Philipp Cimiano

Data and Metadata of Language
Resources as Linked Data on the Web

2nd LIDER Roadmapping Workshop

- **Give me all RESTful services that can add POS annotations in some tagset to my dataset.**

- **Give me services that can disambiguate entities with respect my ontology.**

- **Give me all terminological resources that provide terms and verified translations for the domain of tourism.**

- **Give me all services that can extract terms from text in Latvian.**

- **Give me all German POS-tagged corpora with Creative Common licenses.**

Currently not possible given the state-of-the-art.

Should it be possible?

Where are the use cases?

There is an API that makes all this queries possible!

There is an API that makes all this queries possible!

PARQL

Give me all RESTful services that can add POS annotations in some tagset to my dataset.

SELECT ?url WHERE {

    ?service myvocab:produces ?annotation;

    ?annotation rdf:type myvocab:POS;

    ?service myvocab:implementedby ?url;

    ?url rdf:type myvocab:RESTfulService.

}

Give me services that can disambiguate NER references with respect to my ontology.

SELECT ?url WHERE {

   ?service rdf:type myvocab:NamedEntityLinkingService;

   ?service myvocab:referenceOntology <myOnto>

   ?service myvocab:implementedby ?url;

   ?url rdf:type myvocab:RESTfulService.

}

Give me all terminological resources that provide terms and verified translations for the domain of tourism.

SELECT ?res WHERE {

    ?res rdf:type myvocab:Terminology;

    ?res myvocab:verified true;

