

Präsentation im Rahmen der Ringvorlesung:
"Ferienclub"

Metadaten

Wintersemester 05/06
HAW-Hamburg

Sven Schliesing
sven.schliesing@informatik.haw-hamburg.de

Überblick

- Einleitung: Definition, Einsatzgebiete
- Technologien: MARC 21, Dublin Core, Semantic Web, Prolog
- Mehrwert: Mehr als nur Suchen?
- Risiken: “Metacrap”
- Chancen
- Ausblick

Einleitung – Technologien – Mehrwert – Risiken – Chancen – Ausblick

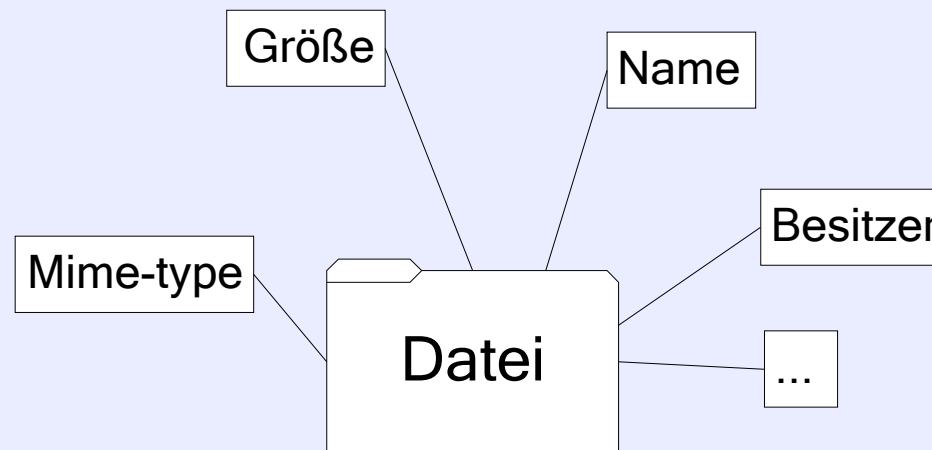
Einleitung

Einleitung – Technologien – Mehrwert – Risiken – Chancen – Ausblick

Sven Schliesing, HAW-Hamburg, Wintersemester 05/06

Definition

- Metadaten sind “Daten über Daten”



Definition

"Metadata is machine understandable information about web resources or other things"

Web architecture: Metadata, Tim Berners-Lee, Januar 1997

"Metakommunikation: die Kommunikation über Kommunikation. Metakommunikation liegt zum einen vor, wenn sich zwei Kommunikationspartner über ihre aktuelle Kommunikation austauschen und diese kommentieren."

Der Brockhaus - Psychologie, Brockhaus GmbH, 2001

Einleitung – Technologien – Mehrwert – Risiken – Chancen – Ausblick

Sven Schliesing, HAW-Hamburg, Wintersemester 05/06

Definition

- Metadaten eventuell auch:
 - Textdokument:
 - Überschrift?
 - Stichwörter aus Abstract?
 - Quellenliste?
 - Bild:
 - Histogramm?
 - Farbtiefe?
 - EXIF-Daten?

Definition

The screenshot shows a Windows-style dialog box titled 'Eigenschaften von Metadaten'. It has four tabs: 'Allgemein', 'Beschreibung', 'Benutzer', and 'Internet'. The 'Beschreibung' tab is selected. Inside the dialog, there are four labeled text input fields: 'Titel' (containing 'Metadaten'), 'Thema' (containing 'Metadaten und was man alles tolles über die erzählen kann'), 'Schlüsselwörter' (containing 'Metadaten, Meta, Daten, Blub, Bla, Fasel'), and 'Kommentar' (containing a multi-line text block). At the bottom, there are four buttons: 'OK', 'Abbrechen', 'Hilfe', and 'Zurück'.

Field	Value
Titel	Metadaten
Thema	Metadaten und was man alles tolles über die erzählen kann
Schlüsselwörter	Metadaten, Meta, Daten, Blub, Bla, Fasel
Kommentar	Dies ist ein Testkommentar damit das Feld nicht so leer ist. Und hier zwei Pangramme: - The quick brown fox jumps over the lazy dog. - Franz jagt im komplett verwahrlosten Taxi quer durch Bayern.

