# AGROVOC and the OWL Web Ontology Language: the Agriculture Ontology Service - Concept Server OWL model

NKOS workshop

Alicante, 2006

# Outline

- Background
- Needs and purposes
- Our approach
- Performance tests
- Current status and Next steps
- Open issues
- Conclusion

# Background (1/2)

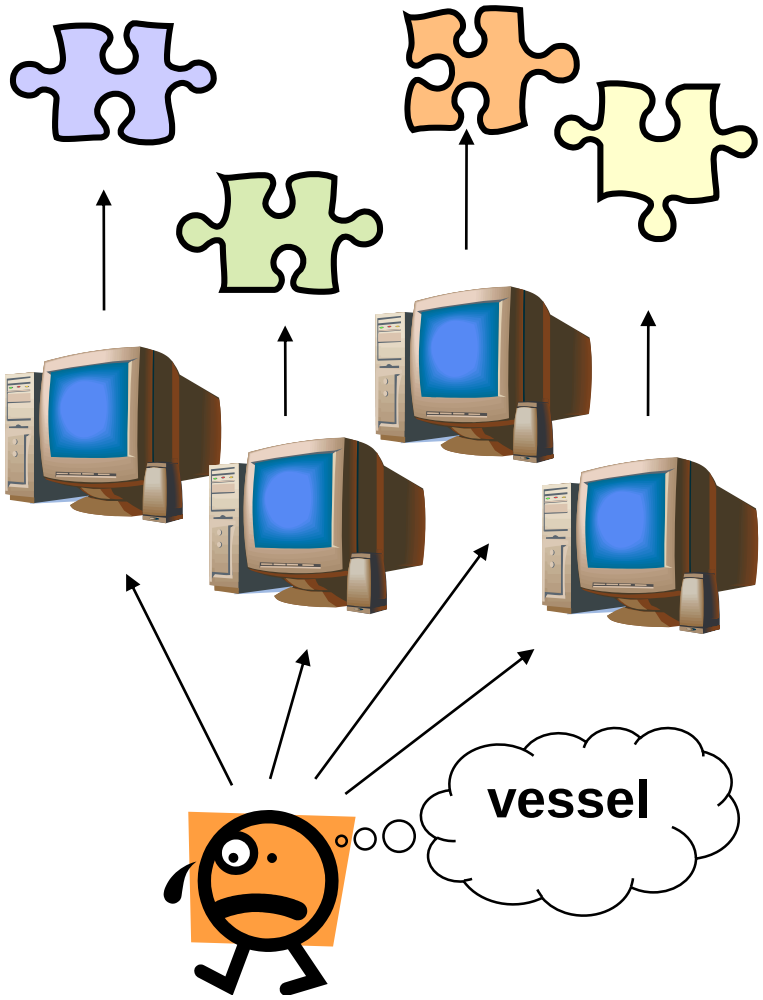
- AGROVOC
  - Used worldwide
  - Multilingual
  - Term-based
  - Limited semantics
  - Maintained as a relational database
  - Distributed in several formats (RDBMS, TagText, ISO2709, ...)

