

COLLEGE STANDAARDISATIE

Summary of Semantic Interoperability for Standardisation Forum**for Logius (Peter Waters)****Jeff Rothenberg****RAND-Europe****March 26, 2010**

This short paper summarizes and extends the results of several recent reports commissioned by the Standardisation Forum on the subject of semantic variability across multiple eGovernment contexts.

Two of the central motivations for creating eGovernment systems are the integration and coordination of interactions among distinct government agencies and the simplification and improvement of citizens' interactions with government. Both of these goals require the increased use of common data, such as registries, in order to reduce redundancy (thereby increasing the consistency and accuracy of the data) and reduce duplication of effort in creating, capturing, and maintaining data. Moreover, both goals require that eGovernment systems performing different tasks or belonging to different agencies work together automatically to coordinate and integrate the activities of these tasks and agencies.

In order for eGovernment systems to work together meaningfully, they must understand each other's terms, concepts, processes, and policies. Achieving semantic interoperability of this kind is far more challenging than simply enabling systems to communicate with each other. Systems that perform specific tasks for different purposes or agencies, or at different levels of government, derive their semantics from the legislation, regulations, procedures, policies, organizational cultures and social traditions related to their missions: each of these factors contributes to the context of a specific task. These contexts can be quite different for government functions as diverse as those related to tax administration, the distribution of social benefits, housing, health, or education. Seemingly simple terms, such as "employer" or "parent" can have distinctly different meanings in these different contexts, and these meanings must be understood and respected by each of a cooperating group of systems, for example to ensure that all of the systems refer to the same person.

A given entity may also have different meanings or different attributes depending on what role it is playing in a specific task. For example, one role of an employer is to withhold taxes from its employees' income, whereas another is to serve as the basis for their eligibility for social benefits. Such multiple roles involve different aspects and activities of the entity and therefore constitute different contexts as well.

Variations in meaning across contexts are not (in most cases) merely artifacts of terminology or jargon: they represent real, underlying differences in the tasks performed in those contexts. For example, tax assessment and collection may define individuals and their relationships to each other (such as dependency) quite differently from the ways these are defined for benefits or issues involving legal custody. These differences are typically defined in legislation, regulations, and policies, but they are even more fundamental than that—inhering in the purpose and nature of the tasks being performed. Although it may be tempting to imagine that these semantic differences can be reconciled by mutual agreement and compromise, this denies the reality that the differences are often irreducible and significant. Such variations in meaning are important and desirable, and they cannot be ignored or wished away.

Understanding and respecting different meanings in different contexts

The fundamental problem of semantic interoperability is how to resolve differences in meaning across distinct contexts while understanding and respecting those differences. Here “resolving” does not in general mean simply finding the common ground between different meanings (i.e., their intersection), since there may be none—or it may not be useful. Instead, resolving differences may mean expanding the context of each meaning until related or identical contexts are found for each—and using these overlapping, expanded contexts to connect the two meanings.

As an example, suppose that a medical ICT system is attempting to automatically derive part of a child’s medical history, for which it may need to find information about the child’s biological father. In order to do this, the medical system might query some non-medical system or registry to find out who the child’s father is. However, in the non-medical system’s context, “father” may mean legal guardian, which may not be the same person as the biological father. The query must therefore not simply ask for the “father” of the child but must specify that the semantics of the query requires the biological father. In response to this query, the non-medical system must expand its normal context for “father” to include other relationships for the child, until it finds one that involves biological parentage; a male relation of the child in this context can then be returned as the child’s “father” in the context of the query.

The recent studies and reports commissioned by the Standardisation Forum on this subject have analyzed specific cases of such variations in meaning, tracing the roots of that variation to underlying legislative and task-related factors. These reports have concluded that in order to make semantic interoperability among eGovernment systems a reality, it is first necessary to identify the semantic variation across the contexts of these systems (and the agencies to which they belong) and to make this variation explicit.

Because semantic variation is inextricably linked to context, one key to making it explicit is to employ a simple but general method of representing and visualizing different contexts and the relationships among them. The MetaPattern modeling technique, developed by Pieter Wisse, is well suited to this purpose and is used in the Standardisation Forum reports to illustrate the complexities of context and the ways they determine variant meanings. In addition, the reports offer the beginnings of some specific models of context for key government concepts, such as

employer-employee relationships. These MetaPattern models show that complex contexts and their associated variant meanings can be specified and understood by human readers, albeit with some effort.

