A Europe-wide Interoperable **V**irtual **R**esearch **E**nvironment
to **E**mpower Multidisciplinary Research Communities
and Accelerate **I**nnovation and **C**ollaboration

Towards contextual awareness and interoperability in an enhanced VRE: mapping individual metadata schemata to CERIF

Maria Theodoridou, Theodore Patkos, Martin Doerr

Foundation for Research & Technology – Hellas (FORTH)

Institute of Computer Science

VRE4EIC Fact Sheet

A Europe-wide Interoperable **V**irtual **R**esearch **E**nvironment to **E**mpower Multidisciplinary Research Communities & Accelerate **I**nnovation and **C**ollaboration

- **Call:** H2020-EINFRA-9-2015, e-Infrastructures for virtual research environments
- **Funding scheme:** RIA-Research and Innovation action over 3 years
- **Proposal acronym:** VRE4EIC
- **Consortium:** 8 partners from 4 countries
- **Started**: October 2015
- **EU contribution**: 4.370.000 €

Consortium

- **ERCIM** management and scientific coordinator;

- ERCIM members (**CWI, CNR, FORTH**) bring in IT development expertise and familiarity with standards and standards development;

- **ERCIM-W3C** brings in expertise in gaining community support for the adoption of standards;

- **euroCRIS** brings in expertise on advanced metadata, metadata interoperability and standards;

- **TU Delft** brings in analytical and policy aspects and expertise in open data and the development of training material;

- ENVRI (**University of Amsterdam**) and EPOS (**INGV Instituto Nazionale Geofisica e Volcanologica**) bring in VRE use cases and evaluation;

The Vision

- Develop an "e-VRE"
- Interoperation across 'silo' VREs
- Achieved by:
  - A VRE Reference Architecture
    - New VREs to adopt the architecture
    - Existing VREs to evolve to the architecture
  - Building blocks of software services to realise an enhanced VRE and to expose / promote / interoperate with / utilise existing e-Research infrastructures providing services, software, data and resources;
  - Standardisation & Training
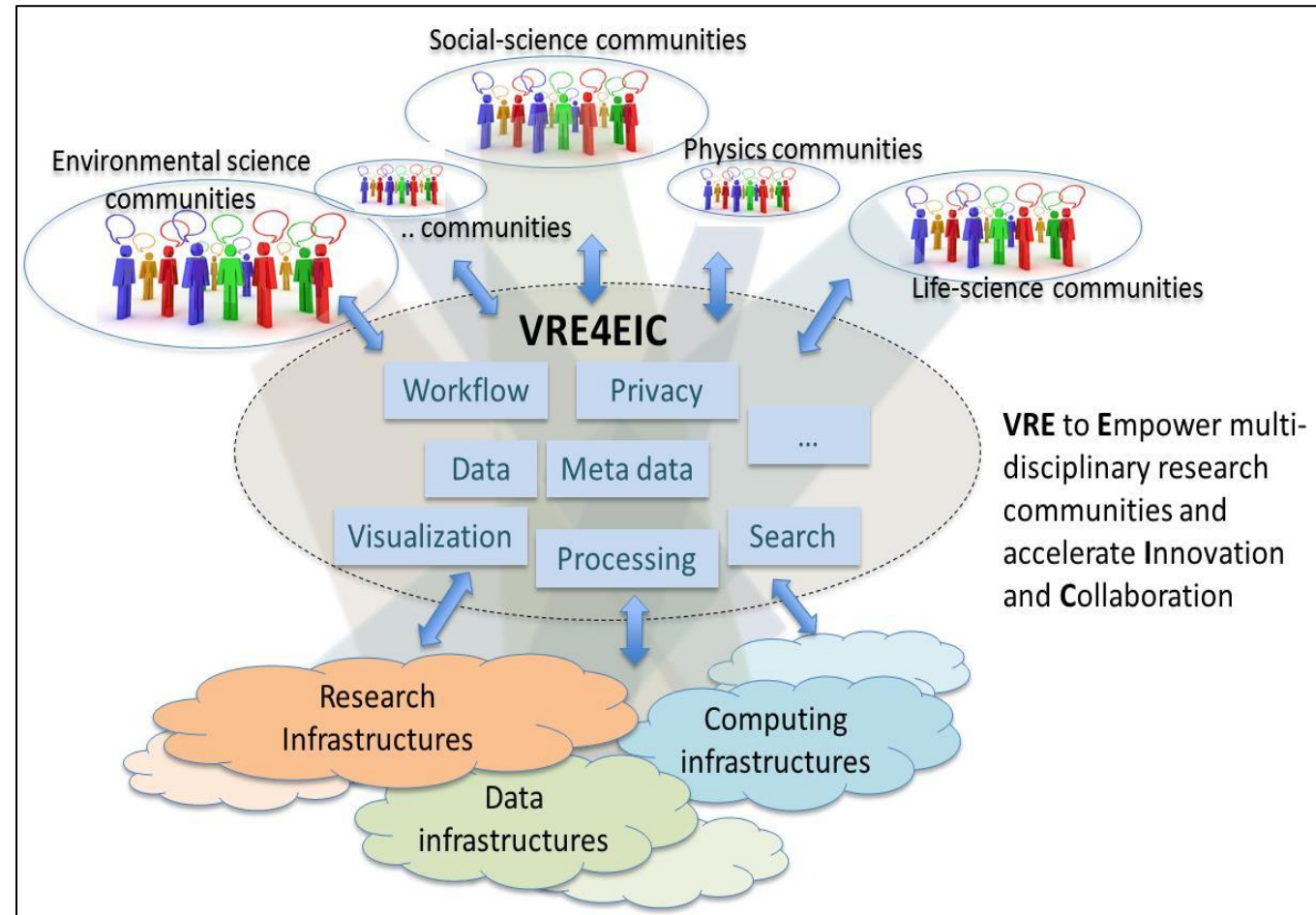
Objectives

