Scientific report of the EHPS-net workshop of Working Group 9 – GIS, 'GIS and the integration of individual level sociodemographic data within large urban areas', 22 April 2014, Vienna Institute of Demography, Austria

Summary

On 22 April 2014, Working Group 9 – GIS held a workshop 'GIS and the integration of individual level sociodemographic data within large urban areas' co-organized by the EHPS-Net and the Vienna Institute of Demography in Vienna http://www.oeaw.ac.at/vid/ within the Wittgenstein Centre for Demography and Global Human Capital of the Austrian Academy of Sciences. The workshop was partially a pre-meeting to the European Social Science History Conference (ESSHC) and a good opportunity to know other similar works taking place on the other side of the Atlantic.

The main purpose of the workshop was to discuss, compare and develop methods and standards for storage, integration, analyses and visualization of data with multiple spatio-temporal representations for large urban environments. These methods are important for a wide range of applications within social science, geography and epidemiology, and are particular crucial to historical demography. Previous workshops were devoted to similar topics but considering municipal level or rural areas data.

The first aim referred to the integration of historic economic and demographic information with historic and modern geographic information. The purpose of this integration is both to enable analysis and visualization. A specific problem in the integration is that the demographic data is individual level longitudinal data whereas the historic maps are only snapshots of the geographic conditions and the change of space within urban areas is very fast. In a second stage, a further methodological challenge is to integrate time-dependent context data on macro level with the individual level demographic and geographic data. This will permit analyses of the impact of common exposure, bound by geography, on individual outcomes. This for example is the case of the identification of deprivation areas within big urban environments.

A second aim is the visualization of demographic and geographic data, especially concerning the problem of their different time representations. This also was discussed from a technological perspective and solutions and methodological choices on how good graphic presentations can be made with the emerging web standards must be compared.

Related to this issue was also the standardization of storage and distribution of historic demographic information. There has been substantial work and progress internationally during the last years. One mission of the workshop was thus to contribute to the standardization work on integration and distribution of geographic data and historic demographic information.

There were 8 presentations by different groups from Europe, United States and Canada. Several of these presentations dealt with spatial analysis of demographic and economic data: infant mortality in Amsterdam and Madrid, urban land property and economic analysis of individual lots in Montreal or health, environment and sanitation in New York City. Other group of presentations where focused on methodologies for building large GIS infrastructures on a semiautomatic system, like the case of John Logan and his work with 40 US Cities connected to the 1940s Census of the United States, the homogenization of municipal boundaries and population data for long periods of time which was the case of the Iberian peninsula, the use within the semantic web of Dutch historical toponyms or the geocodification of historical addresses in Scotland from 1855 to 1974 linking causes of death information and environmental data within the Scottish Longitudinal Study.

The main purpose of the workshop was to discuss, compare and develop methods and standards for storage, integration, analyses and visualization of data with multiple spatio-temporal representations for large urban environments. We decided to devote this workshop to large urban areas, because they represent, firstly the lion share of population in many of the European countries, and secondly, because the continuous transformation of urban space is more intense than the rural one. On the other hand, multitemporal spatial representations are important for a wide range of applications within social science, geography and epidemiology, and are particular crucial to historical demography. Furthermore, making compatible historical-GIS and historical-populations data with actual GIS and demographic data is one of the main ideals within many endeavours and projects all across Europe. Previous workshops were devoted to similar topics but considering municipal level or rural areas data.

The first aim referred to the integration of historic economic and demographic information with historic and modern geographic information. The purpose of this integration is both to enable analysis and visualization. A specific problem in the integration is that the demographic data is individual level longitudinal data whereas the historic maps are only snapshots of the geographic conditions and the change of space within urban areas is very fast. In a second stage, a further methodological challenge is to integrate time-dependent context data on macro level with the individual level demographic and geographic data. This will allow analyses of the impact of common exposure, bound by geography, on individual outcomes. This for example is the case of the identification of deprivation areas within big urban environments. This is especially relevant within the general ideals of building smart cities where much of this historical administrative, population and GIS information is needed, to measure and asses the changes and social transformation within cities derived from local governments' actions on the urban environment and space.

