The IASSIST QUARTERLY represents an international cooperative effort on the part of individuals managing, operating, or using machine-readable data archives, data libraries, and data services. The QUARTERLY reports on activities related to the production, acquisition, preservation, processing, distribution, and use of machine-readable data carried out by its members and others in the international social science community. Your contributions and suggestions for topics of interest are welcomed. The views set forth by authors of articles contained in this publication are not necessarily those of IASSIST.

Information for Authors:
The QUARTERLY is published four times per year. Authors are encouraged to submit papers as word processing files. Hard copy submissions may be required in some instances. Word processing files may be sent via email to jistratford@ucdavis.edu. Manuscripts should be sent to Editor: Juri Stratford, Government Information and Maps Department, Shields Library, University of California, 100 North West Quad, Davis, California 95616-5292. Phone: (530) 752-1624.

The first page should contain the article title, author's name, affiliation, address to which correspondence may be sent, and telephone number. Footnotes and bibliographic citations should be consistent in style, preferably following a standard authority such as the University of Chicago press Manual of Style or Kate L. Turabian's Manual for Writers. Where appropriate, machine-readable data files should be cited with bibliographic citations consistent in style with Dodd, Sue A. "Bibliographic references for numeric social science data files: suggested guidelines", Journal of the American Society for Information Science 30(2):77-82, March 1979. Announcements of conferences, training sessions, or the like, are welcomed and should include a mailing address and a telephone number for the director of the event or for the organization sponsoring the event.

Editors
Karsten Boye Rasmussen, Eckersbergsgade 56, 5230 Odense M, Denmark.
Phone: +45 6612 9811,
Email: boye@get2net.dk

Juri Stratford
Government Information and Maps Department, Shields Library, University of California, 100 North West Quad, Davis, California 95616-5292.
Phone: (530) 752-1624.
Email: jistratford@ucdavis.edu

Production
Laura Bartolo,
Libraries and Media Services, Kent State University, Ohio 44722.
Phone: (330) 672-3024, x31.
Email: lbartolo@kentvm.kent.edu

Walter Piovesan
Research Data Library, Simon Fraser University, Burnaby, B.C.
Canada V5A 1S6.
Phone: (604) 291-5937.
Email: walter@sfu.ca

Title: Newsletter - International Association for Social Science Information Service and Technology
ISSN - United States: 0739-1137 © 1997 by IASSIST. All rights reserved.
A Web-Based Archive of Psychological Experiments: Challenges for Client Server Interactions

Abstract:
The Virtual Psychology Laboratory (VP-Lab) is an ongoing project which will provide psychology educators, students, and researchers with a tool that supports the analysis, modification, and re-execution of previously archived experiments. The archive will include experimental materials, designs, procedures, and results which will be submitted by active researchers and educators. A user will be able to access VP-Lab via the World Wide Web using one of the widely used Web browsers. The paper identifies several requirements of providing psychological experiments in the World Wide Web. In addition, it discusses an approach to Web programming which satisfies all the requirements.

Introduction
With an increasing body of psychological research, archives became interested in providing their services to the psychological community. First attempts in this endeavor involved archiving abstracts of psychological research papers and developing a suitable thesaurus (American Psychological Association 1992, 1994).

An approach which goes far beyond archiving abstracts is currently taken by the Virtual Psychology Laboratory (VP-Lab) project, a collaboration between the School of Psychology at the University of Cardiff, The Data Archive at the University of Essex, and the Computers in Teaching Initiative, Centre for Psychology at the University of York. It is the objective of the VP-Lab project to provide facilities for archiving complete psychological experiments which will be submitted by active researchers and educators. The user will be able to access VP-Lab via the World Wide Web using one of the widely available Web browsers.

The implementation of such a comprehensive service requires an understanding of the elements of psychological research. Given that Psychology is such a diverse discipline, it is not surprising that the field relies on a wide range of methods. Most psychological research methods have the objective of answering empirical questions about behavior or experience by controlled observation. A basic instrument of the empirical approach is the experimental method (Davis 1995, Coolican 1994). We should note that in Psychology as well as in other social sciences other methods are used as well. For example during interviews and surveys, typically the researcher has less control over the environment in which the investigation takes place. The control over the experiment can be viewed as the characteristic difference between the experimental method and other research methods. The advantage of the experimental method in Psychology is that it is potentially capable of providing statistical quantifiable evidence for relationships between cause and effect.

Thus the control of variables is an essential aspect of any experiment. There are two categories of variables: independent and dependent variables. Variables which the experimenter manipulates are referred to as independent variables, whilst variables which describe the experimental result are referred to as dependent variables. The statistical quantifiable evidence for relationships between cause and effect is reflected by these two types of variables. Typically, such relationships are formulated as hypotheses which form part of a larger theory. By making observations, the psychologist can make a statistical judgement as to whether or not a prediction is correct; that is predictions can be tested against the evidence.

The experimental hypothesis proposes that the change in the behavior as it is measured in the dependent variable(s) is actually caused by changes in conditions of the experimental setting. These changing conditions are described in terms of the independent variables. Events which change the experimental setting in an objective way and which are capable of evoking a response from participants of the experiment are referred to as stimuli.

Usually, the evidence cannot be obtained from the total population to which the hypothesis might apply. Therefore a sample of the relevant group is taken. Assumptions need to be made concerning the representativeness of this sample. The selected members of the sample or participants are put into a controlled situation; i.e. a situation where all relevant independent variables are controlled by the experimenter and where the dependent variables can be measured.

However, controlling the independent variables does not suppress the variability in people’s behavior. This variability will influence the measurements taken from the
dependent variable. Therefore the experimenter will be faced with a whole range of different scores by participants. The question to be decided is whether the differences in the scores are the result of manipulating the independent variables, or are the result of chance fluctuations in people’s performance as stated by the null hypothesis. In other words, are the score changes significant in support of the experimental hypothesis? The experimenter decides this question by performing an analysis of data which involves suitable statistical tests.

