

Ingest von Fachverfahren. Werkzeuge des Landesarchivs Baden-Württemberg

13. Tagung des AK Archivierung von Unterlagen
aus digitalen Systemen
27.4.2009, St. Gallen

Dr. Christian Keitel und Rolf Lang

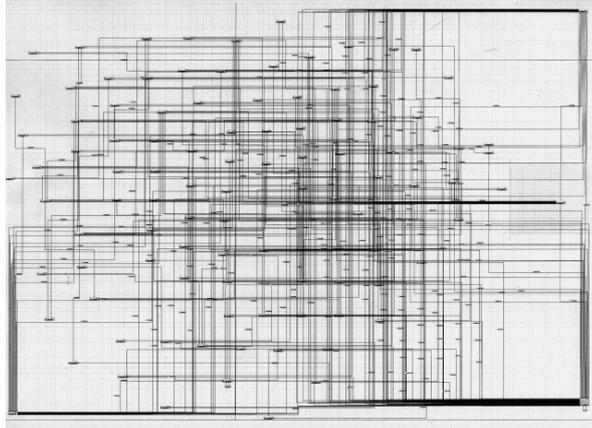


Übersicht

- Fachverfahren
- Standards
- Validierung
- Lösungsansatz des Landesarchivs Baden-Württemberg



Fachverfahren



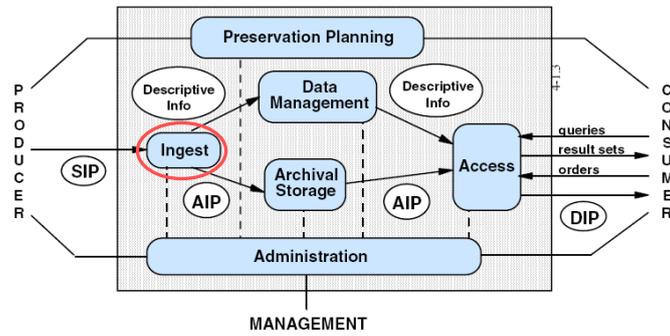
3 Bewertung und Übernahme

Fachverfahren

- IT-Systeme sind konstitutiv für die Datenpräsentation, können aber nicht übernommen werden
- Trennung von Daten und Systemen unabdingbar
- Geringe Normierung
- Minimale Einflussmöglichkeiten der Archive bei der Systemeinführung
- Problemkreise: Prozesse, Authentizität, Dokumentation

4 Bewertung und Übernahme

Ingest-Prozesse in OAIS (einfach)



5

Ingest-Prozesse in OAIS (weniger einfach)

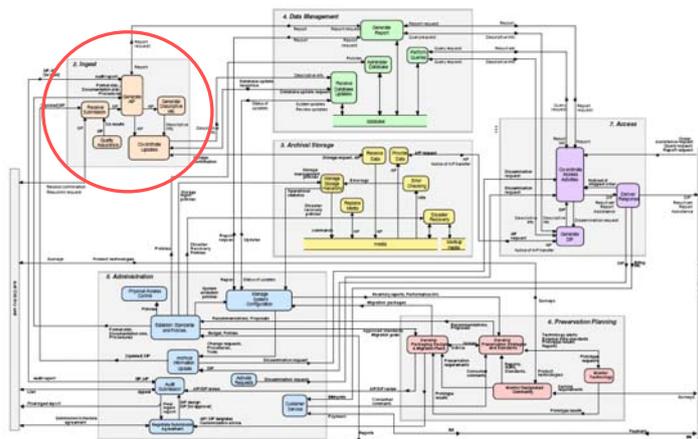
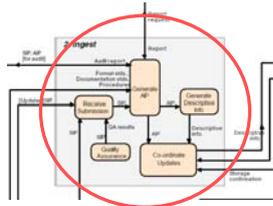


Figure F-1: Composite of Functional Entities

6

Ingest-Prozesse in OAIS and PAIMAS (ziemlich einfach)



