Into the Archive

Potential and Limits of Standardizing the Ingest

iPRES,
San Francisco, 6th of Oct 2009

Jens Ludwig
ludwig@sub.uni-goettingen.de
Beagrie et al 2008: Ingest is biggest cost factor

It is the first step...

and nearly always the first question of interested people!

„Garbage in – Garbage out“: crucial for quality

So what about a standard for ingest?
complexity of ingest

• Is it possible to have a single ingest standard for long-term preservation?

• The variety of data sources is nearly endless and changing (intell., tech. and org. context)

• The usage contexts are nearly unforeseeable

• Main tasks are not tech., but org.

• A single standard, interface, procedure, etc? No. What is the alternative?
basic idea of guide

- Streamline the ingest by providing an introduction and common working basis
  - for archives and information producers/providers
  - for planning the ingest

- nestor working group for standards
  - consists of representatives of libraries and archives

- "Into the Archive - A guide for the information transfer to a digital repository"
  - public draft
  - funding by German Federal Ministry of Education, Federal Ministry for Economics and Technology and the DIN, the German Institute for Standardization.
  - http://nbn-resolving.de/urn:nbn:de:0008-20080710002
main design principle 1

• Do not use the OAIS terminology

• Correct OAIS terminology is too difficult for an introduction (e.g. „Information Object“, „Transformational Information Property“)

• But the merits of the OAIS are undeniable, therefore special attention was paid to be compatible to it
main design principle 2

- Define ingest as transfer of responsibility
  - and not as a technical transfer

- OAIS defines "ingest" only as a functional entity of an archive and not as a process

- The nestor guide excludes some activities of the OAIS (archive internals)
  - and includes additional ones (interactions between archive and producer like appraisal or the definition of authenticity requirements)
main design principle 3

- Organize the process in manageable parts

- Sacrifice comprehensiveness (good enough instead of everything)

- no strict sequences and instructions because of real-life variety and complex dependencies

- So: PAIMAS already exists but it is with its 87 steps too granular and daunting for non-experts
structure

• Objects
  • Information to be archived (intell. and tech.)
  • Metadata
  • Significant properties

• Processes
  • Transfer packages
  • Validation
  • Transfer

• Management
  • Laws and contracts
  • Ingest agreement and documentation
  • Costs and staff (TODO)
First feedback and lessons

- First discussions with potential users have shown demand and interest
- Especially brevity was regarded as positive
- Insight gained in exercises: relevance of people involved (expertise, awareness and common goal)
future work

• Include the viewpoints of more domains (e.g. museum, industry, science)

• Provide more examples, rationales and use cases

• Templates for ingest agreements?

• Tests with more institutions (dp4lib?)
Thank you for your attention!

Questions?