- **Increase the VRE usability for multi-disciplinary research;**

- **Increase the quality of VRE user experiences ;**

- **Increase the deployment of the VRE on different research infrastructures by abstracting and reusing building blocks and workflows;**

- **Improve the contextual awareness and interoperability of the metadata;**

- **Promote the exploitation and standardisation of e-VRE.**
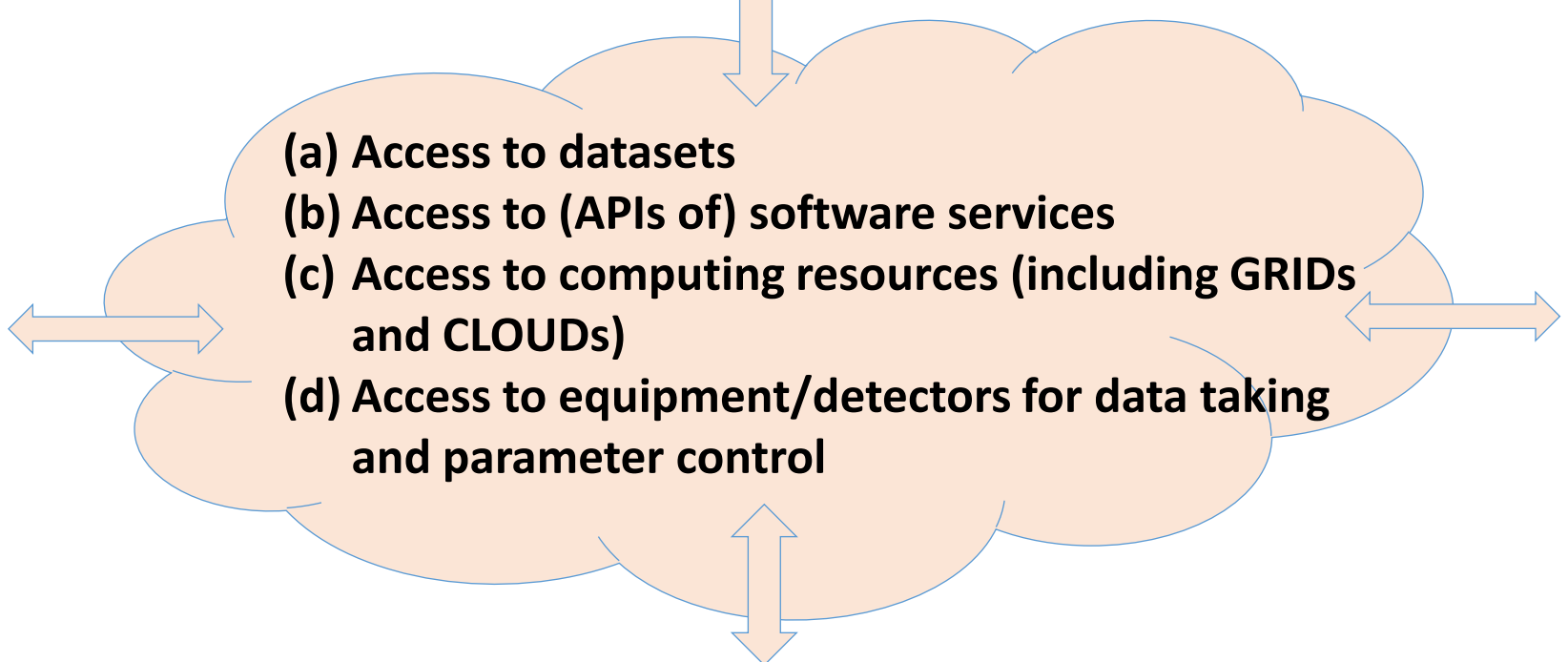
Requirements, Steering, Validation