- Preliminary Phase: First Contact
- Preliminary Phase: Information to Be Archived
- Preliminary Phase: Digital Objects and Standards Applied to These Objects
- Preliminary Phase: Object References
- Preliminary Phase: Quantification
- Preliminary Phase: Security Conditions
- Preliminary Phase: Legal and Contractual Aspects
- Preliminary Phase: Transfer Operations
- Preliminary Phase: Validation
- Preliminary Phase: Schedule
- Preliminary Phase: Permanent Impact On the Archive
- Preliminary Phase: Summary of Costs, Risks
- Preliminary Phase: Critical Points
- Preliminary Phase: Establishment of a Preliminary Agreement
- Formal Definition Phase: Organization of the Formal Definition Phase
- Formal Definition Phase: General Project Context and Definition of Information Objects
- Formal Definition Phase: Creation of a Data Dictionary
- Formal Definition Phase: Construction of a Formal Model
- Formal Definition Phase: Formalization of Contractual and Legal Aspects
- Formal Definition Phase: Definition of Transfer Conditions
- Formal Definition Phase: Validation Definition
- Formal Definition Phase: Delivery Schedule
- Formal Definition Phase: Change Management After Completion of the Submission Agreement
- Formal Definition Phase: Feasibility, Costs and Risks Assessment
- Formal Definition Phase: Submission Agreement
- Transfer Phase: Carry Out the Transfer Test
- Transfer Phase: Manage the Transfer
- Validation Phase: Carry Out the Validation Test
- Validation Phase: Manage the Validation

7

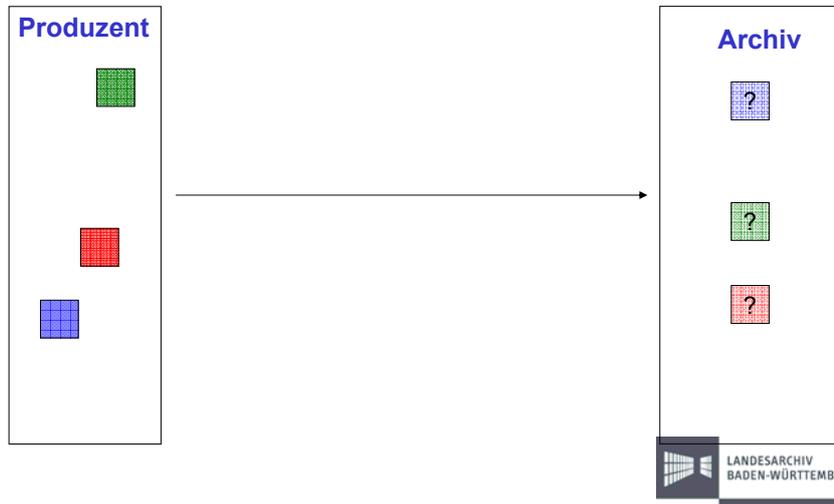
Ingest-Prozesse in PAIMAS (15 Schritte von 88)

- First Contact
- Information to Be Archived
- Digital Objects and Standards Applied to These Objects

- P-1 Identify the contact persons and work organization
- P-2 Exchange of general information
- P-3 Identify the Content Information to be preserved
- P-4 Identify the complementary information
- P-5 Identify the Designated Community
- P-6 Define Consumer access to the information
- P-7 Assess the planned duration of the preservation of this information by this Archive
- P-8 Assess the feasibility and costs induced by the previous actions
- P-9 Make a preliminary identification of the Data Objects related to the different categories of information to be archived
- P-10 Define the rules and standards related to these objects that are accepted by the Archive
- P-11 Describe the tools available for the application of the rules and standards known by the Archive
- P-12 Provide the rules and standards applied to Data Objects by the Producer
- P-13 Describe the tools available for application of the rules and standards known by the Producer
- P-14 Assess the compatibility and study solutions
- P-15 Assess the efforts to be made and the associated costs

