
Archiving of digital content at the Library of the Czech Academy of Sciences

Ing. Martin Lhoták

Library of the Czech Academy of Sciences

12. 3. 2019 National Archive, Prague

Archivierung von Unterlagen aus digitalen Systemen



KNIHOVNA

AKADEMIE VĚD ČR

Czech Academy of Sciences

- **54** public research institutions (separated legal entities)
- **9 000** employees, (> 5 000 R&D)
- > **11 000** research outputs a year (1/3 in impacted journals)
- > **130** titles of journals published by CAS institutes
- > **100** years history

Library of CAS and digital content

- Publications published by institutes of CAS – digitized or e-born → **Digital Library of CAS**
- Publications of CAS authors published by another publisher – digitized or e-born → **Institutional Repository of CAS**
- Research data → **Institutional Repository of CAS**
- **Standardized metadata** for all digital content

Library of CAS and current archiving

- **50 TB**
- bit-level preservation
- **2** copies on HDD storage systems in **2** localities
- **1** copy on LTO tapes



KNIHOVNA

AKADEMIE VĚD ČR

ARCLib – new open source LTP solution

- **Complex Solution for Long Term Archiving of (Library) Digital Collections**
- Applied research grant from Ministry of Culture of CR
- Technologies and methodologies for preservation of culture heritage
- 2016-2020
- 850k Euro



ARCLib – main goals

- Develop complex **open source LTP system** using Archivematica as one of its components
- **Logical preservation** methodology for Czech institutions
- **Bit level preservation** – methodically and technologically
- Compliance with standards **ISO 14721** (OAIS), **ISO 16363** (Audit a certification)
- Experience
- Research papers, etc.

ARCLib – inspiration

Systems

- Archivematica
- RODA
- Commercial solutions (Rosetta, Preservica,...)
- Custom made solutions (NDK, NDA SVK)

