Estonian metadata reference architecture

Hannes Kiivet

Head of Interoperability Solutions Department

26.11.2015 – SharePSI Berlin workshop

Agenda

- Goals

- Process

- Results & recommendations

- Future actions

Goals

Business scenario

- Data suppliers provide information on the public services they provide as close to the source as possible

- That information is collected and stored centrally in a searchable format

- Provide a human-readable display of the collected data

Design goals

- The system must be cheap to implement and easy to scale

- The administrative burden on the data source must decrease

- The benefits to the data source must increase

Task

- Create a proof-of-concept (POC) to validate proposed reference architecture for metadata collection
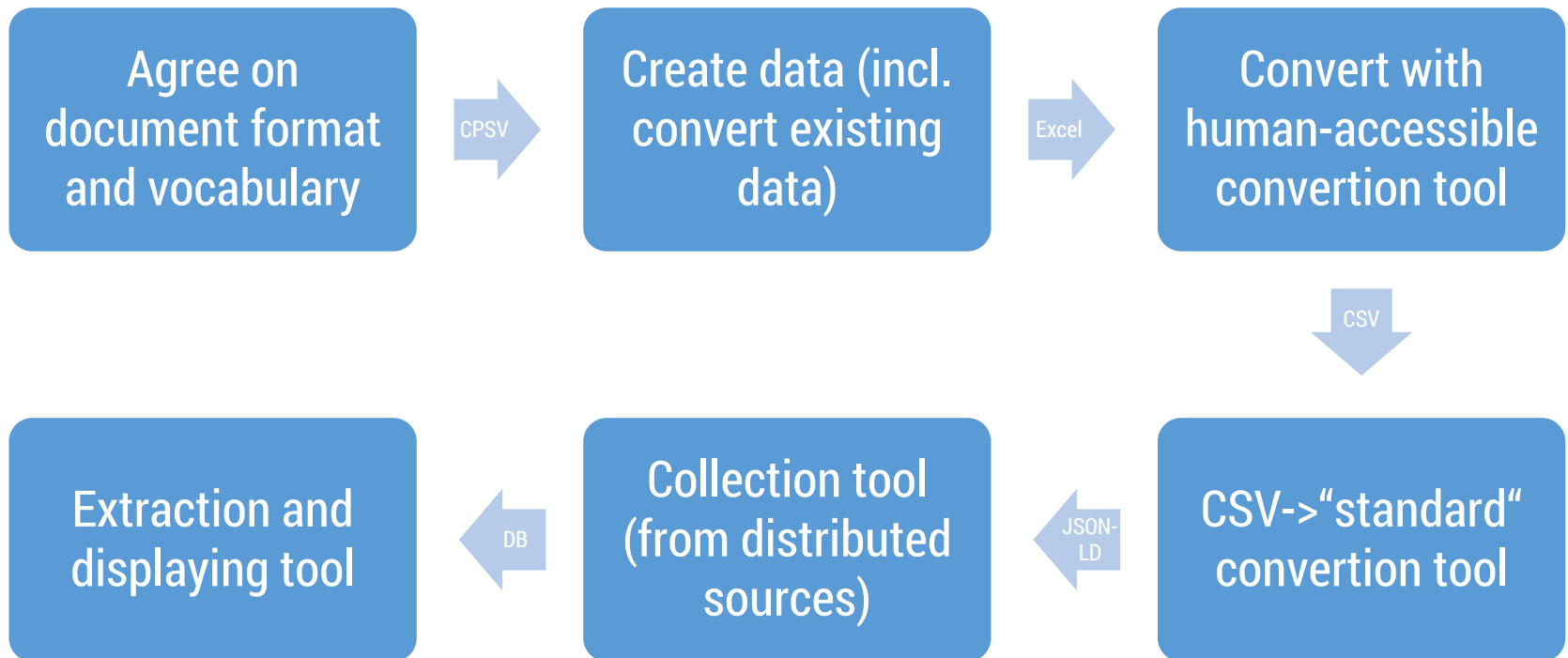
Process

Stakeholders

- EISA - operates the management system of state information systems

- MEAC (Ministry of Economic Affairs and Communication) - policy driver for service governance

- Road Administration - most mature IT and service management functions in Estonian government sector

