Dirk Meusel, Peggy Göpfert, Nicole Wagner, Nora Merker, Wilhelm Kirch

Data Based Web Application for Public Health Relevant Research Data

Datenbank-basierte Web-Applikation für Public Health relevante Forschungsdaten

The demand of giving public account of research activities is met by German Public Health research by the compilation of empirical research data into so called Public Use Files and Public Documentation Files. Since research designs and deriving data sets in Public Health research can be described as heterogeneous as well as they frequently contain sensible data, disseminating these files via Internet requires special considerations concerning the computing architecture as well as software development. The workgroup Public Use Files has set up a pilot web application for testing the dissemination of Public Use Files via Internet. The web application is accessible at URL http://www.public-health.tu-dresden.de and introduced in the paper.
Keywords: Internet, Public Use Files, Research Practice

Die Forderung nach öffentlicher Berichterstattung über Forschungsaktivitäten wird durch die deutsche Public Health Forschung unter anderem durch die Erstellung von so genannten Public Use Files und Public Documentation Files realisiert. Da Forschungsdesigns und die daraus resultierenden Datasets sehr heterogen sind sowie sensible Patientendaten enthalten können, stellt die Dissemination dieser Files über das Internet besondere Anforderungen an die Rechnerarchitektur und die Software Entwicklung. Die Arbeitsgruppe Public Use Files hat eine Pilot-Webanwendung zum Test der Dissemination von Public Use Files bzw. Public Documentation Files über das Internet entwickelt. Diese Applikation ist unter der URL http://www.public-health.tu-dresden.de aufrufbar und wird in diesem Artikel vorgestellt.
Stichworte: Internet, Public Use Files, Forschungspraxis

Introduction

German research in Public Health was funded with approximately 50 million Euro of public incomes. An issue, which faces new attention in Germany too, is giving a public account of research activities. Research teams in numerous European or North American countries meet this demand of accountability by publishing so called *Public* or *Scientific Use Files*. On the one hand, these files provide transparency of research practice as well as research results. On the other, they provide the opportunity of secondary analysis of the scientific data. Common examples of this practice are the Public Use Files of the *Bundesgesundheitssurvey* (Federal Health Survey), which are being carried out and provided by the Robert Koch Institute in Berlin. The compilation of Public Use Files furthermore implies the advantage of promoting research results without additional expenses as well as the opportunity of stimulating research in one's emphasised field of interest.

In accordance with the DGPH (German Association of Public Health) and the project's supporting agency, a workgroup *PUF (Public Use Files)* was founded for publishing data of Public Health Research in Germany. Members of this workgroup are, among others, the commission of Public Affairs of the DGPH and the German Coordinating Agency for Public Health in Freiburg. This workgroup chooses suitable research projects, designs the formal framework for the Public Use File creation of high quality, and preserves the interest of the particular research project groups. One of the main objectives within this PUF workgroup is the dissemination of Public Use Files via Internet.

For the task of disseminating Public Health relevant research data via Internet, the PUF workgroup has set up a pilot web application with a twofold objective. On one hand, the prospective PUF user shall be able to access information regarding the files. On the other hand, the research teams shall be able to keep the information presented in the PUF up-to-date. This paper examines the special characteristics of empirical data in Public Health science and their implications on the dissemination of Public Use Files via internet. Following that, the data based web application is introduced.

Empirical data of Public Health Research and the Internet

Archiving and disseminating empirical data derived from Public Health research is a complex and multifaceted undertaking. As mentioned elsewhere (Meusel/Göpfert/Kirch 2001), several reasons support the strategy to create *public* or *scientific use files* from empirical research data in Public Health science. Firstly, the data deriving from the original research project will not get lost after the research team finished work on the data and on the project. In the past, once the members of a research team left, undocumented data files confronted even experts in secondary analyses with the nearly unsolvable problem of uncovering the connection between variables in the data file and the corresponding items in the field instruments. Hence, the creation of a public use file is an efficient way of archiving the data and preserving the metadata about it. Secondly, utilising empirical data by scientific third party users promotes a more comparative approach to research in Public Health sciences in general. Research methods and results become more lucid and comprehensible. Finally, and most important, public use files provide the opportunity to achieve secondary data analysis without high monetary and time expenditures. They can be utilised for educational purposes in Public Health studies as well as for follow-up research. Furthermore, standardised public use files convey the scientific exchange between national and international research units and, for this reason, present an additional source to promote a project's results.

