Modeling the Statistical Process with Linked Metadata

Franck Cotton and Daniel W. Gillman

INSEE, Paris, France
franck.cotton@insee.fr
US Bureau of Labor Statistics, Washington, USA
Gillman.Daniel@bls.gov

Abstract. The official statistics community has recently put a lot of efforts into the common formalization of its business semantics: information, process or capability models, architectural patterns, etc. One of the first results of these efforts, the Generic Statistical Business Process Model (GSBPM) is now a worldwide reference for the official statisticians. But standard models, like all metadata, need to be formally specified and machine-actionable in order to be used in the most efficient way. This idea drives the preliminary work presented in this paper, which aims at representing the GSBPM as an RDF vocabulary. We explain the approach chosen to link the model to existing ontologies, detail the results obtained so far and present some use cases and future developments.

Keywords: Statistical process, provenance, metadata, linked data, OWL.

1 Introduction

Metadata have always been essential for the statisticians. Structural metadata (codes, concepts, data structure definitions…) specify how statistical products are constructed and organized. Descriptive metadata document the methodology, the quality of the products or how they may be used. Process metadata describe and measure the statistical activities that transform raw data inputs into disseminated products.

In the recent years, the notion of active metadata has gained in importance. In this paradigm, the metadata are not simply after-the-fact documentation of the statistical data or treatments, but rather their "source code": they pilot the processes or generate the statistical tools. This guarantees that the metadata are always in sync with the actual operations. Illustrations of this approach can be found in [1] and [2].

Another trend observed recently in the statistical community is the growth of the international collaboration. The UNECE[1] is playing a key role in this phenomenon, in particular through its High-Level Group for the Modernization of Statistical Production and Services (HLG)[2]. A number of collaborative projects led by the HLG have produced valuable results that are now international standards for the community, in

[1] United Nations Economic Commission for Europe, http://www.unece.org.

[2] http://www1.unece.org/stat/platform/display/hlgbas/High-Level+Group+for+the+Modernisation+of+Statistical+Production+and+Services

particular a statistical information model[3] or a reference business architecture for the statistical industry[4].

In this paper, we focus on another output of the HLG, the Generic Statistical Business Process Model (GSBPM), which has recently become a worldwide de facto standard in the official statistics community. More and more statistical offices describe and organize their business processes in reference to this model.

But the GSBPM is just a Word document at the moment: it is not expressed in a formal and machine-readable form, and thus cannot easily be referenced or used in the current efforts for building a globally shared active metadata corpus. The objective of this paper is precisely to propose a first attempt at building a simple linked (meta)data representation of the model, so that it can offer the benefits of the linked data paradigm (unique global identification, shared semantics, linkability, etc.) and fully play its role as a common vocabulary and central reference for the statistical community.

The paper is structured as follows: in Section 2, we describe in greater detail the GSBPM and how it is organized. Section 3 gives a quick overview of previous works and existing standards that can be leveraged for our purpose. The construction of the GSBPM vocabulary itself is presented in Section 4, and Section 5 lists some intended use cases. Finally, Section 6 provides a summary and conclusion, and discusses possible future work.

2 The statistical process

In some industries, there is a common core business process all producers follow. This is true for official (national and international) statistics, where statistical data are the products of the activities of statistical institutes, and each institute follows the same general practice. Traditionally, statistical offices have used censuses (where every person or business is contacted) or sample surveys (where a select few, or sample, of persons or businesses are contacted) to collect data. These common approaches mean that all statistical offices enjoy similar successes and face the same set of problems. This, in turn, has led to a de facto standard way to conduct the business of official statistics.

More recently, as sample surveys have become significantly more expensive to conduct, statistical offices have looked at other sources of data, such as administrative records and Big Data, to augment or replace the traditional methods. However, the fundamental business process has not changed that much. Statistics and probability are used in the underlying mathematical model for handling survey design, and an advantage to this is that the model can be applied to data collections that are not sample surveys as well.

[3] GSIM, Generic Statistical Information Model, http://www1.unece.org/stat/platform/display/gsim/Generic+Statistical+Information+Model.

