Trusted Digital Archives. Experiences from the Landesarchiv Baden-Württemberg, nestor and DIN

Dr. Christian Keitel

Johannesburg, 27/2/2013

LANDESARCHIV
BADEN-WÜRTTEMBERG

Table of Contents

- Some fundamental concepts
- Solutions of the Landesarchiv Baden-Württemberg
- DIN-Standard 31644 and nestor-certification

LANDESARCHIV
BADEN-WÜRTTEMBERG

Trusted Traditional Archives

- Archives were old and hence trustworthy.
  - „Can I trust the archive?“ >>>
    Ius archivi/concept of unbroken custody (jurisprudence)

- Archival objects were regarded as authentic.
  - „Are the objects what they purport to be?“ >>>
    Examination of data carrier and information (diplomatics)

LANDESARCHIV BADEN-WÜRTTEMBERG

Trusted Digital Archives?

- Archives are young ~~were old and hence trustworthy~~.
  - „Can I trust the archive?“ >>> ~~Ius archivi/concept of unbroken custody (jurisprudence)~~

- Archival objects were regarded as authentic.
  - „Are the objects what they purport to be?“ >>> Examination of ~~data carrier and~~ information (diplomatics)

Report of the Task Force on Archiving of Digital Information (1995)

- "Repositories claiming to serve an archival function must be able to prove that they are who they say they are by meeting or exceeding the standards and criteria of an independently-administered program for archival certification."

Claiming trustworthiness by means of diplomatics.

Any more questions about the authenticity of digital objects?

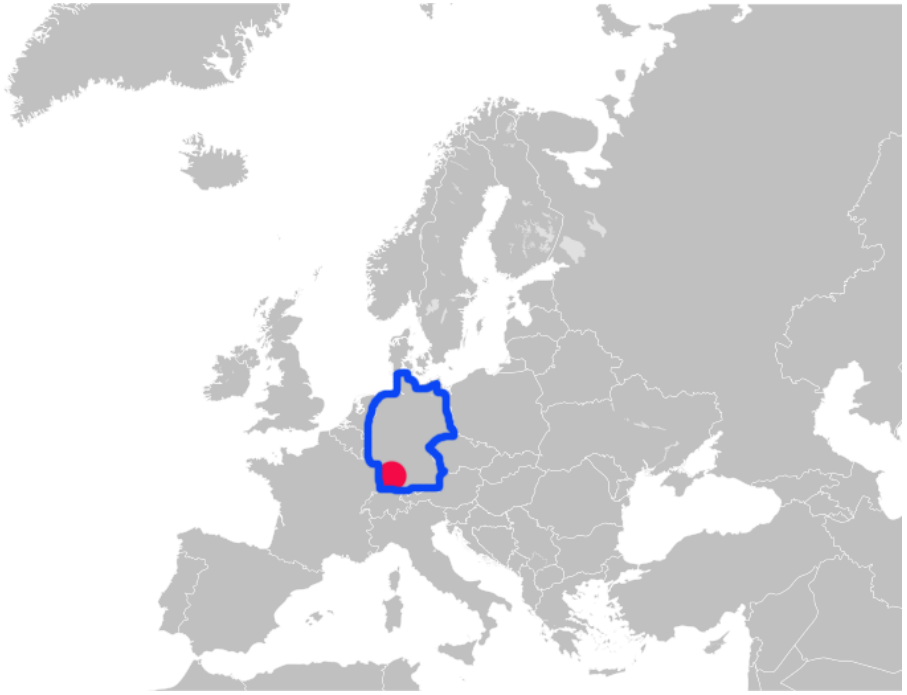
LANDESARCHIV BADEN-WÜRTTEMBERG

The New Landscape

- All users should get the opportunity to check the authenticity of archival objects (digital diplomatics)
  - Landesarchiv Baden-Württemberg
- All users should get information about the trustworthiness of the archive (criteria catalogues)
  - Nestor and DIN

