

Institut für Dokumentologie und Editorik  
ildle

International Center for Archival Research  
ICARUS

## Digitalisierung, LZA und Metadaten

Markus Schnöpf

Folie 1 Digitalisierung, LZA und Metadaten Spring School 6. bis 10. Februar 2012

Institut für Dokumentologie und Editorik  
ildle

International Center for Archival Research  
ICARUS

## Technische Mindeststandards

- Digitalisierungsrichtlinien der Finanzgeber beachten (soweit vorhanden)
- Verwendungszweck beachten
  - Langzeitverfügbarkeit
  - Interoperabilität
- Beispiel DFG (Praxisregeln Digitalisierung)
  - <http://www.dfg.de/foerderung/programme/infrastruktur/lis/veroeffentlichungen/index.html>
  - Digital Library Federation
    - <http://www.diglib.org/standards/bmarkfn.htm>

Folie 2 Digitalisierung, LZA und Metadaten Spring School 6. bis 10. Februar 2012

Institut für Dokumentologie und Editorik  
ildle

International Center for Archival Research  
ICARUS

## Master-Kopie

- Mindestens 300dpi (bezogen auf das Format der zu digitalisierenden Quelle)
- Farbtiefe: mindestens 24bit
- TIFF unkomprimiert
- Verwendung eines Farbkeils
- Keine nachträgliche Qualitätsveränderungen

Folie 3 Digitalisierung, LZA und Metadaten Spring School 6. bis 10. Februar 2012

Institut für Dokumentologie und Editorik  
ildle

International Center for Archival Research  
ICARUS

## Scanner

- Flachbettscanner
- Diascanner
- Mikrofilmscanner
- Buchscanner
- Scanroboter
- 3D-Scanner
- Digitale Fotoapparate
- Röntgenscanner

- Falls ein Dienstleister die Digitalisierung übernimmt, werden die technischen Eckpunkte im Vertrag ausgeführt.

Folie 4 Digitalisierung, LZA und Metadaten Spring School 6. bis 10. Februar 2012

Institut für Dokumentologie und Editorik  
ildle

International Center for Archival Research  
ICARUS

## Vom Bild zum Volltext

- OCR (Optical Character Recognition)
  - Was bedeutet eine Erfassungsquote von 97%?
- Double Key Verfahren
  - Anbieter in Asien und Südosteuropa
- Kollaboratives Bearbeiten und Korrigieren von Quellenmaterial
  - Z.B. wikisource
- Transkription im Stüberl daheim

Folie 5 Digitalisierung, LZA und Metadaten Spring School 6. bis 10. Februar 2012

Institut für Dokumentologie und Editorik  
ildle

International Center for Archival Research  
ICARUS

## Strukturdaten (Druckwerke)

- Textsortenabhängige Gliederungsdaten
  - Kapitel, Unterkapitel, Jahressbände, Aufsätze
- Dazu kommen in beliebiger Erschließungstiefe
  - Inhaltsverzeichnisse, Register, Zeilenumbruch, Spaltenumbruch, Seitenumbruch, Kopfzeile / Fußzeile / Kolummentitel, Seitenzahl, Bilder oder bildähnliche Elemente, Bildunterschriften, Marginalien, Schriftwechsel, z.B. der Wechsel von Fraktur zu Antiqua (etwa für fremdsprachige Zitate), der Wechsel der Schriftgröße, Wechsel der Schriftart (recte, kursiv, fett usw.) u.a.m., Formeln, z.B. mathematische (MathML) oder chemische (CML) Formeln, Fortsetzungsmarkierungen (Kustoden) am Fußende von Seiten (für Anschlussbogen) etc.

Folie 6 Digitalisierung, LZA und Metadaten Spring School 6. bis 10. Februar 2012

Institut für Dokumentologie und Editorik  
ildle

International Center for Archival Research  
ICARUS

## Langzeitarchivierung (LZA) I

- Was bedeutet Langzeitarchivierung (LZA)?
  - Ansprüche an digitale Archive höher als an analoge Archive, aber im Prinzip ähnlich
  - Was zeichnet ein analoges Archiv aus?