8

Glaubwürdigkeitslücke bei der Übernahme

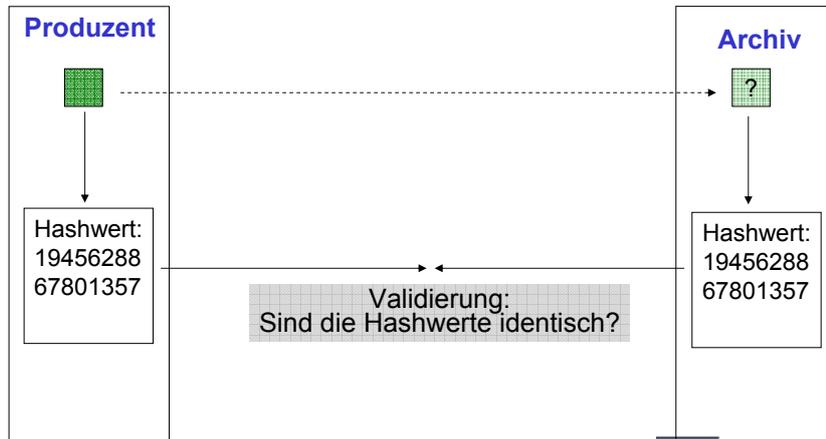


Validierung

- Nach einem Transfer zeigt die Validierung, ob die Objekte das enthalten, was erwartet wird.
- Das Zielobjekt wird dafür verglichen
 - mit dem Ausgangsobjekt
- oder
 - mit formalen oder inhaltlichen Vorgaben

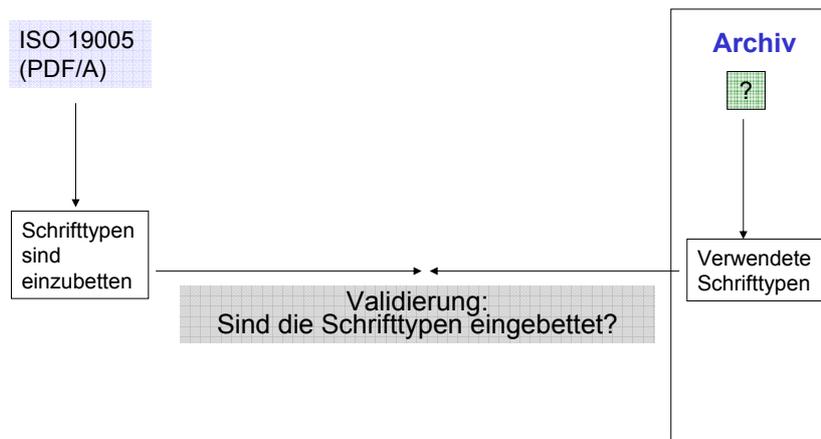
10

Vergleich Zielobjekt - Ausgangsobjekt



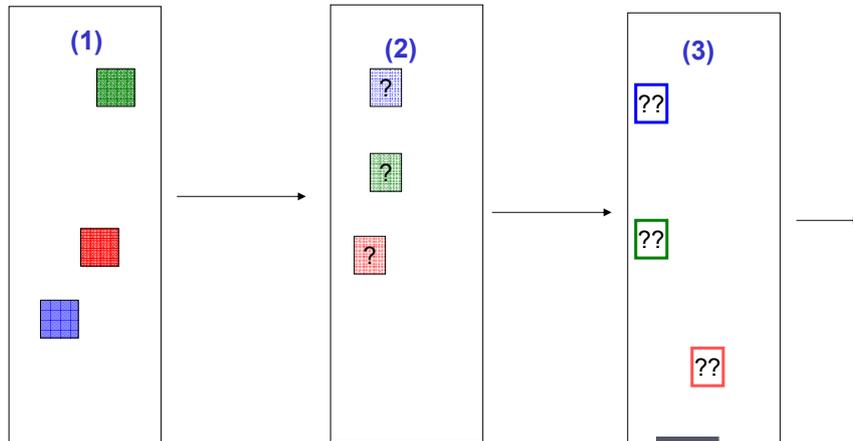
11

Vergleich Zielobjekt - Vorgabe



12

Authentizität



13

Zwischenergebnis

- Validierung nach der Übernahme bedingt eine frühzeitige Dokumentation in der Behörde.
- Validierung ist durch den gesamten Erhaltungsprozess hindurch immer wieder erforderlich.
- PAIMAS ist zu komplex und zu starr in der Abfolge der Schritte.

