



Quality interoperability within DL.org

Giuseppina Vullo

7 October 2010

DL.org Autumn School – Athens, 3-8 October 2010



A question to start

What is a “good quality” DL?

What is Quality?



Quality is something which makes the difference

What is Quality?



Quality means making choices

What is Quality?



Quality needs time, and involves the concepts of standards and best practices

What is Quality?



Quality is always subjective to humans, which are involved in the development & selection of systems

Quality

DELOS RM *Parameters that can be used to characterise and evaluate the content and behaviour of a DL. Quality can be associated not only with each class of content or functionality but also with specific information objects or services*

Quality

But also...

- the degree that the DL conforms to the specified policy that expresses what the **goal** of a DL is. The policy can cover from very general guidelines to very technical issues
- applicable to either **overall** or **single aspects** of any products, services and processes, usually defined in relation to a **set of guidelines and criteria**. Often **implicit**

Quality comprehensive models

Gonçalves et al., 2006

What is a good digital library? A quality model for digital libraries

Table 1
DL high-level concepts and corresponding DL dimensions of quality with respective metrics

DL concept	Dimension of quality	Factors/variables involved in measuring
Digital object	Accessibility	Collection, # of structured streams, rights management metadata, communities
	Pertinence	Context, information, information need
	Preservability	Fidelity (lossiness), migration cost, digital object complexity, stream formats
	Relevance	Query (representation), digital object (representation), external judgment
	Similarity	Same as in relevance, citation/link patterns
	Significance	Citation/link patterns
	Timeliness	Age, time of latest citation, collection freshness
Metadata specification	Accuracy	Accurate attributes, # of attributes in the record
	Completeness	Missing attributes, schema size
	Conformance	Conformant attributes, schema size
Collection	Completeness	Collection size, size of the 'ideal collection'
Catalog	Completeness	# of digital objects without a set of metadata specifications, size of the described collection
	Consistency	# of sets of metadata specifications per digital object
Repository	Completeness	# of collections
	Consistency	# of collections in repository, catalog/collection pair-wise consistency
Services	Composability	Extensibility, reusability
	Efficiency	Response time
	Effectiveness	Precision/recall (search), F1 measure (classification)
	Extensibility	# of extended services, # of services in the DL, # of lines of code per service manager
	Reusability	# of reused services, # of services in the DL, # of lines of code per service manager
	Reliability	# of service failures, # of accesses

