Information and Communication Technologies

EPIWORK

Developing the Framework for an Epidemic Forecast Infrastructure

http://www.epiwork.eu

Project no. 231807

D2.1 Implementation of multi-scale proxies networks for human mobility networks from on-line data sources

Period covered: months 1st - 12th
Start date of project: February 1st, 2009

Date of preparation:
Duration:
Actual submission date: February,

Due date of deliverable: month 12th 8th 2010
Distribution: public Status:

Project Coordinator: Alessandro Vespignani
Project Coordinator Organization Name: ISI Foundation
Lead contractor for this deliverable: Max Planck Gesellschaft zur Foerderung der Wissenschaften E.V.

2010 © Copyright lies with the respective authors and their institutions

Work package participants

The following partners have taken active part in the work leading to the elaboration of this document, even if they might not have directly contributed writing parts of this document:

- MPG
- ISI
- FGC-IGC
- TAU
- BIU
- FBK

Change log

Version	Date	Amended by	Changes
1	**20/02/10**		

Table of Contents

List of figures

2010 © Copyright lies with the respective authors and their institutions

1 Database Implementation

As part of this deliverable, a dedicated high performance MySQL database server was instantiated and runs on a MacOS Server platform. This server manages and maintains large-scale databases of records of mobility related data employed in this project. The server is integrated in a backup system and guarantees data integrity. The server is operated on an user/password basis and can be accessed, updated and managed via TCP/IP protocol from anywhere and from all the participating partners, easing access to teams of the project. In order to manage mobility data from heterogeneous sources mobility specific data syntax was developed with which all the datasets comply. This will render easier the integration of additional datasets into the system as well as the development of analysis tools and software that use the database by implementing an automatic access to the server.

So far three large-scale datasets have been integrated:

1. **Proxy data for mobility based on the geographic circulation of bank notes.** This data was obtained from the online bill tracking website www.wheresgeorge.com and contains the geographic flux of money in the United States, i.e. geographic trajectories of more than 11 million single bank notes in the United States. This data is continually updated. The spatial resolution is that of United States zip codes.
2. **The worldwide air transportation network.** This dataset contains almost all the passenger numbers of all the flights between approx. 3000 airports worldwide over a period of three years. This dataset represents the long distance component of multi-scale mobility networks that are being analyzed in this project.
3. **GPS data on trackable items in geocaching.com**. This dataset was obtained from the international geo-referencing game known as geocaching, containing geographic trajectories of more than half a million GPS trackable items. In this online game, the GPS location as a function of time for trackable items is monitored and reflects a proxy from human mobility on all spatial scales and across international borders.

The raw data obtained from these sources was processed, compiled and embedded the database server. The server is maintained and managed by the MPI-DS team. We are in the process of finishing a web based GUI that will be made available to the project partners for issuing queries to the database and obtaining multi-scale mobility networks that can be incorporated into modeling frameworks.

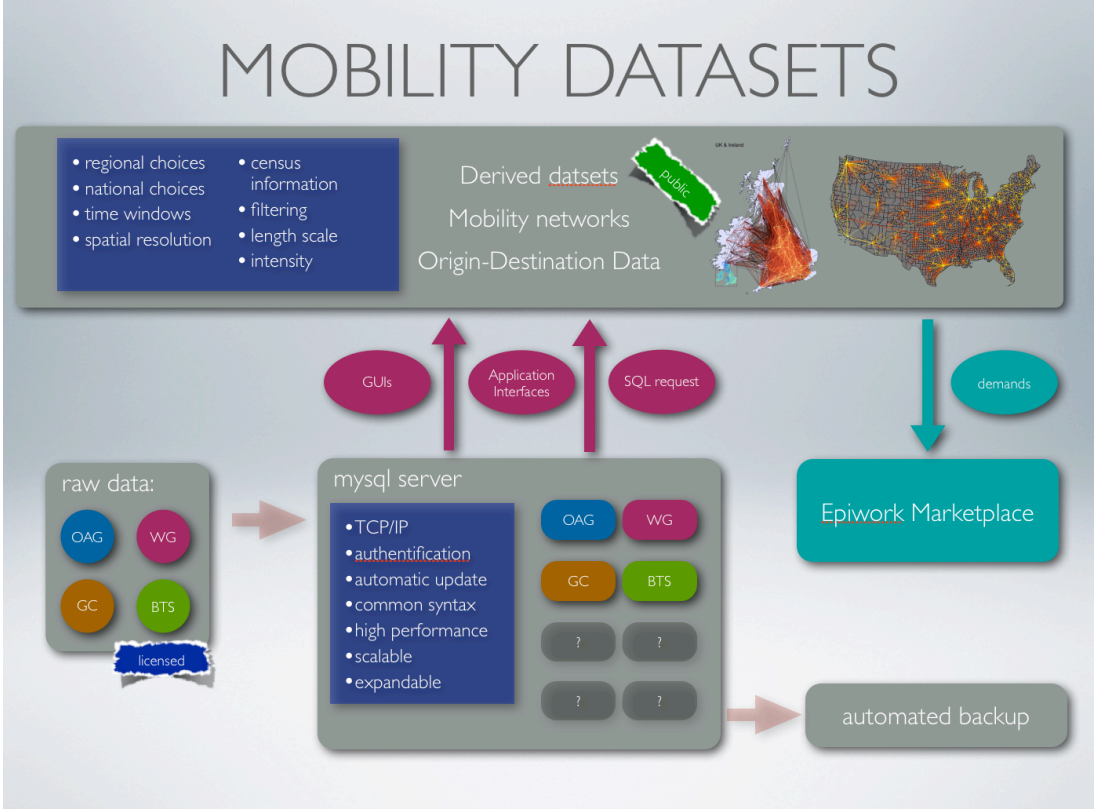

Figure 1: illustration of the structure of the mobility database server and its operation within the project.

