Linked Data: Now What?

Aldo Gangemi
STLab, ISTC-CNR, Rome
aldo.gangemi@cnr.it

What was right

- “Just do it” bootstrapping for the Semantic Web
- Few simple recipes for reengineering, publishing, aligning, and consuming data
- Refreshing attention to practical problems, technologies that scale, and to the webby side of the Semantic Web
- URI-based data integration proven feasible
- Concrete platform and use case for a viral effect in opening data (cf. biology, gov data)

Issues

1. Sparseness of data (unless controlled by a schema)
2. Problems of evolution and versioning (both on schema and data sides)
3. Licensing and policies partly unclear
4. Current topic coverage is scattered in extension and depth
5. Difficulty of exploring data just to know what they are about (exceptions, e.g. RelFinder)
6. Lack of good interaction with linked data
7. What recipes for data created from inference, enrichment, lenses, customized consumption?
8. Semantics of many datasets is suspicious
9. Much knowledge is in literals rather than entities

What should(n't) we do

- Many problems (e.g. 1 through 4) will be solved as a natural evolution of the technology
- Some problems (e.g. 5 through 7) are common to all semantic technologies: are semantic data special from an interaction viewpoint?
- Some problems (e.g. 8/9) depend on limited attention to design aspects
- Let's stand up to the SW vision and to the interdisciplinarity of Web Science
- Please don't reinvent the wheel

Some design directions

- What domain semantics is piped into reengineered linked datasets?
  - are *bridging* approaches (e.g. D2R, Virtuoso Sponger cartridges) sustainable in presence of legacy data or specific requirements?
  - e.g. Freebase Gridworks allows some customization when reengineering DBs
  - e.g. OPPL tool from University of Manchester allows pattern-based refactoring
  - e.g. Semion tool from STLab keeps track of, and enables custom semantic transformations when reengineering
- At the carrefour between informal and formal semantics
  - extensive usage of *metamodels*: SKOS, Lexical, DB, etc.
  - possibilities from OWL2 *punning* mechanism
  - heterogeneous ontology matching techniques can help with semantic conflicts and with the literal vs. entity issue
  - check next workshops: KIELD@EKAW2010 and WOP@ISWC2010
  - check Ontology Alignment Evaluation Initiative on *automatic data interlinking*
- LOD meets Ontology Design
  - good practices and design patterns (just started, but more communication is needed)
  - OWL LOD datasets with task-oriented ontologies
  - e.g. Semantic Scout application (paper@EKAW2010)