

Long-Term Archiving Standards and Best Practices

Philipp Wieder, GWWDG

Sofia, March 12/13, 2019

Very Brief Introduction



Georg-August-Universität Göttingen

- 200 institutes und 13 faculties
- Staff members: 14.000
- Researchers: 3.000, professors: 400
- Students: 34.000

University Medical Center (UMG)

- 7.000 staff members

Göttingen Campus Partners

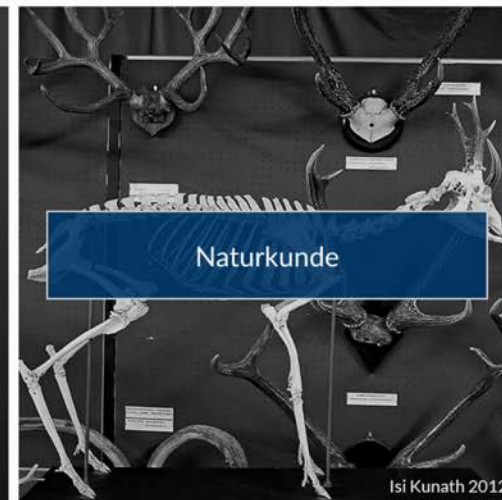
- 5 Max Planck Institutes, Leibniz Institute DPZ, DLR, Academy of Sciences and Humanities, ...

Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

- Data and IT Center for Göttingen University and the Max Planck Society
- IT-Services for over 80.000 users



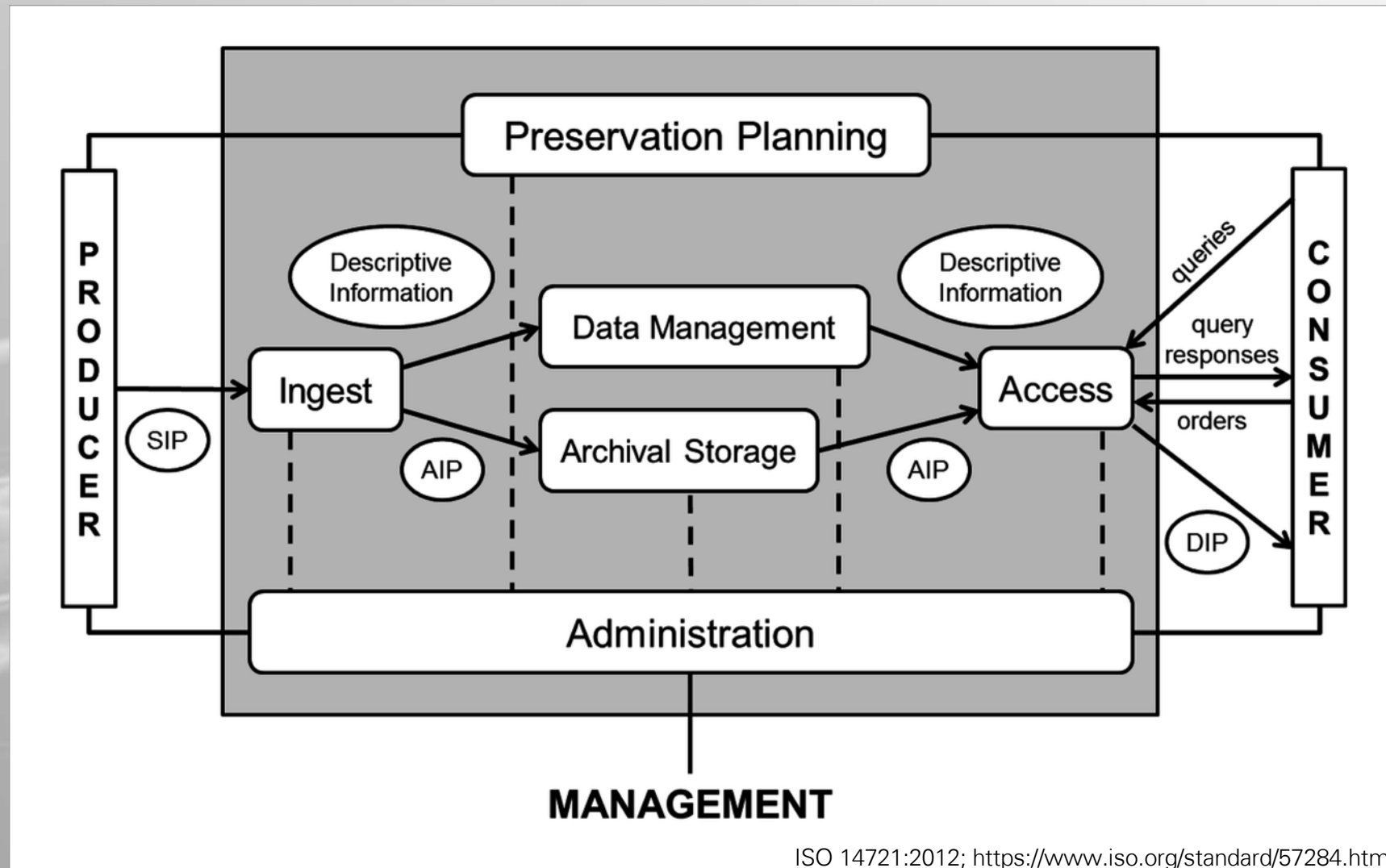
Example: Collection Portal (& DB)



