

Publishing Official Classifications in Linked Open Data

Giorgia Lodi**, Antonio Maccioni**, Monica Scannapieco*,
Mauro Scanu*, Laura Tosco*

* Istituto Nazionale di Statistica – Istat

** Agenzia per l'Italia Digitale – AgID

Motivations

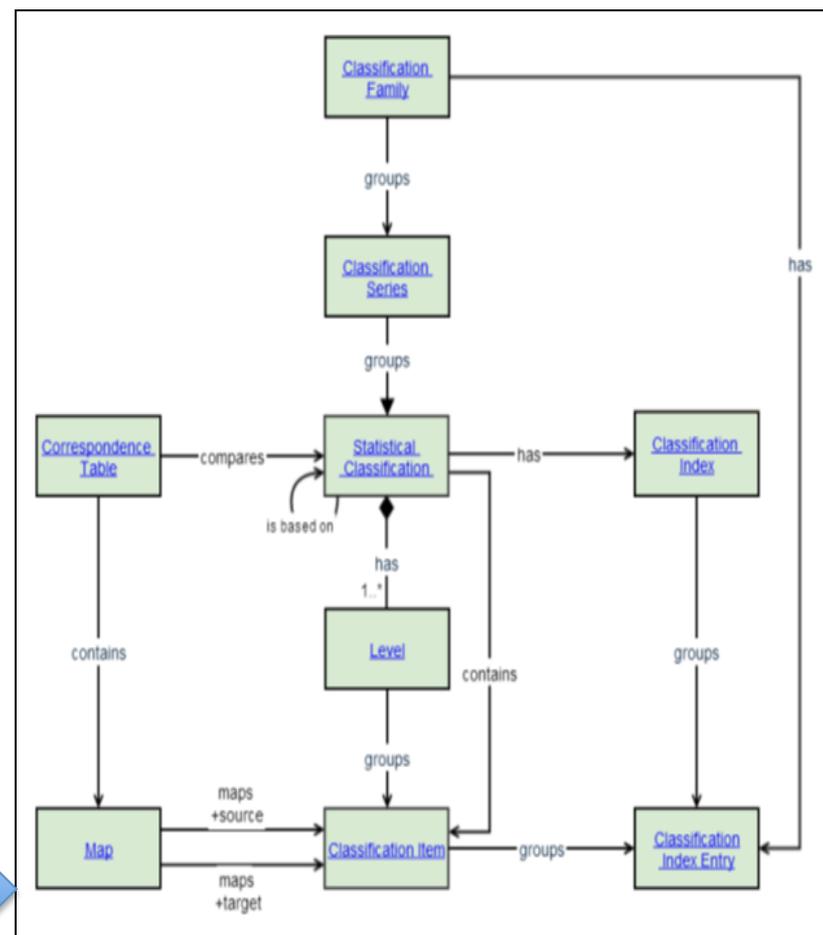
- End of 2012 the Agency for Digital Italy (AgID) published the first version of **National Guidelines on the use of LOD** as the paradigm for enabling semantic interoperability between Public Administrations
- Among the key datasets, AgID identified **official classifications** as cross-domain data to be published in LOD

Aims of the Project

- The Italian National Institute of Statistics - **Istat** & **AgID** launched a joint project:
 - ✓ To model Official Classifications (**Ateco 2007** -Classification of Economic Activity and **COFOG** - (Classification of the Functions of Government) in **LOD** using standard ontologies (SKOS, XKOS, ADMS,...)
 - ✓ To certify data by **provenance** using the **PROV framework**, i.e. document the (i) overall process of creation of a classification and (ii) the creation and publication of the LOD version
 - ✓ publication on a SPARQL endpoint hosted by AgID: <http://spcdata.digitpa.gov.it>

Statistical Official Classifications

- Statistical Official Classifications allow to:
 - ✓ Reuse them in different production processes
 - ✓ Promote harmonization
- Important models:
 - ✓ Neuchâtel Terminology Model Classification (2004)
 - ✓ Generic Statistical Information Model (GSIM) (2013)