[4] CSPA, Common Statistical Production Architecture, http://www1.unece.org/stat/platform/display/CSPA/Common+Statistical+Production+Architecture+Home.

Over time, many statistical offices wrote process models or standards for describing their business activities. These models look remarkably similar across offices, and an effort was started under the UN Economic Commission for Europe (UNECE) to build a standard model. The effort is known as the Generic Statistical Business Process Model (GSBPM).

2.1 The GSBPM

The effort to build the GSBPM started around 2005 under the UNECE Statistical Metadata Working (METIS) Group within its project known as the Common Metadata Framework (CMF). By 2008, the first draft was completed, and GSBPM became the major deliverable of Part C of the CMF. Revisions to GSBPM have been produced periodically since.

Around 2011, the High Level Group for Modernization of Statistics (HLG) was formed, and METIS was replaced by the Modernization Committee for Standardization (MCS) under HLG. Refinements to the GSBPM are being managed under the MCS. The fifth version of GSBPM was released in late 2013.

GSBPM will be described in more detail in the next section, but it is built as a set of major process steps with additional detail under each. The terminology used to name and describe each step is generic, and the terms do not necessarily correspond to the terminology used in any country. Further, the standard is written in English, so most countries have to translate GSBPM to their own language to make it most useful locally. However, many countries have adopted GSBPM, and it is being used in many novel ways. One common use is to classify statistical processing software, and this provides a means to manage and classify statistical software and find areas where statistical offices are unnecessarily duplicating software development activities.

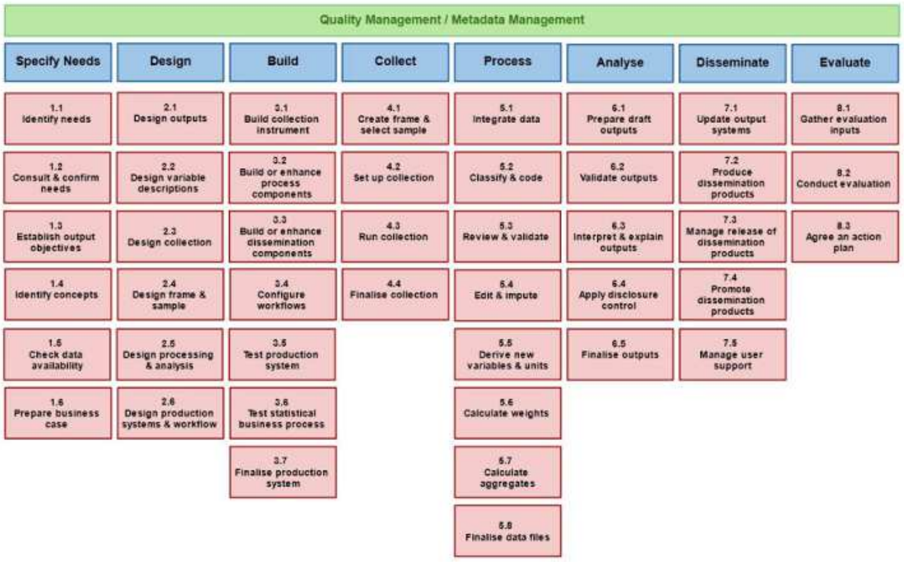
GSBPM has also spawned two other newer standards under HLG/MCS, and these are the Generic Activity Model for Statistical Organizations (GAMSO) and the Generic Statistical Information Model (GSIM). GAMSO, first produced this year, extends GSBPM to include non-statistical activities common to all statistical offices. These include many management activities such as needs for computing infrastructure. GSIM, first released in 2012, is an information model describing the statistical information objects necessary for statistical activities. GSIM and GSBPM are duals of each other in the sense that GSBPM names activities and GSIM names the inputs and outputs of those activities. In this way, GSIM and GSBPM are intertwined.

The GSBPM is not really a business process model. The flow is not described, even if there is a timeline broadly going right and down. It is centered on the naming of the different statistical activities that constitute the statistical process. In fact, it is rather a taxonomy.

