

**The GovStat Ontology**  
**Maria Cristina Pattuelli, Stephanie W. Haas, Ron T. Brown, & Jesse Wilbur**  
**[pattm, haas, browr]@ils.unc.edu, jdwilbur@email.unc.edu**

- Type: Poster
- Demo as well? No
- Author(s): see above
- Address: School of Information and Library Science, University of North Carolina at Chapel Hill, 100 Manning Hall, Chapel Hill 27599-3360

## **Introduction**

The GovStat Project (<http://www.ils.unc.edu/govstat/>) is a joint project developed by the University of North Carolina Interaction Design Lab and the University of Maryland Human-Computer Interaction Lab. The goal of the project is to create an integrated model that facilitates access and use of US government statistical information. The final perspective directing the project is the creation of a unified Statistical Knowledge Network (SKN). The primary research areas under investigation are metadata, interface design, and vocabulary tools (Marchionini, Haas, Plaisant, Shneiderman, & Hert, in review).

People with limited statistical knowledge often have problems finding the statistical information they need and understanding what it means and how to use it. The vocabulary support team within the GovStat Project is developing the Statistical Interactive Glossary (SIG) and the GovStat ontology as online vocabulary tools for supporting users seeking statistical information in government websites.

## **Purpose**

The purpose of the GovStat ontology is to supply semantic support for the SIG. The SIG is an enhanced glossary of statistical terms that users of federal statistical websites often need to understand in order to find or use the information they seek. Relevant characteristics of the statistical glossary include:

- Content - The selection of terms and the content of each explanation are aimed at providing a basic level of statistical literacy. Links to more complete or more technical explanations, such as those found in agencies' technical documents, can be included in the terms' presentations or ontology entries. The content of the explanations may include definitions, examples, brief tutorials, demonstrations, interactive simulations, or combinations of these elements.
- Context specificity - Explanations will be provided at different levels of specificity. When a user invokes help from a term (e.g., by clicking on it), the most specific explanation for the term and context are offered. If there is no explanation appropriate for a specific context, then a more general explanation is offered. The default is a set of context-free, "universal" presentations that can be invoked from any location.
- Format - Explanations in a variety of formats, including text, text plus audio narration, still images, animation, and interactive simulations will be provided. Help facilities of any kind must be attractive enough to entice users.

The ontology provides a means of navigating through statistical terms and definitions linked in a network of relationships. The GovStat ontology is domain-specific in that it deals with concepts from a specific area of knowledge – statistics.

## **Scope**

The GovStat ontology reflects the scope of the glossary, being generally limited to those terms and concepts that a user may encounter on the agency websites. The exception to this is the occasional need to include concepts to connect target concepts. By being user-specific and application-dependent (Fernandez Lopez, 1999; Zuniga, 2000), the GovStat ontology models only that portion of statistics appropriate for its intended use.

### Structure and content

The GovStat ontology represents terms and their relationships, and also informs the content and context of a term's presentation. The two categories of relationships used in the GovStat ontology are taxonomic and domain or functional relationships. The taxonomic relationships are “partial ordering relations” of the type *is-a* and *part-whole*.

The *is-a*, or subsumption relation, is the basis of taxonomy and it is the most common relation for modeling concepts (Guarino & Welty, 2002). Examples in the GovStat ontology include:

Seasonal\_adjustment-Adjustment (See Figure 1)  
Age\_adjustment-Adjustment (See Figure 1)

The *part-whole*, or mereological relation, can be of various types. An example of part-whole relation in the GovStat ontology is:

Sample-Population (See Figure 2)

According to the classification proposed by Winston, Chaffin, and Hermann (1987), the relationship between sample and population would be considered a “portion-mass” or “slice-cake” relationship.

The other category of relationships represented in the GovStat ontology is that of domain or functional relations. These are typed relationships between terms which are able to express rich semantics. Examples in the GovStat ontology include:

Is\_an\_estimate\_of (See Figure 2)  
Smooths (See Figure 1)

The final structure of the GovStat ontology is likely to result in a “forest” (Sowa, 1984) or a family of trees, each expressing specific aspects or facets of the domain of interest (Smith, 2002) rather than in a wide taxonomy composed of one single tree.

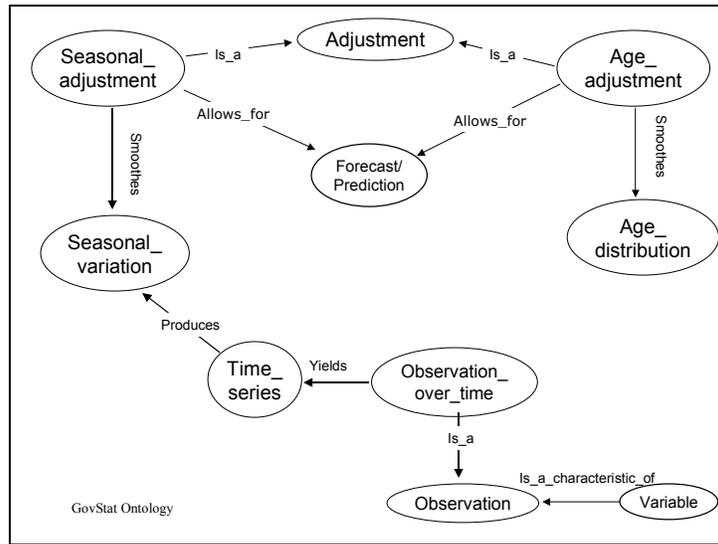
The GovStat ontology supports the design and deployment of the SIG explanations in several ways including:

