



The Policy & Quality Interoperability Surveys Lessons learned from the OAR community

Giuseppina Vullo *HATII at the University of Glasgow (speaker)*

Perla Innocenti *HATII at the University of Glasgow*

Seamus Ross *Faculty of Information, University of Toronto*

“Digital Library and Open Access. Interoperability strategies” Workshop

London, 4 February 2011



Outline

Policy

- Investigating **Policy interoperability**: the **DL.org Policy WG**
- The **Policy Interoperability Survey**

Quality

- Investigating **Quality interoperability**: the **DL.org Quality WG**
- The **Quality Interoperability Survey**

What is policy?



“a course or principle of action adopted or proposed by an organisation or individual” (Oxford English Dictionary)

“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 DL.org REFERENCE MODEL

DL definition

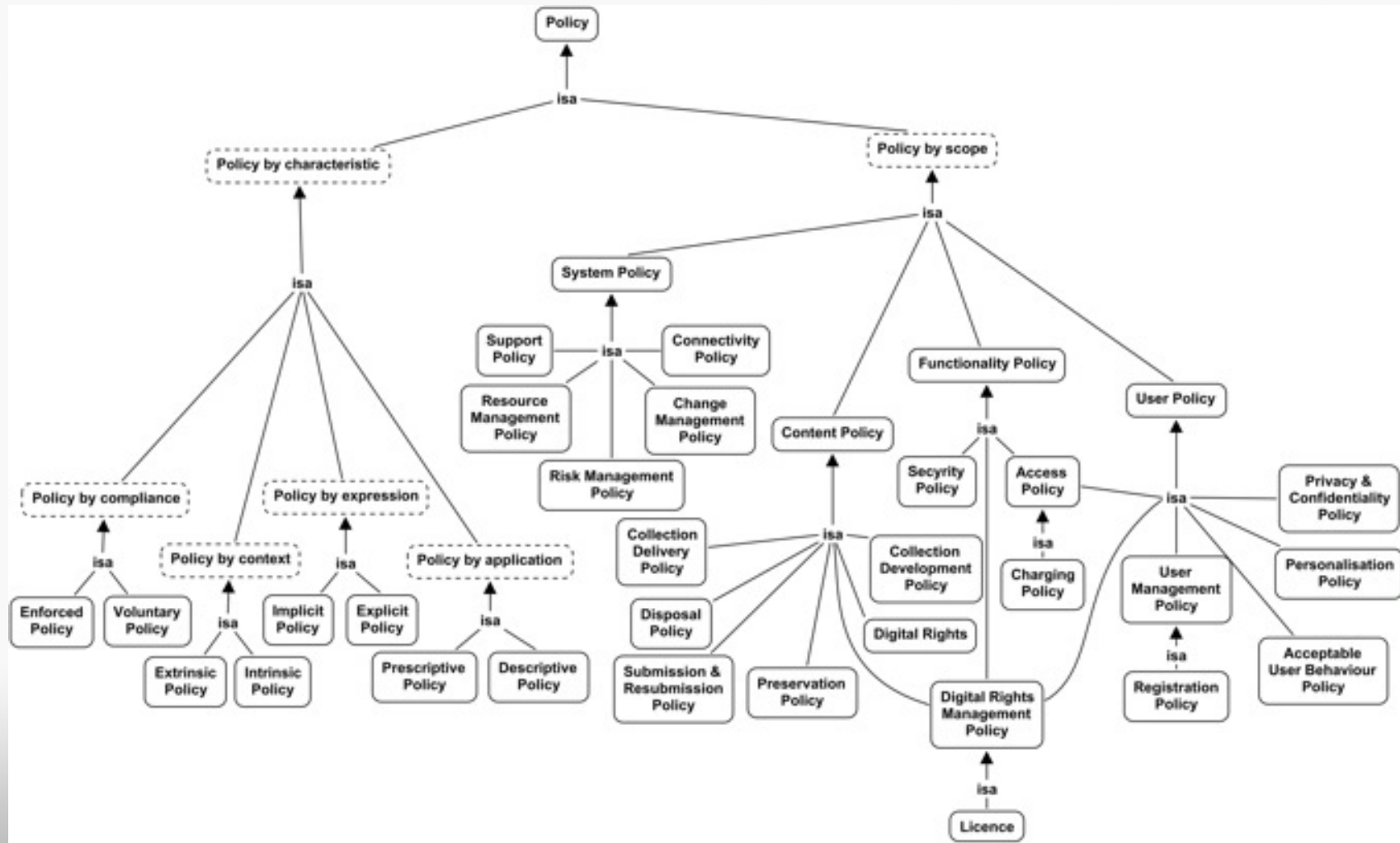
“an **organisation**, which might be virtual, that comprehensively **collects, manages and preserves** for the long term rich digital content, and offers to its users communities specialised functionality on that content, of measurable quality and according to codified policies”

Policy

“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. [...]”

