



The DL.org Policy and Quality Concepts

Giuseppina Vullo, Perla Innocenti and Seamus Ross
HATII – University of Glasgow

DL.org - DILL International Seminar
“Research and education in digital libraries”

Parma, Italy 9 November 2010



Outline

- Policy within DL.org
- Policy interoperability issues & outcomes
- Quality within DL.org
- Quality interoperability issues & outcomes
- Teaching DLs Policy and Quality: some tips from the DL.org Autumn School

What is policy?



“A **policy** is typically described as deliberate plan of action to guide decisions and achieve rational outcome(s). The term may apply to government, private sector organizations and groups, and individuals”

Source: <http://en.wikipedia.org/wiki/Policy>

What is a policy?

From the DELOS REFERENCE MODEL

“The policy concept represents the set or sets of **conditions, rules, terms and regulations governing interactions** between the **Digital Library** and its **users**, whether virtual or real. [...]”

Definition of Digital Library

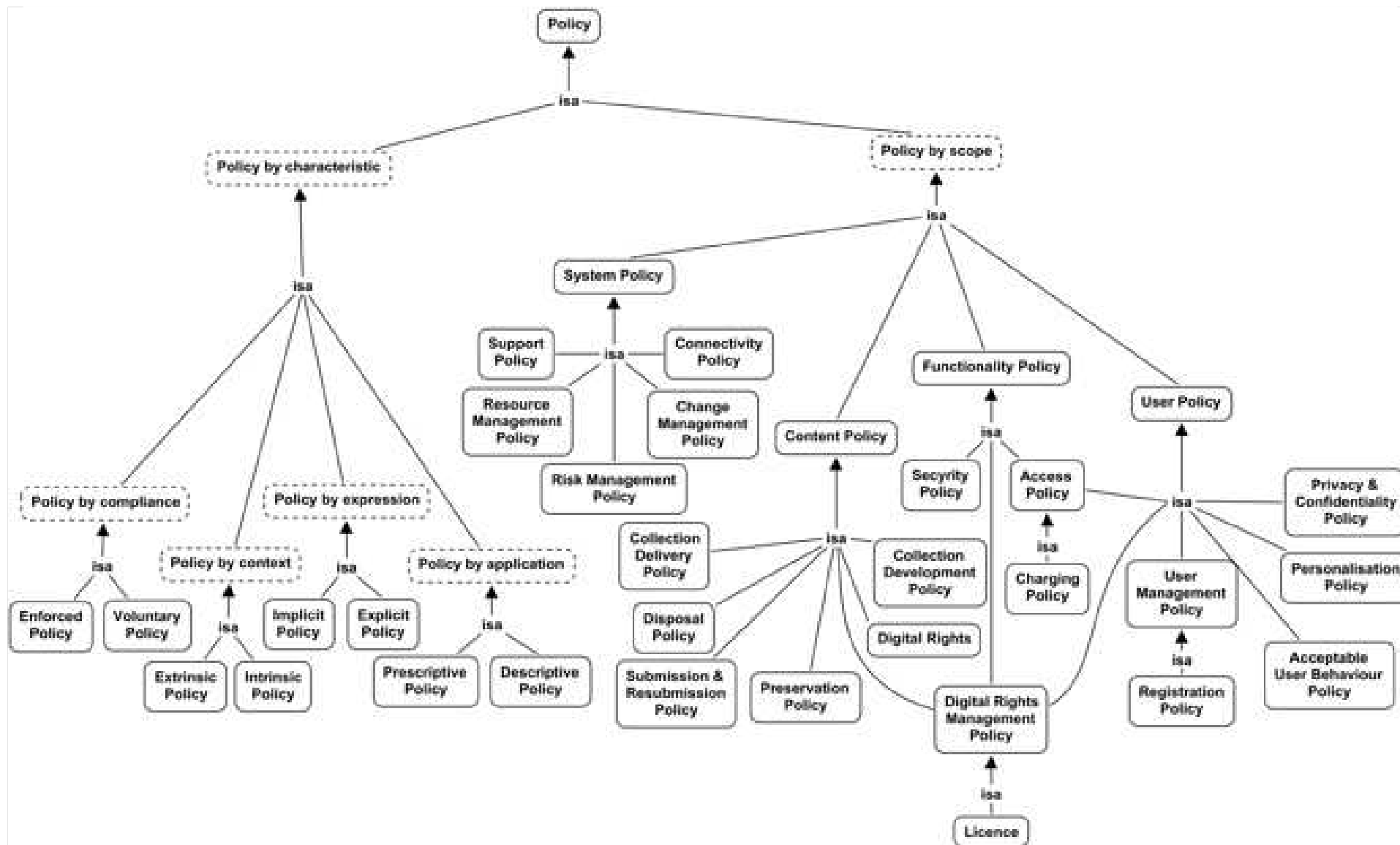
“an **organisation**, which might be virtual, that comprehensively **collects, manages and preserves** for the long term rich digital content, and offers to its users communities specialised functionality on that content, of measurable quality and according to **codified policies**” (**DELOS DIGITAL LIBRARY REFERENCE MODEL**)

