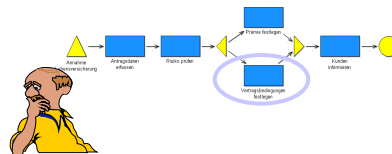


## 5 Metadata and Meta-Knowledge

Knut Hinkelmann

### Limits of Search Engines

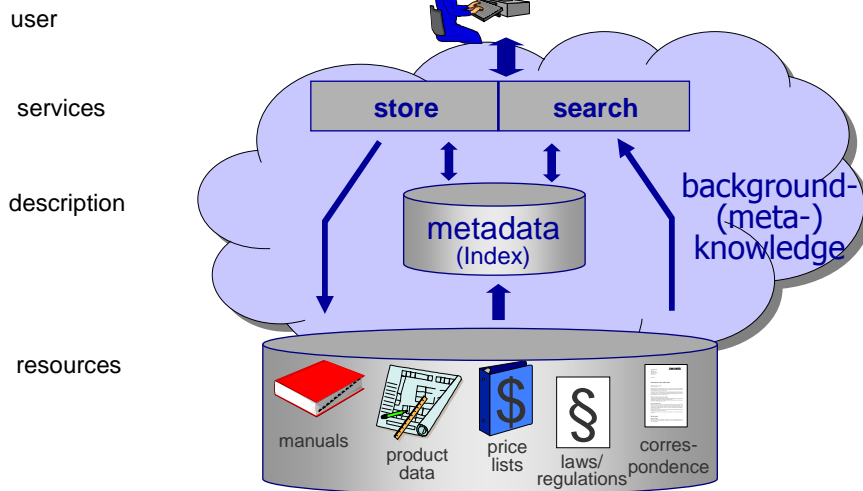
- Fulltext search only works for coded information
- Relevance of document may depend on criteria that are not mentioned in the text, e.g.
  - ◆ document type (article, presentation, treaty, manual, ...)
  - ◆ author of the document
  - ◆ date of publication
  - ◆ context (project, customer, ...)
  - ◆ logical structure (sender/recipient of a letter)
  - ◆ usefulness for a task
  - ◆ appropriateness for a user (level of expertise, ...)



**Additional meta-data and meta-knowledge can help**



## Meta-data and Meta-knowledge



## Use of meta-data

- Search for information resources using suitable criteria
- organisation and effective access to electronic resources (e.g. document management systems, digital libraries)
- unique identification of resources
  - ◆ (storage) location, e.g. URL
  - ◆ unique ID
- distinction of different resources
- data exchange between systems with different data structures and interfaces

## Structured Meta-data – Examples

user data (document)



meta-data

name:	ELENA-Ber
creation:	18.3.2001
modification:	25.6.2001
format:	Word
document type:	project report
recipient:	All Life Insurance Inc.
author:	Smith

Examples for Meta-data:

- ◆ *library catalogue* with description of books: author, title, publication date, publisher, key words, location
- ◆ *document management systems* distinguish between user data (resources, documents) and meta-data
- ◆ *skill databases / yellow pages* contain descriptions of people



## General vs. application-specific metadata

### ■ General metadata

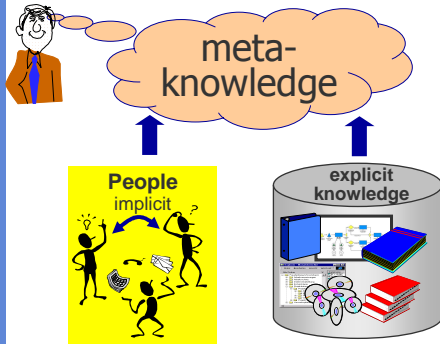
- ◆ can be used for any kind of information
- ◆ Examples: author, date of creation, subject

### ■ Application-specific metadata

- ◆ specific attributes or attribute refinements
  - examples: for a letter: sender  
for a report: project name  
for a piece of music: the style
- ◆ attribute values
  - examples: database of customers  
predefined list of project names  
classification of music styles