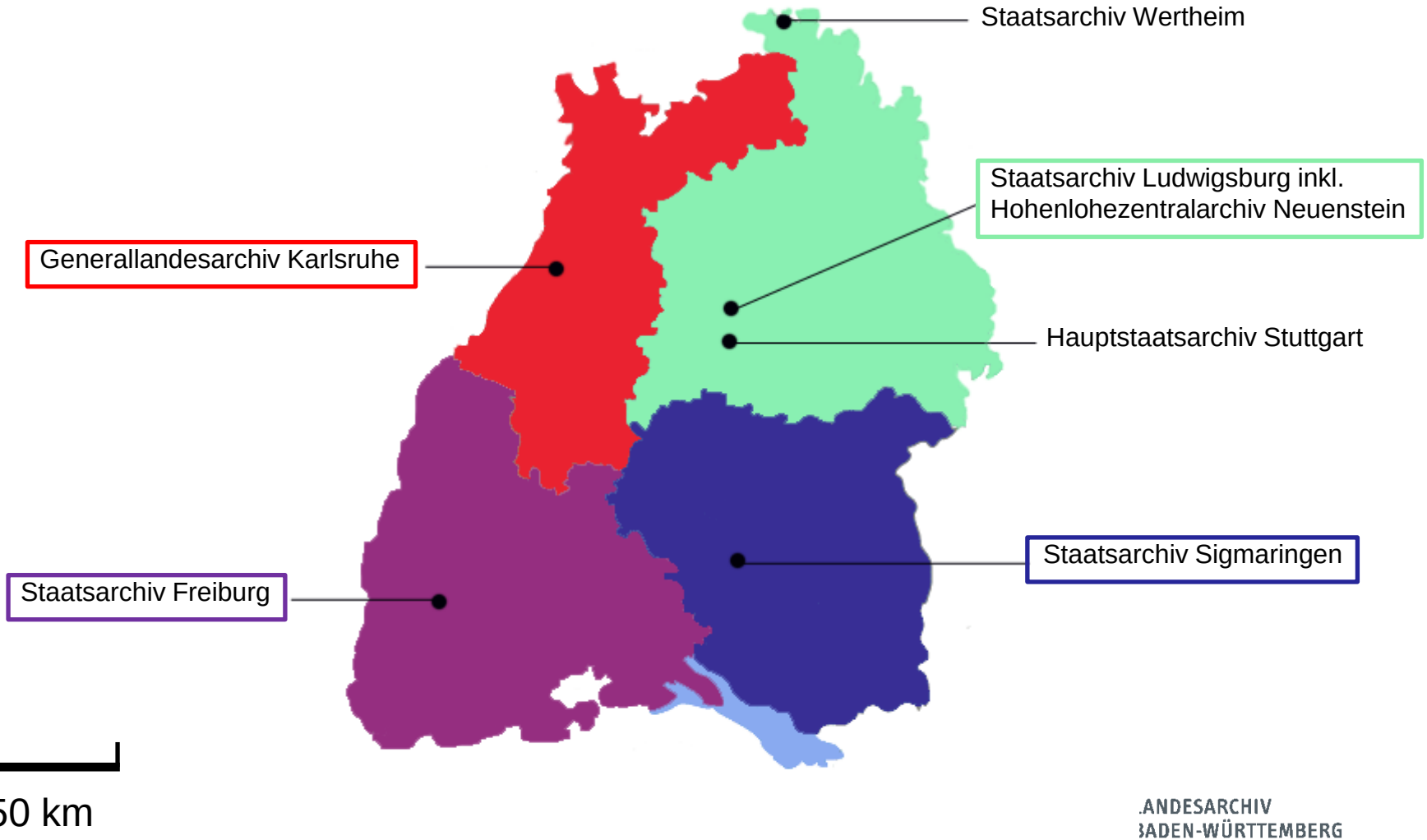
LANDESARCHIV
BADEN-WÜRTTEMBERG

Baden-Württemberg and Germany



LANDESARCHIV
BADEN-WÜRTTEMBERG

Landesarchiv Baden-Württemberg



LANDESARCHIV BADEN-WÜRTTEMBERG

Digital Preservation in Baden-Württemberg

- 1974 First 2-day workshop
- 1987 Archive law: Machine-readable information is equal to paper documents
- 2002 First digital object archived
- 2006 Programming of DIMAG

LANDESARCHIV BADEN-WÜRTTEMBERG

The Problem

- Digital preservation needs to store digital objects in a concrete physical way
  - Specific carrier
  - Specific file
  - Specific file format
  - …
- The physical manifestations have a short lifetime
- Transfers to other manifestations are unavoidable
- How to maintain authenticity?