Quality comprehensive models

Zhang, 2010

*Holistic DL
evaluation
model*

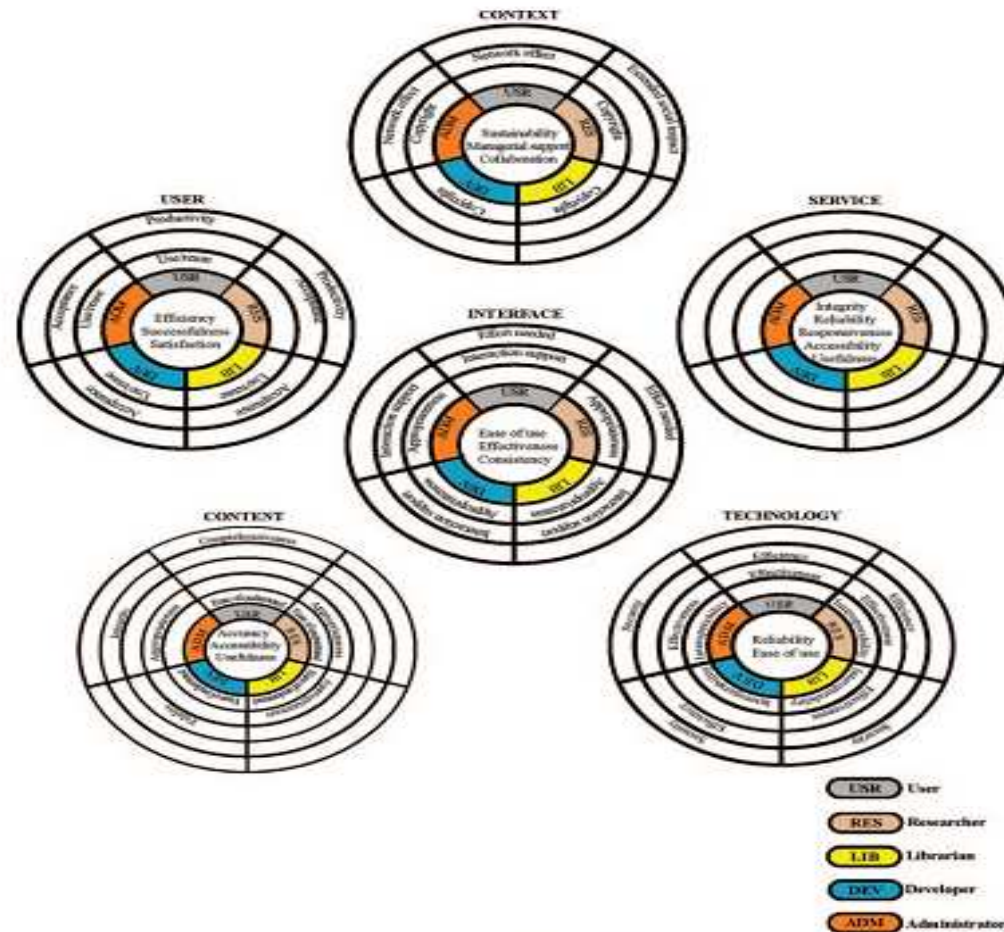
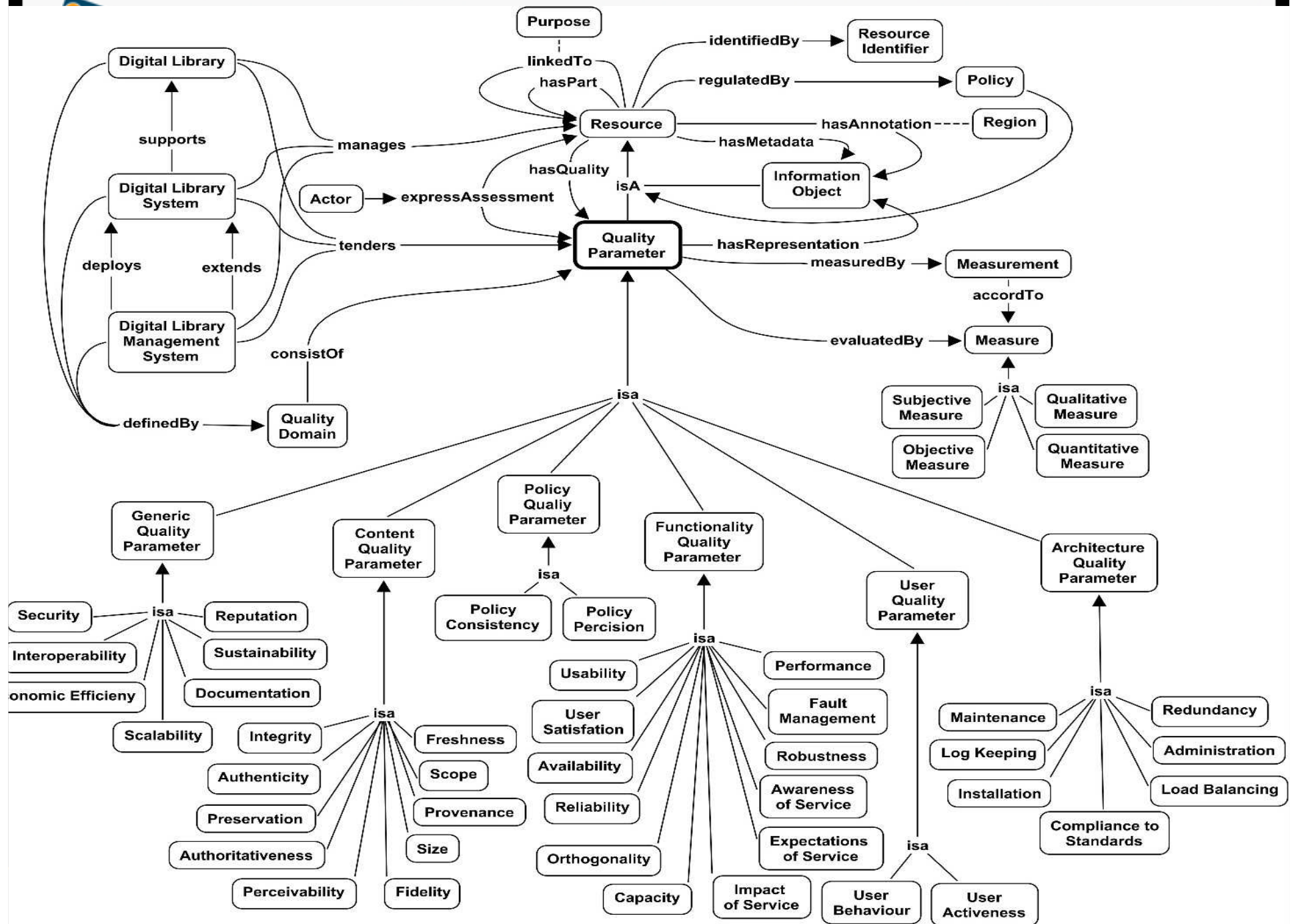