Folie 7 Digitalisierung, LZA und Metadaten Spring School 6. bis 10. Februar 2012

Institut für Dokumentologie und Editorik  
ildle

International Center for Archival Research  
ICARUS

## Langzeitarchivierung (LZA) II

- Vertrauenswürdigkeit
- Langzeit = 5/6 Jahre (Zyklus einer Computergeneration)
- Möglichkeiten der Vergangenheitsherstellung
  - Simulation (Emulation)
  - Dateikonvertierung
    - Verlust des look and feel

Folie 8 Digitalisierung, LZA und Metadaten Spring School 6. bis 10. Februar 2012

Institut für Dokumentologie und Editorik  
ildle

International Center for Archival Research  
ICARUS

## Langzeitarchivierung (LZA) III

- Datenformate
  - Bilder: TIFF, PNG, SVG (verlustfreie Formate)
  - PDF-A
  - XML

Wichtig ist, dass offene Formate gewählt werden.  
Aus einem universitären Repository:  
Wir garantieren die Lesbarkeit für PDF-A Dokumente für einen Zeitraum von fünf Jahren, bei XML-Dokumenten fünfzig Jahre.

Folie 9 Digitalisierung, LZA und Metadaten Spring School 6. bis 10. Februar 2012

Institut für Dokumentologie und Editorik  
ildle

International Center for Archival Research  
ICARUS

## Langzeitarchivierung (LZA) IV

- Speichermedien / Bitstromerhaltung
  - optische Wechseldatenträger wie CD/DVD
  - Bandlaufwerke (streamer)
  - Festplatte
  - Mikrofilm
- Zwei Archivkopien, die an unterschiedlichen Orten gelagert werden

Folie 10 Digitalisierung, LZA und Metadaten Spring School 6. bis 10. Februar 2012

Institut für Dokumentologie und Editorik  
ildle

International Center for Archival Research  
ICARUS

## Langzeitarchivierung (LZA) V

- Doch woher weiß ich in 5 Jahren noch, was in den Dateien steckt? Und wie kann ich das Quellenmaterial rekonstruieren?
  - Deshalb Metadaten für Multimediadateien, (Meta-)Daten in TEI-Dokumenten
  - Interoperabilität von Daten durch Schnittstellen (APIs) und vereinheitlichten Datenformaten oder Data mapping
    - OAI-Schnittstelle, um Harvesting zu ermöglichen

Folie 11 Digitalisierung, LZA und Metadaten Spring School 6. bis 10. Februar 2012

Institut für Dokumentologie und Editorik  
ildle

International Center for Archival Research  
ICARUS

## Langzeitverfügbarkeit I

- Wissenschaft benötigt sichere Daten, Experimente müssen reproduzierbar sein
- Gedächtniswissenschaft benötigt gesichertes Quellenmaterial
- Digitale Quellen müssen referenzierbar sein
  - Persistent Identifiers wie URNs oder DOI
    - URN: Uniform Resource Name
    - DOI: Digital Object Identifier

Folie 12 Digitalisierung, LZA und Metadaten Spring School 6. bis 10. Februar 2012

Institut für Dokumentologie und Editorik  
 iIdle    International Center for Archival Research  
 ICARUS

## Langzeitverfügbarkeit II

- Lage in Österreich:  
 „URN [...] ermöglicht den Zugang zu Dokumenten unabhängig vom Speicherort, wird inzwischen in großen Umgebungen verwendet und ist daher für alle Einrichtungen, die mit der Aufbewahrung elektronischer Dokumente zu tun haben, von zunehmender Bedeutung [...]. In den letzten Wochen ist das schon in kleineren Zirkeln diskutierte Problem einer geeigneten österreichweiten Infrastruktur für URN nun virulent geworden, da einzelne Einrichtungen handeln müssen und sich auf europäischer Ebene eine übergreifende Initiative formiert. Freundlicherweise hat sich die Deutsche Nationalbibliothek bereit erklärt, uns mit Rat und Tat zur Seite zu stehen, sodass wir beste Voraussetzungen haben allerdings die letzten Schritte schon selbst machen müssen.“  
 Email vom 2.11.2011 der voeb-Mailingliste

