Comparing quantitative and qualitative measurements of water quality by using Geographic Information Systems

Camporese, Rina
Iuav University of Venice, New Technologies & Information on Territory and Environment Working Group
via E. Chiaradia, 5
I-35134 Padova, Italy
E-mail: rina.camporese@poste.it

The foundations of this work are the reflections that arose from reading the articles "Thinking spatially in the social science" [Goodchild, Janelle, 2004] and "GIS and geostatistics: Essential partners for spatial analysis" [Burrough, 2001], integrated by the author's past research in Social Statistics and actual study of Geographical Information Science. These, in concise form, the fundamental premises:

- space and territory are essential to study social and environmental phenomena, and also constitute the basis for integrating data from different sources, thanks to geographical coordinates
- the comparison of objective measures on the state of the environment and subjective perceptions of that same environment by people who live there generate new knowledge, analogously of what happens in health statistics when measured and perceived health is combined
- both citizens' individual behaviours and authority decision-making are relevant in determining the state of the environment; therefore, to fully understand environmental phenomena it is important to integrate the views of institutions and citizens
- the integration of disciplinary disparate data creates new synergic information relevant to the study of environmental phenomena

This work aims at highlighting the potential of georeferenced information, managed in a GIS environment, to integrate data that, traditionally, are treated separately by environmental and social statistics. The two disciplines could meet at a geographical working desk to benefit from mutual knowledge and reveal the intertwined relations among human behaviour, territorial policies and environmental status.

In particular, the comparison of quantitative and qualitative measurements of water quality will be taken into consideration, by using Geographic Information Systems' techniques to integrate data coming from:

- institutional agencies, regarding quantitative measurements of water quality, local policies on water resources management
- georeferenced sample data coming from a national household social survey on daily life which covers environmental issues from the citizens' point of view
- qualitative evaluations tagged in web 2.0 environment, with a view to "contaminate" official institutional statistics with volunteer and collaborative information on the internet.

The tentative method is illustrated with partially simulated data and could be generalised to other environmental issues (i.e. air quality, waste management and production, …).

Geography as a working desk

The characterisation of statistical units and measurements in a geographical context – i.e. georeferencing – enables the analysis of the relationships among unconnected data by the use of the territory as a merging key. Overlapped statistical layers on a geographical basis, in the style of Geographic Information Science [Longley et al., 2010], create new synergic information and enlighten the relations among different phenomena occurring in the same area [Goodchild, Janelle, 2004]. GIS and Web can offer effective techniques to analyse and display statistical data on a geographical underlying layer. Moreover,

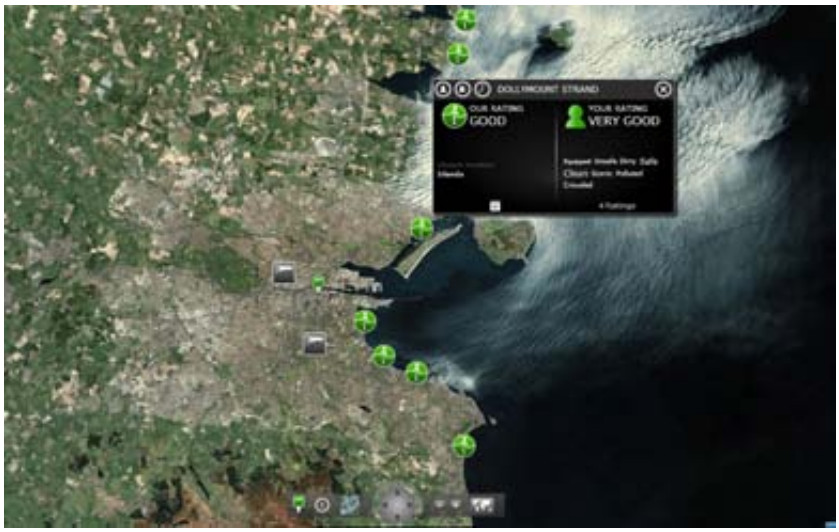
image maps and navigation tools (zoom, pan, view angle, etc.) make the representation of statistics possible into daily life environment, enriching them with the context information about the places they belong to [Batty et al., 2010].

