Interoperability specifications and an automatic validation tool as a key factor for the support of a repository ecosystem

Dr. Houssos, Nikos; Dr. Stathopoulos, Panagiotis, Dr. Stathopoulou, Ioanna-Ourania, Banos, Vangelis, Georgiadis, Harris, Dr. Sachini, Evi

National Documentation Centre/National Hellenic Research Foundation

Vas Kon/nou 48, 116 35, Athens, Greece

http://www.ekt.gr

Implemented in the scope of “Platform for the Deposit, Management and Delivery of Open Metadata and Digital Content”

http://epset.gr/en/SaaS_Services/
Agenda

1. The National Documentation Centre
2. The Problem
3. The Solution, part A: Interoperability specifications
4. The Solution, part B: Automatic Validation Tool
5. Current Status
6. Conclusions and future work
The National Documentation Centre (EKT)

• The national organisation in Greece:
  – for scientific documentation, metrics, online information and support services on research, science and technology
  – Incorporated in National Hellenic Research Foundation (NHRF)

• Core technical and scientific areas of expertise:

• Core OA related EC and EU/National projects:
  – “National Information System for Research and Technology”: http://epset.gr
  – “Platform for the Deposit, Management and Delivery of Open Metadata and Digital Content” http://epset.gr/en/SaaS_Services/ (under development)
  – MEDOANET, EuroRIS-NET, RECODE, OpenAIRE/OpenAiRE+
Background

• On going, country-wide efforts and project to digitise a vast Cultural Heritage
  – Make content (where applicable) open & reusable to new uses: move away from “static” “lo-fi” content
  – Provide interfaces in order to built value added services and applications: make a dynamic ecosystem
  – Provide an integrated view of the digitised content available: overcome content fragmentation
• Clearly Open Repositories are a part of the solution
Environment

• Hundreds of organisations with highly acclaimed content and domain expertise:
  – Museums, Archaeological sites and Schools, Cultural Organisations, Libraries, etc.
  – Over a wide geographical location
• High domain expertise but in many small critical size in the technological domain
  – Fast pace of technological changes
  – Not in their’s core business
  – Needs sustained and invested expertise
Opportunity

- Operational Programme “Digital Convergence” (http://digitalplan.gov.gr) framework program:
  - **60M€** work-programme to fund a variety of Digital Culture services and projects, with a part being digital repositories
  - **4M** items are expected to be documented and digitized, among others
  - **75 different organizations** supported by this work programme
  - Very different needs, domains, and areas for each organization
- Develop an “Ecosystem” not a massive top down solution
  - One size fits all out of the question!
- Each organization defines and runs a separate project
  - Individual projects are usually implemented by an external third parties based on a appropriate Public Tender
Problem

• How to ensure the high quality of implementation:
  – Make content (where applicable) open & reusable to new uses
  – Provide interfaces in order to build value added services and applications
  – Provide an integrated view of the digitised content available
• Over so many different organisations, needs and with a rigid public tender procedures in place
• With significant time constraints in place
• And with a one size fits all approach out of the question
Solution

• Part A: Specify a base **Interoperability Framework and Specifications**
  – Generic and based on international best practices
  – **Funding authority mandating** the implementation of this framework for relevant projects funded in this work programme

• & Part B: Provide an **automatic framework Validation Tool**
  – Specific to this framework
  – The Digital Convergence funding authority signifies **successful projects milestone completion (and relevant payments)** based on the validation tool output
The Interoperability Framework

• Ensures a minimum level of interoperability
• Builds on past experiences, other interoperability frameworks, (DRIVER, Europeana, etc)
• Tailored in order to be **readily included in official Public Tenders**
Main Points

Digitisation practices
  – Need for “uncorrected” OCR everywhere
Interoperability at the level of repository:
  – OAI/PMH or SRU/SRW
Interoperability at the level of syntax and structure,
  – DC, ESE
Interoperability at the semantic level
  – Thesaurus, EDM, CIDOC-CRM
Persisted Identifiers
  – Handle system
Best practices for content presentation
  – According to content kind
User Experience aspects
Value added services

• Services provided by EKT to the repositories:
  – Basic bit-wise remote replication of data ("Safe Deposit" service) for organization: Disaster Recovery based on repositories properties
  – Unified Search for all the items over different organisations
  – Readily interoperable with Europeana
  – Validator tool
The Validator Tool

- Influenced by:
  - the Europeana Validation tools, the OAI PMH, OpenAire validators
- Expands on the specifics of the workprogramme
- Plugable definition of validation rules
- Includes also per digital item validation (digitization, OCR, etc)
Features

• Combination of mature and innovative technologies
  – Support: DC, ESE, EDM
  – Capability to validate/store > 4M items
  – noSQL elements (mongoDB), highly clustered backend
  – Provides validation reports as EARL XML format
  – Validation Domain Specific Language for dynamically defining validation rules

• Details at ACM/IEEE JCDL 2014 (Sept 2014, London)
Components

- **Validator Front End**: user management, validation management, validation reporting
- **Validation backed**: Produces JSON, EARL reports and performs the tests
- **Aggregator**: based on Europeana Corelib, stores EDM data and stores them to a SOLR
- **Harverster**
- **Implemented with a SOA approach**, different language for subsystems according to demands
Architecture
Workflows
Current Status

• Interoperability specifications:
  – Tenders conforming to the Interoperability Specifications.
  – Projects implementation is under way

• Organisational, legal and management infrastructure has been setup among EKT, Digital Convergence funding authority and organisations
  – Very important step
  – Not covered here

• Validator Tool:
  – First version implemented
  – First organisation has submit its systems/content and has get the results of the first validation run!

• 10ths more validations to follow on last semester of 2014
• Validation of 75 different data sets by 2\textsuperscript{nd} semester of 2015
Future work

• Update specifications and validator
• Provide the value added services (unified search/safe deposit)
• Provide them as a continued service
Thanks for the attention!

pstath@ekt.gr, @panstath
www.ekt.gr
www.epset.gr

Implemented in the scope of “Platform for the Deposit, Management and Delivery of Open Metadata and Digital Content”
http://epset.gr/en/SaaS_Services/