WALL-E and the "Many, Many" Maps: Toward User-Centred Ontologies for The National Map

Barbara S. Poore
US Geological Survey / Saint Petersburg / FL / USA

Abstract
The symbolism of the film WALL-E is employed to argue that at the heart of problems with ontologies and the "Semantic Web" lies a misunderstanding of the relationship between man and machine. Originally created to be processed by intelligent machines, top-down ontologies for the Semantic Web have so far proved inadequate to deal with the messiness and emergent properties of current developments in social networking on the Internet. Bottom-up ontologies that incorporate crowd-sourced folksonomies may allow local users to contribute to and enrich national mapping efforts.

Keywords: ontology, folksonomy, GIS, crowd-sourcing, geospatial Web, National Map

Résume
Le symbolisme du film Wall-E est utilisé pour affirmer qu’au cœur des problèmes sur les ontologies et le « Web sémantique » se trouve un quiproquo sur la relation entre l’homme et la machine. À l’origine, les ontologies descendantes avaient été créées pour être traitées par des machines intelligentes. Toutefois, jusqu’à présent, elles ne sont pas adéquates pour faire face au désordre et aux nouvelles propriétés associés au développement actuel du réseau social sur Internet. Les ontologies ascendantes qui incorporent des folksonomies approvisionnées par la foule permettent aux utilisateurs locaux de contribuer aux efforts de cartographie nationale et de les enrichir.

Mots clés : ontologie, folksonomie, SIG, approvisionnement par la foule, Web géospatial, National Map

“Our machines are disturbingly lively, and we ourselves frighteningly inert.”
—Donna Haraway, “A Cyborg Manifesto”

The Disney/Pixar film WALL-E can be seen as a parable of possible human–machine collaborations. WALL-E, the robot whose job is to compact the rubble left on the depopulated Earth of the future, becomes a de facto archaeologist, uncovering and examining the objects of our civilization. He finds a diamond ring in a jewellery box but discards the ring and keeps the box. He takes his favoured objects home each evening and stores them in a revolving industrial shelving system. The items seem to be placed in no discernible order, yet WALL-E is capable of instantly locating exactly what he needs. WALL-E’s
Ontology, or “the science of what is” (Smith 2003, 155), is a branch of philosophy that seeks to establish types of objects and relationships that exist in reality. The notion of ontology was abstracted from philosophy and used to construct knowledge-management tools by researchers in artificial intelligence and computer science. Because they promised to express meanings in collections of documents, ontologies were quickly picked up by other scientific communities as useful in online interdisciplinary collaborations to manage, share, and extract knowledge from semantically divergent data. For example, formal ontologies, “machine readable model[s] of the objects allowed into a formal universe and their associations or relationships between them upon which some automated reasoning tasks can be performed” (Schuurman and Leszczynski 2008, 189), have allowed biologists to make progress in genomic research.

There are six main dilemmas in current ontology research:

- Humans must specify the content and relationships for an ontology.
- Data rarely divulge the contexts of their creation; these contexts must be inferred, making knowledge transfer difficult (Schuurman 2008).
- Different ontologies have emerged from different

information communities, making interoperability even more challenging (Kolas, Hebeler, and Dean 2005).
- Ontologies can succeed only if they are widely used.
- Perhaps most importantly, the ontologies that are written for machines are not necessarily understandable to people (Gahegan and Pike 2006; Schuurman 2006, 2008).

Each of these ontology dilemmas might justify its own research strand, with differing methodologies. The dominant approach to ontologies in GIScience, summarized nicely by Pragya Agarwal (2005), has been to use cognitive techniques and predicate logic to specify and implement ontologies for data integration and interoperability (e.g., Bishr 1998; Bishr, Pundt, and Radwan 1999; Fonseca and others 2002; Kolas and others 2005).

