



CDA Technology and Design Overview

Ľubomír Hribík

CDA DESIGN

HIGHLIGHTS

- Built to serve as national archive for preservation of Slovak cultural heritage
- According to OAIS model, CDA is federated archive with 3 locations: A, B and C (physical LTO storage)
- Open only for designated community – selected memory institutions
- Access, profiles and metrics are based on contract with each memory institution
- System is scalable horizontally and vertically to withstand big data loads or lot of packages on input

CDA PROCESSES

OVERVIEW

- Automated processes are managed by FRAMEWORK component
- Each automated process is set of steps executed in sequence
- Steps are independent and used like plug-ins

3 core processes (semi-automatic):

- INGEST
- DISSEMINATION
- LTP CHECK

CDA PROCESSES

INGEST

- Order -> INPUT method (LTO/HDD/online)
- ImpEx -> Framework (list new SIPs)

FRAMEWORK steps (simplified):

1. Extract package
2. Package identification and structure check
3. Signature verification, Allowed content according to profile
4. Create or update Order data
5. SIP2AIP – check, copy, add PREMIS data, add CDA signature
6. Store AIP -> TSM hierarchical storage
7. Synchronization copy and CDA-C copy
8. Create catalogue record
9. Set SIP as archived , update Order data
10. Send notifications

CDA PROCESSES

INGEST

- Operator is notified when business or technical error occurs
- Process can continue from technical error but cannot from business error
- Typical business errors are wrong file format or errors in METS file
- Technical errors are occasional
- **IMPORTANT:** SIP_ID is unique and reserved for one process so if package needs to be corrected and re-ingested it needs to get a new SIP_ID

CDA PROCESSES

DISSEMINATION

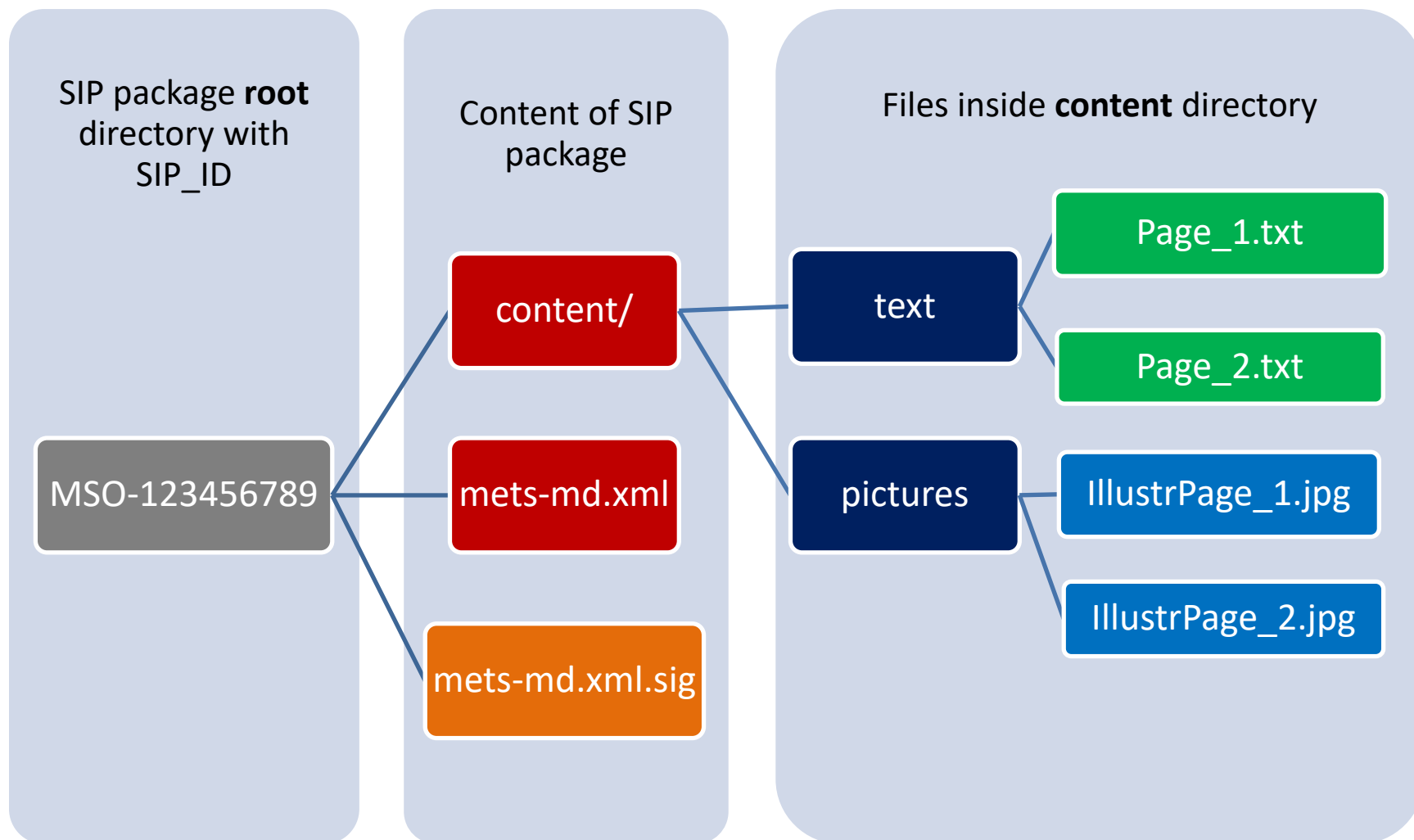
- Very similar to INGEST - input is AIP and output is DIP but without creating any copies of DIP
- User creates Order for each AIP, selects OUTPUT method (LTO/online) and can select subset of AIP data (defined in METS fileSec structure)
- Process is finalized by setting a flag when DIPs are prepared for transport/acquisition
- M.I. is notified by summary e-mail