Policy in the reference model

- Policy is **expressed** by an Information Object
- Policy can be **regulated** by a Resource
- Policy can be **influenced** by Quality Parameters
- Policies can be compound
- Policies can have scope: User, Content, Functionality, System

Policy outside the DL

- Some policies are **intrinsic**: decided by the DL
- Some policies are **extrinsic**: imposed from outside
 - Wider organisational policy
 - Laws
 - Regulations
 - Custom



Interoperability

Standard definitions

- **IEEE (1991):** *the ability of two or more systems or components to exchange **information** and to use the information that has been exchanged*
- **ISO/IEC 2382-2001:** *the capability to communicate, execute programs, or transfer data among various functional units in a manner that requires minimal knowledge of the unique characteristics of those units → **contextualisation?***
- **EIF 2.0 (2008):** *the ability of disparate and diverse organisations to interact towards mutually beneficial and agreed **common goals**, involving the **sharing of information and knowledge between the organizations via the business processes** they support, by means of the exchange of data between their respective information and communication technology (ICT) systems*

Interoperability Framework 2.0

EC 2008

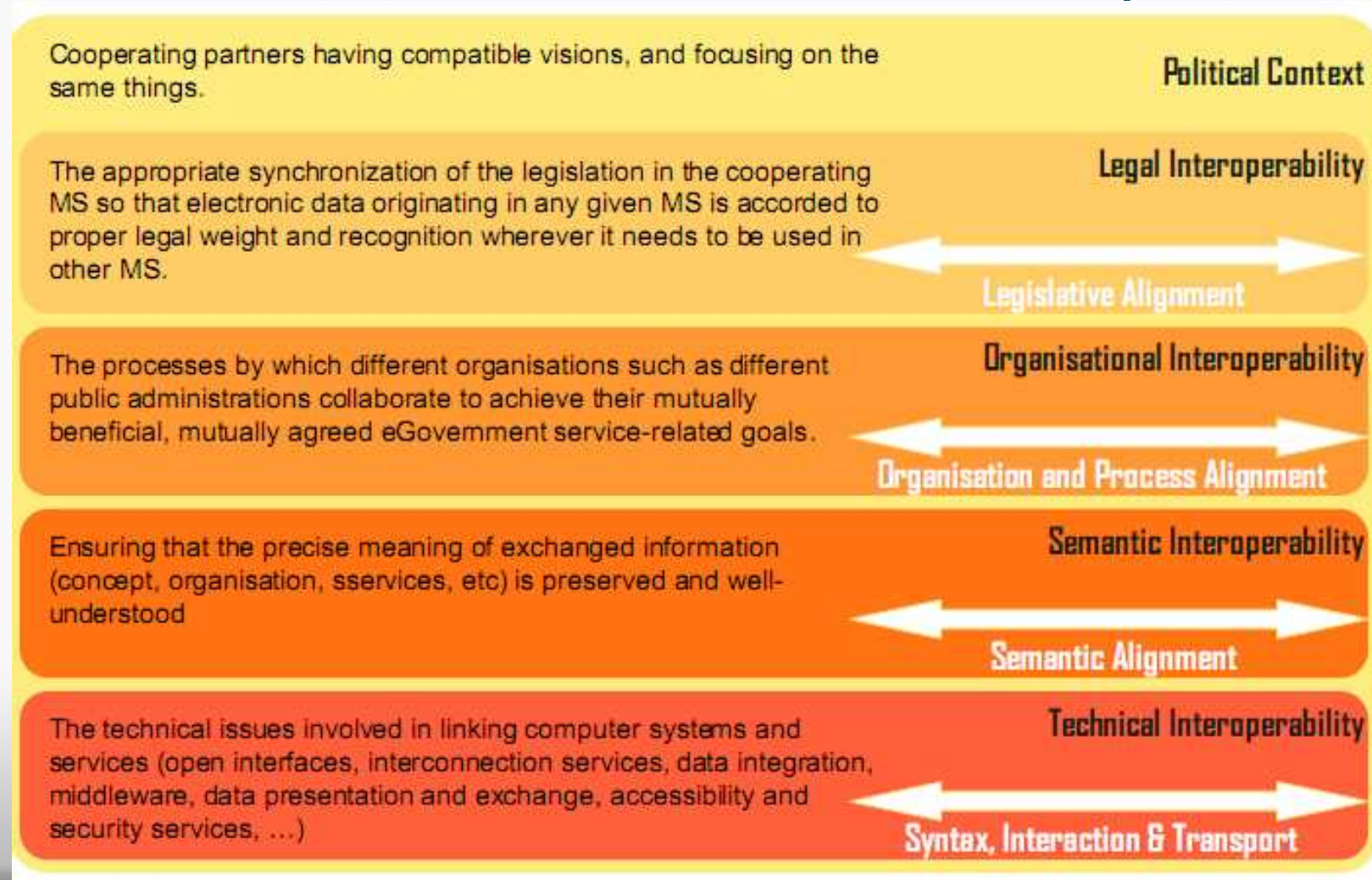
An **Interoperability Framework** describes the way in which organisations have agreed, or should agree, to interact with each other, and how standards should be used. In other words, it provides **policies** and **guidelines** that form the basis for selection of standards