Folie 13 Digitalisierung, LZA und Metadaten Spring School 6. bis 10. Februar 2012

Institut für Dokumentologie und Editorik  
 iIdle    International Center for Archival Research  
 ICARUS

## Was bedeutet das für Sie?

- Ihre wissenschaftlichen Arbeiten müssen in ihrer digitalen Form optimal vorbereitet sein, um
  - Langzeitarchiviert werden zu können
  - Lange Zeit verfügbar zu sein

Folie 14 Digitalisierung, LZA und Metadaten Spring School 6. bis 10. Februar 2012

Institut für Dokumentologie und Editorik  
 iIdle    International Center for Archival Research  
 ICARUS

## Metadaten

- deskriptive (meist bibliographische Beschreibung oder archivarische Erschließung)
  - Dublin Core, TEI-Header, EAD(-MOM), METS, ESE, EDM
- strukturelle (Text-, Dokumentstruktur)
  - TEI, CEI, METS
- administrative (z.B. Rechteverwaltung)
  - METS
- technische (z.B. Dateitypen)
  - METS

Folie 15 Digitalisierung, LZA und Metadaten Spring School 6. bis 10. Februar 2012

Institut für Dokumentologie und Editorik  
 iIdle    International Center for Archival Research  
 ICARUS

## Metadaten: DCMI

- Dublin Core Metadata Initiative
- entstanden 1994, um in der Vorgooglezeit Informationen im www zu erfassen.
  - stabil
  - Bibliothekarische Herkunft
  - Wenig Daten werden erfasst.
  - <http://dublincore.org/>

Folie 16 Digitalisierung, LZA und Metadaten Spring School 6. bis 10. Februar 2012

Institut für Dokumentologie und Editorik  
 iIdle    International Center for Archival Research  
 ICARUS

## Metadaten: EAD

- Erste Entwicklung des Schemas 1993 in Berkeley (CA)
  - Schema: <http://www.loc.gov/ead/eadschema.html>
  - Elemente: [http://www.loc.gov/ead/tglib/element\\_index.html](http://www.loc.gov/ead/tglib/element_index.html)
- EAD ist findbuchorientiert
- EAD ist nicht an mittelalterlichen und frühneuzeitlichen Urkunden orientiert.  
 „The group agreed that pure EAD did not fit the needs of charter description but that there is a strong need of exchanging data between the archives and MOM-CA in a format that has wide acceptance in the archival community.“ (Monasterium – Technical Meeting; Minutes, 6.11.2011).
  - <http://www.loc.gov/ead/>

Folie 17 Digitalisierung, LZA und Metadaten Spring School 6. bis 10. Februar 2012

Institut für Dokumentologie und Editorik  
 iIdle    International Center for Archival Research  
 ICARUS

## Metadaten: METS

- Metadata Encoding and Transmission Standard
- Schema: <http://www.loc.gov/standards/mets/mets.xsd>
- Beschreibung: [http://www.loc.gov/standards/mets/METSOverview.v2\\_de.html](http://www.loc.gov/standards/mets/METSOverview.v2_de.html)
- <http://www.loc.gov/standards/mets/>

Version 1.4 Revised  
 ©2010  
 Digital Library Federation

Folie 18 Digitalisierung, LZA und Metadaten Spring School 6. bis 10. Februar 2012

