Scientific report of the EHPS-net workshop of Working Group 9 – GIS,
'Integrating time, space and individual life stories',
3-4 July 2015, University of Edinburgh,
BioQuarter building, Edinburgh, UK

Summary

In the EHPS-network, several projects and countries have created demographical databases following the Intermediate Data Structure (IDS). EHPS-Net aims at converting a large number of longitudinal historical databases into this common structure. Data extraction programs will thereafter be developed to transform the data into usable datasets for longitudinal analysis. Furthermore, metadata and vocabularies specifying the terminology will be created, which will help the interpretation of data from different countries and communities. Researchers as well as the public will therefore get access to data in a standardized structure explained by metadata, which opens up for more comparative studies.

Despite the importance of space in longitudinal demographic analysis, few spatial analyses have been carried out. The main reason is the lack of geographical information linked to the historical demographic data. Having access to this information may substantially benefit the research. Geographical information is thus important to consider when developing the IDS. To enable the linkage between IDS databases and other geographical information sources, we need to store geometrical representations of geographic objects in IDS. This could for instance be storing a geographic representation of buildings and property units. When individuals are linked to these buildings and property units it will be possible to investigate the environment where these people lived.

Several projects across Europe are working (or plan to work) with linking digitized geographical objects from maps with geographical locations in demographic databases. One method has been to use text labels in the historic maps as well as supplementary text sources to these maps. A limitation with the current solution is that the connection between the demographic and geographic data has only been made on an aggregated level (in terms of individuals). To enhance the possibilities for analysis, a connection on individual level (that each person in the demographic databases is linked to one or several buildings or property units) is also needed. However, this linkage requires methodological development because the two sources by nature often use different time representations. The historical maps can be regarded as snapshots of the conditions at a certain time whereas the demographic database is using continuous object lifelines (a person exists from birth to death).

There were 9 presentations by different groups from Europe and United States. Several of these presentations dealt with spatial analysis of demographic and socioeconomic data: infant mortality in Madrid, historical vital events in Edinburgh and immigration 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 Ian Gregory and his work on methodologies for analysing unstructured texts—including large corpora of historical books, periodicals and official reports—within a GIS environment 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.

Scientific content and discussions at the event

On 3-4 July 2015, Working Group 9 – GIS held a workshop 'Integrating time, space and individual life stories' co-organized by the EHPS-Net and the Edinburgh University in Edinburgh within the BioQuarter building. The workshop was a good opportunity to know other similar works taking place in Europe along with the United States.

In previous EHPS-Net workshops have discussed methods and standards for storage, integration and visualization of data with multiple spatio-temporal representations. This workshop builds on the topics from the previous workshops moving from data integration and storage looking towards research use of the data, for example from boundaries to landownership and households to the application for research. That is assuming that you have got these type of variable available how to associate theses with the historical geographies, what types of area-level variables can be added in for the analysis and how to link these in as often the volume is too big and there is a need to automate the process. These are the variables and measures which can add value to research or analysis (ie localized estimates of pollution, poverty, crime, etc).

In this workshop two main topics have been discussed and explored:

1. Computer science techniques for automatically geocoding historical events and records.
2. Methods for deriving and research uses of historical 'small area' geographical variables.

In other words automation and estimation in historical geographical informatics.

With increasing amounts of transcribed historical information on individuals becoming available, the ability of a research group to geocode events such as place of birth or death, or location during a census, 'by hand', is becoming increasingly untenable. Linkage to larger scale geographical units have been possible through the geocoding administrative units, such as parishes or registration districts, but this does not always produce the fine grain geographical location necessary for certain type of studies. If for example a researcher is interested in records for individuals living near a point of potential contamination (e.g. a contaminated water source) or source of pollution (e.g. a factory), then more precise geolocation data is needed.

