



UNECE

Implementing ModernStats Standards

Linked Open Metadata

Franck Cotton, Insee

Monica Scannapieco, Istat

Modernisation Committee on
Products and Sources



modernstats
by HLG - MOS

WP1: Linked classifications

- Problem

- Classifications are one of the main semantic assets of the statistical community
 - Base to most of the statistical publications
 - Form an internationally coherent system
 - Embed a lot of expertise
- But they are (very) poorly disseminated
 - No central catalogue
 - No common data model
 - Closed and heterogeneous formats: Excel, PDF, Access (?!), XML...



WP1: Linked classifications

- Possible solution: build a harmonised classification system
 - Rich content
 - Classifications and correspondences
 - Structure, history, notes
 - Common format, open and machine-actionable
 - Guidelines and tooling
 - Naming
 - Tools for data integration
 - Dissemination tools (browsing, extraction, querying)
- WP1 was a way to explore and assess this solution



WP1: Linked classifications

- Inputs

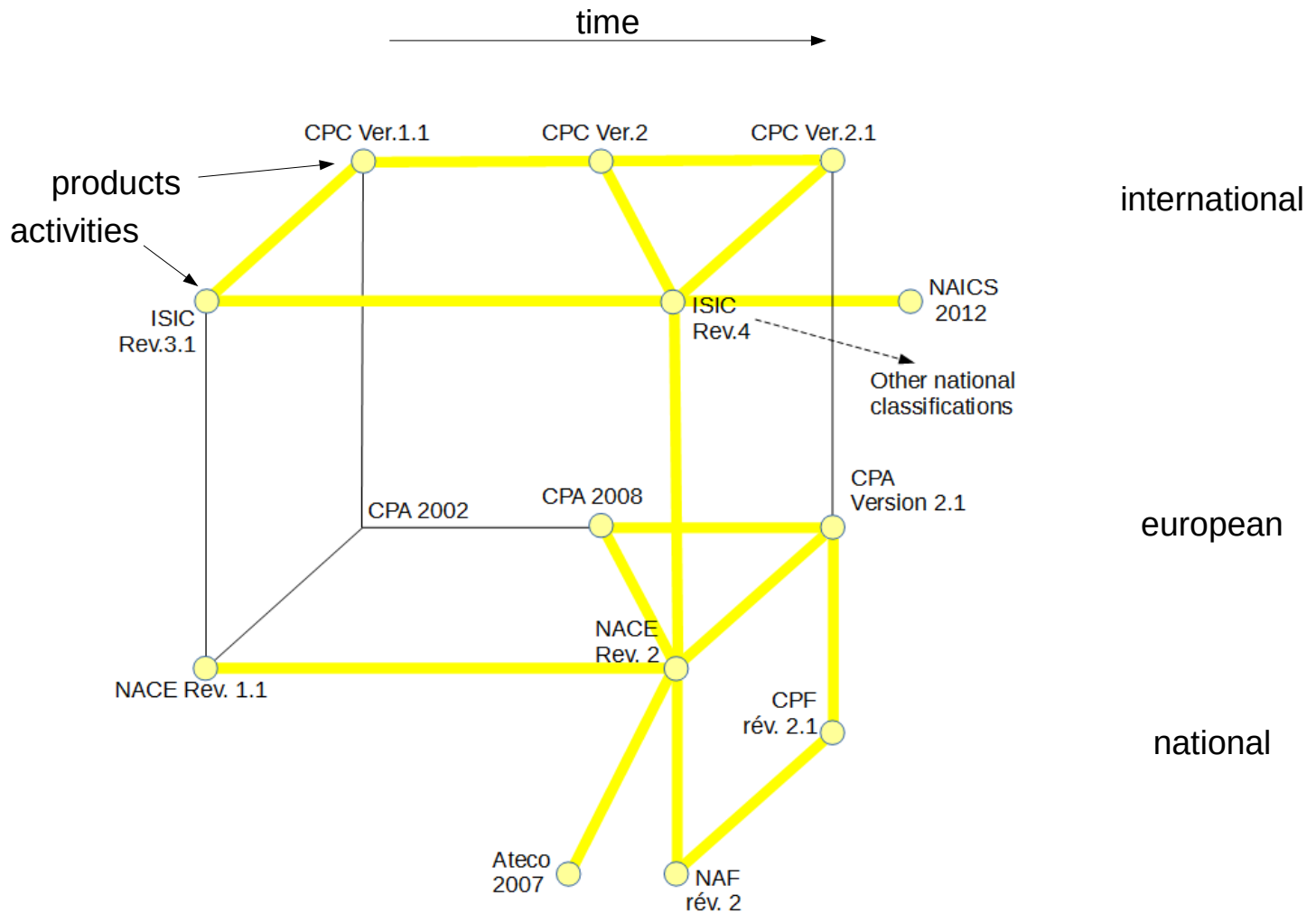
- Sandbox
- Data on classifications / codelists
- XKOS model (compatible Neuchâtel / GSIM)
- Enthusiasm

- Outputs

- Programs to extract and load the data
- Database on the sandbox
- “Classification explorer” (browse, search, export)
- Elements for guidelines
- NB: all programs are open source and on GitHub



WP1: Linked classifications



WP1: Linked classifications

- What we learned
 - Proof of concept successful
 - XKOS is fit for purpose
 - Guidelines are needed
 - Collaboration with classification experts is essential
- Demonstration



WP1: Linked classifications

- What could the HLG do now?
 - Liaise with UNSD and Eurostat to reach international consensus on making international classifications available as linked metadata (*HLG-MOS to CES*)
 - Publish design guidelines for the implementation of classifications as linked metadata (*Supporting Standards*)
 - Add new classifications (*Supporting Standards*)
 - Further develop and support the classification explorer (*Sharing Tools*)
 - Promote the system internally and externally and develop associated capabilities (*Capabilities and Outreach*)



WP2: Linked statistical models

- Problem

- The HLG models are central for in the MOS initiative
 - GSBPM now widely used as reference
 - internationally (e.g. quality indicators)
 - nationally (process description, BPR...)
 - Active CSPA developments (cf. ESSNet)
- But they are not expressed in a formal, coherent and actionable way
 - CSPA and GSBPM are Word documents
 - CSPA references both GSIM and GSBPM, but no formal semantics is used
 - Minor inconsistencies in GSIM



WP2: Linked statistical models

- Possible solution: build an integrated system of MOS models
 - Formalize CSPA and the GSBPM as ontologies
 - Translate GSIM in the same formal framework
 - Link the models
 - Develop a visual client tool for demonstration
- WP2 was a way to explore and assess this solution



WP2: Linked statistical models

- Inputs

- Sandbox
- Specifications
- GSIM UML model
- More enthusiasm

- Outputs

- 2 scientific papers
- A transformation for GSIM (can be replayed)
- “Model explorer” (browse across models, edit services)
- NB: all programs are open source and on GitHub



WP2: Linked statistical models

- What we learned
 - OWL is a good tool for expressing the set of models
 - Other vocabularies can be leveraged (e.g. PROV-O and ORG for the CSPA roles)
- Demonstration



WP2: Linked statistical models

- What could the HLG do now?
 - Adopt linked metadata as a modelling framework for the MOS models (*Supporting Standards*)
 - Add new models: Quality indicators, GAMSO, capability management... (*Supporting Standards*)
 - Further develop and support the model explorer (*Sharing Tools*)