A second aim was the visualization of demographic and geographic data, especially concerning the problem of their different time representations. This also was discussed from a technological perspective and solutions and methodological choices on how good graphic presentations can be made with the emerging web standards must be compared. This is especially relevant within urban areas where big transformations are taking place all the time, with new neighbourhoods, reshaping of old ones, or rapid urbanization processes in periods of time with rapid population growth and high migration.

Related to this issue was also the standardization of storage and distribution of historic demographic information. There has been substantial work and progress internationally during the last years. One mission of the workshop was thus to contribute to the standardization work on integration and distribution of geographic data and historic demographic information.

The workshop started with a brief introduction by Kees Mandemakers to the main purposes of the EHPS Network and specifically in relation with the working group on GIS and Population. After this small presentation, Alessio Fornasin and Diego Ramiro proceeded with a longer introduction to the workshop explaining the main objectives and aims of this specific workshop and they gave a brief overview of previous meetings with the problems detected and the achievements.

The first presentation on HISGIS Amsterdam - Linking 19th century cadastral maps and population register data: A spatial analysis of socioeconomic differences in infant mortality in Amsterdam by Frans van Poppel and Peter Ekamper focused on the analysis of infant mortality in Amsterdam, focusing in some of the problems detected to accomplish this task. Firstly the linkage of individual data of births and deaths and secondly of the linkage of this data with the HISGIS based on cadastral information from midXIX century Amsterdam.

The second presentation, Geocoding 24 million historical addresses in Scotland from 1855 to 1974, divided in two parts, and presented by Chris Dibben, Zhiqiang Feng and Jamie Pearce, dealt with two aspects. Firstly on the methodologies to geocode historical addresses within the Longitudinal Population Register of Scotland and a second part on the possibilities of linking environmental data to use for the sociodemograhic and epidemiological analysis which could be derived from that study.

The third presentation by Ivo Zandhuis on Dutch historical toponyms in the Semantic Web, explained some of the utilities of the website gemeentegeschiedenis.nl. The website presents a uniquely identifiable web page (a so called "uri") for every municipality in the Netherlands since 1812. Each of these web pages provides internal relations to official spelling alternatives of their names, to

toponyms of set elements within these municipalities, between former and current municipalities, and presents maps of all the geographical changes to the official boundaries of these municipalities.

The forth presentation, From Streetscapes to the Individual Lot: The analytical challenges posed by higher resolution for the 19th-century layers of the MAP Historical GIS based on the city of Montreal and presented by Robert Sweeny showed the potentials for the socioeconomic analysis of land property and urban growth in a big urban environment and at the same time the distribution of commerce and economic activity in the cities.

The fifth presentation by Diego Ramiro Fariñas, HISDI-MAD Historical SDI of the City of Madrid, presented an illustrative example on how to integrate individuals and families to geographic objects. Also this presentation discussed the different time representations and the use of data both for demography and geography to solve some of these issues. The presentation also focused on the visualization and on the technical solutions and the possibilities of standardization within this area. The main focus was on an example analysing infant mortality in Madrid in 1905-1906, suing individual level data from the Longitudinal Historical Population Register of the city of Madrid. There were common problems to the presentation related to the city of Amsterdam.

The following presentation by Luís Nuno Espinha da Silveira on Population and GIS at National and Transnational Level: Achievements and Problems, explained the problems and solutions to build transnational GIS system where there are different municipal or administrative boundaries, as it is the case between Spain municipalities and Portuguese parishes and how to build historical GIS across time from late XVIIIth century to the present. Surpassing the national perspective usually adopted, allows the confirmation of the existence of a pattern of population distribution common to the whole Iberian Peninsula in the long run. The observation of the cross-border region has shown that proximity to the frontier has not generated any distinct pattern of population density on either side of the boundary line.

The final two presentations, dealt with Historical GIS in the USA. John Logan from Brown University presented his massive project on GIS and population for 40 cities in the USA in the 1940s, where HIS-GIS will be built for each of these cities in connection with the 1940 Census from IPUMS. He presented a semiautomatic system for building HIS-GIS based on digital recognition of streets and boundaries and the semiautomatic allocation of geographical characteristics and metadata using actual GIS data. Finally, Carlos Villarreal, from the University of Illinois, presented his work on Building a Common Link to the Past Using HGIS based on his analysis of the city of New York, where he presented different aspects. The effects of environment on land prices and health on an urban environment, especially the case of marsh lands and he did a detailed description of problems on constructing historical GIS based on actual cartography and aerial photography, specifically related to block identification and the importance of introducing a tri-dimensional view of the city for socioeconomic and demographic analysis.