To attend to the mismatch between human and machine ontologies, researchers have drawn on philosophies outside the logical positivist tradition, such as pragmatism (Gahegan and Pike 2006) and hermeneutics (Fonseca and Martin 2005). Most recently, qualitative, ethnographic methods have been recommended as a method of expanding metadata to give users a more complete rendition of the contexts of data creation (Schuurman 2006, 2008; Schuurman and Leszczynski 2006).

The latter approach inspired a debate between Agnieszka Leszczynski (2009a) and Jeremy Crampton (2009) over the ontological commitments of GIS critics. This debate gets to the heart of the method question. Leszczynski asserts that a post-structural critical approach ignores the material realities of GIS technologies by conflating epistemology (how something is known) with the ontic (or real, material being) – the bits, bytes, databases, and algorithms that constitute the existence of GIS. With Crampton, I accept that GIScientists must eventually provide implementable ontologies grounded in the material reality of GIS technology (Leszczynski 2009b; Schuurman 2006; Schuurman and Leszczynski 2006).

This stance is bolstered by media philosopher Friedrich Kittler, who lays out the history of philosophers’ neglect of the very media (writing) that performs their ideas: “If an ontology of media wishes to be informed by the technical state of the art, it should know how to read blueprints, layouts, mainboard designs, industrial roadmaps, and so on, in order to learn its very categories from scratch, namely from the hardware of high tech” (2009, 30). But as Crampton (2009, 607) points out, investigating GIS strictly as a set of digital objects does not account for the history of its emergence from mapping, the possibilities of being it affords or denies, or the practices of those who create and engage it. This brings us full circle to the idea of recursivity – that societies affect technologies as much as technologies affect societies (Chrisman 2005). Ontologists must thus be like WALL-E, digging out meaning using anthropological (Crampton 2009) or...
Ontologies and Spatial Data Infrastructures

Ontologies of one sort or another have always played a large role in digital versions of US Geological Survey (USGS) topographic maps. In the late 1980s, the USGS began converting the 55,000 topographic quadrangle maps of the United States to digital formats for use with GIS software. The topographic map is a base map that depicts features at or near the earth’s surface. A distinguishing characteristic of these maps is the use of contour lines to indicate elevation. Other elements mapped include roads, streams, rectified aerial photographs (orthophotography), and geographic names (USGS 2009). Topographic maps, which are in the public domain, have been widely used throughout their history for scientific purposes as well as for disaster search and rescue, flood risk assessment, census taking, land-use planning, resource management, transportation, and “hundreds of applications that contribute to every citizen’s daily life. Accurate and timely map data aid the hunter, the boater, the letter carrier, the schoolchild, the city manager, the law enforcement officer, the scientist, and the President” (NRC 2003, 20). Thus ordinary people have a use for and a stake in these maps. When topographic data began to be freely available online in the mid-1990s, they contributed to the establishment of a strong commercial geospatial industry in the United States (Lopez 1998).

In the conversion from paper to digital mapping, each theme or layer of the topographic map was digitized into a separate database. Common location could be used to position one digital layer on top of the other in a GIS, but issues of the identity and definition of mapped objects could prove problematic. For example, cars and trucks can drive over a bridge across a stream, but water cannot. Thus the ontology, in the information-technology sense, depended on the data creator’s and the user’s points of view. In the late 1980s, to solve some of these semantic issues, the USGS proposed a feature-based data model that would encode all the relationships, spatial as well as attributes, that digital objects such as streams might possess. This was a proto-ontology, and it was structured to provide different entry points into the data depending on the worldviews of users and their purposes (Rossmeisl and others 1988). Such a proto-ontology could have allowed for easy semantic integration among the different layers, but it was only partially implemented, and only for individual layers.

