14 March 2008, University of Tsukuba/Japan

筑波大学
University of Tsukuba

Japanese Symposium Digital Preservation

-- Technical Issues --

(kopal, repository systems, ingest, validation)

Dr. Heike Neuroth
State & University Library, Goettingen
Max Planck Digital Library, Berlin
Germany
neuroth@sub.uni-goettingen.de

official mascot of Tsukuba

ToC

- The Challenges
- Some Basics in Decision Making Techniques
- Criteria Catalog for Comparing and Assessing Products (Software)
- Some Selected Details of the Criteria Catalog and Products
- More Information

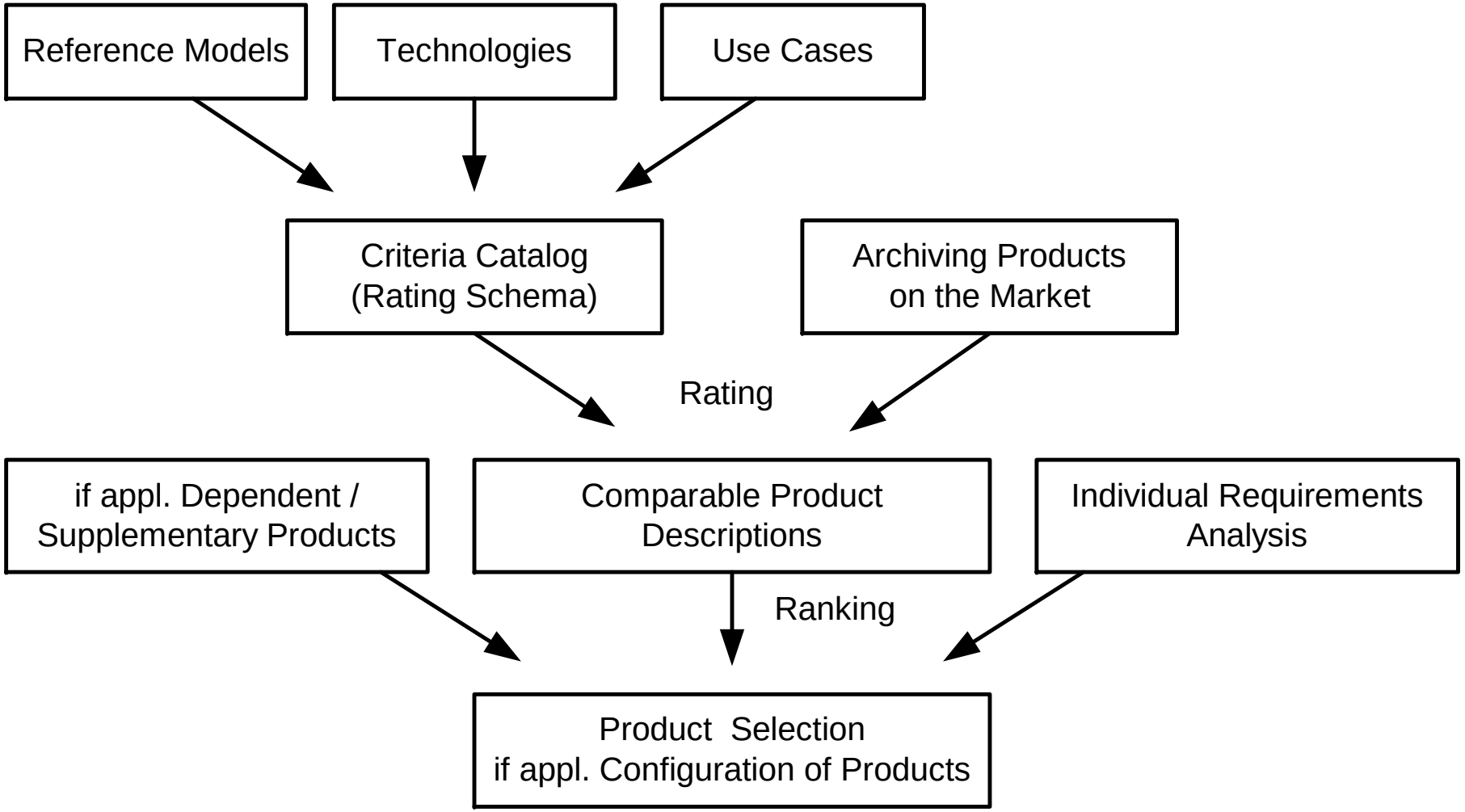
Challenges

- Digital objects are inherently complex
- Individual requirements are quite heterogeneous
- Common criteria for quality (trustworthiness) are rather abstract
- Frameworks may be far away from the implementation level
- Repository software is complex
- System quality strongly depends on individual configurations
- Documentation of products is still an issue
- Further virtualization of technical infrastructure, e.g., GRID computing, raises new challenges

