

DISSEMINATING DISTRIBUTED STATISTICAL DATA: THE ADDSIA PROJECT

Dr Joanne Lamb
University of Edinburgh Scotland, UK

ADDSIA (Access to Distributed Data for Statistical Information and Analysis) is a project funded by the European Union to develop a web-based system to aid the analysis of heterogeneous statistical datasets held at different locations. The impetus for developing this application is the need to compare similar data from the member states of the European Union. The approach can be applied in different settings, to combine data collected in different locations, or data constructed using different physical, logical or conceptual structures.

ADDSIA adopts a hierarchical approach, assuming that only data in a given Domain can be sensibly combined for analysis. A Domain server holds the top level information about domain in question (e.g. Labour Force Surveys, Trade statistics, etc). Providers of data register with the Domain, and provide access to their data. The accompanying metadata is provided in XML format. Users log on to the domain server via a web browser in order to:

- browse descriptions of available data
- recode the data, and apply statistical analysis
- obtain tables for information
- request tables for further publication.

The system relies on a hierarchical model which decomposes statistical queries before sending the relevant subqueries to datasets at their source location. The results of these sub queries are transmitted to the domain server, where they are combined to give required answer. The metadata supplied to the system allows the analysis module to combine responses from heterogeneous datasets.

From the administrative point of view, there are three major tasks – the registering of domain and provider metadata, the registration of users, and the registration of datasets.

The demonstration will show how users can access the system via a standard browser. Access to a Domain Server is provided the ADDSIA home page. The user browses the descriptive metadata to identify the datasets of interest. He can then choose variables from these datasets, and carry out any recodes he requires via a graphical interface. The recodes, including his reasons for them, are stored in his workspace. He then requests a table built out of these recodes. The request is parsed by the Domain, sent to the Providers, who access the native data. The result is assembled at the Domain level, and presented to the user in the style he has requested. Two domains are available in the demonstration, one covering survey data and the other covering timeseries.

Contacts

<http://www.ed.ac.uk/~addsia>

or

email: addsia@ed.ac.uk