Beyond open source: a technology assessment of open standards and validation tools in the era of Cloud computing and a SaaS case study

Panagiotis Stathopoulos, Nikos Houssos, Prodomos Tsiavos, George Stavrou

{pstath;nhoussos}@ekt.gr
National Documentation Centre / National Hellenic Research Foundation,
http:// www.ekt.gr
Athens, Greece

Agenda

- The National Documentation Centre
- Open Source and Grey Literature
- Technology Trends and the Cloud
  - The Software as a Service Cloud Model
  - Issues/Problems and Challenges
- Open Standards
- Validation Tools
- Conclusions

The National Documentation Centre (EKT)

- The national organisation in Greece:
  - for scientific documentation, online information and support services on research, science and technology
  - Objective: **making knowledge accessible to everyone**
- Incorporated in National Hellenic Research Foundation (NHRF)
- Implements the "National Information System for Research and Technology": http://epset.gr
  - National CRIS system, Digital Repositories, e-Publishing, Digital Libraries, Interactive Culture, and more…
  - Open Access advocate: http://openaccess.gr, OPENAIRE/OPENAIRE+ member and NOAD.
  - Greek National Aggregator: http://openarchives.gr
  - Repository as a Service (SaaS) and validation services: http://www.epset.gr/en/SaaS_Services

Grey Literature and Open Source

- Open Source: a critical component of our community's technological infrastructure
- Open Source empowered organisations to easily implement:
  - Digital Repositories and Digital Libraries
  - Infrastructures with reduced cost and increased local "know-how"
  - Reduced initial setup cost
  - Provided solutions & tools to the public, the grey literature professionals and organisations

Open Source Assessment

- Open source has being a disruptive force but:
  - look beyond the initial purchase and installation cost
- Indicative IT systems lifecycle:
  1. ***Datacenter/computer room infrastructure,***
  2. ***Hardware initial purchase cost, depreciation, maintenance and support***
  3. ***Initial design, development, customisation***
  4. ***Software maintenance and support, bug fixes, security fixes, new features requested***
  5. ***System administrators, Monitor and Control Loop***
- Thus Open Source is only a part of a full infrastructure solution
  - Is it possible for every organisation to maintain technological capabilities to support the whole application lifecycle efficiently?

Technology Trends: from Open Source to the Cloud

- Cloud technologies:
  - A variety of technology service offerings, with different definitions but with common core elements:
    - Self service
    - Networked
    - Common pool of resources
  - Service Models:
    - Infrastructure as a Service
    - Platform as a Service
    - Software as a Service
  - Central to EUs Digital Agenda 2020
  - Significant economies of scale

The Software as a Service Model

- Software as a Service (SaaS)
  - Complete solution can fully outsource a system
    - Hardware / Middleware/ software development and maintenance
    - Monitor and Control, Operations, and Management
  - Can resemble hosted services but usually with a increased degree of customisation
  - SaaS applications examples:
    - *docs.google.com*, *Microsoft Live*, *Adobe Connect*, *etc.*
- A promise for cost reduction (?)
  - And focus to each organisations core competencies

Grey Literature and the Cloud

- IaaS and PaaS Service Models:
  - Provide new horizontal capabilities (especially PaaS, big data etc)
  - However largely transparent
- Software as a Service Model:
  - Usually Vertical. Systems that could be available as SaaS:
    1. Digital Repositories
    2. Current Research Information Systems
    3. Integrated Library Systems
    4. Digital Preservation
    5. Repository interoperability
    6. Aggregation Services (as centralised services)
  - Full blown solution

New Issues

- So are our troubles end with the Cloud?
  - Some of them
  - Others, more interesting ones, appear:
- Issues:
  - From s/w vendor lock-in to Cloud vendor lock-in?
    - Prepare migration strategy to different systems in order to avoid “cloud lock-in”
  - Are our data exportable and migration capable?
    - Ensure data are “exportable” and export formats are standards-based
    - Ensure interoperability APIs
  - Plus additional issues: is an open source based SaaS based also on proprietary elements and techniques? Security? Cloud Provider long term viability? SLAs monitoring/enforcement?

Grey Literature and Open Standards

- Open Standards provides the communication tools for interfacing different systems, different content using a common “language”
- Structure content:
  - Flat (DublinCore) or rich (CERIF, EDM, etc)
  - Initial cost of implementation but reduced long term cost and increasing viability
- Independent from Software, Implementation method and Service Model
  - Increasingly critical factor to a number of additional applications/services

Open Standards

- We know the significance of open standards for interoperability/aggregation/etc
- Additionally standards, and standard format increasingly critical for:
  - Migrating our data among SaaS providers
  - Avoid SaaS provider lock-in
    - Create a “SaaS” market
  - Avoiding closed not interworking systems

A Repository as a Service case study

- EKT is Developing a “Repository as a Service” SaaS for eligible Greek organisations:
  - provide Digital Repositories as a Service, for scientific publications, grey literature, cultural institutions and archives
  - Build on EKT’s experience for developing and operating repositories for third parties
  - http://epset.gr/en/saas/ :
    - First pilots (semi automated) ready.
    - Next goal: further automation of tasks
  - Open standards and validation tools in order to increase third party organisations trust