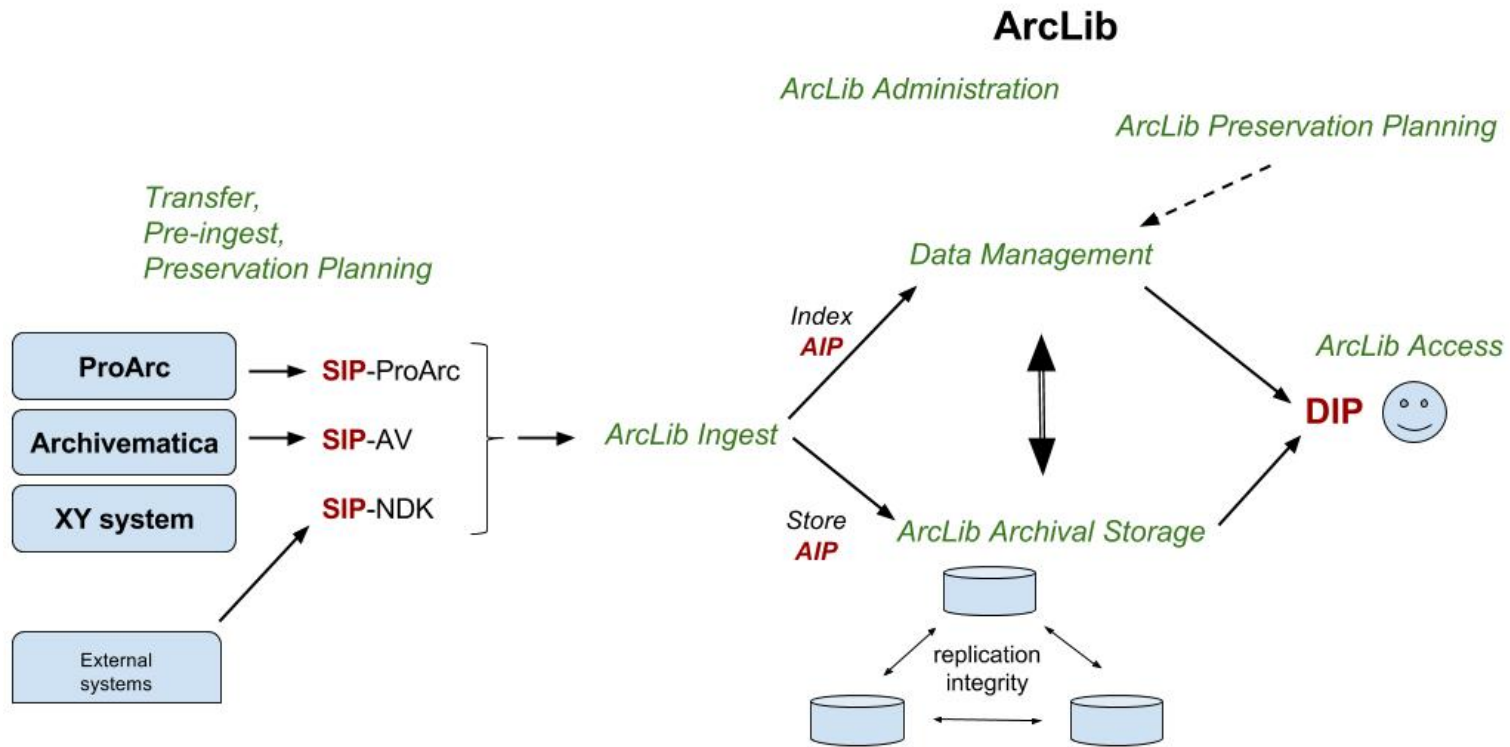
Projects

- CESNET LTP Pilot (testing of Archivematica)
- NDK – Czech National Digital Library
- Czech National Archive (Archivematica)
- Foreign projects (Finland, Germany)

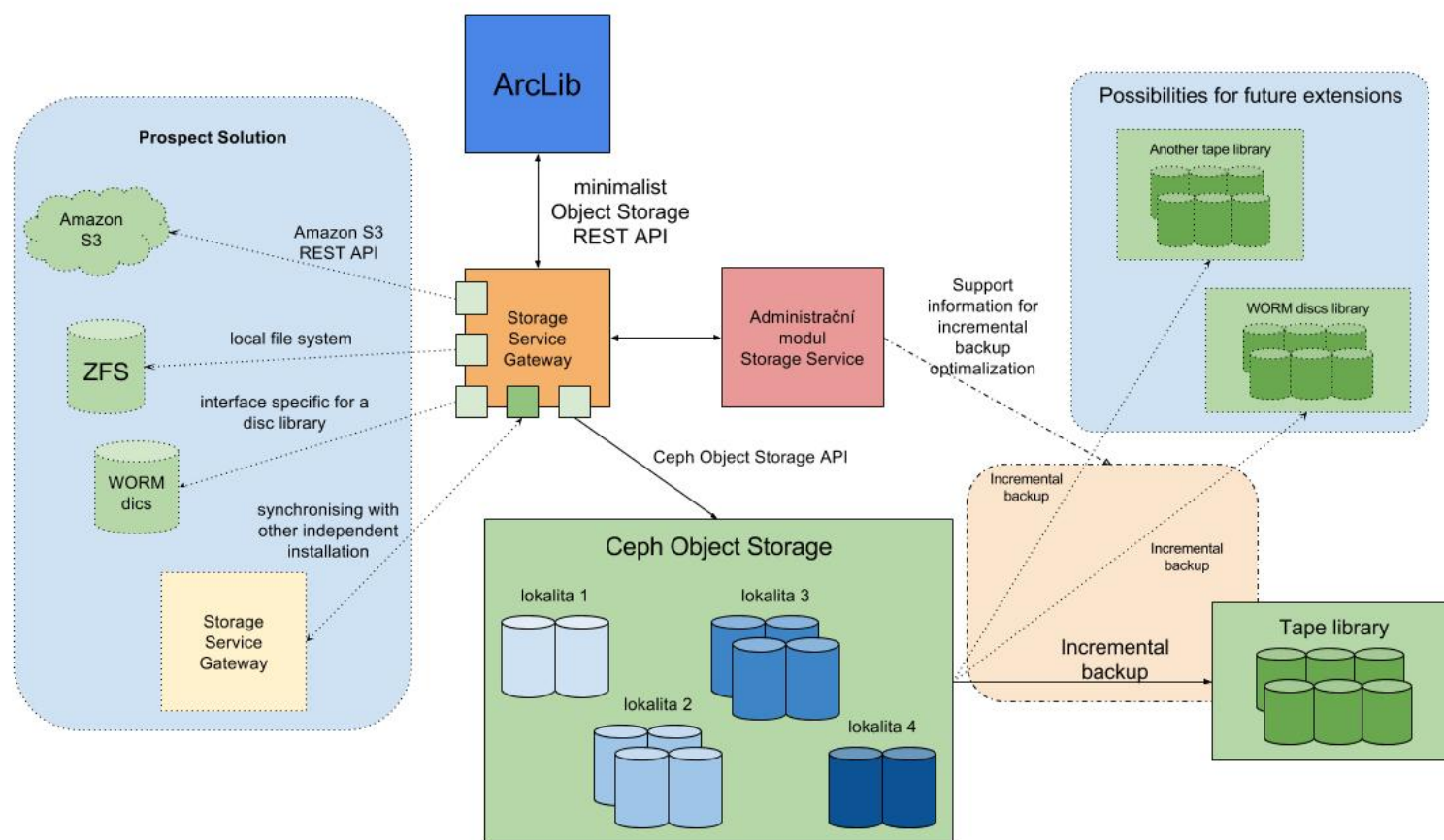
ARCLib – Archivematica?