LANDESARCHIV
BADEN-WÜRTTEMBERG

Enabling the User to Get His Questions Answered

- Task can‘t be postponed

- Digital preservation knows two states and four means to maintain authenticity
  - (1) Maintain integrity via hash values (frozen state)
  - (2) Maintain authenticity via metadata (transition state)
  - (3) Maintain a net of references
  - (4) Document what you do

LANDESARCHIV BADEN-WÜRTTEMBERG

(1) Hash Values (Frozen State)

- Each DIMAG file has its own hash-value-file
- DIMAG recalculates the hash values and compares them with the value in the hash-value-file
  - New files: every day
  - All files: every week

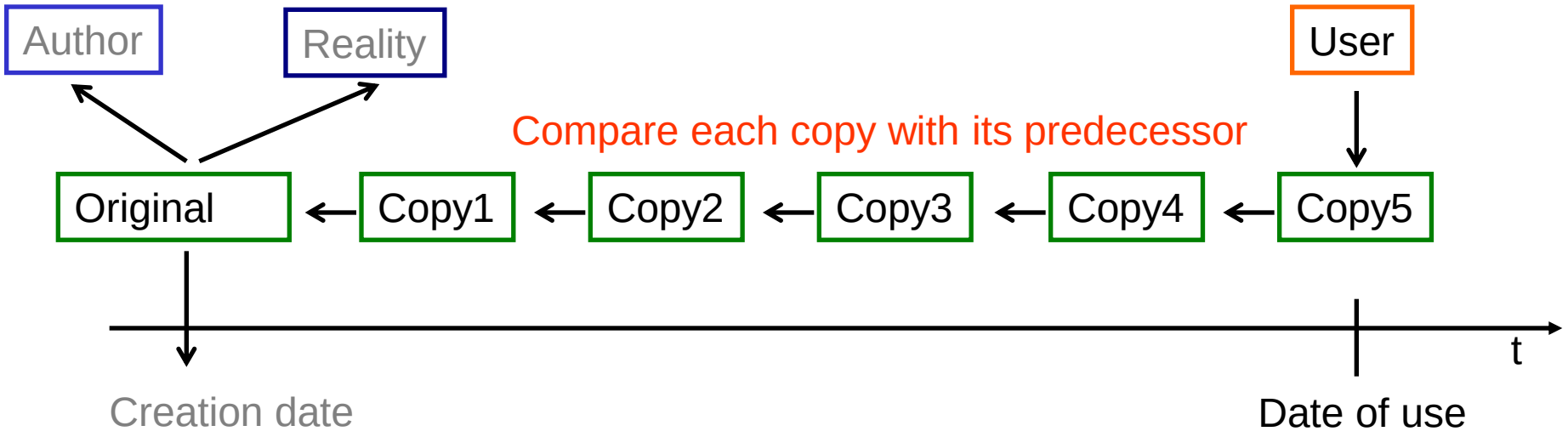
How We Obtain Historical Information

LANDESARCHIV
BADEN-WÜRTTEMBERG

How We Trace Back Our Digital History

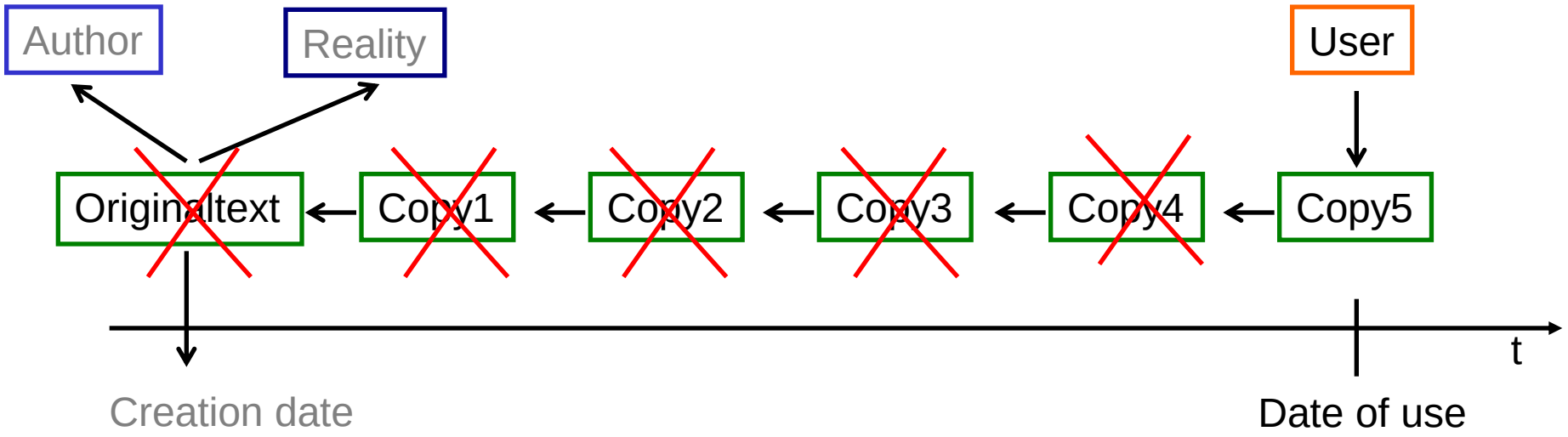
LANDESARCHIV
BADEN-WÜRTTEMBERG

Task...



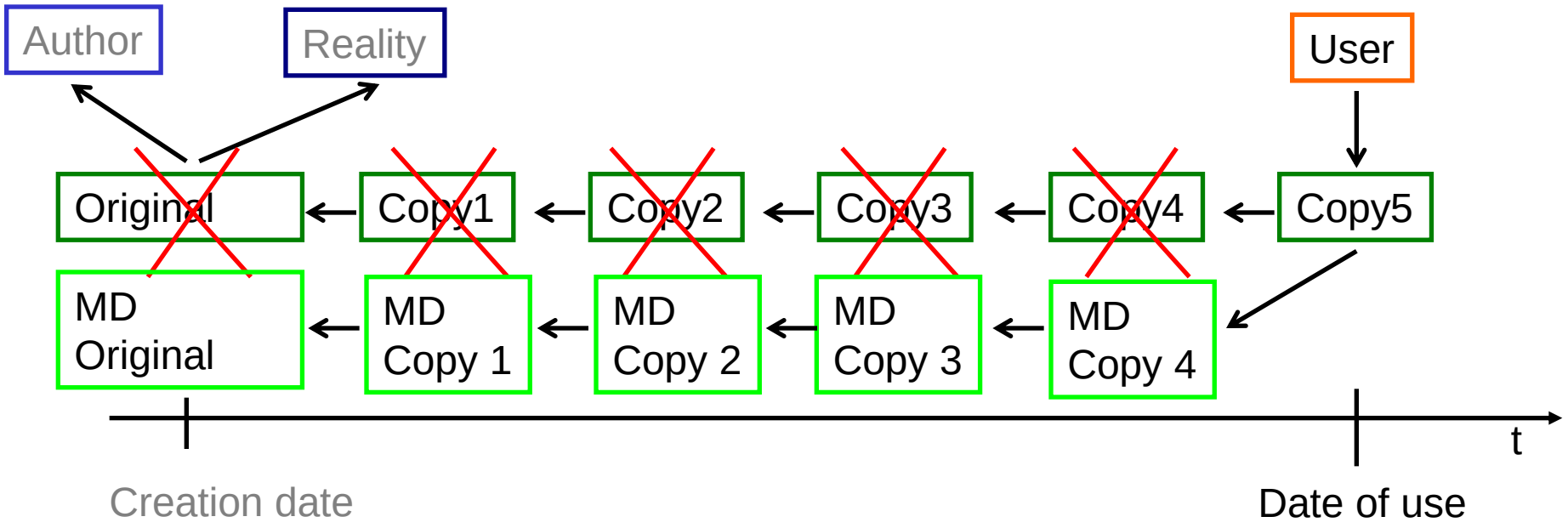
LANDESARCHIV
BADEN-WÜRTTEMBERG

...Problem...