What can be achieved?

- Agreement on set of standards to archive data and to enable FAIR usage
- Common terminology (think well-defined metadata)
- Developing (new) local solutions based on standards
- But
 - Usage of standards does not automatically imply interoperability
 - Existing solutions will remain

OAIS: Foundation, Terminology, ...



ISO 14721:2012; <https://www.iso.org/standard/57284.html>

Metadata (for Archiving)



- Often technical and contextual metadata is distinguished
 - For LTA it is often technical information, but also versioning
 - Rich metadata is good, but is it provided/maintained?
 - Metadata quality will become an issue
 - Talking about “infinite” preservation: semantic change has to be considered

Metadata Standards I

■ PREMIS

- “PREMIS Data Dictionary for Preservation Metadata is the international standard for metadata to support the preservation of digital objects and ensure their long-term usability”
- <http://www.loc.gov/standards/premis/>
- XML Schema; OWL Ontology

■ Dublin Core

- “‘core metadata’ for simple and generic resource descriptions”
- <http://dublincore.org>
- XML, HTML, RDF, Text

Metadata Standards II



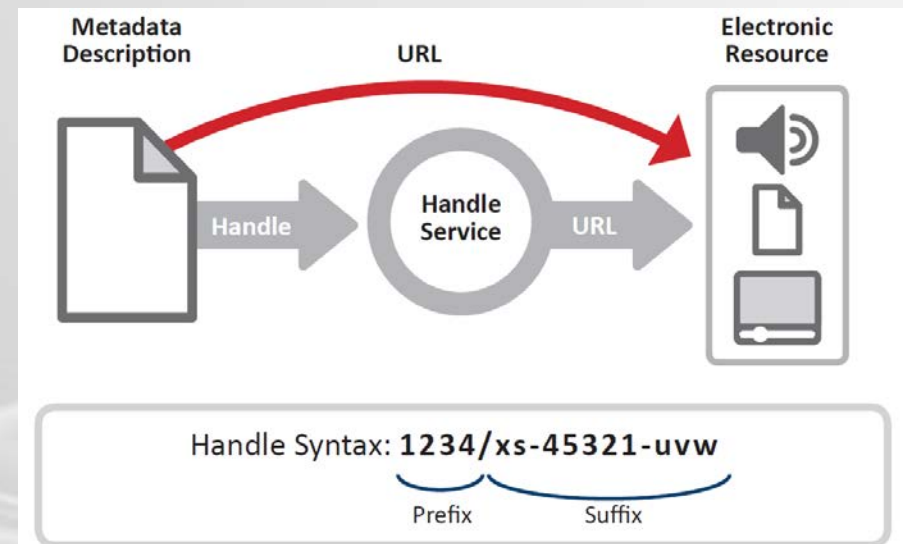
- **LIDO - Lightweight Information Describing Objects**
 - “Format for contributing collection information for resource discovery”
 - <http://network.icom.museum/cidoc/working-groups/lido/lido-technical/specification/>
 - XML Schema
- **More**
 - METS, OAI-PMH, LMER, Spectrum, ...

Packaging

- zip, tar, ..
- Bagit
 - <https://datatracker.ietf.org/doc/rfc8493/>
- Git repository
 - ...

Referencing

- Digital Object Identifier (DOI)
 - ISO/CD 26324
- Handle.Net
 - IETF RFC 3650 et al.
- EPIC PID
 - Based on Handle.Net
- Uniform Resource Name (URN)
 - IETF RFC 2141 et al.