Policy outside the DL

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



Interoperability Standard definitions

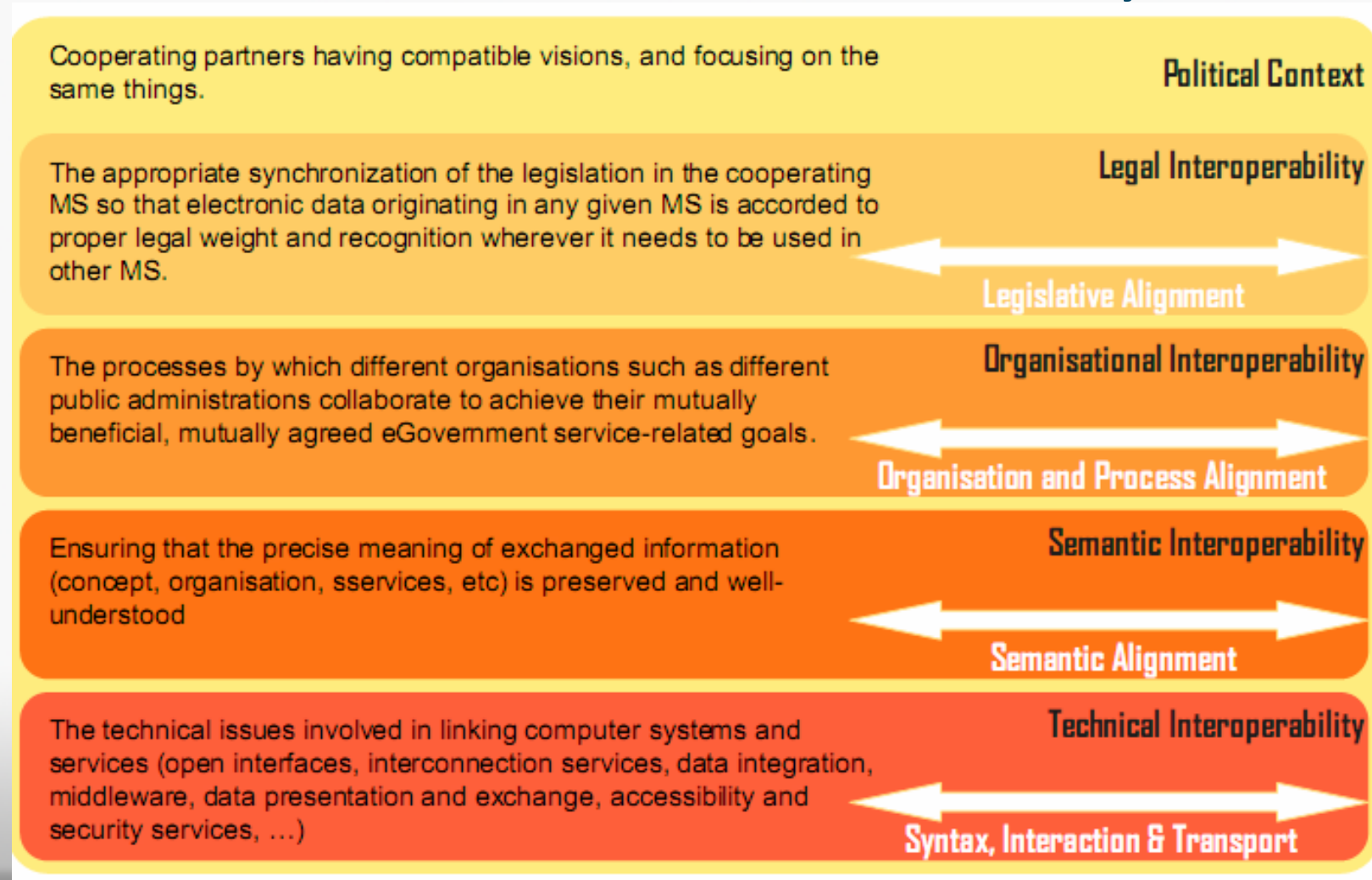
- **IEEE (1991):** *the ability of two or more systems or components to exchange **information** and to use the information that has been exchanged*
- **ISO/IEC 2382-2001:** *the capability to communicate, execute programs, or transfer data among various functional units in a manner that requires minimal knowledge of the unique characteristics of those units*

