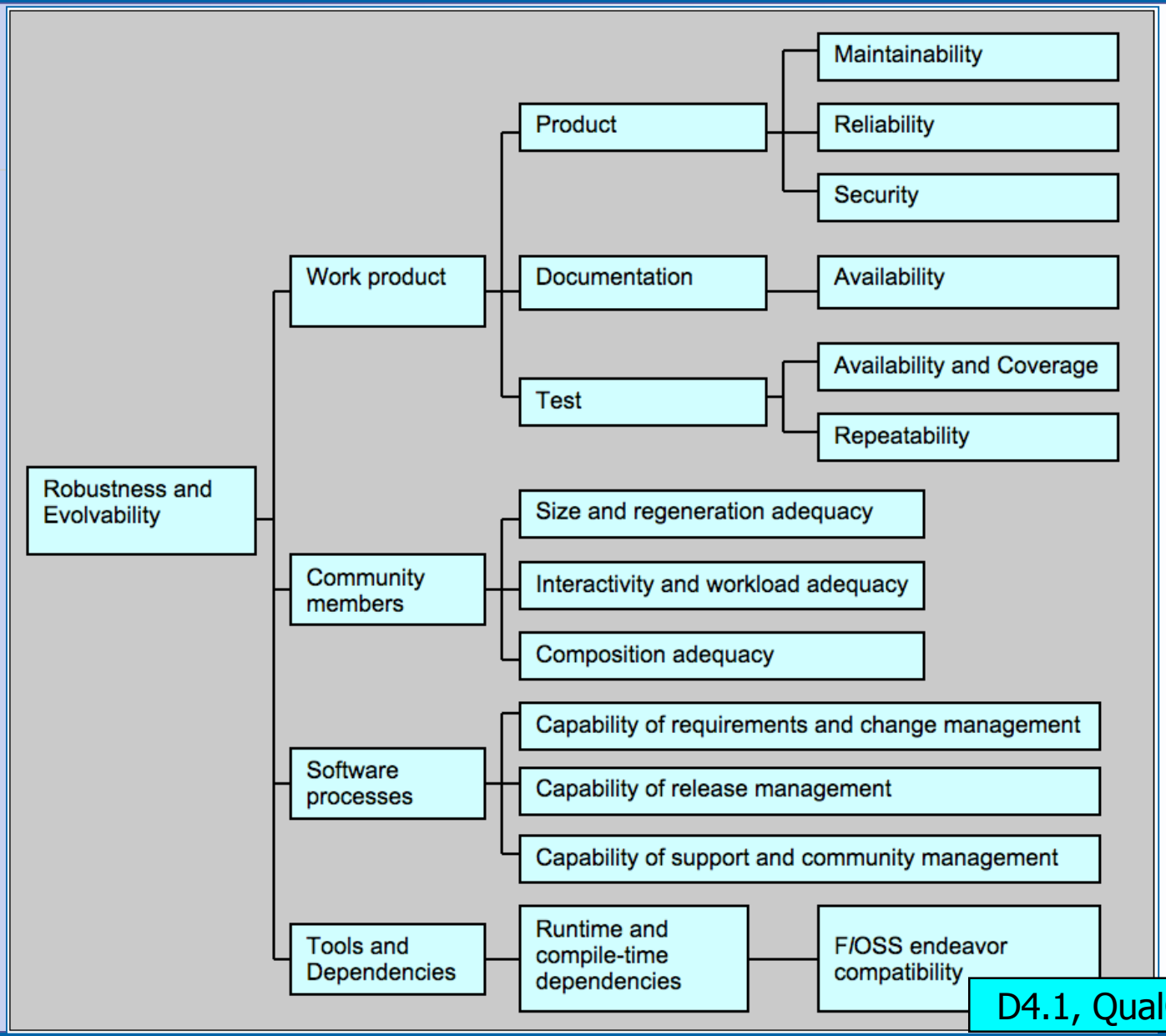
Standard QualOSS Assessment Method

Measuring Documentation Availability

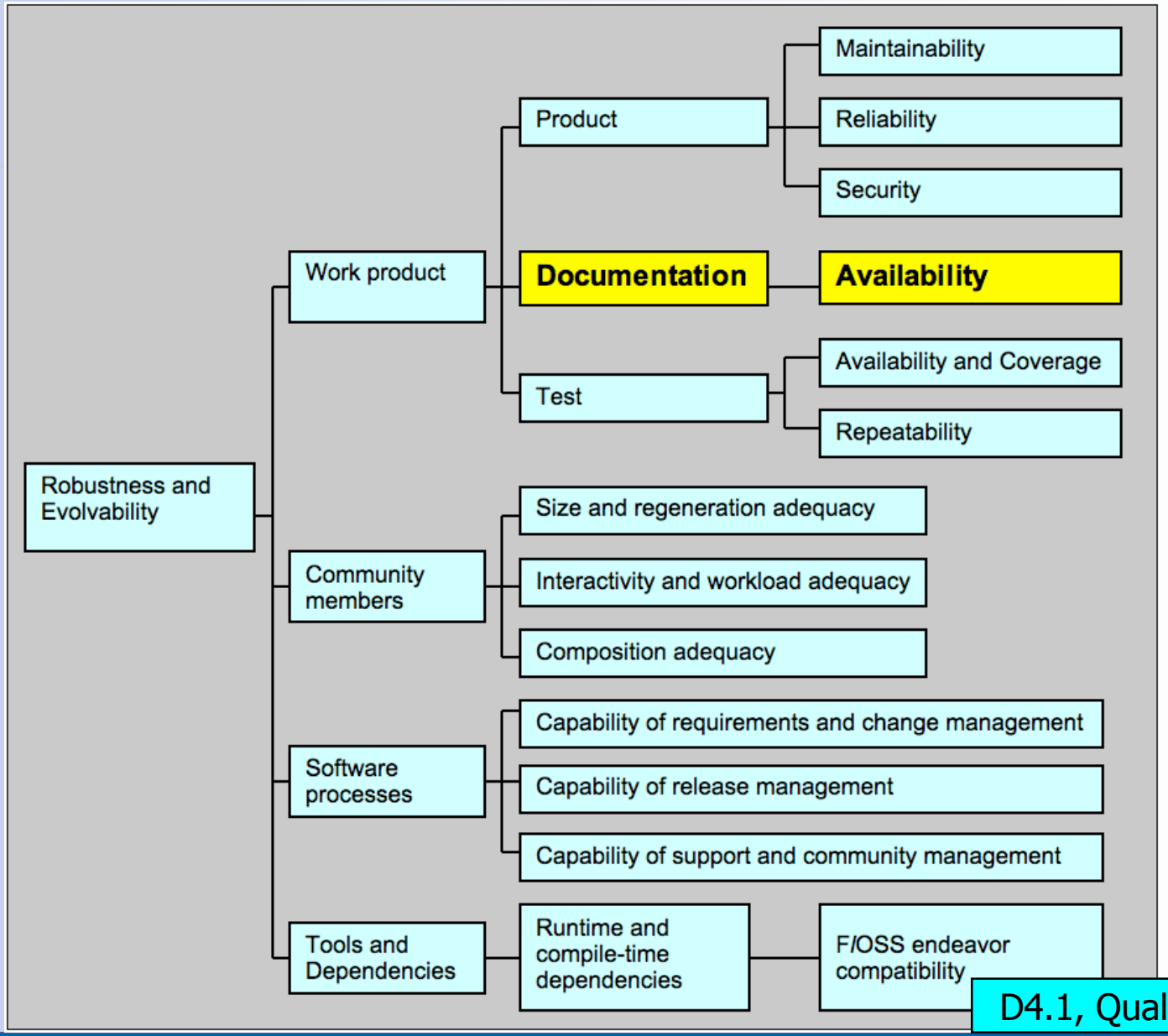
Raimundas Matulevičius,
Flora Kamseu and **Naji Habra**
PReCISE, Computer Science Faculty,
University of Namur, Belgium

Open Source Software Quality

- Work product
- Community members
- Software processes
- Tools and dependencies



D4.1, QualOSS, 2007



D4.1, QualOSS, 2007

Research question

How to evaluate documentation quality of the open source software?