Certification

- ISO 16363:2012 Audit and certification of trustworthy digital repositories (<https://www.iso.org/standard/56510.html>)
- CoreTrust seal (<https://www.coretrustseal.org>)
- Nestor seal for trustworthy digital archives (https://www.langzeitarchivierung.de/Subsites/nestor/EN/Siegel/siegel_node.html)

Records Management

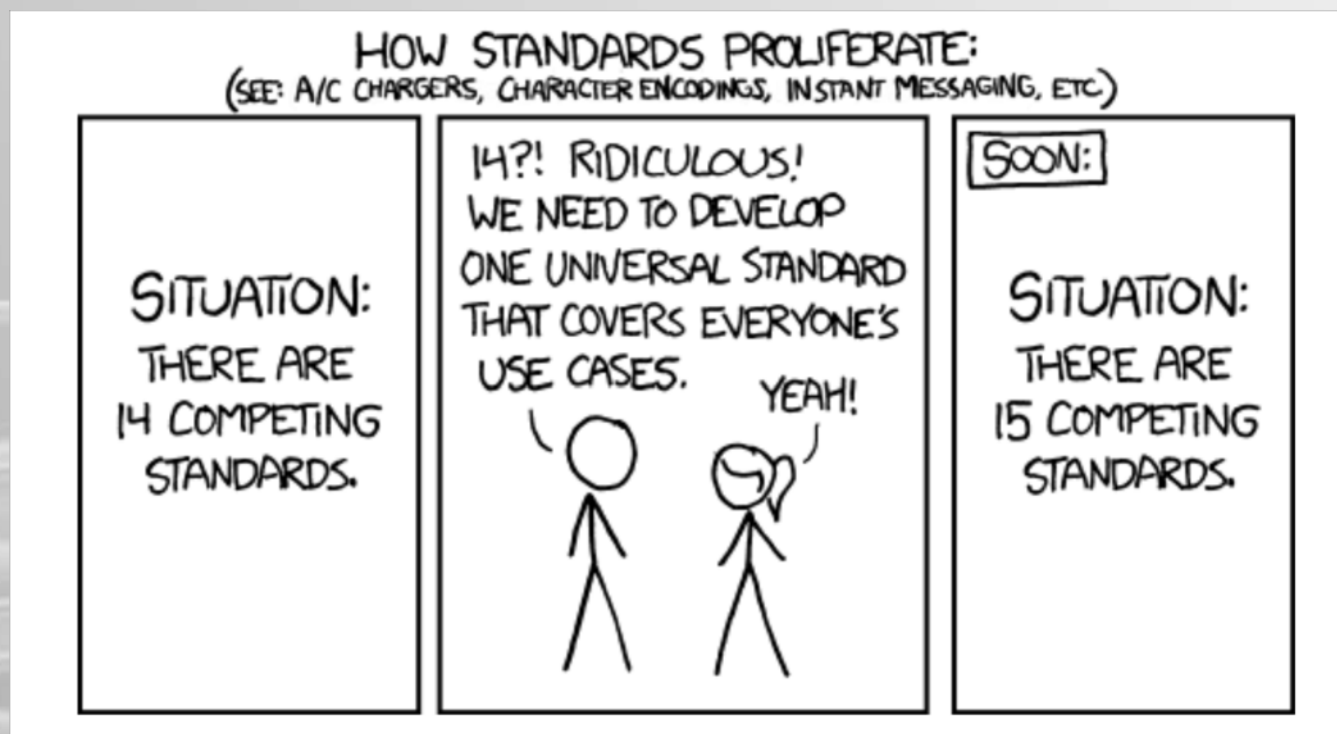


- **ISO 15489-1:2016: Information and documentation - Records management - Part 1: General**
 - Create and capture records to meet requirements for evidence of business activity
 - Take appropriate action to protect the authenticity, reliability, integrity and useability of records, as well as their business context, and to identify requirements for their management over time
- **Plus**
 - ISO 23081-1:2006: Information and documentation - Records management processes - Metadata for records - Part 1: Principles
 - ISO 23081-2:2009: Information and documentation - Managing metadata for records - Part 2: Conceptual and implementation issues
 - ISO/TR 26122:2008: Information and documentation - Work process analysis for records
 - ISO/TR 18492:2005: Long-term preservation of electronic document-based information
 - ISO/TR 15801:2004: Electronic imaging - Information stored electronically - Recommendations for trustworthiness and reliability

Further Links

- Open Preservation Foundation
– <https://openpreservation.org>
- More to come ...

Thank You!



xkcd, CC BY-NC 2.5; <https://xkcd.com/927/>