Empirical data deriving from Public Health research hold two special characteristics with serious implications for their dissemination via internet. Firstly, data sets are very heterogeneous. One end of this continuum presents epidemiological survey data that can be characterised by large samples sizes and an extensive set of quantifiable data variables. Objective of those studies is the population wide description of the examined phenomena. The other end symbolizes case studies with a qualitative approach. Most often, those studies survey interview data to be investigated by different forms of content analysis. Between these ends lay differing research and sample designs with a mixture

378

of quantifiable and qualitative interpretable data scopes. These heterogeneous research and data designs make standardisation, especially if done respectively, a difficult task.

Secondly, research data out of Public Health research often contain medical information, patients information as well as biographical accounts surveyed in interviews. This data has to meet the protection guidelines of data privacy. Opposing to that fact, the internet personifies a medium of free access to all resources. Even password secured information placed on a web server is most often vulnerable for hacking attacks and, therefore, vulnerable for unauthorised access. Furthermore, research teams assured to their research subjects prior to data surveying that any data will not be given away to third parties in any form. Thus, either the data can not be published in form of a Public Use File or the a web application presenting Public Use File information must meet the demand of data privacy by special considerations in the computing architecture.

Following that characteristics, the PUF workgroup decided for a double strategy: Firstly, the compilation of both, *Public Use Files* and *Public Documentation Files*. While the first contain as much information as obtainable from the research team in conjunction with the empirical data files, the latter contain the descriptive information about the research project, field instruments as well as up-to-date contact details that can be used to negotiate the partial use of empirical data with the research team directly. *Public Documentation Files* are being compiled from research projects acquiring qualitative data as, for instance, biographical accounts out of interviews. Secondly, a divided computing architecture has been set up by the PUF workgroup for storing descriptive information and empirical data files. While the latter are stored in a firewall secured network setting, the former are stored on a password secured web server. This strategy provides the possibility that, on one hand, research teams can work directly via a central web application on *Public Use Files* or *Public Documentation Files* compiled out of their own research project. By that, up-to-date information is better granted than with a static approach. On the other hand, empirical data files are not exposed to unauthorised access on the web server. All work on empirical data files is therefore done by the *PUF* workgroup, whereas all other information can be done by both, the *PUF* workgroup and the research teams. For illustrating that scheme, the web application is introduced by the next paragraph.

The Web Application

The pilot web application for disseminating *Public Use Files* and *Public Documentation Files* via the internet is accessible by URL *http://www.public-health.tu-dresden.de*. New software developments, especially for heterogeneous target user groups, always bear a high potential of inconsistencies that can only be observed by feedback of as many users as possible. Therefore the term *pilot* is used in conjunction with this web application to symbolise its developing character. The Public Use Files web application is embedded in the general Public Health Research Association Saxony's web site, which offers the possibility of sharing information. The web application is designed with portal characteristics, meaning that after a successful login by valid username and