Needless to say, strategies to protect data confidentiality are essential: some are easy to imagine (different views related to zoom levels, buffers of proper size to mask identities, ad hoc coordinates' transformations to prevent overlapping on common reference systems), others still have to be invented.

Georeferencing statistical units

Two examples of the use of georeferenced micro-data are illustrated above, one regarding data dissemination and another regarding the integration of independent sources.

The first case is Eye On Earth: a geographic platform where the European Environment Agency shares institutional data and opens them up to the citizen's comments [Goodchild, 2007]. Thanks to network technologies (cloud computing, mash-up, etc.), data of air and water monitoring stations, collected by institutional agencies, are displayed on a navigable map. Users can also geotag their own perception of the quality of air and water. Institutions and citizens gather around a table covered by remote sensing images; the comparison between objective environmental measurements and people's subjective perceptions takes place on a natural representation of the environment. Data and documentation follow strict scientific criteria, but access to information is familiar to web and social networks users: this makes official agencies and citizens closer and provides data for an informed dialog among different stakeholders. Moreover, data displayed on a territorial representation facilitate communication and interaction with end-users. This example regards punctual micro-data that can be published without violating anybody's privacy; with appropriate changes, the idea could be applied also to more sensitive data.



Eyeonearth.eu

The second example derives from the idea of georeferencing sampling units. The topological relationship among various entities on the same territory generates a synergy of information and enables the calculation of indicators that would otherwise have not been known. For example, if the micro-data of National Institute of Statistics sample surveys on daily life were geocoded according to the respondents' address, household habits in the management of waste and local government strategies for recycling could be connected at micro-data level. Coming to water, habits of consuming tap water and the reasons for not drinking it could be connected with pieces of information on local water quality and competent authority policies. Geographic details of individual responses would only be used during the analysis phase: results would be aggregated and respectful of the territorial disaggregation allowed by the sample strategy.

Water quality measurement: quantitative and qualitative, objective and subjective

For the sake of brevity, the arguments that follow relate to water for domestic use, but, by analogy, they can also be extended to the coastal waters or waters for other uses.

There are many institutional sources of data on water quality: Local Environmental Agencies, European Pollutant Release and Transfer Register, Water Authorities, National Institutes of Statistics, etc. Such data cover a wide range of aspects related to water and generally contain quantitative and objective measurements: strategies and policies for water management – collection, purification and release of water for the consumption - purification and pollution load of urban wastewater - hydrography, precipitation and water resources - pollution - soil characteristics - etc.

On the other hand, social surveys produce subjective measurements of water quality and describe citizen's habits with regard to water consumption. In Italy, for instance, daily life and household consumptions sample surveys, conducted every year by Italian National Institute of Statistics, collect information on habits and perceptions related to domestic water: habit of drinking tap water and reasons for not drinking it, evaluation of water supply regularity and cost of water supply service, habit of buying mineral water and average monthly expense.

If all such pieces of information were georeferenced, they could be integrated and analysed thanks to Geographic Information Science in order to reach two knowledge objectives:

- comparison of institutional strategies for the environment and behaviour of citizens: which is the relationship between policies for management and protection of water resources and the daily behaviours of the population?
- comparison between subjective and objective measures of environmental quality: what is the degree of awareness about water quality for domestic use? what is the relationship between the perception and confidence on the quality of water and the quantitative measurements of water properties?