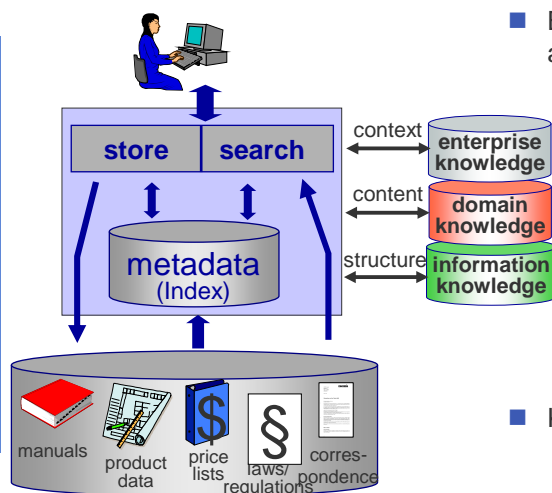
## Knowledge and Meta-Knowledge



- knowledge consists of two levels
  - ◆ knowledge
  - ◆ meta-knowledge
- meta-knowledge is knowledge about existing knowledge or missing knowledge, e.g.
  - ◆ Where do I find which knowledge?
  - ◆ How reliable is the knowledge?
- meta-knowledge is a prerequisite to deal with knowledge itself
- it is a task of knowing, to acquire and adapt meta-knowledge

**Meta-knowledge can be made explicit, too**

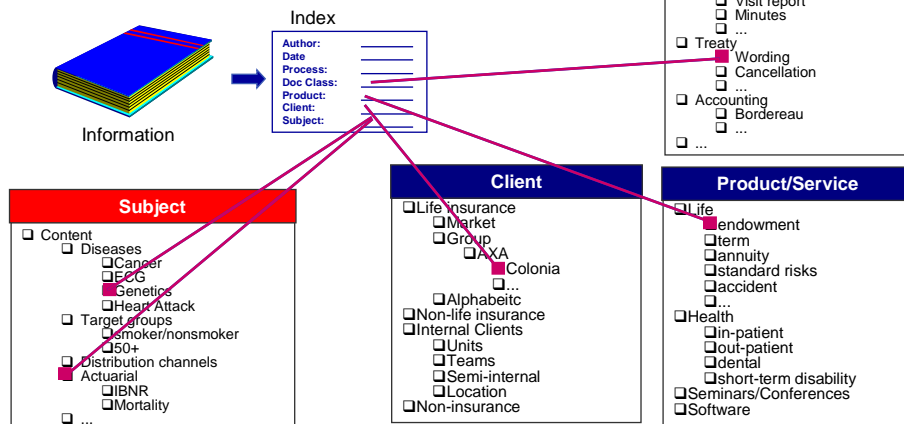
## Types and Use of Meta-Knowledge



- Background (Meta-) knowledge about ...
  - ... enterprise (context)
    - processes
    - organisational structure
    - customers
  - ... domain (content)
    - Terms, concepts
    - relationships
  - ... information
    - document types
    - document structure
- Knowledge is used for
  - ◆ indexing
  - ◆ retrieval

## Meta-data and Meta-Knowledge: Example

Document Management for a Re-Insurance Company



- Meta-data are represented as attribute values
- The domain of possible attribute values represent knowledge

## 5.1 Dublin Core Metadata Initiative (DCMI)

- The Dublin Core Metadata Initiative (DCMI) is an open organization
  - ◆ engaged in the development of interoperable online metadata standards that support a broad range of purposes and business models
  - ◆ dedicated to promoting the widespread adoption of interoperable metadata standards and
  - ◆ developing specialized metadata vocabularies for describing resources that enable more intelligent information discovery systems.



<http://www.dublincore.org>

## ***DCMI Objectives***

- **Simplicity of creation and maintenance**
  - ◆ non-specialist can create simple descriptive records for information resources easily and inexpensively
- **Commonly understood semantics**
  - ◆ a common set of elements, the semantics of which are universally understood and supported.
- **International scope**
  - ◆ versions in many languages available
- **Extensibility**
  - ◆ mechanism for extending the DC element set



## ***DCMI - Overview***

- The Dublin Core standard includes two levels:
  - ◆ Simple
  - ◆ Qualified
- Simple Dublin Core comprises
  - ◆ 15 basic elements  
(DCMES - Dublin Core Metadata Element Set)
- Qualified Dublin Core includes
  - ◆ additional elements
  - ◆ element refinements (also called qualifiers)
  - ◆ Value Encoding Schemes



