

DIE ÖSTERREICHISCHE BIBLIOTHEKENVERBUND
UND SERVICE GMBH

obv sg

FLOATS OUR BOAT A GENESIS OF E-ARK WORKFLOWS WITH LIBSAFE

EARCHIVING INITIATIVE SUMMIT 2023-10-25

BERNHARD HAMPEL-WAFFENTHAL, STEFAN MAJEWSKI

CONTENTS

- Introduction
- Digital preservation at OBVSG
- Choosing LIBSAFE
- Why and how E-ARK?
- Marrying E-ARK and existing METS-based workflows
- General impression and outlook

WHO WE ARE



AUSTRIAN LIBRARY NETWORK AND SERVICE LTD

OBVSG

- Service Center for a network of scientific, humanities and administrative libraries in Austria (OBV)
- A network of more than 90 institutions, still growing
- Member institutions range from large libraries like the Austrian National Library, large university libraries to smaller foundations
- Hosting a shared catalogue with network zone and institutional zones
- Governed by public law:
 - Initially established for the provision and coordination of library-automation <https://www.ris.bka.gv.at/eli/bgbl/i/2002/15/P3/NOR40027275>

DIGITAL PRESERVATION AT OBVSG



- No in-house archiving requirements (so far)
- Library Institutions faced with long-term preservation requirements
 - Digitization of artefacts
 - Metadata management
 - Management of born digital material
- Law extended in 2020 with: Provision of central core infrastructure for digital preservation in scientific and artistic publishing. <https://www.ris.bka.gv.at/eli/bgbl/i/2002/15/P3/NOR40225211>
- Just the right time to start developing and offering digital preservation services

OBVSG AND DIGITAL PRESERVATION

WHAT WE AIM FOR

- Aggregation of Competencies
- Streamlined digital Preservation workflow that allows:
 - A generalized model
 - Tenant specific preservation plans and developments integrating with the same model
 - Wide range of types of archived objects and representations
- Becoming the go-to address for digital archiving and preservation in our domain

CHOOSING LIBSAFE



CHOOSING A DIGITAL PRESERVATION SYSTEM PROCESS AND CRITERIA

- Forming a consortium with Austrian National Library
- Joining forces and expertise in the institutions
 - Austrian National Library has a long track record in digital preservation
 - OBVSG has some staff previously engaged in preservation activities
- Public tender process [\[TED:NOTICE:572204-2020\]](#)
- Posted criteria to find a viable solution
 - delivered with the quality expected from long-term preservation experts
 - allowing the use of established standards, but open to adapt to evolving standards
 - flexible interaction with third party systems
 - scalability & capable of managing multiple tenants
 - clear exit path

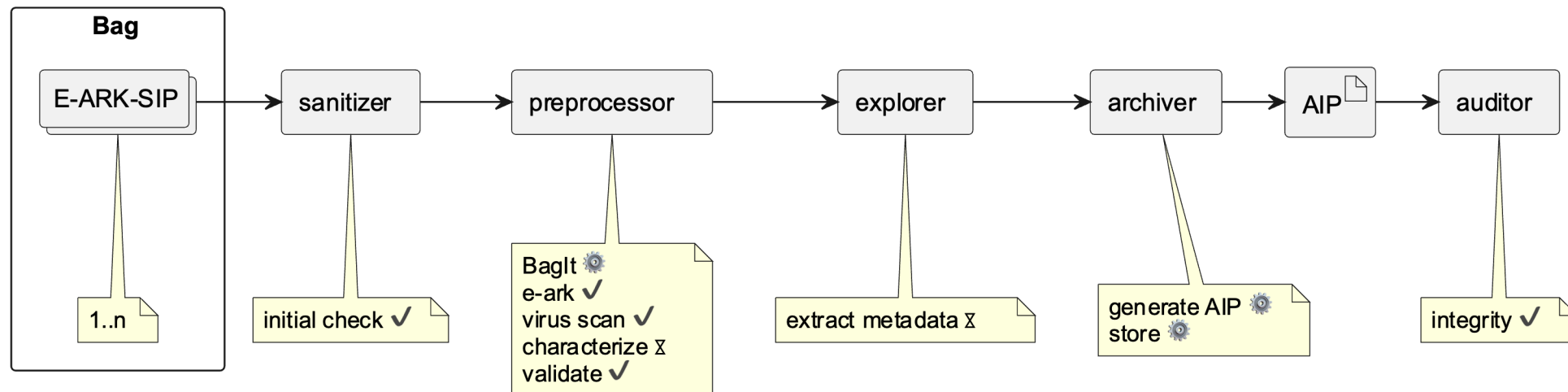
-> LIBSAFE Advanced Pro On-Premise

WHY AND HOW E-ARK



- Searching for a non-proprietary and not system-specific package-format (mainly SIP)
 - Uptake
 - Implementations
 - Strong specification: Clarity and Specificity
- Searching for a specification that promises some longevity
- Implementing the first tenant is a dangerous trap
 - Need to abstract tenant requirements
 - Need to abstract tenant specific third-party system integration
- We wanted to put the package specification upfront. Before specifying the tenants' workflows
- Many aspects depend on package formats
 - Routes from original systems to digital preservation
 - Tools

- We recognized the value in having more of the physical *structure* of objects be specified
- Often only BagIt is used, which only has minimal structure
 - Varying local specifications on metadata fields and inner structure are used
- We still use BagIt on top of E-ARK-SIP to enable the possibility of splitting large objects up for transfer using bag groups
 - We noticed discussion of divided package structure in AIP specification late, more investigation into how this could solve the same need still an open work item