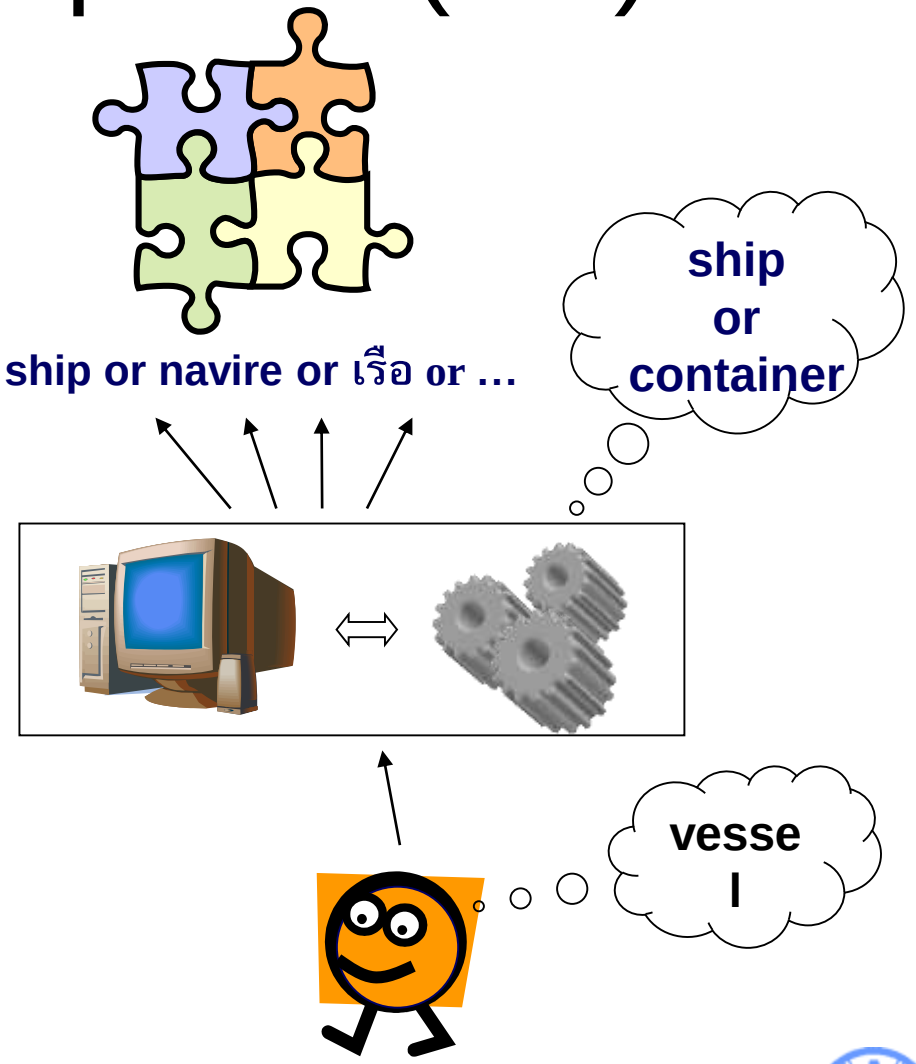
# Background (2/2)

- Draft versions available in TBX, SKOS, OWL
- Access to full thesaurus through Web Services


- Agricultural Ontology Service (AOS)

# Needs and purposes (1/3)



# Needs and purposes (2/3): better serving web applications

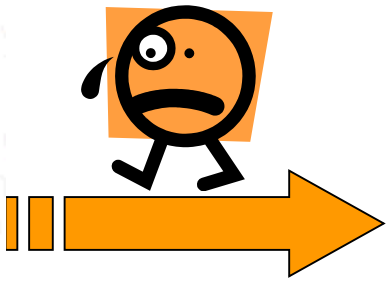
- Semantic navigation of knowledge
- Semantic navigation of resources (bibliographical metadata, etc.)
- Intelligent query expansion
- Terminology brokering
- Improved natural language processing
  - Language recognition
  - Improved parsing (combinatorial)
  - Extended concept resolution
- Inferencing / Reasoning
- Machine learning
- Clustering and ranking

# Needs and purposes (3/3):

- Having a complete structure
  - From which to export any other traditional or different representation in any format
    - Word list, thesaurus, sub-domain ontologies, ...
    - TBX, SKOS, OWL, ...
- Having more
  - More than a thesaurus
  - SKOS: impossible to state information on terms
  - TBX:  XML based

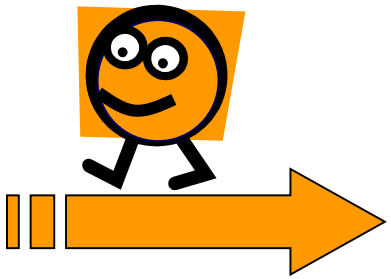
# Our approach (1/12)

Natural resources (5091)
Nonrenewable resources (24056)
- Land resources (4179)
  - Abandoned land (6)
  - Arable land (18948)
  - Common lands (1782)
  - Farmland (2808)
    - Cultivated land (16212)



Better defined structure

Better defined structure



RDBMS

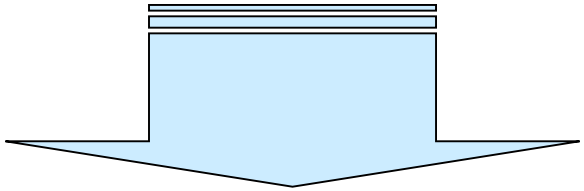
Natural resources (5091) **AGROVOC**
Nonrenewable resources (24056)
Land resources (4179)
- Abandoned land (6)
- Arable land (18948)
- Common lands (1782)
- Farmland (2808)

**XML formats (e.g. TBX)**
**RDFS formats (e.g. SKOS)**

# Our approach (2/12)

- Concept-based
- More semantics
- “Language-independent”
- Easy integration with other KOS
- Easy sharing within the Web