Einleitung – Technologien – Mehrwert – Risiken – Chancen – Ausblick

Definition: Einsatzgebiete

- Bibliotheken
- Wissensdatenbanken
- Suchmaschinen

Technologien

Einleitung – Technologien – Mehrwert – Risiken – Chancen – Ausblick

Sven Schliesing, HAW-Hamburg, Wintersemester 05/06

Technologien: MARC 21

- Machine-readable cataloging
- weit verbreitet
- im Einsatz seit 1968
- Standards: Z39.2, US. national standard for information interchange, ISO 2709

Technologien: MARC 21

```
01041cam 2200265 a 4500
001 ###89048230
003 DLC
005 19911106082810.9
008 891101s1990 maua j 001 0 eng
010 ## $a ###89048230
020 ## $a 0316107514 :
      $c $12.95
020 ## $a 0316107506 (pbk.) :
      $c $5.95 ($6.95 Can.)
040 ## $a DLC
      $c DLC
      $d DLC
050 00 $a GV943.25
      $b .B74 1990
082 00 $a 796.334/2
      $2 20
100 1# $a Brenner, Richard J.,
      $d 1941-
245 10 $a Make the team.
      $p Soccer :
      $b a heads up guide to super soccer! /
      $c Richard J. Brenner.
```

```
246 30 $a Heads up guide to super soccer
250 ## $a 1st ed.
260 ## $a Boston :
      $b Little, Brown,
      $c c1990.
300 ## $a 127 p. :
      $b ill. ;
      $c 19 cm.
500 ## $a "A Sports illustrated for kids
      book."
520 ## $a Instructions for improving soccer
      skills. Discusses dribbling, heading,
      playmaking, defense, conditioning,
      mental attitude, how to handle
      problems with coaches, parents,
      and other players, and the history
      of soccer.
650 #0 $a Soccer
      $v Juvenile literature.
650 #1 $a Soccer.
```

Technologien: MARC 21 - XML

```
<?xml version="1.0" encoding="UTF-8"?>
<collection xmlns="http://www.loc.gov/MARC21/slim">
  <record>
    <leader>01142cam 2200301 a 4500</leader>
    <controlfield tag="001"> 92005291 </controlfield>
    <controlfield tag="003">DLC</controlfield>
    <controlfield tag="005">19930521155141.9</controlfield>

    <controlfield tag="008">920219s1993 caua j 000 0 eng </controlfield>
    <datafield tag="010" ind1=" " ind2=" ">
      <subfield code="a"> 92005291 </subfield>
    </datafield>
    <datafield tag="020" ind1=" " ind2=" ">
      <subfield code="a">0152038655 :</subfield>
      <subfield code="c">$15.95</subfield>
    </datafield>
    <datafield tag="040" ind1=" " ind2=" ">
      <subfield code="a">DLC</subfield>
      <subfield code="c">DLC</subfield>
      <subfield code="d">DLC</subfield>
    </datafield>
    <datafield tag="042" ind1=" " ind2=" ">
      <subfield code="a">lcac</subfield>
    </datafield>
    <datafield tag="050" ind1="0" ind2="0">
      <subfield code="a">PS3537.A618</subfield>
      <subfield code="b">A88 1993</subfield>
    </datafield>
    <datafield tag="082" ind1="0" ind2="0">
      <subfield code="a">811/.52</subfield>
      <subfield code="2">20</subfield>
    </datafield>
    <datafield tag="100" ind1="1" ind2=" ">
      <subfield code="a">Sandburg, Carl,</subfield>
      <subfield code="d">1878-1967.</subfield>
    </datafield>
  </record>
  [...]
</collection>
```