A considerable amount of psychological research follows the general schema of the experimental method as it is outlined above. Currently, researchers obtain information about previous experiments mainly from journal publications and conference proceedings. This method has two disadvantages:

1. Identification and evaluation of relevant previous work can be very time consuming.

2. Even journal publications often may not contain the degree of detail which is required to repeat or to modify a previous experiment.

The aim of the VP-Lab project to overcome these limitations is addressed by making all the details available in the World Wide Web which allows a comparatively fast access to archived information. In addition, this approach enables the archivist to provide more detailed information about an experiment than is possible in a journal paper.

In this paper, we will argue that:

1. the Web-based provision of detailed information about psychological experiments requires
   a) interactive content,
   b) secure access,
   c) interface-database connectivity, and
   d) platform independence.

2. given the current state of the art, the programming language JAVA satisfies these requirements.

In the remainder of this paper we will address these claims by inspecting the various components of a psychological experiment in detail (Section 2). This inspection will indicate the requirements for a Web-based provider of psychological experiments. In Section 3, we will describe an approach to Web programming which is based on the programming language JAVA. Finally, in Section 4, we will discuss the merits of this approach from the perspective of providing executable psychological experiments on the World Wide Web.

Psychological Experiments
In this Section, we will describe the various components which constitute a psychological experiment. From this description, we will then derive technical requirements which have to be satisfied if we wish to provide psychological experiments in the World Wide Web. Typically, descriptions of psychological experiments identify the experimental components: materials, design, procedure, and results. Before we discuss these components, we consider an experiment which will provide the basis for the subsequent sections.

An Example Experiment
The example we choose is an experiment which has been described by Klein (1994). The experiment was based on previous work (Stroop, 1935), in which the reaction of participants who had to identify the colors of ink they saw was measured. In the experimental condition, the participants were presented with words which represented colors such as the words red, blue, and yellow. However, the words were printed in a color which differed from the word’s meaning. For example, the word red was printed in the color blue. In the control condition, participants just had to identify color spots. The reaction time in the experimental condition was significantly longer than in the control condition. This was explained with an interference of the color identification by the processing of the word meaning. Klein studied such inference by varying the meaning of the stimulus words. He compared conditions in which the word was either a nonsense syllable (e.g., BJB), a word that implied a color (e.g., grass), or the name for a color (e.g., red). Then he measured the time his participants required to identify the color under each of these conditions.

Materials
In the introduction, we have emphasized that changes of the experimental setting described in independent variables play a central role in any experiment. Sometimes this setting involves a complete well designed environment. For example, a developmental psychologist might place a child in a play room with particular toys. Often, changes are achieved by presenting participants with pre-defined text, image, or sound samples. This type of materials can be presented by using a computer with multi-media capabilities. For example, the word stimuli we described in the previous section can be displayed on a computer screen. In addition to these types of information, materials include computer programs to generate and present suitable multimedia files. The programs may be stand-alone programs written in a multi-purpose programming language, or they may have the form of script files which can be read and executed by a commercial experiment generator.
The materials characterized in this section are different from data which can typically found in social science archives. First, stimuli which are contained in the materials are suitable to control the independent variables of an experiment. We have emphasized above that this aspect does not play such a central role in studies which are based on surveys rather than experiments. The need to control independent variables will have consequences for indexing stimuli. For example, a researcher who wishes to use a given stimulus in another experiment has to ensure that the independent variables in this experiment are controlled appropriately. Therefore, the variables to be controlled have to be considered during retrieval of the stimuli. The second important difference to data usually found in social science archives is the need to archive computer programs. Moreover, these programs have to be indexed in a suitable way. The support provided by current psychology thesauri such as APA’s thesaurus (American Psychological Association 1994) is very limited. Moreover, the task of archiving programs raises the problem of maintaining programs over long periods of time.

**Design**

For every experiment, it is crucial to group participants and to present stimuli in a way which avoids bias towards a particular experimental result. These general principles of the experimental setup in terms of conditions, participants, and variables used are described as design (Kirk 1995, Leon & Austin 1996). For example the layout of Klein’s experiment mentioned above can be characterized as a three levels of one-factor, between-subjects design. It is based on one independent variable, the word presented to the participant. The three levels are given by the experimenters choice between a nonsense word, an implied color word, and a color name. This design is referred to as between subjects design, because subjects or participants were randomly assigned to each condition without regard to the participants in other conditions, and each participant serves in only one condition. In contrast, a within-subject design would be a design in which the same subjects are repeatedly measured in different conditions, or each subject in one condition is matched with each subject in another condition. In addition to topical retrieval goals, such design characteristics could be used by VP-Lab to retrieve an experiment. However, similar designs are used by a number of different experiments. So design characteristics alone will not uniquely describe a particular experiment, additional characteristics are required.

**Procedure**

Whilst the experimental design describes the general arrangements made for an experiment, the procedure provides the detailed steps to be followed in performing the experiment. Therefore the procedure often follows a temporal sequence. They begin with summarizing the instructions given to the subjects and proceed through the tasks performed by the subjects in the order in which they were performed.

**Results**

During the experiment all the responses of the subjects are carefully recorded. Often this type of information may be obtained automatically using computers. These raw data are then analyzed to determine the significance of the result. The statistical method used for data analysis is determined by the chosen experimental design type. For example, the method used by Klein in his experiment of the Stroop effect is the one way, between-subjects Analysis of Variance. This method should be employed in a between-subject design with two or more levels of one factor. The method of analysis of variance is a procedure which enables the experimenter to determine whether significant differences exist in an experiment involving two or more sample means (Greene & Oliveira 1995). Such statistical methods can be used to index an experiment. A student may later retrieve the experimental data as example data for the given analysis method.