The implementation of the database server and its current service of the above initial datasets reflect the accomplishment of deliverable D2.1. We are currently in the process of development of GUI and application interfaces though which the database can be accessed by other components of the project, for instance the Epiwork Marketplace.

Key to the design of the database and the common syntax for datasets managed by the server is a design that provides maximal flexibility in terms of regional and national parameters and the spatial resolution underlying derived origin destination data

2010 © Copyright lies with the respective authors and their institutions

structures requested from the server as well as important census related data on the specified spatial resolution.

The server is already in live operation and played a key role in the efficient and successful accomplishment of the research described in Theme I of work package 2, that is the study of the structure of human mobility networks, describe in the following section.

2 The structure of multi-scale mobility networks

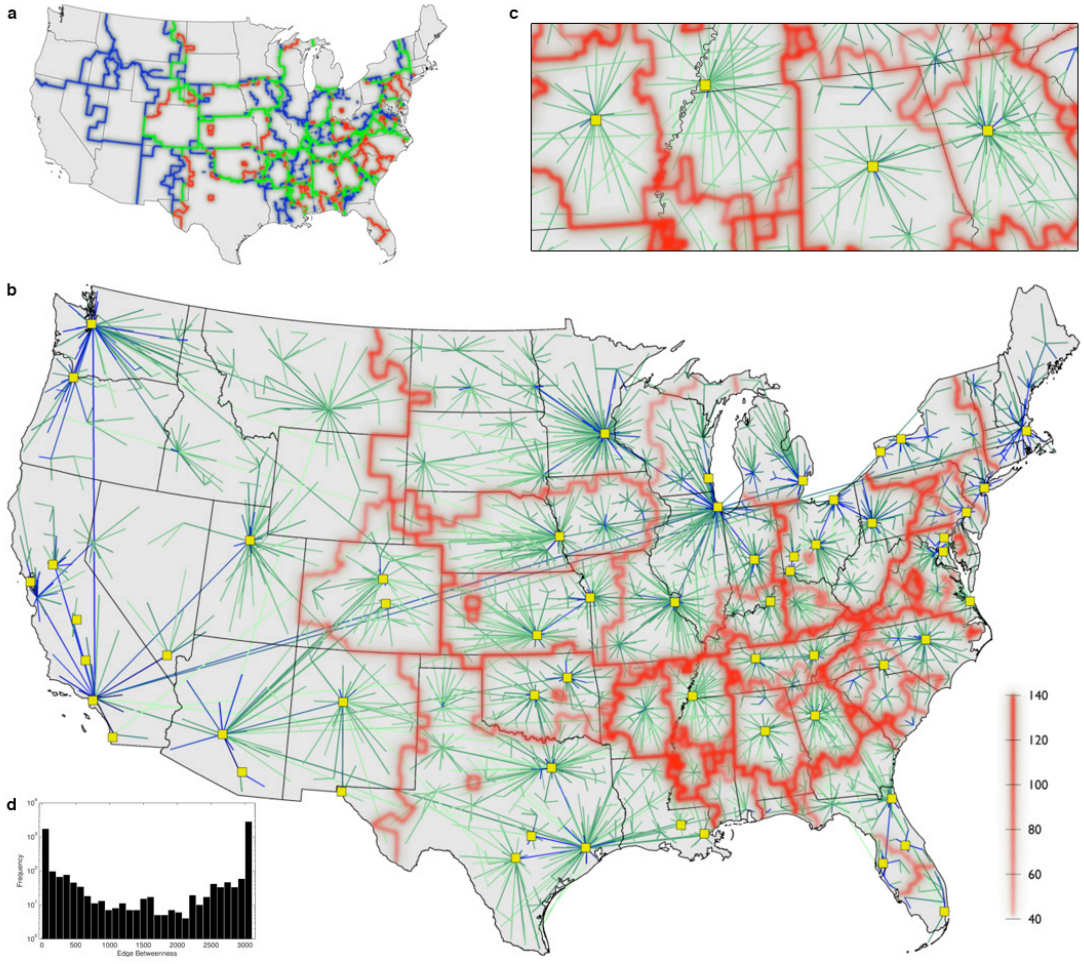

Figure 2: effective geographic borders implicitly encoded in multi-scale mobility networks (red). The discovery and computation of these structures heavily relied on the use of the mobility database server described above.

Considerable progress was made concerning the computation of multi-scale community structure and effective geographic borders as described in Task 2, Theme 1 of the work package. We used a multi scale proxy network obtained from the

geographic circulation for the United States. In this project and while performing the research, we made intense use of the database server that greatly expedited the analysis. In addition to existing methods for community detection in complex networks based on network modularity maximization, we developed a new efficient technique based on the analysis of sets of shortest path trees in the network. The novel approach not only permits the geographic identification of locations of effective borders but also provides a means to quantify their significance and strength. Furthermore it revealed which parts of multi-scale mobility networks are responsible for the shape and location of effective borders. The results of this project are currently under revision at Nature, a preprint of a manuscript is available at http://arxiv1.library.cornell.edu/abs/1001.0943v1. The results represent a promising base for investigations of multi-scale human mobility networks and effective order in Europe, a task that we plan to accomplish next within work package 2 based on proxy data for Europe and direct data on human mobility. Once multi-scale mobility data for Europe is incorporated in the database server, the analysis for Europe can be accomplished almost automatically.

2010 © Copyright lies with the respective authors and their institutions