Einleitung – Technologien – Mehrwert – Risiken – Chancen – Ausblick

Technologien: MARC 21 - MOD

```
<?xml version="1.0"?>
<mods xmlns:xlink="http://www.w3.org/TR/xlink" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns="http://www.loc.gov/mods/" xsi:schemaLocation="http://www.loc.gov/mods/ http://www.loc.gov/standards/mods/mods.xsd">
  <titleInfo>
    <title>Arithmetic </title>
  </titleInfo>
  <name type="personal">
    <namePart>Sandburg, Carl</namePart>
    <namePart type="date">1878-1967</namePart>
    <role><text>creator</text></role>
  </name>
  <name type="personal">
    <namePart>Rand, Ted</namePart>
    <role><text>ill.</text></role>
  </name>
  <typeOfResource>text</typeOfResource>
  <originInfo>
    <place>
      <code authority="marc">cau</code>
      <text>San Diego</text>
    </place>
    <publisher>Harcourt Brace Jovanovich</publisher>
    <dateIssued>c1993</dateIssued>
    <dateIssued encoding="marc">1993</dateIssued>
    <edition>1st ed.</edition>
    <issuance>monographic</issuance>
  </originInfo>
  <language authority="iso639-2b">eng</language>
  <physicalDescription>
    <form authority="marcform">print</form>
    <extent>1 v. (unpaged) : ill. (some col.) ; 26 cm.</extent>
  </physicalDescription>
  <abstract>A poem about numbers and their characteristics. Features anamorphic, or distorted, drawings which can be
restored to normal by viewing from a particular angle or by viewing the image's reflection in the provided Mylar
cone.</abstract>
  [...]
</mods>
```

Einleitung – Technologien – Mehrwert – Risiken – Chancen – Ausblick

Technologien: DC(MES)

- Dublin Core (Metadata Element Set)
- verabschiedet in 1995
- beinhaltet 15 Elemente mit deren Semantik und Definition

Technologien: DCMES – Element Set

contributor

coverage

creator

date

description

format

identifier

language

publisher

relation

rights

source

subject

title

type

Technologien: DCMES – Weitere Elemente

abstract

created

license

modified

valid

...

Einleitung – Technologien – Mehrwert – Risiken – Chancen – Ausblick

Technologien: MARC 21 - DC

```
<?xml version='1.0' ?>
<dc xmlns="http://purl.org/dc/elements/1.1/">
  <title>Arithmetic / </title>
  <creator>Sandburg, Carl, 1878-1967.</creator>
  <creator>Rand, Ted, ill.</creator>

  <type/>
  <publisher>San Diego :Harcourt Brace Jovanovich,</publisher>
  <date>c1993.</date>
  <language>eng</language>
  <description>A poem about numbers and their characteristics. Features anamorphic, or
    distorted, drawings which can be restored to normal by viewing from a particular
    angle or by viewing the image's reflection in the provided Mylar cone.</description>
  <description>One Mylar sheet included in pocket.</description>

  <subject>Arithmetic</subject>
  <subject>Children's poetry, American.</subject>
  <subject>Arithmetic</subject>
  <subject>American poetry.</subject>
  <subject>Visual perception.</subject>
</dc>
```

Technologien: MARC 21 vs./+ DC

- parallele Entwicklung
- gemeinsames Ziel: effiziente Suche/Archivierung
- MARC unterstützt komplexere und präzisere Suche
- sinnvoll: Koexistenz/Kombination

Technologien: BibTeX

- Programm zur Erstellung von Literaturangaben

```
@article{Gettys90,  
  author = {Jim Gettys and Phil Karlton and Scott McGregor},  
  title = {The {X} Window System, Version 11},  
  journal = {Software Practice and Experience},  
  volume = {20},  
  number = {S2},  
  year = {1990},  
  abstract = {A technical overview of the X11 functionality. This is an update  
    of the X10 TOG paper by Scheifler \& Gettys.}  
}
```

- Standard-Typen
 - @book
 - @mastersthesis
- Standard-Felder
 - organization

Einleitung – Technologien – Mehrwert – Risiken – Chancen – Ausblick

Technologien: Semantic Web

"For the Semantic Web, semantic indicates the the meaning of data on the Web can be discovered – not just by people, but also by computers"

Explorer's guide to the Semantic Web, Thomas B. Passin, 2004

- "Auszeichnung" für autonom arbeitende Systeme

Einleitung – Technologien – Mehrwert – Risiken – Chancen – Ausblick

Technologien: Semantic Web

"To find information, a Semantic Web approach would expect to go beyond keyword and alphabetical indexes to let users search by concepts and categories."

Explorer's guide to the Semantic Web, Thomas B. Passin, 2004

- Konzepte und Kategorien durch Metadaten ermittelbar?