Institut für Dokumentologie und Editorik  
ildle

International Center for Archival Research  
ICARUS

## Metadaten: ESE

- Europeana Semantic Elements

„The Europeana Semantic elements (ESE) is the metadata set used to describe cultural heritage objects in Europeana. It is a Dublin Core-based application profile providing a generic set of DC elements and some locally coined terms, which have been added specifically to support Europeana's functionalities.“

- <http://www.europeana.eu/schemas/ese/>
- Schema:  
<http://www.europeana.eu/schemas/ese/ESE-V3.4.xsd>
- Erweiterung des Dublin Core Sets

Folie 19 Digitalisierung, LZA und Metadaten Spring School 6. bis 10. Februar 2012

Institut für Dokumentologie und Editorik  
ildle

International Center for Archival Research  
ICARUS

## Metadaten: EDM

- Europeana Data Model
- Verwendet Elemente anderer Namespaces und eigene
  - Resource Description Framework (RDF)
  - OAI Object Reuse and Exchange (ORE)
  - Simple Knowledge Organization System (SKOS)
  - Dublin Core (elements, terms and types)
  - Offenes System, erweiterbar und integrativ
  - <http://pro.europeana.eu/edm-documentation>

Folie 20 Digitalisierung, LZA und Metadaten Spring School 6. bis 10. Februar 2012

Institut für Dokumentologie und Editorik  
ildle

International Center for Archival Research  
ICARUS

## Metadaten: CEI

- Charter Encoding Initiative

„In April 2004, international scholars met in Munich. Their common purpose was to discuss the possibilities of a standard to encode medieval and early modern charters with XML. They finally built a preliminary working group called "Charters Encoding Initiative" to notify our intention to work continuously together, to spread our proposals in the scientific community and to integrate them into existing standards especially the guidelines of the TEI.“

- <http://www.cei.lmu.de/>
- Schema: <http://www.cei.lmu.de/schema/cei060122.xsd>

Folie 21 Digitalisierung, LZA und Metadaten Spring School 6. bis 10. Februar 2012

Institut für Dokumentologie und Editorik  
ildle

International Center for Archival Research  
ICARUS

## Metadaten: MOM-CA

- <http://www.monasterium.net/>
- <http://www.mom-ca.uni-koeln.de/mom/home>
- Erweiterung der EAD-Metadaten mit CEI Elementen
- Ziel: Offene Plattform für die Edition mittelalterlicher und frühneuzeitlicher Urkunden im Browser.
- Import auch von Excel-Tabellen mit deskriptiven Daten.

Folie 22 Digitalisierung, LZA und Metadaten Spring School 6. bis 10. Februar 2012

Institut für Dokumentologie und Editorik  
ildle

International Center for Archival Research  
ICARUS

## Metadaten: OAI-PMH

- Open Archives Initiative – Protocol for Metadata Harvesting
- Harvester können verschiedene Metadatenformate auslesen
- Metadaten austausch
- Datenprovider muss Informationen über sein repository liefern.
- <http://www.openarchives.org/pmh/>

Folie 23 Digitalisierung, LZA und Metadaten Spring School 6. bis 10. Februar 2012

Institut für Dokumentologie und Editorik  
ildle

International Center for Archival Research  
ICARUS

## Übung

- Welche(s) Metadatenformat(e) würden Sie für Ihr Projekt wählen?
- Erstellen Sie mit dem zugehörigen Schema eine Metadaten-Datei für Ihr Projekt.
  - Für ead siehe Anleitung nächste Folie.

Folie 24 Digitalisierung, LZA und Metadaten Spring School 6. bis 10. Februar 2012

- Some more things
- <http://www.mufi.info/>
  - Medieval Unicode Font Initiative
- [http://wiki.services.openoffice.org/wiki/Bibliographic/Software\\_and\\_Standards\\_Information#Web\\_Application\\_S](http://wiki.services.openoffice.org/wiki/Bibliographic/Software_and_Standards_Information#Web_Application_S)
- <http://www2.archivists.org/standards>
- <http://www.archivists.org/saagroups/ead/tools.html>
  - Ganz unten: oxygen instructions