- open source software
- rapid development
- too big and too general ... (Finland experience)
- dependency, uncertainty
- it could be probably more simple (we don't need universal product, we need system for defined environment and type of data)
- inspiring approaches – microservices, the way of management of ingest

ARCLib – functional requirements



ARCLib – archival storage



ARCLib – AIP

AIP consist of two parts:

provided SIP

+

metadata partly extracted from SIP

- BibMD – DC + MODS
- TechMD – type of scanner, date of scanning, operator, data from JHOVE, etc.

and partly generated by ARCLib on ingest

- AdmMD – data provider, workflow, validation log, validation profile, format identification, date, etc

ARCLib – summary

- ARCLib is an open source solution for **logical and bit-level preservation**/protection of digital data
- ARCLib **doesn't have pre-ingest** or deposit module – it doesn't involve conversion to SIP
- ARCLib is system for **management of archival packages**
- ARCLib is **dark archive**. It is not repository for end users
- ARCLib is open for further development – **another types of archival data**

ARCLib – February 2019 initial release

The screenshot displays the ARCLib web interface. The header shows 'ARCLib' on the left, a language selector 'CZ', and a user profile 'arc-bob'. The sidebar on the left contains navigation links: Users, Producers, Producer profiles, Ingest, Ingest routines, Ingest batches (highlighted), Validation profiles, SIP profiles, Archival storage, Archival storage administration, Logical storage administration, Workflow definitions, Deletion requests, Search queries, AIP search, Preservation Planning, and Issue dictionary.

The main content area is titled 'Ingest batches'. It features a filter dropdown set to 'Updated' and a sort icon. Below this is a table with the following columns: ID, Producer, Transfer area path, Created, Updated, and State. The table contains seven rows of data. At the bottom of the table, there is a pagination control showing '10' items per page and '1 - 7 / 7' total items, with navigation arrows.

ID	Producer	Transfer area path	Created	Updated	State
a2bfa...	Producer 1	fileStorage/prod1	12.03.2019 12:44	12.03.2019 12:47	Processed
807cc...	Producer 1	fileStorage/prod1	12.03.2019 12:28	12.03.2019 12:32	Canceled
b55a3...	Producer 1	fileStorage/prod1	12.03.2019 12:19	12.03.2019 12:22	Processed
ab8e9...	Producer 1	transfer/area/path/3	08.03.2018 10:00	08.03.2018 10:00	Canceled
33bb0...	Producer 1	transfer/area/path/4	08.03.2018 10:00	08.03.2018 10:00	Processing Suspend
0ce8b...	Producer 1	transfer/area/path/2	08.03.2018 09:00	08.03.2018 09:00	Suspended Resume
e1205...	Producer 1	transfer/area/path/1	08.03.2018 08:00	08.03.2018 08:00	Processed

Thank you for attention

Ing. Martin Lhoták
lhotak@knav.cz



KNIHOVNA

AKADEMIE VĚD ČR