Weaving the Web of European social science

Jostein Ryssevik (Keynote Speaker)
Director of Technology and Development, Nesstar Ltd.

In the late 1950s Dr. J.C.R Licklider observed that most of his time as a researcher was spent on getting into a position to think, and not on creative thinking as such. “Much more time went into finding or obtaining information than into digesting it.” (see Howard Reingold: “Tools for Thought – The History and Future of Mind-Expanding Technology”, The MIT Press, Cambridge Massachusetts 2000, p133). A few years later Licklieder became the director of ARPA, the research organization that initiated the forerunner to todays Internet, the ARPAnet.

Licklieder’s observation might be seen as a general justification for the development of any research infrastructure, including the Internet. The overriding goal of a research infrastructure is to facilitate the maximization of the time spent on digesting and thinking over the time spent on finding and accessing.

However, even today nearly 50 years after Licklider’s observation and about 10 years after the invention of the World Wide Web, comparative social science research in Europe is hampered by the fragmentation of the scientific information space. Data, information and knowledge are scattered in space and divided by language and institutional barriers. As a consequence too much of the research are based on data from a single nation, carried out by a single-nation team of researcher and communicated to a single-nation audience. The state of affairs is preventing the development of a comparative and cumulative research process integrating and nurturing the entire European Research Area.

Yesterday’s answers to these challenges would probably have been formulated in terms of centralization and establishment of large-scale European-wide institutions. Today’s answers should rather focus on the power of emerging information technologies to encourage communication, sharing and collaboration across spatially dispersed but scientifically related communities.

What we would like to see is the research infrastructures that could facilitate the following scenarios:

Scenario 1

A user is looking for data on political trust from three different regions of Europe. He uses the geographical interface to circle in the relevant regions on the map and enters the additional search criteria. From the returned hit list he is able to browse layers of increasingly detailed metadata describing potential sources. He is even allowed to perform simple statistical analysis and visualizations on-line to make sure that the data fulfills his requirements. Several datasets can be brought to the desktop at the same time to ease the comparison. As soon as a decision has been made, the chosen datasets can be downloaded and automatically converted to the format of his favorite statistical package. All relevant metadata travels along with the data to assist the researchers in his analysis.

Scenario 2

A researcher analyzing a group of variables in dataset X would like to know if there are similar datasets from other countries that could be used for a comparative study. By hitting the “get comparable dataset" button, a list of potential datasets is immediately returned. By fine-tuning the search criteria to make sure that all datasets fulfills her methodological requirements she is able to circle in on a handful of sources that might be used.

The researchers would also like to have an overview of knowledge products (papers, articles etc.) that are based on these studies and even to browse these objects if they are available on-line. As references and links to derived knowledge products are an integrated part of the metadata of each single dataset a sorted list can be displayed by a single keystroke. Some of these references are also including e-mail and website-addresses to the relevant researchers.

Finding a problem with one of the variables, the researcher writes a note and appends it to the "user experience-section" of the metadata to alert future users about the quality of the data (she also leaves her e-mail address to allow them to contact her). And when the research paper is ready and published in an on-line journal, links to the dataset is added to allow future users to re-visit her analysis

Scenario 3

A researcher who is reading an article in an on-line journal finds a link that connects him to the data that was used by the author to underpin the arguments. By following the link, he is able to load the dataset and to rerun the original analysis. He is even allowed to dig deeper into the same data-source, testing alternative indicators or models. The system is also making him aware of several other comparable data sources published after the article was written and he uses these sources to challenge the conclusion of the author. Links to knowledge products based on these newer data sources is also available. From one of the sources he is even brought to a mail-list that discusses the phenomena in further detail

Scenario 4

A user is looking at a table showing variation in nationalistic attitudes among different educational groups in country X. Through the systems integrated multilingual thesaurus service he is able to pick up the relevant key-words describing this table and to automatically create a multilingual query for datasets that might be used to create comparable tables. He is also leaving the query with the system's "digital research assistant" (an active agent), to make sure that he is alerted by e-mail if a new dataset meeting his requirements is published somewhere in Europe at a later stage. He even ask the agent to look for knowledge products addressing the same topics.

The paper will outline the visions, requirements and architecture of a virtual research infrastructure that would allow these dreams to become true. It will also describe a real life implementation of these visions, the NESSTAR (Networked European Social Science Tools and Resources) system. All of the functionality described in these scenarios are technically feasible using the Nesstar software platform.

Nesstar has been developed according to the following "dream-list":

- all existing empirical data available on-line
- an integrated resource discovery gateway in order to identify and locate relevant resources (including the relevant tools to overcome the language barriers)
- extensive amounts of metadata available (multi-media and integrated with the data)
- the ability to carry out simple browsing, visualisation and analysis of the data on-line
- the ability to subset and download the data and metadata to a favourite analysis too
- „active research agents“ mining the net and informing user when new resources are available
- efficient hyperlinks from the metadata to every relevant report and publication
- current e-mail/web addresses to all relevant researchers, support staff, departments etc.
- an efficient feedback system to the body of metadata allowing the user to add to the collective memory of a data source

This technology is currently being used to develop web-based data services by a variety of data archives, libraries and providers world wide. In Europe, the Council of European Social Science Data Archives (CESSDA) has decided to use the platform to develop a portal (integrated catalogue) to the data resources of all the European data archives. An EC-funded project working toward this goal (MADIERA) will be started up this autumn.

Nesstar might be seen as an example of a new type of application providing the building blocks of what has lately become known as the Semantic Web. These are metadata rich applications where all relevant information is given well-defined meaning, making it easier for computers and people to work in cooperation. Agreed metedata standards are the glue of the Semantic Web, and the most important message form this paper is the need to develop flexible and extensible metadata standards, which will facilitate smooth interoperation across systems and domains.

Contact Information

Jostein Ryssevik
Director of Technology and Development
NESSTAR Ltd.
Norwegian Social Science Data Services (NSD)
Hans Holmboesgt. 22
N-5007 Bergen
Norway

Tel: +47 5558 2654
Fax: +47 5558 9650
e-mail: Jostein.Ryssevik@nsd.uib.no
http://www.nesstar.org