

LusTRE: Linked Thesaurus fRamework for the Environment

P. Podestà, R. Albertoni, M. De Martino (CNR- IMATI)
A. Abecker, R. Wössner (disy)

■ LusTRE Overview

- Context
- Strategy
- Outcomes

■ LusTRE Components

- LusTRE-VOC (knowledge infrastructure)
- LusTRE-WEBE (web exploration tool)
- LusTRE-ES (exploitation services)

■ LusTRE Exploitation

- Stand alone application
- Third party client application

■ LusTRE Indicators

■ Planned Activity and Conclusion

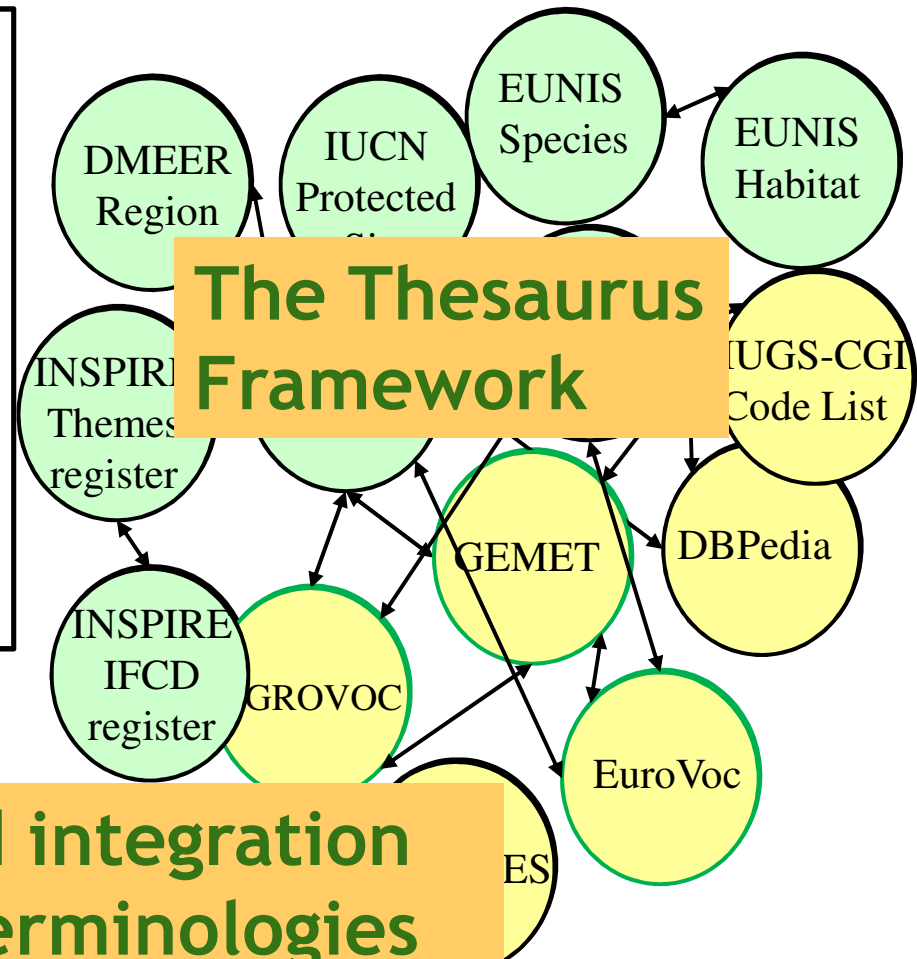
LusTRE Overview

Motivation EU projects Nature SDI and eENVPlus:

- Several environmental terminologies
- Heterogeneity wrt thematic coverage, multilingualism, granularities, popularity in certain communities



- Reuse and integration existing terminologies
- Crosswalking between terminologies



Thesaurus Framework as solution to
the multilingual and multicultural issues in
the environmental data sharing

- ❑ A “common terminology” for the environment to
 - Provide a widely shared concepts
 - Joint exploitation of available terminologies referring to different INSPIRE data Theme
- ❑ A set of services to exploit the Thesaurus Framework
 - To promote uniform data description during metadata provision
 - To improve resource discovery across applications and platforms

INSPIRE Implementation rules

recommend the adoption of (multilingual) thesauri when
compiling metadata for data/services

□ Linked Thesaurus fRamework for Environment (LusTRE)

- A **knowledge infrastructure** of linked Thesauri (**LusTRE-VOC**)
- A personalized user-friendly **Web interface and exploration tool** for search and browse information, like a LD front end (**LusTRE-WEBe**)
- A set of **Web Services** to exploit the knowledge infrastructure (**LusTRE-ES**)

Modular
To add new
plugged
TF

Openness

To easily extend
keeping separate

Interlinking

Linking among the
to the same concept
one thesaurus in order
their usage.