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



Interoperability levels

- **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

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

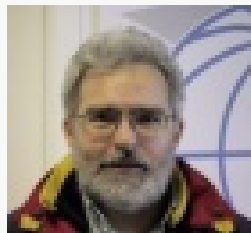


Seamus Ross,
UoT

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

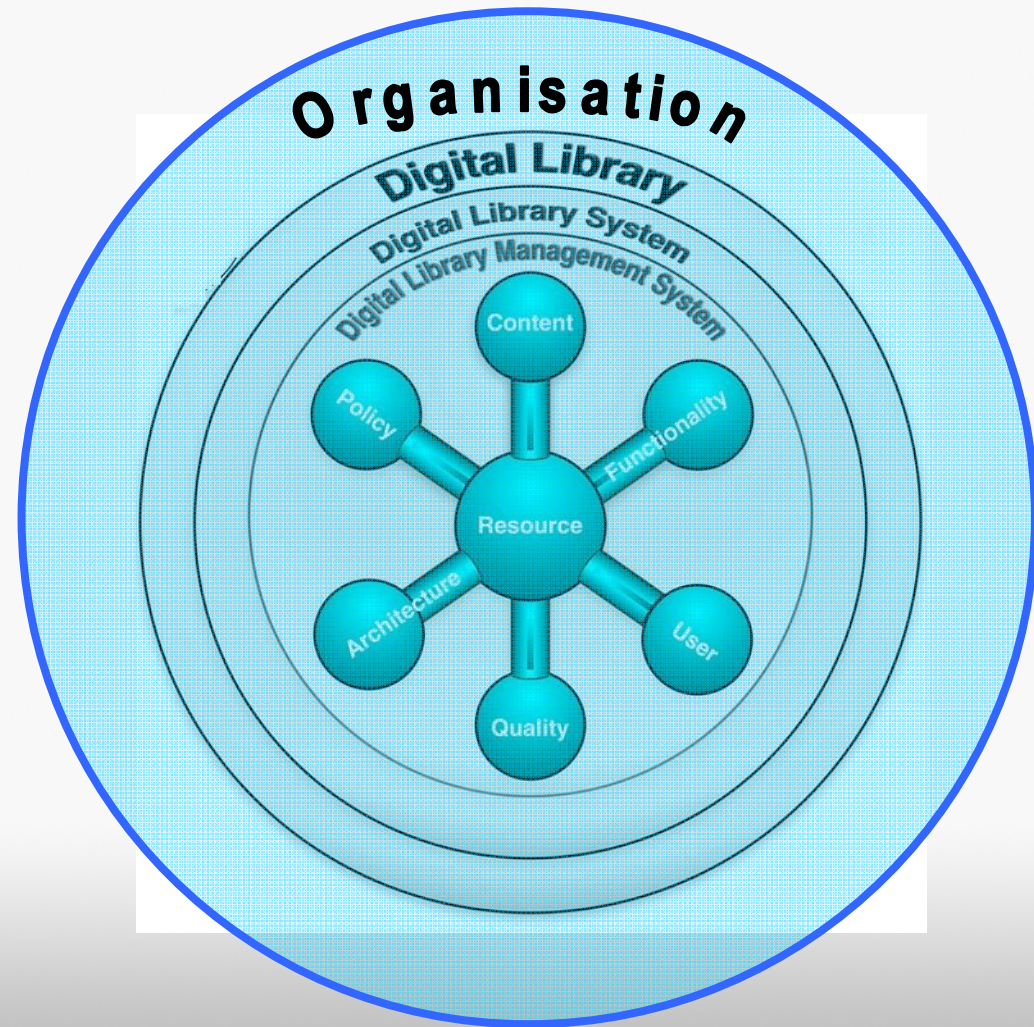
- State of the Art investigation
- Survey of real life DLs
- Suggestion of considering **three** interoperability layers (organisational, semantic, technical)
- Enhancements of the **DL.org RM** and contribution to the **DL.org Cookbook**
- First structure and set of criteria for the **DL.org Checklist**

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

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**

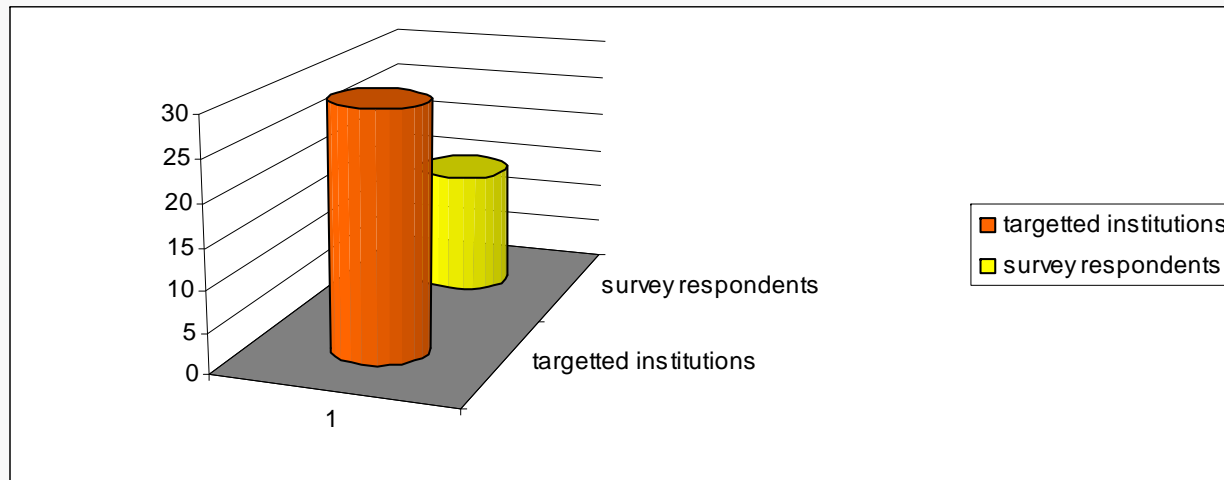


DL.org Policy Interoperability Survey

The survey investigated:

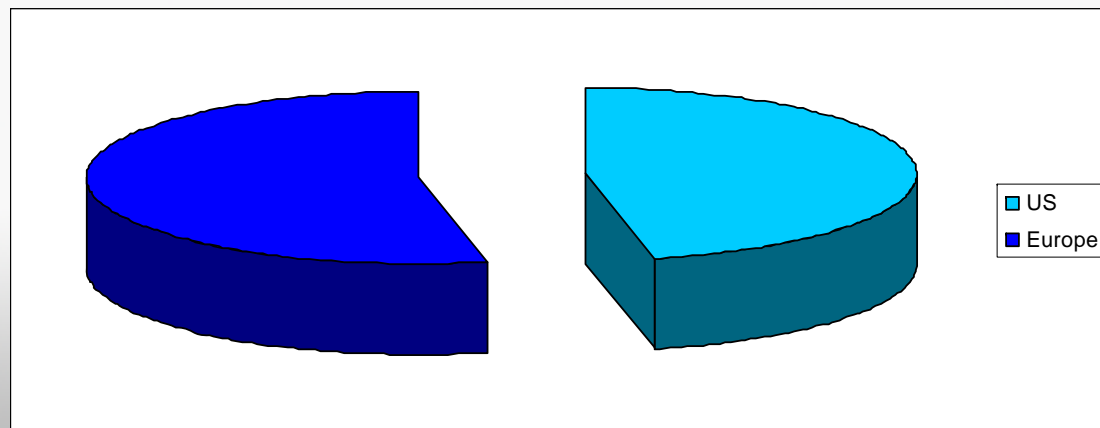
- Any policies, strategies, frameworks, programs, plans, or statements that have been prepared to guide how to **develop and exploit aspects of their digital library/digital repository's information management**
- How these policies, strategies, frameworks, programs, plans, or statements **affects or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

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

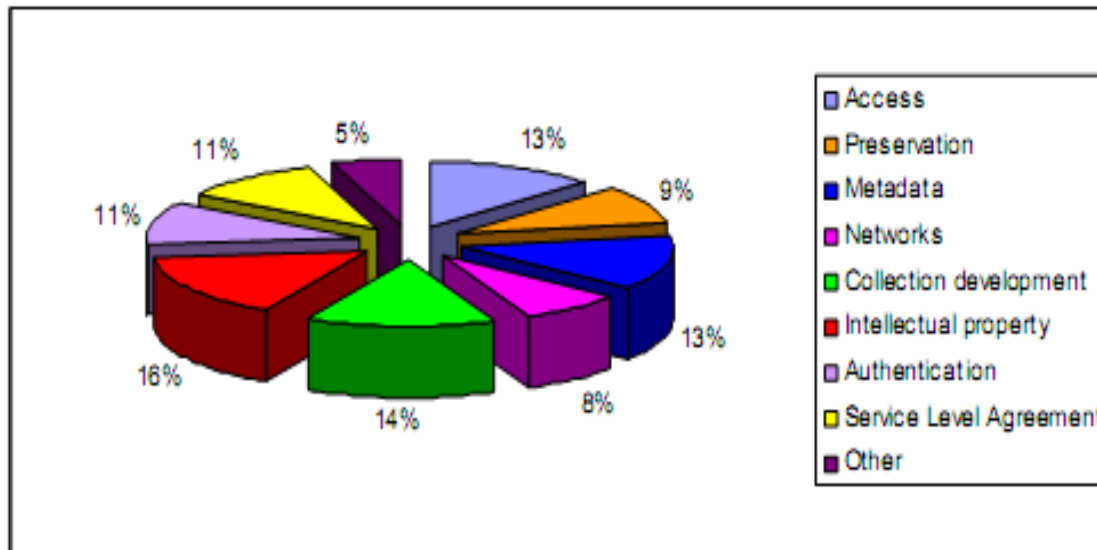
Policy Interoperability Survey: sections