Interoperability levels

The EIF approach:

- Technical: standards for presenting, collecting, exchanging, processing, transporting data
- Semantic: ensuring that transported data preserves meaning
- Organisational: organising processes and structures to enable technical and semantic interoperability

European Interoperability Framework 2.0. EC, 2008



Basics of policy interoperability

- Our policies should speak about the same things
- They should speak about them in comparable ways
- We must be able to reconcile permissions and prohibitions
- We must be able to identify appropriate external as well as internal policies

Policy WG Participants

Scientific
leader



Kevin Ashley,
DCC



Seamus Ross,
UoT

Coordinator
and co-
scientific
leader



Perla Innocenti,
HATII at UG



Hans Pfeiffenberger,
AWI



John Faundeen,
USGS



Antonella De Robbio,
UniPd



Mackenzie Smith,
MIT Libraries



*Steve Knight,
NLNZ

Policy WG public wikipage:

https://workinggroups.wiki.dlorg.eu/index.php/Policy_Working_Group

DL.org Policy WG approach

- State of the Art investigation
- Survey of real life DLs
- Suggestion of considering three interoperability layers (organisational, semantic, technical)
- First set of suggestion for the DELOS DL RM
- Suggestions towards a Policy Interoperability Framework in the DL.org Cookbook
- A checklist for policy?

Identified Policy Interoperability Issues

Concept definition	Underpinning every digital library , there is an organisation governed by an organisational policy framework, that makes the digital library viable . The policy domain is a meta-domain , situated both outside the DL and any technologies used to deliver it, and within the DL
Interoperability level	Policy permeates the digital library from conceptualisation through to operation and needs to be so represented at these various levels https://workinggroups.wiki.dlorg.eu/index.php/Definition_of_Policy_and_Policy_Interoperability
State of the art	Unexplored territory at global organisational (rather than only technical) level & interdisciplinary research
Policy representation	Lack of policy formalisation and representation methods in current DLs https://workinggroups.wiki.dlorg.eu/index.php/Policy_enforcement
Time dimension	Handling policy drift over time

Identified Policy Interoperability approaches

Concept definition	Policy Interoperability defined as Business Level Interoperability
Interoperability level	At high (organisational) level , then instantiated at process level - whether those processes are being handled by human or machine
Policy representation	<ul style="list-style-type: none"> ▪ PLEDGE categorization ▪ Analysis of languages: AIR Policy Language iRODS rules, SWRL, Turtle RDF Triples, REVERSE Policy Language, OWL, KAoS, WSPF-WS, WSPF, WSPL, XACML, Rei
Time dimension & Policy Assessment	<ul style="list-style-type: none"> ▪ Policy user scenarios ▪ Survey of current targeted DLs policies for interoperability ▪ SHAMAN Assessment Framework

