Ann Reg Sci (1999) 33:213–232

The Annals of
Regional Science

Semantic interoperability: A central issue for sharing geographic information

Francis Harvey[1]*, Werner Kuhn[2], Hardy Pundt[2], Yaser Bishr[2], Catharina Riedemann[2]

[1] Department of Geography, University of Kentucky, Lexington, KY 40506-0027, USA (e-mail: fharvey@pop.uky.edu)
[2] Institute for Geoinformatics (IfGI), University of Münster, D-48149 Münster, Germany (e-mail: kuhn,hardy,bishr,riedema@ifgi.uni-muenster.de)

Abstract. Technical interoperability has provided geographic information communities with substantial improvements for constructing GIS capable of very low friction and dynamic data exchanges. These technical advances stand to provide substantial advantages for sharing geographic information, however reaping these advantages in highly heterogeneous operational and organizational environments requires the understanding and resolution of semantic differences. While the OpenGIS consortium has made important progress on technical interoperability, semantic interoperability still remains an unpassed hurdle for efforts to share geographic information across organizational and institutional boundaries at the local, regional, and other levels. Identifying and resolving semantic interoperability issues is especially pertinent for data sharing and considering future developments of standards. This paper presents an overview of semantic interoperability and through case studies shows the breadth and depth of issues and approaches in different countries and at different levels of organizations. These cases illustrate the importance of developing flexible approaches to practical data sharing problems that merge semantical with technical considerations. Based on our examinations of semantic issues and approaches in ongoing research projects, we propose cognitive, computer science, and socio-technical frameworks for examining semantic interoperability.

1. Semantic interoperability, standards, and data sharing

Interoperability is widely recognized as a new paradigm for joining heterogeneous computer systems into synergistic units that facilitate a more efficient use of geographic information resources. This is part of a more comprehensive enterprise-orientated view of information technology in general. Considerable

Please send all correspondence to F. Harvey
This work is supported by a grant from NCGIA's project Varenius. Portions of this paper were presented at the Interop '99 conference (Zurich, March 1999) and will appear in that proceedings volume.

research on technical issues helps to cross many of the barriers that made most information and GIS systems insular data (Bishr 1998). Although the goals of transparent data exchange and remote access have yet to be reached by technical interoperability, work on interoperability provides a basis for facilitating data sharing and helping resolve redundancy problems (McKee 1998). Recent initiatives in the United States and elsewhere underscore the importance of interoperability in data sharing. Efforts in the United States to build an infrastructure for Federal geographic information sharing are now being complemented by efforts to build a national, federal, state, and local infrastructure of geographic information (NSDI) (National Academy of Public Administration 1998). Data sharing is crucial, and technical interoperability will without doubt be significant in building a more dynamic geographic information infrastructure.

These technical opportunities warrant an examination of underlying questions in regards to organizational, institutional, and cultural issues. How will different agencies, organizations, institutions, and, finally, people, meaningfully use and share geographic information from multiple sources? What are the benefits of technical developments without applications? What costs are incurred when the technically feasible is impaired by unresolved organizational difficulties? We know that many information system projects don't work (Ewusi-Mensah 1997). Design and implementation stages, critical phases that meld the social and technical, are cited frequently as the origin of these breakdowns. Although the data sets might be exchangeable, roads in one agency can mean quite something different for another agency. How can the geographic information systems of two agencies with differing understanding and models of roads be made interoperable? What are the semantic differences that should be addressed in constructing data sharing environments and developing cross-standard exchange mechanisms? Data sharing depends on reconciling different meanings. This paper looks at the issues of transferring meaning (or semantics) on hand of three case studies and also sketches out work developing methods to understand and resolve them. The initial work we present here points towards viable directions for considering interoperability not only as a technological issue, but rather as a conflux of social and technical issues.

Data sharing needs to consider the different standards that are now in circulation as well. For some time, the call has gone out for viable standards that lead to a frictionless exchange of data between agencies, regardless of whether they are in the same county government building, or across the world in different nations. This has led to a cacophony of standards, which suggests, perhaps, that the more you want to get people to agree, the more differences will be found. Just as there is no single geographic reality that overrides all others (Nyerges 1991), there is also no singular standard that encompasses the different data models people use. The semantics of geographic phenomena are too broad for standardization. The key issue in standardization is finding ways to minimize information loss (Kuhn 1994).

Semantic interoperability goes beyond attempts to homogenize differences through standards. Accepting the diversity of geography and geographic information technologies, this approach seeks ways to navigate differences in meaning. A central question underpinning our discussions about semantic interoperability is how people and social groups with different perspectives identify and possibly resolve their semantic differences. The construction of information communities in heterogeneous institutional and disciplinary environments calls for frameworks to conceptualize and articulate these semantic differences. Semantic interoperability requires means to resolve complex differences that lurk behind apparently consensual terminology and procedures. This issue was frequently addressed as a crucial scientific issue at the recent Varenius I-20 Interoperability conference and workshop (NCGIA 1997; Yuan 1997).

Semantic problems will persist and hinder the development of interoperable solutions long after technical problems are solved. Several trends seem to work against a universal concept of one unique meaning for every geographic phenomena that technical interoperability and computer networking often seem to engender:

> Previous efforts to specify uniform standards in the GIS community have not met with great success, except where their adoption can be mandated. Even within a single agency there are difficulties in forcing compliance with standards.
>
> The fragmentation of the GIS software industry and increasing overlap with other forms of software such as CAD and DBMS has made it more difficult to promote uniformity. After many years of concerted effort, the failure to arrive at a consensus on a unifying theory of geographic information is frustrating. Because so much has been invested in research, data, and software already, it seems unlikely that a uniform theory could be successfully disseminated and accepted even if one could be found—if one were found, would the communities acknowledge it? Older systems of knowledge dissemination are becoming confused by the comparative ease of access to information on the Internet, the high prices of books and journals, and the budget problems faced by traditional libraries (NCGIA 1997).

Our research sets out to investigate the issues semantic interoperability poses in more depth and move beyond the past experiences with standards towards developing more flexible, non-normative approaches. An important starting point is geographic information communities (GIC) as described in the OpenGIS Abstract specifications (OGC 1998).

In design, implementation, data sharing, standardization, and interoperability, many common terms turn out to carry vastly divergent understandings of the same or related phenomena. Terms such as ‘wetland’ or ‘fallow’ pretend a certain shared understanding, that provides some social coherence but, which on inspection, turns into complex and fragile arrangements of different semantical models and intricate institutional agreements. Our initial research activities focus on understanding what constitutes semantic differences and how other researchers set out to understand and resolve them. As we will show in the following Section 2, when different standards are competing, the chance to develop interoperable solutions is often minimal, though the need is particularly strong.

