Mission and Vision: The Dynamics of Social Research Data Infrastructure

Ekkehard Mochmann
(E.Mochmann@web.de)
Cologne

50 Years Social Research
80 Years Niko Tos

Ljubljana Dec. 4, 2014

A Large Scale Facility for the Catholic Church

Begun 1248

completed 1880

At about same time (1870 +):

Social Conflicts

1872 Verein für Socialpolitik; Kathedersozialisten

1876 First nation wide survey: 7000 respondents

commissioned by German Federal Council on Feb. 19, 1875

Topic: Relations between employers and workers

- Administered by civil servants in 559 communities

Data were analysed on descriptive level: marginal distributions and tables

Max Weber- over the years involved in 6 empirical research projects- made strong efforts to improve Questionnaire design and question formulation and to develop data analysis further

Why did Max Weber fail?

Profession, bureaucracies, funders skeptical and not interested in new methods

Working against business as usual:
Principle of Creative destruction

Lesson: Create interest and visibility for potential of new approaches

same time:

1870 – 1890 Daimler and Otto develop fuel driven engines

1889 First fuel engine driven car by Gottlieb Daimler and Maybach presented in Paris

1892 DMG -Daimler Motor Company sold its first Car

Empirical research in Europe faded at beginning of 20th century
In US:
Gallup: social surveys at mass level
Kish: representative samples

ISR Michigan 1949
ICPR 1962

Mission: develop instruments and provide world wide guidance

Changing the paradigm

Europe: Starting in the late 1950ies:
Stein Rokkan, Erwin K. Scheuch , Warren Miller
Alexander Szalai with ISSC, Paris

Standing Committee on Comparative Research
Standing Committee on Data Archiving

ZA 1960 IZ ZUMA, in 1986 GESIS

Social Research Data Infrastructure

What is data Infrastructure?

A first approximation (1975):
„Secondary Analysis: Organisational and technical prerequisites“

Functions:

Systematic Observation of Data Production and data selection

Methodological and technical Study Description and questions

Data Checking and Cleaning

Preparing Materials for archiving

Retrieval and Analysis tools for Users in interaction with machine

Creation of CESSDA (1976) and IFDO (1977)

Comparative studies: Eurobarometers, ISSP …..

ECPR- Workshops on „ Integrating the European Data Base“ with Eric Tanenbaum (1990)

Villa Vigoni: Integrating the European data base:
Infrastructure?:
„If you come to Brussels never mention infrastructure: This is for us rubbers, rulers and pencils, and we are not going to fund that“

Jim Davis (in Essex and Cologne) :
If we manage to lay the data bases for future generations
we will have done our job!

ESF-SCSS Schloß Ringberg Tegernsee 1992
„Integrating the European Data Base“: The European data base is scattered, data are not well integrated .. … Europe is natural Lab for Social Research“

Paper as basis for ESF-SCSS : Hoffmann-Nowotny, Martinotti, Newby, Smith : Social Research in EC FP4 – TMR/LSF

ESF- CEPS- ZA: Luxembourg Conference Making Data European- Making European Data

Panel on Large Scale Facilities: ----Convincing Sciences:
They can be Virtual and distributed!
Radio Telescope Effelsberg Functional Equivalent 20 Mio €
Eurobarometers

Malacarne (outer space): Data bases are equivalent to LSF in Sciences

1996 ECASS EUROLAB

NESSIE EC High Level Round Tables: Marcia Taylor, Gaston Schaber, Bjorn Henrichsen, EM: Infrastructure more than big pot of money

1999 Strassbourg EC Infrastructure Conference: Social Sciences admitted!

EC funded projects > 5 Mio Euro:

developping methods and tools:

ILSES

NESSTAR

METADATER

BIG Project

ESF SCSS – EC meeting with Martinotti, Smith

Funding of transnational Costs

ESFRI

CESSDA
ESS
SHARE

‘Online survey on scientific information in the digital age July to Sept. 2011

EUROPEAN COMMISSION (2012)