2.2 Model description

GSBPM is called a business process model, but it really is a taxonomy of terms that name the activities conducted by statistical offices. Typically, process models include the flows among all the processes, but GSBPM does not go to that detail. In fact, the statistical process is fluid enough, that the order of many processes is not that important. Sometimes, steps can be conducted in any order with the same results. Therefore, the higher level taxonomy is the lowest level of detail the model can provide and still be useful world-wide.

Figure 1 below contains the top two levels of the GSBPM.



Quality Management / Metadata Management

Specify Needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
1.1 Identify needs	2.1 Design outputs	3.1 Build collection instrument	4.1 Create frame & select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update output systems	8.1 Gather evaluation inputs
1.2 Consult & confirm needs	2.2 Design variable descriptions	3.2 Build or enhance process components	4.2 Set up collection	5.2 Classify & code	6.2 Validate outputs	7.2 Produce dissemination products	8.2 Conduct evaluation
1.3 Establish output objectives	2.3 Design collection	3.3 Build or enhance dissemination components	4.3 Run collection	5.3 Review & validate	6.3 Interpret & explain outputs	7.3 Manage release of dissemination products	8.3 Agree an action plan
1.4 Identify concepts	2.4 Design frame & sample	3.4 Configure workflows	4.4 Finalise collection	5.4 Edit & impute	6.4 Apply disclosure control	7.4 Promote dissemination products	
1.5 Check data availability	2.5 Design processing & analysis	3.5 Test production system		5.5 Derive new variables & units	6.5 Finalise outputs	7.5 Manage user support	
1.6 Prepare business case	2.6 Design production systems & workflow	3.6 Test statistical business process		5.6 Calculate weights			
		3.7 Finalise production system		5.7 Calculate aggregates			
				5.8 Finalise data files			

Fig. 1. Generic Statistical Business Process Model

There are eight main process steps, which are labeled in the blue boxes at the tops of each of the columns. The items below each of the blue boxes are the sub-processes under the main process labels. Provided across the top are two over-arching steps, quality and metadata – as these are important activities that go along with every process in the model. In other words, every process generates metadata needed to describe that process and describe its usage; and every process has a quality component attached to it.

The main processes, or phases, are defined as follows:

- Specify Needs - the initial investigation and identification of what statistics are needed and what is needed of the statistics
- Design - the development and design activities, and any associated practical research work needed to define the statistical outputs, concepts, methodologies, collection instruments, and operational processes
- Build - build and test the production solution to the point where it is ready for use in the "live" environment
- Collect - collects or gathers all necessary information (data and metadata), using different collection modes (including extractions from statistical, administrative and other non-statistical registers and databases), and loads them into the appropriate environment for further processing

- Process - the cleaning of data and their preparation for analysis
- Analyze - statistical outputs are produced, examined in detail and made ready for dissemination
- Disseminate - manage the release of the statistical products to customers
- Evaluate - manages the evaluation of a specific instance of a statistical business process

Under each main process are defined several sub-processes. These define specific kinds of processes under each of the main headings. Further detail can be described, but at this level those details are probably institute specific.

The main result of the GSBPM is a taxonomy of business production activities. This is made possible in part by the acceptance of common practices across the official statistical community world-wide.

3 Choosing base vocabularies

When designing an RDF vocabulary, it is always good practice to relate it to existing ontologies, in order to leverage field-proven models. It helps to clearly specify the semantics, facilitates the creation of links between models or entities that is at the heart of the Linked Data paradigm, and allows the reuse of existing knowledge.

The GSBPM being a business process model, a first line of work consist in looking towards existing standards in the field of business process modeling (BPM). But we also noted that the GSBPM could be seen as a taxonomy of statistical activities, which leads us to examine the domain of knowledge organization systems. Finally, we can refer to the transformative character of the statistical process and associate that to information provenance models.

3.1 Business process modeling standards

One of the most prominent standards in the BPM domain is the Business Process Model Notation (BPMN), maintained by the OMG[5].

BPMN is a diagramming standard: it defines a way to represent process objects like activities, data objects, events or gateways (where activity flow paths are combined or separated). The most recent version of the standard is version 2 [3], following versions 1.0, 1.1 and 1.2.