FIG. 5. The proposed holistic DL evaluation model.



Interoperability Framework 2.0

EC 2008

An **Interoperability Framework** describes the way in which organisations have agreed, or should agree, to interact with each other, and how standards should be used. In other words, it provides **policies** and **guidelines** that form the basis for selection of standards

European Interoperability Framework 2.0. EC, 2008

Cooperating partners having compatible visions, and focusing on the same things.

Political Context

The appropriate synchronization of the legislation in the cooperating MS so that electronic data originating in any given MS is accorded to proper legal weight and recognition wherever it needs to be used in other MS.

Legal Interoperability

Legislative Alignment

The processes by which different organisations such as different public administrations collaborate to achieve their mutually beneficial, mutually agreed eGovernment service-related goals.

Organisational Interoperability

Organisation and Process Alignment

Ensuring that the precise meaning of exchanged information (concept, organisation, services, etc) is preserved and well-understood

Semantic Interoperability

Semantic Alignment

The technical issues involved in linking computer systems and services (open interfaces, interconnection services, data integration, middleware, data presentation and exchange, accessibility and security services, ...)

Technical Interoperability

Syntax, Interaction & Transport

Quality interoperability

- Establishment, adoption and measurement of quality requirements and performance indicators... **How these requirements/ indicators can interoperate?**
- Interrelations → low quality services can affect the degree of interoperability among different components, preventing the successful cooperation among different systems
- **The possibility for DLs to share a common quality framework**
- Decentralised paradigm on how to link heterogeneous and dispersed resources keeping reliability of services, data precision, homogeneous experience for the end user