14

Aktionstypen

- **Archivische Bewertung**
- **Bestandsaufnahme:** Zentrale Eigenschaften einer Datei oder eines Objekts werden erhoben.
- **Transfer:** Übertragung von Informationen auf einen anderen Datenträger oder ein anderes Datei- oder Zeichenformat.
- **Validierung:** Überprüfung der Transferergebnisse.

15

Aktionen und Prozessschritte

Aktionstypen	Ingest 1	Ingest 3	PAIMAS
Bewertung	1	1	1 ...
Bestandsaufnahme	2	2	2 ...
Transfer	3	3	3 ...
Validierung	4	4	...
	5	5	85 ...
	6		86 ...
	7		87 ...
			88 ...

16

Ein Beispiel



17

IngestList

- Dokumentiert die Zustände während des Ingest-Prozesses.
- Ermöglicht eine Validierung der Zustände.
- Protokolliert die Ergebnisse der Validierung
Protokoll einer Archivalie setzt bereits in der Behörde ein.
- Kann als frei heruntergeladen und eingesetzt werden.

18

Download

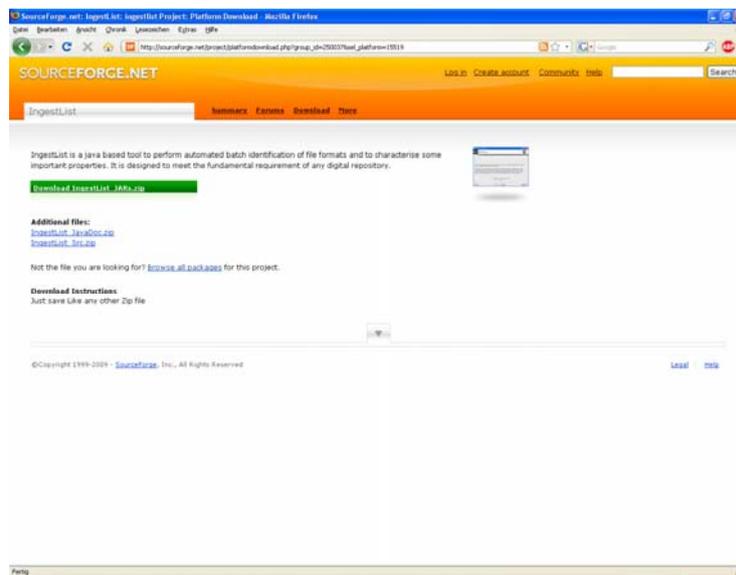
Wiki:

<http://ingestlist.wiki.sourceforge.net/>

Project:

<http://sourceforge.net/projects/ingestlist/>

19



20

SourceForge.net: IngestList Files Results | Firefox

Home | Register | Login | Search | Account | Help

SourceForge.NET

You have selected the 2009-04-27 release.
Please choose the file that best matches your architecture or operating system from the list of files contained in the release.