2. Examples of semantic interoperability issues

This section documents the case studies we have investigated. Each case is distinct and is approached differently. The first case, roads in Europe, stands out for its emphasis on finding mappings between divergent meanings as a basis for data sharing in an interoperable environment. The study of fallow lands in Europe points to the influence institutional mandates assert on the semantics, even when the phenomena in question are known under a common name in identifying semantic differences. The difficulties in negotiating a mutually acceptable definition of wetlands point to the problems finding viable solutions between agencies. The last case study connects all these issues and illustrates how difficult particular understandings make it for legal procedures to be implemented.

2.1 What are roads?

Europe has a vast and extensive ground and water transportation network. Several public and private organizations deal with transportation information, e.g., suppliers of data for car navigation systems, logistics transportation, and traffic control, management, and analysis. These agencies usually require transportation information that stretches beyond national borders. For example, traffic management and control agencies sometimes require transportation information collected by mapping agencies.

There are several efforts to standardize transportation definitions and classification, e.g., ATKIS and GDF. Developed between 1985 and 1989, the Official Topographic-Cartographic Information System (Amtliches Topographisch-Kartographisches Informationsystem) of the Federal Republic of Germany, ATKIS, is a topographic and cartographic model of reality.

Geographic Data Files, GDF, are a European standard released in October 1988 that was updated several times before 1995 (ERTICO 1998). It aims to provide a reference data structure for describing road networks for car navigation, vehicle routing, traffic analysis and other applications. It has been created in order to improve efficiency in the capture and handling of data for geographic information industry.

The objectives of each standard differ because of differences in the cultural settings of each standard. These differences correspond to social groups called geographic information communities (GICs'). In this case we consider a German topographic GIC and a pan European traffic management GIC. Different constellations of agencies and institutions may belong to multiple groups. Both respectively take advantage of the ATKIS and GDF standards. We call them here for convenience, ATKIS GIC and GDF GIC, respectively. The ATKIS GIC conceptualizes transportation networks as artifacts that are part of landscapes, which are presented in topographic maps. The GDF GIC conceptualizes transportation networks as a section of the earth, which is designed for, or the result of *any* vehicular movement. Each GIC has a distinct point of view – not only of their own data, but also of the point in exchanging data. Roads are not always roads (ATKIS/GDF comparison) From the GDF GIC point of view, the main purposes of a connection between their information system and the ATKIS information system are to provide the most recent and up to date information about new roads and status, e.g., to provide an online service for car navigation systems.

From the ATKIS point of view, the main purposes of a connection between their information system and the GDF information system, is to take advantage of the GDF's traffic flow information and routing information, and provide it for local applications that adopt ATKIS as their base model and require more information about the traffic flow, direction, rules, etc.

Any consideration of this problem may start by asking the question "Does transportation network mean the same thing in the two GICs?"

In GDF, the term "road" encompasses roads, railways, waterways, junctions, rail junctions and water junctions, while in ATKIS waterways are not considered a part of "roads." In the ATKIS GIC roads refer solely to ground transportation networks. A road element is the smallest part of a road that has a consistent width, e.g., does not change within a certain threshold. In GDF a road network also encompasses ferry connections which are not implied in ATKIS. A road element does not only depend on its width but also on traffic rules in GDF. For example, a new road element will be created in GDF if the direction of flow changes. In ATKIS this would be just one road element.

Even the term "ferry networks" in ATKIS refers only to ferryboats, while in GDF a ferry is a vehicle transport facility between two fixed locations on the road network and which uses a prescribed mode of transport, for example, ship or train. Considering the ground transportation road network, we find that ATKIS includes pedestrian zones and bike paths as part of a road feature, while in GDF, a pedestrian zone is not part of "roads" and a "bike road" is a type of a road network.

Figure 1 illustrates these differences by showing a hypothetical road. Baker Street is a two-direction street. In ATKIS it is viewed as one road element that has two intersection points. In GDF the same road is presented as two road elements, one for each direction of traffic flow. If you ask ATKIS GIC about Backer Street you will get one road as shown in Fig. 1b. If you ask the same question to the GDF GIC you will get two roads for Backer Street, one for each traffic flow direction (Fig. 1c).

2.2 *What are fallow lands?*

Land use classifications are especially prone to differing semantics of assumed similar land use types. Fallow land (often called "set-aside land" as well)

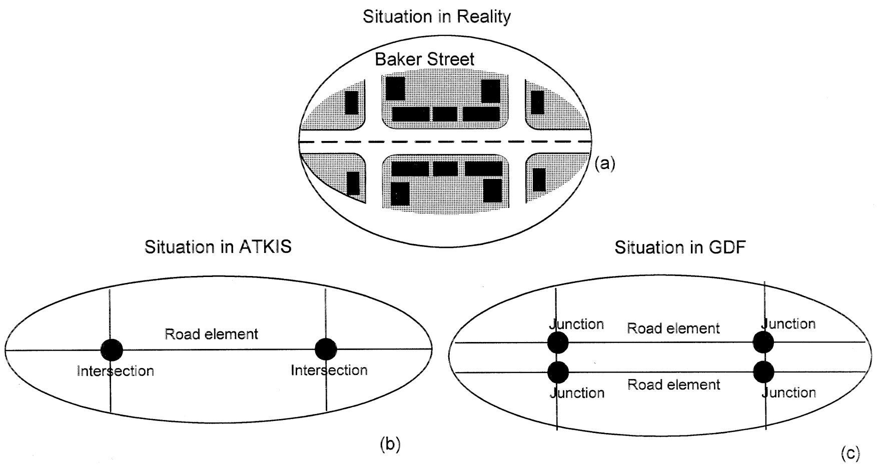

Fig. 1. a–c Roads are not always roads (ATKIS/GDF comparison)

represents a special kind of land use subject to quite unique interpretations. Which land parcels in certain areas are assigned to the class "fallow land" differs between different geospatial information communities (GICs). Differences depend on the rules and mandates by which parcels are designated. Such rules are defined by specific GICs such as Nature Protection, Agriculture or Landscape Ecology.

In Germany, land use information is provided in the official topographic and cartographic information system of Germany (ATKIS 1995). ATKIS is based on an object catalogue where every geo-object is defined by several attributes. Objects and attributes have identifiers. "Fallow land", for example, is defined as object number 4110 in the ATKIS object catalogue.