LANDESARCHIV
BADEN-WÜRTTEMBERG

...and Solution

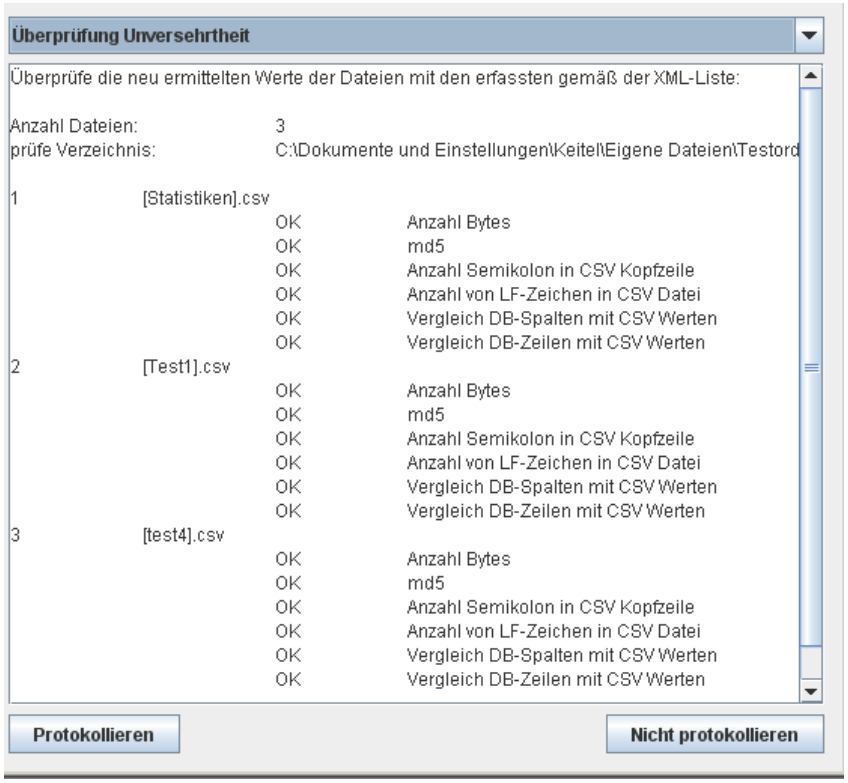


MD = Metadata

LANDESARCHIV BADEN-WÜRTTEMBERG

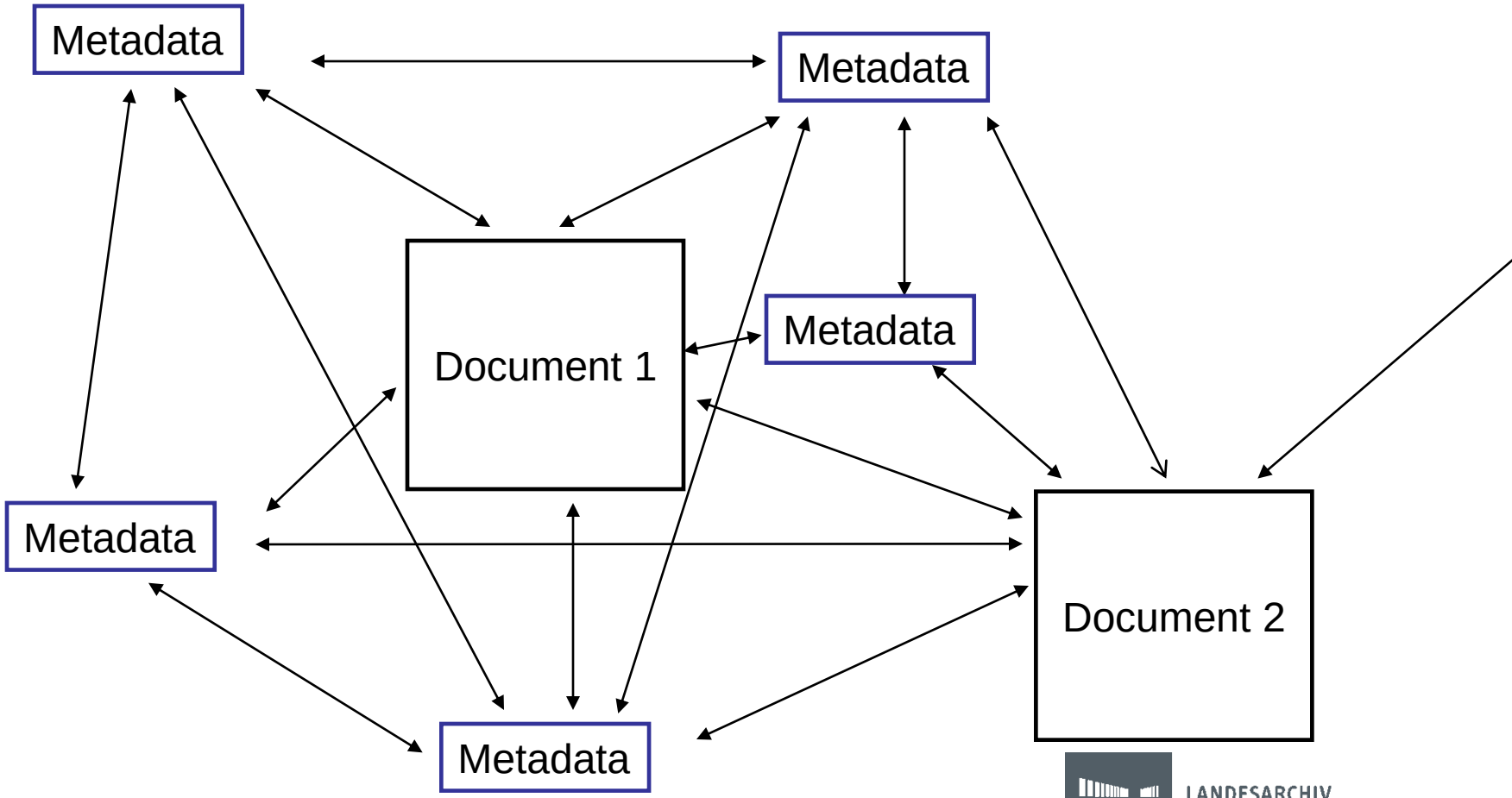
(2) Metadata (Transition State) - IngestList