**Better defined structure: the CS**



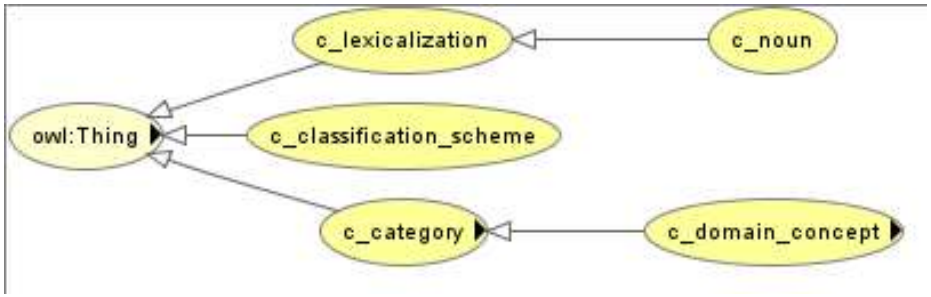
ontology + OWL

# Our approach: The OWL model (3/12)

- Why OWL?
  - Built on top of RDF, increased interest, future support
  - W3C recommendation
  - Represented as triples
  - Interoperable and web-enabled (linking multiple ontologies)
  - Reuse of existing tools, no proprietary RDBMS
  - Reasoning is possible: to arrive at conclusions beyond what is asserted + consistency checks
  - Revision was needed → better semantic and refinement
- Problems
  - Backward compatibility with legacy systems
  - Many desirable kinds of information must be represented tortuously or cannot be represented at all