## Simple Dublin Core

- The Dublin Core Metadata Element Set is a vocabulary of fifteen properties

<b>creator</b>	<b>title</b>	<b>subject</b>	<b>publisher</b>
<b>contributor</b>	<b>date</b>	<b>description</b>	<b>type</b>
<b>coverage</b>	<b>rights</b>	<b>relation</b>	<b>format</b>
<b>source</b>	<b>language</b>	<b>identifier</b>	

- Each element is optional and can be repeated
- Elements correspond to structured meta-data attributes

- ◆ Example:

```
Title="Studienfuehrer Hochschule fuer Wirtschaft"  
Source="http://www.fhnw.ch/wi"  
Subject="Bachelor,Wirtschaftsinformatik,Betriebsoekonomie"  
Type="Image"  
Type="Text"
```

<http://www.dublincore.org/documents/dces/>



## Elements in Simple Dublin Core (1/3)

**Title:** The name given to the resource

**Subject:** The topic of the content of the resource.

- Typically expressed as keywords, key phrases or classification codes.
- Recommended best practice: to select a value from a controlled vocabulary or formal classification scheme.

**Description:** An account of the content of the resource.

- abstract, table of contents, reference to a graphical representation of content ...

**Type:** The nature or genre of the content of the resource.

- general categories, functions, genres, ...

**Source:** Reference to a resource from which the present resource is derived

<http://www.dublincore.org/documents/dces/>



## ***Elemente in Simple Dublin Core (2/3)***

**Relation:** a reference to a related resource

**Coverage:** the extent or scope of the content of the resource.

- spatial location (a place name or geographic co-ordinates),
- temporal period (a period label, date, date range) or
- jurisdiction (administrative entity).

**Creator:** an entity primarily responsible for making the content of the resource

- person, organisation, service

**Publisher:** an entity responsible for making the resource available

- person, organisation, service

**Contributor:** an entity responsible for making contributions to the content of the resource

- person, organisation, service



## ***Elements in Simple Dublin Core (3/3)***

**Rights:** information about rights held in and over the resource

- Intellectual Property Rights (IPR), Copyright

**Date:** a date associated with an event in the life cycle of the resource

- creation, modification, access
- Recommended best practice: defined in a profile of ISO 8601 and follows the YYYY-MM-DD format

**Format:** the physical or digital manifestation of the resource

**Identifier:** an unambiguous reference to the resource within a given context

- URI, Signatur, ISBN

**Language:** language of the intellectual content of the resource

- z.B. RFC 3066 [RFC3066]





## Qualified Dublin Core

- Additional elements:
  - ◆ Audience, Provenance, RightsHolder, InstructionalMethod, AccrualMethod, AccrualPeriodicity, AccrualPolicy, ...
- Element Refinements:
  - ◆ qualifiers making the meaning of an element narrower or more specific  
(e.g. distinguishing date of creation and modification)
- Value Encoding Schemes:
  - ◆ schemes that aid in the interpretation of an element value



## Additional Elements in Qualified Dublin Core

- All Elements are described in the document "DCMI Metadata Terms", <http://www.dublincore.org/documents/dcmi-terms/>
- Here are some examples of additional elements
  - Audience:** class of entity for whom the resource is intended or useful
  - Provenance:** statement of any changes in ownership and custody of the resource since its creation that are significant for its authenticity, integrity and interpretation
  - RightsHolder:** person or organization owning or managing rights over the resource
  - InstructionalMethod:** process, used to engender knowledge, attitudes and skills, that the resource is designed to support  
→ ways of presenting instructional materials
  - accrualMethod:** method by which items are added to a collection



## Element Refinements

Element	Refinement
<i>Title</i>	Alternative
<i>Description</i>	Table Of Contents Abstract
<i>Date</i>	Created Valid Available Issued Modified Date Accepted Date Copyrighted Date Submitted
<i>Format</i>	Extent Medium
<i>Identifier</i>	Bibliographic Citation
<i>Coverage</i>	Spatial Temporal

Element	Refinement
<i>Relation</i>	Is Version Of Has Version Is Replaced By Replaces Is Required By Requires Is Part Of Has Part Is Referenced By References Is Format Of Has Format Conforms To
<i>Rights</i>	Access Rights License
<i>Audience</i>	Mediator EducationLevel