In the 1990s, the Federal Geographic Data Committee (FGDC) proposed a National Spatial Data Infrastructure (NSDI) for all federal agencies as an Internet-based mechanism for data sharing (FGDC 1993). The digital version of USGS’s topographic maps became The National Map (TNM), a Web portal that served digital topographic data and maps as part of the NSDI. TNM was based on the concept that state and local governments have the best data at each level, and thus data-sharing partnerships between the USGS and many government agencies were established for the creation and sharing of data at many levels (NRC 2003). The development of spatial data infrastructures, in which many different data sets from many different agencies had to be integrated, posed even greater obstacles to semantic data integration than integrating data within one agency’s databases. In the geospatial community, the Semantic Web was seen as a necessity, not just for the discovery of geospatial resources that might be housed in different servers on the Internet but also to allow semantic integration of different data (Fonseca and Sheth 2002).

Aside from issues of integration, ontologies for geographic information are difficult to craft. Geographic concepts can be indeterminate and fuzzy. Where do the walls of a canyon start and stop? Any ontology must support mapping, whether on the screen or on paper, but maps are scale-dependent; at what scale does a river need to be represented as double lines on a map, as opposed to a single line? Finally, there is the concept of changes over time. The USGS has historical maps of changes to the landscape covering a century and more; how can this change be represented by an ontology? In the federated data model envisioned by the NSDI, what happens when features are updated by one of the partners? Is a copy of the old feature kept or discarded, and does the ontology reflect this?

Unfortunately, the proposed solutions for a geospatial Semantic Web follow the generic Semantic Web in relying too heavily on the promise of intelligent machines. As Yasr Bishr and Werner Kuhn point out, “the semantic web technology of today rests on an illusory view of top-down semantic modeling through ontologies . . . idealistic views of semantics need to be complemented with social approaches to become useful” (2007, 384). Extending the parable of WALL-E to the Geospatial Web, this article suggests that, following the example of critical GIS (Schuurman 2006; Wilson and Poore 2009), a critical
An ontology approach that embraces the social is needed. A critical approach to ontology could take some of the same pathways that critical GIS has taken (O’Sullivan 2006), but a most effective direction would be to combine ontologies with an examination of the local practices of users as they struggle with the technology (Kwan 2002). The question of non-professional users is an important one. GIS is a technology that matters to many ordinary citizens, whether they are aware of it or not. Rescue from a hurricane, finding the nearest place to buy a hamburger, directions to a dinner party on a mobile phone— all these everyday activities and more are made possible by the use of geographic information in some type of machine configuration. As Dawn Wright and others (1997) have pointed out, the data in a GIS are useful in the world, perhaps more useful than other types of scientific data— therefore, these ontologies matter.

**Working with Machines**

In WALL-E — to return to the movie with which we started— the humans, having abandoned the dying Earth, float through the universe on a spaceship, confined to motorized lounge chairs and completely dependent on machines. Likewise, the initial vision of the Semantic Web presents end users as passive recipients, while lively software agents (machines) constantly scour the Web for items of interest (Berners-Lee and others 2001). But that is not the message of WALL-E. In the end, WALL-E and his girlfriend-robot, Eve, persuade the humans to get out of their chairs, seize control of the spaceship, and return to Earth. Earth is restored to life by active humans and machines working together.

It is clear that the Semantic Web and ontologies have not become a panacea, although certain aspects of Semantic Web technologies, such as the Resource Description Framework (RDF), are seeing wider adoption (Lasilla and Hendler 2007). Mark Gehegan and William Pike (2006) have developed operational ontologies with geospatial data that are responsive to different semantic understandings and offer views of project databases from different perspectives, much like the conceptual plans of the USGS discussed above, but these applications are largely aimed at professional scientists.