GIS techniques to elaborate and integrate data

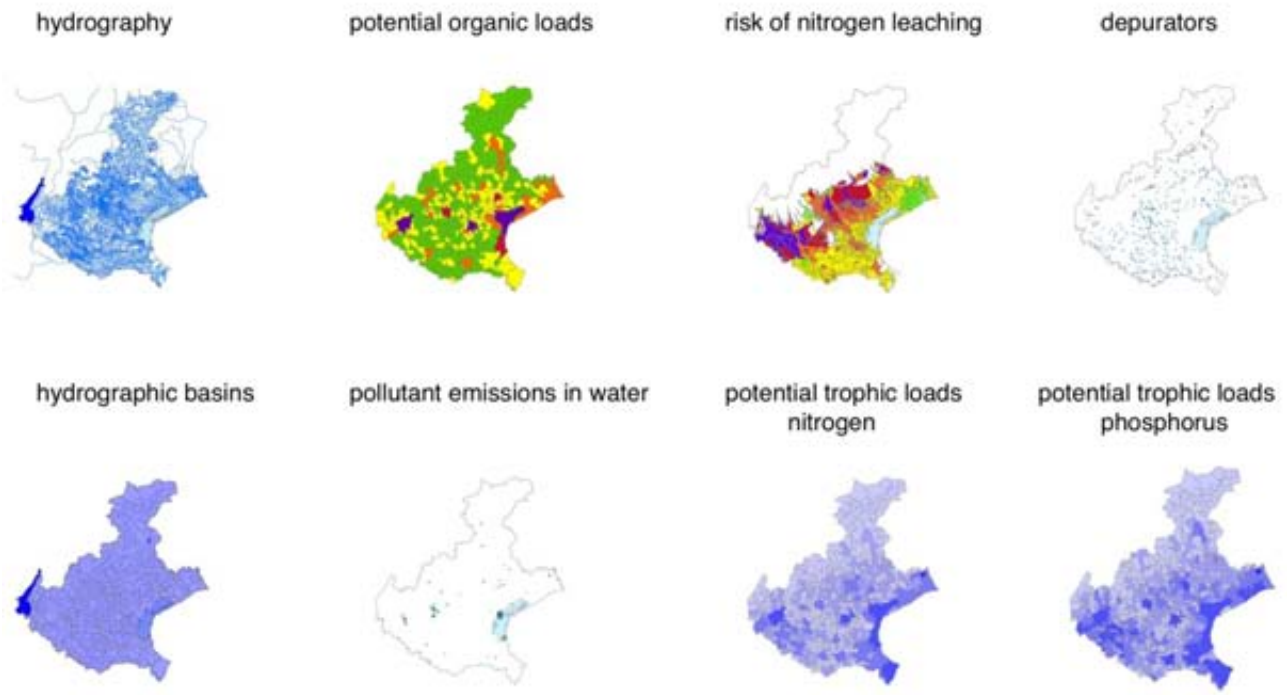
Geographical location enables to connect the people's evaluations and information supplied by institutions. Once integrated such different data layers within a single GIS environment, it is possible to calculate a lot of derived information such as:

- the relationship between confidence in drinking tap water and objective measures of water quality in the area of residence
- habit of buying mineral water and average monthly expenditure for it versus water rationing strategies or overall quality of the water.

Data and examples illustrated hereby come from Knowledge Framework of Veneto Region, European Pollutant Release and Transfer Register and Daily Life and Household Consumptions sample surveys. The work, for the sake of brevity, will be limited to the area of Veneto region.

Since institutional data on water quality are various and not homogeneous, they can be synthetised thanks to a Multi-Criteria Evaluation analysis applied to raster data [Eastman et al., 1995] [Cromley, Tyler Huffman, 2006]. Different available information on water quality can be translated into several raster data layers, overlapped on the selected territory. Then, thanks to the Multi-Criteria Evaluation (MCE) technique, multiple raster layers can be weighted and combined into a synthetic layer which represents an overall assessment of the water quality. In GIS applications each data layer is a criterion and can be transformed into either factors or constraints. In the traditional use of MCE for decision making processes, a factor influences the suitability of a decision according to its value in a scaling function, while a constraint produces an absolute result in terms of inclusion or exclusion of possible outcomes (scaling transformation: zero/one dichotomy). Data layers can also be distinguished in site versus situation characteristics. Site characteristics regard absolute qualities of a specific location, while situation characteristics refer to relative qualities of a location, with respect to other places or features (e.g. distance from a point of interest). Data layers can also be characterized either by a positive or a negative influence on the investigated phenomenon; i.e. a positive