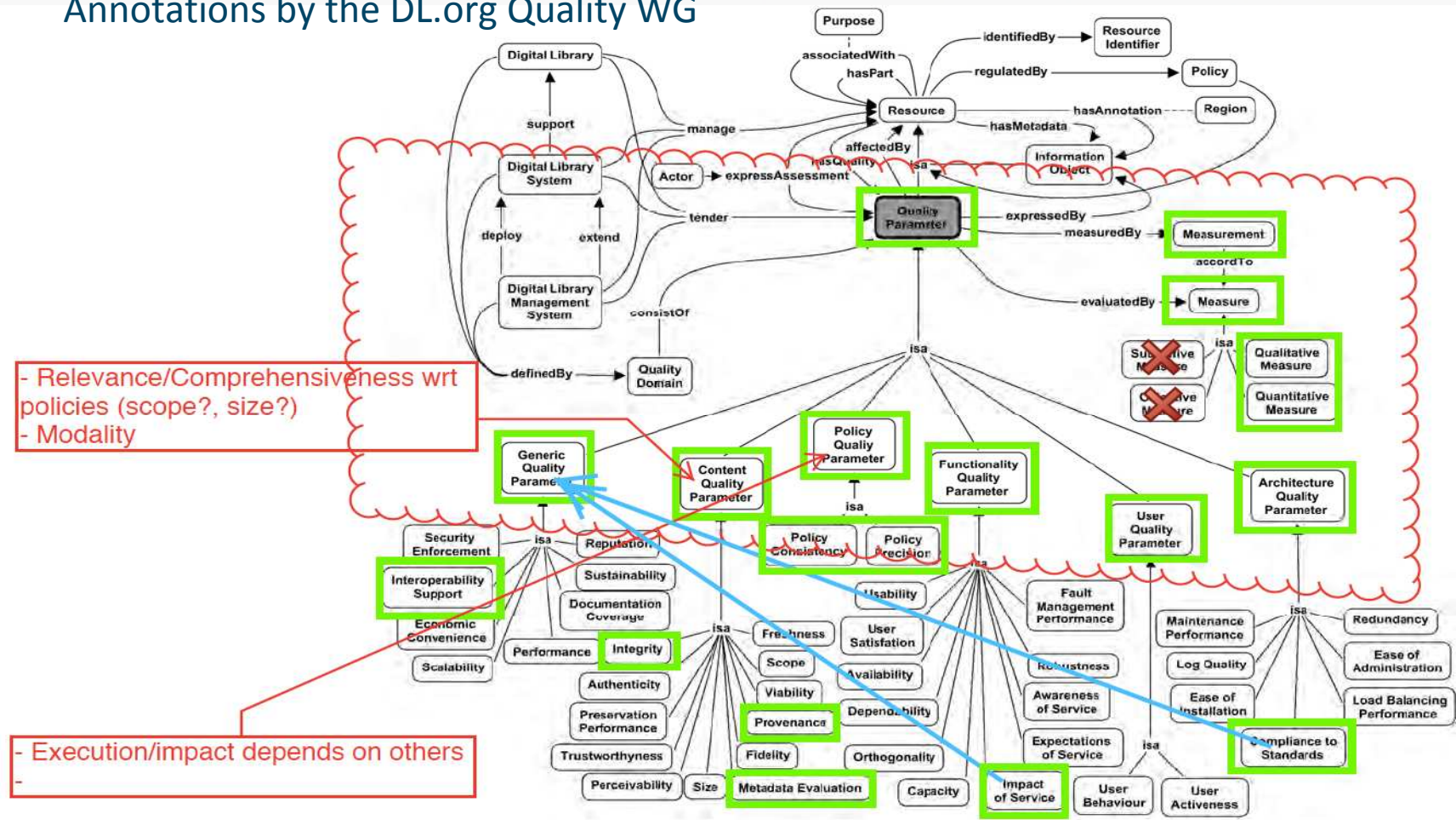
Quality WG motivating interoperability scenario

Our motivating scenario: consider that representatives of two (or more) DLs have a round table to negotiate a service level agreement (SLA) defining their interoperability requirements and for this establish a quality threshold that each individual DL has to meet or exceed; “Quality” would provide transparent qualitative or quantitative parameters for defining the threshold