Package	Release	Filename	Size	Architecture
IngestList				
Latest	2009-04-27	(2009-04-29 11:54)		
		IngestList_Web.zip	6637330	none
		IngestList_JavaDoc.zip	118051	none
		IngestList_Src.zip	6570246	none
Totals:	1	3	13325627	

©Copyright 1999-2009 - SourceForge, Inc., All Rights Reserved

LandesARCHIV BADEN-WÜRTTEMBERG

21

Funktionen

1. Erkennen und Festhalten von Primär- und signifikanten Metadaten an der Quelle
2. Verwendung von jHove und DROID Methoden
3. Sichere, vollständige und überprüfbare Übertragung auf md5 / XML Basis
4. Unterstützung des Formierens und der Migration
5. Anbindung an ein Langzeitarchiv

22

technische Details

- Plattform übergreifend, da Java basiert
- Entwickelt mit NetBeans IDE 6.5
- Grafische Oberfläche für Bedienung durch Behörde / Archivar
- Kommando Zeilen Version für die Automatisierung

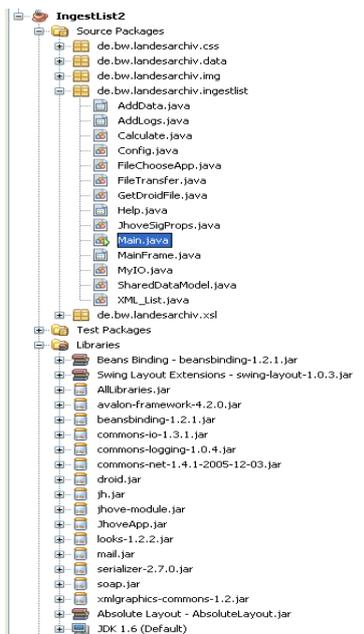
23



Projekt Struktur

Eigenentwicklung
main
JhoveSigProps
...

Importierte Libraries
sun
jhove
droid
...



24



de.bw.landesarchiv.ingestlist (IngestList2) - Mozilla Firefox

Package [Class](#) [Use Tree](#) [Deprecated](#) [Index](#) [Help](#)

PREV PACKAGE NEXT PACKAGE

Package de.bw.landesarchiv.ingestlist

Class Summary

AddData	add Data into jTable, check and transfer
AddLogs	handle the log files
Calculate	For one file create MD5, signatures droid Formats
Config	Loads an external or internal XML document and provides methods to access the data
FileChooseApp	Open and save files, using a JFrame dialog
FileTransfer	Handles the filetransfer.
GetDroidFile	This will refresh the PRONOM signatures.
Help	Help as JDialog form
JHoveSigProps	Class to get the JHove significant properties.
Main	This is the main entry for IngestList2. It will parse for arguments and starts either the CLI or the GUI version.
MainFrame	This is the Main entry for the GUI.
MyIO	Some IO Utilities

Package [Class](#) [Use Tree](#) [Deprecated](#) [Index](#) [Help](#)

PREV PACKAGE NEXT PACKAGE

Fertig

LANDESARCHIV BADEN-WÜRTTEMBERG

25

Calculate (IngestList2) - Mozilla Firefox

analyseText

```
public java.lang.String analyseText ()

Try to analyse the codepage.
Process the following rules in sequence:

Rule0: is identified by BOM
return BOM Code

Rule1: nothing above 0x80
nothing below 0x20, except of Tab 9, LineFeed 10, CarriageReturn 13.
return "ASCII"

Rule2: check Sequence of Bytes
0xxx xxxx
110x xxxx    followed by 10xx xxxx
1110 xxxx    followed by 10xx xxxx 10xx xxxx
1111 0xxx    followed by 10xx xxxx 10xx xxxx 10xx xxxx
return "UTF-8"

Rule3: nothing below 0x20, except of Tab 9, LineFeed 10, CarriageReturn 13.
amount of SPACES > amount of $
Note: if EBCDIC Code, $ is the SPACE:
return "Latin-1"

Rule4: nothing below 0x90, except of HT 5, SHM 10, VT 11, FF 12, CR 13, NL 21, BS 22, LF 27
return "EBCDIC"

Rule5: number of Hex 00 >= 1/3 filesize
return "UTF16"

Rule6:
return ""

Returns:
A String for the analysed Text, which is BOM identified, ASCII, UTF-8, Latin-1, EBCDIC, UTF-16
```

Fertig

LANDESARCHIV BADEN-WÜRTTEMBERG

26

IngestList Demo

Abgabe

INGEST

Archiv