- Content and Accuracy
- Structural Organisation and Completeness

Outline

- Documentation typology
- Documentation availability
  - Metrics
  - Measurement procedure
  - Indicators
  - Indicator interpretation model

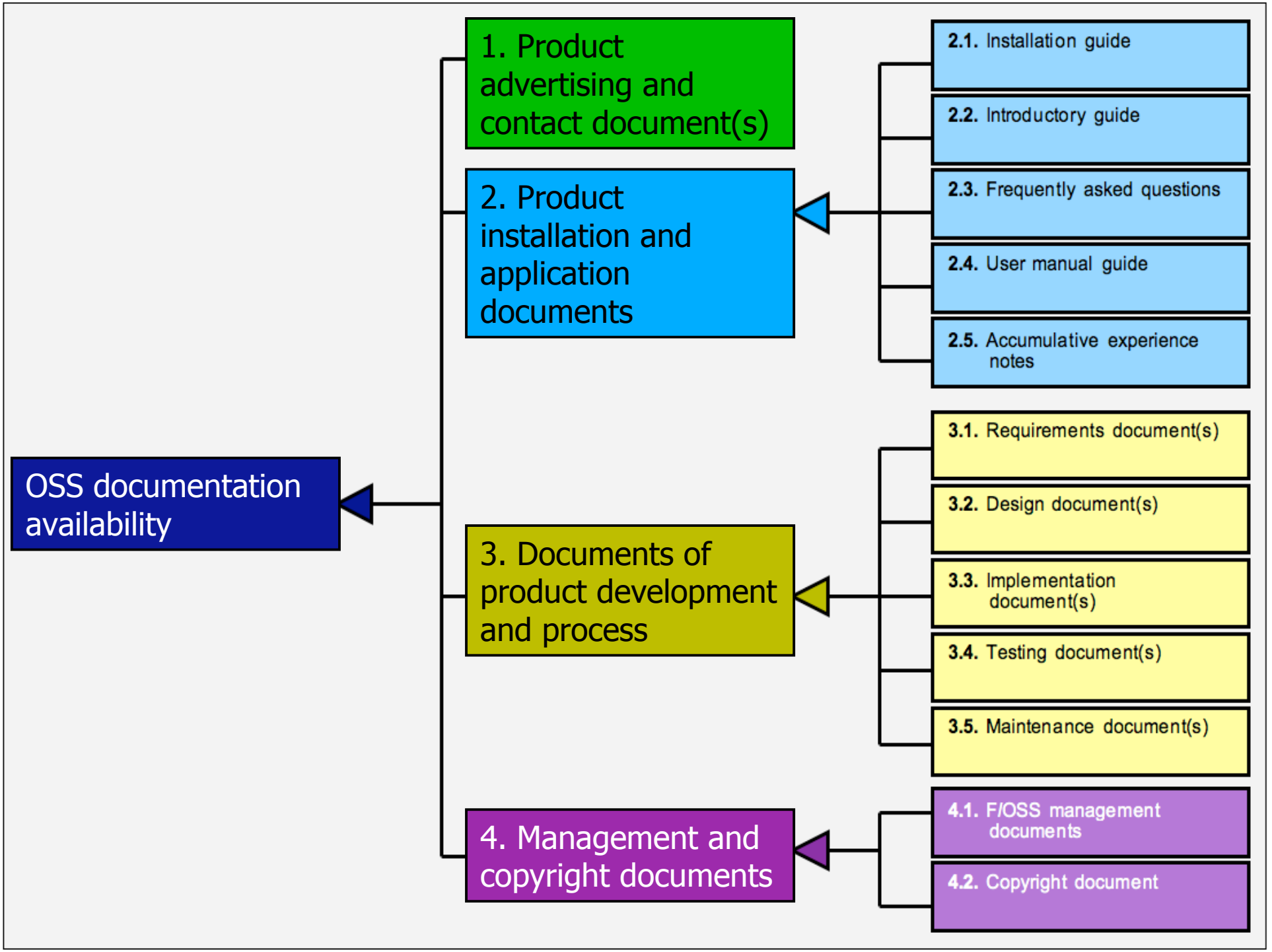
Outline

- Documentation typology
- Documentation availability
  - Metrics
  - Measurement procedure
  - Indicators
  - Indicator interpretation model

OSS product

- Source code
  - compiled results in the tool
- Tool (compiled source code)
  - functions supporting work activities
- Documentation
  - how the source code is written
  - how the tool functions

What are we measuring?


OSS documentation availability
1. Product advertising and contact document(s)
2. Product installation and application documents
2.1. Installation guide
2.2. Introductory guide
2.3. Frequently asked questions
2.4. User manual guide
2.5. Accumulative experience notes
3. Documents of product development and process
3.1. Requirements document(s)
3.2. Design document(s)
3.3. Implementation document(s)
3.4. Testing document(s)
3.5. Maintenance document(s)
4. Management and copyright documents
4.1. F/OSS management documents
4.2. Copyright document

Outline

- Documentation typology
- Documentation availability
  - Metrics
  - Measurement procedure
  - Indicators
  - Indicator interpretation model

Documentation availability

- A quality characteristic:
  - if there ***exists a set of* OSS *documents*** needed for the OSS ***stakeholders*** to achieve their ***goals***;
  - if the physical or electronic ***location*** of these documents ***is known*** to the OSS stakeholders;
  - if these documents contain ***organised*** and ***complete*** information, presented at the ***complete level of detail***.