Another GIC that is interested in information about "fallow land" is agriculture. For the purpose of paying subsidies to farmers, the German National Chambers of Agriculture need information about the sizes of parcels which are declared as "set-aside" or "fallow land" (Große-Enking 1994; Wehland 1994). However, the rules, defined within the chamber of agriculture GIC, to determine which parcels are "fallow land" are different from the ATKIS rules.

Other GICs, e.g. Nature Protection, are seriously interested in information about "fallow land" within the framework of specific environmental projects. Many national nature protection laws define the term "fallow land" clearly, and yet differently (Klein et al. 1997).

As a consequence of the different semantics, it would not be possible to use ATKIS data for nature protection projects, and even the nature protection GIC could not share data with the agriculture GIC without considerable additional work. This case study represents a typical example of the limitations varying data semantics cause for interoperability between GICs.

In North Rhine-Westphalia (Germany) two agencies recently merged because of political and financial reasons. The agriculture agency was merged with the agency for ecology, landscape development and forestry. This merging process gave rise to the different types of semantic heterogeneity not across GICs, but also within the same GIC. Although, it may now be one agency, the process of constructing a single GIC did not automatically follow this administrative action.

2.3 What are wetlands?

Data sharing encounters semantic issues head on in ways that frequently go beyond technical and organizational issues and turns them into political issues. This is especially true for environmental phenomena that involve property rights. In the United States a very contentious category is "wetlands" (Shapiro 1995). "Wetlands" mean different things to different agencies and individuals. The apparently consensually agreed to term wetlands may conceal fundamental semantic differences.

At the core of these debates lies the question, "What exactly is a wetland?" The contentions surrounding Federal attempts to proclaim a wetland classification scheme, the Cowardin system developed by the U.S. Fish and Wildlife Service 20 years ago (Cowardin et al. 1979) as a "standard" for use by the U.S. federal government (a process completed in 1997) illustrate well the intricate relationships between technology and institutions.

At least six agencies are involved in wetlands mapping activities through over 19 programs and projects (Department of the Interior & Fish and Wildlife Service, 1990). The agencies include the Fish and Wildlife Service (FWS), Geological Service (USGS), National Oceanic and Atmospheric Administration (NOAA), Natural Resources Conservation Service (NRCS), and Corps of Engineers (CE). The programs range from activities required for hazardous waste clean-ups (Superfund), inventories of estuaries, flood control, to various regional-federal-state cooperative programs with multiple goals. As President Clinton's executive order leading to the organization of the Federal Geographic Data Center (FGDC) pointed out, it is necessary to improve the coordination of these various activities.

Wetlands have seen much FGDC activity following this executive order with the specific Clinton policy goal to reconcile and integrate all federal agency wetland inventory activities (Shapiro 1995). A Wetlands subcommittee specifically targeted this issue. To estimate the extent of the problems resulting from multiple definitions and establish ways to resolve them, comparisons of different wetland mapping techniques used in the Federal government were made for an area in Maryland close to Washington DC. With the agenda to reconcile differences and integrate results, this report comes to the conclusion that the four data sets compared "disagree in more than 90% of the area that at least one of the four data sets delineates as wetland" (Shapiro 1995, p. xiii). The report cites various reasons for this extreme disagreement, but acknowledges that even if the areas were extended by 50 m in every direction (buffered in GIS) the disagreement is still 60%. The maps included in the report demonstrate that the social agreement about definitions does not result in actually delineating the same areas on the landscape.

In conjunction with this analysis there was an attempt to standardize wetlands definitions and classifications, using the Fish and Wildlife Service Cowardin Classification Methods (Cowardin et al. 1979). This attempt met contention, particularly from the US Corps of Engineers (CE), who was apparently not consulted by the FGDC. The CE has developed another methodology and found substantial differences to the wetlands delineated by the Cowardin classification primarily because of the ecosystems orientation and delineation of a wetland when just one positive wetland indicator is present for any parameter (vegetation, soils, and hydrology). The guidelines used by the CE require all three parameters (Federal Geographic Data Committee & Wetlands Subcommittee 1997).

For groups with different mandates and semantics, such as the Corps of Engineers, the Cowardin classification is simply not a standard. As the proponents from the Fish and Wildlife Service are forced to acknowledge, the Cowardin system does not supersede existing law or agency policy, nor, most importantly, "Application of the standard is not regulatory" (Federal Geographic Data Committee & Wetlands Subcommittee 1997). Other federal agencies embrace the Cowardin system, but the Corps of Engineers could resist this definition. In contrast to the Fish and Wildlife Service, the Corps of Engineers characterize wetlands by combinations of vegetation, soils, and hydrology aspects, not any single characteristic in isolation.

2.4 Noise abatement from sports grounds

Sport activities lead to noise. These noise emissions are addressed in the German "Lärmminderungsplanung" (noise abatement planning), an administrative procedure dealing with multiple noise emissions coming from different sources, e.g. traffic, industry and sports. When assessing the noise emitted from sports grounds, the understanding of terms plays an important role.

Noise abatement planning is designed as a small-scale instrument for urban areas. Only those sports grounds that contribute significantly to the noise situation are examined in detail. Consequently, as a first step the following question must be answered: "Which sports grounds qualify as relevant to noise abatement planning?" The term to be examined in this context is relevant sports ground. When the selected relevant sports grounds are looked at in detail, it is sensible to use existing information if it meets acoustic requirements. This leads to the question: "What data sets are used in modeling sports grounds for noise abatement planning?" In other words: "Do data sets exist which have the same semantic understanding of a sports ground as an acoustic expert?"

The local administration is responsible for noise abatement planning; thus instructions contained in the law must first be taken into consideration in determining which data sets are relevant. Further, it is useful to have a look at how such a task is executed, because this is done by people who did not create the laws and because it usually needs refinement and interpretation to put legal instructions into practice. The following Table 1 shows how laws and other instructions in the federal state of North Rhine-Westphalia consider sports grounds in noise abatement planning.

The table reveals that there is only one definition of sports grounds for noise abatement purposes. The world of noise assessment seems to have a common basic understanding of the term sports ground. However, differences soon show up. As the actual case becomes more practical, more criteria are found and they become much more precise. In addition, an important change in criteria surfaces: the law refers to size (certainly taking it as a proxy for aspects of usage), but practice refers mainly to usage. Size is not mentioned, but instead refers to several more precise aspects of usage. A large number of criteria with high precision might be useful for understanding, but can make information access difficult. For a sports ground without scheduled use, nobody keeps a record of the hours of usage and the number of present people. In this case it becomes necessary to define fewer criteria and dispense with accuracy. Still, the same attributes have to be included. Probability of usage instead of exact training and competition hours would be considered in this case (see table, rows 5 and 6).