Exploitability

To encode in a standard and flexible format
in order to encourage the adoption and its
enrichment from third party system

LusTRE Components

LusTRE- VOC

- Identification of suitable environmental vocabulary according with “reusability” criteria (Licence openness and LD compliance: dereferenceable uri, 5 star classification)
- Vocabularies processing and publication (LodRefine tool - SKOS/RDF)

Automatic generation interlinking (SILK tool) and validation with the Expert community

- Interlinking among vocabularies inside LusTRE
- Interlinking towards external LD vocabularies

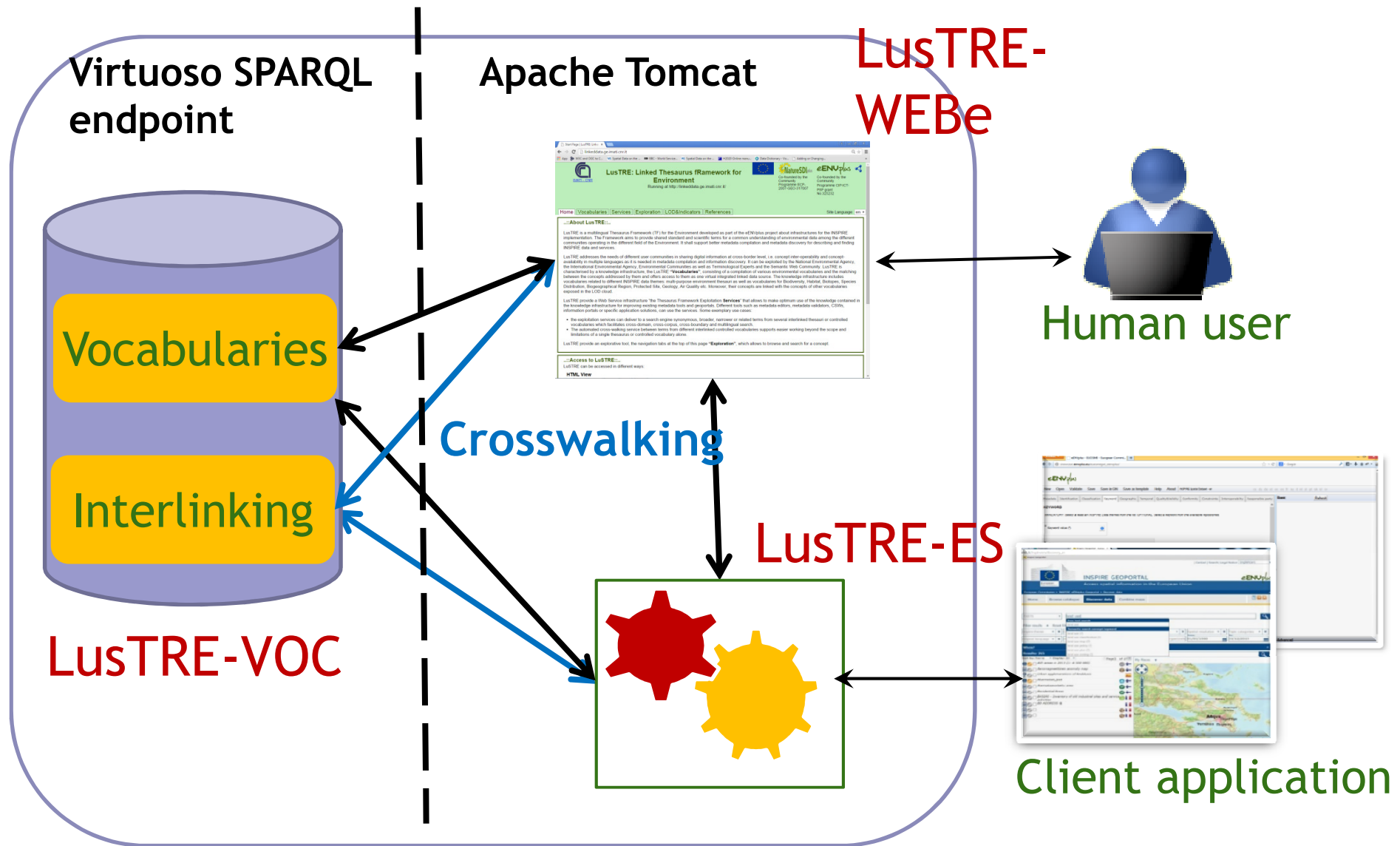
- server setting up and maintenance based on VIRTUOSO