**Requirements:**

I have provided information about the typical components of a psychological experiment because this provides important indications of the requirements a Web-based experiment provider has to address.

**Interactivity**

Interacting with stimuli may not be just a two step process of presenting stimuli and recording the response; it may be a sequence of interactive steps. The presentation of a stimulus may even depend on a previous response. Furthermore, a participant might be required to interact with a specified part of a stimulus such as a particular area in an image. Therefore the presentation has to include executable content. This is a feature which allows different responses and supports different reactions of the stimulus depending on the response.

**Platform Independence**

Typically, an experiment is developed by using a particular computer system, usually one known or accessible to the researchers. There is no need for them to consider portability to other systems. However, a Web-based experiment will be used by a large number of users with different computers and operating systems.

**Interface-Database Connectivity**

The various experimental components have to be stored in a database. Rather than just addressing the issues of interactivity and platform independence, a Web-based provider of psychological experiments has to address the question of how to connect an interactive and platform independent browser with the database. A researcher who interacts with VP-Lab may be merely interested in a particular type of stimuli rather than a complete experiment. Therefore the connection has to be achieved in
a flexible way which supports the interaction between user interface and single experimental components.

Security
In all Web applications, security issues need to be taken very seriously because networked computer systems are by definition more vulnerable to attacks. In addition to these general considerations, providers of psychological experiments have to ensure the confidentiality of participants. Often psychological data obtained from a single participant reveal highly personal information. This problem appears only to a smaller degree in journal publications because they contain the results of the data analysis rather than the raw data. In addition to the protection of the rights of participants, we have to consider the protection of the rights of the researcher who deposits the experiment. Sometimes in the course of research, highly sophisticated software has been developed over a long period. The researcher may wish to distribute the experiment in terms of stimuli, design, procedures and results, whilst restricting the distribution of the software for generating the stimuli.

Providing Psychological Experiments via the Web
The requirement that we have identified in the previous section can be addressed by programs written in the programming language JAVA.

Interactivity
JAVA is a programming language developed in 1995 by Sun Microsystems (Gosling, Joy, & Steele 1997). It is used to create executable content which can be distributed through networks. The JAVA concept distinguishes between two types of code; stand-alone programs referred to as application and pieces of code that are linked to a Web page and sent as executable content through the Internet. This type of program is referred to as applet. In order to view JAVA content on the Web, a user’s browser must be JAVA enabled; i.e., the browser has to be integrated with a JAVA interpreter. An increasing number of Web browsers have this feature. An early example of this type of browser is Netscape Navigator. The content downloaded by a Web browser can include a variety of multimedia documents.

If the browser receives a user request, it downloads content that describes a Web page (Figure 1). The Web page can contain a particular hypertext tag called APPLET. When downloading a Web page containing an APPLET tag, the JAVA-enabled browser knows that a special kind of JAVA program called an applet is associated with that Web page. The browser than downloads another file of information, as named in an attribute of the APPLET tag, that describes the execution of that applet. This file of information is written in what are called bytecodes. The JAVA-enabled browser interprets these bytecodes and runs them as an executable program on the user’s computer. The resulting execution then drives the animation, interaction, or further communication which is again displayed by the browser. The overall pattern for the use of content is selection of content by the user, downloading, executing, and displaying of content by the browser. This process in itself already provides considerable support for interactions between a user and psychological experiments described as executable content. However, JAVA has two other supportive features: It is object-oriented and multithreaded. The term object-oriented means that components of an experiment can be represented as separate objects (classes in the JAVA terminology) which store the functionality of the components in the form of methods. This concept supports a highly modular approach because changes in one experimental component could be made independently of other components. Moreover, experimental components would be the building blocks of an experiment on the implementational level rather than just on the conceptual level. The term multithreaded refers to a pseudo-parallel approach. Typically, computers have only a single processor; therefore parallel execution of several programs in the strict sense is not possible. However, JAVA programs can direct the processor in those parts of the program to be executed next. These parts can be very small and a fast switch from one part to the other gives the impression that these parts are executed in parallel. Therefore multithreading also supports rapid interactions between a user and a pre-stored experiment.

Platform Independence
We pointed out that JAVA applets are transported over the Internet in a particular form which we referred to as bytecode. Every computer that can execute JAVA code has a machine specific interpreter that translates bytecode into machine code. This is the reason why JAVA programs are machine independent; i.e. the same program can be executed on a UNIX system, a PC, or a Macintosh computer.

Interface-Database Connectivity
The JAVA language includes several tools that extend the language to different tasks. One of these tools supports connections between JAVA code and relational database systems. This tool is referred to as Java Database Connectivity (JDBC) and defines every aspect of making data-aware JAVA applications and applets (Jepson 1996, Patel & Moss 1996). In using this tool, the developer does not need to be concerned about the database-specific syntax when connecting to and querying different databases. Another advantage of this approach is that changes to the applet code can be minimized when the database system is changed. The JDBC concept has recently been extended to a JAVA based three-tier architecture. Typical client server interactions are based on two types of computer systems: a central server that maintains the database and a number of clients that maintain the user interface to the database. In a three-tier architecture this concept is enhanced by an
additional middleware server that connects with the database server on one side and with a number of clients on the other (Symantec 1996). This approach has several advantages. For example, the client systems are easy to manage because no application or database software is required to be installed on the client side. Any system with a JAVA enabled web browser will work as a client with no additional software. The system is easy to program because all code is executed on the client system. The developer of application programs accesses neither the middleware server nor the database server.

**Security**

It is the basic principle of the applet approach that JAVA code is downloaded to and executed by the client computer. This involves a considerable potential security risk. Therefore the JAVA language includes the most sophisticated security system of any programming language (Breedlove et al. 1996, Morrison et al. 1996). The complexity of this system is clearly beyond the scope of this paper. However we will discuss some of the basic ideas. JAVA programs can be viewed as a set of classes which in turn can be built-in or user-defined. We focus on the security check for user-defined classes. Each class is checked by three different sub-systems: the JAVA verifier, the class loader, and the security manager. The JAVA verifier is used to check the bytecode to make sure that the safety features of the JAVA language are followed.