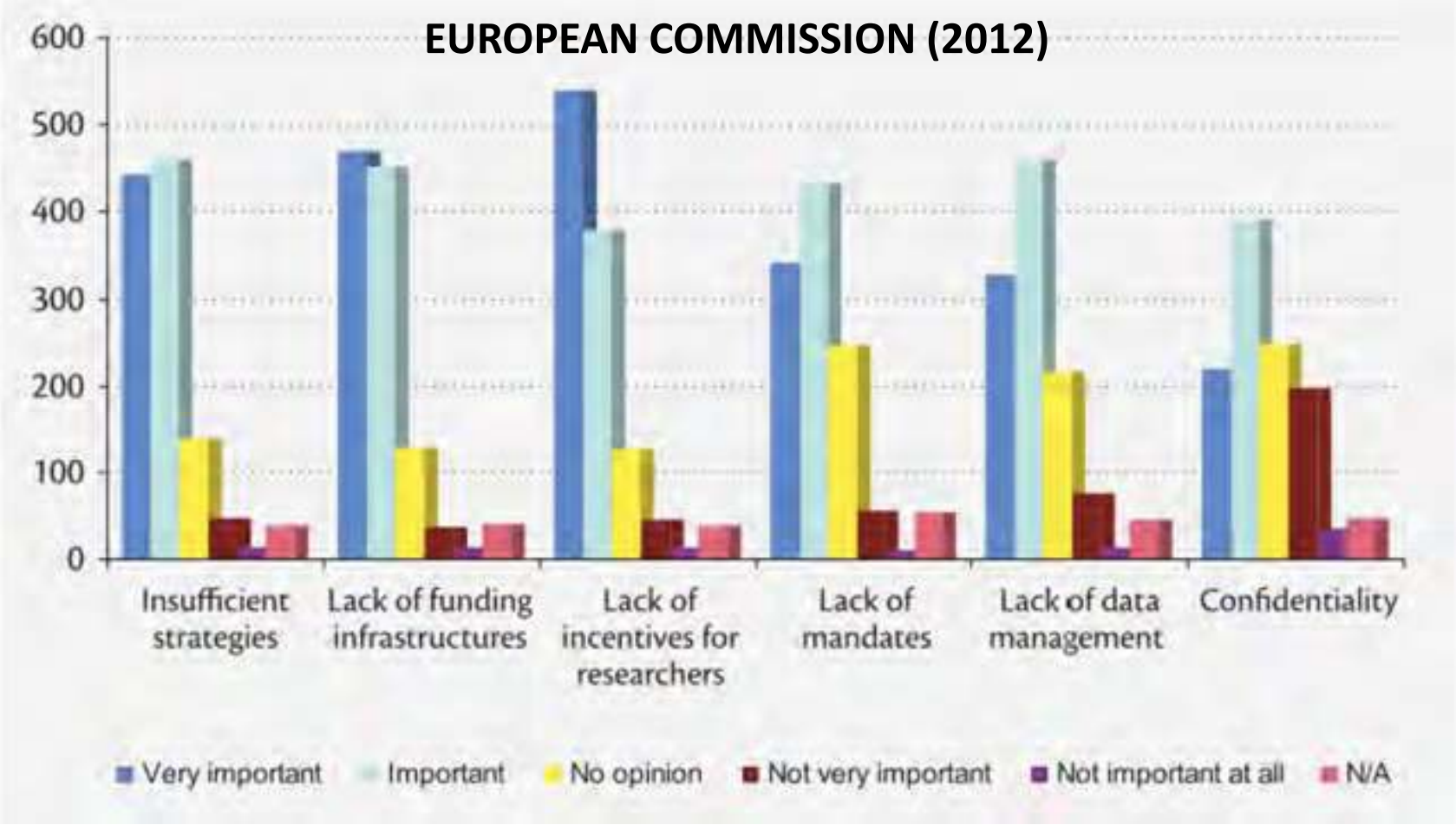
„As for the question of access to research data, the vast majority of respondents (87 %) disagreed or disagreed strongly with the statement that there is **no** access problem for research data in Europe.”

„The barriers to access research data considered very important or important by respondents were:
lack of funding to develop and maintain the necessary infrastructures (80 %);
the insufficient credit given to researchers for making research data available (80 %);
and insufficient national/regional strategies/policies (79 %).”

„There was strong support (90 % of responses) for research data that is publicly available and results from public funding to be, as a matter of principle, available for reuse and free of charge on the Internet.”