Einleitung – Technologien – Mehrwert – Risiken – Chancen – Ausblick

Technologien: RDF

- Aussagen aufgebaut als Subjekt-Prädikat-Objekt Triple
- Beispiel:
 - Aussage: „Der **Autor** von <http://dietlweiss.de> ist **Tobias Dietl**“
 - RDF Statement in N-Triples Notation:
 - <<http://dietlweiss.de/>> subject
 - <<http://terms.org/author>> predicate
 - <<http://persons.org/TobiasDietl>> object
 - Bedeutung: <http://dietlweiss.de/> hat den [Autor Tobias Dietl](#)

Technologien: RDF

```
<?xml version="1.0"?>
<rdf:RDF      xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:terms="http://terms.org/">
  <rdf:Description rdf:about="http://dietylweiss.de/">
    <terms:author rdf:resource="http://persons.org/TobiasDietl" />
  </rdf:Description>
</rdf:RDF>
```

Technologien: Prolog

```
reads(john, 'XML.com').  
reads(mary, 'XML.com').  
reads(mary, xmlhack).  
reads(cristina, xmlhack).
```

```
?-reads(john, 'XML.com').  
    yes
```

```
?-reads(mary, X).  
    X = 'XML.com'  
    yes
```

```
?-reads(Person, 'XML.com').  
    Person = john;  
    Person = mary;  
    No
```

```
xml_insider(Insider) :-  
    reads(Insider, 'XML.com'),  
    reads(Insider, xmlhack).
```

Einleitung – Technologien – Mehrwert – Risiken – Chancen – Ausblick

Technologien: RDF + Prolog

- Objekt: 'http://www.xml.com/' und 'http://www.xmlhack.com/'
- Prädikat: 'http://www.xml.com/reads' (gekürzt)
- Subjekt: 'mailto:mary@xml.com', 'mailto:john@xml.com'
'mailto:christina@xml.com'

Prädikat	Subjekt	Objekt
reads	john	'XML.com'
reads	mary	'XML.com'
reads	mary	xmlhack
reads	cristina	xmlhack

Prädikat	Subjekt	Objekt
'http://www.xml.com/reads'	'mailto:john@xml.com'	'http://www.xml.com/'
'http://www.xml.com/reads'	'mailto:mary@xml.com'	'http://www.xml.com/'
'http://www.xml.com/reads'	'mailto:mary@xml.com'	'http://www.xmlhack.com/'
'http://www.xml.com/reads'	'mailto:cristina@xml.com'	'http://www.xmlhack.com/'

Technologien: RDF + Prolog

```
'http://www.xml.com/reads'('mailto:john+xml.com',  
                             'http://www.xml.com/').
```

```
?-'http://www.xml.com/reads'(Person,  
                               'http://www.xml.com/').  
    Person = 'mailto:john+xml.com'  
    Yes
```

```
rdf_triple('http://www.xml.com/reads',  
           'mailto:john+xml.com',  
           'http://www.xml.com/').
```

```
?-rdf_triple('http://www.xml.com/reads',  
             Person,  
             'http://www.xml.com/').  
    Person = 'mailto:john+xml.com'  
    Yes
```

```
reads(Person, Website) :-  
    rdf_triple('http://www.xml.com/reads',  
              Person,  
              Website).
```

Mehrwert

Einleitung – Technologien – Mehrwert – Risiken – Chancen – Ausblick

Sven Schliesing, HAW-Hamburg, Wintersemester 05/06

Mehrwert

- Verarbeitung/Abspeichern
- Aufbereitung/Generierung
- Anreicherung/Analyse/Interpretation
- Suche
- Auswertung/Statistik

Einleitung – Technologien – Mehrwert – Risiken – Chancen – Ausblick

Risiken

Einleitung – Technologien – Mehrwert – Risiken – Chancen – Ausblick

Sven Schliesing, HAW-Hamburg, Wintersemester 05/06

Risiken

- Zu viele Dokumente
 - google: 8 Milliarden+
 - yahoo: 3 Milliarden+
- Viele Dokumente nur für temporären Zweck
- Kontrolle über die Richtigkeit der Metadaten (“People lie”)

Risiken

- unzureichende Metadaten ("People are lazy")
- falsche/unvollständige Angaben ("People are stupid")
- subjektive Bewertung/Gewichtung/Benennung
 - "Schemas aren't neutral"
 - "There's more than one way to describe something"

"It's not a bald spot, it's a solar panel for a sex-machine."

unknown entity, unknown year

Chancen

Einleitung – Technologien – Mehrwert – Risiken – Chancen – Ausblick

Sven Schliesing, HAW-Hamburg, Wintersemester 05/06

Chancen

- Implizite Daten – schwer fälschbar?

