The role of E-ARK in meemoo’s ingest

10 years of E(-)ARKiving
Today’s (story) arc

Who is meemoo?

A bit of context

Our search for a SIP

Introducing the ‘meemoo SIP’

Ingest SIPs into our archive

Lessons learned & way forward
meemoo’s service provision

- Digitisation, digital archiving and management of archival materials
- Make content accessible and usable
- Actively gather and share expertise on digital archive operations
- Advise on digital heritage processes
Collaboration

182 content partners (CP) in culture, media and government

These figures are from 31 December 2023.
### Content partners in different sectors

<table>
<thead>
<tr>
<th>Sector</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>performing arts</td>
<td>54</td>
</tr>
<tr>
<td>museums</td>
<td>50</td>
</tr>
<tr>
<td>archives</td>
<td>24</td>
</tr>
<tr>
<td>heritage societies</td>
<td>20</td>
</tr>
<tr>
<td>regional broadcasters</td>
<td>10</td>
</tr>
<tr>
<td>government institutions</td>
<td>12</td>
</tr>
<tr>
<td>heritage libraries</td>
<td>7</td>
</tr>
<tr>
<td>national broadcasters</td>
<td>3</td>
</tr>
<tr>
<td>sector institutes</td>
<td>2</td>
</tr>
</tbody>
</table>

These figures are from 31 December 2023.
In figures

> 8 million objects in our archive system

> 540,000 audiovisual carriers transferred to our archive system

nearly 185,000 user accounts at The Archive for Education at end of 2022-2023 academic year

All these figures except for education are from 31 December 2023.
Media Asset Management System

Interaction
- The Archive for Education
- News of the great war
- Art in Flanders
- Internal tools
- Applications of Content partners

Integration
- SEARCH
- REST API
- GRAPHQL
- OAI-PMH

Archive + metadata management
- Media Asset Management System
- Digitization
- Batch
- Continuous
- Contentpartner

metadata + material

News of the great war

hetarchief.be

Art in Flanders

Internal tools

Applications of Content partners

Contentpartner
A bit of context

how we came to the realization that we needed a SIP
First SIP practice in Digitisation

Digitisation via external digitization companies

Manageable & simple packaging rules

Vendor-imposed SIP (with a descriptive metadata twist)
Rise in number of partners

Large scale digitisation

Digitisation + “Born Digital” archiving
Influx of digital collections

FTP/S3

DROID

Descriptive metadata templates
More complex object structures

Newspaper issue
- xml
- Page 1
  - xml
  - alto xml
  - tiff
  - jp2
- Page 2
  ...

Postcard
- picture front
- picture back
- xml

Single digitised picture
- JPEG
- xml

Performance
- video
- subtitles
- xml

heterogeneous...
file types
structure
object types
metadata
More heterogeneity in metadata

Content partners from different domains

Collection management model A
Collection management model B
... model ...
Collection management model Z

Increasing demand for descriptive non-AV metadata fields

Diverse metadata intake
Mitigation strategies (sort of)

- Divide digital influx into different workflows
- New vendor-specific SIPS for e.g. newspapers
- Providing custom services via project manager

All of this works (kind of), but increases complexity and overhead
Setting sail for a (new) SIP

how we came aboard the (E-)ARK
Requirements

A new uniform SIP design should ...

- have a **formal and mature specification**
- be embedded in **standards**
  - be technologically aligned with the **media and cultural heritage sector**
- be **flexible and extensible**
## Options we explored

<table>
<thead>
<tr>
<th></th>
<th>Standards</th>
<th>Specification</th>
<th>Mature</th>
<th>Verbosity</th>
<th>Community support</th>
<th>Sector alignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>RO-Crate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCDM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BagIt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-ARK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Closing the E-ARK and meemoo gap

Flexibility can lead to complexity, esp. with not very tech savvy partners

Need for more rules wrt. to object structure, but mostly descriptive metadata

Translate E-ARK spec into our own meemoo SIP specification

- `terminology` + `core concepts` sections
- narrowly following the hierarchical structure
- conceived the concept of ‘content profiles’: additional use-case driven constraints, often domain or material specific
A (deep) dive into the ‘meemoo SIP’

how we put it in a Bag only to remove it afterwards
meemoo SIP Specification

The meemoo SIP specification is available in the following versions:

**Version 2.0**  EDITOR'S DRAFT
This major release ensures adherence of the meemoo SIP to the EARK CSIP specification. This version is being drafted and is not yet supported.