CDA PROCESSES

LTP CHECK

- Process designed to check cold storage data
- Periodically checks date of last check (catalogue) / tape
- Extracts all AIPs from tape
- Checks each AIP using same steps as for INGEST
(antivirus, fixity, formats)
- Stores results in catalogue
- If error is detected then restoration process should be run
- Restoration – manual process by operator

CDA PACKAGE STRUCTURE



METS

METADATA ENCODING & TRANSMISSION STANDARD

XML document describing structure and physical location of your digital content. It can also contain technical and descriptive metadata about each object.

7 main sections:

- Mets Header (institution ID, package ID)
- Descriptive Metadata (DublinCore)
- Administrative Metadata (optional, PREMIS events)
- File Section (physical structure, fileGrp)
- Structural Map (logical hierarchical structure)
- Structural Links (links between objects in Map)
- Behavior (not used)

FILE FORMATS

TOOLS & PLUG-INS

- Format identification
 - DROID, puid from PRONOM (NA UK)
 - Puid in Contracts and Profiles
- Format validation (pairing to mime-type)
 - JHOVE plug-ins, mediaConch (server), veraPDF (PREFORMA)
 - Plug-ins in Profiles
- Format database (FMT DB)
 - Risk formats
 - Version history (DROID signature files)
 - Add proprietary format (own puid & identification)

CDA INTERFACES

GRAPHICAL UI

Web GUI for Operator and Users:

- Orders (ingest, dissemination, single or mass)
- Catalogue (search for package, file or format)
- Dashboard (today, total, just M.I., both locations, compared)

Only for the Operator:

- Logistics and stock management (any medium, CDA-C tapes)
- FMT DB (risk formats, actual format versions and history)
- Tasks (history of done ingests, disseminations & ltpchecks)
- Monitoring (HW vendor software)
- Reporting (SpagoBI)
- User management

CDA INTERFACES

OTHERS

CMD line like (Operator must be logged on server):

- Certificates and keys generator
- Profiles (upload, read-only, test profile)
- ImpEx (managing campaigns)
- Format identification and verification (except mediaConch)
- Administrative tools (configure, start/stop manually)

Webservices (for M.I.):

- IngestOrder, DisseminationOrder
- OAI-PMH

LESSONS LEARNED

HOW TO BUILD DIGITAL ARCHIVE

Purpose:

- Local archive or Central (open) archive
- Just archiving digital content or also LTP archiving

Major components:

- STORAGE
- INTERFACES
- METADATA

LESSONS LEARNED

HOW TO BUILD DIGITAL ARCHIVE

STORAGE

- LTP archive – LTO tapes, more locations synced
- Open archive - staging area for inputs/outputs
- Local archive – disk arrays and backup storage

LESSONS LEARNED

HOW TO BUILD DIGITAL ARCHIVE

INTERFACES

- Open archive – Web app for Users and Operators
- LTP archive – monitoring apps, file format
- Local archive – manually or cmd line like

SERVICES

- Open archive – metadata (OAI-PMH service)
- LTP archive – format validation and conversion
- Local archive – only for data migration

LESSONS LEARNED

HOW TO BUILD DIGITAL ARCHIVE

METADATA

- Outside of type, they need to be in high quality and in metadata standard
- Descriptive vs Technical/structural
- Search vs Publishing

INDEX

- Lot of data = need to implement “ranking system”

DAP

DIGITAL ARCHIVE PLATFORM

 **Tempest**

IT makes sense

DAP DESIGN

HIGHLIGHTS

Integrates modules from CDA and DDP projects into one software solution:

- Supports both archiving and bibliographic work
- MARC21 as metadata standard (native)
- Modular architecture (core / add-ons)
- Performance scaling (horizontal/vertical)
- Web app user interfaces (redesign, translations)
- Automated workflow and distribution of tasks