There has been a lot of work realized in the area of semantic BPM (SBPM), in particular around BPMN. For example, [4] presents a very comprehensive work made on the semantic formalization of BPMN 1.1. There is also an ongoing work by the same team for building a BPMN 2.0 ontology, but it is not achieved yet, to the best of our knowledge.

The authors clearly indicate that the BPMN ontology models graphical symbols, and not the represented process components. For example, all main classes extend a "graphical-element" class. This is not the notion that we are looking for.

[5] Object Management Group, http://www.omg.org.

In addition, the conformity to the BPMN model leads to introduce some constraints that are not suited for our purpose (for example, the "activity" class must have a type, start and completion quantities, etc.).

Another important BPM standard is WS-BPEL [5], or BPEL in short. Here also, some SBPM references exist, for example [6]. This BPEL ontology introduces an "Activity" class at the base of a hierarchy of more specific activities. There again, in conformity with the objective of the ontology, what is modeled is the BPEL activity (basically a combination of web services): the Activity is described by "Concept of being a BPEL activity", whereas GSBPM phases or sub-processes, or even statistical production activities, are of course not all activities in the sense of BPEL. Moreover, the BPEL ontology does not seem to be available as OWL.

As a conclusion, both the BPMN and BPEL ontologies can be very useful in the representation or annotation of the detailed statistical processes that the GSBPM is meant to classify, but neither provides the notion of generic statistical activity (or even simply activity) that we would need here. The authors of [4] refer to the DOLCE[6] upper ontology for the modeling of the processes themselves, so we could envision to use it also, or its lighter OWL manifestation DOLCE+DnS Ultralite (DUL)[7], but the Process as defined in DOLCE is a very generic concept, and DOLCE is based on an ontological approach, very well described in [7], that needs to be studied further to see if it is appropriate to our use case.

3.2 knowledge organization systems

We saw in the introduction that GSBPM activities and sub-processes are rather activity categories where actual statistical activities can be classified. In that sense, the model is close to a taxonomy of statistical activities, and one can think of describing it as a concept scheme as defined in the Simple Knowledge Organization System (SKOS) [8], by far the most used knowledge organization vocabulary in the RDF world.

In such a representation, the GSBPM phases and sub-processes, as well as the whole statistical production process itself, would be viewed as concepts belonging to the GSBPM concept scheme. In SKOS terms, a sub-process would have for broader concept the phase to which it belongs, this phase having itself the whole statistical process as broader concept.

Using SKOS also has the advantage of being able to leverage the different documentation properties defined in the recommendation (`skos:note` and all its sub-properties), in order to store the abundant textual material contained in the GSBPM specification. It also directly enables the use SKOS extensions, like the XKOS vocabulary [9] which is already known in the statistical community and which defines in particular semantic relations between concepts that refine those defined in SKOS, for example partitive or generic/specific relations.

[6] https://en.wikipedia.org/wiki/Upper_ontology#DOLCE_and_DnS.

[7] http://ontologydesignpatterns.org/wiki/Ontology:DOLCE%2BDnS_Ultralite

3.3 Provenance models

When searching for the term 'activity' in the Linked Open Vocabularies (LOV)[8], the first answer returned belongs to the PROV ontology, or PROV-O [10].

PROV is a set of W3C Recommendations that defines a data model and various serializations to support the representation and exchange of provenance information.

It is based on a very simple and elegant core model:

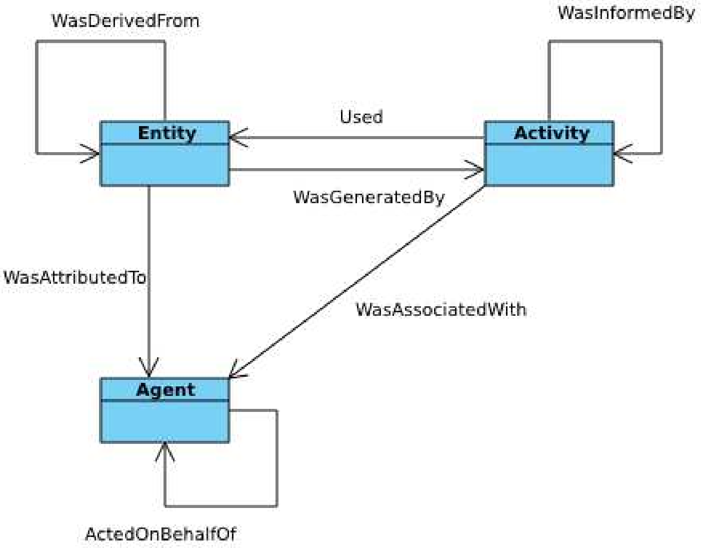


Fig. 2. PROV Core Structures (copyright © 2011-2013 W3C®)

In this model, an entity is "a physical, digital, conceptual, or other kind of thing with some fixed aspects", and an activity is "something that occurs over a period of time and acts upon or with entities". These definitions are quite generic, but the model represented above gives them a semantic context that seems immediately relevant to the statistician. Statistical production is all about informational entities that derive from each other through a succession of activities informed by one another.

Looking more precisely into the complete PROV model, it is easy to verify that it contains all the information needed to implement process traceability. This is a very important feature for the statistical organizations, which put more and more efforts into quality insurance.

3.4 Conclusion

To conclude this section, we see that the specialized ontologies defined in the SBPM field in relation to well-known standards like BPEL or BPMN are too specialized to be used in the representation of the GSBPM, but could be very useful for the

[8] http://lov.okfn.org

modeling or annotation of more fine-grained activities that would be classified according to the GSBPM.

In contrast, more high-level vocabularies like PROV and SKOS can easily be related to the model, and they bring valuable additional semantics that can add to the expressivity and usefulness of the model.

4 Building the vocabulary

This section builds on the previous ones and explains concretely how we design the GSBPM vocabulary. Given that PROV-O and SKOS/XKOS are expressed in OWL, we adopt the same formalism, although in practice only RDFS statements are used at this early stage, except for the ontology object itself (a `owl:Ontology`). The code snippets[9] given below as examples use Turtle for the representation of RDF 1.1 statements, and the usual prefixes for the well-known vocabularies as documented in `http://prefix.cc/`.

For now, the vocabulary defines only classes and individuals: datatype or object properties could be introduced in future versions.

4.1 Namespaces

The GSBPM is published and maintained under the auspices of the UNECE, so we tentatively use as base namespace `http://rdf.unece.org/models/gsbpm#` for the ontology, and `http://id.unece.org/models/gsbpm/` for the individuals. Using two different namespaces emphasizes the special importance of global identification for the GSBPM individuals. This design choice can easily be changed and the two namespaces merged in the future. In any case, the namespaces must be discussed and validated with the UNECE and HLG, in particular to implement dereferenceability of the URIs.

The prefixes associated with the namespaces mentioned above will be respectively `gsbpm` and `igsbpm`.

4.2 Classes

As we saw in the description of the GSBPM, the central notion is "statistical production activity". In order to capture this conclusion, we create a base class named `StatisticalProductionActivity`. As a consequence of the study made in the previous section, this class will be defined as a sub-class of both `prov:Activity` and `skos:Concept`:

```
gsbpm:StatisticalProductionActivity
  a rdfs:Class, owl:Class ;
  rdfs:label      "Statistical production activity"@en ;
```

[9] The complete ontology is available at: https://github.com/FranckCo/Stamina/raw/master/doc/gsbpm.ttl.zip

```
  rdfs:subClassOf prov:Activity , skos:Concept .
```

This class is further specialized into `gsbpm:Phase` and `gsbpm:SubProcess`. The aggregation relation between phases and sub-processes is not represented at this level (it could be in OWL), but rather at the level of the individuals.