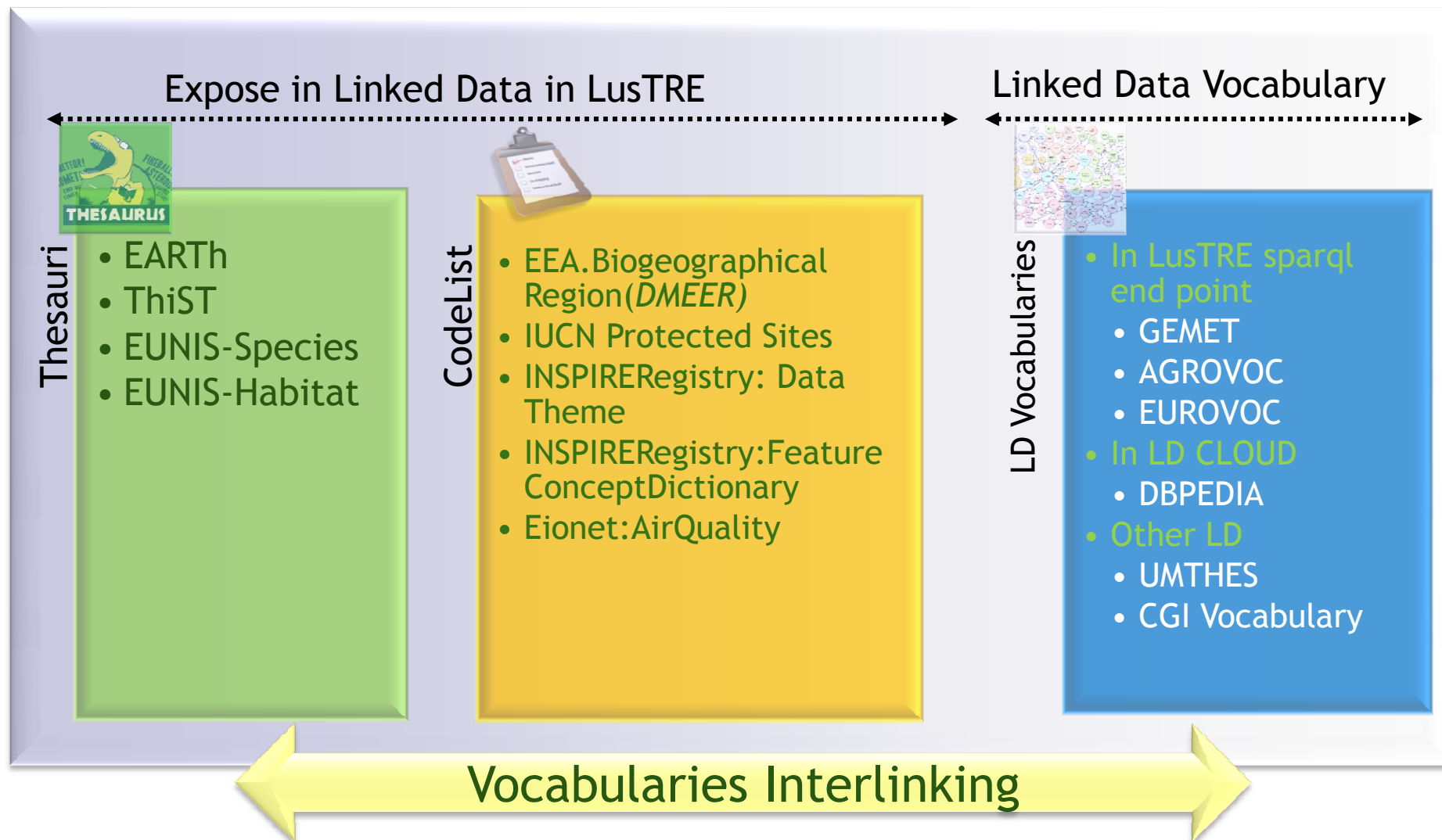
LusTRE- WEBe

Design and development of smart Web exploration interface suitable for human-readable browsing, deployed in Apache Tomcat

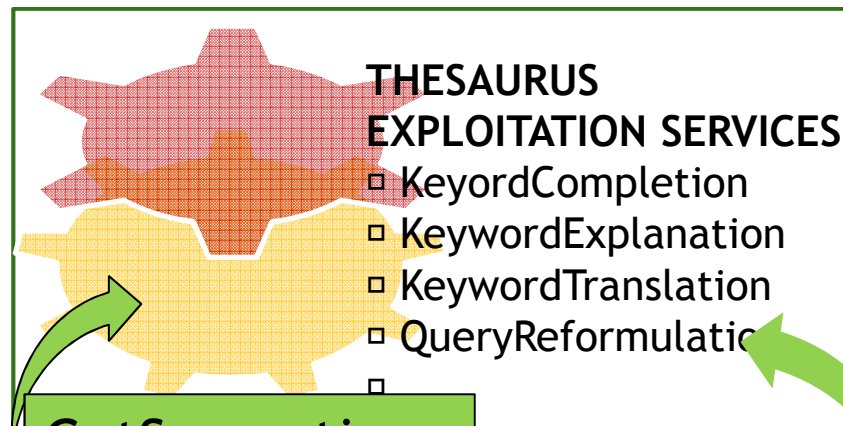
LusTRE- ES

In-house java-based web services, deployed in Apache Tomcat

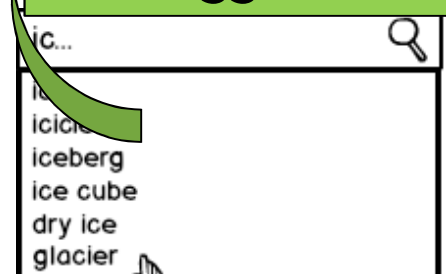




LusTRE-ES



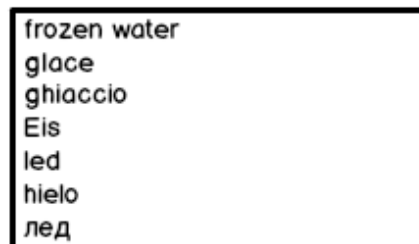
GetSuggestions



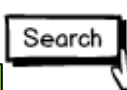
definition of "glacier" ...

broader: glaciology
narrower: glacier tongue,
glacier cave

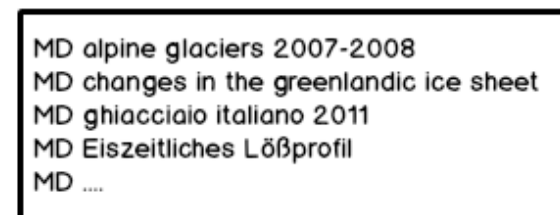
keyword completions
(metadata
compilation)



Search automatically
synonyms, translations
(metadata discovery)



Improving Metadata
Data Search



Results from different
multilingual, cross-domain
sources (query refinement
with crosswalking)

HTTP REST interface

GetCapabilities
GetSuggestions
GetSynonyms
GetRelatives
Resolve Thesaurus
GetTopMostConcepts