Obviously different answers to the question "Which sports grounds qualify as relevant to noise abatement planning?" will produce different results in emission assessment. The administration must avoid this as much as possible. Further, there is a point of communication. If you have to ask other authorities for data, which happens regularly with noise abatement planning, you must be able to give a precise description of what you want. This requires that you have a clear set of requirements and be able to express them adequately for people outside of your discipline. If you ask for a list of all larger sports grounds and take the law literally, you might be presented with many sports grounds that are not used regularly and are consequently of no interest. On the other hand important smaller sports grounds being used regularly will not appear.

The "Sportzentrum Roxel", a sports complex in Münster, was chosen as a case study to illustrate different views of sports grounds. It is a larger sports

Table 1. Sports grounds in North Rhine-Westphalian noise abatement planning

Regulation and information source	Aim, contents, definition of sports grounds	Criteria for relevance of sports grounds
1. Federal law for protection against noise impact (§47a BImSchG)	introduces noise abatement planning; does not mention special noise sources like sports grounds; no definition	no criteria mentioned
2. Instructions concerning the federal law for administration; model (LAI 1992) and version for North Rhine-Westphalia (VwV NRW)	makes instructions by law more concrete and practicable; mentions sports grounds as noise sources; no definition	• larger sports grounds
3. Instruction for protection against the noise of sports grounds (18. BImSchV)	defines a uniform method for the assessment of noise impact caused by sports grounds; assessment procedure; stationary facilities intended for doing sports, including facilities that have close proximity, spatially and operationally	no criteria mentioned
4. North Rhine-Westphalian guide for making noise impact plans (Hillen 1993)	refines legal instructions to put them into practice; no definition	• competitions • preparation of competitions • run by municipalities, clubs, enterprises • considerable noise emission examples: football fields with more than 200 spectators, tennis complex with more than 3 courts
5. Practice, version for sports grounds with scheduled use	carries out noise impact assessment	• outdoors • ball games like soccer or tennis • regular usage • used by clubs • competitions on Sundays between 1 p.m. and 3 p.m. • many spectators • training after 8 p.m. • usage of loud-speakers • residential buildings in the neighborhood (within a radius of 200 m in case of a usage before 8 p.m. or after 10 p.m., otherwise within a radius of 100 m)
6. Practice, version for sports grounds without scheduled use	carries out noise impact assessment	• outdoors • noisy forms of sports • acceptance by population • high probability of usage after 8 p.m. • much usage between 8 a.m. and 8 p.m. (counted in 25, 50 or 75%)

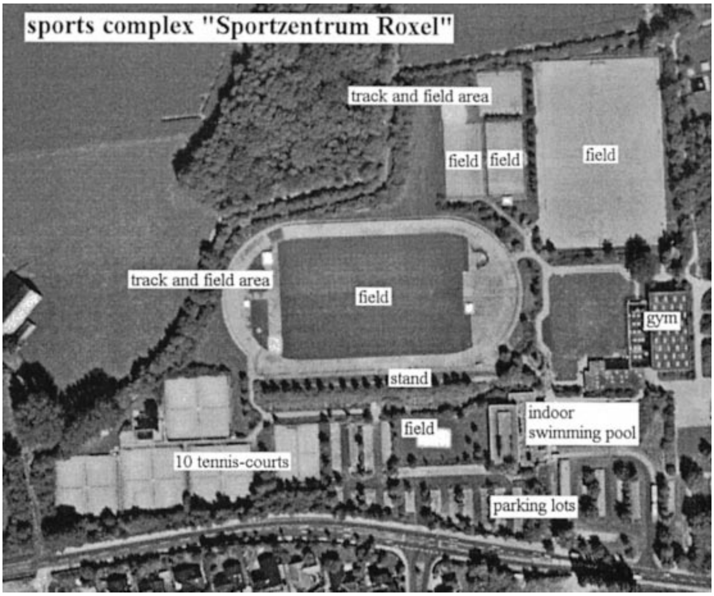

Fig. 2 Aerial photograph of "Sportzentrum Roxel"

ground comprising several playing-fields of various sizes dedicated to different forms of sport, a stand for spectators and parking lots which, according to a legal definition (18. BImSchV), belong to the sports ground. The following descriptions only take into account spatial characteristics and leave out attribute issues.

The aerial photograph (Fig. 2) is meant to give an "objective" view of the sports ground, as far as this is possible.

There is a large playing field in the center of the complex. This is the main field, where soccer is mostly played. It is surrounded by track and field areas, and, south of it, a stand for spectators adjoins. In the northeast another large field is located, which is also mainly used for soccer. Between these fields you can see two smaller multipurpose fields (for basketball, handball, and volleyball), which – in contrast to the large fields – are free to be used by everyone and are not subject to any schedule. North of them, there is another track and field area. Ten tennis-courts are situated in the southwest corner and east of them, parking lots can be seen. A beach volleyball field lies next to the parking lots, and east of it there is an indoor swimming-pool. The building at the eastern edge is a gymnasium.

Figure 3 shows how an expert in acoustics models the noise immissions from this sport ground. You find only those parts which are important from an acoustic point of view, the sources of noise, are included. Noise can be emitted by players, spectators and cars. But some of these areas, where these sources originate, do not appear in the model due to a low frequency of usage, the form of sports, or because they represent indoor facilities. They are irrelevant for noise abatement planning. The two large fields, the tennis-courts, the stand, and the parking lots are relevant sources that must be looked at.

In Fig. 4, cadastral data ("Automatisierte Liegenschaftskarte", ALK) of

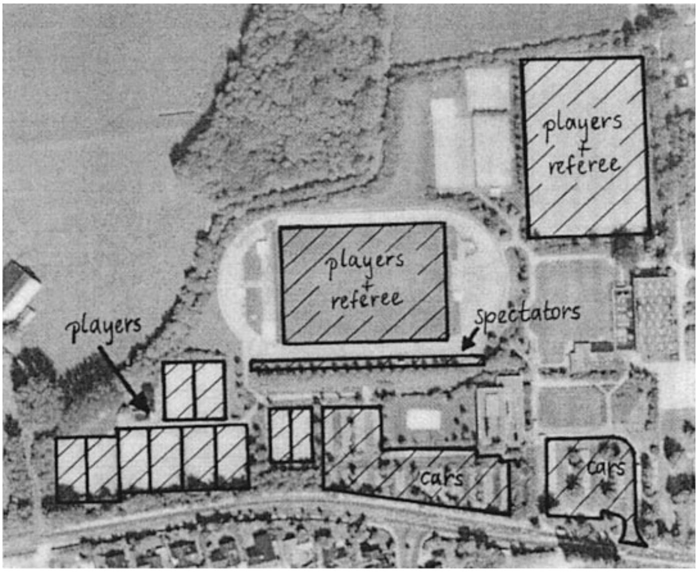

Fig. 3. An acoustic model of the "Sportzentrum Roxel"

the city of Münster are shown. Cadastral data are meant to provide information about location, shape and size of parcels as a basis for property documentation and taxation.

