

How to keep distributed efforts for developing semantic vocabularies together?

Piek Vossen, VU University Amsterdam

Flarenet Forum 2010, 11-12 February, 2010, Barcelona

Overview

- Semantic vocabularies are not only developed in the linguistic community → e.g. Open data project
- Provide semantic interoperability across external semantic vocabularies, general language resources, domains and across languages
- Kyoto project: community platform for developing domain resources, applied to the environment domain, that anchor these resources to generic lexicons/ontologies and background vocabularies and databases:
 - Define mapping relations and inferencing
 - Provide tools for automatic alignment

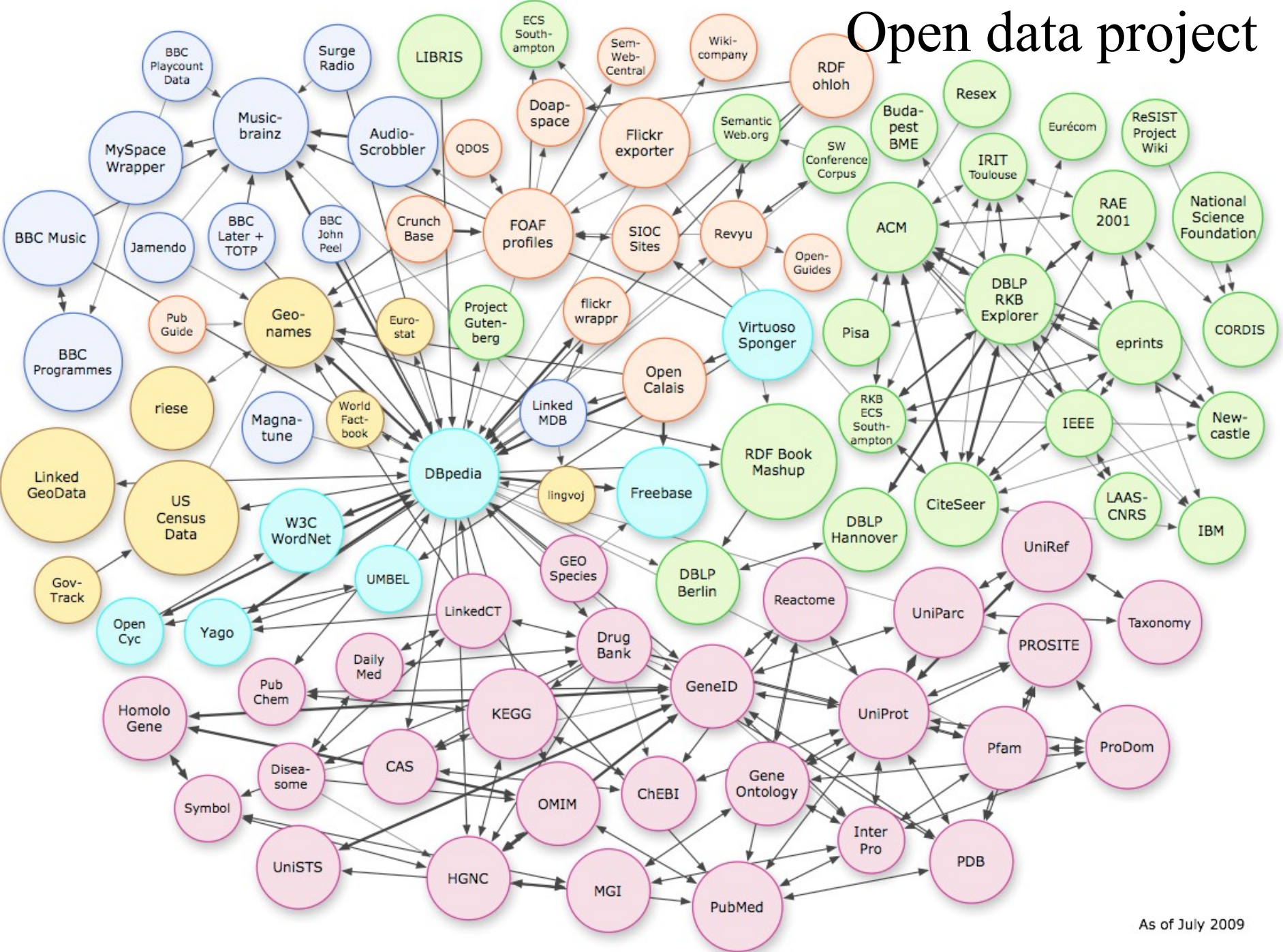


Provide a platform for community based extension

Flarenet Forum, Feb-2010, Barcelona



Open data project



Conceptual versus factual

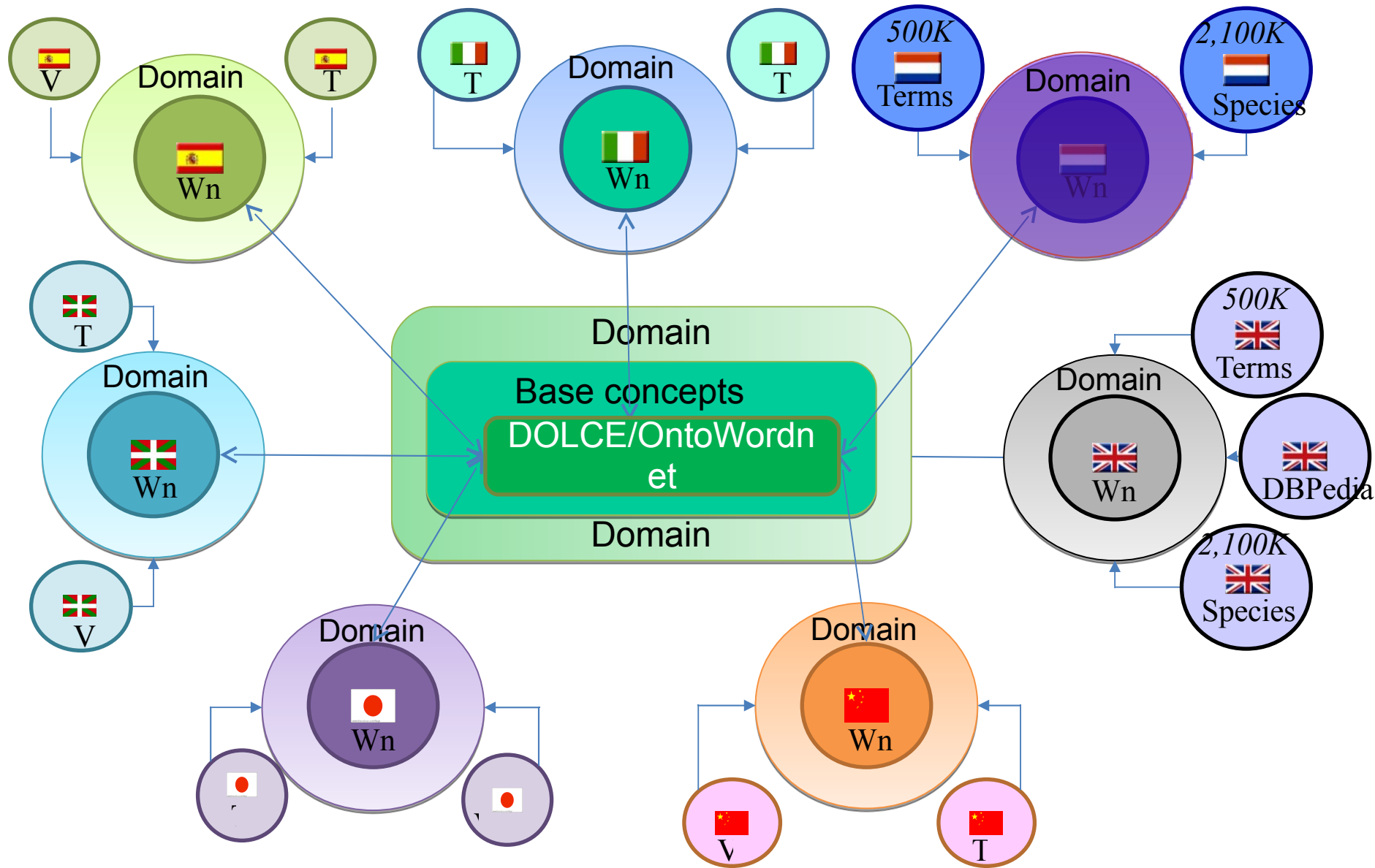
- Conceptual:
 - W3C Wordnet, DBPedia, Open CYC, Gene ontology....
- Factual:
 - DBPedia, Flickr, BBC music,....

Repositories for Kyoto project on the environment domain

- **Term database:** 500,000 terms per 1,000 documents per language
- Open data project:
 - **DBPedia:** 2.6 million things, including at least 213,000 persons, 328,000 places, 57,000 music albums, 36,000 films, 20,000 companies.
The knowledge base consists of 274 million pieces of information (RDF triples).
 - **GeoNames:** 8 million geographical names and consists of 6.5 million unique features whereof 2.2 million populated places and 1.8 million alternate names.
- **Domain thesauri and taxonomies:** Species 2000: 2,1 million species
- **Wordnets** for 7 languages: about 50,000 to 120,000 synsets per language
- **Ontologies:** SUMO, DOLCE, SIMPLE



Kyoto Knowledge Base



Combining the work of many

- Keep knowledge distributed over different layers:
 - External data:
 - Open data project
 - Domain vocabularies
 - Acquired domain terms
 - Generic resources: wordnets and ontologies
- Define the relations between the layers
- Provide tools for alignment for relating the layers