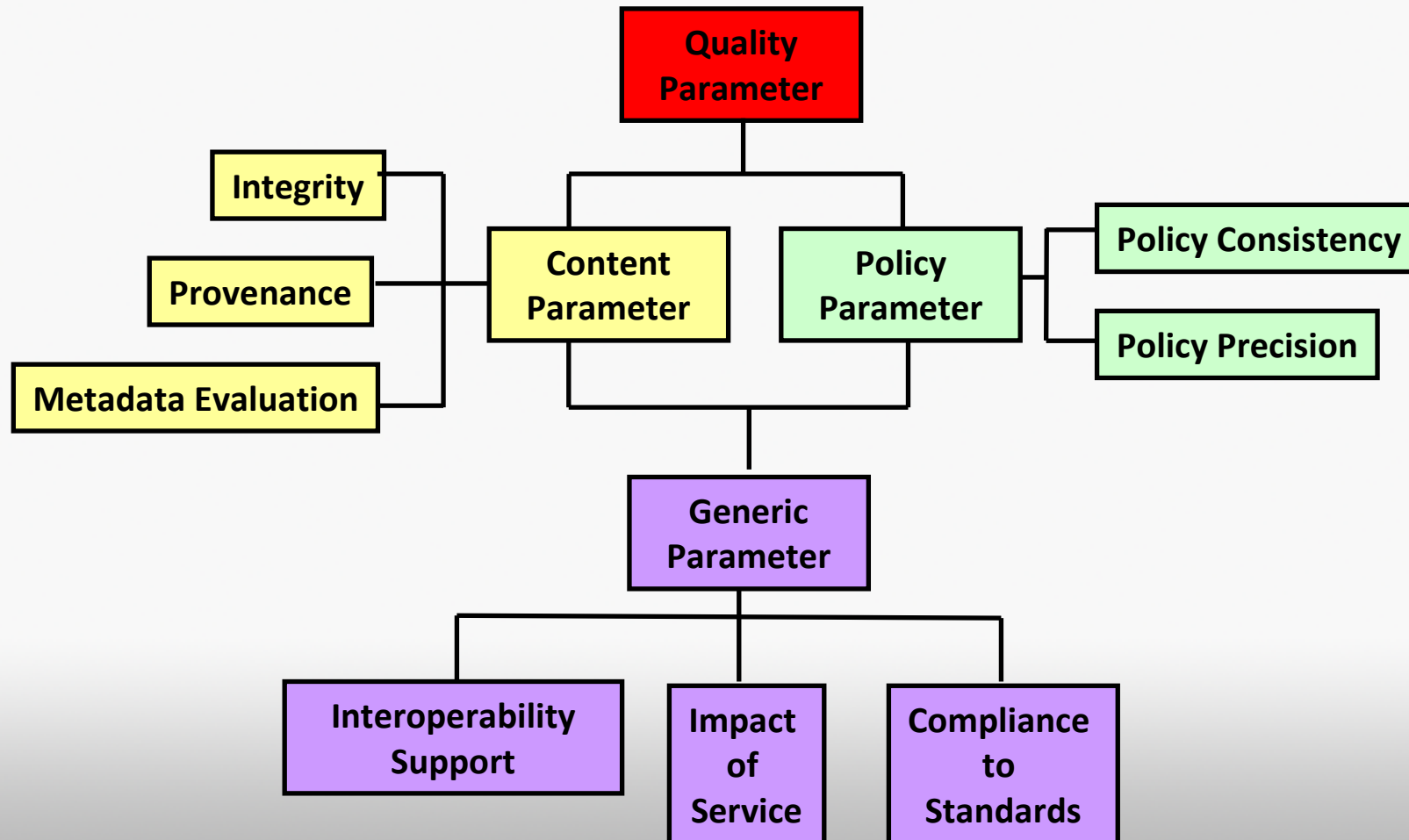
Annotating the Quality Concept Map

Quality Concept Map

Annotations by the DL.org Quality WG



The Quality Core Model



Generic Parameter Interoperability Support

Capability of a digital library to interoperate with other digital libraries as well as the ability to integrate with legacy systems and solutions

Approaches to interoperability:

- Define generic interchange protocols – OAI-PMH
- Set up research infrastructures which define a framework for participants eg. D4SCIENCE

Possible parameters:

- OAI-PMH compliance
- Use of persistent identifiers
- Metadata specifications
- Authorisation and authentication procedures

Related to:

- Compliance to standards

Generic Parameter

Compliance to Standards

The degree to which standards have been adopted in developing, managing and delivering a digital library service

