Semantic technologies for the Production and Publication of Open Data in ACI - Automobile Club d’Italia

Lorenzo Lepore
(on behalf of many al.)

ACI Informatica
Sapienza - Università di Roma
Okkam

October 27, 2019
Introduction to the Project

- Joint project by:
  - Automobile Club d'Italia (ACI)
  - Sapienza - Università di Roma
  - Okkam S.r.l. (spinoff of the University of Trento)

- Objectives:
  - Definition of an ontology describing ACI’s Public Vehicle Register (PRA) and car tax domains
  - Development of an OBDM system to access ACI’s data through such ontology
  - Creation of a web portal for the publication of (a fragment of) ACI’s data in Linked Open format
Administrations from both European and Italian governing bodies have recently pushed public administrations to store and publish data about their sectors so that such data could be easily understood, redistributed and reused.

In such context, standard languages and paradigms from the Semantic Web area are highly recommendable (LOD, RDF, OWL, …)
Project Goals

ACI (*Automobile Club d’Italia*) was committed to comply with such recommendations by:

- defining a formal semantic representation of its principal information assets
- adopting data management techniques allowing linking entities of such formal representation to existing relational data collected by ACI applications
- improving both quality and interoperability of data
Project Challenges

- Choice of:
  - knowledge representation language
  - data to be published (large portion of data not disclosable due to privacy issues)
  - paradigms and technologies for Linked Open Data preparation

- Link published data to that of other administrations

- Harmonization of new paradigms with pre-existing methods and processes in ACI
Achieving the Goals

- OWL 2 Ontology of Public Vehicle Register and Car Tax domains
- Mapping between ontology elements and pre-existing application data stored in relational sources
- Entity identification and linking process to establish matching between ACI’s data and that of other related administrations (ISPRA¹, ISTAT², . . .)
- Use of standard shared vocabularies for modelling of statistical multi-dimensional data (Data Cube ontology³) and geographical data (GeoNames⁴)

¹ [http://dati.isprambiente.it/](http://dati.isprambiente.it/)
² [https://www.istat.it/](https://www.istat.it/)
³ [https://www.w3.org/TR/vocab-data-cube/](https://www.w3.org/TR/vocab-data-cube/)
⁴ [http://www.geonames.org/ontology/documentation.html](http://www.geonames.org/ontology/documentation.html)
Data Publication Paradigm

- OWL 2 Ontology + Mapping → OBDM specification managed by the Mastro system:
  - RDF datasets extracted from the sources by means of SPARQL queries over the ontology
  - Knowledge of the ontology used to semantically annotate the data
- Entity reconciliation obtained by adopting the Okkam’s *Entity Names System* capable of:
  - identifying the same resource among different data flows
  - reconciling different references to the same resource by assigning a global identifier

Lorenzo Lepore (on behalf of many *al.*)

ACI InformaticaSapienza - Università di RomaOkkam

ACI Linked Open Data (ISWC 19)
Project Outcome

The ACI Linked Open Data Portal\(^5\)

- Ontology inspection (both in graphical and OWL axioms forms)
- Ontology documentation (*wiki* format enriched with hypertext links for easy navigation)
- RDF/CSV datasets access and manipulation:
  - each dataset provided with a description, a distribution license and downloadable files
- Knowledge graph navigation by dereferencing resources to unique URIs corresponding to web pages containing resource description plus related RDF triples

\(^5\)http://lod.aci.it/
THANK YOU!