After incoming code has been checked by the verifier, the protections in the Java class loader are invoked. An important function of the class loader is to avoid overriding of filesystem source classes. It also makes it impossible for a file system source class to access a network source class by accident. Finally, the security manager is used to provide a flexible access control mechanism. Any time a
the interactive interface-database computer.

JAVA
experiments
experiment,
experimentation archive
advantages.
Web
will
Interactivity
requirements
language
JAVA
experiments
the
browser
language
Platform
argued
Conclusions
In the introduction to this paper, we have described the objective of the VP-Lab project as to provide an archive of psychological experiments which can be performed over the World Wide Web. In context of this task we have argued that

1. the Web-based provision of detailed information about psychological experiments requires
a) interactive content,
b) secure access,
c) interface-database connectivity, and
d) platform independence.

2. given the current state of the art, the programming language JAVA satisfies these requirements.

We have considered typical components of psychological experiments to illustrate the need for interactive content, platform independence, database connectivity, and secure access. Finally, we have addressed these requirements from the perspective of the general purpose programming language JAVA. We will now evaluate this perspective in terms of the identified requirements. This evaluation will be biased because our viewpoint is how useful the language is for supporting the provision of psychological experiments rather than JAVA’s general usefulness for other types of applications. Nevertheless, several of our requirements can be regarded as general requirements for numerous Web applications.

Interactivity
JAVA code is downloaded rather than executed on a Web server. Therefore performing a psychological experiment will not depend on the load imposed by other users on the Web server. It just depends on the capabilities of the client computer. Downloading JAVA applets can be used to modify the interface of the Web browser. This has two advantages. First, VP-Lab can integrate a standard Web browser with a state-of-the-art interface for browsing its archive and supporting the psychologist in various experimentation tasks. Second, in performing a pre-stored experiment, the interface can be changed automatically to the type of screen used in the original experiment.

Platform Independence

JAVA programs are executed by the client machine; therefore a very important issue is whether they can be executed by machines of different types under different operating systems. Otherwise, the usability of these programs would be very limited. JAVA applets are already available in the Internet and their platform independence can be tested. The same applet can be downloaded and executed by a PC, a Macintosh, and a UNIX system. In addition, an increasing number of Web browsers have become JAVA-enabled.

Interface-Database Connectivity
The JDBC-based three-tier approach reduces the programming effort for the actual link between the application program and database because a considerable proportion of this link is already provided by the middleware server and needs only to be adapted to the particular application. This adaptation is performed on the client side rather than on the side of the middleware server or database server. In addition, the middleware server is embedded in a high-level development environment that reduces the programming effort further.

Security
JAVA controls have by definition no capabilities to read or write to the local file system or to make any operating system calls. Furthermore, JAVA uses a runtime verification of its code. Using bytecode has two advantages from the security perspective. Bytecode can be checked for security violations and it allows accessing of methods and variables by name rather than by number. This makes it easier to determine what is being used and to protect from misuse.

All these considerations led to the decision to implement VP-Lab in JAVA. We should note however that VP-Lab is a short-term project that requires a decision about facilities which are immediately available. In a few years time the situation may change. However, we expect that future Web programming languages will incorporate JAVA concepts such as machine independence and multi-layer security mechanisms. However, even more important than these technical aspects will be the fact that psychologists do not need to be aware that they are interacting with JAVA applets or with the World Wide Web. All they will see is an archive of experimental materials, designs, procedures and results.

References


1 This description follows closely a review given by Heiman (1995).

* Paper presented at IASSIST/IFDO ‘97, Odense, Denmark, May 6-9,1997. Ruediger Oehlmann,University of Essex, The Data Archive, Psychology Unit Colchester CO4 3SQ, UK oehlmann@essex.ac.uk
ILSES (Integrated Library and Survey-data Extraction Service) project

ILSES General overview
This document describes the ILSES (Integrated Library and Survey-data Extraction Service) project.

ILSES: introduction
General information
ILSES is a project of a number of Dutch, German, French, and Irish institutes. ILSES has been accepted by the European Commission under the Fourth Framework (Telematics Applications of Common Interest, Telematics for Libraries) programme as project LB-4050/C and will as such run from September 1996 to September 1999.

Project Goals
The ILSES project aims to develop a service that enables individuals users to access and retrieve documentary information and empirical data related to large-scale surveys such as the binannual Eurobarometer surveys.

ILSES is designed to serve both end-users and content-providers of socio-economic information.

End-users
For end-users ILSES facilitates the integrated access and retrieval of two different kinds of information:

1. documentary information as is commonly available from libraries, and
2. empirical data as archived by data archives.

ILSES allows end-users to extend literature research and searches with a focused access of empirical data which can or have been used in empirical research of the kind reported in the literature they review. This allows them to extend classic literature research with their own original empirical analyses of relevant data.

ILSES of course also offers the other route: directing data-analysts searching for specific empirical data to be used in secondary analysis to literature in which results and outcomes of previous and similar analyses have been reported.

Content-providers
For content-providers ILSES offers tools and procedures for the normalization, cataloguing and controlled distribution of (distributed) holdings of the documentary and data resources mentioned above. Such content-providers are the database administrating staffs of librarian and data-archival institutions. ILSES will enable them to drastically increase and improve the utilization of their information resources while at the same time reducing their support burden per information request or retrieval.

General design
ILSES will be designed as an open system which can be applied to different kinds of library and data holdings. In this project, however, a pilot-application will be focused on socio-economic information as collected by large scale surveys, and on the associated literature.