DAP ARCHITECTURE

LIST OF COMPONENTS

Core

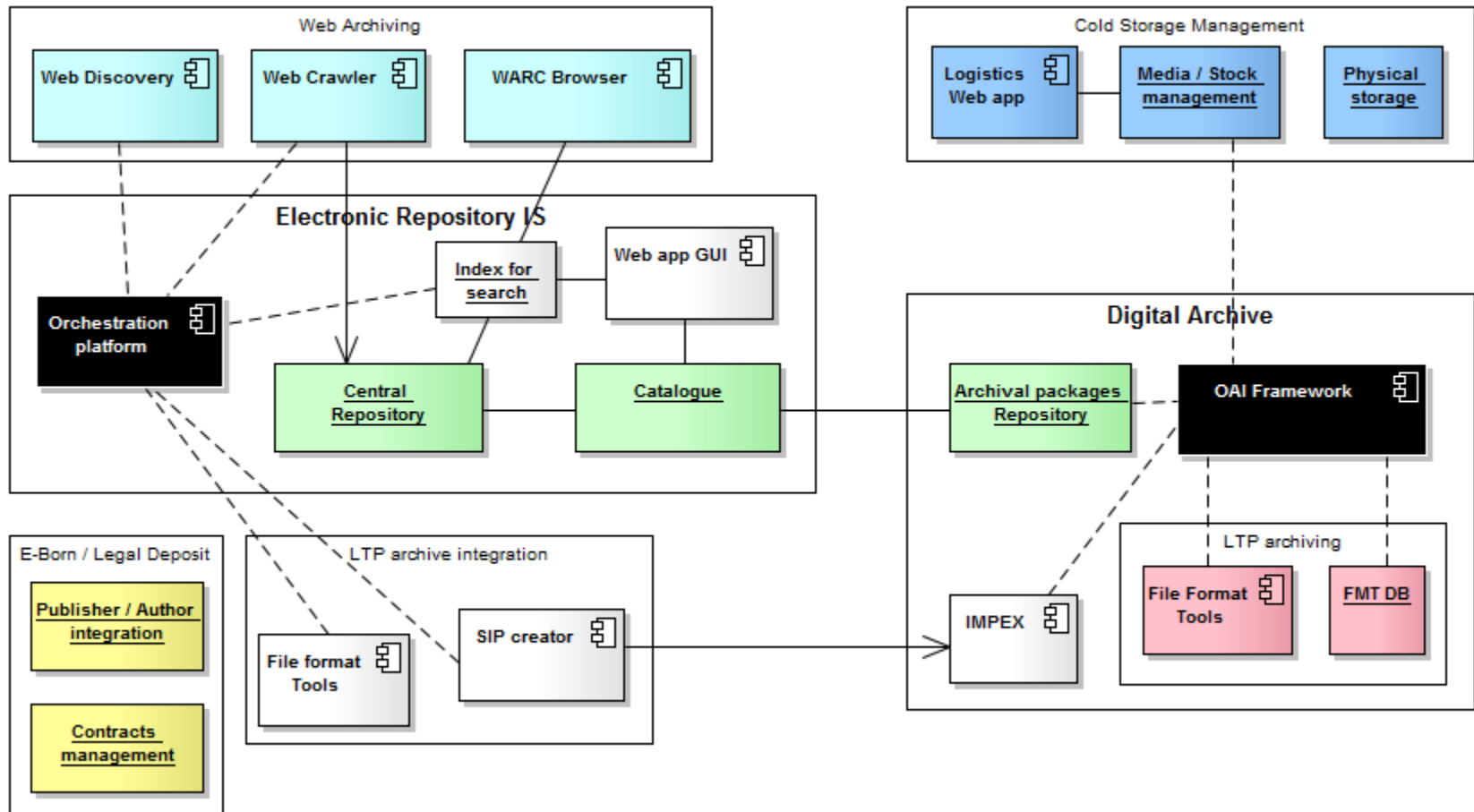
- **Repository** with orchestration platform and interface for its object curators
- **Digital archive** with framework, LTP module

Add-ons

- Webarchive with discovery, web crawler and browser
- Legal deposit / E-Born bibliographic records (FRBR)
- Logistics and stock management for cold storage

DAP ARCHITECTURE

LOGICAL MODEL



DAP HOMEPAGE

WWW.DIGITALPRESERVATION.SK/EN

DIGITAL ARCHIVE PLATFORM

DAP will provide you with a central
search repository



Solution features



Modular and scalable



Flexible workflow



OAIS standard



LTP archive



Cold Storage
Management



Advanced Repository
integrated with
webharvesting

THANK YOU
FOR YOUR ATTENTION

Ľubomír Hribík

IT Business Analyst

e-mail: lubomir_hribik@tempest.sk

mobile: +421 917 493 588

Company reception phone

+421 (2) 502 67 111

Company reception fax

+421 (2) 502 67 100

Information

info@tempest.sk

Sales

obchod@tempest.sk

www.tempest.sk

TEMPEST a. s.

Galvaniho 17 / B

821 04 Bratislava 2

Slovenská Republika

 **Tempest**
I T m a k e s s e n s e