Modern computer science techniques may offer help. Probabilistic record matching techniques, allow address information, of variable quality (i.e. with transcription errors or varying spelling), to be matched to geocoded lookup tables. Natural Language Processing (NLP) allow address information be extracted from free text, such as newspaper articles. Some techniques are still being developed. Optical Character Recognition may be used to identify place names from geotagged scanned image of maps. Techniques used for remote sensing may be used to automatically classify areas on maps based on an interpretation of their graphic qualities (e.g. colours). These types of questions have been addressed in this workshop.

Once records are geocoded, the possibility of linking area and environment data to them arises. This then allows questions of impact and effect to be asked. However unlike contemporary research situations, for historical data there is rarely readily available area or environment data. Instead this data to needs to be estimated from a variety of historical sources. So for example air pollution levels needs to be estimated from information on coal sales in local areas. The second theme for the

workshop was methods and examples of the estimation of area or environmental data and its use in historic micro data research.

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 9 on historical GIS and Population. After this small presentation, Chris Dibben 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 *Historical address geocoder GIS (HAGGIS) - An advanced system for geocoding historical addresses in Scotland* by Konstantinos Daras focused on the methodology for geocoding 24 million historical addresses in Scotland from 1855 to 1974 by introducing the matching algorithms and open tools implemented for the needs of the Digitising Scotland project. Apart from geocoding the historical addresses, the pilot historical addresses from death event records in 1950 and 1951 was used for producing pseudo registration district boundaries. This way the new historical geographical dataset creates new insights into the Scottish geographies of the past.

The second presentation, *Life course of place: using GIS to integrate longitudinal environmental measures with Scottish cohort data* presented by Catherine Tisch, dealt with the epidemiological Life course study, specifically on how the GIS tools have been used to integrate longitudinal environmental measures with Scottish cohort data for exploring how physical, built and social environments evolve over time and how they impact on inequalities in health-related mobility as people move into older age.

The third presentation by Eric Grosso and Richard Roger on *Mapping Edinburgh's Social History (MESH)*, explained some of the 'open source' tools – free available to all members of the public – implemented for mapping historical data and creating a new online Atlas of Edinburgh 1000-2000 AD based on thematic historical maps. The final product will be an open digital data resource for historians and the general public, these will make it possible for anyone to map historical data.

The forth presentation, *Automated approaches to understanding the geographies in historical texts* presented by Ian Gregory showed a step-change in how place, space and geography are explored in the Humanities focused on methodologies for analysing unstructured texts—including large corpora of historical books, periodicals and official reports—within a Geographic Information Systems environment.

The fifth presentation by Jesper Zedlitz, *Place identification in historic sources*, presented a place identification approach for implementing an online Historical Gazetteer that provides comprehensive historic administrative information for a correct and accurate identification elaborating web interface and Linked Open Data access.

The following presentation by Diego Ramiro Fariñas on *the spatial distribution of childhood infectious disease in Madrid in early XXth Century*, focused on the visualization and on the technical solutions and the possibilities of standardization within an illustrative example on how to integrate individuals and families to geographic objects. 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.

The seventh presentation, *"Kleindeutschland," the Lower East Side in New York City at Tompkins Square in the 1880s: Exploring Immigration Patterns at the Street and Building Level* presented by Kurt Schlichting showed the spatial distribution of immigrant population in New York City in 1880 focused on Lower East Side – ten block area to the south of Tompkins Square – and examined residential patterns by block and building occupied by German immigrants.

The eight presentation by Finn Hedefalk on *the influence of micro-level geographic factors on mortality in southern Sweden, 1813-1914* presented methods for building longitudinal geographic and demographic databases on the micro-level focusing on the influence of the soil types spatial coverage on mortality in southern Sweden.

The final presentation, dealt with Historical GIS in the France. Eric Grosso presented *the ALPAGE project: Paris and its urban area at the intersection of history and geography (9th – 18th century)*, where aims at producing open data and tools to understand the long term relationships between spaces and societies in Paris. The goal of this project is to study the urban morphology on a long term period in order to show that social practices have an effect on the urban morphology in addition to the classical explanations of planning: the way people lived in spaces, their needs and what they did, can partly explain the maintenance of some shapes of roads and parcels.