Assessment of the results and impact of the event on the future

There were several areas of discussion within this workshop that can be summarized as follows:

1 - As in previous meetings there was a common agreement of using OGC standards in web-services (standard SDI techniques) and at least for the European case, to follow INSPIRE regulations which solve many of the problems regarding common methodologies and comparability with other works in different European countries and as well as to favour the comparability through time with actual Spatial Data Infrastructures.

2 - The semiautomatic collection of huge amount of data, in some cases for 40 cities at a time, implies the application of new methods with some benefits and limitations. There was discussion about the level of detail and the accuracy needed for the analysis. There are some benefits of a higher level of detail, such as it increases versatility of use but at the cost that are time-consuming to construct. There was an agreement that the importance of accuracy depends on the spatial scale of research questions and that in any case data strengths and limitations must be clearly stated.

3 - There was a consensus of the need for supplementing geographic data linkage through the use of data from other sources, especially economic and environmental data.

4 - There was a consensus on creating common methodologies and sharing experiences between the work done in Europe and in the USA on this topics, with the intention of applying further funding for the comparative study of historical GIS in both sides of the Atlantic.

The final discussion of the result of the workshop identified these as the main points derived from the presentations and discussions. In many cases common approaches to these issues were found and in other cases specificities of data had led to different solutions. In the latter case a continuous discussion on common methodologies and standards were seen as crucial. The North American colleagues expressed their interest on having another meeting in the USA at the time of the next SSHA meeting in Toronto.

Annex 1 - Programme of the meeting

Tuesday, 22 April 2014

08.45 – 09.00 **Registration**

09.00 – 09.15 **Welcome** by Kees Mandemakers, chair EHPS-Net

09.15 – 09.45 **Introduction to the Workshop** by Diego Ramiro and Alessio Fornasin

Session 1. Chair: Kees Mandemakers

09.45 – 10.30 **HISGIS Amsterdam - Linking 19th century cadastral maps and population register data: A spatial analysis of socioeconomic differences in infant mortality in Amsterdam**
Professor Frans van Poppel (NIDI) and Peter Ekamper (NIDI)

10.30 – 11.15 **Geocoding 24 million historical addresses in Scotland from 1855 to 1974**
Professor Chris Dibben (U. Edinburgh) and Dr. Zhiqiang Feng (U. Edinburgh)

11.15 – 11.30 Coffee break

Session 2. Chair: Frans van Poppel

11.30 – 12.15 **Dutch historical toponyms in the Semantic Web**
Ivo Zandhuis, Menno den Engelse and Edward Mac Gillavry

12.15 – 13.00 **From Streetscapes to the Individual Lot: The analytical challenges posed by higher resolution for the 19th-century layers of the MAP Historical GIS**
Professor Sherry Olson (Geography, McGill) and Professor Robert Sweeny (History, Memorial University of Newfoundland)

13.00 – 14.00 Lunch

Session 3. Chair: Alessio Fornasin

14.00 – 14.45 **HISDI-MAD Historical SDI of the City of Madrid**
Dr. Diego Ramiro, Head of Department of Population (CSIC)

14.45 - 15.30 **Population and GIS at National and Transnational Level: Achievements and Problems**
Luís Nuno Espinha da Silveira Associate Professor (U. Nova de Lisboa)

15.30 – 16.00 Coffee break

Session 4. Chair: Diego Ramiro

16.00 – 16.45 **GIS and Population in US cities**

John R. Logan, Professor of Sociology, Director, Spatial Structures in the Social Sciences, Department of Sociology (Brown University)

16.45 - 17.30 **Building a Common Link to the Past Using HGIS**
Carlos Villareal, PhD Candidate, Department of Economics (University of Illinois, Chicago)

17.30 – 18.00 **Final conclusions**

19.30 Dinner

Annex 2 - Full list of speakers and participants

Professor Chris Dibben, Male, chris.dibben@ed.ac.uk
Department of Geography & Sustainable Development, School of Geography & Geosciences, University of St Andrews and Institute of Geography & the Lived Environment School of GeoSciences University of Edinburgh, Geography Building, Drummond Street, Edinburgh EH8 9XP, UK