Stakeholders and their Purposes

Stakeholder role	Goal	Information needed about
Product acquirer	Obtain a OSS product	Product advertising and contact documents
Product users	Use the OSS tool	Product installation and application documents
Product developers	Develop new improve and existing functionality	Documents of product development process
Product contractor	Become a member of the OSS community	Management and copyright documents

Documentation location

- ***Physical location***
  - physical place where the OSS documentation is stored
    - ✓ desks, cases, document archives, and libraries
- ***Electronic location***
  - electronic place where the OSS documentation is stored
  - Private (e.g., personal computer, CD/DVD, USB key)
  - Public (Internet Web sites)
  - Protected (accessible with required security procedures)

Outline

- Documentation typology
- Documentation availability
  - Metrics
  - Measurement procedure
  - Indicators
  - Indicator interpretation model

Metric

Document organisation

- Content is arranged so that information can be *easily* located

No:	Question	Metric	Range
1	Is document divided to **chapters**?	Chapter	Yes / No
2	Is document organised to **sections** and/or **subsections**?	Section / subsection	Yes / No
3	Is there a **table of content** in the document?	Table of content	Yes / No
4	Is there a **term index** in the document?	Term index	Yes / No
5	Is there an **acronym glossary** in the document?	Acronym glossary	Yes / No
6	Is there a **bibliography** (cited **literature references**) in the document?	Bibliography (literature references)	Yes / No / NA
7	Are all the **tables** annotated with a unique **identification number**?	Table identification numbers	Yes / No / NA