From the cadastral perspective, sports grounds are not an object of primary interest. They belong to the supplementary topography, which is not registered systematically. This is why only few parts of the sports complex appear. There are three objects, each of which represents a generalization of two tennis-courts. In contrast to that, the parking lots are modeled in detail. The data model contains in addition objects for small and large fields, which is why we can expect extensions of the current map at some time, but no object for a stand.

The intention of topographic data is to show the surface of the earth and the objects on it. Depending on scale and purpose, the modeling of real world objects in topographic maps varies. Figure 4 contains two examples for this kind of data. On the one hand, a part of the German base map 1 : 5000 ("Deutsche Grundkarte 1 : 5000", DGK5, here not shown to scale) appears in black lines. This map is available digitally only in raster format. On the other hand, digital vector data of the Authoritative Topographic-Cartographic Information System ("ATKIS") are shown in gray. The contents of ATKIS is comparable to a topographic map with a scale of 1 : 25000.

In the DGK 5 *(see Musterblatt DGK 5 1983)*, all fields are represented, but there are generalizations: the playing-field in the oval area is not a separate object, and two adjacent tennis-courts are represented by one object. The parking lots are generalized as well by depicting just their outlines. Since the outlines are not closed, they merge with the street to one object. The DGK5 does not have an object "stand".

In ATKIS *(see ATKIS-OK 1995)*, the whole sports complex is depicted by one large object. ATKIS also comprises objects like playing-field, stand and

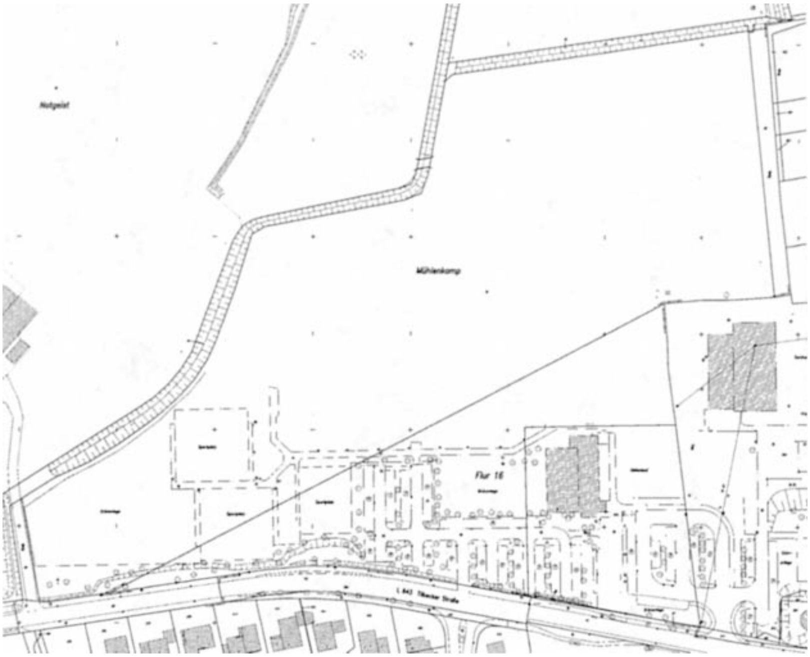

Fig. 4. "Sportzentrum Roxel" in ALK

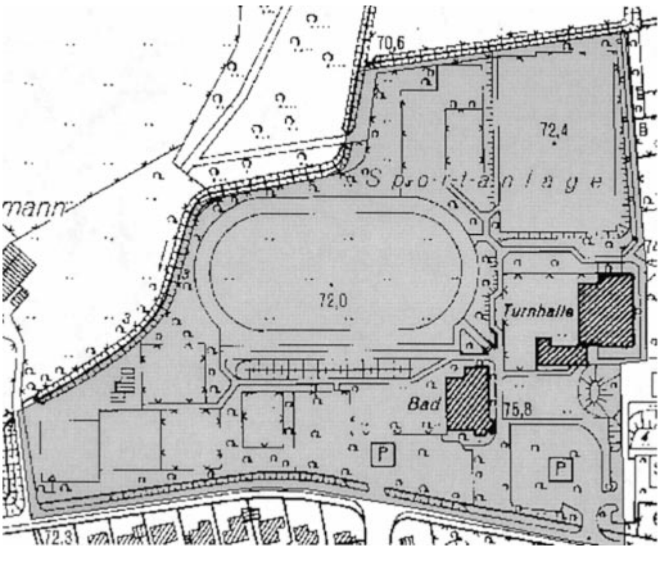

Fig. 5. "Sportzentrum Roxel" in DGK 5 and ATKIS

parking lot, which could show more details, but they have not been included here yet. In the future, there will be a more detailed model of the sports complex.

The acoustic model used in noise abatement planning requires the existence of certain objects at a certain level of generalization. The large fields, e.g., must be represented by rectangular objects, which means that the generalized oval object in the DGK5 is not acceptable. The parking lots just require outlines. In this respect the generalization of the DGK5 is better than the detailed cadastral data. For the consideration of spectators, which is only necessary for the large fields and not for tennis-courts, there are several possibilities. The easiest way is to assume spectators on the field together with the players and the referee. When the number of spectators increases to more than 500, more precision is needed and consequently geometry must change. Either the spectator's stand must be introduced as an additional object, or the rectangle of the field must be enlarged by several meters on the long sides. This will somewhat improve results near the field.

All data sets show weaknesses for noise abatement planning. The cadastral data (Fig. 3), on the one hand, misses many objects, whereas parking lots are depicted in too much detail. The DGK5 and ATKIS data models (Fig. 4) are not detailed enough regarding the playing field. ATKIS lacks any useful objects at this time, although it is the only model that provides an object for the spectators stand.