Some critics have argued that ontologies will never be an appropriate organizing paradigm for the messiness of the Internet because classification schemes are inevitably constrained by the social contexts of their creation (Shirky 2005). It may have been conceivable for top-down, expert-driven classification schemes for the Web to succeed in 2000, but the once-formidable barriers to posting data on the Internet have become radically lower, and the users have become hyperactive. Millions of people are blogging, Facebooking, wikiing, Flickring, tagging, and mashing up data, a phenomenon that has been dubbed “Web 2.0” by the business community (O’Reilly 2005). Web 2.0 puts users, many of them non-technical, at the centre of the Internet via a convergence of digital media, open-source software, the use of mobile devices, accessible application programming interfaces (API) for reusing and combining data, and the “long tail,” which makes it profitable for organizations to cater to extremely specialized interests. In recommending against ontology as an organizing principle for the Web, Clay Shirky (2005) cites this influx of active but amateur users: “If you’ve got a large, ill-defined corpus, if you’ve got naive users, if your cataloguers aren’t expert, if there’s no one to say authoritatively what’s going on, then ontology is going to be a bad strategy.”

Geographic applications have recently become prominent in the Web 2.0 environment, spawning numerous new start-ups and disrupting the business models of national mapping agencies and old-line geospatial technology vendors alike. Since the opening of the military-grade GPS signal to civilians in 2000, and Google’s publication of its Google Maps API in 2005, individuals and groups have made “many, many maps” online by georeferencing, annotating, and/or mashing up data streams from different sources to display on a map backdrop, typically from Google Maps or Google Earth (Tulloch 2007; see also Goodchild 2007; Haklay, Singleton, and Parker 2008). Entirely new maps of the world are being made by volunteers. For example, OpenStreetMap (OSM), a wiki-based, open-source editable map of the world, aims at nothing less than using volunteers to replicate the data traditionally provided by national mapping agencies, especially in areas of the world where mapping data are sparse or restricted by licence.

An ontology could be generated out of each of these many, many maps, but, as discussed above, integration problems would only multiply, as mappings between different ontologies would be required. User-generated data present an extreme case for ontology building. Ontology issues arise even between scientists from different disciplines or GIS specialists from different agencies, who at least share a common commitment to a scientific worldview or to software routines; user-generated geographic data, by contrast, are typically produced by non-geographers for a wide variety of reasons. Some maps are clearly intended to provide a public service— for example, the online maps that were generated to help connect people after Hurricane Katrina in 2005 (Singel 2005)but many map mashups are made for personal and recreational purposes (e.g., a beer map of the United States; Fairbrother 2005). In fact, the amateur status of users who dub themselves “neogeographers” is a badge of distinction, although many of these neogeographers are highly skilled computer programmers (Turner 2006; see also Goodchild 2007).

Because these maps are made by non-professionals, they might seem to have nothing to offer the professional map-
Now tags. Traditional GIS needs about 5 years to design the transport format for the ontology for the map. What I decided in OSM was that OSM should be like del.icio.us and allow people to just add attributes as they see fit. The traditional way would be to design an ontology and tell people something has to be a Road, a Motorway or a Railway. I decided on tags because I wanted OSM to be multi-national. How the hell was I going to design an ontology for the whole world? It would be impossible (or maybe with 5 years and a budget from the OGC) to do. So just let people tag. But really it wasn’t about my time or the technical problems of implementing ontologies vs. tags, it was about community. It’s the community, stupid. See if you let people design their own ontology and create processes to define it, like we have, you’re building a strong community where people interact, argue, debate and this beautiful thing comes out – a map. (Coast 2007, 1)

It might be argued that left to their own devices, users are undisciplined and frequently generate random, ambiguous, or useless tags, but researchers have found a regularity to collaborative tagging systems (Halpin, Robu, and Shepard 2007; Golder and Huberman 2005), and many computer scientists are experimenting with statistical methods to extract ontologies from crowdsourced data (Chen, Liu, and Qin 2004). Others have suggested aggregating folksonomies into ontologies by constraining them with controlled vocabularies such as gazetteers (Van Damme, Hepp, and Siorpaes 2007; Chen and Qin 2008) or by using topic maps. Topic maps – a standardized form of concept or mental map that is tied to user conceptions rather than to the information object – could mediate between subject identities that are meaningful to the user and the names for objects that appear in an ontology (Park and Cheyer 2006). A promising European project has suggested semantic annotation for geospatial data sets based on tagging and relevance feedback (Maué 2009.)