But …

- Initiatives and projects around the world deal with long-term preservation
- First frameworks, partly formally standardized, are available, e.g. OAIS (ISO)
- Specific tools are designed, under development, or running, e.g. KoLiBri for ingest, JOVHE for validation
- Software makers expect a market for (long-term) archiving products and services
- A set of lessons learned helps us to optimize future work

Decision Process Example



筑波大学 University of Tsukuba

A Glance at Rating and Ranking Techniques

- Complex systems require multiple attributes (criteria) to be rated (MADM: Multi Attribute Decision Making)
- For each attribute the degree of fulfillment may depend on individual requirements, e,g. a perfect TIFF-viewer will be useless if there are no plans to store TIFF-files in your repository
- Ranking is a linear ordering of choices like

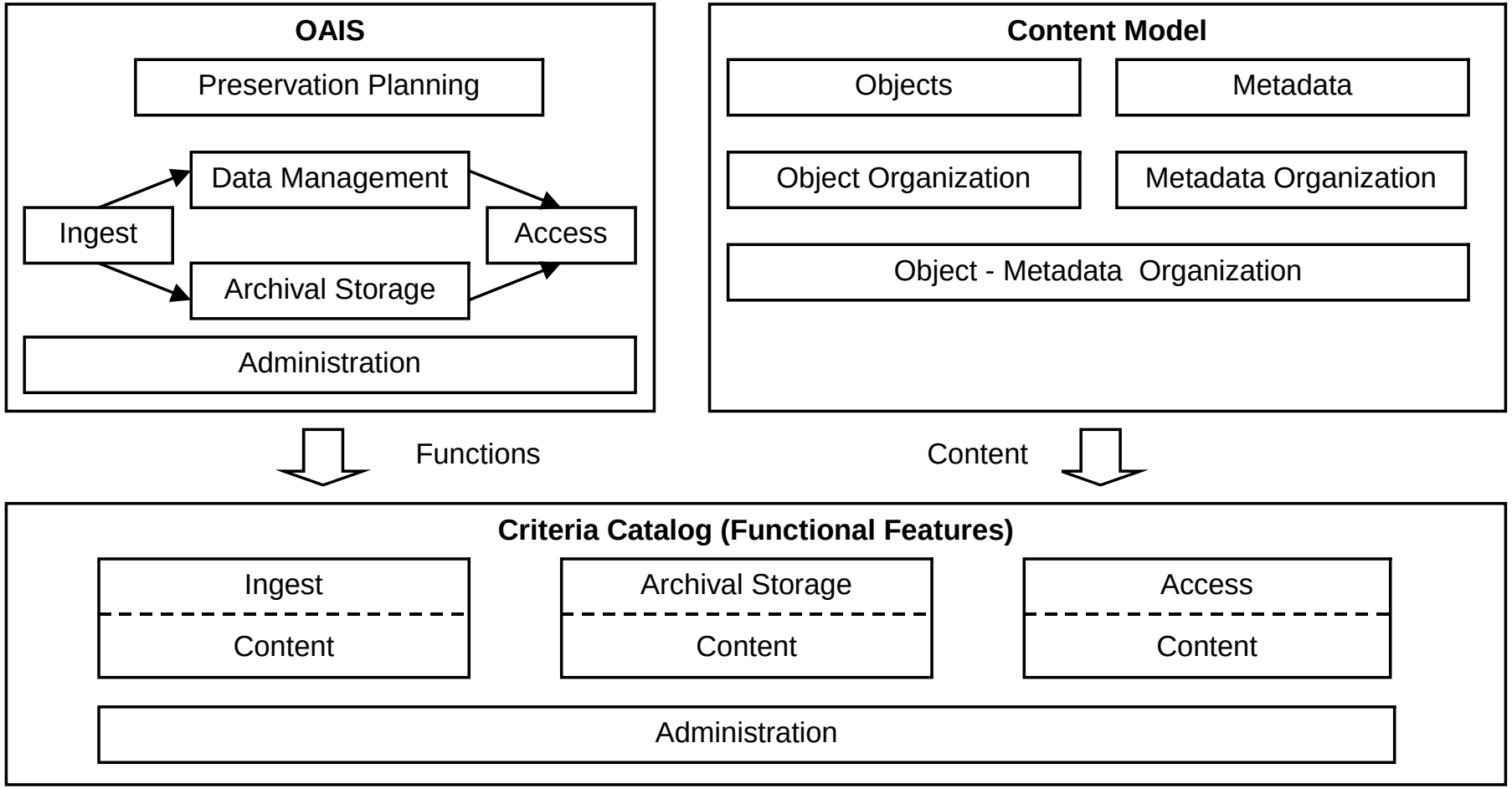
  product 1 is better than product 2 and product 2 is better than Product 3 …