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

Access
Preservation
Metadata
Networks

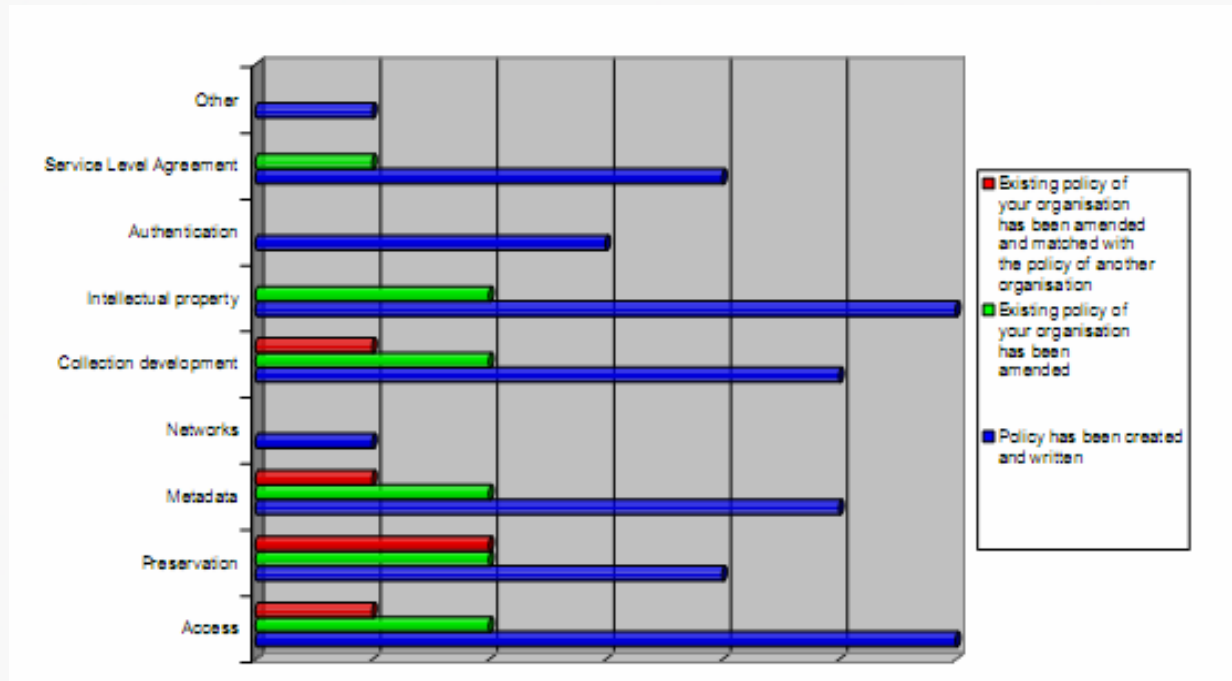
Collection
development
Intellectual property
Authentication
Service level
agreements

Existence of written policies per areas of interest



Almost all respondents indicated that their digital library/repository/archive had a written strategy or plan, either as part of a library strategic plan or as independent entity within the organisation

How policies reflect interoperability needs per areas of interest



In terms of policy exchange and reuse with other entities, only in the areas of Preservation, Access, Collection Development and Metadata the existing policies of the respondent organisations were amended and matched with the policies of other organisations

Interoperability



All respondents indicated an interest or need to interoperate with peer and smaller/bigger organisations, both in the public and private sector. But interestingly few written policies were indicated as available to regulate this interaction

What we learned so far

- **Lack** of policy formalisation & representation. Limited formal specifications are supported, e.g. for network management, security and privacy
- **'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 technical interoperability of policy is possible, but only for **very specific and technical cases** (e.g., access control via Shibboleth)

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. Research should usefully focus on human-machine interaction, e.g. how licensing policy interoperability might be achieved automatically in the near future
- **Making policies machine-readable** would make them easier to manage

Quality



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)*

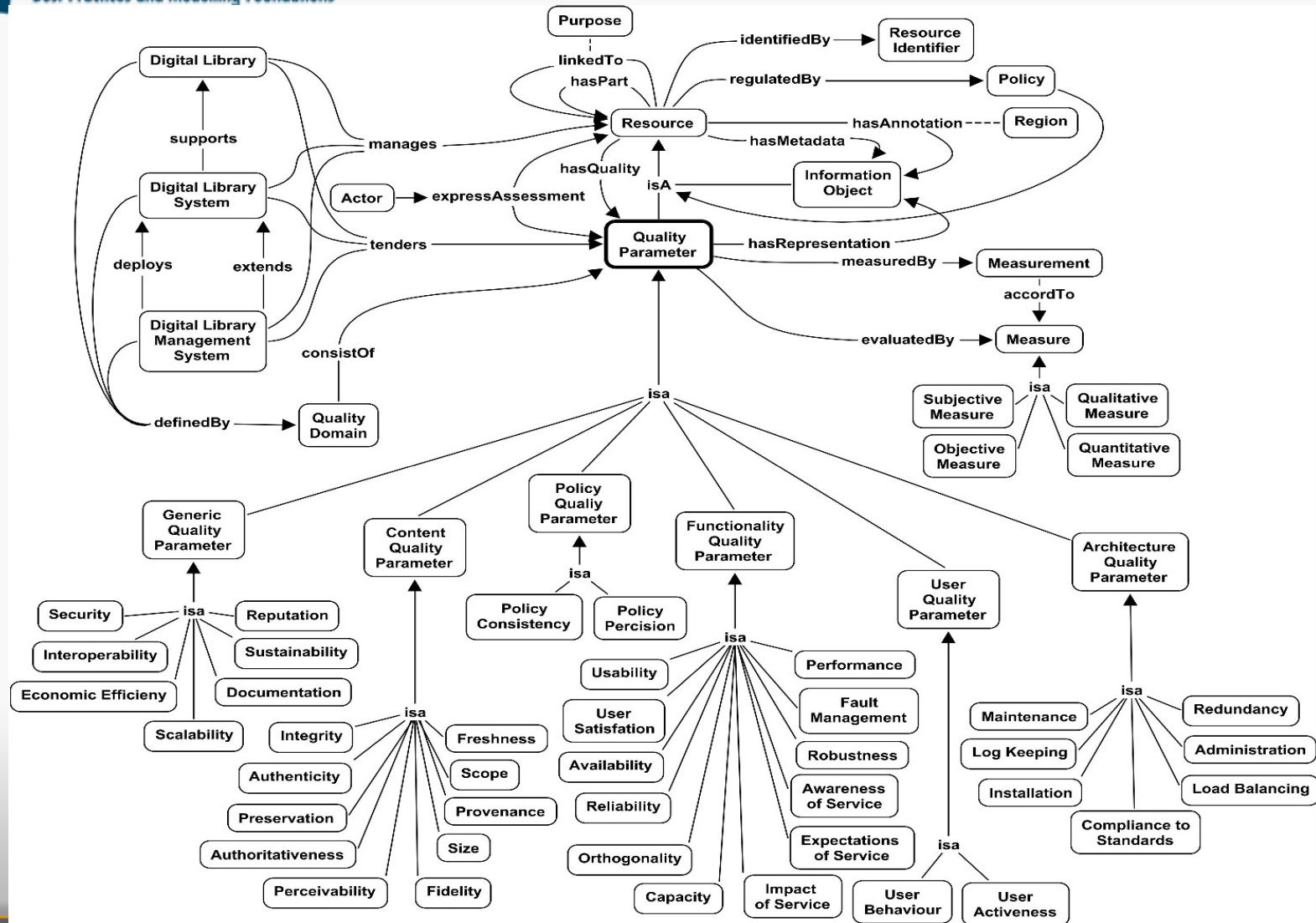
DELOS RM 2008 *parameters that can be used to characterise and evaluate the content and behaviour of a DL. Quality can be associated not only with each class of content or functionality but also with specific information objects or services*