4.3 Individuals

The GSPBM model, considered as a global taxonomy of statistical activities, is represented in the model as a `skos:ConceptScheme` (only a few properties are reproduced below). This materializes a clear distinction between the model itself and the statistical process which is modeled.

```
gsbpm:GSBPM
  a skos:ConceptScheme ;
  rdfs:label "Generic Statistical Business Process
              Model"@en ;
  foaf:homepage <http://www1.unece.org/stat/
                platform/display/GSBPM/GSBPM+v5.0> .
```

The statistical process as a whole is represented as an instance of `StatisticalProductionActivity`:

```
igsbpm:StatisticalProductionProcess
  a                 gsbpm:StatisticalProductionActivity ;
  rdfs:label        "Statistical Production Process"@en ;
  skos:topConceptOf igsbpm:GSBPM ;
  skos:narrower     igsbpm: 1 , igsbpm: 2 , igsbpm: 3,
   igsbpm:4 , igsbpm:5 , igsbpm:6 , igsbpm:7 , igsbpm:8 .
```

In the same manner, all the phases and sub-processes defined in the GSBPM are also instances of `StatisticalProductionActivity`. We give below example definitions of a phase and a sub-process (only a few properties are shown, and texts have been shortened for brevity):

```
igsbpm:4 a gsbpm:Phase ;
  skos:notation   "4" ;
  skos:prefLabel  "Collect"@en ;
  skos:definition "This phase collects..."@en ;
  skos:inScheme   igsbpm:gsbpm ;
  skos:narrower   igsbpm:4.1 , igsbpm:4.2 , igsbpm:4.3
                  igsbpm:4.4 ;
  skos:broader    igsbpm:StatisticalProductionProcess ;
  xkos:isPartOf   igsbpm:StatisticalProductionProcess .

igsbpm:6.2 a gspbm:SubProcess ;
  skos:broader     gsbpm:6 ;
  xkos:isPartOf    gsbpm:6 ;
  skos:definition  "This sub-process is..."@en ;
  skos:inScheme    gsbpm:gsbpm ;
  skos:notation    "6.2" ;
  skos:prefLabel   "Validate outputs"@en .
```

The `skos:broader` and `skos:narrower` properties are used to formalize the hierarchical structure of the GSBPM, doubled with partitive relations like `xkos:isPartOf` for the use cases that understand XKOS.

We chose not to express relations of causality, sequencing or temporality between phases or sub-processes because, as explained several times in the GSBPM textual material, they can overlap or be altogether missing in some cases. In the description of more precise processes, statistical activities could be connected by these kind of relations using XKOS associative properties like `xkos:causal`, `xkos:sequential`, `xkos:disjoint` or their sub-properties.

The reader may have notices that the over-arching processes of the GSBPM have not been included in the model. This is intentional, because quality and metadata management need to be considered in a broader perspective that is now covered by the GAMSO. The articulation between the GSBPM and the GAMSO still need a bit of work, and semantic formalization of both models will undeniably help in that respect.

5 Use cases

As we indicated in the introduction, the formalization of the semantics of the GSBPM is an objective in itself because the model is a reference framework for official statistics. Representing it as a simple, yet precise, RDF model opens the possibility to discuss within the community in order to improve this representation and come to an agreement on a shared semantics.

The use of RDF also provides a global naming scheme, so that we can refer unambiguously to the model or its specific components. Also, RDF being multilingual by design, we can enrich the GSBPM vocabulary with translations of the textual descriptions that it contains. Spanish and French translations of the previous version of the GSBPM are available and could easily be upgraded and integrated.

Beyond this "intrinsic" utility, the work initiated here can be leveraged in different use cases. We give below a few examples.

First, the vocabulary provides an anchor for national refinements of the GSBPM. Several countries have developed more precise models where the sub-processes are broken down into more precise activity descriptions. The approach exposed in this paper can be reused to model these extensions and to attach them to the main model. This will in particular allow for comparisons between the different variations that can exist at national level.

Similarly, the GSBPM ontology can be used to organize and share formal descriptions of actual statistical processes. If a country models a process, either with the SBPM tools and techniques that we briefly reviewed in section 3 or by means of any other formal representation, it can refer by URI to (and be referred by) the relevant GSBPM sub-process. We see here the possibility to build a global reference of statistical process descriptions, classified according to the GSBPM, that would be fully in line with the CSPA approach (see below).

