



DDP Technology and Design Overview

Ľubomír Hribík

REQUIREMENTS

- Catalogue with .sk domains and manually added
- Archive with almost identical copies of webpages
- Archive for e-born documents
- Support for future e-born legal deposit
- Integration with CDA archive

Challenge:

- Catalogue vs Archive
- Metadata standard?
- SK legislative vs EU Author's act



WEB ARCHIVE

Our decisions:

- Catalogue filled by Discovery process
- Rich metadata editor using MARC21 standard
- 1 crawler -> 1 URL from catalogue
- Fast archiving process = set of workers running web crawler
- Orchestration platform (MQ, priority, other system tasks)
- Browsing archive divided into 3 areas (public, library, curator)

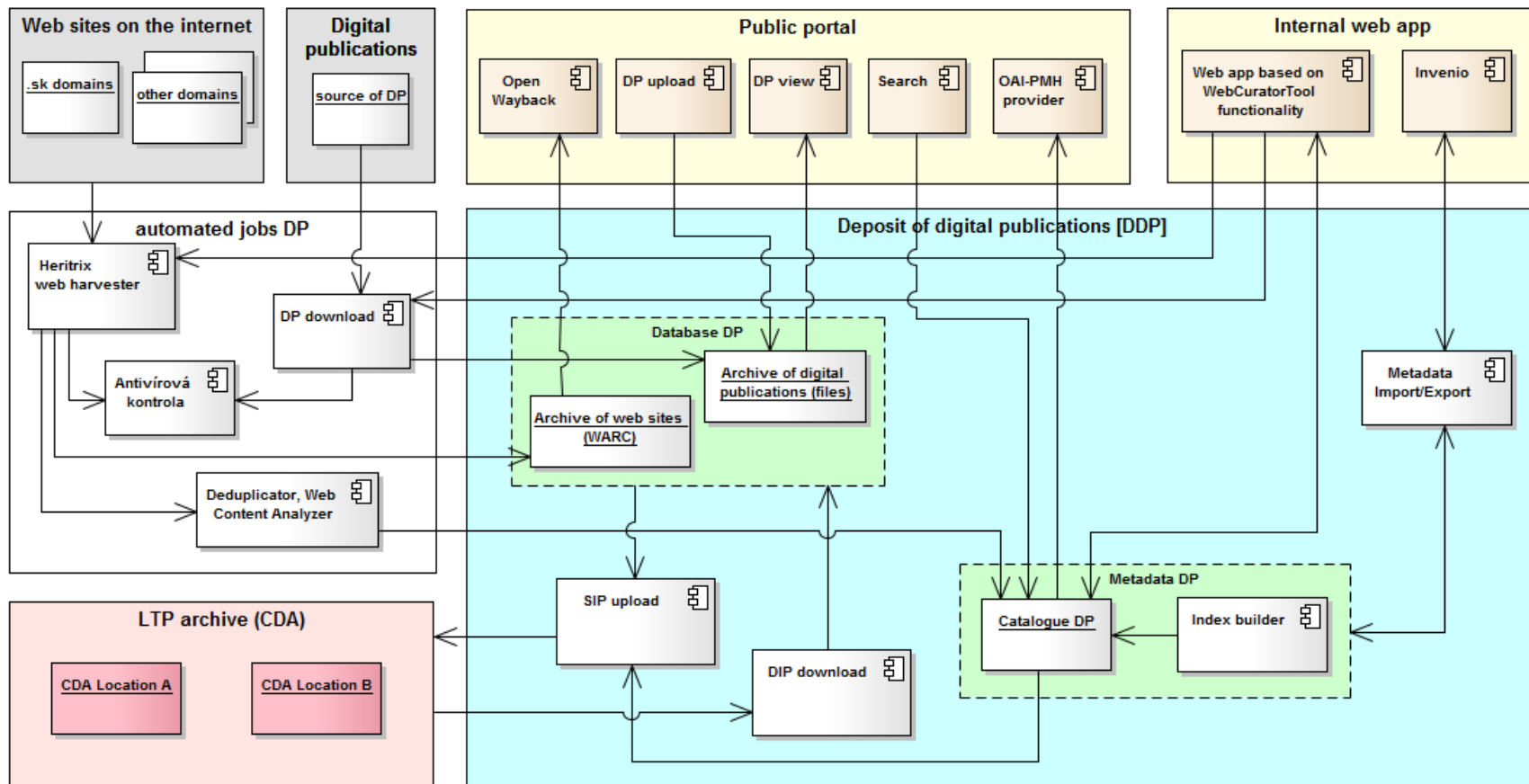
Inspiration from others:

- Heritrix and OpenWayback were the best alternatives (UK)
- Respecting robots.txt, crawler traps (CZ)

E-BORN DEPOSIT

- Implement FRBR / WEMI model (manifestation-item)
- Subjects, contracts, access rights (new role Producer)
- Cooperative/assisted work with Producers
- Prepared for legal deposit (announced for 1.1.2018 but was postponed)
- Integration with CDA - formats (especially PDF versions)
- Serials as continually updated web pages (harvested, but metadata are structural, issue is linked to harvest run) – new type: webserial

ARCHITECTURE



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**

IT makes sense