  => you need a technique to map the rated criteria to a single scale

...

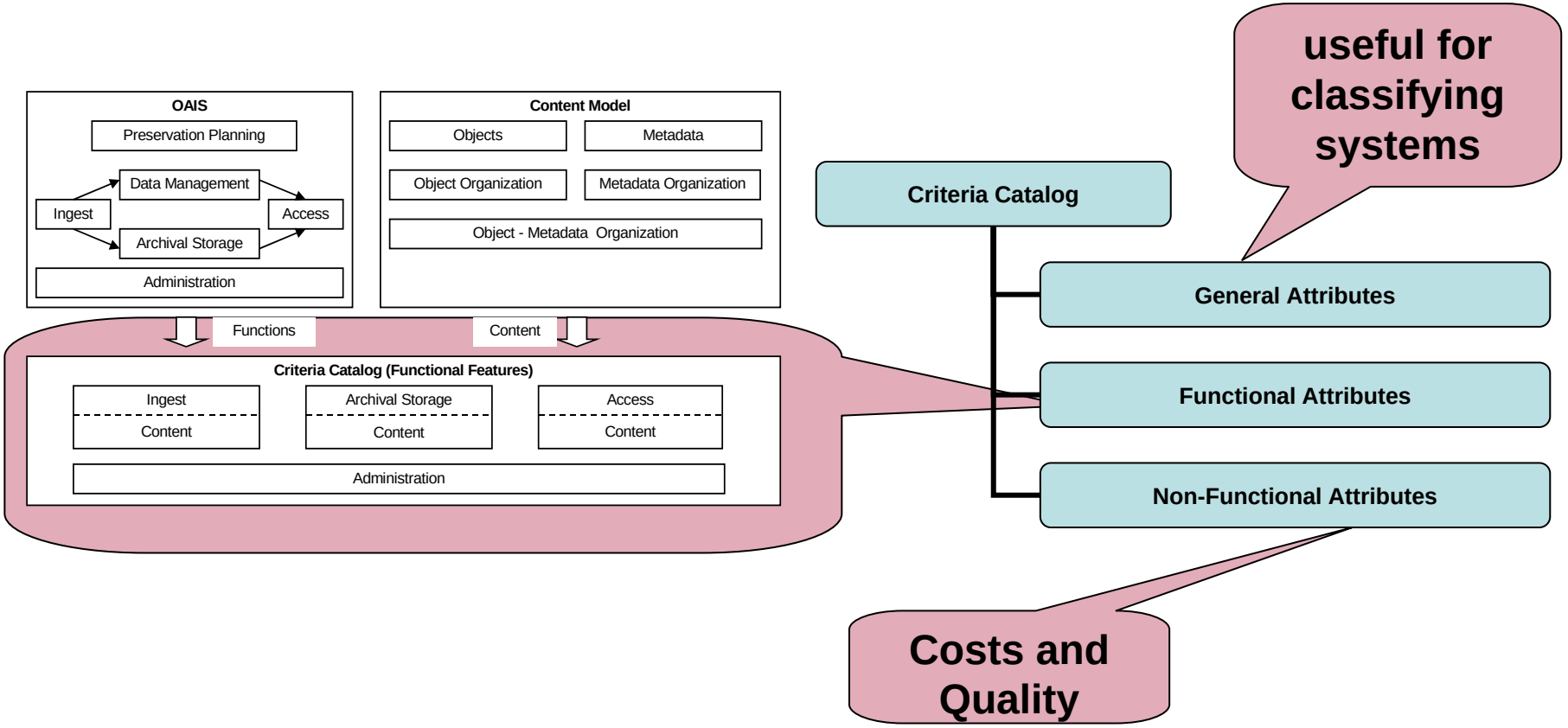
- This mapping may also depend on individual requirements, e.g. to which degree can a perfect TIFF-viewer compensate for the shortcomings of poor metadata schemes necessary to describe the TIFF-files