All data models show that this is an area dedicated to sports, but each data set has its own perspective differing from the emissions legal perspective. Consequently each provides, at best, part of what is needed for emission planning. Currently, emissions studies rely on existing data that are often digitized again, or approximative data are used, although often more appropriate data exist. This occurs because users are not presented with exactly the data they need and because it is often too complicated to build the needed data out of existing data sets. Support for transferring attributes and feature semantics would help improve this situation. This would mean, on the one hand, improving the use of the available data and, on the other hand, giving access to the most appropriate data. To get an appropriate digital acoustic model in the future, parts of future cadastral and future ATKIS data together can be taken as a basis and completed by digitizing supplementary information out of the DGK 5.

3. Semantic interoperability issues

Semantic interoperability, as the above cases illustrate, must account for a vast range of issues and approaches to resolve situations with complicated histories. Standardization aids in documenting and reducing differences, but cross-standard data sharing issues always remain. Clearly, each implementation may need to combine data in different ways. The development of an interoperable GIS for a telecommunications company and utilities company may require only a limited range of transfer points between GICs. The organizational requirements may be well known and the technical issues may be readily identifiable for the required process automation. Interoperable GIS for global change research may apparently present the opposite end of the spectrum, but even something as apparently mundane as data sharing in a county GIS, may be unsolvable if standards are used to entrench and construct data castles. Semantic interoperability is a way to avoid administrative data wars and improve access by linking different groups in discussions about substantive technical issues (Harvey 1997b).

Technical issues remain crucial in delimiting the resources needed to develop viable interoperable solutions, but should be examined in light of semantic issues. First of all, based on our preliminary findings, it is important to distinguish between open and closed systems. This differentiation refers to the degree in which the information systems in question are canonically defined and the system components documented. The information system of a nuclear power plant is a closed system that eliminates outside factors (or controls them) in order to provide exact control over system functions. An open system, is, as the name indicates, not well defined, not rigorously documented and subject to change. Most GIS used in public administrations probably fall into this category.

Clearly, closed systems with their formalized descriptions and mathematical rigor are ideally suited for intra-system interoperable solutions, but for semantic interoperability between different closed systems (Bishr 1997, 1998; Harvey 1997b). Semantic interoperable solutions for open systems (which can also involve different systems) call for a much wider range of considerations that we will discuss in the following section.

An important issue in determining semantic issues is assessing different viewpoints and perspectives. Understanding each groups semantics is more fundamental and a more viable solution than drafting canonical data descriptions and seeking to enforce them through regulations, that may, in fact, lead to more trench digging than otherwise. This is also an issue we discuss further in the next section.

Finally, resolving semantic differences calls for strong technical and social prowess in mitigating differences and presenting robust technical solutions. How different groups understand geographic phenomena may be central to their institutional role and possibly even the broader social issues. Any National Mapping Agency has a long-standing definition of roads, often chiseled in deeper in bureaucratic stone than any standard. It will be far harder to alter their semantic model than developing a translation package to exchange data between their data model and others.

The cases above show how technical and organizational issues are enmeshed. Resolving semantic differences requires that trenchant differences be resolved in a holistic fashion. The two European road databases need to be understood in the context of the purposes for which they were constructed. Fallow lands designation is connected to an institutional mandate. Wetlands are intricately part of agency agendas. The meaning of data comes through use. The semantics of interoperability and data sharing presents a means to address these issues and get to the differences between groups.

4. Frameworks for semantic interoperability and data sharing

Although semantic issues are very broad, there has already been quite an amount of research in computer science, cognitive science, and anthropological linguistics that we should consider. We see this as a first phase in a broader research effort that sets out to describe formal and computational models for semantic interoperability. We are still in the process of considering this work for semantic interoperability, but we want to document the frameworks we have been investigating, and point out our initial and expected results in the following sections.

Our research is considering three frameworks that we will refer to with the labels cognitive, computer scientific, and linguistic anthropological. Obviously, each of these three will have a distinct emphasis that we believe are not only interesting and pertinent to examine in their own right, but even more so for eventual synergies between different frameworks. The rest of this section briefly presents each framework. We defer any discussion about synergies to the following, and final, section of this paper.

4.1 Cognitive framework

The cognitive framework we are examining focuses on the work of Fauconnier and metaphorical mapping (Fauconnier 1994, 1997; Lakoff 1997). In Fauconnier's concept of mental spaces, developed in the 1980s, the mind creates multiple cognitive "spaces" to mediate its understanding of relations and activities in the world. We are looking at Lakoff's work on metaphorical mappings and Fauconnier's on more general mappings among conceptual domains which may well combine into an outstanding way to analyze semantic differences and break-them down into integral components which could well become the base for more formal and rigorous models of semantic interoperability. The essence of both these cognitive approaches are *partial* mappings from multiple sources structuring a target concept. Many of the semantic differences we encounter in our case studies appear to result from such partial structuring. Also, such mappings are amenable to mathematically rigorous formalization and implementations, thus allowing for experimental testing of semantic models (Frank and Raubal, 1998). This work could complement computer science work, although the process of formalization has yet to be tested.

4.2 Computer science framework

Substantial work on semantics as been published in computer science, which provides us with the strongest starting point for considering how to resolve semantic differences. At the present, we are perhaps furthest in building on Amit Sheth's computer science approach which is quite developed. Sheth's rich publications on interoperable computer system semantics provide us with a framework we don't feel to have yet exhausted (Kashyap and Sheth, 1996, 1997; Sheth 1996, 1997; Sheth and Gala, 1989). Semantics for Sheth need to be assessed in terms of the context. The concept of semantic proximity refers to an abstraction or mapping between the domains of two objects. Establishing similarities calls for comparing the intensional (contextual) descriptions of the two objects, described in a description logic language that links the semantic and schematic level. Conceptually, semantic integration in this approach consists of two phases. In the first phase objects are identified in different databases that are conceptually similar. In the second phase, the semantic differences are resolved between semantically related objects (Kashyap and Sheth, 1996).

Sheth's approach provides the most rigorous framework we have found to date for dealing with semantic differences in a practical, engineering-orientated approach. We have also begun to investigate the work of other computer scientists. The procedures outlined for dealing with partial incompatibilities between objects and domains are very detailed and beyond the focus of this paper to deal with in a just manner. This overview will hopefully serve as a suitable introduction to what we feel is a very promising approach to addressing semantic differences.

The concept in this approach is semantic proximity, which calls for declarative language to articulate the definitions of objects, and very strong ontological definitions. Semantics is acknowledged to consist of more than this, and thus, this approach needs to be extended by a broader consideration of context, rather than just database ontologies and declarative descriptions.