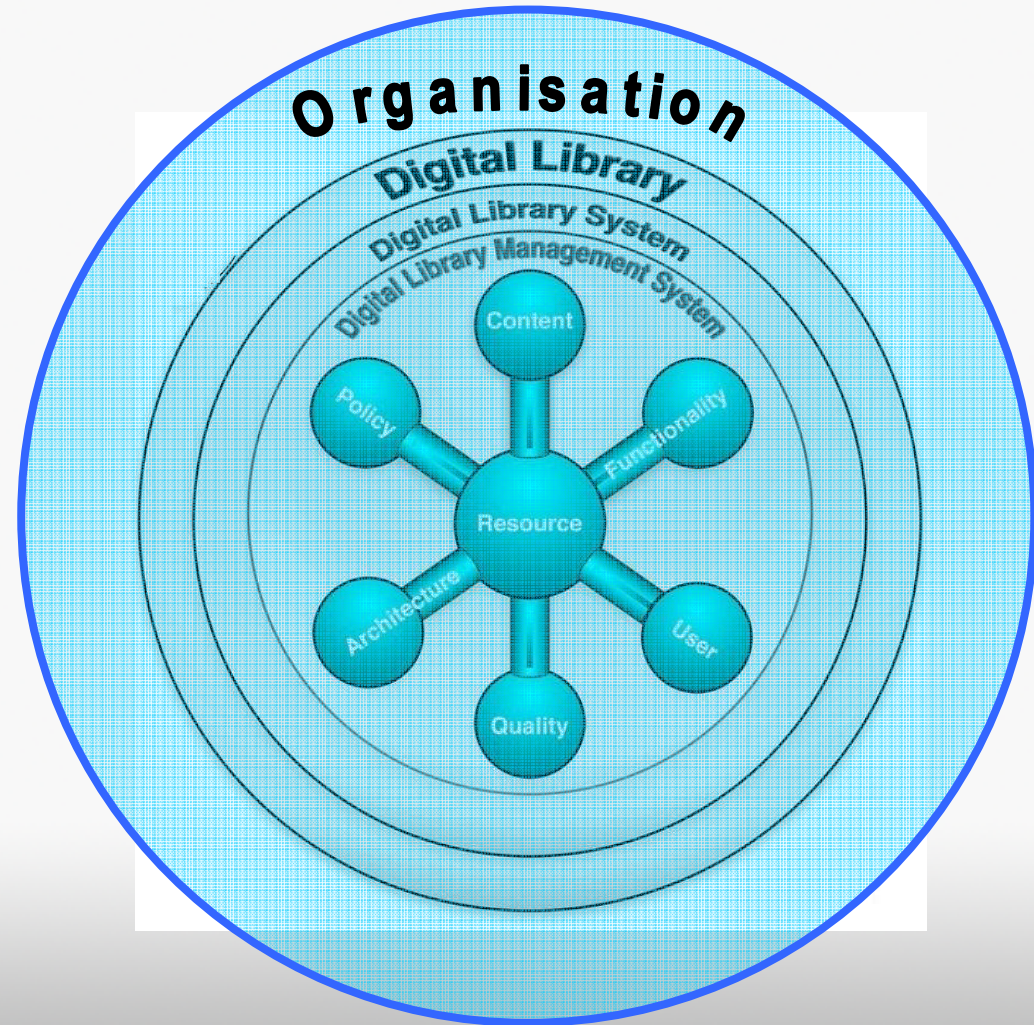


Towards a **Policy Interoperability Framework**

https://workinggroups.wiki.dlorg.eu/index.php/Policy_Interoperability_Approaches_Summary

Shared Quality/Policy WGs Organisational Issues

A DL may operate within an **organisation** which defines over-arching policies (not necessarily specific to Digital Libraries) which affect **interoperability**

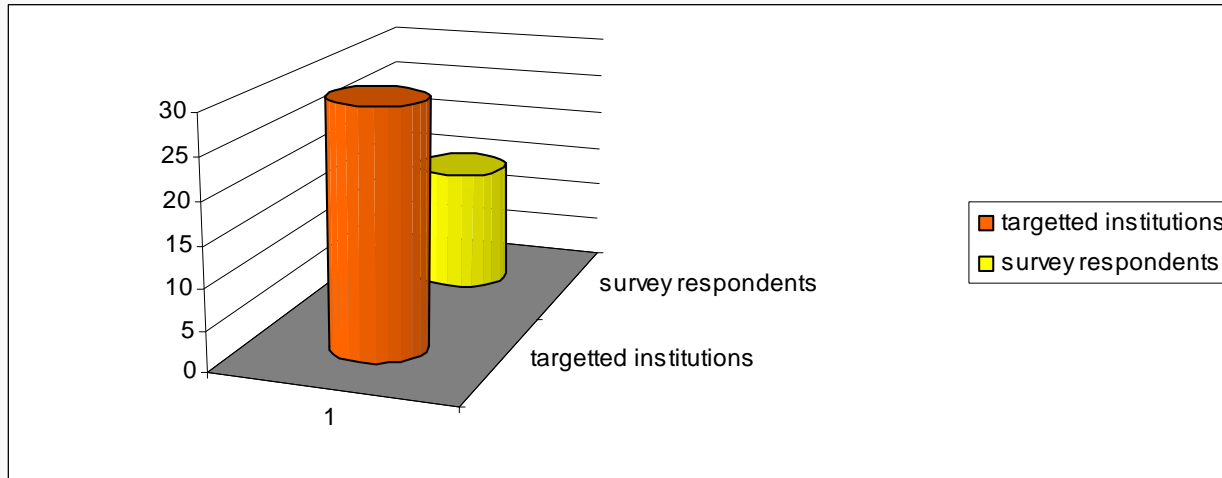


DL.org Policy Interoperability Survey

The survey investigated:

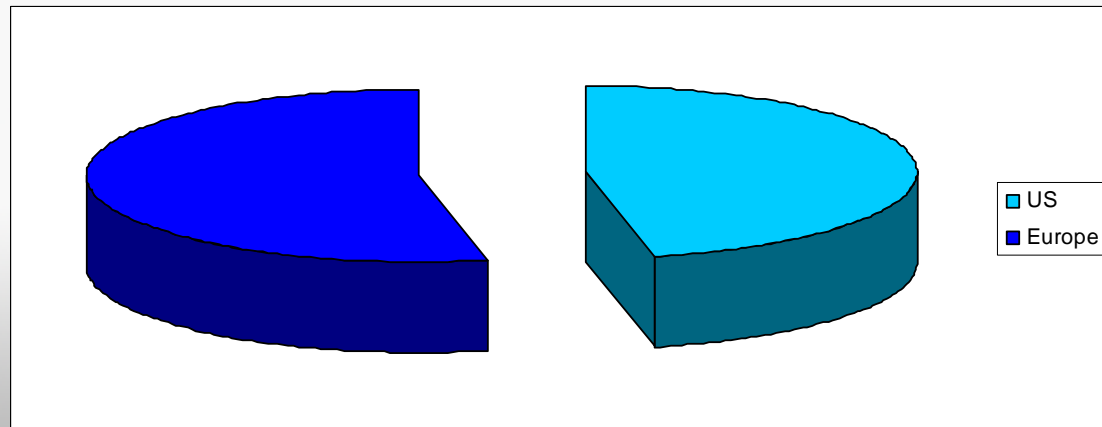
- Any policies, strategies, frameworks, programs, plans, or statements that have been prepared to guide how to **develop and exploit aspects of their digital library/digital repository's information management.**
- How these policies, strategies, frameworks, programs, plans, or statements **affects or are affected by interoperability.**