# Our approach: The OWL model (4/12)

- Concept / Term / term variants



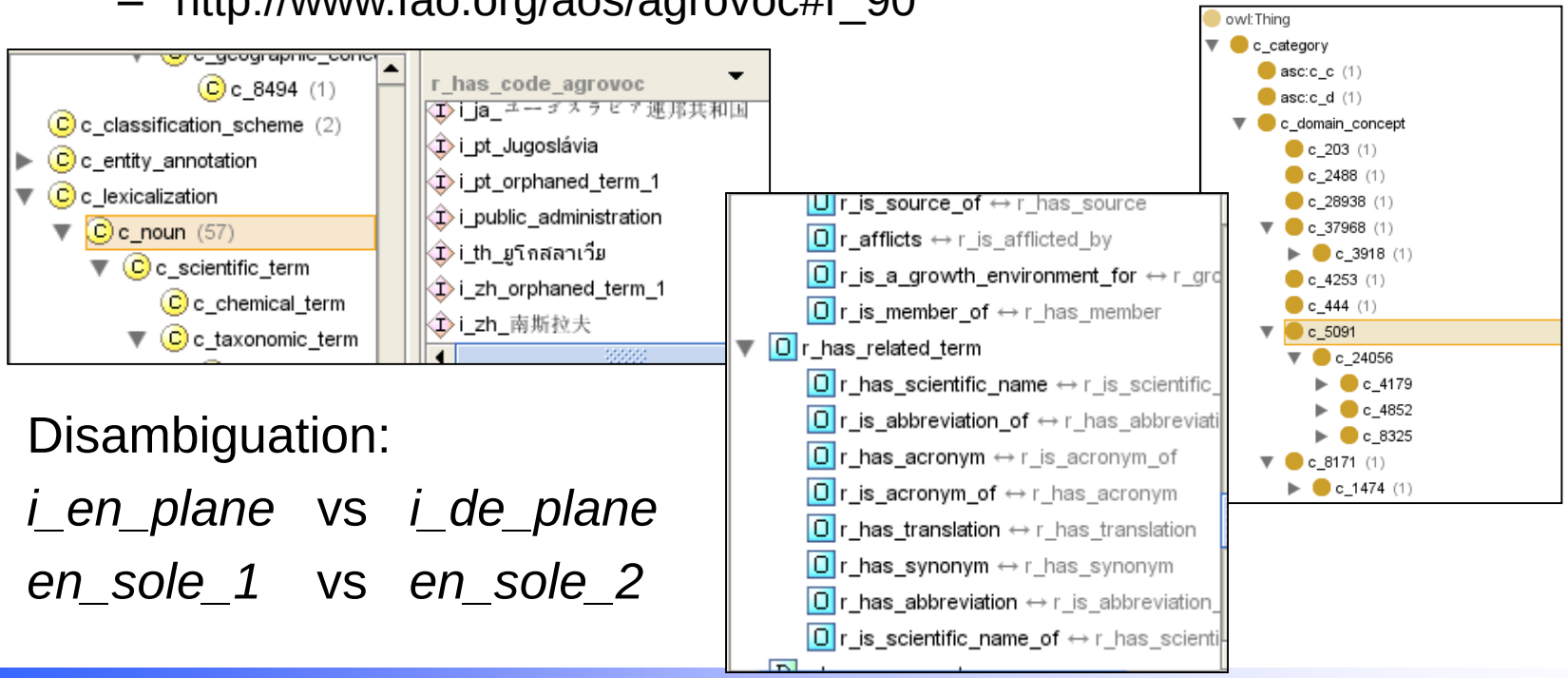

- Language issue
  - 'has_lexicalization'/ 'lexicalized_with' functional
- AOS/CS base URI: http://www.fao.org/aos/agrovoc

# Our approach: The OWL model (5/12)

- Concepts are classes AND instances
  - Classes → to support hierarchy and inheritance
  - Instances → to keep OWL DL
- Terms are instances of a specific class

# Our approach: The OWL model (6/12)

- URI and class name: “c_”, “r_”, “i_”
  - http://www.fao.org/aos/agrovoc#c_28938
  - http://www.fao.org/aos/agrovoc#i_en_public_administration
  - http://www.fao.org/aos/agrovoc#r_90



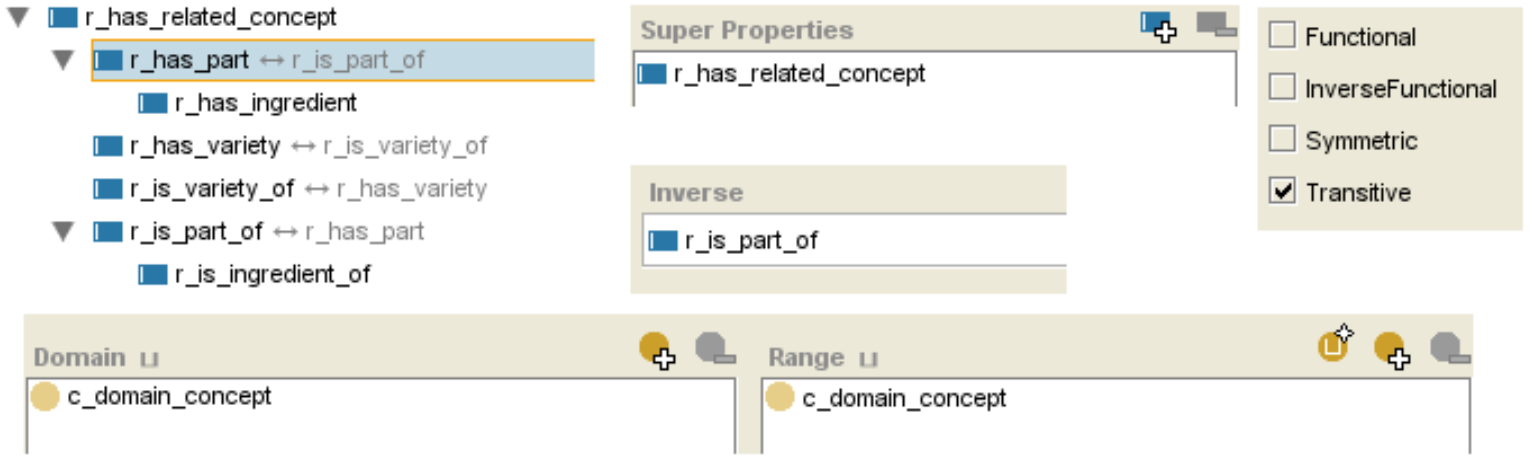

Disambiguation:

*i_en_plane* vs *i_de_plane*

*en_sole_1* vs *en_sole_2*

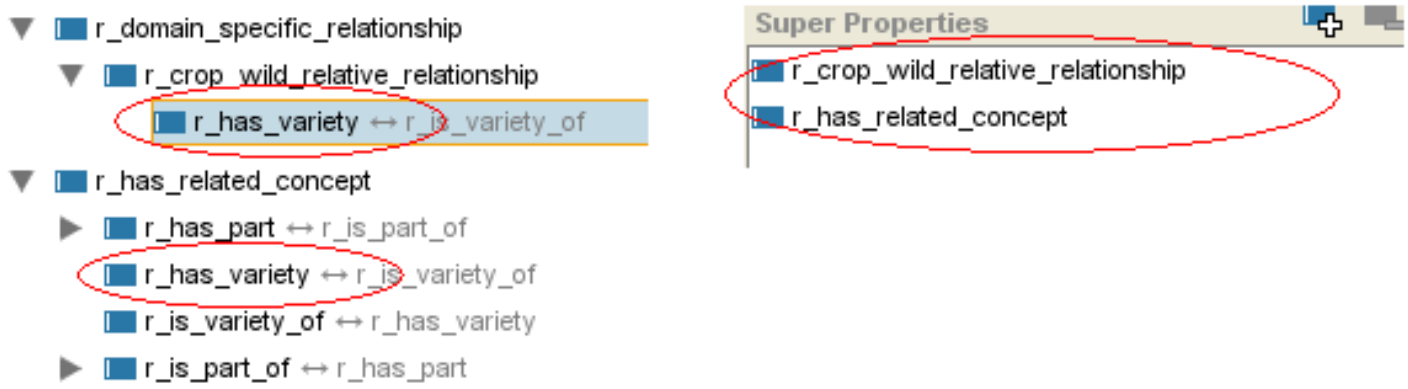
# Our approach: The OWL model (7/12)

- Concept-to-concept relationships
  - **to be refined**



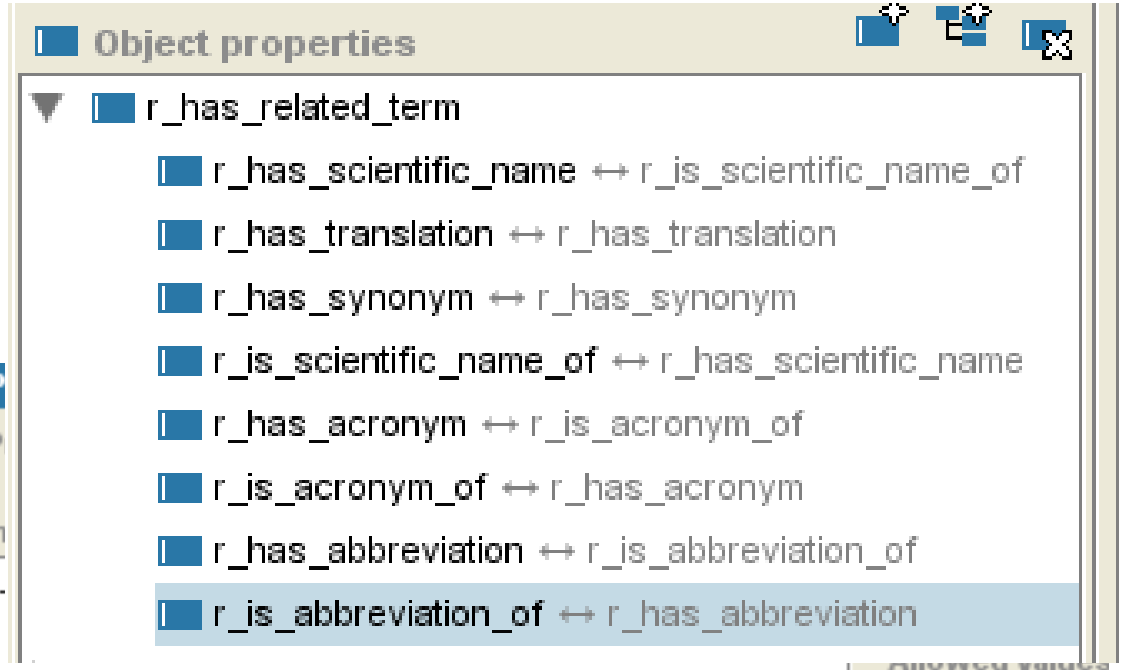
# Our approach: The OWL model (8/12)