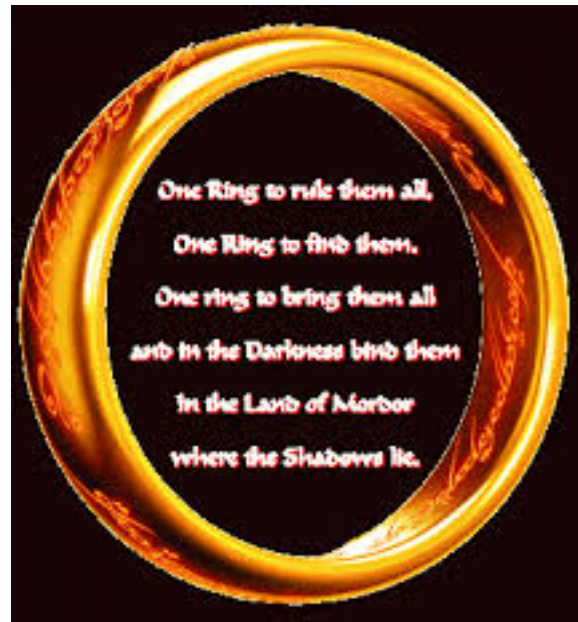   **?res dc:subject "tourism".**

}

Give me all services that can extract terms from text in Latvian.

```
SELECT ?url WHERE {
    ?service myvocab:produces ?annotation;
    ?annotation rdf:type myvocab:Term;
    ?service myvocab:implementedby ?url;
    ?service myvocab:supportedLanguage "lv".
}
```
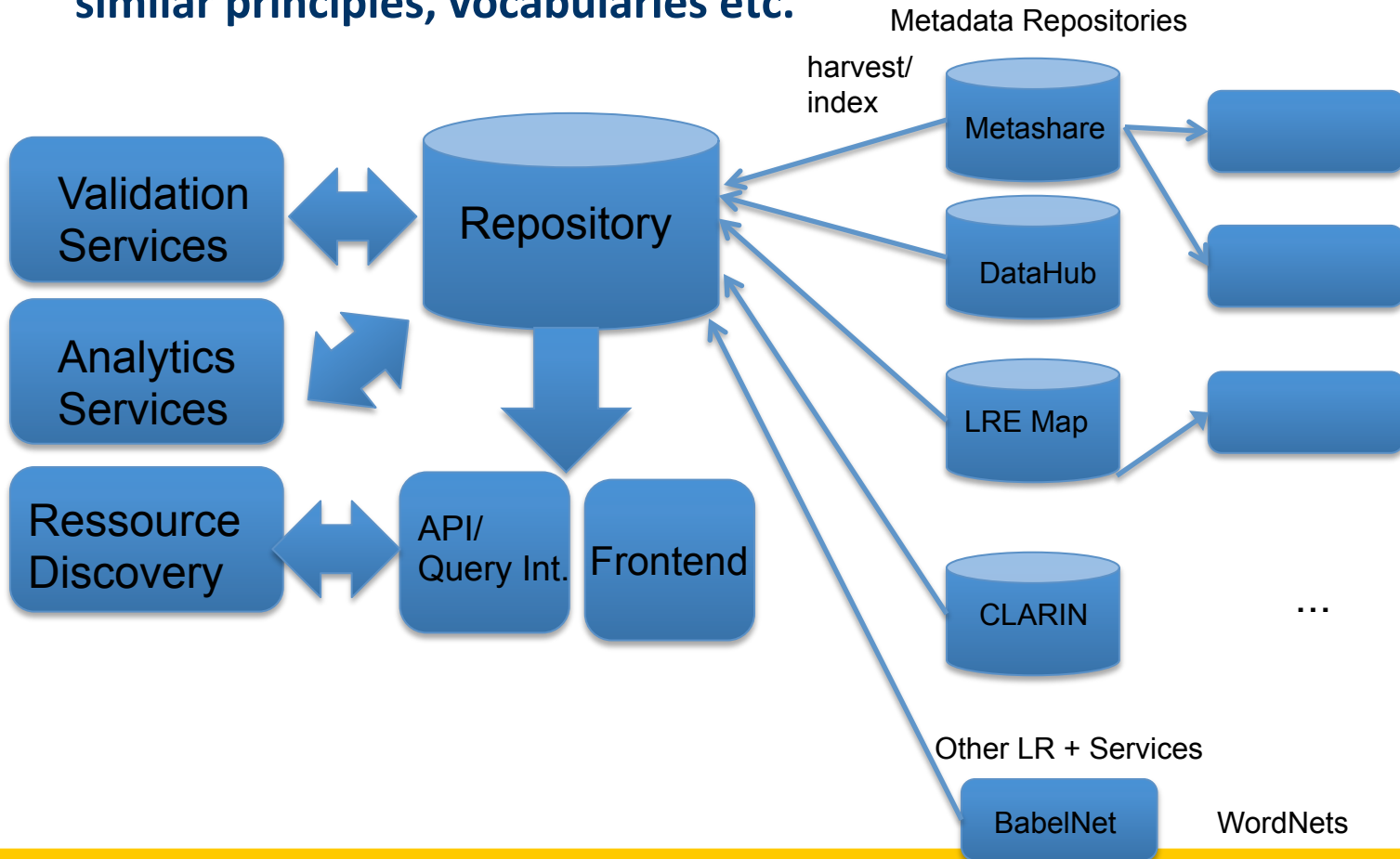
Three key steps!

1 CHOOSE YOUR SAUCE

2 SELECT YOUR MEAT OR VEG

3 COOK UNTIL PIPING HOT AND SERVE

- **Step1: One repository to query them all!**

- **Step1: One or several SPARQL repositories containing the metadata, using similar principles, vocabularies etc.**

Metadata Repositories

harvest/ index

Validation Services

Analytics Services

Ressource Discovery

Repository

API/ Query Int.

Frontend

Metashare

DataHub

LRE Map

CLARIN

…

Other LR + Services

BabelNet

WordNets

- **Step2: Interoperability of Metadata**

- **Requires: some agreement on metadata both about resources and services**
- **Contribution of LIDER: i) foster reuse of W3C standards in metadata provision, ii) provide mappings between properties / data elements used in different metadatamodels**
- **Provide W3C compliant data models for metadata specification**

- **Step3: Link to actual data, data in RDF format, exploiting standard vocabularies**
- **Corollary: i) Slicing and repurposing, ii) merging etc.**

1. Let's apply standard web formats for linguistic linked data publishing!

2. Let's harmonize metadata vocabularies, accepting diversity of datamodels!

3. Let's reuse standardized vocabularies!

4. Let's exploit possibilities of SPARQL for discovery, exploitation, filtering etc. of language resources!