Professor Jamie Pearce, Male, jamie.pearce@ed.ac.uk
Department of Geography & Sustainable Development, School of Geography & Geosciences, University of St Andrews and Institute of Geography & the Lived Environment School of GeoSciences University of Edinburgh, Drummond Street, Edinburgh EH8 9XP, UK

Dr. Zhiqiang Feng, Senior Lecturer, Male, zf2@st-andrews.ac.uk
Department of Geography & Sustainable Development, Irvine Building, University of St Andrews, North Street, St Andrews, KY16 9AL Fife, Scotland, UK

Dr Konstantinos Daras, Research Fellow, Male, kd54@st-andrews.ac.uk
Department of Geography & Sustainable Development, Irvine Building, University of St Andrews, North Street, St Andrews, KY16 9AL Fife, Scotland, UK

Associate Professor Luís Nuno Espinha da Silveira, Male, luis.silveira@netcabo.pt
Department of History, Contemporary History Institute, Faculdade de Ciências Sociais e Humanas Universidade Nova de Lisboa, Avenida de Berna, 26-C / 1069-061 Lisboa

Professor Kees Mandemakers, chair of EHPS-net, Male, kma@iisg.nl
International Institute of Social History, Amsterdam en Erasmus University Rotterdam, Erasmus School of History, Culture and Communication, P.O. Box 2169, 1000 CD Amsterdam. Cruquiusweg 31, 1019 AT Amsterdam, The Netherlands

Professor Frans van Poppel, Male, poppel@nidi.nl
Nederlands Interdisciplinair Demografisch Instituut, Lange Houtstraat 19, NL-2511 CV The Hague, The Netherlands

Dr. Peter Ekamper, Male, ekamper@nidi.nl
Nederlands Interdisciplinair Demografisch Instituut, Lange Houtstraat 19, NL-2511 CV The Hague, The Netherlands

Ivo Zandhuis, Independent researcher, Male, ivo@zandhuis.nl
Rijksstraatweg 79, 2024 DB Haarlem, The Netherlands

Dr. Alessio Fornasin, Risercatore Universitario, Male, alessio.fornasin@uniud.it
Department of Economics and Statistics, University of Udine via Tomadini, 30/A I-33100 Udine, Italy

Dr. Diego Ramiro, Head of Department of Population, Male, diego.ramiro@cchs.csic.es
Department of Population, Institute of Economics, Geography and Demography, Spanish National Research Council, calle Albasanz 26-28, Madrid 28037, Spain

Dr. Bárbara Revuelta, Postdoctoral Researcher, Female, barbara.revuelta-eugercios@ined.fr
Institut national d'études démographiques INED**, 133 boulevard Davout - 75020 Paris, France**

Sara García Ferrero, PhD. Female, sara.garcia@cchs.csic.es
Department of Population, Institute of Economics, Geography and Demography, Spanish National Research Council, calle Albasanz 26-28, Madrid 28037, Spain

Professor Robert Sweeny, Male, rsweeny@mun.ca
Department of History, Memorial University of NewfoundlandLost in Dingle, St. John's, NL A1C 5S7, Canada

Professor Sherry Olson, Female, sherry.olson@mcgill.ca
Dept. of Geography, McGill University, Room 611, Burnside Hall, 805 Sherbrooke St. W. Montreal, QC, H3A 2K6, Canada

Dr. Ramon Bauer, Male, ramon.bauer@oeaw.ac.at
Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/ÖAW, WU), Vienna Institute of Demography (Austrian Academy of Sciences), Wohllebengasse 12-14, 6th floor, 1040 Vienna, Austria

Markus Speringer, PhD., Male, Markus.Speringer@oeaw.ac.at
Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/ÖAW, WU), Vienna Institute of Demography (Austrian Academy of Sciences), Wohllebengasse 12-14, 6th floor, 1040 Vienna, Austria

Professor John R. Logan, Professor of Sociology, Director, Spatial Structures in the Social Sciences,Department of Sociology, Male, John_Logan@brown.edu
Department of Sociology, Brown University, Providence, RI 02912, USA

Carlos Villareal, PhD Candidate, Male, cvilla22@uic.edu
Department of Economics, University of Illinois at Chicago, 1740 W Taylor St, Chicago, IL 60612, USA