GSIM: UML- schema

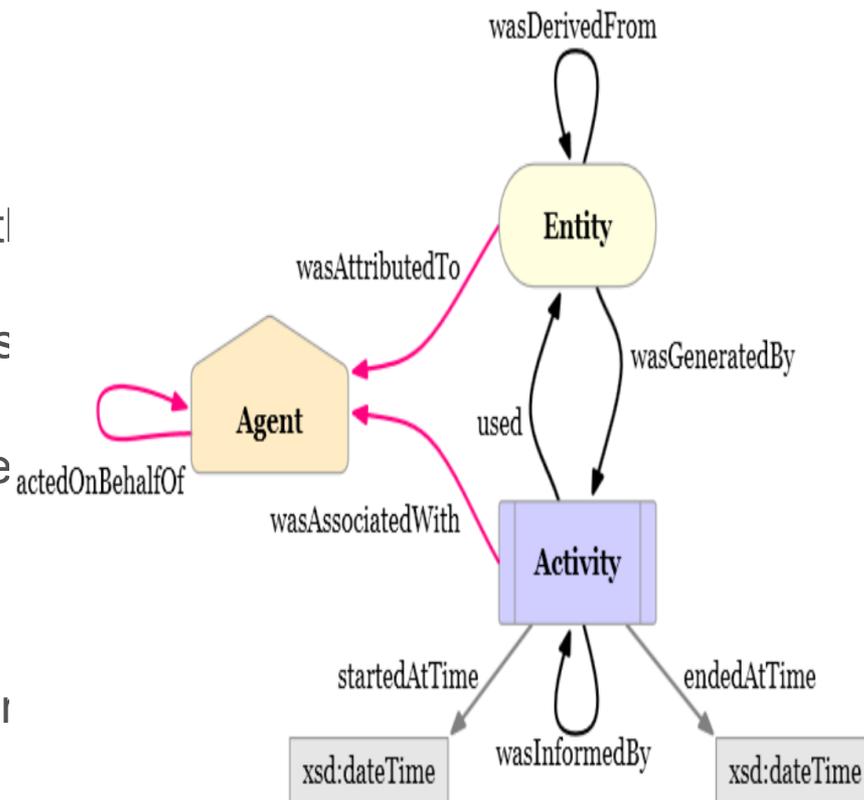
ATECO 2007

- Classificazione delle Attività Economiche (ATECO)
 - Version 2007
 - Italian counterpart of NACE classification (<http://epp.eurostat.ec.europa.eu/portal/page/portal/nacerev2/introduction>)
 - Consisting of the following levels:

ATECO	NACE
Sezioni	Sections
Divisioni	Divisions
Gruppi	Groups
Classi	Classes
Categorie	-
Sottocategorie	-

PROV Ontology

- W3C recommendation, OWL2 ontology
- Provides a set of classes, properties, and restrictions to represent **provenance information**
 - Publishing, together with data, who is **responsible of the data** - person/ institution/ administration that manages creates/ manipulates the data
 - **Certifying data** - data are certified if they are published by their responsible
 - **Provenance of data** in terms of information regarding entities and processes involved in the production and publication of data



XKOS Ontology

- XKOS - eXtended Knowledge Organization System
 - Data Documentation Initiative (DDI)
 - Extension of SKOS - Simple Knowledge Organization System
 - Manages statistical classifications
 - Implements requirements laid out in ISO standards (ISO 704:2000 revised 2009 and ISO 1087-1:2000)
 - Represent statistical classifications with their structure, textual properties, and relations between classifications