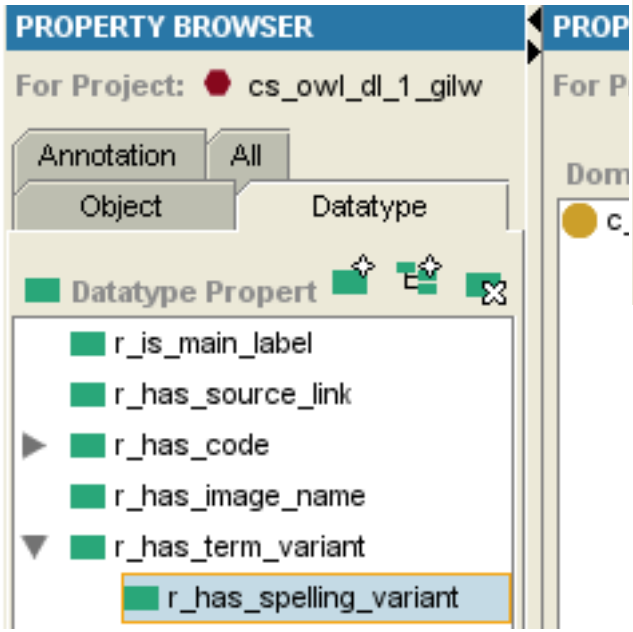
- Domain-specific relationships



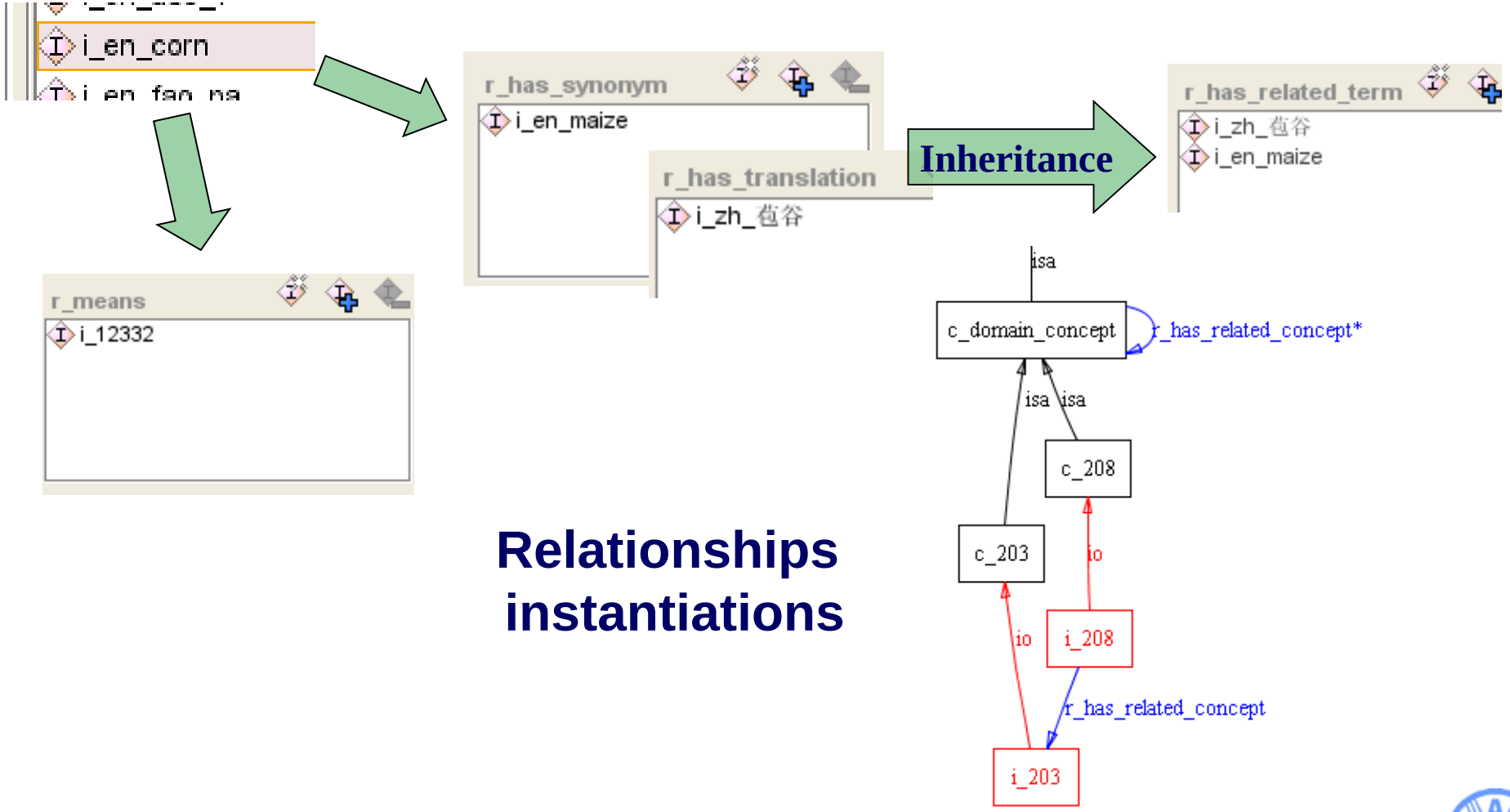
# Our approach: The OWL model (9/12)