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 huge amount of historical data used in the majority of the works presented in this workshop imply 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 socio-economic and environmental data.
4. There was a discussion on the continuation of developing open source tools and sharing experiences between the work done in Europe and in the US on methodologies such as address geocoding and data linkage with the intention of investigating further the benefits and limitations for the studies 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.

Annex 1 - Programme of the meeting

Friday 3rd July

12:00 – 13:00 Lunch

13:00 - 13:10 Welcome by Kees Mandemakers, chair EHPS-Net

13:10 - 13:20 Introduction to the topic by Chris Dibben, Zhiqiang Feng & Lee Williamson

13:20 - 14:05 Presentation group 1 (30mins) followed by discussion (15mins)
Konstantinos Daras, *Historical address geocoder GIS (HAGGIS) - An advanced system for geocoding historical addresses in Scotland*

14:05 - 14:40 Presentation group 2 (20mins) followed by discussion (15mins)
Catherine Tisch, *Life course of place: using GIS to integrate longitudinal environmental measures with Scottish cohort data*

14:40 - 15:15 Presentation group 3 (20mins) followed by discussion (15mins)
Eric Grosso / Richard Roger, *Mapping Edinburgh's Social History (MESH)*

15:15 - 15:35 Coffee

15:35 - 16:15 Presentation group 4 (25mins) followed by discussion (15mins)
Ian Gregory, *Automated approaches to understanding the geographies in historical texts*

16:15 - 16:50 Presentation group 5 (20mins) followed by discussion (15mins)
Jesper Zedlitz, *Place identification in historic sources*

16:50 - 17:00 Day 1 closing remarks Diego Ramiro Fariñas, chair of EHPS-Net Working Group 9 GIS

18:30 Workshop dinner

Saturday, 4 July 2015

10:00 - 10:35 Presentation group 6 (20mins) followed by discussion (15mins)
Diego Ramiro Fariñas, *The Spatial distribution of childhood infectious disease in Madrid in early XXth Century*

10:35 - 11:15 Presentation group 7 (25mins) followed by discussion (15mins)
Kurt Schlichting, *"Kleindeutschland," the Lower East Side in New York City at Tompkins Square in the 1880s: Exploring Immigration Patterns at the Street and Building Level*

11:15 - 11:35 Coffee

11:35 – 12:15 Presentation group 8 (25mins) followed by discussion (15mins)
Finn Hedefalk, *The influence of micro-level geographic factors on mortality in southern Sweden, 1813-1914*

12:15 - 12:50 Presentation group 9 (20mins) followed by discussion (15mins)
Eric Grosso / Laurent Costa / Sandrine Robert, *Paris and its urban area at the intersection of history and geography (9th – 18th century) - ALPAGE project*

12:50 - 13:10 Final conclusions Diego Ramiro Fariñas, chair of EHPS-Net Working Group 9 GIS

13:10 - 14:30 Lunch

Annex 2 - Full list of speakers and participants

Professor Michael Anderson, Honorary Professorial Fellow, Male, m.anderson@ed.ac.uk
School of History, Classics and Archaeology, William Robertson Wing, Old Medical School, Teviot Place, Edinburgh, EH8 9AG, Scotland.

Associate Professor Peter Christen, Male, peter.christen@anu.edu.au
Research School of Computer Science, College of Engineering and Computer Science, CSIT Building (108), North Road, The Australian National University, Canberra ACT 0200 Australia

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

Dr. Joseph Day, Research Associate, Male, jd466@cam.ac.uk
Cambridge Group for the History of Population and Social Structure, Sir William Hardy Building, Department of Geography, Downing Place, CB2 3EN

Professor Chris Dibben, Chair in Geography, Male, chris.dibben@ed.ac.uk
Research Institute of Geography & the Lived Environment, School of GeoSciences, University of Edinburgh, Geography Building, 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. Eilidh Garrett, Research Fellow, Female, eilidh.garrett@btinternet.com
Department of History, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK.