Metric

Document organisation

$$g = \frac{\sum_{i=1}^{N} r_i}{N}$$

$\{ r_1 \ldots r_N \}$: answers to each question about document organisation

N : number of questions about document organisation

g : organisation of a single document

$$dor_j = \frac{\sum_{i=1}^{K} g_i}{K}$$

K : number of documents belonging to a document type j

dor_j : organisation of documents belonging to document type j

Metric

Content completeness

- Information necessary for the stakeholder

*Template for **ID3.1 Requirements document(s)***

Content of the document	Description
2. Overall description	Describe the general factors that affect the product and its requirements. Provides a background for those requirements
2.1. Product perspective	Put the product into perspective with other related products or clearly state that it is independent.
2.2. Product function	Provide a summary of the major functions that the software will perform
2.3. User characteristics	Describe those general characteristics of the intended users of the product including educational level, experience, and technical expertise
2.4. Constraints	Provide a general description of any other items that will limit the developer's options
2.5. Assumptions and Dependencies	List each of the factors that affect the requirements stated in the SRS. These factors are not design constraints on the software but are, rather, any changes to them that can affect the requirements
2.6. Apportioning of requirements	Identify requirements that may be delayed until future versions of the system

Metric

Information completeness

- All information units are defined at the high level of detail

No:	Question	Metric	Range
1	What is the degree of information presentation in a document?	Information presentation degree	0 – when question is not analysed; 1 – low presentation degree; 2 – average presentation degree; 3 – hight presentation degree

Metric

Documentation completeness

$$dco_j = \frac{\sum_{i=1}^{M} c_i d_i}{3M}$$

{ c_1 ... c_M } : estimates of *content completeness*

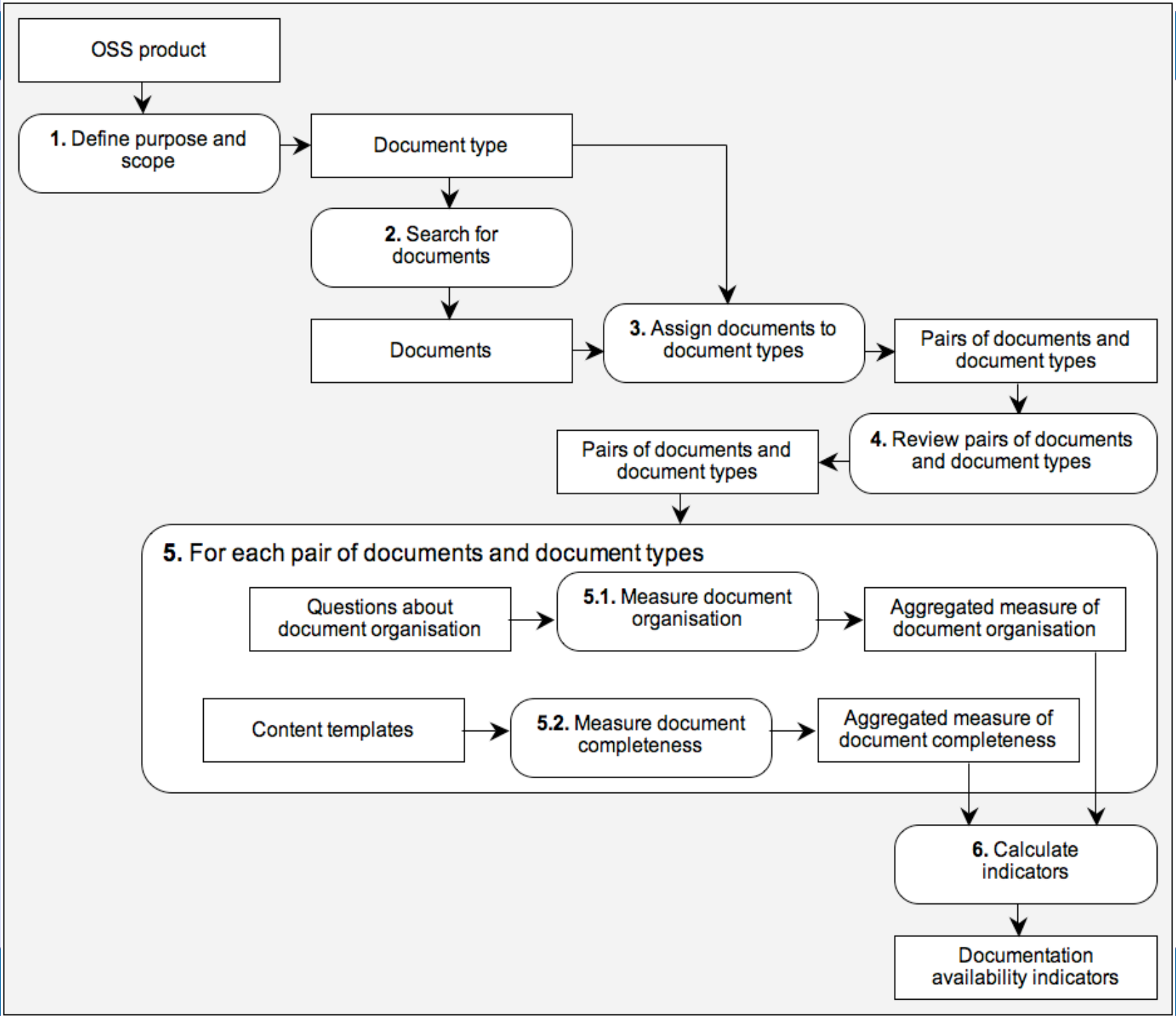
{ d_1 ... d_M } : estimates of *information completeness*