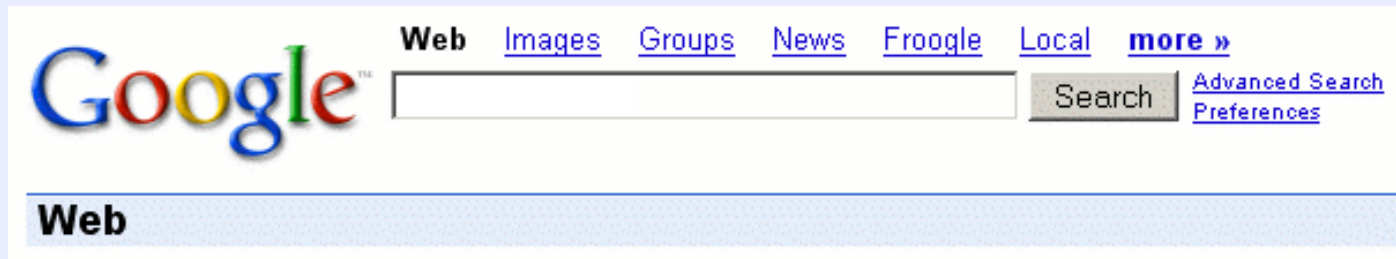
"Certain kinds of implicit metadata is awfully useful, in fact. Google exploits metadata about the structure of the World Wide Web: by examining the number of links pointing at a page [...]"

Metacrap: Putting the torch to seven straw-men of the meta-utopia, Cory Doctorow, August 2001

Einleitung – Technologien – Mehrwert – Risiken – Chancen – Ausblick

Chancen

- Google-Hacking



Einleitung – Technologien – Mehrwert – Risiken – Chancen – Ausblick

Ausblick

Einleitung – Technologien – Mehrwert – Risiken – Chancen – Ausblick

Sven Schliesing, HAW-Hamburg, Wintersemester 05/06

Ausblick

- Einsatz im Ferienclub
 - Sammeln der Metadaten verschiedenster Applikationen
 - Interpretation und Anreicherung der Daten
 - Suchfunktionen sowie Statistiken

Ausblick: Masterarbeit

- Richtungen:
 - Semantic Web?
 - RDF?
 - Prolog?
 - Psychologie?
 - OAIS?

