

DATA PROCESSING PROTOCOLS IN THE FRAMEWORK OF A TERRITORIAL INTELLIGENCE SYSTEM

Antonio Moreno Moreno

Observatorio Local de Empleo, Universidad de Huelva
amm@ole.uhu.es , + 34 959 21 94 12

Adresse professionnelle

Universidad de Huelva, Campus de El Carmen. Pb.7

Summary: This paper presents results about the modelling of the data processing protocols within community systems of territorial intelligence. The CSTI are territorial information system accessible to the territorial actors and respectful of a sustainable development ethic. This research activity is worked out in a team of the coordination action of the European network of territorial intelligence (caENTI) dedicated to the CSTI (Wp6s), within the work package 6 "Tools for and with actors", which design and disseminate methods and tools.

Keywords: territorial intelligence, protocols, data analysis, data processing.

Data Processing Protocols in the framework of a Territorial Intelligence System

1. THE FRAMEWORK: TERRITORIAL INTELLIGENCES SYSTEMS

An Information System is an organized set of features, which can be persons, data, activities or resources in a general aspect. These elements interact with each other to process information and disseminate it in an appropriate manner. If we add to this definition the concept of territory as an element which is linked to the information, we will have a Territorial Information System (TIS). If, moreover, that system is accessible and at the service of territorial actors that want to develop democratic governance at the service of sustainable development [Girardot (2006)] we're talking about a **Community System of Territorial Intelligence (ICTS)**.

This particular kind of community information systems provide us some positive aspects:

- It favours the information sharing within a territorial development partnership;
- It instruments the data cooperative analysis and the results participative interpretation;
- It introduces the citizens' participation in the process of decision-making;
- It provides the actors with the useful information to draft projects, and then to manage them and evaluate them.

Is this concept that is being developed under the coordination action of the European Network of Territorial Intelligence (caENTI) [Girardot (2001), Dumas (2004), Bertacchini (2004)]: It exploits the potential that computer science offers to gather information, share this information, favour the actors' partnership from the information cooperative analysis, and to increase the citizens' participation by improving their information and their access to information. It respects the constraints that sustainable development, particularly participation and partnership, imposes to information processing and publication protocols. During this research, we became aware of the editorial function importance of this convivial analytical chain that is directed towards online results edition. In the context of the territorial intelligence systems, this **editorial chain** concerns production of digital documents that should be shared within a partnership. It is also a work flow which cannot only be organized according to data analysis technical protocols but also to allow participation within the partnership and above all beyond it, within the territorial community. That is why we call them Territorial Intelligence Community Systems (TICS) [Girardot (2006)].

It's in this model of the community system of territorial intelligence developed in the research of the CAENTI where it makes sense to consider, within the editorial chain, the concept of processing and analysis protocols, plus others (see Figure 1 . Territorial Intelligence Community System model).

2. DATA PROCESSING PROTOCOLS

In the proposed community system of territorial intelligence model shows a set of actions (marked in red on the Chart 1) that represent the actions of processing the data and that is perfectly modelled. Following the editorial chain concept developed, protocol this set of processes represents another step in the accessibility of IT tools to the end user (territorial actor), because through this automation, is possible go to next top level in the chain without the prerequisite experience and expertise in the use of the tools used.

Therefore the data processing protocols aims at automating the data processing to make it more accessible to territorial actors. They respect scientific and technical criteria as the statistical representation of the data, or the rules of communication according to publics, that assume the quality of the analysis. They had to correspond to the uses of the actors. In the particular case of Catalyse methodology used by caENTI, the automatic already existing processes with a new set of actions to select, to recount, to calculate and to represent the information, so that the human interventions limit themselves to the stages of capture of decision, all this under the beginning of the representation (quantitatively) and of the sense (qualitative).

This paper shows, specifically, the work done to establish the protocol of data processing used by the CATALYSE observatories, particularly those designed and put in practice by ACCEM and Integra for their observatories, which there will be necessary to adapt to the European Guide designed and agreed by caENTI.

The data processing protocol can be in four general phases of more or less automated form:

1. Control of data quality.
2. Codification of open questions.
3. Recodification of closed questions.

4. Selection of the characters for the qualitative analysis.

Protocols can be established for each question or indicator of the guide to facilitate territorial information analysis by inexperienced actors.

The present paper tries to establish, on the one hand, a series of procedures obliged to be accomplished as for the general treatment of the information and that will shape a General Protocol for the data processing and, on the other hand, and being based specifically on the European Guide, to establish for each indicator the basic elements for the codification and recodification, a few guidelines for the help on technical aspects to the analysis of the information and to the interpretation of the results.

2.1 Suppression of bad quality questions.

The absence of answer to a question complicates the processing and it is generally advisable to draw aside the guides in which answers are missing. The reconstitution of missing answers is absolutely proscribed. However, it is necessary to differentiate the "failure to reply", what means that the person did not answer, from the answer "without object" which logically results from a former answer .

The control of responses absent affects both individuals and questions. In general would have to eliminate all individuals for whom lack a response or more. Nevertheless, the data acquisition cost and the concern in not introducing bias in a sample can oppose to a so drastic information loss. In this case, it is necessary to eliminate the individuals and the questions for which the rate of missing answers is too important, for example higher to 5%.