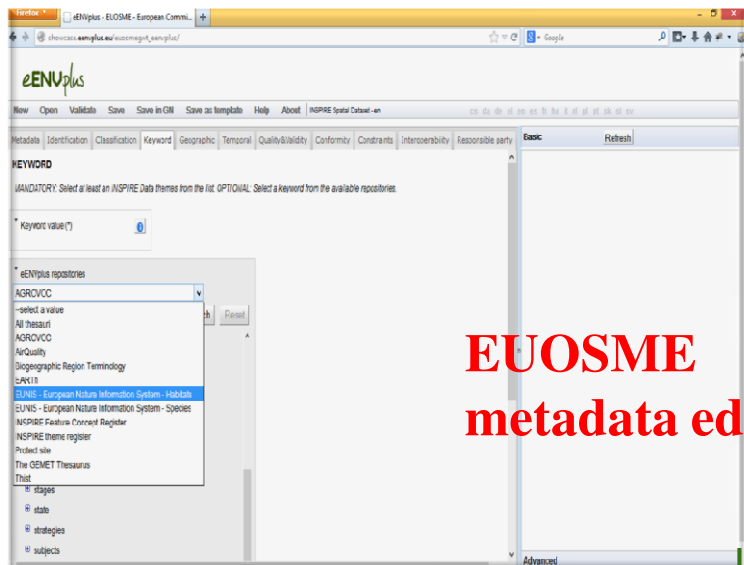
The screenshot shows the LusTRE-WEBe Exploration tool interface. The browser address bar displays 'data.ge.imati.cnr.it/exploration.jsp'. The page header includes the eENVplus logo, the European Union flag, and logos for NatureSDIplus and eENVplus. Text on the right side of the header states: 'Co-founded by the Community Programme ECP-2007-GEO-317007' and 'Co-founded by the Community Programme CIP-ICT-PSP grant No.325232'. The main navigation bar contains links: Home, Vocabularies, Services, Exploration, LOD&Indicators, and References. The 'Exploration' link is highlighted. Below the navigation bar, there are three main sections: 'Search&Browse Concepts', 'Translation', and 'Results'. The 'Search&Browse Concepts' section includes a 'Browse' list with links like 'BiogeographicalRegion', 'EARTH', 'EUNISHabitat', etc., and a search form with a text input (containing 'mare'), a language dropdown (set to 'it'), a 'num max results' dropdown (set to '25'), a 'vocabulary' dropdown (set to 'EARTH'), and a 'Search' button. The 'Translation' section has a 'Translate: Keyword' input, a 'vocabulary' dropdown (set to 'All'), a 'CrossWalking Info' checkbox, and a 'Translate' button. The 'Results' section shows a list of results for the keyword 'mare' under the 'EARTH' vocabulary, including terms like 'marea nera (it)', 'marea verde (it)', 'marea (it)', 'mareggio (it)', 'mare profondo (it)', 'marea rossa (it)', 'mareografo (it)', and 'maremoto (it)', each with a corresponding URL. Three yellow callout boxes are overlaid on the image: one pointing to the 'Vocabularies' link in the navigation bar, one pointing to the 'Exploration' link, and one pointing to the 'Translation' section.



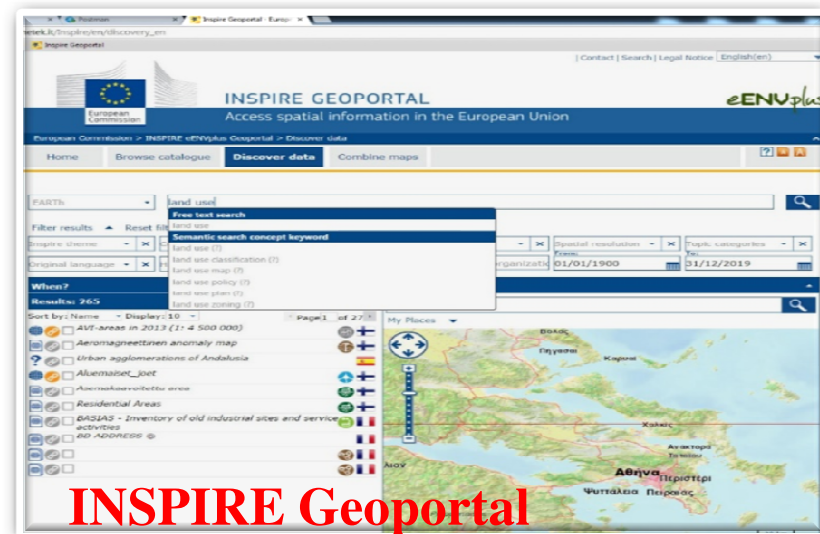
□ For metadata management:

■ Metadata description INSPIRE Compliant

- EUOSME Metadata editor: to select a concept of LusTRE and its associated URI as metadata keyword (<http://showcase.eenvplus.eu/client/editor.htm>)
- “INSPIRE” Geoportal Data (awarded to Planetek): to search by keyword (<http://inspire-geoportal-eenvplus.planetek.it>)