The Element Refinements can be used like the 15 core elements

## Encoding Schemes

- Encoding schemes aid in the interpretation of an element value
- These schemes include
  - ◆ controlled vocabularies (a term from a classification system or set of subject headings) and
  - ◆ formal notations or parsing rules (e.g., "2000-01-01" as the standard expression of a date )
- Examples of Encoding Schemes
  - ◆ Uniform Resource Identifier (URI) for "identifier", "relation"
  - ◆ classification systems for "subject":
    - Library of Congress Classification (LCC),
    - Dewey Decimal Classification (DCC)
  - ◆ W3C-DTF for date

## Encoding Schemes

Element	Encoding Scheme(s)
Subject	LCSH, MeSH, DDC, LCC, UDC
Date	DCMI Period, W3C-DTF
Type	DCMI Type Vocabulary
Format	IMT
Identifier	URI
Source	
Language	ISO 639-2RFC 3066
Relation	URI
Coverage	Spatial: DCMI Point, ISO 3166, DCMI Box, TGN Temporal: DCMI Period, W3C-DTF

## Dublin Core Principles

- The One-to-One Principle
  - ◆ Dublin Core metadata describes one manifestation or version of a resource.
  - ◆ Example: A jpeg image of the Mona Lisa is not the same as the original painting. The relationship between the metadata for the original and the reproduction is part of the metadata description.
- The Dumb-down Principle
  - ◆ Qualification is supposed only to refine, not extend the semantic scope of a property
  - ◆ A client should be able to ignore any qualifier and use the value as if it were unqualified: the remaining element value must continue to be generally correct and useful for discovery.
- Appropriate values
  - ◆ metadata must be useful for discovery

## *Implementation of Dublin Core*

- There are two ways to connect metadata and resources
  - ◆ metadata can be embedded within the resource itself (e.g. in the <head> of HTML)
  - ◆ metadata can be stored in any kind of database with a link provided to the described resource
- Example of syntax for Dublin Core
  - ◆ HTML/XHTML
  - ◆ plain XML
  - ◆ RDF/XML
- Syntax rules
  - ◆ Elements and Qualifiers start with lowercase letters (e.g. title)
  - ◆ Encoding schemes start with upper case (e.g. URI, DCMIType)



## *XML schemes and Namespaces for Dublin Core*

- Dublin Core defines Namespaces for Metadata terms
  - ◆ **dc:** (e.g. <http://purl.org/dc/elements/1.1/creator>) is the legacy namespace variant for the 15 core elements
  - ◆ **dcterms:** (e.g., <http://purl.org/dc/terms/creator>) is the new namespace containing all the terms including also the core elements
- Implementers may freely choose to use the 15 core elements either in their legacy dc: variant or in the dcterms: variant depending on application requirements.
- Over time, however, implementers are encouraged to use the semantically more precise dcterms: properties



## ■ Elements are represented as XML elements *Dublin Core in XML* using namespaces dc or dcterms

- ◆ dc for the 15 core elements ,
- ◆ dcterms also for additional elements and refinements

## ■ More details can be represented using attributes:

- ◆ Attribute `xsi:type` for encoding schemes
- ◆ Attribute `xml:lang` for language

## ■ Examples:

```
<dcterms:modified xsi:type="dcterms:W3CDTF">
2001-07-18 </dcterms:modified>
<dc:title>Dublin Core in XML</dc:title>
<dc:subject xml:lang="en">information services,
public library networking</dc:subject>
<dc:subject xml:lang="de">Informationsdienste,
```



