UNCERTAINTIES IN LAND INFORMATION SYSTEMS DATABASES

Yvan Bédard, Ph.D.
Université Laval
Département des Sciences géodésiques et de Télédétection
Pavillon Casault, #1344
Québec, Qc
Canada G1K 7P4

ABSTRACT

Based on the communication paradigm of Land Information Systems, this paper presents (1) how uncertainty is inevitably introduced in LIS databases, (2) four resulting types of uncertainty, and (3) different means to deal with uncertainty. Finally, the paper suggests that there exist two classes of land data with regards to their reliability.

INTRODUCTION

Land Information Systems (LIS) are useful only to the extent that the information they provide effectively reflect the Reality. However, several limitations affect the veracity of the data stored in spatial databases and concerns about their reliability are highlighted in recent literature (see for example Bouillé 1982; Craig 1983; Blakemore 1983; Dutton 1984; Dangermond, Derrenbacher and Harnden 1984; Robinson and Strahler 1984; Robinson and Frank 1985; Zwart 1985; Bédard 1986 a-b). The following pages relate to these concerns, they present (1) the causes affecting the veracity of land data, (2) four resulting orders of uncertainties, and (3) different solutions to better deal with these uncertainties.

The paper is based on Communication Sciences, Information Theory and Computer Sciences concepts. The overall analysis is built upon the "communication paradigm of LIS" as described by Bédard (1986 a-b) where (1) the terms "data" and "information" are respectively used for physical and cognitive models, and (2) a LIS is seen as an indirect sequential communication process between collectors and users of data.

THE COMMUNICATION PARADIGM OF LAND INFORMATION SYSTEMS: AN OVERVIEW

According to the general framework introduced by Bédard (1986 a-b), Land Information Systems begin with observations of the *real world* which is assumed objective, independent of the observers and which is very complex.

To make decisions about this world, *abstraction* is necessary. To do so, humans selectively perceive the reality where living beings, objects, places, events, or their surrogates emit or reflect different *signals* (light, sounds, odors, etc.). LIS observers, like land surveyors, pick up those signals through their five senses which are sometimes assisted by technical extensions such as amplifiers and translaters.

The detected signals travel to the observer's brain to be *recognized*. This consists of matching the detected signals with previously stored "referents" in order to give meaning to those signals. Then, the observer mentally reconstructs the observed part of the reality. This *cognitive image* is the first model of the observed reality in the LIS communication process. It is assumed partly subjective since it depends not only on the reality, but also on the observer and the context.

Afterwards, the observer must *communicate* his cognitive model to the LIS central agency and the LIS users. However, mental models cannot be communicated directly; physical counterparts must be created like sounds, drawings, and writings. Using the right physical counterpart to communicate the desired meaning is called *encoding*; it is the processus used to create data. Those encoded signals, when put together, form a second model of the reality. They form a *physical* model ready for communication. However, this physical model stems from the observer's perception of the reality, not the reality.

Those encoded signals, which are transmitted, have two components: a *content* and a *form*. When we communicate, only one component is transmitted: the form. Nevertheless, those signals convey an *intended* content (which is not built in the data).

Those signals or land data are usually transmitted to an intermediary, an LIS central agency which performs several gatekeeping functions such as implementing modelization and communication rules and checking for data quality. This central agency usually stores the transmitted data in its database, i.e. the LIS database.

Afterwards, copies of the stored land data or new ones created from those previous ones are sent to the LIS users. Those physical signals reach the receiver's sensory organs (or technical extensions) and travel to his brain to be *decoded*; that is, the receiver must perform the inverse operation of the sender's encoding, but with his *own* referents. He must "guess", among potential meanings, the one conveyed by the received signals. To do so, he must know the context and the language used by the encoder because the same content may have different forms and the same form lead to different contents. This interpretation of the message symbolically sent to him allows for the creation of his own cognitive model of a part of the reality, a part that he has probably *not* observed himself. This is the fundamental basis of the communication paradigm of LIS.