Developing the Criteria Catalog



筑波大学 University of Tsukuba

Criteria Catalog: Main Structure



筑波大学 University of Tsukuba

Criteria Catalog: General Attributes

- Overall system architecture
  - Design principles: Compliance with standards or recommendations (e.g., OAIS, OAI)
  - Explicit long-term features, e.g. file format registry, preservation metadata scheme, VM
  - Object organization, e.g. single objects, collections, identification
  - Metadata organization, e.g. supported metadata schemes
  - Rights, e.g. Object-related rights management
  - Roles, e.g. Consumer / producer / archive operator
  - Functions, Pre-ingest / ingest / access / archival storage / administration

...

- System / application integration

  Library system / publishing system /
  product data management system / other archives
- Organizational integration

  Federation / cooperation / user communities
- Software architecture
- Hardware basis

筑波大学
University of Tsukuba

Criteria Catalog: Functional Attributes - Ingest

- Accepted submission formats
  - Object format / identification, e.g., file format restrictions
  - Object organization, E.g., hierarchies, links, versions, variants
- Access procedures for producers
  - metadata scheme incl. meta-data entry procedure
  - batch ingest / conversion / (formal) quality checking / dedicated workflow
  - for metadata: manually / automatic extraction / 3rd party services
- Creation of Archival Storage Organization, e.g. final step of ingest like generating IDst

Criteria Catalog: Functional Attributes – Access

- Access procedure for consumer (Remote vs. local / multilingual / help system / notification services / communication protocols)
- Search / retrieval (Metadata indexes / navigation / full text search / inspection of class methods)
- Dissemination form of objects / metadata (Conversion on the fly / on demand)
- Accounting (e.g. as part of a Digital Rights Management)
- Federation (Access or replication transparency)
- Interoperation (Explicit exchange of objects and metadata)

Criteria Catalog: Functional Attributes - Archival Storage

- Archival Storage Organization - conceptual organization of objects, metadata, and their relations
  - Object formats
  - Object relationships
  - Object identifications
  - Object versions and variants (manifestations)
  - Metadata
  - Relationships objects – metadata
- Logical Storage Organization - Mapping of conceptual organization to logical elements (e.g. files / file systems / database tables)
- Physical storage - media / interfaces / abstraction
- Limits, e.g. number / size of objects (or relations)

筑波大学 University of Tsukuba

Criteria Catalog: Functional Attributes - Admin

- Access procedures for administrators

  Local / remote / special protection
- Administration of objects and metadata

  Deletion of collection / reorganization / updates / controlled vocabulary
- Administration of user access

  OAIS-roles like producer / consumer / admin / management
- Object-related rights
- Administration of physical storage

  e.g. allocation of storage for objects / collections / roles

...

- Access to internal interfaces
  e.g. to basic database schemes / storage system
- Configuration / scaling
  e.g. scalability transparency
- Disaster management
  e.g. backup / recovery
  Redundancy / replication / fragmentation for availability
- Monitoring / reporting
  Trouble ticket systems / error reports / statistics / metrics

筑波大学 University of Tsukuba

Criteria Catalog: Non-Functional Attributes - Costs

- Product Costs - initial purchase / license / leasing / maintenance / updates / training
- Human resources
  - Initial installation
  - Operating
  - End user / producer support (hotline, newsletter, FAQ / producing submissions)
  - Long-term preservation (monitoring of designated community / monitoring of applied (embedded) technologies / media migration)
- Material Resources (hardware and additional software)