password, the client user is granted access to various functionalities within the web application depending on the rights set for the username. Those functionalities are accessible by a personalised option page (to be found at the *peoples* icon in the top right corner).
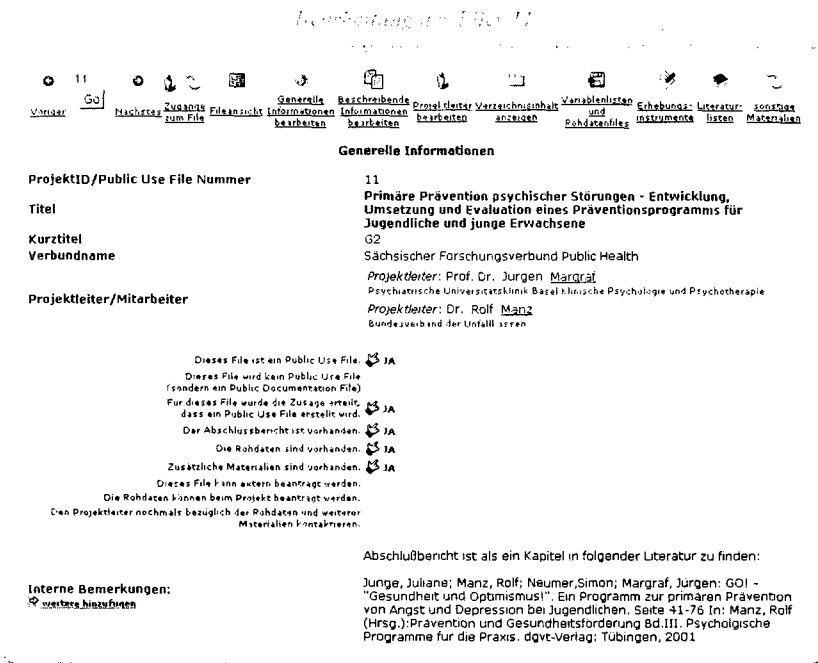
The Public User

The term *public user* applies to both, client users that did not log in by username and password as well as users, whose user rights are set to *guest* (Gast_allgemein). Clicking the index card *Public Use File* the public user already finds numerous information. Firstly, the public user can browse or search by single words the file archive for *Public Use Files* or *Public Documentation Files* of particular interest. Secondly, the public user is able to search all variables in all *Public Use Files* for specific terms or expression (e.g. *Kind* or *Alkohol*). Via the result listing the particular research project's description can be accessed. Thirdly, a forum on different issues regarding the work with *Public Use Files* can be accessed for reading. Finally, the download area provides various resources for downloading. Once the public user is logged in as *guest* (a new username can be registered by clicking the *Registrieren* button), additional functions are accessible as, for instance, providing one's own contact details, news, event announcements and web links with Public Health relevance on the web site. Furthermore, both the extended online view as well as the use of particular files can be applied for. After approval of application the extended view is accessible on the personalised option page too.

The Authorised User

The term *authorised users* applies to members of research teams, whose research project has been compiled to a *Public Use File* or a *Public Documentation File*. After registering the username, the workgroup evaluates the purpose remarks in the registration form. Following a telephone contact the user rights are set to work on the corresponding file or files. Figure 1 shows the dialog that is accessible via the personalised option page. Every single information can be altered, deleted or added. General information concerning the research project, descriptive information, and information about research team members are to be changed directly via HTML forms. Additional files can be used to provide information about field instruments, questionnaires, published and unpublished literature etc. These files may have any file format that is supported by most of the prospective client users. Therefore, Acrobat's *portable document format (pdf)* is the most appropriate. Once the files are generated, research team members are able to upload these file to the web server and to associate those with the corresponding *Public Use File* or *Public Documentation File*. For the latter, a mere database entry by the authorised user (research team member) is sufficient to describe the file more detailed and to make it visible to client users.

Figure : The work on Public Use Files dialog

References

Meusel, D./Göpfert, P./Kirch, W. (2001). 'Public Use Files Project – Archiving and Dissemination Empirical Data of the Public Health Research Networks in Germany'. In: Merker, N./Göpfert, P./Kirch, W. (ed.). Public Health Research and Practice: Report of the Public Health Research Association Saxony 2000-2001, Volume III. Regensburg: Roderer, pp. 375.

Correspondence:
Dirk Meusel
Public Health Research Association Saxony
Fiedlerstr. 33
D-01307 Dresden
Tel.: +49-351-458 5200