It is only when this latter model of the world, created by LIS data users from physical models instead of the reality, is made that the LIS communication process is completed. Thus, a LIS is a sequence of cognitive and physical modeling processes (where the number of model increases with the number of intermediaries in the LIS communication process) (see figure 1).

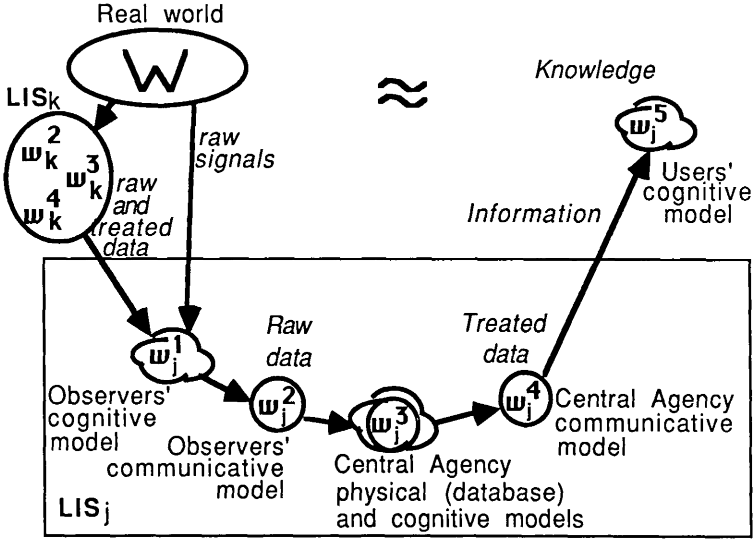

Figure 1: General sequence of model buildings in LIS (from Bédard 1986a) (circles and clouds respectively represent physical and cognitive models).

UNCERTAINTY IN LAND INFORMATION SYSTEMS DATABASES: THE CAUSES

Because uncertainty is introduced each time a model is built, the LIS database w^3 cannot be a perfect model of the reality. As stated by Bouillé (1982), "cartographic results are fuzzy and have been since the earliest beginning of cartography"; such a point of view is endorsed by Robinson and Strahler (1984) who also wrote that "GIS's should be thought of as containing inexact representations of segments of reality".

The uncertainty introduced each time a model is built comes from *two sources*. *First*, it comes from the intrinsic limitations of the modelization process itself; *second*, it comes from the model-makers.

Limitations inherent to the modelization process itself

There are several limitations which are intrinsic to the modelization process, for example:

1- loss of details,
2- purposefulness or goal dependency,
3- model-maker dependency,
4- context dependency,
5- translations between cognitive and physical models are not straightforward,
6- modelization and communication rules are rarely unequivocal.

However, the most important limitation probably is that models are only approximative estimations. This is explained by two kinds of *estimation limitations*: (1) fuzziness in the identification or labeling of an entity, and (2) limitations in the measurement of the properties of those entities.

1- Fuzziness in identification takes place when humans classify realities into discrete groups of entities that do not have sharp boundaries. This can happen (1) on the boundary of an entity type between existing and not existing (e.g. when to attach the label "tree" to a tree? the label "fence" to a fence? "wetland" to a wetland?), and (2) on the border between two types of entities (e.g. in the Quebec assessment file, a multipurpose building which can be classified as entity types "commerce" or "industry" will be described by different attributes depending upon the final choice).

The problem is that the world generally is a continuum while labels are discrete; however, "for many aspects of the world, a discrete set of concepts is adequate.... Yet such distinctions break down when pushed to extremes" (Sowa 1984, 344). The mistake is to suppose that realities are clear-cut and self-identifying. On the contrary, many of our perceptions involve the ambiguous and the intangible. This fuzziness in the identification introduces uncertainty in the existence of an entity or in its classification in the good entity type.

