Geography as a working desk: a meeting place for data, statisticians and users

Rina Camporese

Key words: georeferencing microdata, data analysis.

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 [2], create new synergic information and enlightens the relations among different phenomena occurring in the same area [1]. 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. 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 have to be invented. Three examples of the use of georeferenced micro-data are illustrated below, (a) to disseminate data, (b) to define reference areas, and (c) to integrate independent sources. The first case (a) is Eye On Earth: a geographic platform where the European Environment Agency shares institutional data and opens them up to the citizens' comments. Thanks to network technologies (cloud computing, mash-up, etc.), data from air and water monitoring stations, collected by institutional agencies, are displayed on a navigable map. Users can also geotag their own perception of air and water quality. Institutions and citizens gather around a table made covered by remote sensing images; the comparison between objective environmental measurements and people's subjective perceptions takes place on a natural representation of the

1 Rina Camporese, Iuav University of Venice – New Technologies & Information on Territory and Environment Research Group; rina.camporese@poste.it

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. The above 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. A second example (b) is based on the idea that statistical units belong to an administrative area by convention: a merely formal link connects them to the institutions governing their area of pertinence. The administrative boundaries are invisible and materialize themselves only at the time of accessing services provided by agencies in their jurisdiction. On the contrary, individuals, and statistical units in general, cross many of these boundaries in everyday life and, furthermore, they locate themselves into administrative areas according to criteria of convenience, not always according to the *de facto* situation. Many statistics are, therefore, prisoners of administrative territorial fragmentation. Furthermore, Modifiable Areal Unit Problem [4] and Ecological Fallacy [3] affect data aggregated by area. Administrative boundaries should be bypassed, when they obscure the spatial distribution of phenomena. Actual technologies can deal with complex data, taking also into account their geographical location. The geocoding of micro-data may bypass the administrative boundaries and allow for greater freedom in defining areas of analysis, e.g. continuous urban areas crossing separate administrative units, proximity / distance from points of interest, etc. By using this kind of geographical perspective, spatial trends could be better highlighted. The third case (c) 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 Istat 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. Environmental noise could be studied by comparing the household proximity to the sources of noise (e.g. congested roads, etc.) with the perception of noise in the area where respondents live. Besides, habits to consume tap water and the reasons for not drinking it could be connected with administrative pieces of information on available water resources 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.

References

1. Goodchild, M.F., Janelle, D.G. (eds.): Spatially Integrated Social Science. Oxford University Press, New York (2004)
2. Longley P.A., Goodchild M.F., Maguire D.J., Rhind D.W.: Geographic Information Systems and Science, 3rd Edition, Wiley, New York, (2010)
3. Robinson, W.S.: Ecological Correlations and the Behaviour of Individuals. American Sociological Review, 15 (3), 351–357 (1950)
4. Unwin, D.J.: GIS, spatial analysis and spatial statistics. Progress in Human Geography, 20, 540-551 (1996)