XKOS and SKOS

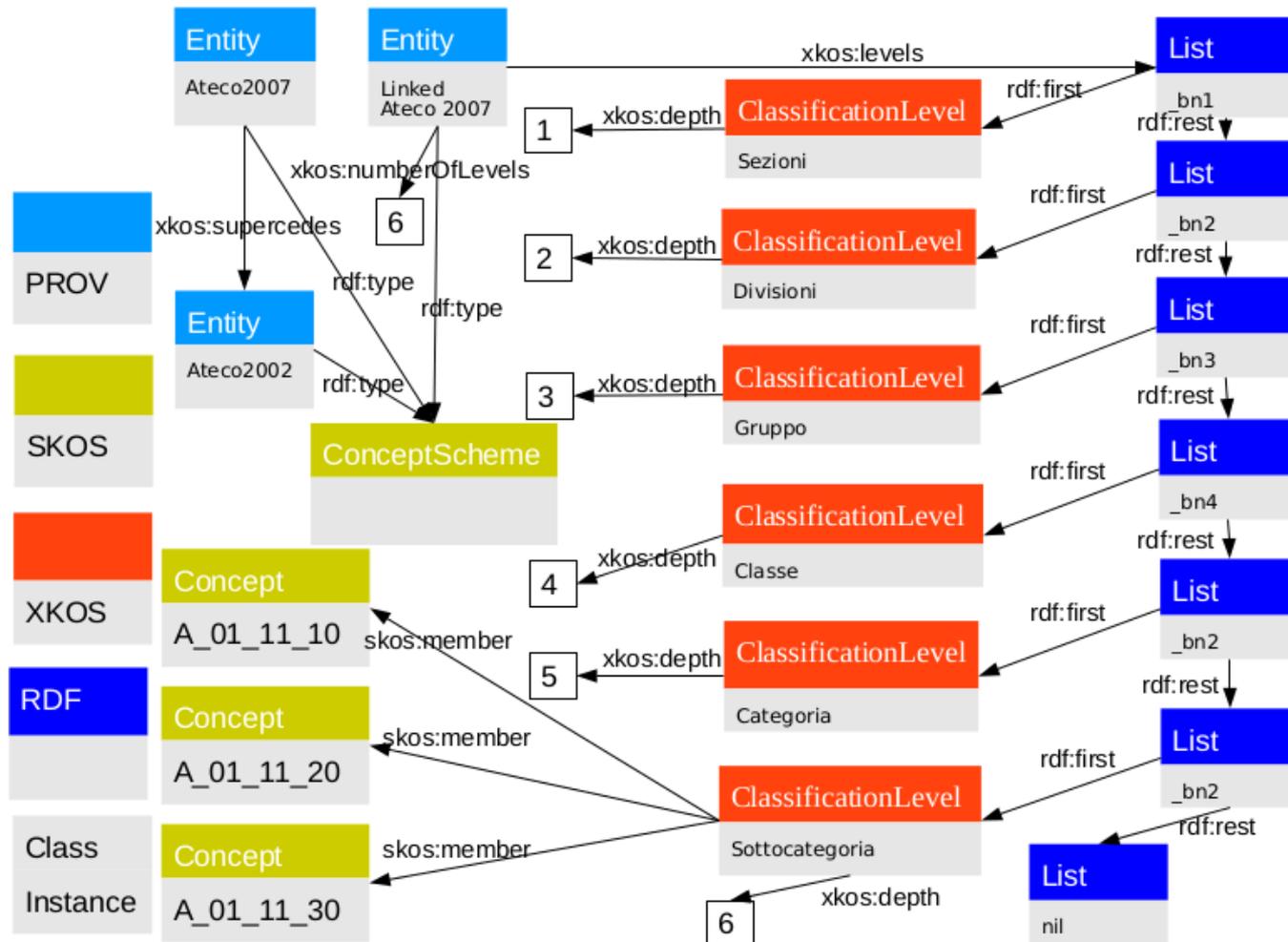
- SKOS vocabulary: loosely defined notions of ‘broader than’, ‘narrower than, and ‘related to’ relationships.
- Statistical classifications - XKOS:
 - Formally defined **hierarchical relations**, which are referred to as generic (generic-specific) and partitive (whole-part).
 - **Levels** that correspond to increasingly detailed views of the field covered
 - **Use of associations** that are more specific than the generic ‘related to’. Causal, sequential, and temporal relations defined.
- XKOS **uses** SKOS classes and related properties (codes, labels, etc.) to represent:
 - Classification items (*skos:Concept*)
 - Classification
 - Classification
 - *skos:inScheme*
 - *skos:member*
 - *skos:broader*
 - *skos:narrower*
 - *skos:Collection*
 - *skos:Concept*
 - ...
- XKOS **adds** properties to describe the links between these objects i.e.:
 - *xkos:belongsTo*
 - *xkos:follows*
 - *xkos:correspondence*
 - *xkos:madeOf*
 -

skos:classification scheme may be attached to a classification using xkos:belongsTo