Data supply chain

Agree on document format and vocabulary	→CPSV→	Create data (incl. convert existing data)	→Excel→	Convert with human-accessible convertion tool

↓ CSV

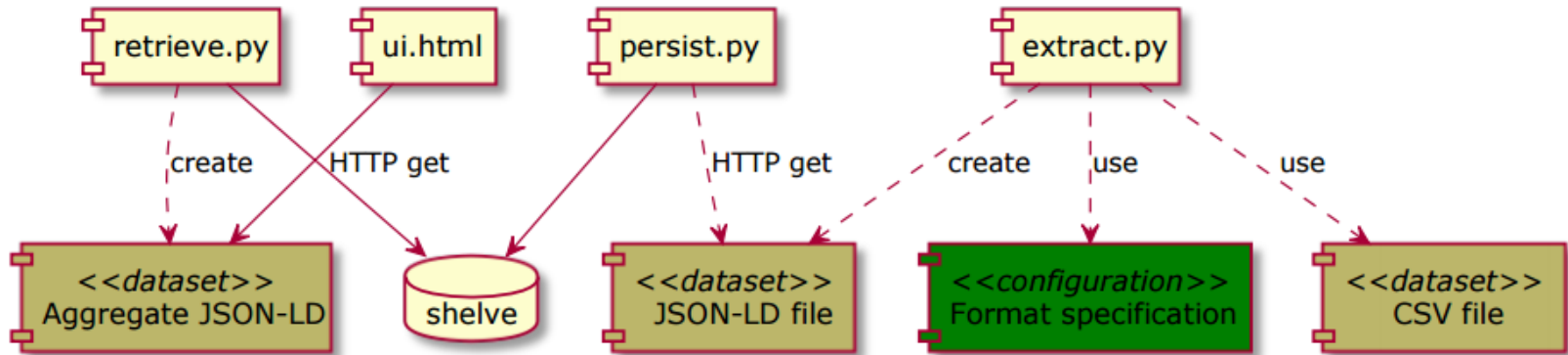| Extraction and displaying tool	←DB←	Collection tool (from distributed sources)	←JSON-LD←	CSV->"standard" convertion tool

Created output

- File format and vocabulary was defined

- Several public organisations produced data that could be included in the common data set

- Small-scale system was implemented

File format - service description file (SDF)

Relies heavily on CPSV and JSON-LD:

- A legal identity (CPSV type LegalEntity, Agent type

- A service description (CPSV type PublicService)

- An URL of another SDF file

Implemented architecture

Republic of Estonia
INFORMATION SYSTEM AUTHORITY

Results & recommendations

Key findings

- Public agencies are both willing and able to participate in distributed system

- Publishing a static machine-readable file on organisational web sites is difficult

- Content providers produced invalid JSON

- JSON-LD documentation and tools is lacking

- Discrepancy between the file encoding and HTTP headers led to transcoding issues

Advice

- Decentralise content supply + provide tools for extracting data from delimited files

- Decouple from organisational web sites + build measures of validating content origin

- Minimum standardisation + rapid technical and organisational feedback loops

- JSON-LD is practical

- Hardcode UTF-8? :)

Future actions

Service catalogue project

- Describing services in Github: https://github.com/MKM-ITAO/riigiteenused, (deadline 2016 end of Q1)

- Current list: https://www.mkm.ee/et/teenuste-otsing, including in JSON (not yet JSON-LD) https://www.riigiteenused.ee/api/et/all

Information systems catalogue project

- Results are used to update the current management system of information systems (RIHA, https://riha.eesti.ee/)

- Business analysis started and plan to release new version with distributed architecture in 2017

Further reading

- Description: https://www.w3.org/2013/share-psi/wiki/images/0/0e/Pilot_Description_(Extract).pdf

- Final report: https://meta.ria.ee/metadata_pilot_final_report.pdf

- Applications' source code and test data: https://github.com/andreskytt/ee_service/

- Estonian state IT-architect blog post of the topic: https://www.ria.ee/riigiarhitektuur/blog/2014/12/17/68/

REPUBLIC OF ESTONIA
INFORMATION SYSTEM AUTHORITY

Thank you for listening! Any questions?

Hannes Kiivet

Hannes.Kiivet@ria.ee