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**

The Quality Domain



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

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

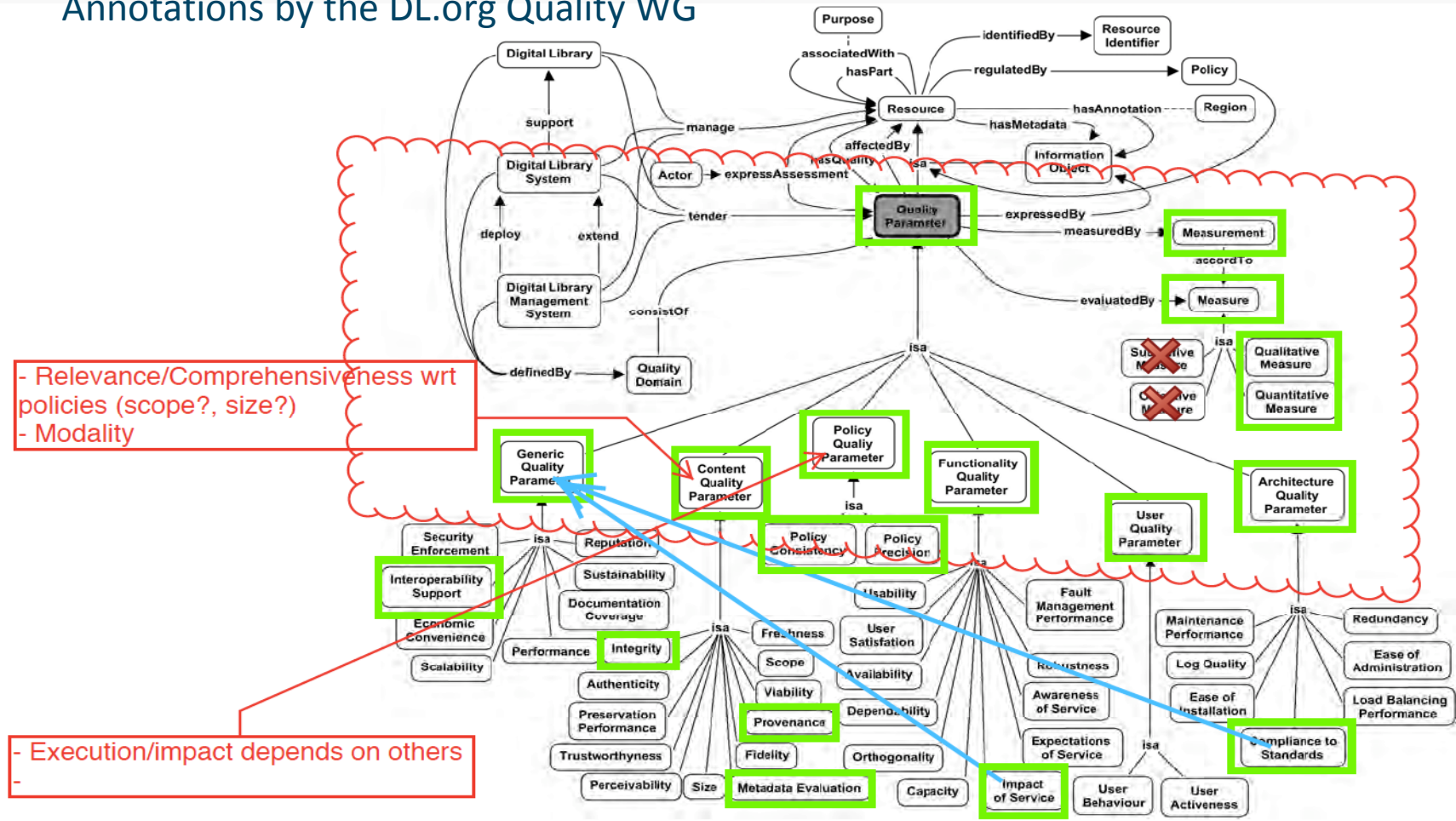
Our motivating scenario: consider that representatives of two (or more) DLs have a round table to negotiate a service level agreement (SLA) defining their interoperability requirements and for this establish a quality threshold that each individual DL has to meet or exceed; “Quality” would provide transparent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