- Quality interoperability depends on the extent a DL adheres to a set of pre-determined rules or codes, which include:
 - Data / content standards
 - Metadata standards
 - Web interface standards
 - Data sharing protocols
- Which framework to adopt depend on the community or discipline involved
- Establish a measurable standards compliance agreement
- Related to:
 - Interoperability support
 - Sustainability

Generic Parameter

Impact of Service

The influence that a digital library service has on the users' knowledge and behaviour

Impact of service can be measured by:

- Increase of user knowledge
- Improvement in DL practical skills over time

Content Parameter

Integrity

The quality of being whole and unaltered through loss, tampering, or corruption

- Completeness & accuracy of the Information object

Related to:

- Metadata integrity
- Regular content update
- Accurate format migrations

Content Parameter

Integrity

User scenario

Collection of journal articles:

- Does the final version of each article appear in DL?
- Are all the pages and figures available?
- Does the scanning quality mean that all pages are clear?
- Has OCR scanning been proof-read and corrected
- For merged collections:
 - Is there only one entry in the catalogue?
 - Have all entries copied correctly?
 - Does the collection only contain what is expected?

Content Parameter

Provenance

Information regarding the origins, custody, and ownership of an item or collection (the resource story, how to establish quality)

- Tracking origins and history of the Information Object to know if it is fit for purpose:
 - Transformations? Cleaning? Rescaling? Modelling? Mergers?
 - Authorship, IPR, integrity and authenticity
- Issues for quality provenance information:
 - metadata standards for tracking provenance?
 - How to capture
 - What to capture
- Related to: Metadata, Annotation, Preservation Policy

Content Parameter Provenance

User scenario

A bioinformatics DL, which supports the analysis of gene expression and analysis, requires tools to be applied to the raw data in a defined workflow.

Are the following maintained?

- Results of workflow
- Intermediate steps of the workflow
- Configuration of tools and algorithms

Content Parameter Metadata Evaluation

The measurements of metadata schemas and their individual fields to support the collection, management, discovery and preservation of digital library content

- Metadata evaluation should look the support in all classes of metadata:
 - Descriptive, Technical, Administrative, Use, Preservation
- Evaluation of metadata for:
 - Use of structure standards
 - Use of content standards
 - Metadata creation
- Related to: Content Quality Parameter , Policy Quality Parameter, Compliance to Standards, Interoperability Support, Scalability, Sustainability

Policy Parameter

Policy consistency & precision

- **Policy consistency** the extent to which a policy or a set of policies are free of contradictions - eg consistency across *Content Policy* and *Registration Policy* (real case DRIVER)
- **Policy precision** the extent to which a set of policies have defined impacts and do not have unintended consequences

Policies should be detailed and defined enough to constrain behaviours, deal with consequences and enforce:

- Envisage aspects of governance
- Sufficient knowledge of technology – architecture and software

Quality Interoperability Survey

- Quality Interoperability Survey
- Survey Pilot
- Disambiguation (Glossary) & Collection strategy
- Data analysis and interpretation
- Best practices & checklist with practical recommendations

Quality Interoperability Survey

Some participants:

German Digital Library

Max-Planck DL

E-prints for Library and Information Science (E-LIS)

Europeana

E-Archivo: Institutional Repository of University Carlos III of Madrid

The European Library (TEL)

DRIVER D-NET

The World Digital Library (WDL)

...