Bruce M Gittings, Director GIS Programmes, Male, bruce@ed.ac.uk
Institute of Geography and the Lived Environment, School of GeoSciences, University of Edinburgh, Geography Building, Drummond Street, Edinburgh, EH8 9XP, UK

Professor Ian Gregory, Professor of Digital Humanities, Male, I.Gregory@lancaster.ac.uk
Department of History, Lancaster University, Lancaster, LA1 4YT, UK

Eric Grosso, MESH research and technical officer, Male, eric.grosso@ed.ac.uk
Mapping Edinburgh's Social History (MESH) project, School of History, Classics and Archaeology, University of Edinburgh, William Robertson Wing, Old Medical School, Teviot Place, Edinburgh, EH8 9AG, UK

Dr Masih Hajiarabderkani, Postdoctoral Research Fellow, Male, masih.h.derkani@st-andrews.ac.uk
School of Computer Science, Jack Cole Building, North Haugh, University of St Andrews, St Andrews, Fife KY16 9SX, Scotland.

Finn Hedefalk, PhD Candidate, Male, finn.hedefalk@nateko.lu.se
Deptartment of Physical Geography and Ecosystem Science, Lund University.

Zengyi Huang, Database Manager, Male, zengyi, huang@ed.ac.uk
Longitudinal Studies Centre - Scotland, School of GeoSciences, University of Edinburgh, Ladywell House, Ladywell Road, Edinburgh EH12 7TF, UK

Dr Graham Kirby, Senior Lecturer, Male, graham.kirby@st-andrews.ac.uk
School of Computer Science, University of St Andrews, North Haugh, St Andrews, Fife KY16 9SX, Scotland

Mr. Owen Macdonald, GIS Support Officer, Male, owen.macdonald@ed.ac.uk
School of GeoSciences, University of Edinburgh, Geography Building, Drummond Street, Edinburgh EH8 9XP, UK

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

Sophie McCallum, MESH project administrator, Female, sophie.mccallum@ed.ac.uk

Mapping Edinburgh's Social History (MESH) project, School of History, Classics and Archaeology, University of Edinburgh, William Robertson Wing, Old Medical School, Teviot Place, Edinburgh, EH8 9AG, UK

Dr Mary Murphy, Adjunct Professor of English, Female, mmurphy@fairfield.edu
Dept. of English, Fairfield University, 1073 North Benson Rd. DMH 139, Fairfield, CT 06824-5195 U.S.A.

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 Alice Reid, Lecturer in Human Geography, Female, amr1001@cam.ac.uk
Cambridge Group for the History of Population and Social Structure, Department of Geography, Downing Place, Cambridge CB2 3EN, UK

Dr Corinne Roughley, Affiliated Lecturer of the Dept. of Archaeology & Anthropology, Female, cfa1001@cam.ac.uk
Division of Archaeology, Downing Street, University of Cambridge, Cambridge, CB2 3DZ

Dr. Mattias Sandström, Research Investigator, Male, mattias.sandstrom@umu.se
Centre for Demographic and Ageing Research, Demographic Data Base, Umeå University, SE-901 87 Umeå, Sweden

Dr. Kurt Schlichting,E. Gerald Corrigan '63 Chair in Humanities and Social Sciences Professor of Sociology, Male, Kurt@fairfield.edu
Dept. of Sociology & Anthropology, Fairfield University, 1073 North Benson Rd. DMH 216, Fairfield, CT 06824-5195 U.S.A.

Catherine Tisch, Research Assistant, Female, catherine.tisch@ed.ac.uk
Research Institute of Geography & the Lived Environment, School of GeoSciences, University of Edinburgh, Drummond Street, Edinburgh EH8 9XP, UK

Dr Lee Williamson, Research Fellow, Female, lee.williamson@ed.ac.uk
Research Institute of Geography & the Lived Environment, School of GeoSciences, University of Edinburgh, Drummond Street, Edinburgh EH8 9XP, UK

Dr Jesper Zedlitz, Wissenschaftlicher Mitarbeiter (Scientific Assistant), Male, j.zedlitz@email.uni-kiel.de
AG Kommunikationssysteme, Inst. f. Informatik, CAU Kiel, Germany.