2- Limitations in the measurement of the properties of the observed entities are related to its qualitative or quantitative character (Robinson and Frank 1985). Those limitations are of two kinds: (1) *imprecision* (in its mathematical sense) which is related to the quantitative measurement of attribute values (e.g. standard deviation of 3 cm) and (2) *fuzziness in the qualitative measurement* of attribute values where classes of values do not have sharply defined boundaries (e.g. the building quality codes used in Quebec assessment files: mauvais état, médiocre-, médiocre, standard-, standard, standard+, bon, bon+, excellent état). This fuzziness is in fact the same phenomena than fuzziness in identification but applied at a different level of abstraction. Those limitations in measurement introduce uncertainty in the attribute values stored in a database.

Thus, there are two kinds of estimation limitations which affect LIS users in "knowing what they talk about" (fuzziness in identification) and "describing and locating it" (measurement limitations). Those estimation limitations have different consequences in the content of a database: while fuzziness in identification affects the type of an entity and consequently which properties are measured for this entity, limitations in measurement affect the values of the measured properties. Thus, a fuzzy identification affects the reliability of the entity as a whole while limitations in measurement individually affect the attribute values of an entity.

Limitations related to model-makers

Land data uncertainty is also related to the model-makers involved in the LIS communication process; i.e., the collectors, intermediaries and users of LIS data.

As we can deduce from figure 1, a large amount of LIS data involve human judgments. However, there is a gap between humans' imprecise knowledge of the reality and the crisp representations of the reality which are stored in LIS databases.

Humans, as information processors, have limited capabilities and introduce subjectivity in data. It is commonly accepted that even in the best conditions, a same reality will *not* be modeled the same way by different persons or by the same person in different times and contexts. As stated by Dutton (1984), "our models of reality, including cartographic databases, are highly conditioned by our cultural and institutional consensus concerning 'what exists'. But that consensus varies across space, differs among groups, shifts over time, and is colored by our concerns". Also, "there is an implicit assumption that the information contained in land information systems is objective, quantified and correct (i.e. scientific or rational data)…. the data contained within land information systems is neither totally objective nor necessarily based on acceptable statistical measures" (Zwart 1985).

Communication scientists explain this phenomena by the influence of everyone's *frame of reference* particularities, meaning the influence of someone's history, experience, learning, needs, aspirations, beliefs, values, and personality. This also includes someone's group norms: cultural, professional, and familial.

Also responsible is the *concept of "satisficing"* where someone does not automatically analyse all the possibilities to select the best representation of the real world. Instead, he limits his search for the best solution by accepting the first alternative satisfying all the given requirements (Davis and Olson 1985, 169). This method is very frequent in LIS activities (e.g. differences in cost and time between Quebec subdivisions and "bornages", two operations delimiting rights to the land but with only the latter one having a legal value).

Everyone's frame of reference and use of the satisficing concept influence directly the reliability of LIS databases. This influence may happen during the perception (detection and recognition) of the reality raw signals, during the perception (detection and decoding) of raw and treated data, during the creation of cognitive models, and during the encoding of cognitive models into physical models (e.g. LIS databases).

Thus, uncertainties stemming from the modelization process itself and the model-makers are unavoidable. Consequently, there is an inherent uncertainty in land data which cannot be avoided and LIS cannot deliver perfect information. At best, LIS databases can only be workable approximations of the real world W.

THE PROBLEM

Land information systems communicate models of parts of the real world to identify land-related entities, to describe them and to locate them in space and time. However, even in the best conditions, there are uncertainties affecting the reliability of LIS databases. To better understand the consequences of those uncertainties, i.e. the resulting problem, the following classification has been done:

1- First order *(conceptual)* uncertainty: refers to the fuzziness in the identification of an observed reality (e.g. being or not being such an entity? Being an entity of type A or of type B?).
2- Second order *(descriptive)* uncertainty: refers to the uncertainty in the attribute values of an observed reality (i.e. imprecision in quantitative values and fuzziness in qualitative values).
3- Third order *(locational)* uncertainty: refers to the fuzziness in the qualitative values and imprecision in the quantitative values used for the location in space and time of an observed reality (e.g. error ellipses in geodesy).
4- Fourth order uncertainty *(meta-uncertainty)*: refers to the degree to which the preceding uncertainties are unknown (e.g. absolute error ellipses with a probability of 39.3%; being pretty sure that a building quality is "standard +").