Einleitung – Technologien – Mehrwert – Risiken – Chancen – Ausblick

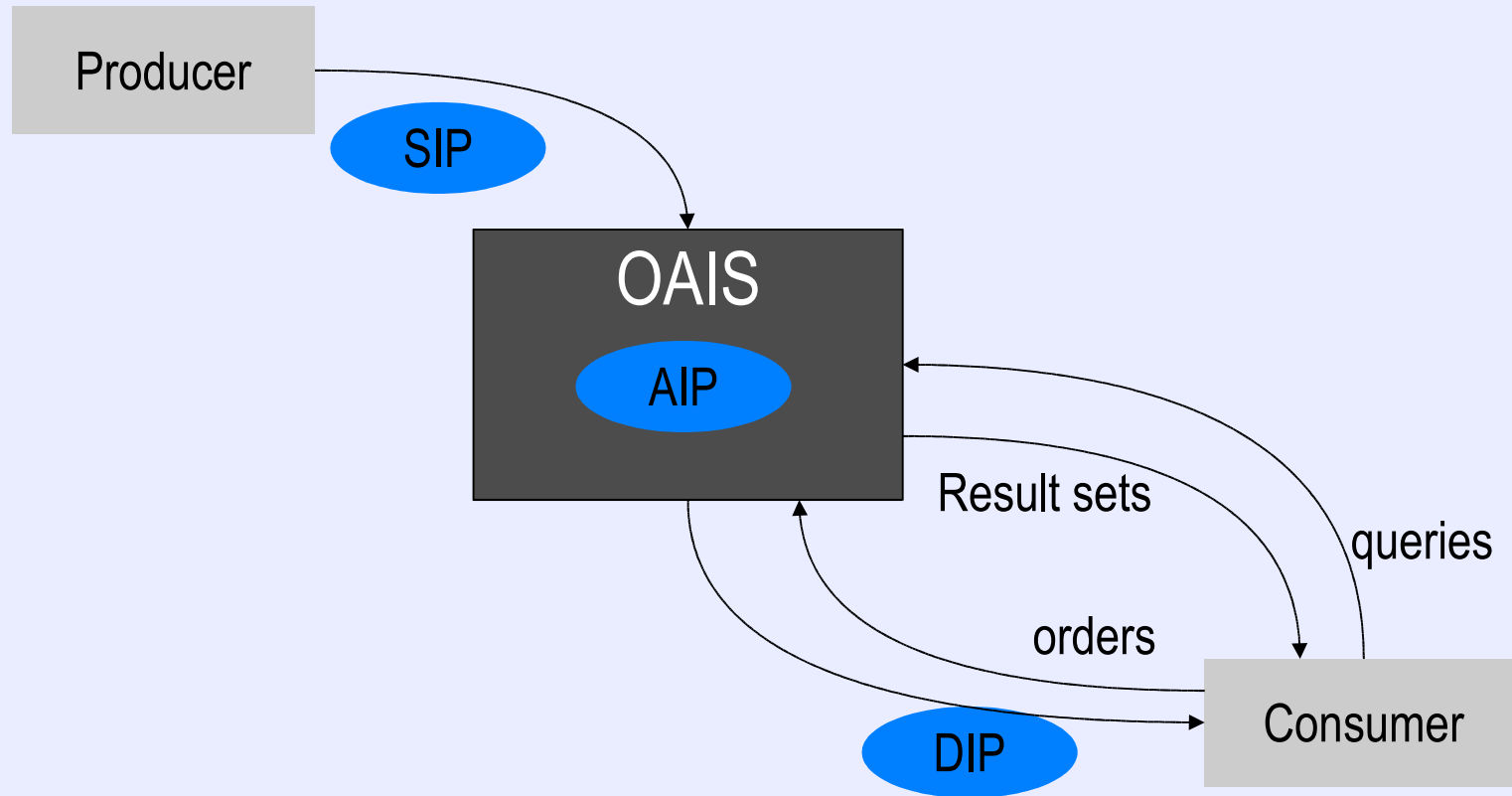
OAIS - Basics

- "Open Archival Information System"
- Entwickelt von CCSDS (Consultive Committee for Space Data Systems)
- ISO 14721 (2002)
- (offener) Standard sowie Modell
- anwendungsfähig und skalierbar

"Digital Information lasts forever – or for five years, whichever comes first."