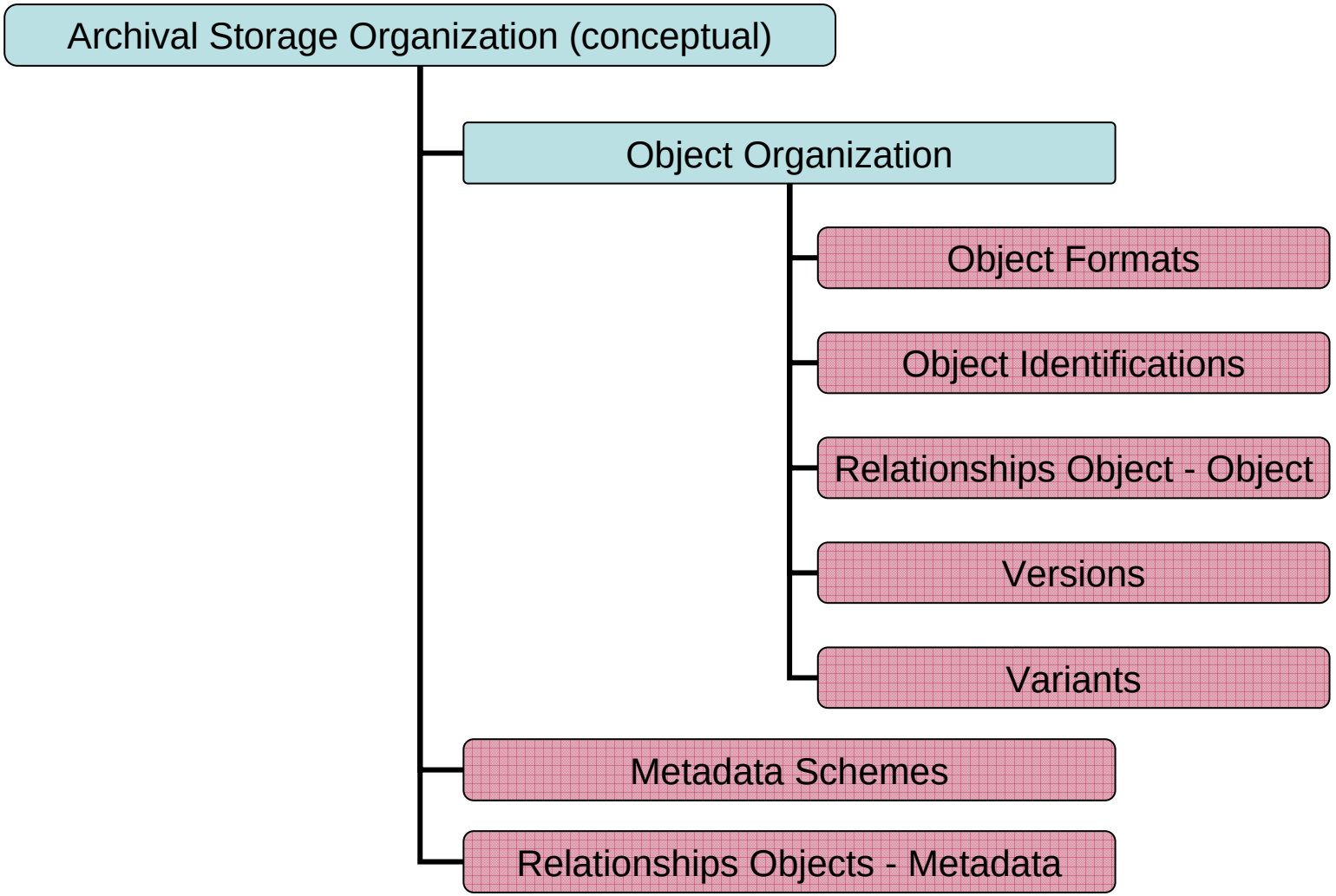
Criteria Catalog: Non-Functional Attributes - Quality

- Maturity of manufacturer
- Maturity of product

  Development status (ratio of implemented features to announced ones)
- Stability

  Quality of implemented features
- Documentation
- Support
- Market penetration / user community

Matrix?