ILSES is based on integrated relational databases of meta-information pertaining to both library and data-archive holdings, both of which are typically distributed over many different institutions. In order to productively connect such holdings with each other and with end-users, ILSES provides a wide-area network interface utilizing Internet and supporting browsing and retrieval tools such as WWW.

ILSES partners
The following institutes are partners in the ILSES project:

- icer ProGAMMA, Groningen
- SWIDOC, Amsterdam
- University of Amsterdam
- Zentral Archiv, Cologne

The following institutes are associate partners in the ILSES project, working in association with the University of Amsterdam:

- Trinity College, Dublin
- CIDSP, Grenoble


**iec ProGAMMA**

*Introduction*

The interuniversity expertise center ProGAMMA is a not_for_profit cooperation of eight Dutch, British, and Belgian universities, established in 1989. The center promotes the development, quality, and use of computer applications for the social and behavioral sciences. ProGAMMA is recognized by the Netherlands Organization for Scientific Research (NWO) and supported by the Dutch government.

**Role in the ILSES project**

ProGAMMA is the Coordinating Partner in the ILSES project. The Coordinating Partner is responsible for communication with the European Commission. ProGAMMA is also responsible for the management and financial administration of the project.

ProGAMMA is also responsible for the development of many of the ILSES software tools.

**SWIDOC**

*Introduction*

The Social Science Information and Documentation Centre (SWIDOC) of the Royal Netherlands Academy of Arts and Sciences, promotes and facilitates the exchange and efficient use of information in social science research.

**Role in the ILSES project**

SWIDOC will deliver the necessary expertise in the library field for the development and evaluation of the ILSES bibliographic data. SWIDOC will also function as content provider of bibliographic data, and will as such function as a test site for the content-provider tools of LIB-ILSES.

**University of Amsterdam**

*Introduction*

The University of Amsterdam, more specifically the Faculty of Political, Social and Cultural Sciences (PSCW) is since long one of the most prominent locations of advanced academic empirical social research in the Netherlands. The Faculty contains strong and productive nuclei of methodological expertise, and of innovative large_scale survey research. The combined expertise is particularly strong in the area of electoral research, public opinion studies, multi_level research and methodology of comparative research, all of which are of eminent importance to the ILSES project.

**Role in the ILSES project**

The University of Amsterdam will function as the main end-user representative. They will research user demands prior to the creation of the ILSES tools, coordinate end-user testing at the CIDSP and Trinity College, and evaluate the results. Throughout the project, the University of Amsterdam will provide end-user feedback for the developing ILSES tools.

**Zentral Archiv**

*Introduction*

The Central Archive for Empirical Social Research at the University of Cologne, Germany (ZA) is a resource, research, and teaching centre for national and international comparative research. The ZA is responsible for archiving, processing and distributing the Eurobarometer data, in co-operation with the Inter-university Consortium for Political and Social Research (ICPSR) and the Swedish Social Science Data Service (SSD).

Within the framework of the International Federation of Data Organisations for the Social Sciences (IFDO) and CESSDA, ZA has access to the data contained in social science data archives worldwide.

**Role in the ILSES project**

The Central Archive, on the background of rich research and data management experience and acknowledged competence in integrating international comparative data sets, will help develop the content-provider ILSES tools to access survey-data and guarantee the consideration of international standards for data documentation.

**Trinity College**

*Introduction*

Trinity College in Dublin, is one of the most important academic institutions in the Republic of Ireland. It is renowned for its library and documentary infrastructure, and for its sound empirical social research. The Department of Politics, which will be the most involved in the organization of the user validation workshops of the ILSES project, is unusually strongly embedded in European social research and collaboration networks, owing to which it is particularly well placed for contributing to such validation studies. This Department can be regarded as a prototypical environment for many of the end_users of ILSES.

**Role in the ILSES project**

Trinity College will serve as end-user test site for the ILSES tools. The College will also help disseminate ILSES through the organisation of workshops.

Trinity College is an associated contractor, associated to the University of Amsterdam.

**CIDSP**

*Introduction*

The Centre d'Information des Données SocioPolitiques, as intermediaries and suppliers, represent a rapidly developing new type of interactive data and documentation resource and interactive service providers. Their experience with international organisations such as CESSDA will enable the input of international best practice and standardization to the ILSES project.
Role in the ILSES project

CIDSP will function as end-user test site for the ILSES tools. CIDSP will also help disseminate ILSES through the organisation of workshops.

CIDSP is an associated contractor, associated to the University of Amsterdam.

ILSES tools
In the ILSES project the following tools will be developed:

Administrator-ILSES
The technical heart of the system

LIB-ILSES
For library content-providers

DAT-ILSES
For content providers with data archives

E-ILSES
For end-users on standalone machines

NET-ILSES
For end-users on the Internet

Administrator-ILSES
Administrator-ILSES will be the technical heart of the system. It consists of the datadictionary, containing all metadata about available archive data and bibliographic data, plus the interfaces that access this metadata.

Administrator-ILSES will consist of a formal specification, plus a set of (MS-Windows-) DLLs that implements this specification. Content-providers or other interested parties seeking to expand upon the ILSES service may get access to Administrator-ILSES.

The specification of Administrator-ILSES will be developed by the University of Amsterdam, in consultation with the other project members. The implementation of the software modules will be done by iec ProGAMMA.

According to the ILSES planning, the first version of Administrator-ILSES will be ready in March 1997.

LIB-ILSES
LIB-ILSES will enable the library content-provider to produce and maintain the metadata needed for the ILSES end-users to access bibliographic information through NET-ILSES and E-ILSES. LIB-ILSES will contain a set of tools to set up and maintain the metadata and will contain also connections to give on-line access to existing bibliographic and documentary databases and their interrelations.

LIB-ILSES will consist of a set of mainly PC-based programs, that will be available for library content-providers.