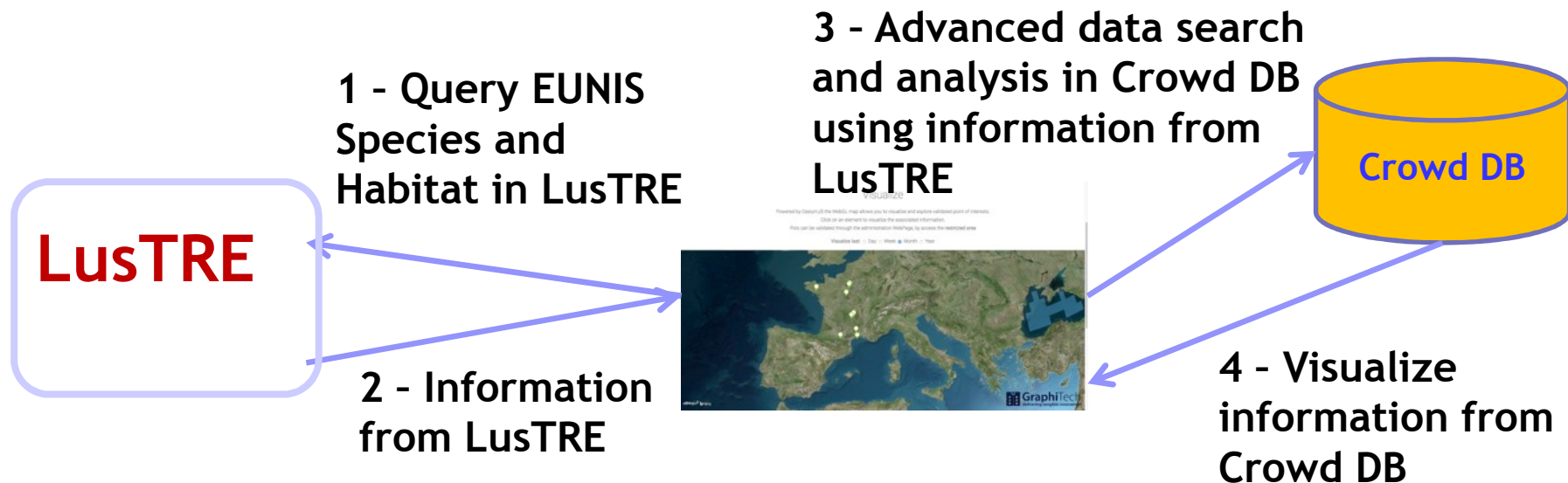


**EUOSME
metadata editor**



INSPIRE Geoportal

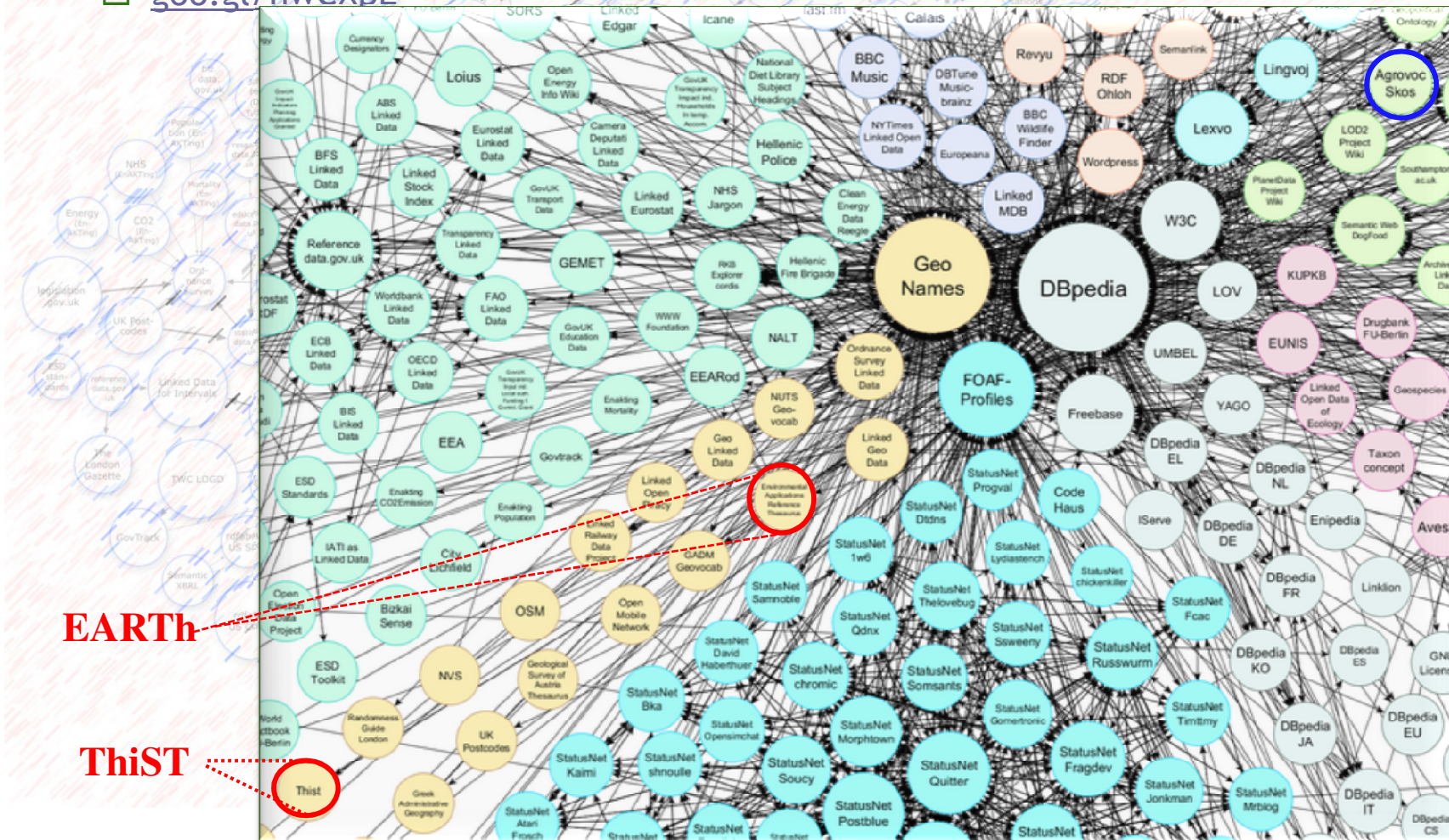
- For data analysis (pilot of EU project eENVplus)
 - (under development) Advanced data search and analysis of Species crowdsourced data in the eENVPlus Crowdsourcing DB (CrowdDB)