- **Requirements** condensed from many existing and developing VREs

- **Steering** from two very large VREs (EPOS, ENVRI+) each providing homogeneity over ~20 heterogeneous RIs (Research Infrastructures)

- **Validation** by EPOS and ENVRI+ then wider – any interested VRE

- **Architectural approach** generalised from EPOS and ENVRI+ (and influenced by RDA)
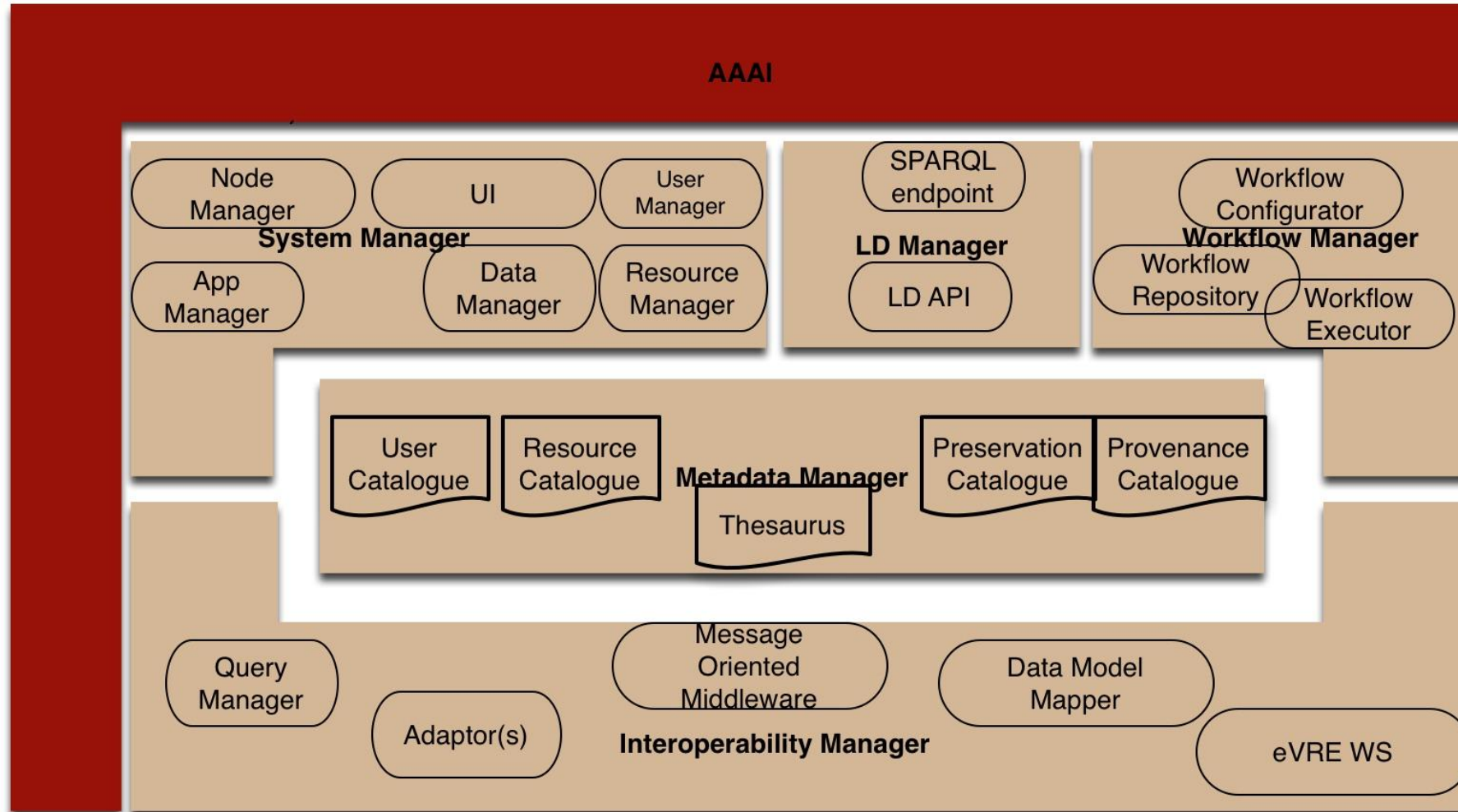
Impact