## *Example: Metadata in XML*

```
<?xml version="1.0"?>
<metadata xmlns="http://example.org/myapp/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:dcterms="http://purl.org/dc/terms/">

<dc:title> UKOLN </dc:title>
<dcterms:alternative>
UK Office for Library and Information Networking
</dcterms:alternative>
<dc:subject>
national centre, network information support, library community,
awareness, research, information services, public library networking,
bibliographic management, distributed library systems, metadata,
resource discovery, conferences,lectures, workshops
</dc:subject>
<dc:subject xsi:type="dcterms:DDC"> 062 </dc:subject>
<dc:subject xsi:type="dcterms:UDC"> 061(410) </dc:subject>
<dc:description>
UKOLN is a national focus of expertise in digital information management.
</dc:description>
<dc:description xml:lang="fr">
UKOLN est un centre national d'expertise dans la gestion de l'information digitale.
</dc:description>
<dc:publisher> UKOLN, University of Bath </dc:publisher>
<dcterms:isPartOf xsi:type="dcterms:URI"> http://www.bath.ac.uk/ </dcterms:isPartOf>
<dcterms:modified xsi:type="dcterms:W3CDTF"> 2001-07-18 </dcterms:modified>
</metadata>
```



## Dublin Core in HTML/XHTML

- Metadata should be embedded into the <head> section of an XHTML Web page using the <meta> and <link> elements
- Elements and their values

```
<meta name="DC.date" content="2001-07-18" />
<meta name="DC.title" content="Dublin Core" />
<meta name="DCTERMS.audience" content="students" />
```
- Refinements

```
<meta name="DCTERMS.modified" content="2001-07-18" />
```
- Encoding schemes

```
<meta name="DC.date" scheme="DCTERMS.W3CDTF" content="2001-07-18" />
```
- Linking to other resources

```
<link rel="DC.relation" href="http://www.dublincore.org/" />
```
- Namespaces, z.B..

```
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
<link rel="schema.DCTERMS" href="http://purl.org/dc/terms/" />
```



## Example: Meta-data in HTML

```
Quelltext von: http://knut.hinkelmann.ch/ - Mozilla Firefox
Datei Bearbeiten Ansicht Hilfe
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
"http://www.w3.org/TR/html4/loose.dtd">
<html>
<head>
  <meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
  <title>Homepage of Knut Hinkelmann</title>
  <link href="knut.css" rel="stylesheet" type="text/css">
  <link href="navigationbar.css" rel="stylesheet" type="text/css">
  <link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
  <link rel="schema.DCTERMS" href="http://purl.org/dc/terms/" />
  <meta name="DC.title" content="Homepage of Knut Hinkelmann" />
  <meta name="DC.subject" content="Management; Wissensmanagement; Wissensnutzung; Engineering" />
  <meta name="DC.type" scheme="DCTERMS.DCMIType" content="Text" />
  <meta name="DC.format" content="text/html" />
  <meta name="DC.format" content="10148 bytes" />
  <meta name="DC.identifier" scheme="DCTERMS.URI" content="http://knut.hinkelmann.ch" />
</head>
<body id="page" onload="init()">
```



## Dublin Core for Webpages: Extension of Firefox

Prof. Dr. rer. nat.  
**Knut Hinkelmann**  
Fachhochschule Nordwestschweiz  
Hochschule für Wirtschaft  
Riggenbachstrasse 16  
4600 Olten  
knut.hinkelmann@fhnw.ch  
++41 62 286 00 80 (direkt)  
++41 848 821 011 (CCC)  
++41 78 896 84 24 (mobil)

**Funktionen**

- Studiengangleiter *Wirtschaftsinformatik* an der FHNW
- Leiter Informations- und Wissensmanagement an der FHNW
- Wissenschaftlicher Beirat der *STEAG & Partner*

**Kompetenzschwerpunkt: Knowledge and...**

- Geschäftsprozessmanagement, Workflow-Management, Business Rules
- Informations- und Wissensmanagement
- Wissenstechnologien (Semantic Web, Ontologien, Topic Maps, Wissensverarbeitung)
- E-Government

Prof. Dr. Knut Hinkelmann

Information Retrieval and Knowledge Organisation - 5 Metadata

29

## Tools for DCMI

- Meta-data editors
- Automatic Extraction/Gathering of from HTML, Word etc. (e.g. DC-dot)
- Automatic production of metadata
- Conversion between metadata formats
- Integrated tool environments
- see <http://www.dublincore.org/tools/>

## ***Adpating DCMI for specific application***

- DCMI has general attributes with broad applicability
- DCMI can be adapted for specific applications
  - ◆ multiple subject elements
  - ◆ refinements
  - ◆ encoding schemes
    - define controlled vocabulary
    - knowledge structures (classification systems, ontologies)