Those four orders of uncertainty combine to each other to generate the total uncertainty in LIS databases, leading to an uncertain information about the real world. The result is a user who doubts if a given reality is in fact such an entity in the real world, if it really has the given attribute values, if it really is where it is depicted, and if the level of those uncertainties is high or low.

THE SOLUTIONS

As previously seen, there is an inherent uncertainty in LIS databases which cannot be avoided. We can take means to (1) reduce this uncertainty and to (2) absorb partially or completely the remaining uncertainty. The right balance among those alternatives depends upon political, cultural, and economical concerns; it is an institutional choice to be done within each jurisdiction.

Uncertainty reduction

Uncertainty reduction takes place when modelisation rules (defining the content of a model, i.e. what to observe and how) and communication rules (defining the form of a model, i.e. which graphical and literal languages to use) are established either (1) to decrease the fuzziness associated with the identification of a spatial entity or (2) to insure precision and crispness in the description and location in space and time of this spatial entity.

This can be done by appropriate technical, procedural, organizational and legal requirements such as geodetic tying of surveys, use of mathematics such as adjustments for repetitive quantitative measurements and fuzzy logic for qualitative measurements, good professional training, high precision standards, mandatory marking of property corners, use of standard symbols, inclusion of lineage in digital maps, mandatory registration of all the rights to the land, etc. Such methods increase the likelihood that the several models which are built in the LIS communication process will correspond more closely to the observed reality.

Any LIS reduces the uncertainty inherent in land data to a certain degree. However, this is limited by fundamental concepts as well as practical and economical conditions. Furthermore, although we can reduce the uncertainty inherent in land data, we cannot eliminate all of it. Thus, there remains, in the LIS communication process, someone who absorbs, in whole or in part, the effects of the remaining uncertainty.

Uncertainty absorption

Uncertainty absorption takes place when a model maker guarantees his model of the reality and compensates the users damaged by poor data. Uncertainty absorption also takes place when non-guaranteed models are utilized. Here, the user and not the provider of data absorbs the uncertainty. In fact, the level of uncertainty absorption is defined as the level of (monetary) risk in providing or using data. When errors in data cause damages to users, the ones who pay for these damages are the ones who absorb the uncertainty.

Uncertainty absorption is very different from uncertainty reduction. In the latter case, the uncertainty is literally reduced (ex. requiring a precision of 3cm instead of 10cm, asking for the opinion of two or three land surveyors instead of only one). In the former case however, there is someone who guarantees the data as the "truth" and who is willing to take the inherent risk (e.g. guarantee of titles and boundaries with indemnity funds in the Massachusetts registration system).

Only the absorption performed or made official by the LIS central agency (or its extension like a tribunal) applies to all the participants in the LIS communication process. In such cases, the LIS central agency has the power and means to impose a specific model of the world as the "good one". When this happens, an LIS database (or part of it) becomes the "official" view of the reality, a kind of "artificial truth" binding every participant in the LIS. Although these models do not necessarily represent exactly the reality, they represent the "official" version, the "official" model of this reality and they are guaranteed.

Such an alternative almost eliminates the uncertainty inherent in the original nature of data. In fact, it really absorbs the remaining uncertainty and decisionmakers can better rely on such data. Users of those data must and can rely on them.

It is interesting to note at this point that most of the ways to reduce uncertainty are technical, while most of the ways to absorb the remaining uncertainty are institutional. Finally, the higher the uncertainty reduction, the lower the uncertainty absorption needed.