Quality Interoperability Survey

QCM Covered areas

- Formats
- Format compliance checking tools (and results)
- Metadata standards
- Metadata compliance checking tools (and results)
- Communication protocols
- Communication protocol compliance checking tools (and results)
- Web guidelines / standards in the areas of accessibility, usability, multilingualism
- Policies and legal obligations (eg for web standards or DRM)

Quality Interoperability Survey

Monitoring, interoperability, more general info

- Multi-level guidelines and certifications
- User satisfaction
- Current interoperations
- Quality interoperability and the RM

Quality Interoperability Survey

Validations

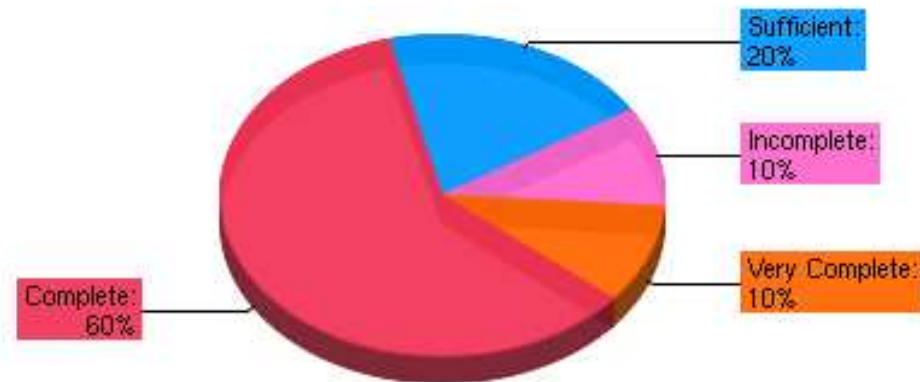
Do you use any validation tools to check

- **Information object format** compliance (eg. Pdf/A Validator)? YES 60%, NO 40%
- **Metadata format** compliance (eg. DC Validator) YES 80% NO 20%
- **Communication protocols** compliance (OAI/PMH & DRIVER Validators) YES 50% NO 50%

Quality Interoperability Survey

Metadata completeness

On a scale 1-5 [1 very incomplete; 2 incomplete; 3 sufficient; 4 complete; 5 very complete], how complete is your metadata?



Quality Interoperability Survey

Metadata creation

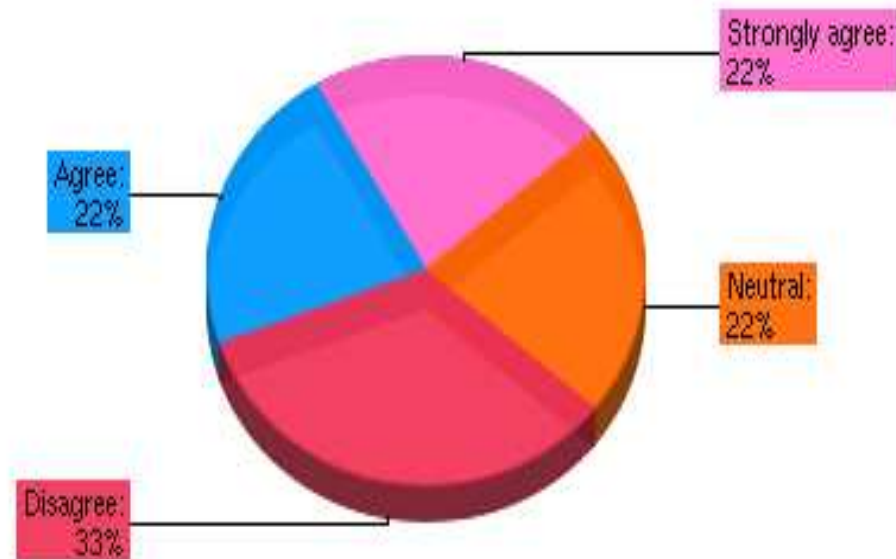
In your opinion, what is the single greatest barrier to metadata creation?

- Time
- Accuracy
- Missing or too complex or contradictory guidelines
- Not having enough humans involved in the process
- Not understanding its real value, reason and purpose
- Review is required by qualified personnel

Quality Interoperability Survey

Is interoperability technical?