Term-to-Term and Term-to-Variants Relationships



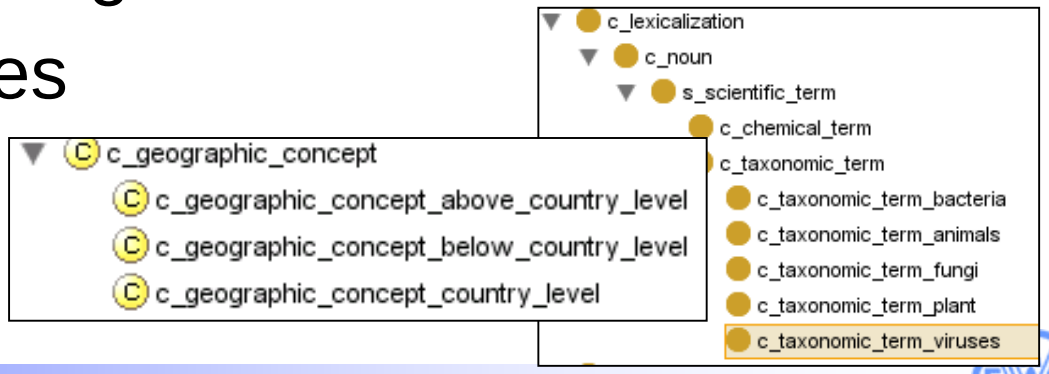
# Our approach: The OWL model (10/12)




**Relationships instantiations**

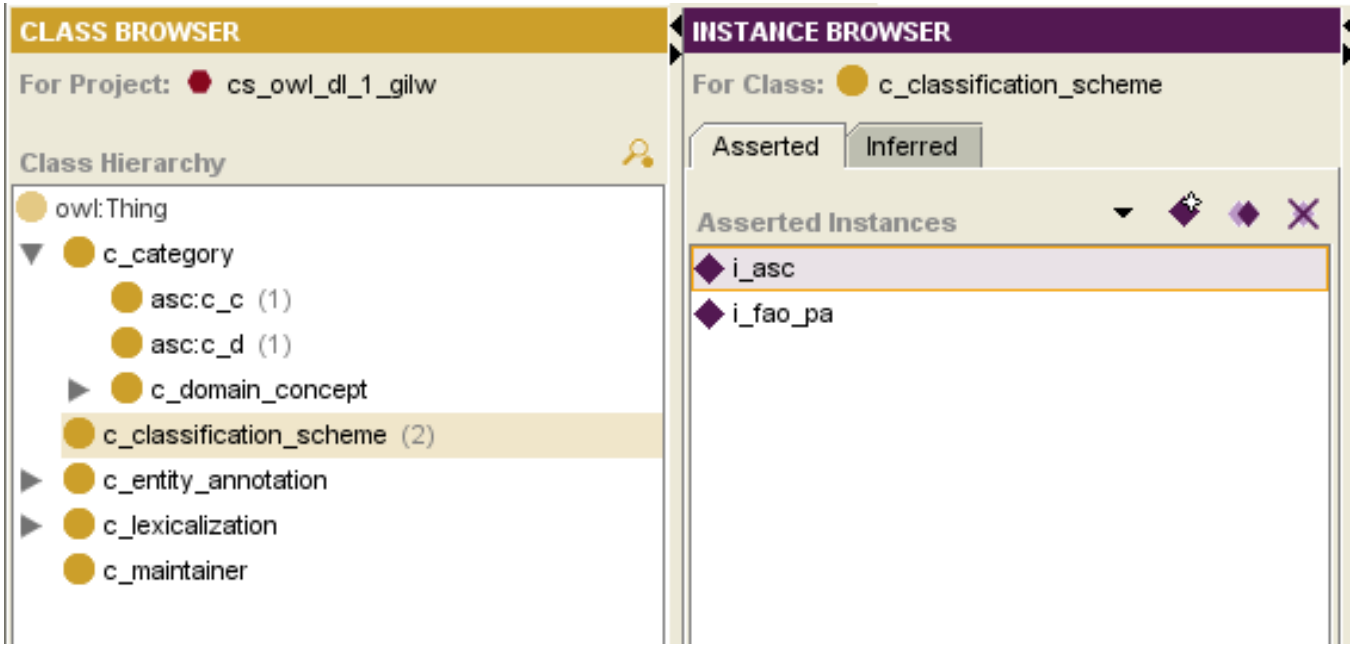
# Our approach: The OWL model (11/12)

- Other elements
  - Status for concepts and terms (suggested, approved, reviewed, deprecated)
  - *r_has_date_created*
  - *r_has_date_last_updated*
  - Scope notes / images / definitions
  - Sub-vocabularies



# Our approach: The OWL model (12/12)

- Classification schemes and categories



# Backward compatibility

- Backward compatibility with a traditional thesaurus
  - Main descriptor (*r_is_main_label*)
  - Term codes references
  - UF+
  - Scope notes
  - etc.