筑波大学 University of Tsukuba

dSpace

Object Formats	any bitstream (computer file)
Object Identifications	PI for *Communities, Collections, Items* (CNRI handles)
Relationships Object - Object	(multi-)hierarchies: *Community – opt. Sub-Cummunity* (arbitrary depth) – *Collection – Item – Bundle – Bitstream; Item* can belong to more than one *Collection*
Versions	no predefined structures
Variants	no predefined structures
Metadata Schemes	basic schema for descriptive metadata (qual. DC), administrative metadata, structural metadata (organization of objects within *Items, Bundles)*
Relationships Object - Metadata	basic schema for descriptive metadata and any schema (as serialized bitstream) at level *Item*, reduced descriptive metadata at level *Community, Collection;* assignment of *bitstreams* to one support level in: *supported, known, unsupported*

DigiTool

Object Formats	any bitstream (computer file)
Object Identifications	PI, user-definable syntax and internal resolving, add. PI can be defined
Relationships Object - Object	hierarchies of arbitraray depth, relationship types: part-of, includes
Versions	no predefined structures (implicit by object history)
Variants	*manifestations*
Metadata Schemaes	1. predefined: DC, MARC21, Z39.87 Mix, PREMIS *objects* and *events,* METS Text MD, LOC AMD&VMD, object history, access rights; 2. user defined: local fields and categories
Relationships Object - Metadata	n : m (metadata-objects get an ID)

Kopal / DIAS

Object Formats	any bitstream (computer file)
Object Identifications	PI: URN (at package level)
Relationships Object - Object	hierarchies (as modelled by file system folders), object links as expressible by METS (with some constraints) folders together with METS-file within one package called *UOF (Universal Object Format)*
Versions	predefined structures at package level (see METS)
Variants	no predefined structures (see METS)
Metadata Schemes	METS (with defined constraints, e.g., preservation metadata denoting the format of each file is required) based on a subset of LMER (Long-term preservation Metadata for Electronic Resources)
Relationships Object - Metadata	mechanisms of METS (with defined constraints); assigment of administrative MD at package level and file level

And very new: eSciDoc

- Federal Ministry of Education and Research (BMBF), funded project with Max Planck Society and FIZ Karlsruhe
- Running until 2009
- first release this year: 1. European eSciDoc user group meeting in June in Berlin
- eSciDoc is as a joint project with the aim to realize a next-generation platform for communication and publication in research organizations (including resaearch data, cross-disciplinary)

eScience Framework

- To exempt researchers from barriers of physical location of knowledge (data) and know-how (individuals), a solid infrastructure to provide data storage, interoperability and seamless integration into community-specific working environments is necessary
- Cross-disciplinary
- Research data & publications