What was added compared to the previous version?
- Conformance with the EARK CSIP 2.2 specification. For more details, check out changes to 1.X section.

What was removed compared to the previous version?
- The Bag level was removed from the SIP, thereby also removing its dependency on the BagIt specification.

**Version 1.2**  RELEASE CANDIDATE
This third release broadens support for written works and extends the metadata for the basic profile.

[https://data.hetarchief.be/id/sip/](https://data.hetarchief.be/id/sip/)
## Content profiles

<table>
<thead>
<tr>
<th>Limited metadata (DC + schema.org)</th>
<th>Basic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bibliographic metadata (MODS)</td>
<td></td>
</tr>
<tr>
<td>High-level metadata (schema.org)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1 Intellectual Entity</th>
<th># Intellectual Entities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Representation</td>
<td># Representations</td>
</tr>
<tr>
<td># Files</td>
<td># Files</td>
</tr>
</tbody>
</table>

**Material artwork**
Open editorial process using Github pull requests and JIRA
Open editorial process using Github pull requests and JIRA.
# Support level through publication statuses

<table>
<thead>
<tr>
<th>Status</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EDITOR’S DRAFT</strong></td>
<td>This version of the specification is still being developed.</td>
</tr>
<tr>
<td><strong>RELEASE CANDIDATE</strong></td>
<td>This version of the specification is ready for public review and implementation. The text can still undergo changes and meemoo does not give any guarantees that material can be supplied in this way.</td>
</tr>
<tr>
<td><strong>STABLE</strong></td>
<td>This version of the specification will no longer change and has been fully implemented. Meemoo guarantees that material supplied in this way can be archived.</td>
</tr>
<tr>
<td><strong>DEPRECATED</strong></td>
<td>This version of the specification is outdated and will no longer be supported in the future. Its use is not recommended by meemoo.</td>
</tr>
<tr>
<td><strong>RETIRED</strong></td>
<td>This version of the specification is no longer supported by meemoo. Material can no longer be supplied in this way.</td>
</tr>
</tbody>
</table>
Versioning

- **1.0** (RETIRED) - simple video & audio
- **1.1** (STABLE) - digitised artworks
- **1.2** (RELEASE CANDIDATE) - written works
- **2.0** (EDITOR’S DRAFT) - ‘bridging the E-ARK gap’
Move editing & versioning into governance

Release cycle to avoid ad-hoc alterations

Connect with partners, gather input & cover their needs

Predictable support for both sides

Still ongoing work
Ingesting a ‘meemoo SIP’

how we transform it to a graph for semantic validation
Ingesting a ‘meemoo SIP’

- Delivery via S3
- BAG validation
- XSD validation
- Transform to RDF graph
- SHACL validation
- MAM SIP
Validating the SIP as a graph

Graph Transformation

Representation

Validation

content profile

RDF Graph
PREMIS OWL + Schema.org/bibframe
Tides of wisdom

aka lessons learned and the way forward
Lessons learned

A meemoo-SIP specification is not enough for our partners

- non-technical manuals for creating specific meemoo-SIP
- support in metadata mappings

Information about E-ARK is scattered over websites and specs

- avoidable breaking changes & custom tooling
- too little contact with the E-ARK community
The way forward

Reducing the complexity of current meemoo-SIP

Bridging the E-ARK gap

Usage of E-ARK tooling and contributing

Specific documentation for users

Thinking about a SIP lite using JSON-LD
The way forward (2)
Thanks for listening!

Miel Vander Sande (miel.vandersande@meemoo.be)
Mattias Poppe (mattias.poppe@meemoo.be)
Matthias Priem (matthias.priem@meemoo.be)
Milan Valadou (milan.valadou@meemoo.be)