches
- Infringes autonomy, strong in effectiveness

Consumer-oriented schema

Provider-oriented schema

Interoperability approaches: Mediator-based

- An intermediary service linking Provider and Consumer
- Strong in autonomy, development and maintenance cost

Consumer-side schema

Provider-side schema

Third-party schema

Interoperability Framework: Issue and Solution

- **Interoperability Issue**
  - the factor hindering interoperability scenario
  - task preconditions are not satisfied
- **Interoperability Solution**
  - it can be applied under certain conditions
  - if effective it removes the interoperability issue
  - it contains a “transformation function” (even in its most abstract case)
- A solution might be “wider” than the issue
  - this is not a problem if its transformation is scenario “safe”
Interoperability approaches: Blending and Compound solutions

- Combining Agreement-based & Mediator-based approaches, e.g.
  - a Mediator implementing a Standard and complementing it
  - two mediators or two standards form a new solution
- Compatibility issue among solutions
- Alternative solutions
- No solution exist vs for each problem there exists at least a solution

How is an interoperability solution described?

1. **Overview**
   Context of the proposed approach including pointers to detailed description of it

2. **Requirements**
   Conditions under which the solution might be used

3. **Results**
   Changes resulting from the usage of the solution

4. **Implementation guidelines**
   How the changes are produced

5. **Assessment**
   Qualitative evaluation of the proposed solution

Current Solutions Sample

- **Content**
  - ID publishing (e.g. OAIPMH, OA-ORE), metadata (e.g. DublinCore, Europeana, Application Profiles), identifiers (e.g. DOI), ...
- **User**
  - User modeling (e.g. GUMO), model conversion (e.g. GUC), AuthN/AuthZ (e.g. OpenID, SAML), ...
- **Functionality**
  - Interface specification and reconciliation (e.g. WSDL, adapters), behaviour specification and reconciliation (e.g. WS-BPEL, automatic composition)
- **Policy**
  - Almost huge and uncharted territory in DL
  - Policy languages (e.g. XACML)
- **Quality**
  - Frameworks (e.g. DAQuinCiS), ontologies (e.g. WS-QuS), guidelines (e.g. DRIVER)
- **Architecture**
  - Largely overlap with the rest
  - Component profile (e.g. WS-I, WSDL), protocols (e.g. SRU, OpenSearch), ...
- **Cross-domain**
  - Provenance (e.g. OPM)
Cookbook Next Steps

- RFC feedback and contributions (ongoing)
- Cookbook 1.0 Release (February ‘11 -> …)
- Beyond DL.org
  - Cookbook wikification (editorial control)
  - Awareness raising

Thank you
wiki.dlorg.eu/index.php/Cookbook