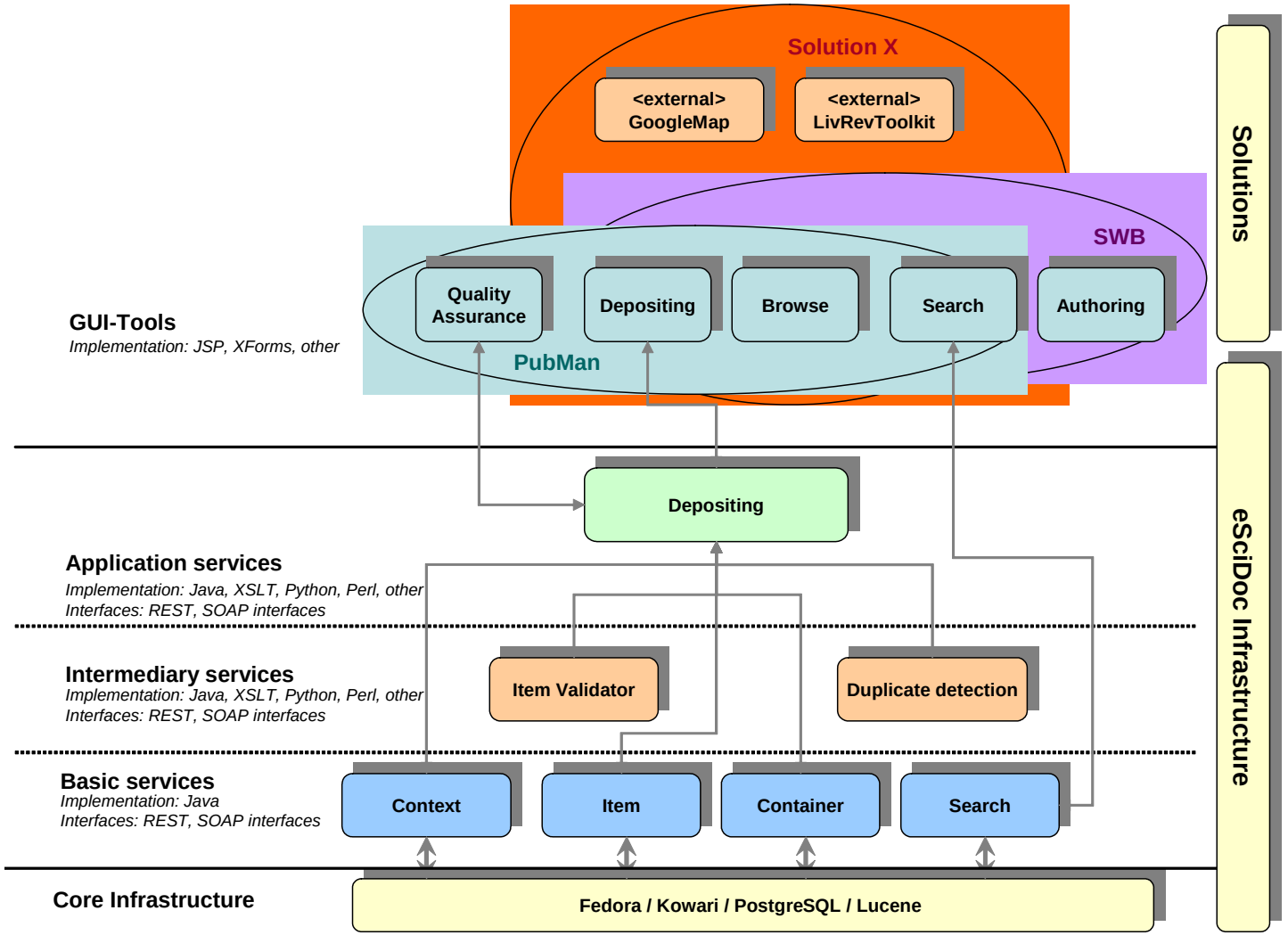
eSciDoc Infrastructure

- eSciDoc Infrastructure includes the following services:
  - Object Manager
  - Organizational Unit Handler
  - Search
  - Security
  - Content Model Manager
  - Statistics
  - Semantic Store Handler
  - PID Manager
  - Workflow Manager
  - Validation

http://192.129.1.106:8080/ir/item/escidoc:363

筑波大学 University of Tsukuba

eSciDoc Architecture



筑波大学 University of Tsukuba

eSciDoc Information

- Open Source Software, download:
  http://www.escidoc-project.de/JSPWiki/Wiki.jsp?page=Download

- Contact:
  Malte Dreyer at Max Planck Digital Library
  malte.dreyer@mpdl.mpg.de

- Big interest worldwide: eSciDoc federation?

筑波大学
University of Tsukuba

More Information

- nestor Subject Gateway at: http://nestor.sub.uni-goettingen.de/nestor_on/index.php

筑波大学 University of Tsukuba

筑波大学
University of Tsukuba

Title:	**Digital Preservation Management: Implementing Short-term Strategies to Long-term Problems**
URL:	**http://www.library.cornell.edu/iris/tutorial/dpm/index.html**
Creator:	Kenney, Anne R.Kenney, Anne R. — McGovern, Nancy Y. — Entlich, Richard
Publisher / Organiser:	Cornell University LibraryCornell University Library [**Contact**]
Contributor:	Kehoe, William R.; Olsen, Erica
Document Type:	**Online Tutorial**
Subject:	**General Resources — Management**
Keywords:	training
Description:	"Cornell University Library is offering an innovative new digital preservation training program with funding from the National Endowment for the Humanities. The program consists of an online tutorial and a series of one-week workshops held in Ithaca, NY. The primary goal of this program is to enable effective decision making for administrators who will be responsible for the longevity of digital objects in an age of technological uncertainty."
Language:	English
Region:	Global
File Format:	text/html
Date record last updated:	2005-07-15

[top]

Export
[Print] [Send page by e-mail]

14 March 2008, University of Tsukuba/Japan

筑波大学
University of Tsukuba

Japanese Symposium Digital Preservation

Thank your for your attention!
Time for discussion, questions …

Dr. Heike Neuroth
State & University Library, Goettingen
Max Planck Digital Library, Berlin
Germany
neuroth@sub.uni-goettingen.de

official mascot of Tsukuba