- Overall, the VRE4EIC partners have access to **43 VRE related projects and initiatives**, from which they will obtain knowledge

- **25 real use cases** will be developed, addressing relevant societal and economic challenges,

  e.g., unemployment, social exclusion, healthcare, security, climate change, energy sustainability, capital provision and others.

- **>15 workshops** and **>20 training sessions** will be organized during the course of the project

- The project will contribute to standardisation by setting up a **W3C Community Group**

Contextual awareness & interoperability

(a) Access to datasets
(b) Access to (APIs of) software services
(c) Access to computing resources (including GRIDs and CLOUDs)
(d) Access to equipment/detectors for data taking and parameter control

e-VRE architecture

CERIF-Common European Research Information Format

The CERIF metadata schema will act as the 'switchboard' in the e-VRE's metadata manager enabling **interoperability** and **contextual awareness** among RIs.

Diverse metadata schemata describing RIs, datasets, software (services) and publications from different VREs will be mapped to a common metadata catalogue using CERIF.

Mapping one schema to another

A sufficient specification for the transformation of each instance of a source schema into an instance of a target schema while preserving as much as possible its initial 'meaning'

- interpretation of source schema as semantic model (nodes and links),
- mapping each element of that to an equivalent path of the target schema,
- such that each instance of an element of the source semantic model can be converted into a valid instance of the target with the same meaning.
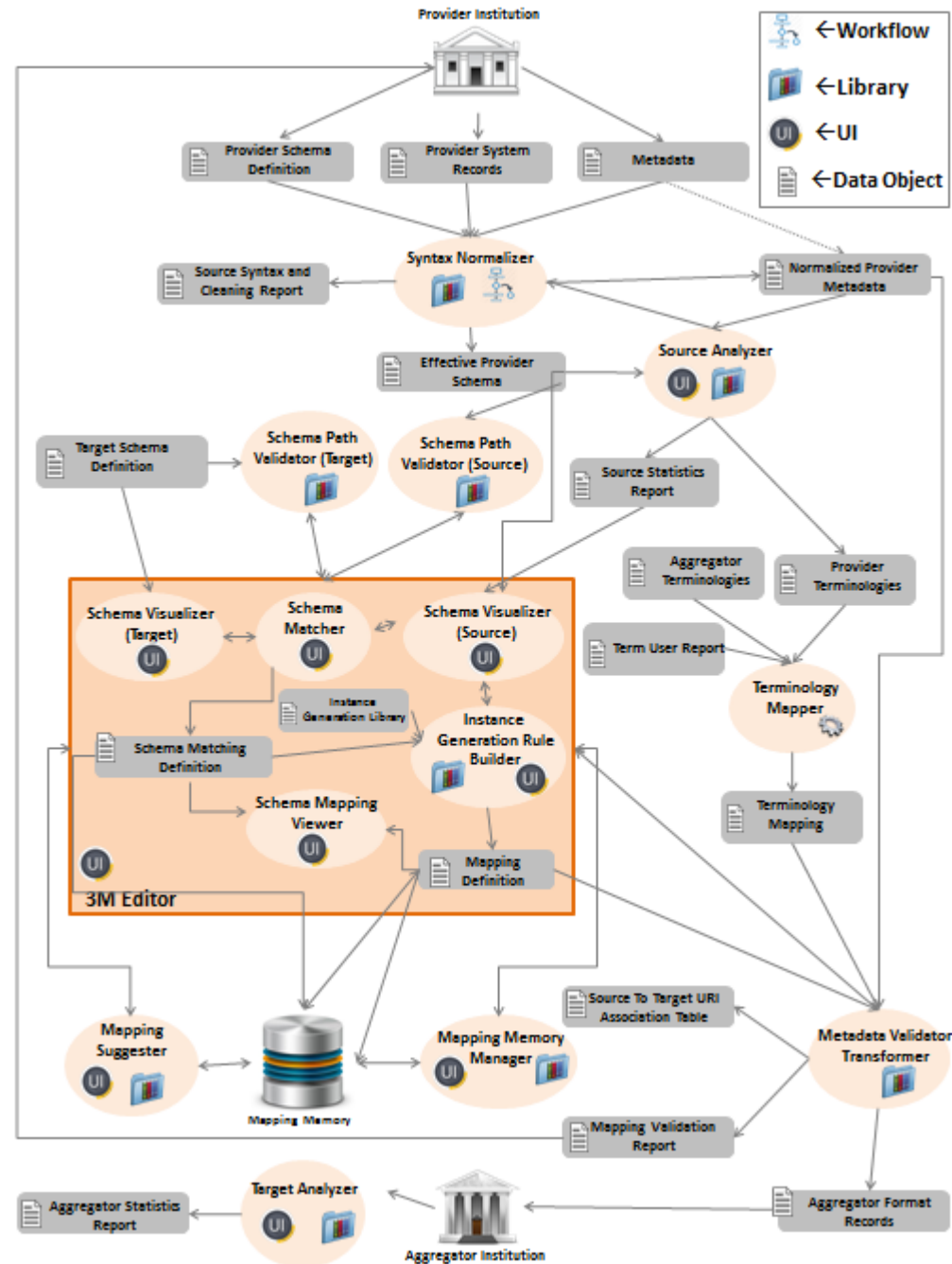- CERIF as target schema
- CERIF as source schema