- Catch central metadata of the objects to be archived (significant properties) as early as possible
  - Count fields and rows of the original data base table
  - Export tables as csv
  - Count rows and field delimiters of the csv
  - Compare metadata (number of rows and fields/field delimiters)



LANDESARCHIV BADEN-WÜRTTEMBERG

(3) Maintain a Net of References



LANDESARCHIV
BADEN-WÜRTTEMBERG

(4) Document What You Do - DIMAG

Protokoll-Metadaten

Prozess Ende	Prozess Ausführender	Bezug	Prozess	Nähere Angaben
2008-06-25 09:03	Dr. Kai Naumann	D A StAL EL 413_4 1 DO_1 1	Anlage Objekt	Digitales Objekt 'Daten aus Summenkarten' ist angelegt
2008-06-25 09:03	Dr. Kai Naumann	D A StAL EL 413_4 1 DO_1 1 R_1	Anlage Repräsentation	Repräsentation 'Festbreitenformat' ist angelegt
2008-06-25 09:03	Dr. Kai Naumann	D A StAL EL 413_4 1 DO_1 1 R_2	Anlage Repräsentation	Repräsentation 'CSV-Format' ist angelegt
2008-06-25 09:03	Dr. Kai Naumann	D A StAL EL 413_4 1 DO_1 1 R_3	Anlage Repräsentation	Repräsentation 'CSV-Format entschlüsselt' ist angelegt
2008-06-25 15:50:03	n	D A StAL EL 413_4 1 DO_1 1 R_2	Anlegen Dateiliste	Übernehme Dateien von Verzeichnis D:\Zug\LB EL\VZ61\DO_01_01-83\Rep_02 Verzeichnis . lese D:\Zug\LB EL\VZ61\DO_01_01-83\Rep_02\01_rep02.csv lese D:\Zug\LB EL\VZ61\DO_01_01-83\Rep_02\02_rep02.csv lese D:\Zug\LB EL\VZ61\DO_01_01-83\Rep_02\03_rep02.csv lese D:\Zug\LB EL\VZ61\DO_01_01-83\Rep_02\04_rep02.csv lese D:\Zug\LB EL\VZ61\DO_01_01-83\Rep_02\05_rep02.csv lese D:\Zug\LB EL\VZ61\DO_01_01-83\Rep_02\06_rep02.csv lese D:\Zug\LB EL\VZ61\DO_01_01-83\Rep_02\07_rep02.csv lese D:\Zug\LB EL\VZ61\DO_01_01-83\Rep_02\08_rep02.csv lese D:\Zug\LB EL\VZ61\DO_01_01-83\Rep_02\09_rep02.csv lese D:\Zug\LB EL\VZ61\DO_01_01-83\Rep_02\10_rep02.csv lese D:\Zug\LB EL\VZ61\DO_01_01-83\Rep_02\11_rep02.csv lese D:\Zug\LB EL\VZ61\DO_01_01-83\Rep_02\12_rep02.csv lese D:\Zug\LB EL\VZ61\DO_01_01-83\Rep_02\13_rep02.csv lese D:\Zug\LB EL\VZ61\DO_01_01-83\Rep_02\14_rep02.csv lese D:\Zug\LB EL\VZ61\DO_01_01-83\Rep_02\15_rep02.csv lese D:\Zug\LB EL\VZ61\DO_01_01-83\Rep_02\16_rep02.csv lese D:\Zug\LB EL\VZ61\DO_01_01-83\Rep_02\17_rep02.csv lese D:\Zug\LB EL\VZ61\DO_01_01-83\Rep_02\18_rep02.csv lese D:\Zug\LB EL\VZ61\DO_01_01-83\Rep_02\19_rep02.csv lese D:\Zug\LB EL\VZ61\DO_01_01-83\Rep_02\20_rep02.csv lese D:\Zug\LB EL\VZ61\DO_01_01-83\Rep_02\21_rep02.csv lese D:\Zug\LB EL\VZ61\DO_01_01-83\Rep_02\22_rep02.csv lese D:\Zug\LB EL\VZ61\DO_01_01-83\Rep_02\23_rep02.csv lese D:\Zug\LB EL\VZ61\DO_01_01-83\Rep_02\24_rep02.csv lese D:\Zug\LB EL\VZ61\DO_01_01-83\Rep_02\25_rep02.csv lese D:\Zug\LB EL\VZ61\DO_01_01-83\Rep_02\30_rep02.csv lese D:\Zug\LB EL\VZ61\DO_01_01-83\Rep_02\31_rep02.csv lese D:\Zug\LB EL\VZ61\DO_01_01-83\Rep_02\32_rep02.csv