LusTRE Indicators

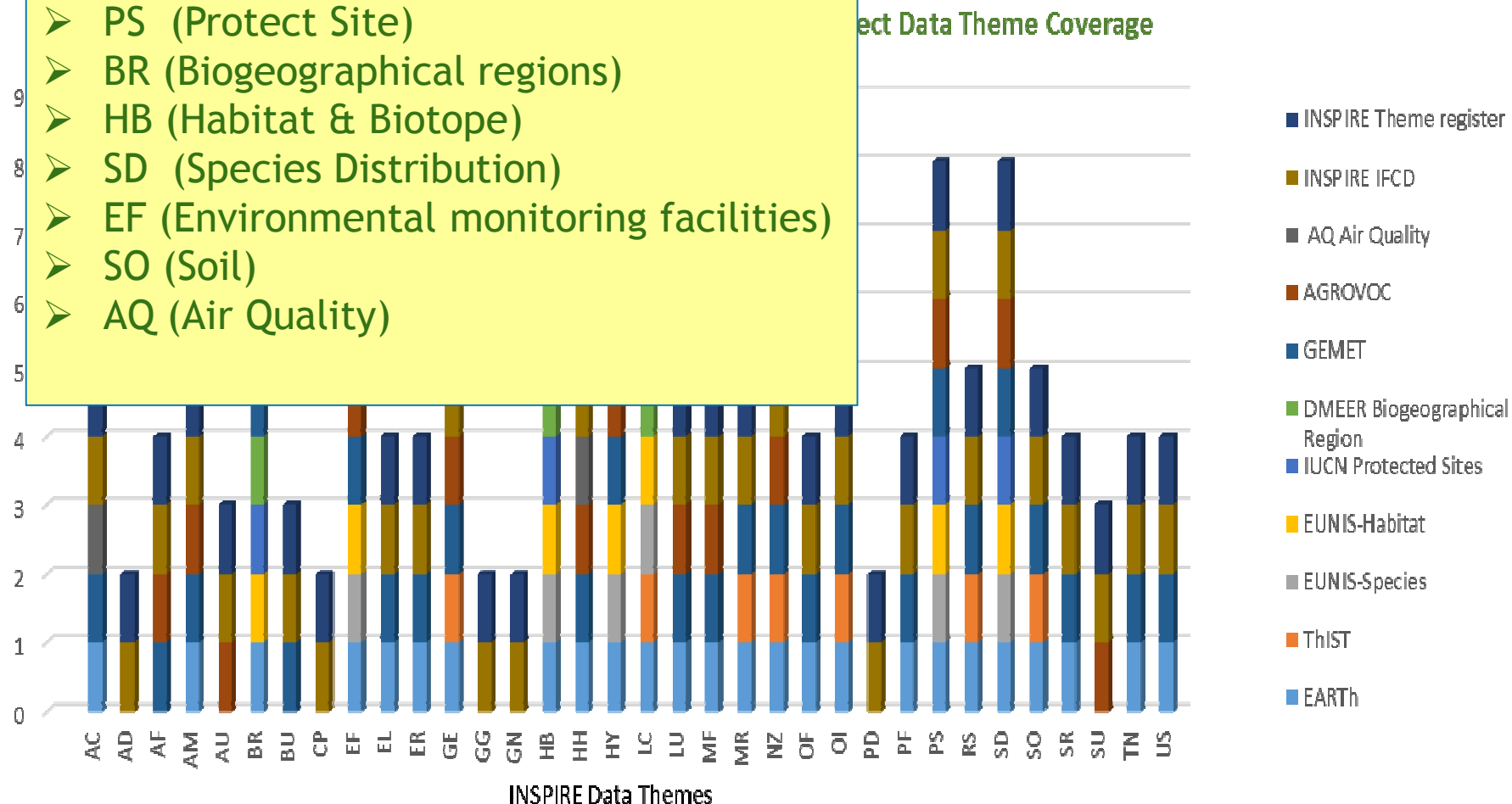
- EARTH and ThiST are included in the latest LOD Cloud 2014