Synergy Reference Model

A **reference model** for a better practice of **data provisioning** and **aggregation** processes
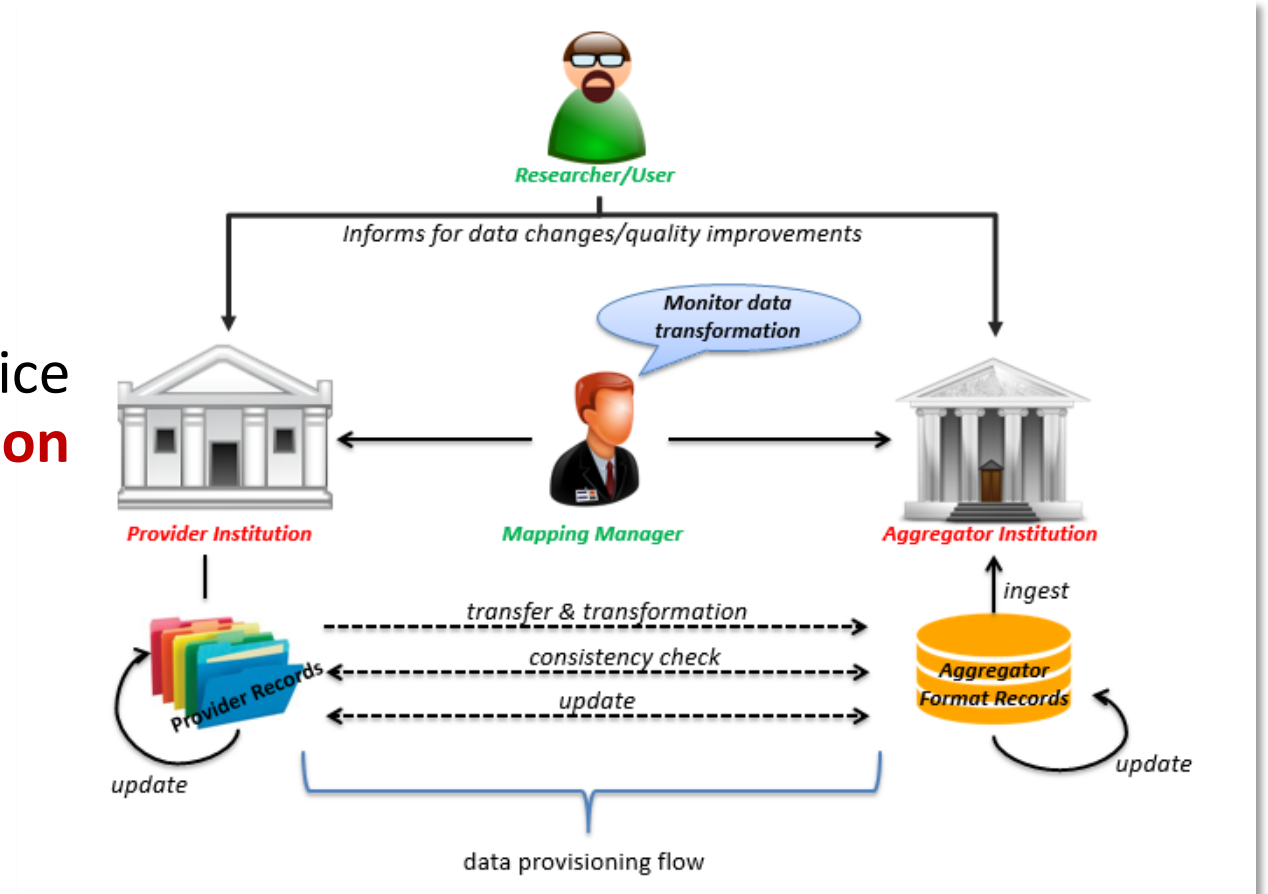
Synergy Reference Model

X3ML mapping framework

- X3ML mapping definition language

- 3M Mapping Memory Manager

- X3ML engine

Y. Marketakis, N. Minadakis, H. Kondylakis, K. Konsolaki, G. Samaritakis, M. Theodoridou, G. Flouris, M. Doerr (2016). X3ML mapping framework for information integration in cultural heritage and beyond. International Journal on Digital Libraries, June 2016, DOI: 10.1007/s00799-016-0179-1

N. Minadakis, Y. Marketakis, H. Kondylakis, G. Flouris, M. Theodoridou, G. de Jong, M. Doerr: X3ML Framework: An Effective Suite for Supporting Data Mappings. Proceedings Workshop EMF-CRM2015, Poznań, Poland, September 17, 2015, http://CEUR-WS.org, online http://CEUR-WS.org/Vol-1656/paper1.pdf