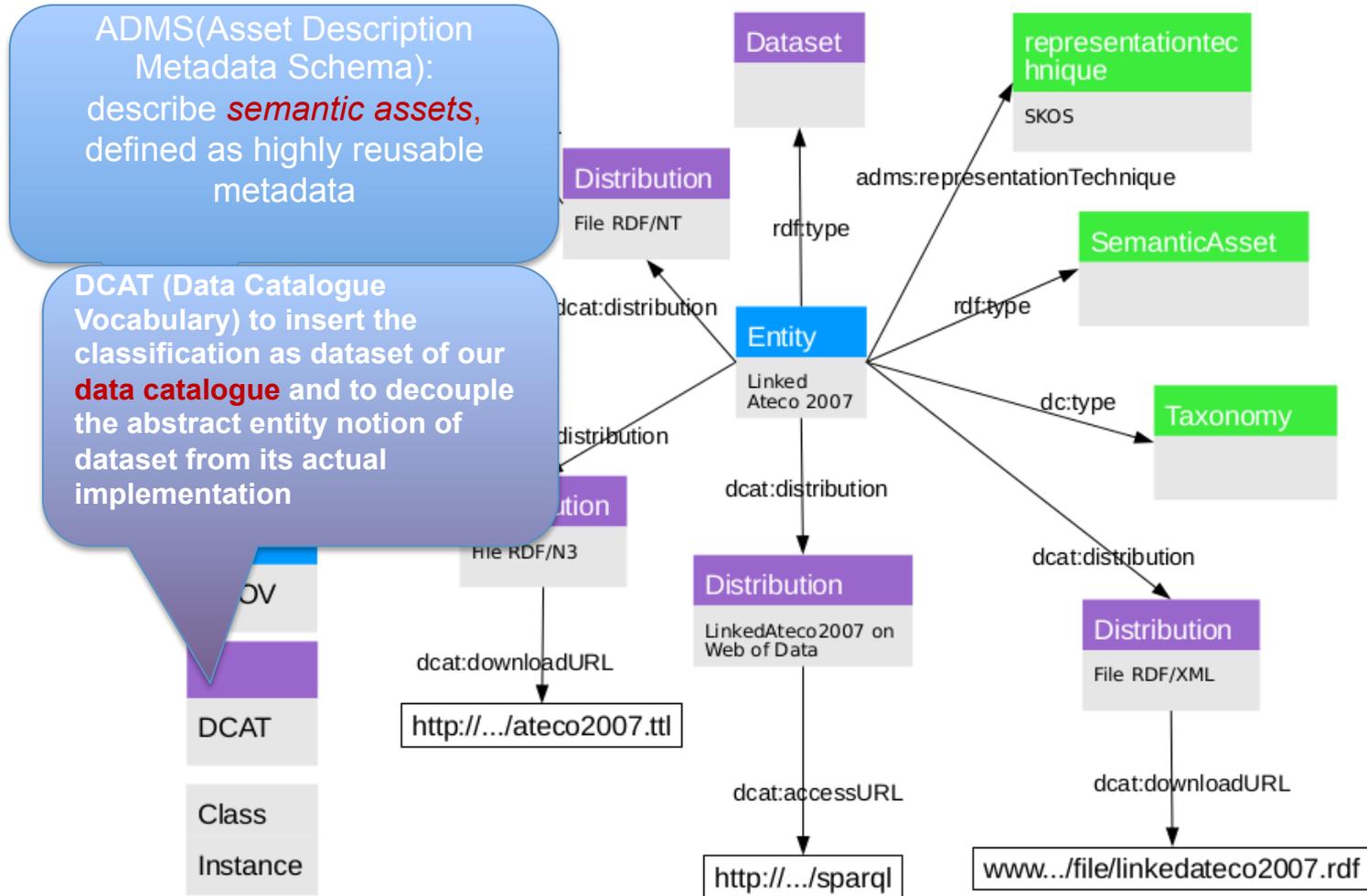
LinkedATECO 2007

- Available at AgID SPARQL endpoint:
 - <http://spcdata.digitpa.gov.it/>
- Uses different ontologies and vocabularies
- Realization steps:
 - Provenance Modeling
 - Classification Modeling
 - Classification Distribution
 - Implementation