„Lower support (72 % of responses) was given for data resulting from partly publicly and partly privately funded research.”

How would you rate the importance of the following potential barriers to enhancing access to research data?



Existing Data Infrastructure in Germany

Data Archive for Empirical Social Research

(ZA since 1960, now department of GESIS)

Ongoing Data Collection Programmes

like Allbus, ISSP, Bildungspanel, SHARE, SOEP ….

KVI Commission on the Improvement of the Informational Infrastructure:

RDCs and DSCs for statistical micro data RatSWD

25 RDCs and DSCs

The research data centers (RDCs) and data service centers (DSCs) are accredited and supported by the German Data Forum (Rat SWD) with the aim of improving the research data infrastructure for the social, economic, and behavioral sciences, both at a German as well as international level.

Whilst pursuing this goal, the German Data Forum also bears in mind that infrastructure also has to be installed in areas that go beyond the scope of traditional infrastructure as given by governmental statistics (for example, departmental research, evaluati on studies, research based surveys, and research focal points using public funding).

http://ratswd.de/en/data-infrastructure/rdc

Principles for the Handling of Research Data

These principles were adopted by the
Alliance of German Science Organisations on 24 June 2010:

Alexander von Humboldt Foundation
German Academy of Sciences Leopoldina
Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)
German Academic Exchange Service (DAAD)
Fraunhofer-Gesellschaft
Helmholtz Association
German Rectors' Conference (Hochschulrektorenkonferenz - HRK)
Leibniz Association
Max Planck Society
Wissenschaftsrat (German Council of Science and Humanities)

Alliance Principles

Preamble

Quality-assured research data are a cornerstone of scientific knowledge and, independent of the purpose for which they were originally obtained, can often serve as the basis for further research. This applies especially to the aggregation of data from various sources for combined utilization. Preserving research data over the long term and making them available therefore does not only serve the verification of prior results, but also, to a large extent, the obtaining of future ones. It is a strategic task to which science and the humanities, politics as well as other parts of society, must contribute. With the objective of supporting the quality, productivity and competitiveness of science and academia, the Alliance of German Science Organisations has adopted the following data policy for a coordinated further course of action.

Preservation and Accessibility

In accordance with important international organisations involved in funding and performing research1, the Alliance supports the long-term preservation of, and the principle of open access to, datafrom publicly funded research.

This principle shall be balanced against the scientific and legal interests of researchers. The protection of the personal data of participants, patients and others affected by the collected data, as well as obligations to third parties — e.g. cooperation partners — have to be taken into account. The principles of good scientific practice must also to be observed2.

Differences between the scientific disciplines

The ways of and conditions for access to research data must be developed separately for the individual scientific disciplines, taking into account the methods of data acquisition, the volume and potential for integration of the data, as well as its practical usability. At the same time, the respective lifecycles and usage scenarios of the data in the specific research fields have to be considered.

Scientific recognition

The provision of research data for further use is a service which benefits the sciences and humanities in their entirety. The Alliance encourages the recognition and support of this additional costly and time consuming effort.

Teaching and qualification

For those involved in research, an appropriate range of training and support services for ***professional data management must be made available, meeting the specific requirements of the different disciplines.

Use of standards

Proper use of research data requires that the data are documented and provided with appropriate metadata in a standardised manner. Observing subject-specific requirements, standards, metadata catalogues and registries are to be developed in such a way that interdisciplinary use is also possible.

Development of infrastructures

Sustainable research data management imposes a wide range of technical and organisational requirements. These requirements must be defined through the cooperation of researchers and information specialists. Infrastructures are to be developed according to these requirements and, if possible, interoperably integrated in international and interdisciplinary networks from the start.

Challenges

Systemic research

Contextualisation and integration of data

New data types and management challenges

Capacity building and data science

And world awareness analytics in a digital world with - - electric cars driven by auto pilot-- need methodological grounding

ISR is at the forefront in developing new ways of thinking and new methods required to understand and solve these problems. More and more, ISR's traditional social science research has expanded to include environmental factors, biometric and biological data, including genetic material. We are now able to link survey, administrative, health and genetic data, and information about where people live from global positioning information.