X3ML Workflow
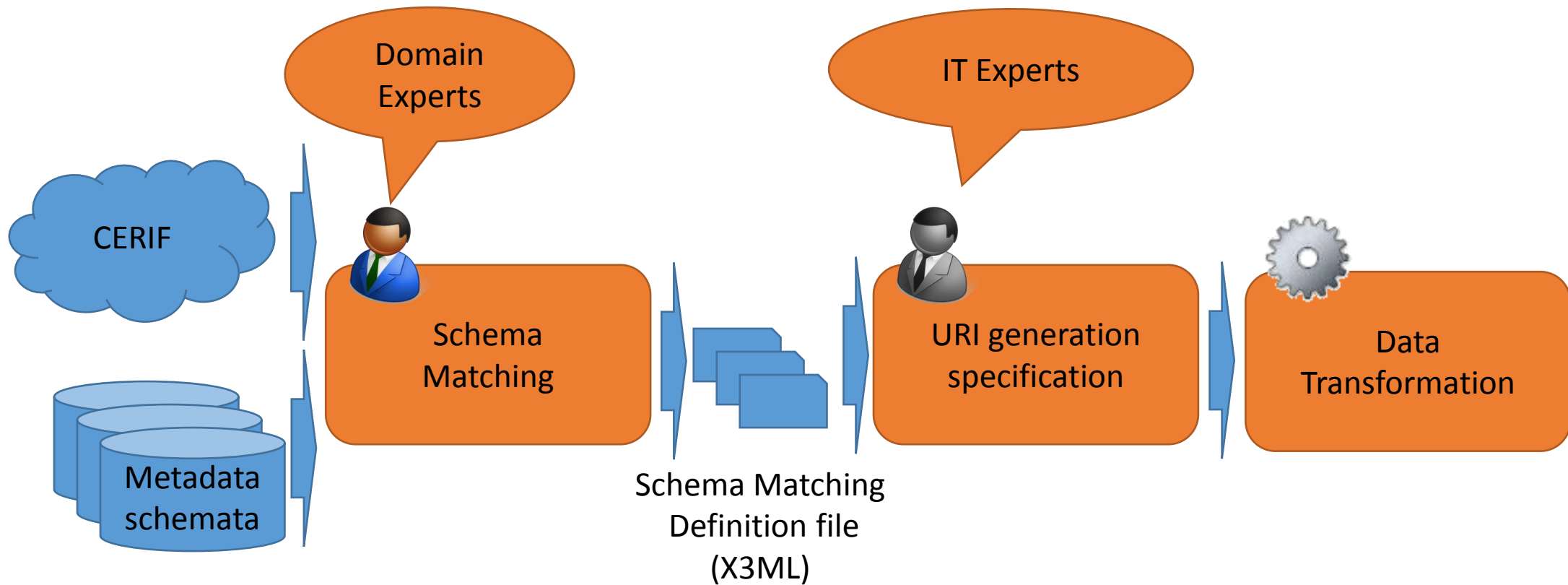
X3ML mapping definition language

X3ML is an XML based language which describes schema mappings in such a way that they can be collaboratively created and discussed by experts.

Mappings have been done in very many custom ways in the past.

In practice mappings are produced manually by Domain/IT experts:

- labor-intensive
- error prone
- time consuming

Emphasis is on establishing a standardized mapping description which lends itself to collaboration and the building of a mapping memory to accumulate knowledge and experience.

3M- Mapping Memory Manager

3M is an open source system that allows domain experts to build and discuss mappings with little resource to any particular software skills. It offers administrative facilities for **searching, filtering, sorting and exporting** mappings and a graphical interface for **generating, visualizing and validating** mappings.

It provides:

- Source and target agnostic mapping facility
- Guided mapping according to deployed ontology's logic
- Comment and justification facility
- Mapping storage
- Separated instance generation practice for IT professionals

X3ML engine

- The X3ML engine performs the data Transformation based on the mapping definition and produces a set of valid target records.

- The transformation process may run completely automatically.

- If any issues arise:

  - the aggregator can resolve them on a temporary or permanent basis or

  - records are send back to the provider for further analysis and resolution

Mappings to CERIF

- In progress: definition of CERIF encoded in RDF

- Automatic download of CERIF XML instances from EKT http://cc-refim.ekt.gr/cerifcris/ using the CERIF REST API to use them for testing the services

- In progress: mapping of CERIF XML to CERIF RDF in order to transform the EKT XML data to RDF instances.