The first version of LIB-ILSES will be developed by iec ProGAMMA, in consultation with the SWIDOC, the Central Archive and the University of Amsterdam. The LIB-ILSES tools will then be installed at SWIDOC to be integrated with existing library systems. At least 250 new documentary records will be compiled at SWIDOC and included into ILSES.

According to the ILSES planning, the first version of LIB-ILSES will be available April 1998.

DAT-ILSES
DAT-ILSES will enable the data-archive content-provider to produce and maintain the metadata needed for the ILSES end-users to access statistical data through NET-ILSES and E-ILSES. DAT-ILSES will contain a set of tools to set up and maintain the metadata and will contain also connections to give on-line access to existing statistical data archives. End-users will also be able to download selected sets of statistical data.

DAT-ILSES will consist of a set of PC- and UNIX(tm)-based programs, that will be available for data-archive content-providers.

The first version of DAT-ILSES will be developed by iec ProGAMMA, in consultation with the Central Archive, the University of Amsterdam, and SWIDOC.

The DAT-ILSES tools will then be be installed at the Central Archive and integrated with the local retrieval system. A test base for selected Eurobarometers will be provided.

According to the ILSES planning, the first version of DAT-ILSES will be available November 1997.

E-ILSES
E-ILSES is the tool for the end-user to access networked information resources, as services to be provided by both LIB-ILSES and DAT-ILSES. However, it should also be capable of operating stand-alone. This is a strict requirement from the research-user who wants to be able to use E-ILSES stand-alone as information- and data extractor for distributed SPSS study files and documentation. This implies that some functionality of the LIB-ILSES and DAT-ILSES will also be incorporated into the ILSES user tool. When operating stand-alone E-ILSES can use any PC-based common database format, including those available through ODBC. When operating in client-server mode any relational database capable of understanding SQL can be used and manipulated.
Through E-ILSES, the user may:

- Examine content of archives and retrieve references to articles relating to the holdings
- Combine the contents of several holdings to set up a new study
- Retrieve the selected data

E-ILSES will be developed as a stand-alone PC-program, running under Windows 3.x or Windows 95/NT.

E-ILSES will be developed by iec ProGAMMA, in consultation with the University of Amsterdam, the Central Archive, and SWIDOC.

According to the ILSES planning, the first version of E-ILSES will be available in June 1997.

**NET-ILSES**
Next to the stand-alone version for end-users, E-ILSES, a WWW-based service for end-users will also be developed. This service will offer (part of) the functionality of E-ILSES, and possibly more. As the technology of the Internet is very rapidly progressing, the form NET-ILSES will eventually take, and the relation with E-ILSES, will continually be re-evaluated in order to follow technology and end-user demands.

NET-ILSES will be developed by iec ProGAMMA, in consultation with the Central Archive, the University of Amsterdam, and SWIDOC.

NET-ILSES will be developed in the form of one or more HTML-pages with corresponding code, tools, applets etc.

According to the ILSES planning, the first version of NET-ILSES will be available August 1998.

**More Information**
More information about ILSES can be found at the ILSES home page, at URL http://www.gamma.rug.nl/ilses

There is a mailing list that will keep interested people updated about the progress of the project, and encourages discussions about ILSES.

It can be subscribed to via the ILSES home page, or by sending a mail message to listserv@nic.surfnet.nl, with the message text SUBSCRIBE ILSES-L YourNameGoesHere (put your name in the location YourNameGoesHere.)

You can also contact the project coordinator at iec ProGAMMA: David A. Schweizer, iec ProGAMMA, P.O. Box 841, 9700 AV Groningen, The Netherlands, tel: +31 50 363 6900, fax: +31 50 363 6687, e-mail: gamma.post@gamma.rug.nl

* Paper presented at IASSIST/IFDO '97, Odense, Denmark, May 6-9,1997. -
Meaningful Relationships

by Ken Miller

INTRODUCTION
The Data Archive’s thesaurus, HASSET (Humanities and Social Science Electronic Thesaurus), is based upon the UNESCO thesaurus compiled by Jean Aitchison (Paris; UNESCO 1977) and has been built up over 18 years so that its coverage reflects the subject matter of the 5,000 datasets held at The Data Archive.

This paper will describe the construction, maintenance and use of the thesaurus as a controlled vocabulary for indexing and a retrieval tool in The Data Archive’s on-line catalogue BIRON (Bibliographic Information Retrieval ON-line). It will also outline The Data Archive’s proposed developments for HASSET as an on-line thesaural resource for the social science community in general and as a multilingual free-text retrieval tool within, among others, the NESSTAR (Networked European Social Science Tools and Resources) project.

THESAURI
Dictionary definitions of thesauri describe them as “a storehouse of information” e.g. a dictionary, or “a list of concepts or words arranged according to sense” e.g. Roget’s, or “a list of concepts or words chosen for use in indexing” e.g. the UNESCO thesaurus.

HASSET is all this and more and is why we at The Data Archive consider it just that, a Huge ASSET. Its use as a controlled vocabulary means that every dataset whose question or variable covers the same subject material will be indexed by the same concept term. The structured relationships allow the indexer to view candidate terms within a concept hierarchy. The fact that it is machine readable allows instant, easy and consistent maintenance and flexibility in displaying terms in various different ways. It also means that it can be used as a retrieval tool in BIRON helping the searcher to better define, expand or focus their search.