The need for guidelines

- Metadata and functionality validation tools are critical for:
  - Repository as a Service development
  - Digital content interoperability
  - Aggregation mechanisms
- EKT has specified basic interoperability guidelines for digital repositories:
  - http://hdl.handle.net/10442/8887
    - English translation under way
  - Applicable both if development by EKT, provided as SaaS or developed externally
- Ensures a minimum level of interoperability independent of
  - Software
  - Implementation method
  - Delivery method (in house, outsourced, or Cloud)

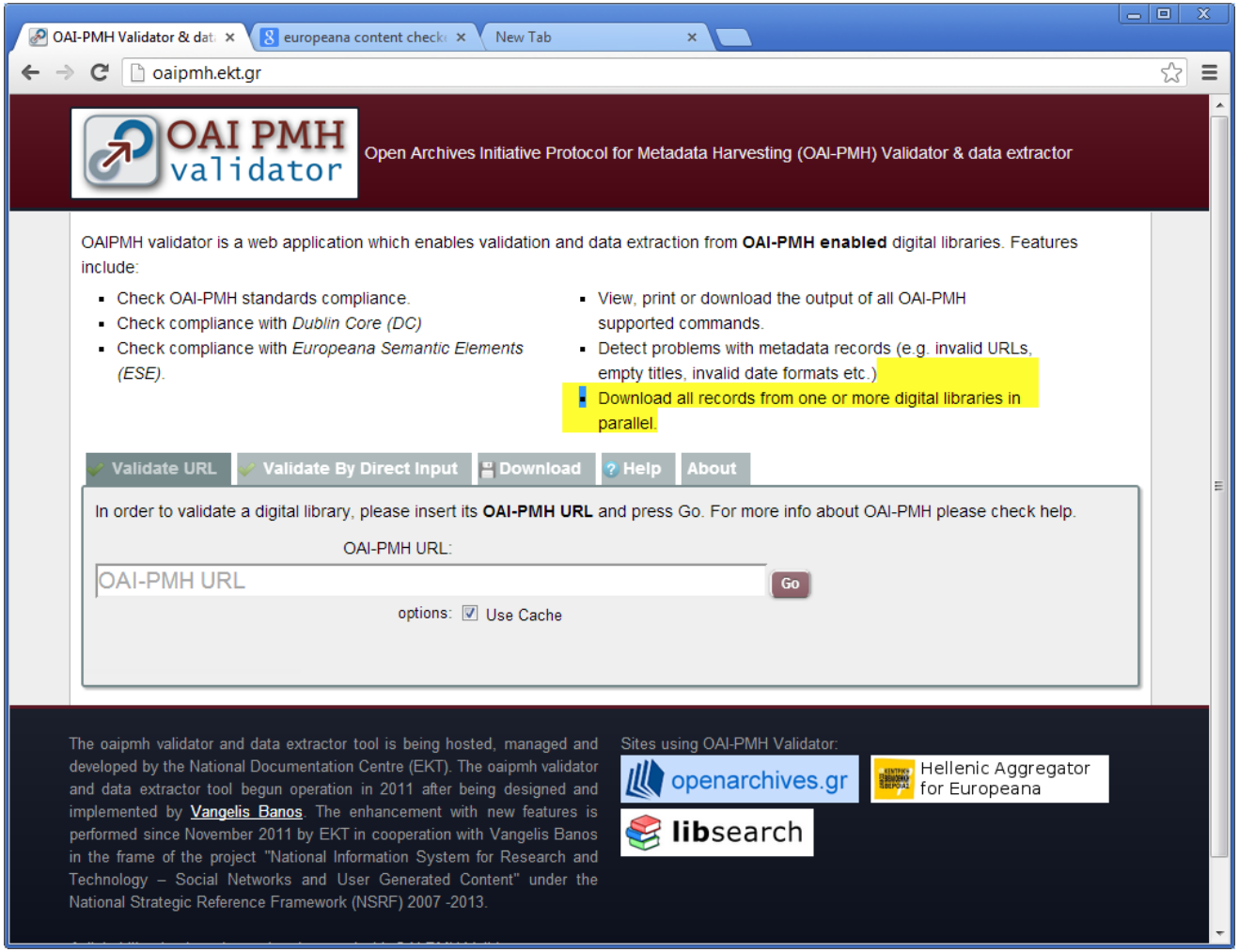
Mandates

- Could combine guidelines, funder's mandates and funding in order to guarantee high quality projects/outcomes
- Case study:
  - The Greek Digital Convergence Funding authority mandated that **digital repositories must implement Digital Content Interoperability guidelines**
    - Call mandated: http://www.digitalplan.gov.gr/portal/resource/Prosklhsh-31-Politismos
    - >75 funded organisations, 65M€ of funding.
    - Focused on digital culture but includes Grey Literature related content