Policy Interoperability Survey: targeted institutions



- large/medium DL, repository, archive
- public and commercial sector
- at least some policies in place

- US
- Europe (UK, Italy, Greece, European initiatives)



Policy Interoperability Survey: first set of organisations

- ACM - Digital Library
- California Digital Library (CDL) - Calisphere
- DANS
- DRIVER
- ELis
- Europeana
- ITHAKA: JSTOR, PORTICO
- Liber Liber
- NARA
- Nemertes
- National Science Digital Library (NSDL)
- Padua@Research
- UK Data Archive
- University of Chicago Digital Repository
- USGS Digital Library

Policy Interoperability Survey: sections

1. **scoping** the digital library and organisation staff involved in the digital library policies
2. **questions focused on policies** for:

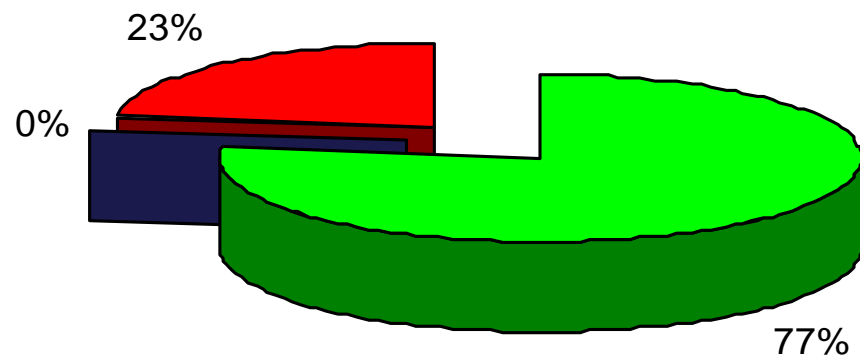
Access
Preservation
Metadata
Networks

Collection
development
Intellectual property
Authentication
Service level
agreements

Policy Interoperability Survey: staff responsible for policies

- *'The Management team (Director / Deputy Director, and three coordinators in the field of (a) data archiving (b) data infrastructure and (c) systems development. Next to that small (temporarily) workgroups on specific topics provide information to formulate policies on different topics.'*
- *'The whole consortium!'*
- *'Archive Services Product Manager, Director of Data Technology, Content Management Systems Architect, Vice President for Content Management, Senior Research Developer'*
- *'The Policy Group including all associate directors and the director with support from a Strategy Group which includes representatives of all sections and services.'*

Policy Interoperability Survey: Is policy useful for your DL?



- We considered policy for improving our interoperability, and find it helpful/necessary
- We considered policy for improving our interoperability, but did not find it helpful/necessary
- We did not consider policy as a mean to improve our interoperability within our organisation or with other organisations

What did we learn so far

- Policy formalisation & representation
- ‘hot areas’ needed for policy interoperability (e.g. Machine-encoding, Interoperability assessment)
- What DLs are currently using and what might be used for policy interoperability

Some thoughts on improving policy interoperability

- Rather than ‘solutions’, for policy interoperability it would be more appropriate to talk about a **‘future’ state**
- Some **active areas** for policy interoperability are e.g. related to access, authentication and licensing policies
- **Making policies machine-readable** would make them easier to manage.
- Need to focus on **human-machine interaction**

Quality



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

ISO 8402-1994 *the totality of characteristics of an entity that bear on its ability to satisfy stated and implied needs*

ISO 9000-2005 *The degree to which a set of inherent characteristics fulfils requirements (needs or expectations stated/implied/obligatory)*

DELOS RM 2008 *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

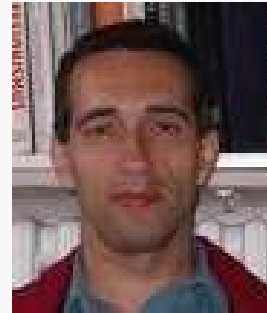
The Quality WG



Nicola Ferro
University of Padua
WG Scientific Chair



Sarah Higgins
Aberystwyth University (UK)



Sarantos Kapidakis
Ionian University



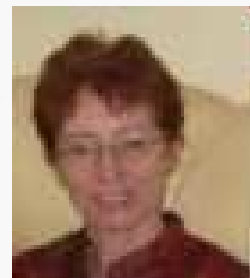
Wolfram Horstmann
University of Bielefeld
DRIVER



Seamus Ross
University of Toronto



René Van Horik
Data Archiving and
Networked Services
(NL)



Genevieve Clavel
Swiss National Library



Giuseppina Vullo
HATII, University of
Glasgow

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 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**, eg. how to link heterogeneous and dispersed resources keeping reliability of services, data precision, homogeneous experience for the end user