HASSET
There are six basic relationships between the terms held in the thesaurus and these are held in one database table with the simple format of CONCEPT TERM - relationship type - RELATIONSHIP TERM. They are 1) Use 2) Use For - UF 3) Narrower Term - NT 4) Broader Term - BT 5) Top Term - TT 6) Related Term - RT.
Hence :-

DRINKING HABITS - use - ALCOHOL USE
i.e. “drinking habits” is a non-preferred synonym of the preferred term “alcohol use”
There will also be the reciprocal entry :- ALCOHOL USE - use for - DRINKING HABITS

ALCOHOLISM - broader term - ADDICTION
with the reciprocal entry :- ADDICTION - narrower term - ALCOHOLISM
i.e. There is a narrower concept “alcoholism” to the subject term “addiction”
Both subject terms “alcoholism” and “addiction” are in two hierarchies, one from the top term “diseases” and one from the top term “social problems”. Hence the following entries are found in the database table:

ALCOHOLISM - top term DISEASES
ADDICTION - top term - DISEASES
ALCOHOLISM - top term SOCIAL PROBLEMS
ADDICTION - top term - SOCIAL PROBLEMS

N.B. there is no reciprocal entry for a top term relationship, which acts as an aid to understand the scope and meaning of the subject concept under review, and in programming to build up the correct hierarchies.

![Hierarchical listing of term ADDICTION](image)

The final relationship is that between two preferred subject terms which are related to each other but are not covered by the NT, BT or TT relationships. The reciprocal entry is also included in the database table. Hence:-

ALCOHOL USE - related term - ALCOHOLISM

ALCOHOLISM - related term - ALCOHOL USE

HASSET does actually have two other database tables, one which holds a textual clarification of the subject term, known as a scope note (SN), and the second holds a classification code which places the subject term in one fixed hierarchy, so that HASSET could be used as a shelving scheme for hard copy documentation, and a marker to show whether the term was taken from the UNESCO thesaurus or is a Data Archive new term.
There are, at present, approximately 8,650 terms in HASSET; 2,500 of which are non-preferred terms or synonyms. 38,600 relationships exist between these terms and they form 296 hierarchies. Approximately half of the terms have been taken from the UNESCO thesaurus.

**MAINTENANCE & INTERFACES**

The tables described above are held in an INGRES database and updates to the thesaurus are performed through ‘C’ programs, written at The Data Archive, which employ embedded SQL calls to the underlying tables. The same interface also performs the indexing of the actual datasets with terms from the controlled vocabulary.

The program ensures that reciprocal entries are automatically included, terms are correctly positioned in hierarchies with the most appropriate allocation of classification code. The duplication of terms is impossible as is the creation of incorrect relationships between terms.

To aid the allocation of terms to the datasets held at The Data Archive, the indexer has recourse not only to the thesaurus, hierarchical and classification listings described above, but also the scope notes, listings of datasets previously indexed by the term under consideration and a KWIC (keyword in context) listing of words from the candidate term. The example below shows the kwic listing for the term “alcohol use”.

**Classified listing of term ADDICTION**

- R73 50/90 ADDICTION
- R73 55... GAMBLING
- R73 60... DRUG ABUSE
- R73 60 10... DRUG SOURCES
- R73 60 20... NARCOTIC DRUGS
- R73 60 30... STIMULATING DRUGS
- R73 60 40... HALLUCINOGENS
- R73 70... ALCOHOLISM
- R73 80... SMOKING
- R73 90... SOLVENT ABUSE

**KWIC listing of terms containing words in ALCOHOL USE**

- ALCOHOL
- ALCOHOL USE
- BLOOD ALCOHOL TESTING
- ALCOHOL USE
- BOOK USE
- BOOK USE FREQUENCY
- COSMETICS USE
- CREDIT CARD USE
- DRUG USE DETECTION
- ILLEGAL DRUG USE
BIRON & HASSET

The WWW interface to both HASSET and BIRON is through dynamically produced html forms from a cgi-bin ‘C’ program with embedded SQL calls to the underlying INGRES database tables.

How then does HASSET aid the searcher of The Data Archive’s on-line catalogue BIRON. First of all the indexing program ensures that the same subject concept in any dataset held is assigned the same controlled vocabulary term. BIRON’s first task then is to point the user to the preferred term, if the keyword searched on is not in itself a preferred term; it does this in three ways.

Firstly it searches the synonyms from the USE and UF relationships to see if it can find a match. If it does it will automatically substitute the preferred term and carry out a search immediately. If it cannot match against a non-preferred term then the program produces a KWIC listing from the word or words in the search term.

Finally, if the second option fails, BiRON will produce another KWIC listing, but this time from progressively truncating the search string until a listing is produced, even if it has to be a list of all terms in the thesaurus. Hence entering a slight misspelling of “alchol” results in :-

![KWIC listing of terms starting ALC%](image)

Therefore the searcher is always offered some candidate terms no matter what is entered as the search string. Selection is carried out by just clicking on the required term and then the on “search” icon to perform the search.

Once a search has been carried out the thesaurus is also available to help the searcher redefine their search through the displays described above, by changing to broader or narrower concepts, adding more terms to their search or combining the results from their present search, so that the datasets retrieved also cover the concept of another subject term or terms displayed.
Consider the following search:

<table>
<thead>
<tr>
<th>Search options</th>
<th>Fill in one or more</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject keyword</td>
<td>alcohol use</td>
</tr>
<tr>
<td>Geographical location</td>
<td>denmark</td>
</tr>
<tr>
<td>Year from</td>
<td>1990</td>
</tr>
<tr>
<td>Person/organisation</td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td></td>
</tr>
</tbody>
</table>

Which results in:

**BIRON 4.1 internal**

497 studies found for period 1990 to 1997
189 studies found for subject Alcohol Use
129 studies found for location Denmark
9 unique studies found from integrated search:

1. [3229] Euro-Barometer 39.0: European Community Policies and Family Life
2. [3228] Euro-Barometer 36: Regional Identity and Perceptions of the Third
3. [3227] Euro-Barometer 34.1: Health Problems, Fall 1990
7. [2968] Euro-Barometer 34.1: Health Problems, Fall 1990
8. [2959] Euro-Barometer 36.0: Regional Identity and Perceptions of the Th
9. [2830] Euro-Barometer 37.0: Awareness and Importance of Maastricht
Clicking on the thesaurus help icon displays the thesaural entry for “alcohol use”.