M: number of questions

dco$_j$ **:** completeness of documents belonging to document type *j*

Outline

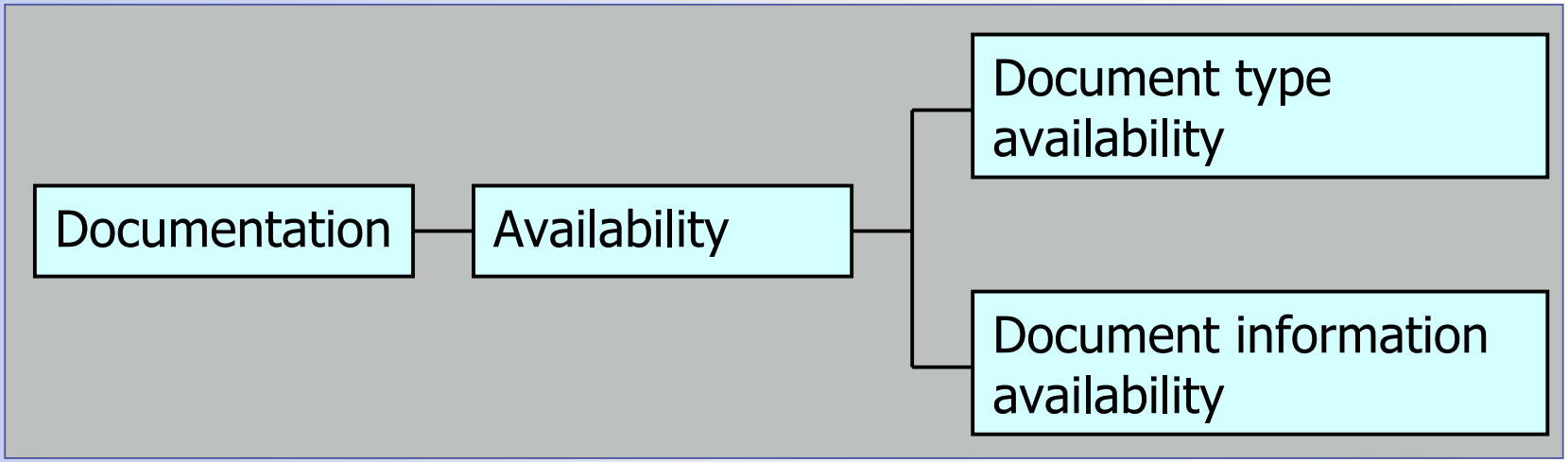
- Documentation typology
- Documentation availability
  - Metrics
  - Measurement procedure
  - Indicators
  - Indicator interpretation model