Quality WG motivating interoperability scenario

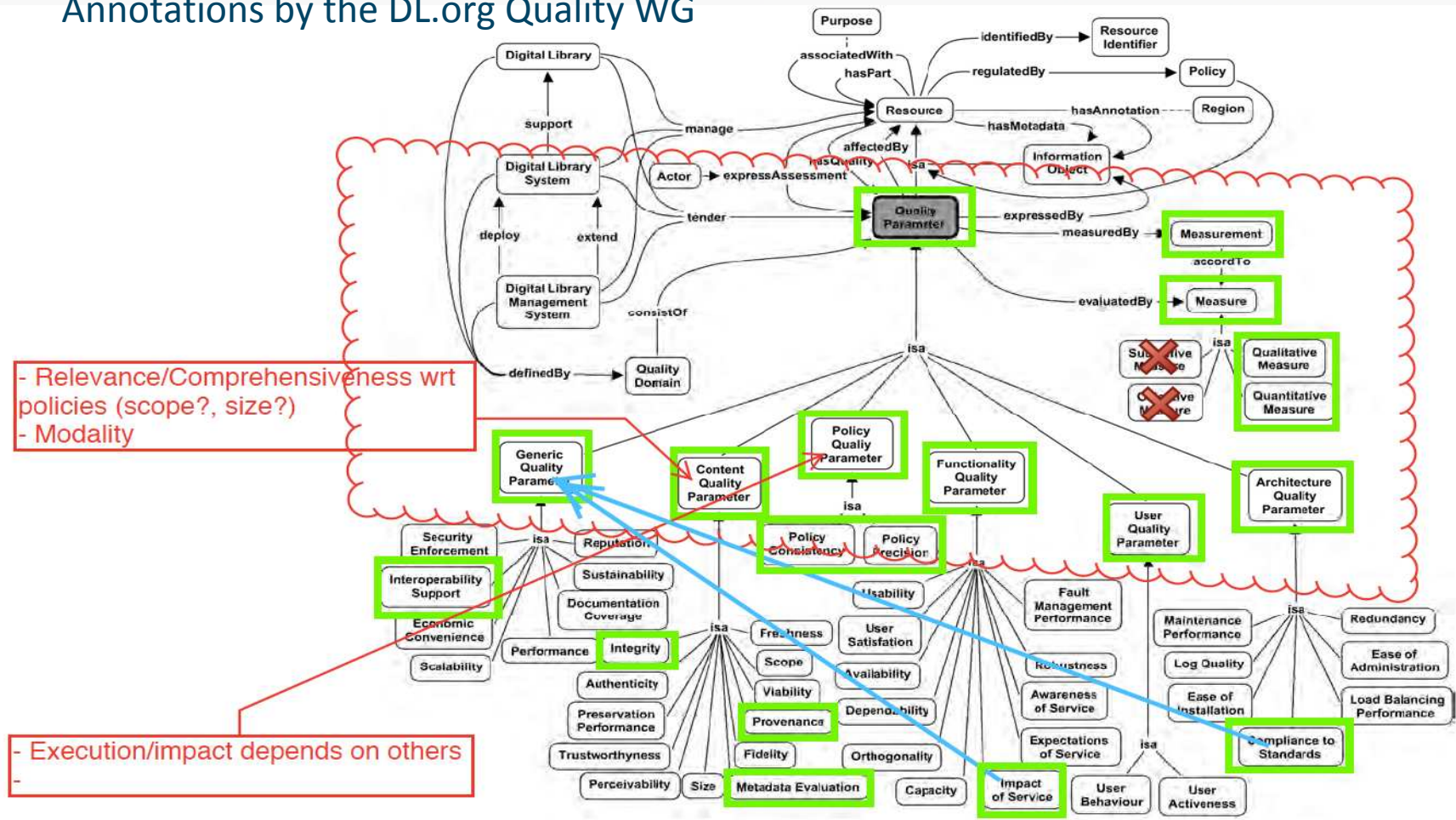
Our motivating scenario: consider that representatives of two (or more) DLs have a round table to negotiate a service level agreement (SLA) defining their interoperability requirements and for this establish a quality threshold that each individual DL has to meet or exceed; “Quality” would provide transparent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