Annotating the Quality Concept Map

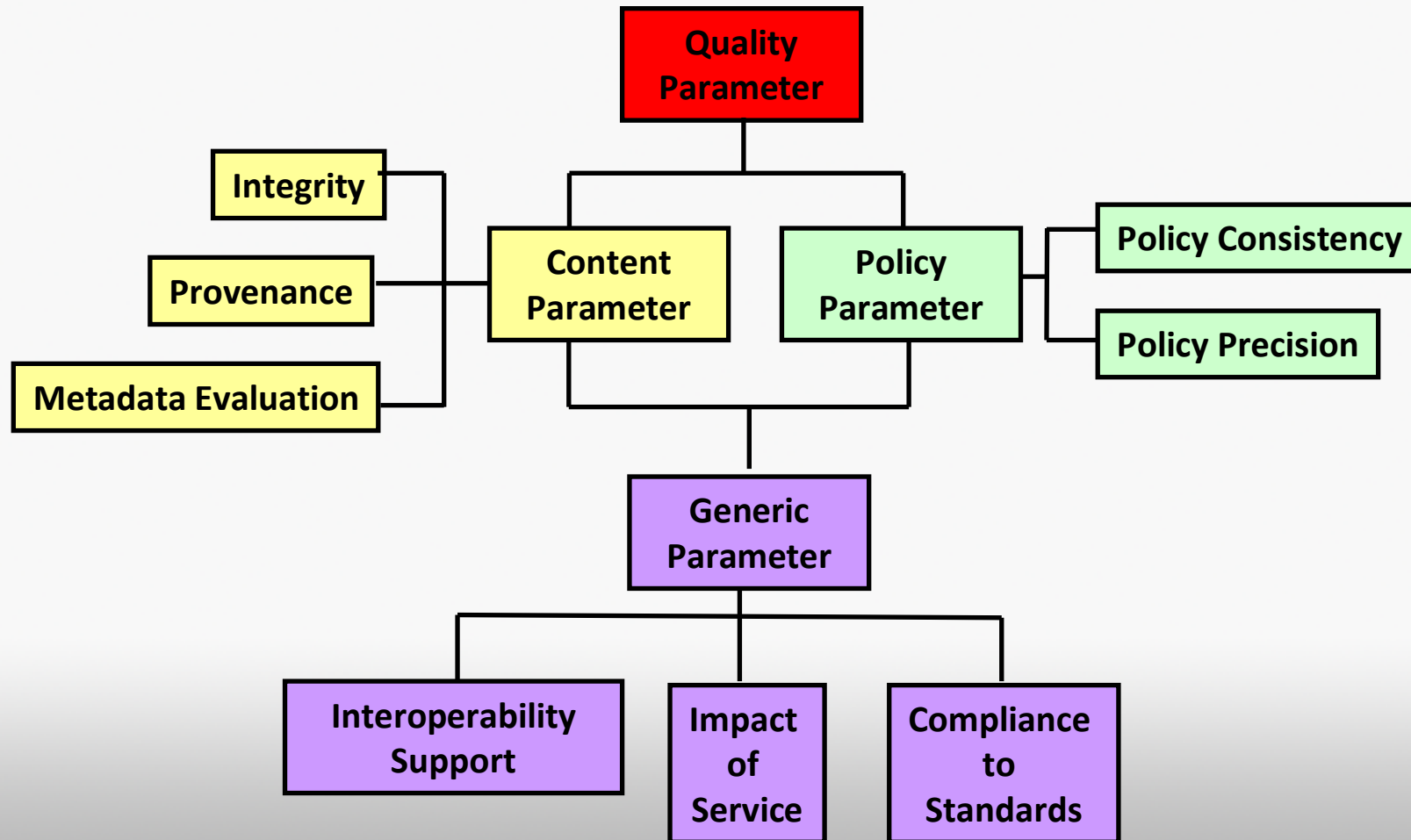
Map

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
- **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

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

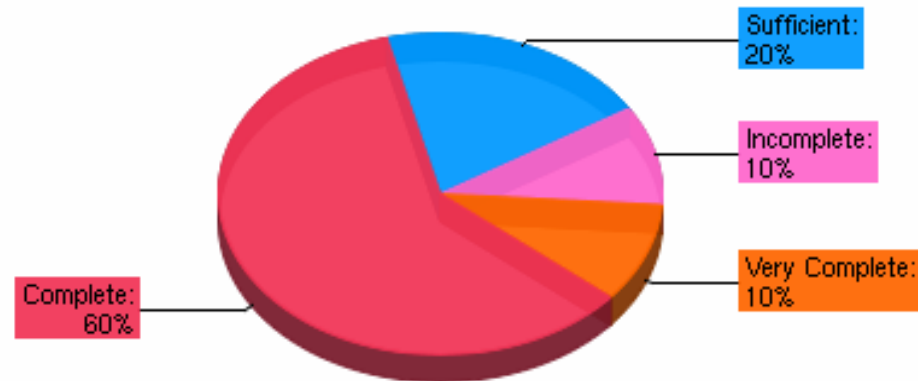
Do you use any validation tools to check