2.2 Coding of the open questions.

A question is called "closed" if the possible answers are expressed in the shape of a modalities list among which you should choose. A question is called "open" if the answer is free. When the question is closed, its answer is a code which indicates the answer. When the question is open, its answer is called a measure if it has an order; in the other cases it is called a text. The codes are the only ones that can directly be the subject of a quantitative processing (that is called sorting), and then of a qualitative analysis of data by means of a recoding. That is why it is advisable to avoid the open questions. If that is not possible, we should keep in mind that the measurements are easier to automatically code, whereas the texts imply a long hand coding.

The open questions coding precedes the steps that are made on closed questions. The coding aims at changing the open questions into closed questions, that is to say the values, measures or texts into codes. It is an indispensable operation that will allow making statistic sorting on the data, and from these quantitative counts making histograms or possibly geographical maps. It prepares the selection of characters for a qualitative analysis.

Traditionally is normal to coding the questions according to their order number, and then to their modalities and row. This coding is enough for the quantitative processing and for cartography. However, it is extremely difficult to make the guide evolve, because the suppression or the addition of a question or a modality generally results in a codes shift. Consequently, we prefer having a coding that would be specific to the data analysis which consists in indicating the questions and their modalities by a two-letter alphanumeric code for the questions and with four letters for the modalities. To be more exact, the modalities code corresponds to the question code for the first two letters and the last two letters more specifically identify the modality. Of course each modality is identified by a single code at the whole guide level. We can remove and add questions and modalities without any consequence, if we create new codes in the event of addition.

2.3 Recoding of the closed questions.

The processing protocol is characterized by a series of recodings. A first recoding generally allows gathering an important list of modalities in a shorter list of categories (if the question is qualitative) or of classes (if the question is quantitative). As is well known people have difficulty to simultaneously memorize more than seven items, we usually limit the modalities lists to six choices. When it is impossible, for example for the incomes which include around twenty different types, it is necessary to envisage gathering the modalities in six categories at the maximum. In the case of a European questionnaire, this regrouping can allow harmonizing. Generally, we can envisage a second tightened regrouping with only two or three categories to answer the evaluation needs.

Another kind of recoding process are linked to the representativeness lack of a modality, for example, are not of that kind, but it is possible to anticipate them, for example by fixing a representativeness threshold below which a modality, or a category, will be removed or gathered with another one. But the improvement of the modalities representativeness does not constitute the only aim of recoding anymore. This one is now often motivated by the will to effectuate comprehensible tables by various publics. The communication aims became also diversified with useful synthesis to prepare the decisions of the operational group, a more detailed vision of results for the work of workshops or cooperative spaces, accessible tables for public presentations and publication on Internet.[Girardot (2004)]

2.4. Selection of characters for the qualitative data analysis.

All these operations we have seen are designed to prepare the data for qualitative analysis. In this logical way, one first globally select the questions to be analyzed, then code the open questions in order to select the characters from the modalities, possibly by regrouping several modalities.

The characters selection has to be performed now as a distinct function, following the coding and recoding operations that can also satisfy aims related to quantitative processing or to the results communication before or after having contributed to the making of data table for the qualitative analysis.

It should also become a simple selection of characters among the recoded modalities. The experience also shows that we perform several data analysis, the first one having more and more a function of assessment about decisions beforehand made during the coding and recoding and an exploratory function.

2.5 Other data processing considerations.

The questions with multiple answers also complicate the processing, as a consequence it is necessary to limit its number by using, as often as possible, questions with single choice. When it is not possible, it will be necessary "to break up" the question by creating as many questions with yes/no answers as there are modalities.

The coding, the recoding and the selection of characters can involve the suppression of answers. Then they make missing answers. The control of missing answers has consequently to be repeated after each of these operations.

Conclusion

Figure 1. Territorial Intelligence Community System Model.

Source: CAENTI (2008): “CATALYSE Toolkit” Specifications for the processing and editorial chain from territorial data to results. Deliverable N° 58

Bibliography

Girardot, J.-J. (2003), “Concept, method and tools of territorial intelligence”, *invited conference in the Social Worker National Training School*, Nantau (Taiwan).

Sánchez, C., Girardot, J.-J. (2006): “Specifications of the contents of the European Guide of Diagnosis and Evaluation.”, in: *Acts of International Conference of Territorial Intelligence*, ALBA IULIA (Romania), <URL: <http://www.territorialintelligence.eu/telechargement/albaiulia2006/Alba06-Sanchez.pdf>>

Sánchez, C., Asensio Coto, M.J., Girardot, J.-J. (2007): “Specifications for a European on-line Inclusion Itinerary Accompaniment File”. CAENTI Deliverable N° 57.

- Appendix C – “Updated protocols of data analysis for the Guide within European on-line Inclusion Itinerary Accompaniment File”.

- Appendix E- “ACCEM Data processing and personal data analysis protocols in ACCEM Experimentation”.

Girardot, J.-J. (2003), “Specifications for the “PRAGMA” data collection and quantitative treatment software.” CAENTI. Deliverable N° 54.

Sánchez, C., Girardot, J.-J., Masselot, C. (2006): “CATALYSE Toolkit. European contents specifications of an online territorial repertory of services. CAENTI Deliverable N° 52.

Asensio Coto, M.J., Girardot J.J. (2006): “Guidance notes for the use of CATALYSE information and tools. Part III: Data analysis protocols of the guide. Uses of CATALYSE in development partnerships. Conclusion”. CAENTI Deliverable N° 56.