OSS product
1. Define purpose and scope
Document type
2. Search for documents
Documents
3. Assign documents to document types
Pairs of documents and document types
4. Review pairs of documents and document types
Pairs of documents and document types
5. For each pair of documents and document types
Questions about document organisation
5.1. Measure document organisation
Aggregated measure of document organisation
Content templates
5.2. Measure document completeness
Aggregated measure of document completeness
6. Calculate indicators
Documentation availability indicators

Outline

- Documentation typology
- Documentation availability
  - Metrics
  - Measurement procedure
  - Indicators
  - Indicator interpretation model

Indicators


Documentation
Availability
Document type availability
Document information availability

Indicator

DTA: Document type availability

- Availability of documents belonging to a certain document type

$$DTA = \frac{DF}{DN}$$

DF : number of documents types for which documents are found

DN : number of considered document types

Indicator

DIA: Document information availability

- Organisation and information completeness of the document content

$$DIA = \frac{\sum_{i=1}^{DN} (dor_i + dco_i)}{2DN}$$

dor_i : *organisation of documents* belonging to document type i

dco_i : *completeness of documents* belonging to document type i

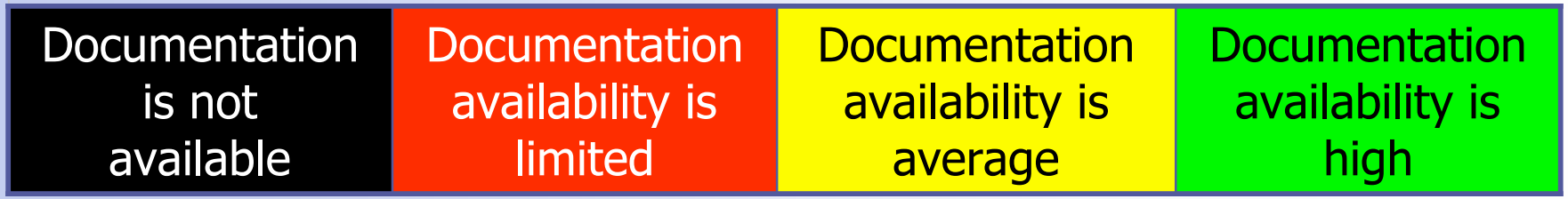
DN : number of considered document types

Outline

- Documentation typology
- Documentation availability
  - Metrics
  - Measurement procedure
  - Indicators
  - Indicator interpretation model