# Performance tests

- Sesame / Jena
- PostgreSQL / MySQL / Native db
- With Sesame:
  - Loading 9 MB ontology
    - Processed 273644 statements in 463 seconds
  - Querying
    - 21858 results found in 13030 ms

# Current status

- What exists concretely of the model:
  - Description of the model
  - Relationship definition (in collab. with CNR)
  - Test project
  - Full AGROVOC conversion procedure
  - Performance tests
  - AOS/CS Workbench construction

# Next steps

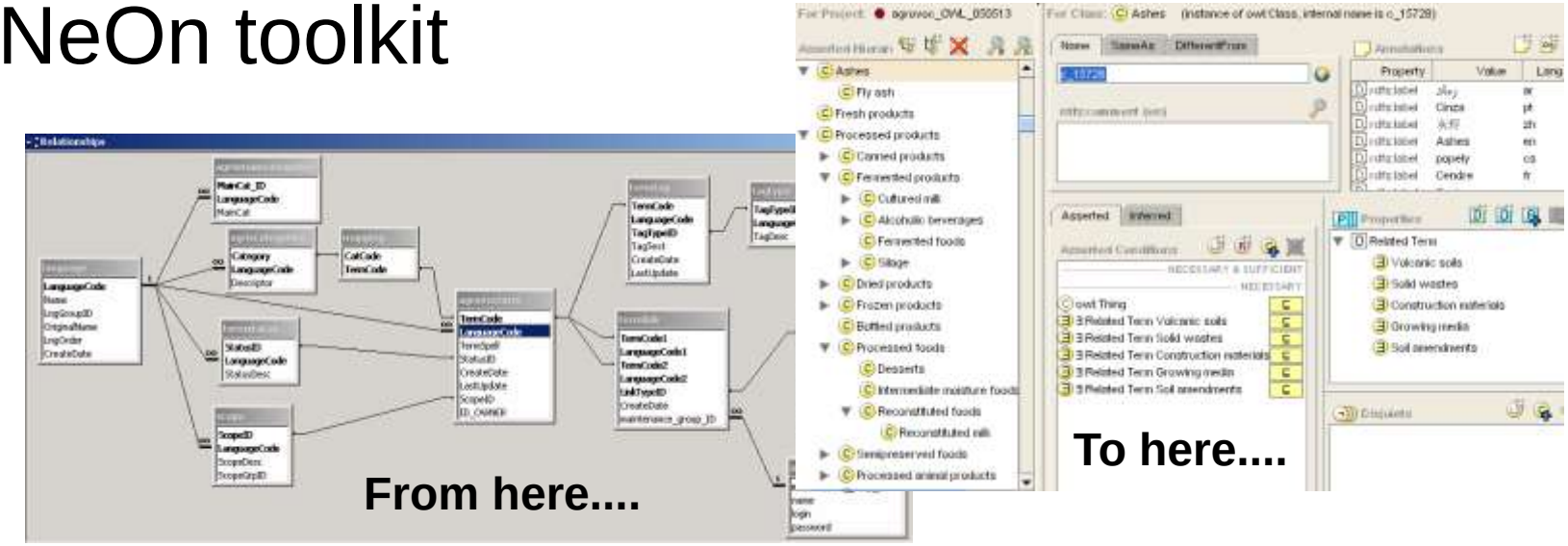
- AGROVOC refinement and conversion
- Build the AOS/CS Workbench
- Extensive tests
  - scalability at storage and operational level
  - performance at the maintenance and data retrieval level
  - integration of and linkage to datasets
- Create a network of ontology experts
  - Workshops/Trainings
- NeOn results

# Open issues

- Assign attributes to relationships
- Distinguish concepts instances from individuals
- Validity of relationships (or context)
- Ontology lifecycle, versioning
- Ontology mapping and merging
- No more words but URIs in IS
- Better exploitation of the potentiality at the application level: powerful IR
- Ontology Web services (OWS)

# Conclusion

- AOS is still a success story and is gaining terrain in private sector
- More ontologies in FAO
- NeOn toolkit



acliang@alum.mit.edu
boris.lauser@fao.org
margherita.sini@fao.org
johannes.keizer@fao.org

# Questions?

Thank you