Jeff Rothenberg

OAIS



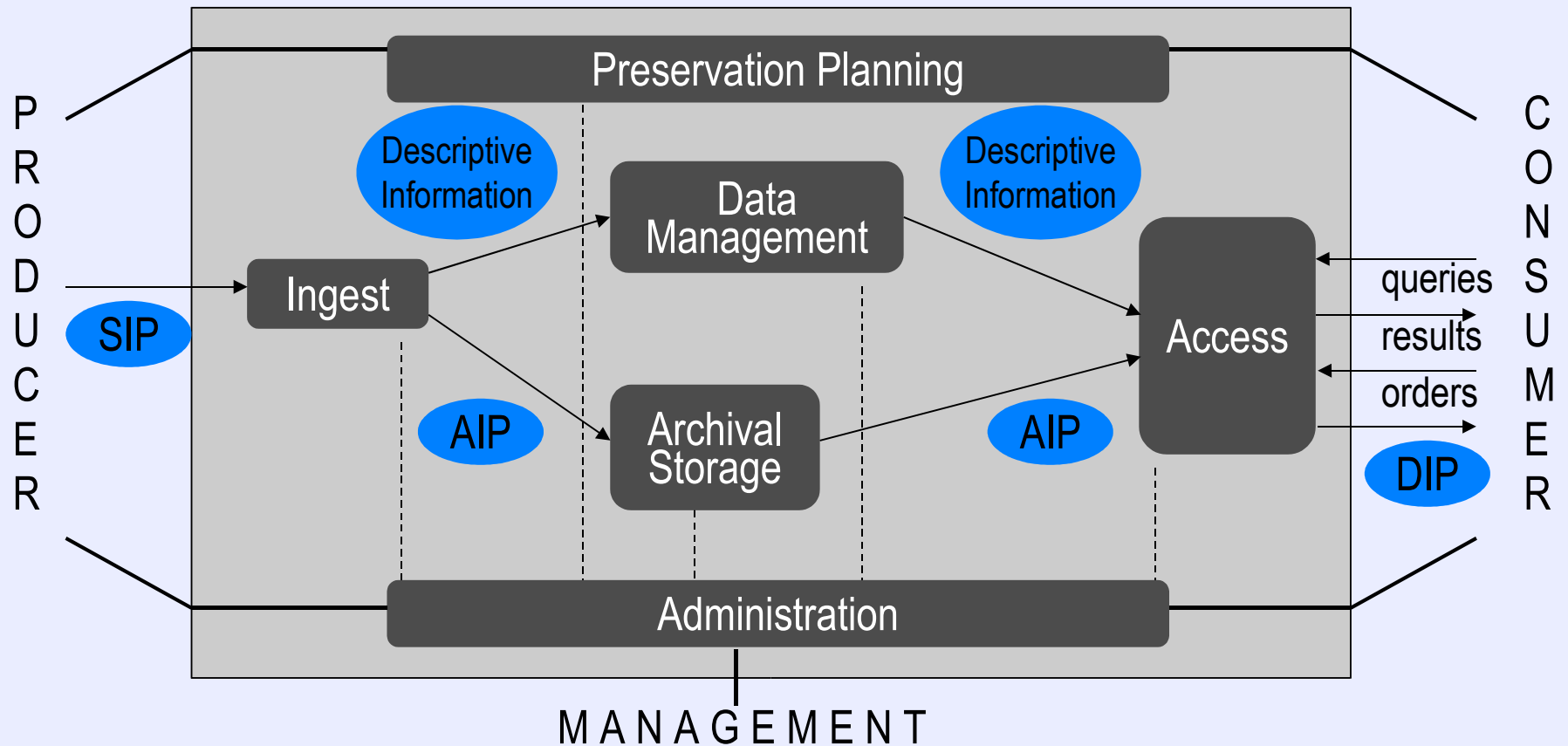
MANAGEMENT

Einleitung – Technologien – Mehrwert – Risiken – Chancen – Ausblick

OAIS – SIP, DIP, AIP

- SIP: Submission Information Package
- AIP: Archive Information Package
- DIP: Dissemination Information Package

OAIS



Einleitung – Technologien – Mehrwert – Risiken – Chancen – Ausblick

OAIS und Metadaten

- METS: Metadata Encoding & Transmission Standard
- beinhaltet z.B.:
 - “MARC”-Standard
 - DC-Komponenten
 - GDM (A Descriptive Metadata Extension Schema for the Metadata Encoding and Transmission Standard)
- METS Java Toolkit
- <http://www.loc.gov/standards/mets>

OAIS und Metadaten

OAIS + METS + x =
Das perfekte Archivierungs-System?

Einleitung – Technologien – Mehrwert – Risiken – Chancen – Ausblick

Quellen

- Einführung in Metadaten, März 2001: <http://www2.sub.uni-goettingen.de/intrometa.html>
- Cataloging the Web, Wayne Jones et al, Scarecrow Press 2002
- Web architecture: Metadata, Tim Berners-Lee, Januar 1997
- Der Brockhaus - Psychologie, Brockhaus GmbH 2001
- Explorer's guide to the Semantic Web, Thomas B. Passin, Manning, 2004
- An Introduction to Prolog and RDF, April 2001:<http://www.xml.com/pub/a/2001/04/25/prologrdf/index.html>
- Metacrap: Putting the torch to seven straw-men of the meta-utopia, August 2001: <http://www.well.com/~doctorow/metacrap.htm>
- <http://www11.informatik.tu-muenchen.de/lehre/seminare/seminarSW-SS2002/extension/sprachen.ppt>
- OAIS – Das „Open Archival Information System“: Ein Referenzmodell zur Organisation und Abwicklung der Archivierung digitaler Unterlagen, Nils Brübach, http://www.sachsen.de/de/bf/verwaltung/archivverwaltung/pdf/pdf_onlinepublikationen/pp_bruebach.pdf
- A Reference Model for an Open Archival Information System, Richard Ullmann, Juli 2004, <http://apan.net/meetings/cairns2004/presentation/escience-ullman.ppt>

Vielen Dank
für eure
Aufmerksamkeit!

Einleitung – Technologien – Mehrwert – Risiken – Chancen – Ausblick

Sven Schliesing, HAW-Hamburg, Wintersemester 05/06