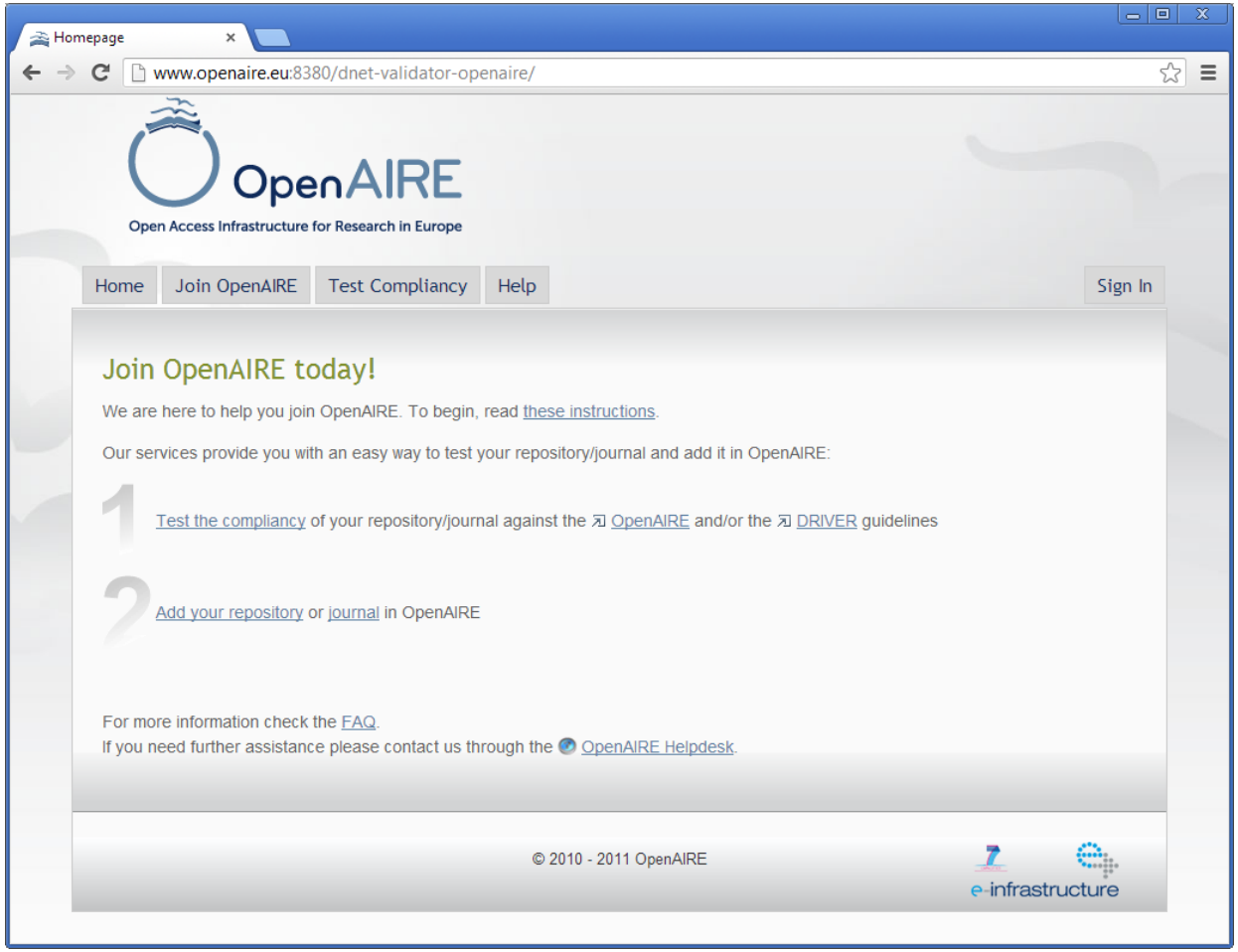
The need for validation tools

- Automatic validation of guidelines critical
  - Link automatic validation to value added services
  - E.g.: aggregator harvesting (OpenAIRE+, EUROPEANA) to funder mandate (Digital Convergence/EKT)
- Various validation tools for various standards:
  - Free, open, or project specific
  - E.g. http://oaipmh.ekt.gr , http://www.openaire.eu:8380/dnet-validator-openaire/, Europeana Content Checker, etc.
    - (*also CERIF validation tool under development*)
- Multi-level Validation of repositories
  - Cultural, archival, etc
  - Generic and specific cases

Validator Example (1)



Validator Example (2)



Validation Benefits

- Ensure wide interoperability and aggregation
- Avoid “data” lock-in, ensure capability to transfer content and service among:
  - Different Digital Library/Repository software
  - Different SaaS cloud providers
    - Exploit advantages of Cloud without “lock-in” dangers
- Ensure high quality of funded Digital Repositories and Libraries
- Continuous implementation of the chosen guidelines

Conclusions

- Open Source was (and is) a key driving factor for Digital Libraries
- We must expand the interoperability and flexibility capabilities Open Source has provided
  - While exploiting where applicable Cloud and SaaS resources
- How?
  - Open Standards for content
  - Interoperability specifications and guidelines
  - Automatic validations tools
  - Aggregation Services

Thank you for your attention!

pstath@ekt.gr
nhoussos@ekt.gr