spatial externality implies that low distances generate high benefits and vice versa. In view of obtaining an overall assessment of water quality, each available piece of information can be interpreted as a factor which influences the global quality of water available for domestic use in a given portion of territory. The layers have to be weighted according to their relative importance and normalised to a common scale; then they can be combined to obtain an overall raster map whose values represent a synthetic measure of water quality, obtained through the simultaneous evaluation of many different characteristics. The entire process includes a great part of subjectivity in the decisions related to rasterization of data layers, and to weighting and scaling of factors. This represents both the critical point and also the potentiality of such a method. Anyway, thanks to GIS tools, the output raster map containing a global water quality measure can be easily compared with its various components, e.g. via small multiples visualization, so as to have a better understanding of the different pieces of information underlying.



Some possible input layers for MCE

Unfortunately micro data on citizens' perceptions and behaviours are not georeferenced. Consequently, the following analysis are simulated because in micro-data files available for research purposes, only a regional code (NUTS2) expresses the territoriality. If during the data collection phase units would have been georeferenced thanks to the address of residence, each individual response could have been considered as a measurement made by a "human sensor"; this perspective would open up new possibilities for understanding environmental phenomena seen through the prism of citizen. Knowing the location of the respondents' residence would allow the calculation of many derived variables using the geographical key to link individual responses to indicators on the state of the territory in which they live.

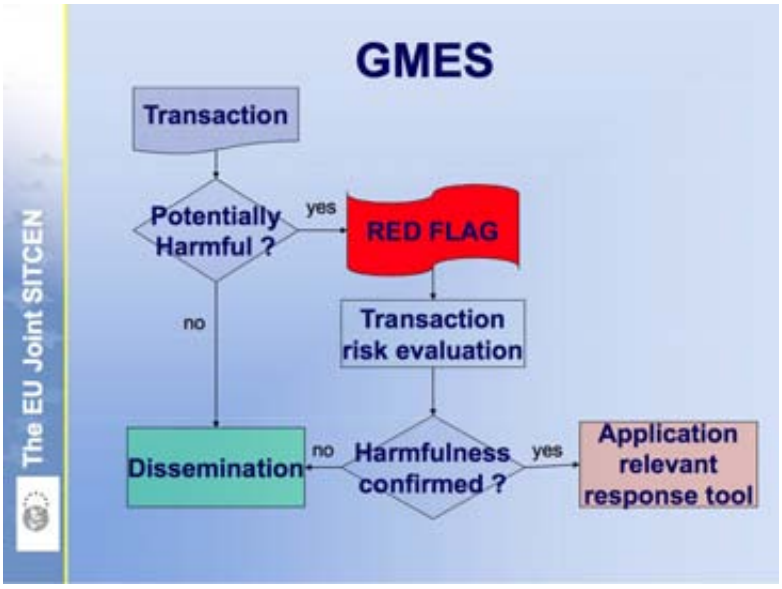
For the purposes of this test, the desired situation has been simulated thanks to the laws of probability and sampling, by using the marginal distributions of observed frequencies and the layers of information on land cover of the European Urban Atlas project. Hypothetical addresses of respondents' households have been selected within urban residential areas with a probability proportional to the density of the urban built environment. To each household random responses to environmental questions have been attributed, so as to reproduce the marginal distributions observed in the real sample of 2009.

The entire work is in progress.

Why are some data simulated?

It was not possible to access individual micro data for daily life survey because the law protecting confidentiality of respondents inhibits access to this information even to universities and researchers that are not part of National Statistical System which gather statistical offices of many Italian public institutions. There is still a strong resistance to give access to such information even for research purposes aiming to produce shared knowledge.