- Support for context-specific presentations. Ontology as a navigation device from specific to general level allows users to identify related terms. Inheritance of taxonomic relationships between concepts will support the provision of context-specific presentations.
- Identification of related terms. Associations among concepts also suggest opportunities for combining related concepts into a single more comprehensive explanation, such as a tutorial. For example, the *part-whole* relationship between *sample* and *population* suggests that an explanation of *sample* should include a mention of the population from which a sample is drawn.
- Support for explanation templates. Once an example of a concept has been established, then definitions or examples of subclasses of the concept can follow the template, with minor

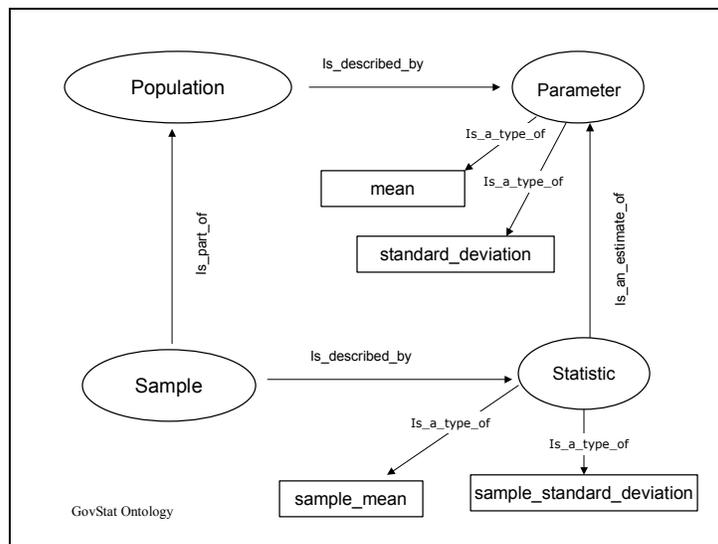
adjustments. For example, explanations for *adjustment* can include a template that illustrates the idea of smoothing statistics to remove predictable variation. Explanations of subclasses of *adjustment*, such as *seasonal adjustment* or *age adjustment* can also be incorporated into this template. Templates streamline the creation of additional presentations for other subclasses or for additional contexts.

In addition to the functions described above, the GovStat ontology offers an exploration space where users can manipulate it directly, increasing their understanding of statistical concepts by exploring the hierarchy of the concepts and the semantics of the domain relationships between them.

### Examples



**Fig. 1.** Graphical representation of *Adjustment*.



**Fig. 2.** Graphical representation of *Sample* and *Population*.

## Development

Ontology development is an iterative process composed of a series of activities. The methodology adopted for the development of the GovStat ontology includes the following activities:

- **Specification**
- **Knowledge Acquisition**
- **Conceptualization**
- Formalization
- Implementation
- Integration
- Evaluation
- **Documentation.**

The activities in bold represent those completed or in process.

**Acknowledgements.** This research was supported by NSF grant (EIA 0131824). We would also like to thank the entire GovStat team for helpful comments and discussions.

## References

- Fernandez Lopez, M. (1999). Overview of methodologies for building ontologies. In Proceedings of the IJCAI-99 Workshop on Ontologies and Problem-Solving Methods: Lessons Learned and Future Trends. CEUR Publications, 1999. *Intelligent Systems*, 16(1):26-- 34, 2001.
- Guarino, N. & Welty, C. (2002). Evaluating ontological decisions with Ontoclean. In *Communication of the ACM*. 45:61-65.
- Haas, S. W., Pattuelli, M. C., & Brown, R. T. (in review). Understanding statistical concepts and terms in context: The GovStat Ontology and the Statistical Interactive Glossary. Submitted to the *ASIST 2003 Annual Conference*.
- Marchionini, G., Haas, S. W., Plaisant, C., Shneiderman, B., & Hert, C. (in review). Toward a statistical knowledge network. Submitted to the *dg.o2003 National Conference on Digital Government Research*.
- Smith, B. (2002). Ontology and information systems. Retrieved December 3, 2002 from <http://ontology.buffalo.edu/ontology.doc>.
- Sowa, J. F. (1984). *Conceptual structures: Information processing in mind and machine*. Reading, MA: Addison Wesley.
- Winston, M. E., Chaffin, R., & Hermann, D. J. (1987). A taxonomy of part-whole relations. *Cognitive Science*, 11:417-444.
- Zuniga, G. L. (2000). Ontology: Its transformation from philosophy to information systems. In *Papers from the Second International Conference* (pp. 187-197). ACM Press.