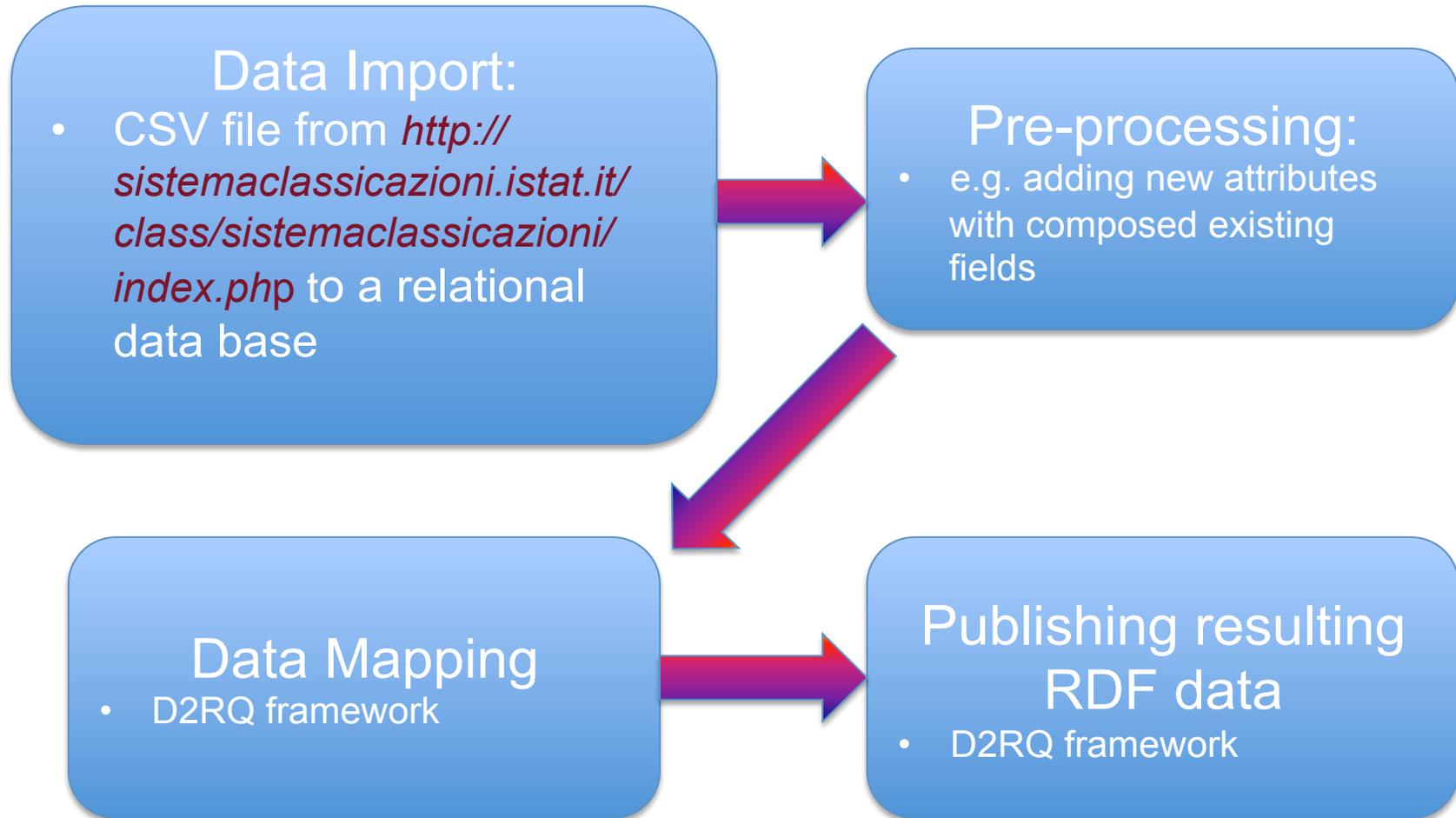
Classification Modeling



Classification Distribution



Implementation



Conclusions

- Contribution of the work:
 - Technical: modeling aspects of OS classification by making use of ontologies that are specific of the statistical domain, like XKOS, as well as of more general purpose ontologies, like PROV.
- Methodological: data interoperability is made possible by twofold efforts
 - Shared formats and modeling standards like RDF framework and XKOS
 - Content harmonization, i.e., common domain-specific information concepts.

OS classifications in LOD
as a first step towards
content harmonization