LANDESARCHIV BADEN-WÜRTTEMBERG

(4) Document What You Do- DIMAG

- Internal procedures
- XML-schemes
- File format descriptions
- …

- DIMAG handles these documents as if they were genuine archival objects

LANDESARCHIV
BADEN-WÜRTTEMBERG

The International Landscape

- ISO 16363: 2012 – 109 criteria
  - Certification process by external experts (ISO 16919)
- DIN 31644: 2012 – 34 criteria
  - Extended peer-reviewed self-assessment or certification by external experts
- Data seal of approval – 16 criteria
  - Peer-reviewed self-assessment

- *European Framework for Audit and Certification of Digitial Repositories*
  http://www.trusteddigitalrepository.eu

LANDESARCHIV BADEN-WÜRTTEMBERG

The National Landscape

- Working groups at
  - nestor, the German competence network for digital preservation (founded in 2002)
  - DIN, the German Institute for Standardization

- Main developments
  - Criteria catalogue (DIN 31644, predecessor at nestor)
  - Certification versus DIN 31644 (nestor)

DIN 31644

- Purpose: Trustworthiness of the archive
- No compulsory technical implementation
- 4 principles: Documentation, transparency, adequateness, measurablity
- 34 criteria

C 1	Selection of information objects and representations
The digital archive has defined criteria for the selection of information objects and representations. The scope is limited by legal regulations, by the institution‘s mission, and by further objectives.	
Published criteria for the selection	

LANDESARCHIV BADEN-WÜRTTEMBERG

- C1 Selection of information objects and their representations
- C2 Responsibility for preservation
- C3 Designated communities
- C4 Access
- C5 Interpretability
- C6 Legal and contractual basis
- C7 Legal conformity
- C8 Funding
- C9 Personnel
- C10 Organisation and processes
- C11 Preservation measures
- C12 Crisis / successorship management
- C13 Significant properties
- C14 Integrity: Ingest interface
- C15 Integrity: Functions of the archival storage
- C16 Integrity: user interface
- C17 Authenticity: Ingest
- C18 Authenticity: Preservation measures
- C 19 Authenticity: Use
- C20 Technical authority
- C21 Transfer packages
- C 22 Transformation of the transfer packages into archival packages
- C 23 Archival packages
- C 24 Interpretability of the archival packages
- C 25 Transformation of archival packages into access packages
- C26 Access packages
- C27 Identification
- C28 Descriptive metadata
- C29 Structural metadata
- C30 Technical metadata
- C31 Logging the preservation measures
- C32 Administrative metadata
- C33 IT infrastructure
- C34 Security