Annotating the Quality Concept Map

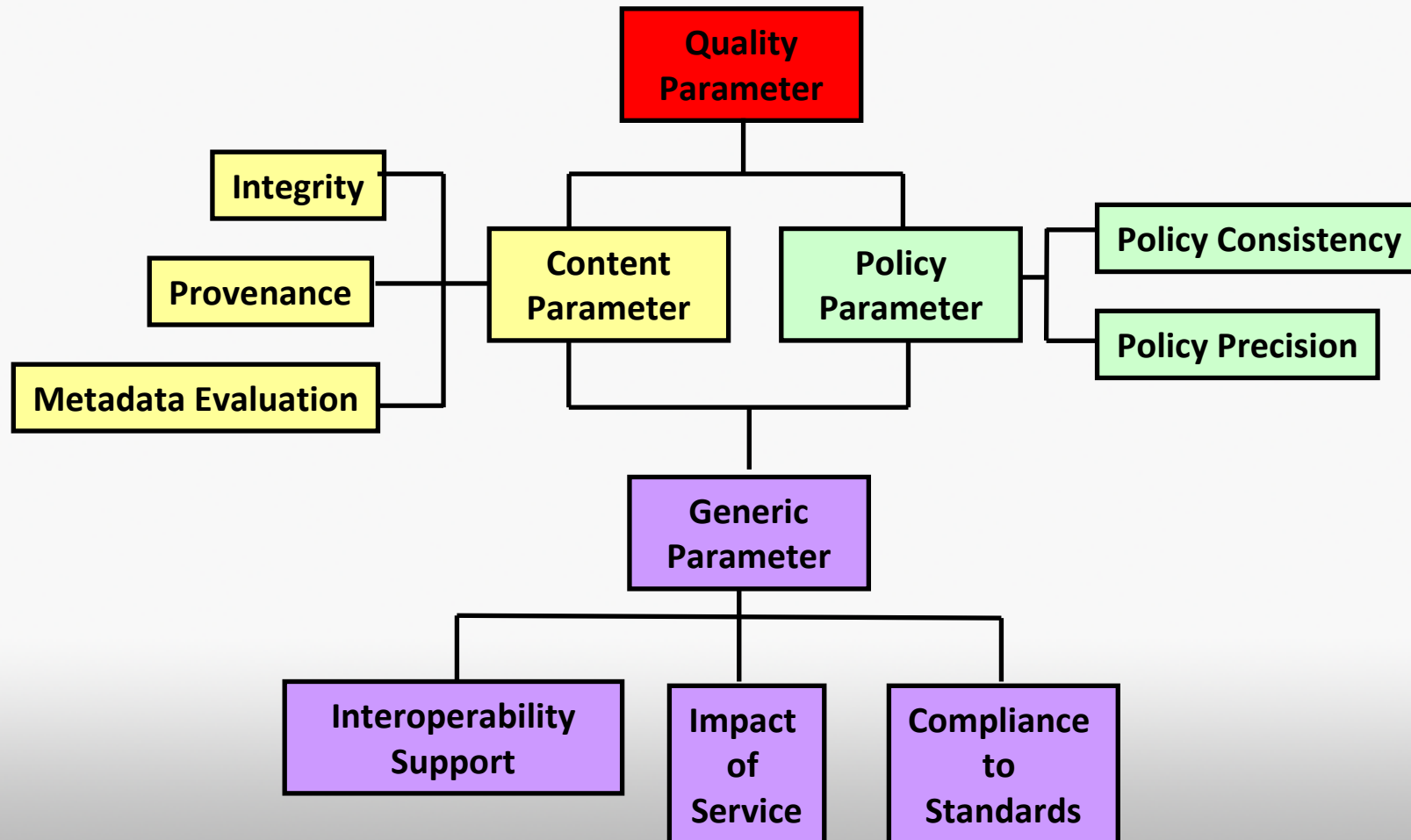
Map

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

DLs Information Objects:

- Consistency of actions, values, methods, measures, principles, expectations and outcomes
- Completeness, accuracy

Related to:

- Metadata integrity
- Policy consistency
- Regular content update
- Accurate format migrations

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: 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

- **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

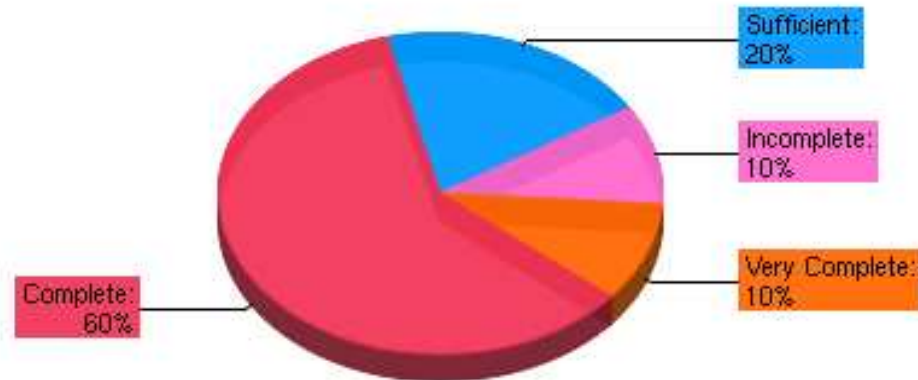
Do you use any validation tools to check