CONCLUSION

Land data are physical and formal symbolic surrogates created by humans to communicate information about the description and location of land-related realities. Thus, LIS databases contain the symbols of the LIS communication process with their inherent uncertainty. However, the four resulting orders of uncertainty can be reduced and absorbed by appropriate means. This gives rise to two types of data in LIS databases:

a) *Second class* land data: this is the typical data found in LIS databases, they have a certain degree of remaining uncertainty which has not been absorbed. These data are approximative surrogates which lie on a spectrum going from *vague* on one end (data with a lot of uncertainty remaining) to *exact* on the other end (little uncertainty remains).

b) *First Class* land data: this type of land data is rare in LIS databases, this is the data for which the uncertainty has been absorbed. These land data are the *official model* binding every participant in the LIS. The original nature of these land data has been changed to "artificial truth" and no uncertainty remains for them.

Exact land data are very good approximations of the reality. But, only the First Class land data can be considered as having a complete reliability.

CITED REFERENCES

Bédard, Y. 1986a, A Study of the Nature of Data Using a Communication-Based Conceptual Framework of Land Information Systems. Ph.D. Dissertation, University of Maine, Orono, 260 pp.

Bédard, Y. 1986b, A Study of the Nature of Data Using a Communication-Based Conceptual Framework of Land Information Sytems, Updated Version. First presented at the XVIII Congress, Fédération Internationale des Géomètres, Toronto, June 1-11, Volume 3. Updated version reprinted in The Canadian Surveyor, Winter Issue.

Bédard, Y. 1986c, Comparing Registration Systems Using a Communication-Based Criterion: Uncertainty Absorption. XVIII Congress, Fédération Internationale des Géomètres, Toronto, June 1-11, Volume 7.

Blakemore, M. 1983, Generalization and Error in Spatial Data Bases. Proceedings of the Sixth International Symposium on Automated Cartography, Ottawa, October 16-21, Volume 1, pp. 313-322.

Bouillé, F. 1982, Actual Tools for Cartography Today. Cartographica, Volume 19, No.2, pp. 27-32.

Craig, W.J. 1983, Problems of Spatial Accuracy. Decision Support Systems for Policy and Management. Annual Conference of the Urban and Regional Information Systems Association, August 14-17, Atlanta, pp. 1-10.

Dangermond, J., B. Derrenbacher, and E. Harnder 1984, Description of Techniques for Automation of Regional Natural Resources Inventories. Appendix 2: Error Potential Related to Integrated and Parametric Mapping in GIS. In Seminar on the Multipurpose Cadastre: Modernizing Land Information Systems in North America. B.J. Niemann Ed. Madison, Wisconsin: the University.

Davis, G. B. and M. H. Olson 1985, Management Information Systems: Conceptual Foundations, Structure, and Development. 2nd ed. New York: McGraw-Hill.

Dutton, G. 1984, Truth and its Consequences in Digital Cartography. Technical Papers, the American Congress on Surveying and Mapping, 44th Annual Meeting, Washington, D.C., March 11-24, pp. 273-283.

Minsky, M. L. 1965, Matter, Minds, and Models. Proceedings of the International Federation of Information Processing Congress, Vol. 1, pp. 45-49. Reprinted in Marvin L. Minsky, ed. Semantic Information Processing. MIT Press, 1968.

Robinson, V. B. and A.H. Strahler 1984, Issues in Designing Geographic Information Systems under Conditions of Inexactness. Tenth International Symposium, Machine Processing of Remotely Sensed Data, West Lafayette, Indiana, June 12-14, pp. 198-204.

Robinson, V. B. and A. U. Frank 1985, About Different Kinds of Uncertainty in Collections of Spatial Data. Proceedings of the Seventh International Symposium on Automated Mapping: Digital Representations of Spatial Knowledge. American Society of Photogrammetry and American Congress on Surveying and Mapping. Washington, D.C. March 11-14, pp. 440-449.

Sowa, J. F. 1984 Conceptual Structures: Information Processing in Mind and Machine. Reading, Mass.: Addison-Westley.

Zwart, P. 1985, Response Paper on Assessing User Requirements and System Design Approaches to Land Information Systems. Workshop on Fundamental Research Needs in Surveying, Mapping, and Land Information Systems. November 17-20, Virginia Polytechnic Institute and State University.