B03.3

Thinking beyond the data: Creating knowledge from geospatial chaos

Judith Jerome, Information Services Manager, AGI

Abstract

This paper advocates data as simply a starting point. It is the transformation of data into knowledge that is important. This transformation is aided by appropriate technology but it is the visionary that sees this transformation as an intellectual process that goes beyond mere hardware and software. Such issues as the Semantic Web will be discussed along with the real meaning and value of metadata, thesauri, taxonomies, ontologies and persistent identifiers. In order for the GeoSpatial community to move beyond printing maps, it needs to internalise these essential processes.

So What's the Problem?

As members of AGI, we discuss the entire universe of geographic information in a rather take-for-granted way. The terms datasets, GIS and spatial this and that comprise the common parlance of our working lives. But if we have moved from being a data-poor society to a data-rich one, we have yet to move to a knowledge one. The problem has now become not finding the data, but filtering through large volumes of data to find geographic knowledge. And this is way more than funnelling vast quantities of data through GIS software and observing what comes out the other end.

We speak of information being a "valuable commodity", but that value lies only in its potential to be turned into knowledge. Ultimately, knowledge is what makes the difference in what we can do, and the value of information depends upon the value of the knowledge to which it can lead.

Being GIS experts, we toss around various data attributes with a high degree of confidence. Essentially, geographic information systems take the data and morphs it in such a way as to have substance with a certain structure. It can be studied mathematically using techniques such as spatial data mining, multi-media spatial data typing, and online analytic processing (OLAP). But real knowledge is information put into practice, or at least possessed in a form that makes it immediately available to be put into practice. In particular, knowledge requires a knower. The focus therefore becomes the user of the outputs, not the technology nor the bits and pieces of data.

Knowledge Information

Let's explore the differences between knowledge and information. Although the terms are often used interchangeably, they are not the same. Nor is information the same as data, though those two terms are also often confused. When we acquire data and fit it into an overall framework or previous acquired information, that data becomes information. To quote Peter Drucker, a popular management guru, information is "data endowed with relevance and purpose." Data becomes information when its creator adds meaning. We can therefore say:

Information = Data + Meaning

When a person internalises information to the degree that he or she is able, the information becomes knowledge.

Consultants Thomas Davenport and Laurence Prusak, authors of a seminal study called Working Knowledge (1998), define knowledge this way:

Knowledge is a fluid mix of framed experiences, values, contextual information, and expert insight that provides a framework for evaluating and incorporating new experiences and information. It originates and is applied in the minds of knowers. In organisations, it often becomes embedded not only in documents or repositories but also in organisational routines, processes, practices, and norms.

The resultant equation is:

Knowledge = Internalised Information + Ability to Utilise the Information

What Does that Mean in a Commercial World?

We hear those familiar clichés pertaining to 'the Knowledge Economy.' This or that practice, or law, or personnel implementation, or government funded program is necessary for the Knowledge Economy. We also hear about Information being the glue that knits organisations together. And yet usually the subtext of this glib statement is not Information as we defined it previously but information technologies or more precisely a collection of information technologies. The usual corporate response is: "If x, y and z isn't working appropriately, we need bigger, faster, better technology. This is similar to saying that the solution to London's traffic problems is the building of two or three more M25 like motorways. It's shortsighted and limited in focus.

The answer is of course an expert to manage a company's various information assets. Just as companies hire solicitors or legal experts to advise on contracts and similar legal matters, an information/knowledge manager is needed to handle a company's information assets.

What Does this Mean in the World of Geographic Information?

If you think about the human intermediation discussed above which results in the morphing of data into knowledge, you could easily theorise the answer. GIS is often defined by the vendor or the system from which it originates, and to be fair, it was the advent of these systems which promoted the world of geographic information beyond the simple map output. However, that map bias persists. The "create-a-map" uber-focus forces the drawing of a line or the placing of a point somewhere and the uncertainty or imprecision of that location is lost from view (even if it is carried in the metadata somewhere). This is a major knowledge inhibitor.