According to Sheth and Kashyap semantics involves vocabulary, content and structure (Sheth 1996). Reporting on a workshop, they point out there is no clear definition of semantics among participants, but academics consistently refer to semantics as the similarities between objects, relationships, and context. If semantics are cultural agreements between independent agents observing the real world, then we expect that illuminating insights will come from the examination of the group processes that lead to 'accepted' understandings, and the role of language as the most fundamental way of finding and assuring agreement.

4.3 Linguistic framework

Anthropologists, sociologists, and linguists have also examined semantics along these lines and have developed some very interesting insights that provoke us to think beyond the normal bounds of computer systems when considering semantic interoperability. Their work is wholly outside the bounds of computer science, which, although a limitation, does not diminish its relevance for any activity we engage in involving language and social groups.

We have been examining approaches that consider the processes social groups engage in to assure collaborative action. Bruno Latour's well know works examine the historical process of consolidating power across divergent interest groups, and the role of artifacts in enabling, effecting, and replacing human action – thus, acting themselves (Latour 1987, 1992, 1993, 1996; Latour and Bastide 1986). While most of this work is not directly pertinent to our project, its relevancy is expressed in the related work of lesser know, but more empirical work that evaluates the processes of constructing information technology in heterogeneous settings.

Barry's and Callon's work, for instance, points to the relevance of multiple groups finding their interests supported by the information technology under development (Barry 1997; Callon et al. 1986; Coyne 1995). Without it, ultimately, a group's opposition leads to its failure. The role of technology has been more extensively examined by other authors (Neumann and Star 1996; Star 1995a,b; Star and Griesemer 1989), which has also been recently applied to GIS and geography (Harvey 1997ab). This work corresponds to similarity orientated work on participative design that clarifies the importance of involving different groups in articulating their differences in order to find robust

Table 2. Components and brief descriptions of semantic proximity

Context representation	partial or full representation, structural and semantic components
Semantic taxonomy	semantic similarity is a qualitative measure to distinguish semantic relationships
Semantic heterogeneities in Multidatabases	Synonyms, data representation conflicts, scaling, precision all need to be dealt with
Data value incomparability	State differences, time lags, etc
Abstraction level incomparability	Different generalizations
Schematic discrepancies	Attributation conflicts
Structural similarity	Representation of structural similarities

solutions to semantical, social, and political problems of information systems design (Suchman 1987).

These points are all the pertinent when we bring linguistic anthropology's findings regarding the process of language development to bear on the issue of semantic interoperability. Many apparently simple problems of different meanings are intricately tied together with the specific language used by a cultural group. This is perfectly obvious when referring to different national languages (German and English for instance) which are perfectly unable to translate one-to-one common phrases into the other language. *Versteht Ihr mich?* has no direct equivalent in English. Translated as *Do you understand me?*, it loses the indication through the pronoun *Ihr* (second person plural) in German that the question is directed to more than one person, that the speaker feels he/she can informally address.

These issues in dealing with one shared language (if English should be counted as such, its regional and disciplinary differences provide the best counterpoints) are more subtle, but multiple understandings can be associated with terms used by disciplines in different ways. Such common terms used for geographic information are subject to myriad descriptions, disciplinary definitions, and regional variations. The linguistic methods for studying differences may be equally applicable for analytically determining crucial semantic differences (Holland and Quinn, 1987a,b; Salzmann 1993).

We believe these frameworks can each aid us find valuable insights into identifying and resolving semantic differences which can be crippling for data sharing and impair the successful dissemination and use of standards. Our work in progress specifically targets these two issues with the aim of making scientific and research contributions.

5. Expected results and future research

This paper provides an initial literature review of relevant frameworks for considering semantic interoperability and resolving cross-standard issues and going beyond lexical standards. We find there is ample evidence already that semantic issues underlie many of the difficult questions surrounding GIS data sharing and that as we conduct research on these questions we should strive to develop frameworks for further work in this domain. Considered individually, the frameworks are helpful (Table 2), but their complementary use may be much more effective.

Table 3. Tentative strengths and weaknesses from initial findings

Approach	Understanding	Resolving
Cognitive	+	+
Computer Science	±	+
Linguistic	+	−

The three frameworks could be linked together in various ways together. Taking our case studies, we hope to explore issues in developing more robust frameworks for addressing semantic interoperability issues.

Our initial results let themselves be summarized as an extension of technical interoperability to include persistent data sharing issues with a focus on semantics. Interoperability issues can obviously be seen in a number of different lights. We find it is important to differentiate between technical and semantic issues and simultaneously find different ways to connect the two. Technical and semantic issues complement each other in a myriad of ways calling for solutions to the problems at hand as well as research into the reasons for these problems. Clearly, most applications call for a more pragmatic engineering approach to constructing interoperable geographic information technologies. There is equally an outspoken need to grapple with underlying problems and map out the landscape where technical issues meet social differences.

There are many uncharted uses of terms in a variety of settings that no standardization will ever effectively resolve. Although a mapping for resolving different semantics in SAIF and SDTS may be technically possible, the myriad derivatives which will come on the tails of these standards will regularly break these standards. They call for more viable solutions with the flexibility for dealing with differences on the basis of algebraic approaches (Kuhn 1997). Semantic interoperability aims to improve our understanding of the meanings people associate with geographic information and so help overcome barriers for data sharing and work towards developing more robust interoperable solutions in the future.

References

Arbeitsgemeinschaft der Vermessungsverwaltungen der Bundesrepublik Deutschland (AdV) (1995) Amtliches Topographisch-Kartographisches Informationssystem (ATKIS). Landesvermessungsamt Nordrhein-Westfalen, Bonn

Barry A (1997 10–11 July 1997) Real time: European devices and London air. Paper presented at the CSTT Workshop 'Actor Network and After', Keele, UK

Bishr Y (1997) Semantic Aspects of Interoperable GIS. ITC, Enschede

Bishr Y (1998) Overcoming the semantic and other barriers to GIS interoperability. International Journal of Geographic Information Science 12(4):299–314

Callon M, Law J, Rip A (eds) (1986) Mapping the Dynamics of Science and Technology. Sociology of Science in the Real World. Macmillan, London

Cowardin L, Carter V, Golet FC, LaRoe ET (1979) Classification of Wetlands and Deep Water Habitats in the United States (FWS/OBS-79/31 (http://www.nwi.fws.gov/classman.html)). U. S. Fish and Wildlife Service, Washington, DC

Coyne R (1995) Designing Information Technology in the Postmodern Age. From Method to Metaphor. MIT Press, Cambridge, MA