An interesting approach to overcome the normative limits to access sensitive data, while preserving the confidentiality of respondents, is the one being currently developed by EU Joint Situation Centre in the context of Security for European Space Tools. The decision whether or not to disseminate data is not based on the intrinsic nature of the data (sensitive or not, dangerous or not ...), but on the evaluation of the transaction applied to the data. The question to be answered is then: is the transaction to be applied to data potentially dangerous or not? If not, the data are released. In this perspective, the access to sensitive personal data would be possible as long as the processing and the published results (transaction) don't violate the confidentiality of respondents.



EU Joint Situation Centre - Security for European space tools *[Claeys, 2011]*

Drink tap water? Buy mineral water?

Integrating different sources, i.e. integrating multiple points of view, offers new knowledge frameworks that arise from the relationship between institutional strategies, state of the environment, behaviours and perceptions of the population. The potential of integrating different pieces of information in a geographic environment is not limited to the production of thematic maps. The relationship existing between different entities that belong to the same territory creates synergic information and enables the calculation of measures and indicators that would otherwise have not been known. The idea will be illustrated hereafter thanks to a simple example, which contains some key elements.

Is there a relation between habits of drinking tap water or buying mineral water and quantitative measures of domestic water quality? The procedure described before could offer some answers to such a question; it would allow, for instance, to produce the following table which is very elementary, but still not available in Italy.

domestic water overall quality in residence area	households drinking tap water			
		no		
	yes	not drinkable or not trusted	other reason	*total*
very good	あい %	うえ %	おう %	げこ %
good	けこ %	がぎ %	げこ %	ざぬ %
fair	さす %	せそ %	ざぬ %	せそ %
bad	ひふ %	ほぴ %	まみ %	がぎ %
very bad	がぎ %	ざぬ %	あい %	あい %
total	67%	28%	5%	100%

	households buying mineral water	
	yes	average montly expense
very good	あい %	うえ
good	けこ %	がぎ
fair	さす %	せそ
bad	ひふ %	ほぴ
very bad	がぎ %	ざぬ
total	68%	€18

A possible table: drinking tap water and buying mineral water vs domestic water quality

Similar procedures would enable to compare strategies and resources used for the management of water resources and the evaluations of citizens about water supply irregularities and costs of water supply services, thus comparing institutional policies and the citizens' perception.

Conclusions

The integration of disparate data sources always poses methodological challenges and problems of interpretation, but it is still very intriguing and shows positive research perspectives.

REFERENCES

Batty M., Crooks A., Hudson-Smith A., Milton R., Anand S., Jackson M., Morley J. (2010) Data mash-ups and the future of mapping. JISC, Bristol

Burrough P.A. (2001) GIS and geostatistics: Essential partners for spatial analysis. Environmental and Ecological Statistics, 8, 361-377

Claeys C. (2011) Ensuring security whilst keeping access. Principles of security for the European space tools, International Conference Data Flow from Space to Earth, Venice, 21-23 March 2011

Cromley R.G., Tyler Huffman F. (2006) Modeling Situation Factors Used in MCE Procedures for Raster GIS, Transactions in GIS, 2006, 10 (2), pages 239–251

Eastman J.R., Jin W.G., Kyem P., Toledano J. (1995) Raster Procedures for Multicriteria Multiobjective Decisions, Photogrammetric Engineering and Remote Sensing, 61 (5), pages 539-547

Goodchild M., Janelle D.G. (Eds.) (2004) Spatially Integrated Social Science. Oxford University Press, NY

Goodchild M., (2007) Citizens as sensors: web 2.0 and the volunteering of geographic information (Editorial). Geofocus International Review of Geographical Information Science and Technology, Volume 7, pages 8-10

Longley P.A., Goodchild M.F., Maguire D.J., Rhind D.W. (2010) Geographic Information Systems and Science, 3rd Edition. Wiley, New York