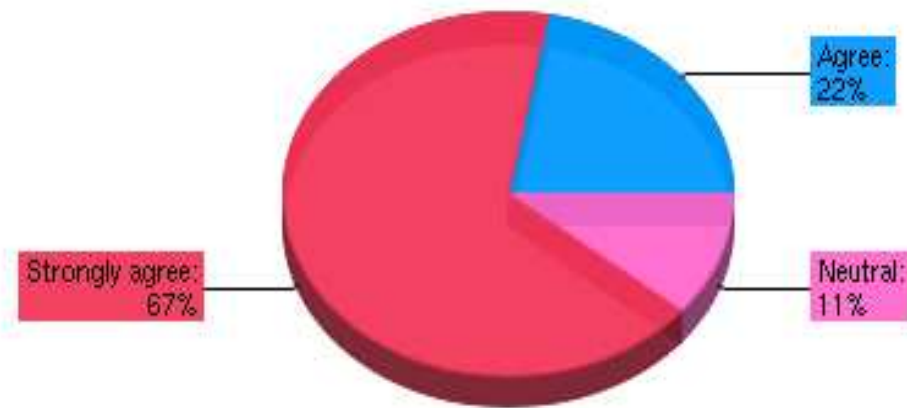
Successful interoperability is largely a technical issue



Quality Interoperability Survey

Quality and interoperability

Quality aspects are crucial for successful interoperability



Quality Interoperability Survey

DELOS RM

Some DLs are **already using** the RM

- Design and operation of processes
- Business and organisational models
- Changes of institutional repositories
- Revision of DL policies

Quality Interoperability Survey

A “good quality” DL

What do you consider to be a “good quality” Digital Library (DL)?

- *A high organisational level of interoperability between objects and people concerning interoperability aspects of embedded devices and process management*
- *I would tend to say that a "good quality" digital library cannot be measured only through the metadata quality or interoperability level. In my eyes Quality is a combination of Content, User satisfaction, Functionality, Policy, Quality, and Architecture of the system. A good level for each of these can lead to a good quality Digital library*
- *Containing consistent and complete metadata; valid identifiers to full-text and other material*
- *Usefulness for the end user, all the functions working, understandable (language and functions), user finds what he/she was looking for (if it can be found), user do not have to print anything*
- *A good quality DL has a strategy and clear target to be compliant to the technical standards mostly accepted in the network, to be easy for its patrons/users, to be oriented to improve something every year*
- *One that provides the services that end users demand and are in line with best practices at the international level*

Some preliminary evidences

- **Metadata-centric** world
- Role of **guidelines** (eg DRIVER, MINERVA, etc.), **certifications** (eg. DINI, Drambora) and **validators**
- Different meanings of **Quality** and **Interoperability**
- Lack of formalised and well-analysed **policies**
- Need to be **supported**

Work in progress

- Complete the **survey** and analyse/interpret data
- Identification and selection of **best practices** and **recommendations** for the Cookbook
- Enhancing the **Quality domain** in the RM
- Elaborating more our **definitions**

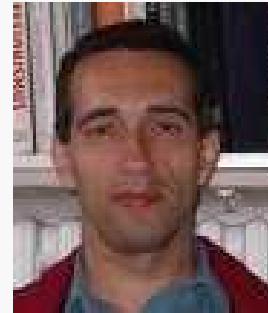
Thank you



Nicola Ferro
University of Padua
WG Scientific Chair



Sarah Higgins
Aberystwyth University (UK)



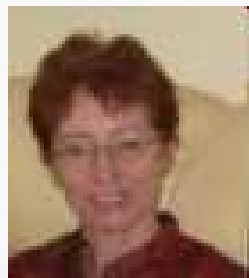
Sarantos Kapidakis
Ionian University



Wolfram Horstmann
University of Bielefeld
DRIVER



René Van Horik
Data Archiving and
Networked Services
(NL)



Genevieve Clavel
Swiss National Library



Seamus Ross
University of Toronto



Giuseppina Vullo
HATII, University of
Glasgow

Hands-on exercise

Exercise

*Build your own **Quality Core Model!***

By creating a hit-list of DELOS RM parameters and prioritising them according to your group interoperability scenario, please present the outcomes explaining us the rationale behind your choice.