From the human problem solving point of view i.e. people who need answers to questions and need to deal with data that doesn't fit well within the GIS mathematical and geospatially-oriented structure, GIS information can often be very opaque. Two examples of these limitations involve analysing changes through time in boundaries and land use is one. Dealing with fuzzy data is another.

The answer lies with openly developed systems which allow tinkering outside of the traditional commercial marketplace. And of course the issue of open standards which will lead the Semantic Web beyond to concept to reality.

The Issue of Language: You Say Tomahto, I Say Tomaeto…..

What about the cryptic treatment of placenames in GIS, often treated as a simple table of labels for features which has limited usefulness beyond providing some label for the features of a GIS data layer. These labels can be something like 'pond'. Even when they are real placenames, there is only one name. The problem with this is that there are often alternative names for places a different spellings, in different languages etc.

If geographic information systems structured their feature labels as a gazetteer then there would be a link between the common way that people refer to places and the way that GIS represents them geospatially. Again see the focus on the 'person' needs and not the technology driven needs.

There are limits to gazetteers though. Gazetteers are woefully inadequate in geospatial footprints, that is, boundary polygons. The knowledge transformation here happens when the polygons of GIS can be linked to the toponymic and descriptive richness of gazetteers. Doing this presents problems. There are only a limited set of unique identifiers (again the issue of standards) for geographic names. Generally, you can't count on the name or the location or the type to uniquely identify a particular feature (e.g. multiple names, the same name for different places, different coordinates, etc.)

Onomies and Ologies: Taxo, Onto and otherwise

Intrinsically tied to the issues of language and gazetteers, is the issue of labelling hierarchies especially at the feature level. For instance, Linda Hill, University of California, Santa Barbara describes the following relationships in the Feature Type Thesaurus – Top Term Report:

Physiographic features ➩ Tectonic features ➩ Faults ➩ Rift zones

This is a very basic hierarchy starting at the most general level and refining to the specific. A linguistic hierarchy exists when the essential "language" of a discipline is extracted and related in a way which can universally describe that area. This is called a taxonomy. It is a language that can be universally accepted for the communication of that particular discipline.

We are familiar with the taxonomies of the biological sciences. Deriving from that model, a taxonomy is defined as:

`A system of language which derives from a common genus and are hierarchically represented according to their common differences. This genus to species nomenclature is often characterised by terminology which moves from generic terms to specific terms.

In the world of geospatial information, we often derive geo terms by referencing the discipline from which they derive. For instance, environmental sciences, archaeology, historical anthropology, transportation are all disciplines which are defined by a common language which also has a geographic attribute. When you take the universe, or, essentially ask the question "What is there?", an ontology is the answer to that question. An ontology reflects the commonly used and trusted breakdown of categories.

Ontologies typically organise taxonomies at a high level. Where the taxonomy shows a hierarchical relationships within a system, an ontology shows the relationships and connections between the taxonomies and is usually visually depicted in a more three dimensional graphic.

If we accept that a major driver for geographic information in a technologically structured world, is the sharing of knowledge, then sophisticated structures like ontologies are good candidates for respresenting and abstracting geographic data. In 1991, Nunes pointed out that the first step to build a next generation GIS would be the building of a systematic collection and specification of geographic entities, their properties and relations.[1]

Fonseca and Egenhofer of the National Center for Geographic Information and Analysis, University of Maine, suggest that the only way forward for interoperability and knowledge sharing in the world of GIS is via ontology construction and portability.[2]

In Conclusion

Metadata, which needs no explication, is one way for transforming information capital into knowledge. However, it is a baseline. Underlying issues of metadata creation are issues of linguistic adoption and of course....standards. It also involves acknowledgment and awareness on the part of the GIS community of the knowledge in the user community. GIS shouldn't be a specialist enterprise.

---

[1] Nunes, J., "Geographic Space as a Set of Concrete Geographical Entities", in *Cognitive and Linguistic Aspects of Geographic Space*, Kluwer Academic Publishers, 1991, pp. 9-33.

[2] Fonseca, Frederico T. and Max J. Egenhofer, "Knowledge Sharing in Geographic Information Systems", in: P. Scheuerman and F. Bastani (eds.) *Third IEEE International Knowledge and Data Engineering Exchange Workshop*, November, 1999.