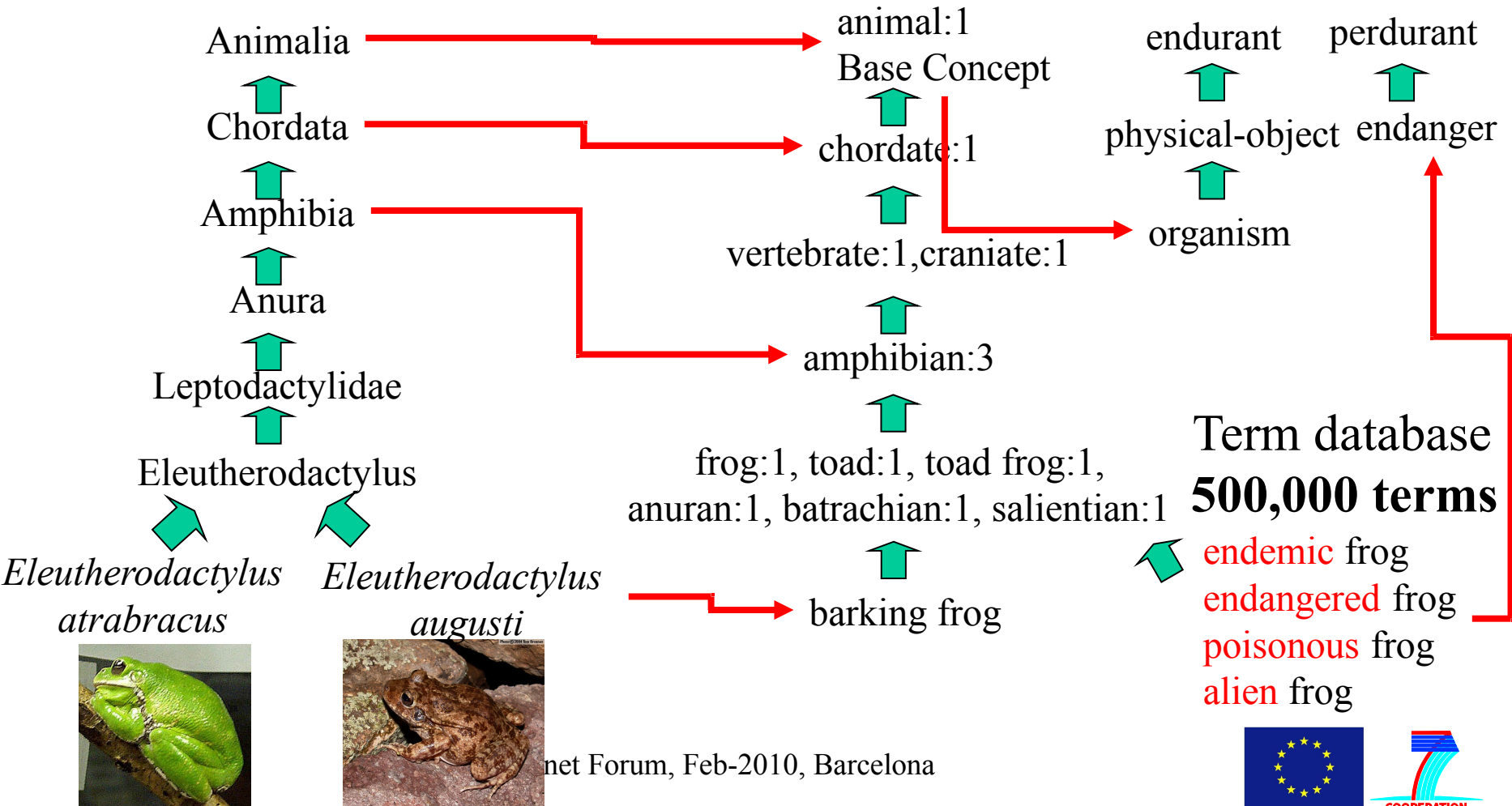


Division of labor in knowledge sources

Skos database
2.1 million species

Wordnet-LMF
100,000 synsets

Ontology-OWL-DL
2,000 types



net Forum, Feb-2010, Barcelona





Keyword Search



Semantic Wiki



Wikiplanet



Linguistic Processor

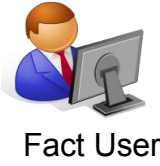


Multilingual Knowledge Base

Wordnets
Ontologies
interlinked



2



Kybot

Fact Extractor



Semantic & Syntactic Base
Kyoto Annotation Format (KAF)

Wikyoto



Wiki Term Editor



Fact Base



3



Term Extractor

Tybot



Term Base



Semantic Search



SEPLN09, September, 7-10th, 2009, San Sebastian