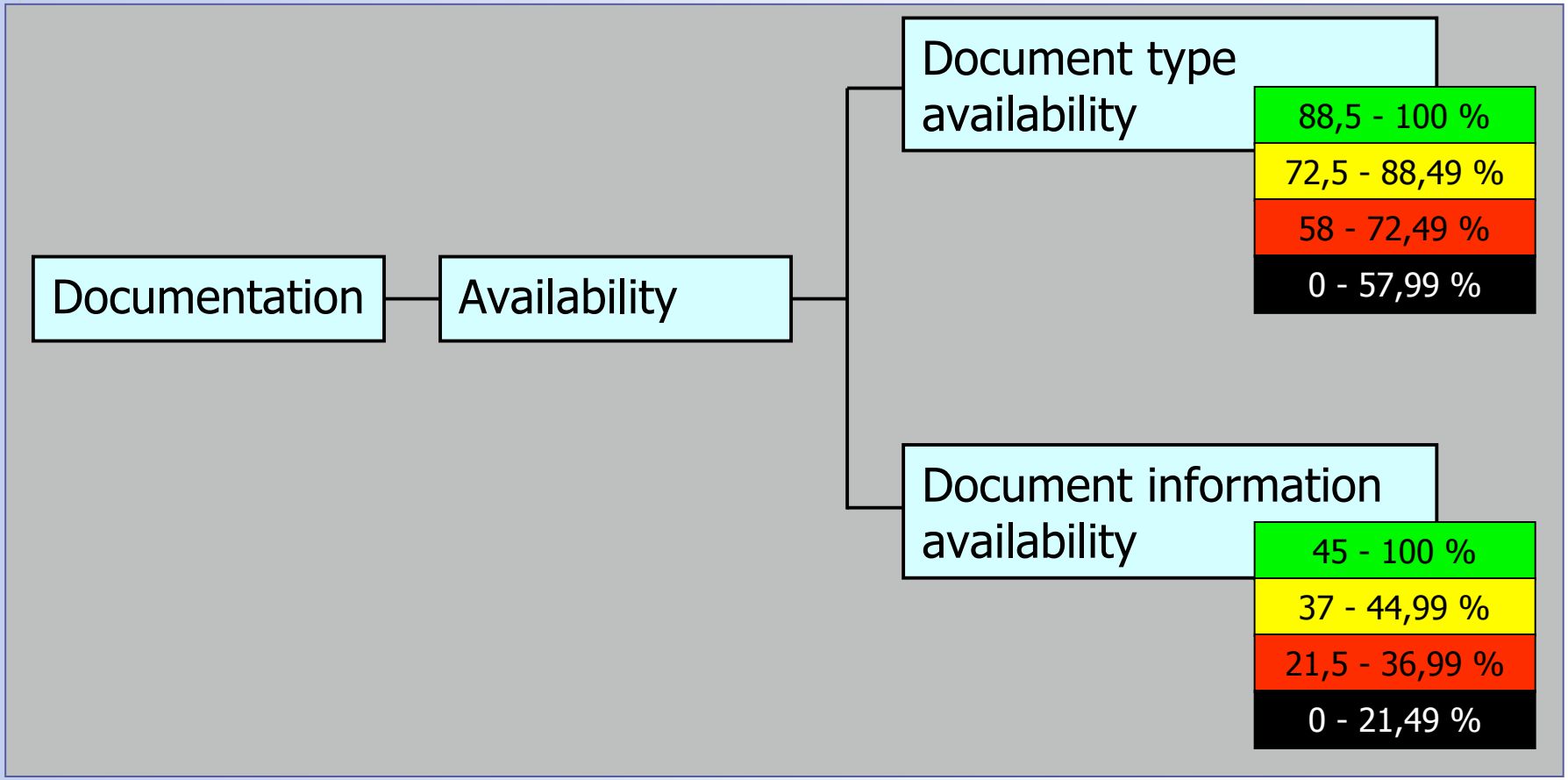
Indicator interpretation

How can indicators be interpreted?



How can interpretation of the documentation availability (DA) indicators be estimated?

Interpretation model for the documentation availability indicators



Indicator interpretation

	Projects	Document type availability	Document information availability
1	Thunderbird (Mozilla)	100.00%	50.46%
2	Hadoop	100.00%	41.72%
3	Xemacs	100.00%	41.28%
4	Python	92.31%	52.45%
5	Evolution	92.31%	32.72%
6	Jmeter	92.31%	32.08%
7	PLONE	84.62%	52.08%
8	Zope	84.62%	49.47%
9	FreeBSD	84.62%	47.31%
10	NetBSD	84.62%	43.25%
11	Eclipse	84.62%	42.98%
12	Jetspeed	84.62%	34.83%
13	Evince	84.62%	30.85%
14	PhPMyAdmin	76.92%	46.56%
15	Writer (OpenOffice)	76.92%	42.20%
16	VLC	76.92%	38.74%
17	Jakarta Struts	76.92%	34.27%
18	Nautilus	76.92%	25.44%
19	k3b	69.23%	31.52%
20	xPDF	69.23%	28.96%
21	Findbugs	69.23%	27.62%
22	CVSanaly	69.23%	24.54%
23	Galeon	61.54%	17.01%
24	Sup	61.54%	16.15%
25	Omnitux	53.85%	15.56%
26	yanolc	46.15%	18.41%

- Availability of documents of a certain type does not guarantee availability of the information
- A document can contain high availability degree of information, although the overall project documentation availability might be of a low degree

Standard QualOSS assessment Method

- Version 1.1