These technical solutions remain to be explored, but ontologies derived from folksonomies, if implemented, would be more responsive than top-down ontologies to emergent behaviour. Michael Curry (2000) contends that the descriptive and normative ways of building ontologies in the original Semantic Web do not allow the possibility of emergent properties and relationships. Emergence depends on the narrative construction of time; producing an ontology that allows for time requires using the tools of the social sciences and the humanities, with the understanding that no knowledge can be completely captured by any one method. Mei-Po Kwan (2008), among others, has demonstrated methods to introduce an ontology of time into GIS; similar methods should be explored for geospatial ontologies.

User-centred ontology would put the community first, before the ontology. Recent research recognizes that the practice of map-making is differentially engaged by different people in different locations and situations (Kitchin and Dodge 2007). These differences affect their views of the world and thus their ontologies (Brodaric and Gahegan 2001; Fonseca and Martin 2005). Users “perform” mapping bodily, as they move through space (DelCasino, Vincent, and Hannah 2006), while maps, in turn, structure users’ knowledge of urban space (Vertesi 2008). Users articulate different forms of knowledge and ontologies to bring their maps into alignment with the maps of others (Poore 2003); indigenous knowledge can be preserved by a blend of participatory GIS and geospatial ontologies (Sieber and Wellen 2007). The interaction, argument, and debate that, in Coast’s words, produce a “beautiful map” may also produce a richer ontology. Allowing users to form a community around contributing local data to a national map would also enrich that map. Ontologies don’t just happen. Developing, ensuring the adoption of, and sustaining any type of standard over time is a social process (Bowker and Star 1999). Complex negotiations are undertaken to develop metadata, data models, and ontologies for data sets; when the final data set is released to the world, however, these complex
negotiations – the social wrappers for the data – are lost (Schuurman 2008). Often, making full use of data for a geographic analysis depends on recovering the lost social wrappers through historical or ethnographic analysis (Comber, Fisher, and Wadsworth 2004). Emerging research on cyber-infrastructures shows that ontology development is a learning trajectory for achieving common ground among scientists from various disciplines (Ribes and Bowker 2009). Thus, paying attention to common users as important sources of information, not just in the requirements phase of ontology development but as the ontology is rolled out and adopted, could result in wider acceptance and use. As Gahegan and Pike point out, C.S. Peirce is sometimes credited with the observation that: “The textbook is the funereal urn of a discipline.” Perhaps the modern equivalent of the textbook for us is the computational ontology, for without the ability to evolve and adapt in response to conceptual changes, ontologies will effectively lock us into the “static, lifeless, purposeless world” (Sowa 2002) in which discovery, contested meaning and re-conceptualization are not facilitated. (2006, 747 n2)

By himself, WALL-E could disassemble the world but could not put it back together. But working with active humans, he could help create the world anew. Deriving an ontology from crowdsourced data is not a panacea, but it might open up new possibilities for a different kind of national map.

Acknowledgements

The author is grateful to Lynn Usery and Dalia Varanka for organizing the USGS/UCGIS meeting at which these ideas were first presented, and to three anonymous reviewers for suggestions that substantially improved the paper.

Author Information

Barbara S. Poore is a research geographer at the US Geological Survey, 600 4th Street South, Saint Petersburg, FL 33705 USA. E-mail: bspoore@usgs.gov. She specializes in the use of geographic information, volunteered geographic information, spatial data infrastructures, and data communities. She received her PhD in Geography from the University of Washington, and previously worked for the Federal Geographic Data Committee on the National Spatial Data Infrastructure.

Notes

1. For OpenStreetMap see http://www.openstreetmap.org/.

References