Department of the Interior, & Fish and Wildlife Service (1990) Federal Coastal Wetland Mapping Programs (Biological Report 90(18)). U.S. Department of the Interior, Washington, DC
Ewusi-Mensah K (1997) Critical issues in abandoned information systems development projects. Communications of the ACM 40(9):74–80
Fauconnier G (1994) Mental Spaces. Cambridge University Press, Cambridge
Fauconnier G (1997) Mappings in Thought and Language. Cambridge University Press, Cambridge
Federal Geographic Data Committee, & Wetlands Subcommittee (1997) Proposal to Establish the Cowardin System as the Federal Wetland Classification Standard (WWW Document http://www.nwi.fws.gov/qanda.html). USGS, Washington DC
Forman RTT, Godron M (1986) Landscape Ecology. John Wiley & Sons, New York
Frank AU, Raubal M (1998, 11–15 July 1998) Specifications for interoperability: Formalizing image schemata for geographic space. Paper presented at the SDH '98, Vancouver, Canada
Große-Enking J (1994) Kontrollen bis 31. August. Landwirtschaftliches Wochenblatt 32(94):16–17
Harvey F (1997a, 19–21 November 1996) Agreeing to disagree: The social construction of geographic information technology. Paper presented at the GIS/LIS '97, Cincinnati, OH
Harvey F (1997b). Improving Multi-Purpose GIS Design: Participative Design. In: Hirtle S, Frank AU (eds) Spatial Information Theory. A Theoretical Basis for GIS, Vol. 1329, pp 313–328. Springer, Berlin Heidelberg New York
Holland D, Quinn N. (eds) (1987a) Cultural Models in Language and Thought. Cambridge University Press, Cambridge
Holland D, Quinn N (1987b) Culture and cognition. In: Holland D, Quinn N (eds) Cultural Models in Language and Thought, pp. 3–40. Cambridge University Press, Cambridge
Kashyap V, Sheth A (1996) Semantic and schematic similarities between database objects: a context based approach. VLDB Journal 5:276–304
Kashyap V, Sheth A (1997) Semantic heterogeneity in Global Information Systems: The role of Metadata, context, and ontologies. In: Papzoglou M, Schlageter G (eds) Cooperative Information Systems: Current Trends and Directions, pp. 1–30. Academic Press, New York
Klein M, Riecken U, Schröder E. (1997) Begriffsdefinitionen im Spannungsfeld zwischen Naturschutz und Landwirtschaft. Naturschutz und Landschaftsplanung 29(8):229–237
Kuhn W (1994, 5–9 September) Defining semantics for spatial data transfers. Paper presented at the Sixth International Symposium on Spatial Data Handling (SDH '94), Edinburgh, Scotland, UK
Kuhn W (1997) Approaching the issue of information loss in geographic data transfers. Geographical Systems 4(3):261–276
Lakoff G (1997) Contemporary Theories of Metaphor. In: Ortony A (ed) Metaphor and Thought. Cambridge University Press, Cambridge
Latour B (1987) Science in Action. Harvard University Press, Cambridge, MA
Latour B (1992) Where are the missing masses? The sociology of a few mundane artifacts. In: Bijker W, Law J (eds) Shaping Technology/Building Society, pp. 265–257. MIT Press, Cambridge, MA
Latour B (1993) We Have Never Been Modern (Porter, Catherine, Trans.). Harvard University Press, Cambridge, MA
Latour B (1996) Aramis or the love of technology (Porter, Catherine, Trans.). Harvard University Press, Cambridge, MA
Latour B, Bastide F (1986) Writing science – fact and fiction: The analysis of the process of reality construction through the application of socio-semiotic methods to scientific texts. In: Callon M, Law J Rip A (eds) Mapping the Dynamics of Science and Technology. Sociology of Science in the Real World, pp. 51–66. Macmillan, London
McKee L (1998, February) Open for business. Mapping Awareness 12:18–20
National Academy of Public Administration (1998) Geographic Information for the 21st Century. Building a Strategy for the Nation. National Academy of Public Administration, Washington, DC
NCGIA (1997, 3–4 December 1997) Paper presented at the INTEROP '97 International Conference on Interoperating Geographic Information Systems, Santa Barbara, CA
Neumann LJ, Star SL (1996, 13–15 November 1996) Making Infrastructure: The dream of common language. Paper presented at the PDC'96, Participatory Design Conference, Cambridge, MA, USA

Nyerges T (1991) Representing geographical meaning. In: Buttenfield BP, McMaster RB (eds) Map Generalization: Making Rules for Knowledge Representation, pp. 59–85. Longman Scientific and Technical, Harlow, Essex

OGC (1998) Geospatial Information Communities – Addressing the Geodata Semantics Problem, [HTML]. Available: http://www.opengis.org/techno/guide/guide980615/Chap6.htm [1998, 5Oct1998]

Salzmann Z (1993) Language, Culture, and Society. Westview Press, New York

Shapiro C (1995) Coordination and Integration of Wetland Data for Status and Trend and Inventory Estimates (Technical Report 2). Federal Geographic Data Committee, Wetlands Subcommittee, Washington, DC

Sheth A (1996) Data Semantics: What, Where and How? Paper presented at the 6th IFIP Working Conference on Data Semantics (DS-6), Atlanta, GA

Sheth A (1997, 3–4 December 1997). Semantic Interoperability in Infocosm: Beyond Infrastructural and data interoperability in federated information systems. Paper presented at the International Conference on Interoperating Geographic Systems (Interop'97), Santa Barbara

Sheth AP, Gala SK (1989, December 11–13, 1989) Attribute relationships: An impediment in automating schema integration. Paper presented at the Workshop on Heterogeneous Database Systems, Chicago

Star SL (ed) (1995a) The Cultures of Computing. Blackwell Publishers/The Sociological Review, Oxford

Star SL (ed) (1995b) Ecologies of Knowledge. Work and Politics in Science and Technology. State University of New York Press, Albany

Star SL, Griesemer JR (1989) Institutional Ecology, "Translations," and boundary objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology. Social Studies of Science 19:387–420

Suchman L (1987) Plans and Situated Actions. The Problem of Human-Machine Communication. Cambridge University Press, Cambridge

Wehland W (1994) Neues zur Flächenstillegung. Top Agrar 1(94):33

Yuan M (1997, 3–4 December 1997). Development of a global conceptual schema for interoperable geographic information. Paper presented at the INTEROP '97 International Conference on Interoperating Geographic Information Systems, Santa Barbara, CA