- C1 Selection of information objects and their representations
- C2 Responsibility for preservation
- C3 Designated communities
- C4 Access
- C5 Interpretability
- C6 Legal and contractual basis
- C7 Legal conformity
- C8 Funding
- C9 Personnel
- C10 Organisation and processes
- C11 Preservation measures
- C12 Crisis / successorship management
- C13 Significant properties
- C14 Integrity: Ingest interface
- C15 Integrity: Functions of the archival storage
- C16 Integrity: user interface
- C17 Authenticity: Ingest
- C18 Authenticity: Preservation measures
- C 19 Authenticity: Use
- C20 Technical authority
- C21 Transfer packages
- C 22 Transformation of the transfer packages into archival packages
- C 23 Archival packages
- C 24 Interpretability of the archival packages
- C 25 Transformation of archival packages into access packages
- C26 Access packages
- C27 Identification
- C28 Descriptive metadata
- C29 Structural metadata
- C30 Technical metadata
- C31 Logging the preservation measures
- C32 Administrative metadata
- C33 IT infrastructure
- C34 Security

- C1 Selection of information objects and their representations
- C2 Responsibility for preservation
- C3 Designated communities
- C4 Access
- C5 Interpretability
- C6 Legal and contractual basis
- C7 Legal conformity
- C8 Funding
- C9 Personnel
- C10 Organisation and processes
- C11 Preservation measures
- C12 Crisis / successorship management
- C13 Significant properties
- C14 Integrity: Ingest interface
- C15 Integrity: Functions of the archival storage
- C16 Integrity: user interface
- C17 Authenticity: Ingest
- C18 Authenticity: Preservation measures
- C 19 Authenticity: Use
- C20 Technical authority
- C21 Transfer packages
- C 22 Transformation of the transfer packages into archival packages
- C 23 Archival packages
- C 24 Interpretability of the archival packages
- C 25 Transformation of archival packages into access packages
- C26 Access packages
- C27 Identification
- C28 Descriptive metadata
- C29 Structural metadata
- C30 Technical metadata
- C31 Logging the preservation measures
- C32 Administrative metadata
- C33 IT infrastructure
- C34 Security

LANDESARCHIV BADEN-WÜRTTEMBERG

The User‘s Questions

C 19	Authenticity: Use
The digital archive allows the user to evaluate the object’s authenticity and enables the administration of the archive to preserve it….	

Docun

C 31	Logging the preservation measures
The digital repository logs the preservation measures and any changes to the representations.	
Logs	

LANDESARCHIV
BADEN-WÜRTTEMBERG

Adequateness

- Implementations must be consistent with the archive's objective.

- An example: An archive wants to preserve just the text of web sites (no pictures, no sounds…)
  - For a political archive of East Asia studies, it might be sufficient.
  - For the archive of an academy of arts, it's not enough.

Nestor Certification - Overview

- Discussed since 2004
- Nestor working group on certification started in 2010
- Part of the European framework for certification of digital repositories
- Pilot certification of the German National Library
- Extended certification starts in 2013

LANDESARCHIV
BADEN-WÜRTTEMBERG

Extended Certification -Process

- Defining the scope
- All criteria applicable?
- Questions for each criterion
  - Current conditions?
  - Rating?
  - Documentation?
- Evaluation by 1st reviewer
- Fixing by 2nd reviewer

LANDESARCHIV
BADEN-WÜRTTEMBERG

C 21	Transfer packages

The digital repository has issued specifications regarding its transfer packages. The digital repository agrees with the producers on the transfer packages to be ingested (content data and metadata). The transfer packages are checked on the basis of the specifications.

- Which specifications does the digital repository have regarding transfer packages? Which content data are acceptable? Which metadata are required? Are there special requirements and processes for the creation of transfer packages?
- Which measures exist for validating the conformity of transfer packages?
- Will the repository reject defective transfer packages before ingesting, or will it take corrective measures within a defined work area?

LANDESARCHIV BADEN-WÜRTTEMBERG

Prospects

- Extended certification rollout in 2013
- The nestor working group on certification will address formal certification in 2013
- English translation of DIN standard and/or guidance document in 2013

LANDESARCHIV
BADEN-WÜRTTEMBERG

Questions?

- christian.keitel@la-bw.de
- ++49 711 212 4276

LANDESARCHIV
BADEN-WÜRTTEMBERG