□ goo.gl/hwCXpE

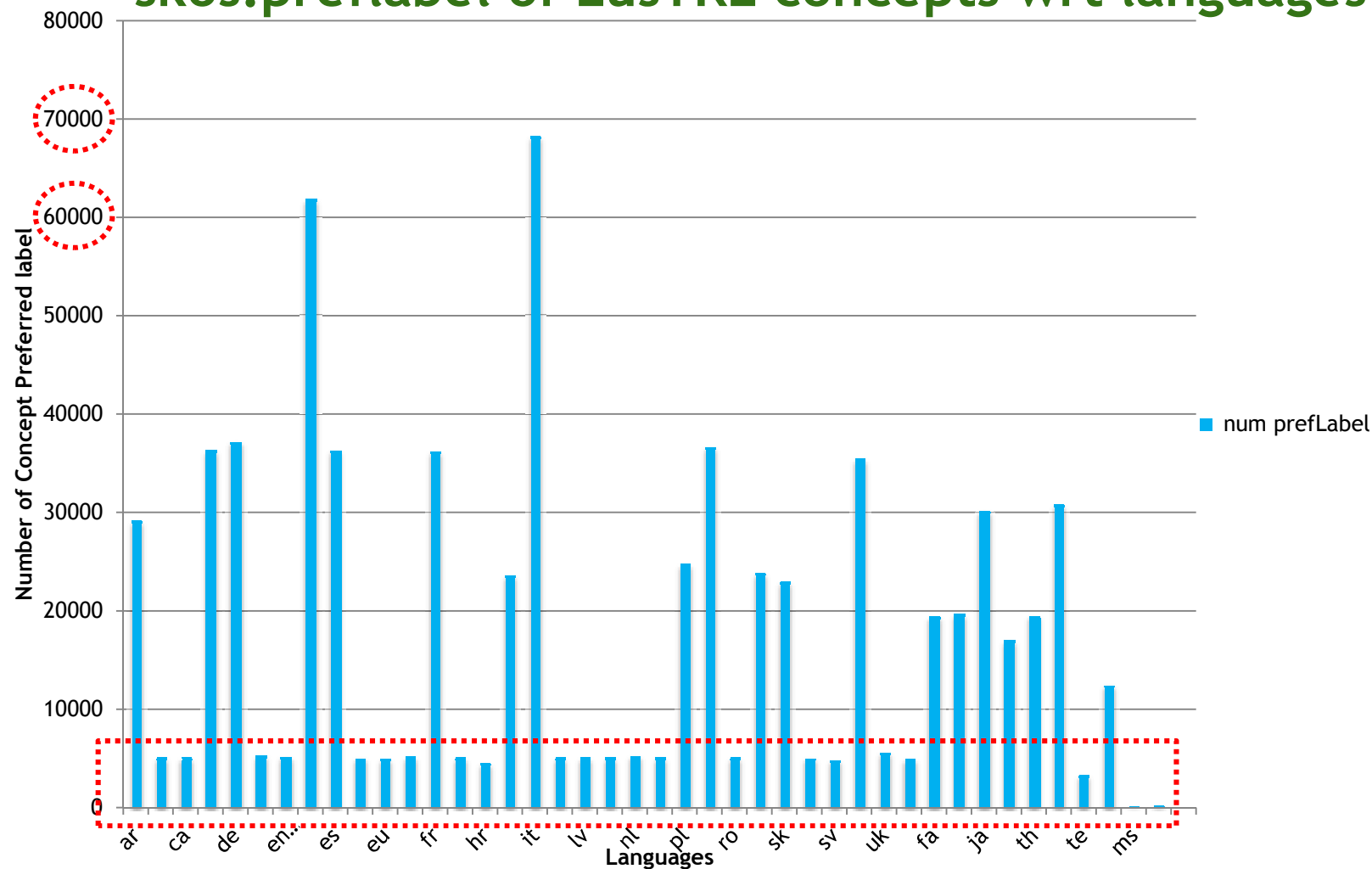


Specific Data Theme coverage

- PS (Protect Site)
- BR (Biogeographical regions)
- HB (Habitat & Biotope)
- SD (Species Distribution)
- EF (Environmental monitoring facilities)
- SO (Soil)
- AQ (Air Quality)



skos:preflabel of LusTRE concepts wrt languages



□ log access numbers :

131033 access to LusTRE - period 25 September to 9 December

3962 access to services - period 4 November to 9 December



Map of access locations done during December 2014

LusTRE Conclusion

What is Next ...

☐ Outcomes: LusTRE

- ☐ Knowledge Infrastructure for a “common” terminology for “All”

- ☐ Web Services and web interface and exploration tool

 - ☐ to metadata compilation

 - ☐ to information discovery

☐ Next release

- ☐ New services LuSTRE - ES: Semantic explorative search with semantic visualization (June)

- ☐ Web Application exploiting LuSTRE to analyse Species crowdsourced data

■ For more information

- ☐ Attend the workshop: «eENVplus» next Friday
- ☐ Visit the stand «eENVplus» at the exhibiton
- ☐ Whenever you meet us
- ☐ Enjoy LusTRE: <http://linkeddata.ge.imati.cnr.it/>

Thanks for
your
attention!

Contact Persons:

CNR-IMATI

demartino@ge.imati.cnr.it

albertoni@ge.imati.cnr.it

podesta@ge.imati.cnr.it

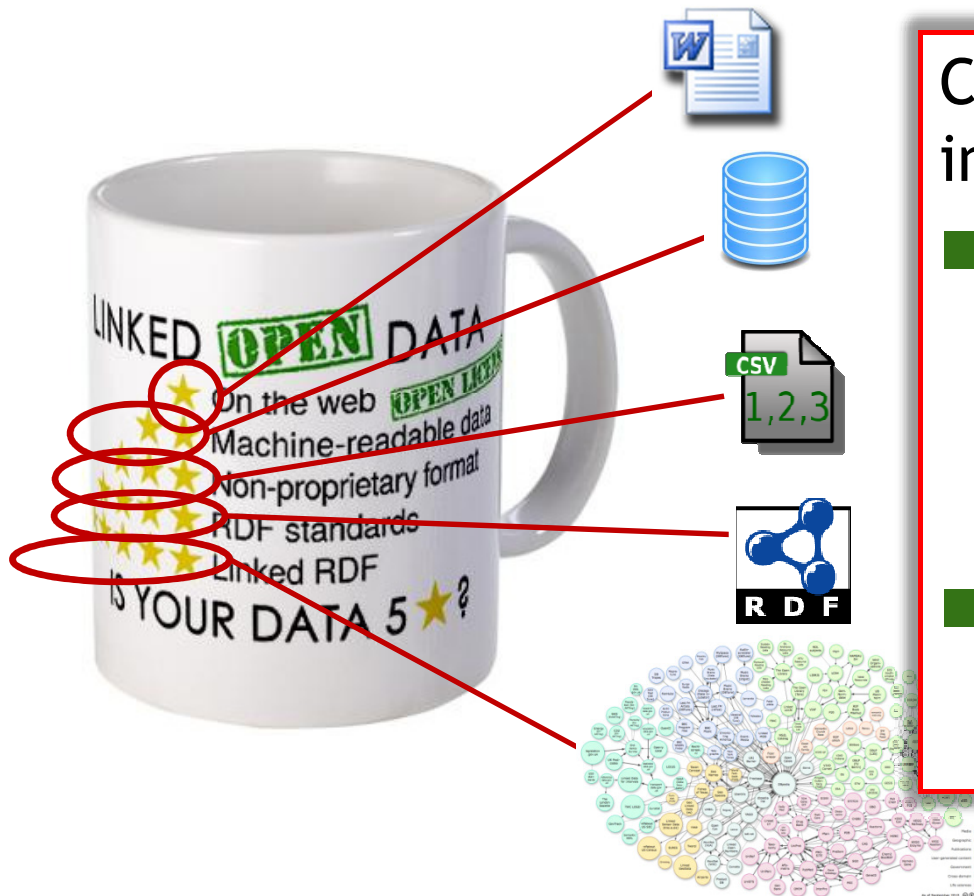
Disy

andreas.abecker@disy.net

roman.woessner@disy.net

APPENDIX

- Tim Berners-Lee 5 stars classification (LOD compliance)
(<http://www.w3.org/DesignIssues/LinkedData.html>)



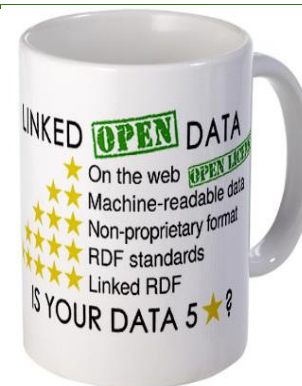
Critical issues in order to be included in TF :

- **Dereferenceable URI**

- ☐ are the basic prerequisite to have Linked Data, without them, it is not possible to check what a URI stands for, and so identifiers are not truly reusable.

- **Open licence with right to publish derivative works**

- 5 Stars classification of LD by Tim Berners-Lee
- HTTP dereferenceability of the URI mandatory LD prerequisite



1 star	resources available on the web (whatever format)
2 stars	resources available as machine-readable structured data (e.g., Excel)
3 stars	as 2 stars plus non-proprietary format (e.g., CSV instead of Excel)
3,5 stars	resources available as RDF dump without dereferenceable HTTP URI
3,9 stars	resources provided as RDFa (RDF embedded in XHTML) or SPARQL end point which are very close to be LD ready but without dereferenceable HTTP URI
4 stars	all the above plus, use open standards from W3C (RDF and SPARQL) and HTTP dereferenceable URI to identify things, so that people can point at published resources
5 stars	all the above, plus interlinks to other data to provide context

Framework Creative Common

Licence (acronym)	Characteristics	Licence reusability evaluation
Public Domain (CC0)	All the rights have been waived	5
Attribution (CC-BY)	Attribution is required	4.5
Share alike (CC-SA)	Copyleft licence	4
With restrictions (CC-NC , CC-ND, CC-NC-ND)	More severe restrictions	3.5
Closed (CR)	Closed licence	3
In progress (Pr)	Licence is going to be defined soon	2
Not found (NF)	No licence has been found in the website	1

STEP #1 PREPARE STAKEHOLDERS:

Prepare stakeholders by explaining the process of creating and maintaining [Linked Open Data](#).

STEP #2 SELECT A DATASET:

Select a dataset that provides benefit to others for reuse.

STEP #3 MODEL THE DATA:

[Modeling Linked Data](#) involves representing data objects and how they are related in an application-independent way.

STEP #4 SPECIFY AN APPROPRIATE LICENSE:

Specify an appropriate open data license. Data reuse is more likely to occur when there is a clear statement about the origin, ownership and terms related to the use of the published data.

STEP #5 GOOD URIs FOR LINKED DATA:

The core of Linked Data is a well-considered URI naming strategy and implementation plan, based on [HTTP URIs](#). Consideration for naming objects, multilingual support, data change over time and persistence strategy are the building blocks for useful Linked Data.

STEP #6 USE STANDARD VOCABULARIES:

Describe objects with previously defined [vocabularies](#) whenever possible. Extend standard vocabularies where necessary, and create vocabularies (only when required) that follow best practices whenever possible.

STEP #7 CONVERT DATA:

Convert data to a Linked Data representation. This is typically done by script or other automated processes.

STEP #8 PROVIDE MACHINE ACCESS TO DATA:

Provide various ways for search engines and other automated processes to access data using standard Web mechanisms.

STEP #9 ANNOUNCE NEW DATA SETS:

Remember to announce new data sets on an authoritative domain. Importantly, remember that as a Linked Open Data publisher, an implicit social contract is in effect.

STEP #10 RECOGNIZE THE SOCIAL CONTRACT:

Recognize your responsibility in maintaining data once it is published. Ensure that the dataset(s) remain available where your organization says it will be and is maintained over time