Mappings to CERIF

- OIL-E to CERIF and CERIF to OIL-E
  (OIL-E is the conceptual model used within the ENVRI+ project by the underlying eRIs)
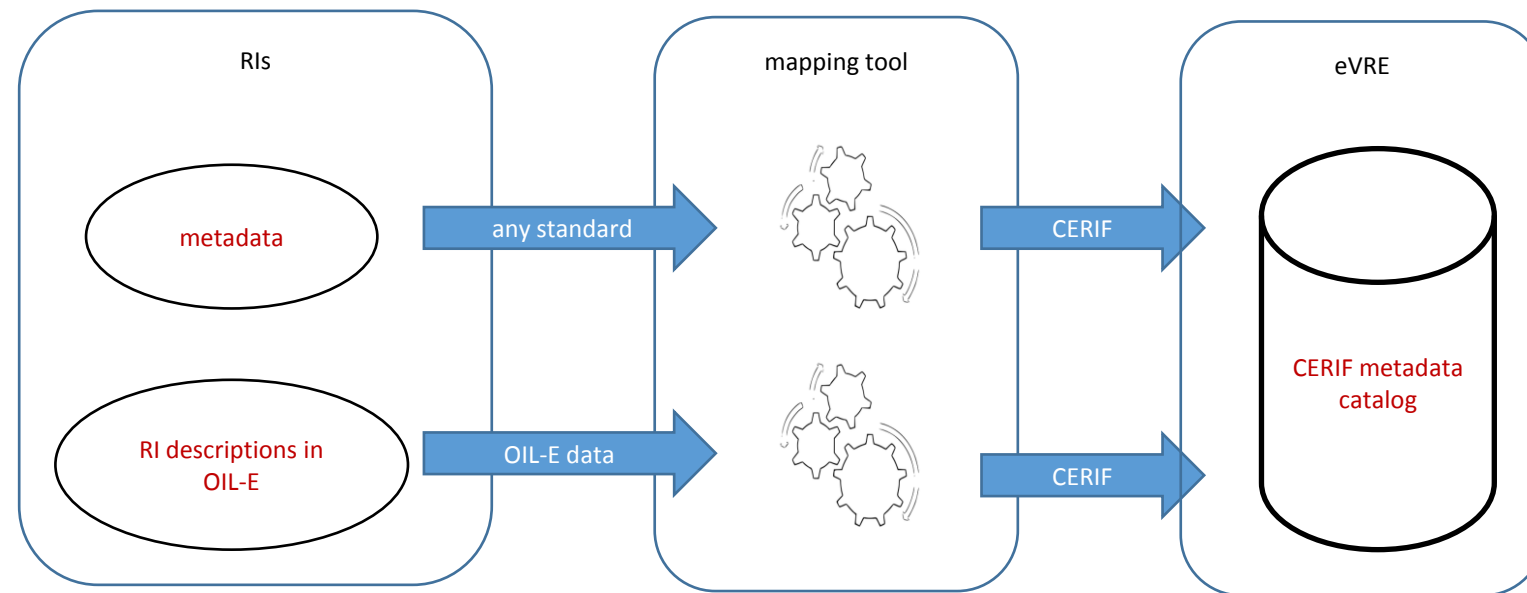

figure credit to Laurent Remy

Mappings to CERIF

Planned mappings for:

CKAN: the metadata structures of CKAN (Comprehensive Knowledge Archive Network) are widely used for Open Data.

DCAT – the W3C Data Catalog Vocabulary

INSPIRE / GeoDCAT-AP

GeoDCAT-AP provides a DCAT-AP compliant representation for INSPIRE / ISO 19115 metadata

Dublin Core metadata

Conclusions

- Mappings to CERIF are an significant step to
  - expose ;
  - promote;
  - interoperate with ;
  - utilise

existing e-Research infrastructures providing services, software, data and resources in an enhanced VRE .

- The aggregation of the heterogeneous data from existing e-Research infrastructures in the e-VRE, has the potential to create rich data resources useful for a range of different purposes, from research to education and public interests.

- The X3ML framework supports the aggregation process . It provides tools for:
  - describing both schema mappings and URI generation policies;
  - managing, editing, visualizing and executing the mappings.

Useful links

Visit VRE4EIC at www.vre4eic.eu

Follow VRE4EIC on Twitter: @VRE4EIC

Survey:

http://www.vre4eic.eu/publications/news/92-vre4eic-online-survey-to-elicit-the-user-requirements-for-a-virtual-research-environment

X3ML mapping framework (including links to github)
http://www.ics.forth.gr/isl/index_main.php?l=e&c=721

VRE4EIC Instance of 3M: http://139.91.183.3:8084/3M/

Contact us: Maria Theodoridou maria@ics.forth.gr

Acknowledgment

The VRE4EIC project has received funding
from the European Union's Horizon 2020 research and innovation
programme under grant agreement No 676247

THANK YOU!