- **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

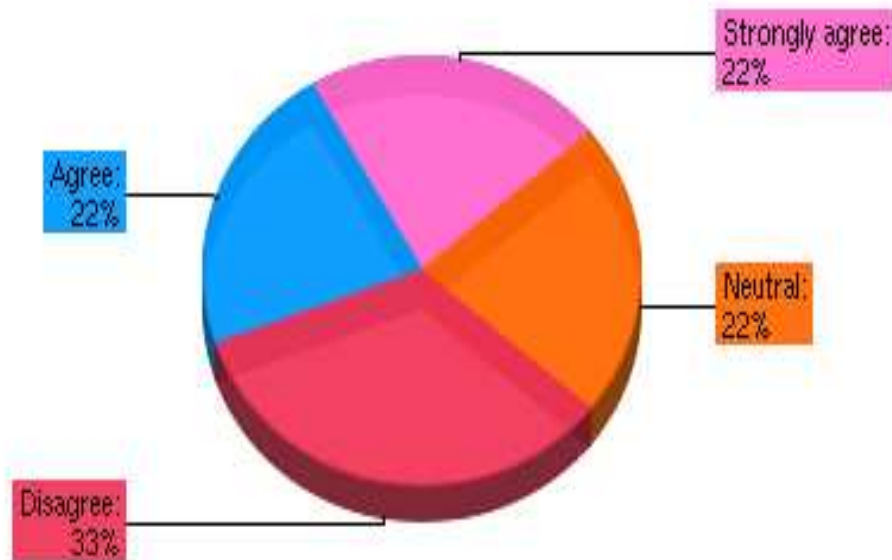
In your opinion, what is the single greatest barrier to metadata creation?

- 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?

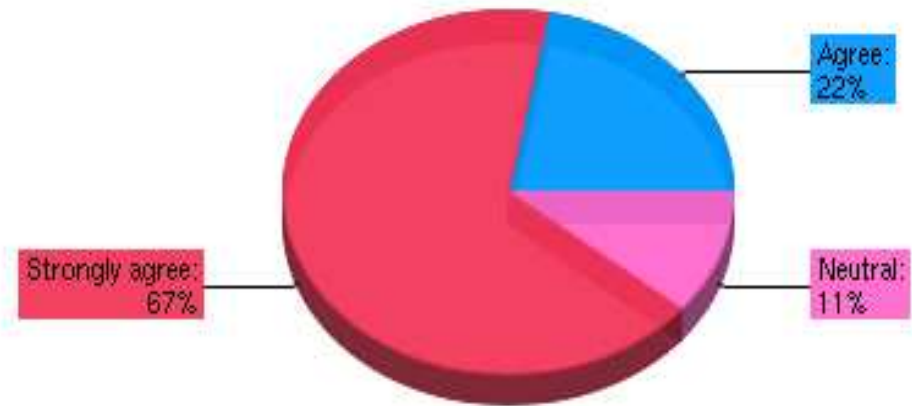
Successful interoperability is largely a technical issue



Quality Interoperability Survey

Quality and interoperability

Quality aspects are crucial for successful interoperability



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*
- *Containing consistent and complete metadata; valid identifiers to full-text and other material*
- *A DL that includes consistent, authoritative data within a user-centred website*
- *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*

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: contexts** and **objectives**
- Lack of formalised and well-analysed **policies**
- Need to be **supported**

Work in progress

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

Some hands-on tips from the DL.org AS...



DL.org Autumn School

Training on interoperability scenarios

- **SCENARIO 1** - *Producing summary versions of compound multimedia historical documents*
- **SCENARIO 2** - *Metadata evaluation and preservation policies in a new EU consortium of digital repositories*
- **SCENARIO 3** - *Designing an information discovery service that integrates Wikipedia, Amazon and Europeana user models*

Hands on Policy at the DL.org Autumn School

Exercise

Look at DL existing accessible policies

Analyse those policies in relation to the RM

Reflect on how those policies support or hinder interoperability

Each student group reports back at the end of the exercise

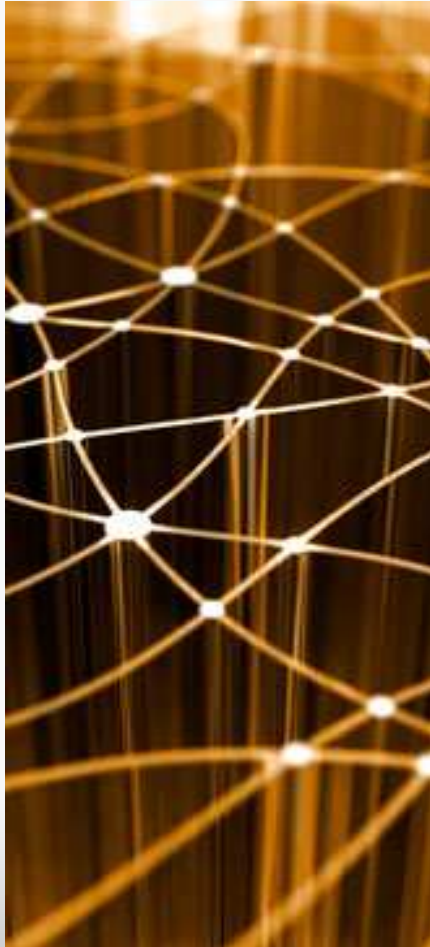
Hands on Quality at the DL.org Autumn School

Exercise

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

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

Thank you 😊



POLICY WG WIKI

https://workinggroups.wiki.dlorg.eu/index.php/policy_Working_Group

QUALITY WG WIKI

https://workinggroups.wiki.dlorg.eu/index.php/Quality_Working_Group