Automating semantic interoperation

The ultimate purpose of such modeling efforts (in the context of the Standardisation Forum) is to enable meaningful automatic interoperation among eGovernment systems. Prior to the advent of such systems, agencies interoperated with each other by means of human mediation. Personnel in each agency understood their agency's context and were able to translate and interpret that context appropriately when interacting with personnel from other agencies, in order to perform cooperative tasks. Different individuals might apply somewhat different interpretations, but training, documentation, and experience were typically enough to ensure consistent and meaningful interoperation.

In the age of eGovernment, however, such human mediation among agencies must largely be replaced by automation. This should enable faster and more consistent government response to citizens' requests, at any hour of the day or night. Yet automating this process requires a deep and pervasive understanding of semantic variation across contexts, as well as the embodiment of traditional human mediation processes in eGovernment computer programs.

The MetaPattern modeling technique offers a means of specifying the complex contextual relationships that define variant meanings. However, in order to enable eGovernment systems to utilize these contextual relationships to resolve semantic variation, it is necessary to represent contextual semantics in actual ICT systems. A next step toward enabling automated interoperation among eGovernment systems would therefore be to represent the contextual specifications of some selected use cases in an actual database or knowledge base, using one of a number of existing candidate formalisms, such as the Web Ontology Language (OWL), Topic Maps, or Cyc, for use in an actual ICT system. Examples of semantic variability that could be explored in such use cases might include concepts such as "employer" or "child" that have been modeled in the recent Forum studies.

Appropriate use cases for this purpose might involve answering simple queries for citizens or performing simple procedures for them. A sequence of increasingly complex steps of this sort could be undertaken, starting with use cases involving semantic variants within a single eGovernment system belonging to a single agency and progressing to cases involving variants in multiple systems within a single agency and ultimately to systems belonging to multiple agencies.

Beyond merely representing the variant meanings of concepts in ICT systems, this prototype effort should develop strategies and techniques for combining, translating, and mediating variant semantics across multiple contexts for the selected use cases, as in the "biological father" example, above. These strategies and techniques, along with the contextual semantics representations underlying them, should form the basis for the future expansion of interoperability among eGovernment systems.

Conclusions

The interoperability studies undertaken by the Standardisation Forum to date have shed considerable light on the sources of semantic variation, while demonstrating the conceptual utility of using contextual models to identify and understand this variation.

MetaPattern provides a useful modeling tool for conceptualizing and understanding complex relationships among semantic contexts. Although many other tools have been developed in the Knowledge Representation sub-field of Artificial Intelligence that also recognize the importance of context, the MetaPattern formalism focuses a unique blend of simplicity, flexibility, and visual clarity on the task of contextual specification. Unlike many other modeling methods that are designed to implement their representations in ICT systems, MetaPattern remains aloof from such implementation concerns, making it well suited to the conceptual modeling of semantic contexts prior to actually building eGovernment systems. It would therefore be useful to encourage the use of MetaPattern among ICT practitioners and across agencies that are involved in the design of eGovernment systems.

One way to encourage this might be to present papers and organize panels or sessions at ICT and eGovernment conferences and workshops, focusing on the impact of semantic variation on interoperability among eGovernment systems and the use of contextual modeling as a way of understanding and dealing with that variation. In such fora, MetaPattern should be offered as a useful tool, but other approaches should be included and discussed as well, to avoid inevitable resistance from the “not invented here” syndrome among modelers and system designers.

In addition, however, in order to promote acceptance of the MetaPattern technique—and moreover, pave the way toward effective eGovernment—it is necessary to demonstrate that the use of MetaPattern supports the design and implementation of semantically interoperable eGovernment systems. As suggested above, MetaPattern contextual specifications of a sequence of increasingly complex use cases that embody semantic variability (in concepts such as “employer” or “child”) should be implemented in prototype ICT systems, using some existing formalism, such as OWL, Topic Maps, or Cyc. These prototypes should develop strategies and techniques for combining, translating, and mediating variant semantics across multiple contexts for the selected use cases, thereby forming the basis for future interoperability among eGovernment systems.