- **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

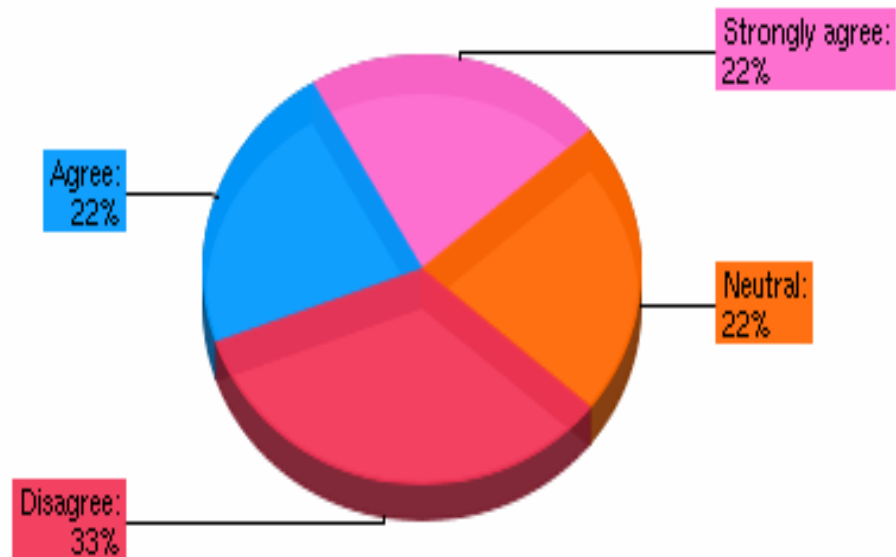
In your opinion, what is the single greatest barrier to metadata creation?

- 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?

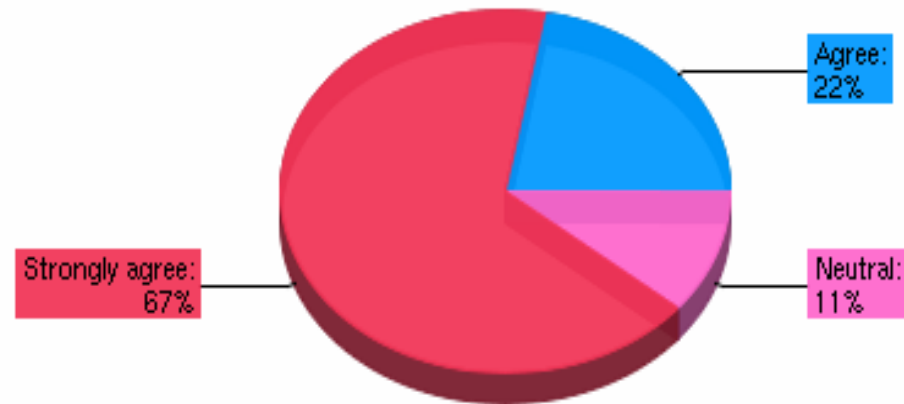
Successful interoperability is largely a technical issue



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

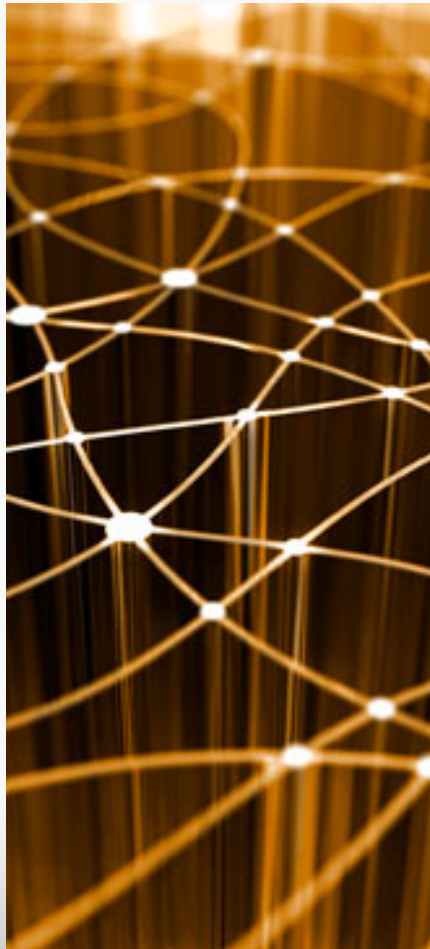
Some preliminary evidence

- **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

- Publication of the **survey's results**
- Identification and selection of **best practices** and **recommendations** for the Cookbook
- Enhancing the **Quality domain** in the RM
- Contributing to the **DL.org Checklist**

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