Another interesting use case has already been mentioned and is offered by the articulation with the PROV ontology. PROV's very rich model opens great perspectives for the documentation of the statistical process, which is basically a very complex transformational process. The possibility to capture provenance information in a standard way along the whole flow, even at a relatively coarse-grain level, is a very interesting one.

Finally, the simple fact to actually store the model in a RDF database with basic capabilities like querying, full-text search or simple visualization would greatly improve the dissemination of the GSBPM.

6 Conclusion and future work

In this paper, we presented a preliminary work on representing the Generic Statistical Business Process Model (GSBPM) as an RDF vocabulary. We studied different standards existing in the field of Semantic Business Process Modeling like the BPMN and BPEL ontologies, and concluded that they could be useful for representing actual "real-life" statistical processes, but that the GSBPM itself was too high-level to be directly related to them. On the contrary, more generic models like those of SKOS and PROV can easily be used in that respect, and they bring valuable enrichments and generate new use cases.

The work presented here is only a first step. The big test, of course, will be whether the statistical community wants to use the GSBPM vocabulary. For that, we need to consolidate the vocabulary, have the model validated by the community, make the URI scheme and license model official, etc. We also should implement a quick and easy use case like the basic visualization application mentioned at the end of Section 5. It would also be important to study in more detail the work done in the SBPM field in order to provide advice to the statistical community on modeling processes and attaching them to the GSBPM.

The work ahead must be put in the perspective of the international collaborations going on. The Common Statistical Production Architecture (CSPA) is an effort to standardize statistical production software across all statistical offices. This effort involves the idea that production systems are built up from small modules that everyone shares. However, GSBPM will guide the development of which modules must be built, and GSIM will guide which metadata are inputs and outputs to each process. GAMSO will extend the description to activities that are outside of statistics proper. As these standards are formalized as linked metadata, it will be possible to move towards the ultimate objective of a completely formalized semantics of the statistical domain, in order to be in the position to implement the "Active meta-data vision".

References

1. Rivera, A., Wall, S. and Glasson, M.: Metadata-driven business process in the Australian Bureau of Statistics (2013), downloaded from the web on 9 July 2015 at http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.40/2013/WP21.pdf.
2. Sigaud, É., Tailhurat, R., Cotton, F. and van der Vlist, É.: XForms generation, a real world example (2013), downloaded from the web on 9 July 2015 at http://www.balisage.net/Proceedings/vol11/html/Cotton01/BalisageVol11-Cotton01.html.
3. OMG, Business Process Model and Notation (BPMN), Version 2.0 (2011), specification downloaded from the web on 9 July 2015 at http://www.omg.org/spec/BPMN/2.0/PDF.
4. Rospocher, M. Ghidini, C. and Serafini, L.: An Ontology for the Business Process Modelling Notation (2014), downloaded from the web on 9 July 2015 at https://dkm-static.fbk.eu/people/rospocher/files/pubs/2014foisbpmn.pdf.
5. OASIS, Web Services Business Process Execution Language v2.0 (2007), specification downloaded from the web on 9 July 2015 at https://www.oasis-open.org/standards#wsbpelv2.0.
6. Nitzsche, J., Wutke, D., and van Lessen, T.: An Ontology for Executable Business Processes (2007), downloaded from the web on 9 July 2015 at http://ceur-ws.org/Vol-251/paper8.pdf.
7. Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., Schneider, L.: Sweetening Ontologies with DOLCE (2002), downloaded from the web on 9 July 2015 at http://www.researchgate.net/profile/Nicola_Guarino/publication/221630979_Sweetening_Ontologies_with_DOLCE/links/0046352ce74538d87a000000.pdf
8. W3C, SKOS Simple Knowledge Organization System Reference (2009), recommendation downloaded from the web on 9 July 2015 at http://www.w3.org/TR/skos-reference/.
9. Data Documentation Initiative (DDI) Alliance (2012), XKOS - Extended Knowledge Organization System, specification downloaded from the web on 9 July 2015 at http://www.ddialliance.org/Specification/RDF/XKOS.
10. W3C, PROV-O: The PROV Ontology (2013), recommendation downloaded from the web on 9 July 2015 at http://www.w3.org/TR/prov-o/.