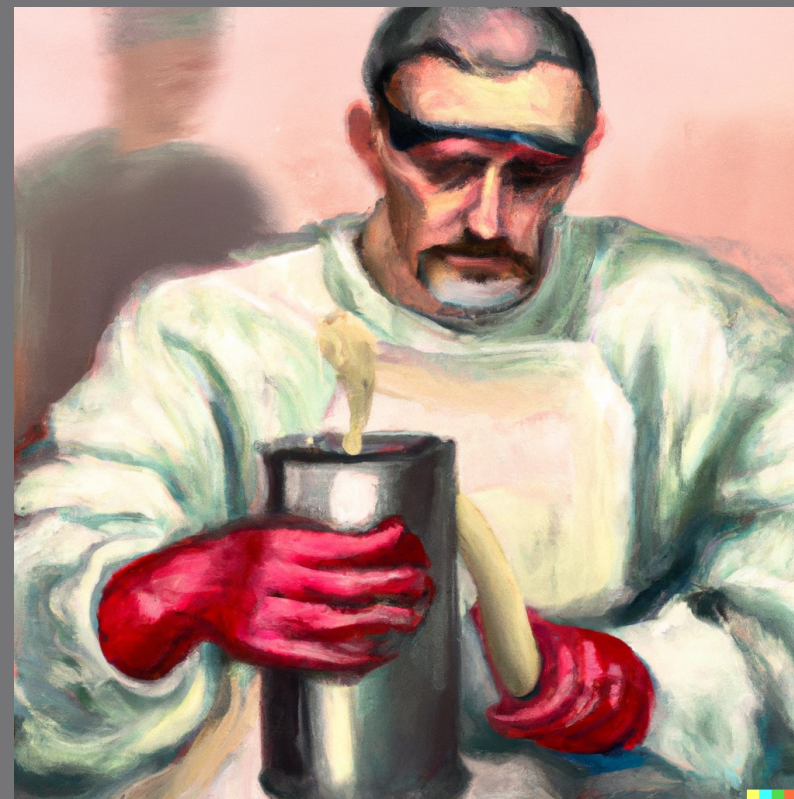
- LIBSAFE allows to integrate own tools and processing steps
- Recommendation to keep metadata files physically separate from METS proved to make some steps easier
 - Extraction of metadata parts for LIBSAFE's "extended metadata" only needed file reading without METS-awareness
 - Metadata updates during objects' lifetime can be handled on that level and be written back including update history when a DIP is generated, reducing storage need of whole-object updates given the current architecture
- We decided not to follow the recommendation to store XML schemas inside each object

WHY AND HOW E-ARK INTEGRATION IN LIBSAFE

- LIBSAFE offers a lot of flexibility in configuring preservation workflows, areas & plans, making it effectively packaging standard agnostic
- Puts power to build processing steps around E-ARK in the ingestion and dissemination pipelines into our hands...
- ...but to make the most of E-ARK support for some of its concepts needs to be extended.
- Most importantly the system is not natively aware of representations and doesn't surface these for metadata management or generic targeted dissemination

MARRYING E-ARK

AND EXISTING METS-BASED WORKFLOWS
WILL IT BLEND



ADAPTING TO E-ARK SIP WITH PRE-EXISTING METS-BASED EXPORTS

- Important question early during ingest: Will it blend validate?
- Integrating commons-ip was deemed best option within our constraints
- Vienna University of Economics uses Goobi as the source system for the digital objects
- Already heavily reliant on METS, which was built upon for the newly developed E-ARK-Export integration
- Validation logic as implemented frequently turned up issues when confronted with (allowed) METS usage not explicitly discussed in requirements
- Many bugs were reported and fixed since, but some remain...

ADAPTING TO E-ARK SIP WITH PRE-EXISTING METS-BASED EXPORTS

- Not all issues turned out to be implementation issues, but pointed to need for clarification or further discussion in the specification
- Open issues
 - How to handle or constrain package/object identifiers containing characters not representable on common file systems? (e.g. DOIs, which contain slashes)
 - METS documents have no way to unambiguously state what specification version they are intended to comply with
- Issues clarified or fixed
 - Values for MIME type attributes seemingly being limited to only the small set on the IANA list doesn't play well with the reality of diverse formats and their MIME types we sometimes get
 - Validity scope of component identifiers was unclear enough to result in overly strict validation of uniqueness across METS documents

IMPRESSIONS & OUTLOOK



- Overall complexity relatively high
 - Relationship between CITS and PREMIS/Archival Information specifications unclear, but we see this is being actively discussed
 - Hard to get a grasp on in its entirety for a small outside team like ours
 - Coherency between specifications could probably still be improved
 - Some aspects are being discussed for unexpected types (cf. divided package structure in AIP when "This is not only relevant for storing the AIP, it also concerns the SIP which might need to be divided before the data is submitted to the repository.")
- E-ARK DIP specified mainly for single representations?
- How to test an archival system migration scenario towards practical package compatibility? Against which systems?

- Uptake of E-ARK package formats and generally eArchiving specifications
- Long-term funding perspective for the initiatives behind E-ARK
- Strengthening of the community processes
- Long-Term commitment for maintaining and improving E-ARK as a standard

THANK YOU!