From which you could select the related term

Select the button and click on the icon which results in:

497 studies found for period 1990 to 1997
18 studies found for subject Drinking Offences
129 studies found for location Denmark
2 unique studies found from integrated search:

Or you could select more than one related term and click on the extend icon which results in:

497 studies found for period 1990 to 1997
55 studies found for subject Alcoholism, Drinking Offences
129 studies found for location Denmark
5 unique studies found from integrated extended search:

Then from “alcoholism” you could select the top term “social problems”.

and click on the followed by the to display the full hierarchy
Full subject hierarchy Social problems

Select one or more terms to extend search

and select up to ten terms from the listing of 269 terms. N.B. n4 indicates a narrower term 4 levels below the selected term. The maximum level for this hierarchy is n6.

All 269 terms can be selected for a search by returning to the thesaurus listing and selecting the following options before clicking on the extend icon.

Which results in:

497 studies found for period 1990 to 1997
6537 studies found for subject Social Problems and narrower terms to level 6
129 studies found for location Denmark
32 unique studies found from integrated extended search:

FUTURE DEVELOPMENTS
It must be remembered that HASSET has been constructed based on the 5,000 datasets held at The Data Archive as an indexing tool. So therefore the coverage only reflects the subject coverage of these datasets themselves, and because it is a controlled vocabulary it has not been specifically designed as a free text retrieval tool. However, its use within BIRON has seen an increase in the number of USE and UF relationships. So although the study descriptions and dataset documentation have not been trawled for candidate terms, which are then structured into a thesaurus, The Data Archive and several external
organisations are experimenting with using HASSET as a retrieval tool for free text searching.

The CESSDA (Council for European Social Science Data Archives) IDC (Integrated Data Catalogue) is based on a Z39.50-WAIS protocol and uses freeWais-sf and SFgate as its search engine and gateway.

One of the options when creating an index for a WAIS database is to have present a synonym file, however since WAIS indexes every word, apart from stop words such as and, the etc., the synonyms have to be single words themselves. There is also no facility for the narrower / broader type relationships or control over when to apply the synonyms to a search. Hence we have selected only single word terms with a USE relationship to another single word term and single word top terms that have a RT relationship with other single word top terms. Part of the NESSTAR project will be to investigate how HASSET can be employed more fruitfully across the distributed databases of the European data archives and whether a multi-lingual version is a viable option.

Other organisations have also shown an interest in HASSET, namely SOSIG (Social Science Information Gateway), MIDAS (Manchester Information Datasets and Associated Services), QUALIDATA (Qualitative Data Archival Resource Centre), the Steinmetz Archive for the EU-funded ILSES project, IBSS (International Bibliography for the Social Sciences) and the Office for National Statistics in the UK.

The most advanced of these is SOSIG who have a test interface on the WWW which they hope to incorporate into their search facility by June 1997. They have matched terms in the HASSET thesaurus against keywords used in their own database records.

By keying in a search string and selecting the ‘Any related terms’ button and clicking on ‘Do Look up’ the present test interface will return the number of direct matches and also any term from the HASSET relationships that are guaranteed to result in a match in SOSIG.
The Data Archive hopes to undertake a project later this year where these participating organisations help convert HASSET into a thesaurus resource for the whole of the social science community. Control and maintenance of HASSET will still remain the responsibility of The Data Archive, but the other organisations will offer up candidate terms and suggestion position in the hierarchies through a new WWW interface to HASSET. As well The Data Archive will also review the contents and structure of HASSET through analysis of the search logs from BIRON and a trawl of the study descriptions and recently digitised dataset documentation. It is hoped that this will also make HASSET a valuable, universally available, free-text retrieval tool.

Norwegian Social Science Data Services
A Large Scale Facility for European Research and Data Analysis

RESEARCH GRANTS - CALL FOR PROPOSALS

LSF funds of the European Union are available for supporting access to large comparative databases in the social sciences. The Large Scale Facility for the social sciences at the Norwegian Social Science Data Services (NSD) invites researchers to work with comparative data sets in the research environment of NSD, located at the University of Bergen, Norway.

What can NSD offer

NSD is a resource- and competence centre, providing data, substantive knowledge, methodological and technical assistance to social scientists, with a particular emphasis on the spatial dimensions of societal problems. Besides being one of the largest data archives in the world, NSD assists researchers with respect to data gathering, data integration, questionnaire design, selection of software tools, data analysis, methodology and privacy issues. A distinguishing characteristic of NSD is its multidisciplinary orientation. The data holdings, as well as the educational and intellectual background of the staff cover all major fields and disciplines of the social sciences.

NSD has a variety of data resources. Among these are all the major international survey collections, like the Eurobarometer, the International Social Survey Programme, World Values Surveys and international election studies. NSD also holds large databases with regional and historical data, census microdata, economic data, data about political elites, political parties and institutions.

Who can apply

All researchers from the EU member states and the Associated States Iceland, Liechtenstein and Israel, wanting to do comparative research may apply for participation. They must be entitled to publish the results of their work at the LSF in the open literature.

Financial support

Within the TMR program financial support is given to cover international travel and subsistence expenses in accordance with travel expenditure regulations for researchers participating in LSF-activities. Access to the LSF will be provided free of charge and will include access to training seminars as well as all infrastructural, logistical and scientific support that is normally provided to external users. Researchers may apply for a period ideally between two and four weeks.

How to apply

You have to fill in an application form. The proposals will be selected by an independent peer review procedure. There are no deadlines for application, but please be aware that the formal selection procedure may take some time. For application or further information, please contact:

Norwegian Social Science Data Services (NSD)
Hans Holmboesgt. 22
N-5007 Bergen, Norway
Tel: +47 55 58 21 17, Fax: +47 55 58 96 50
E-mail: nsd@nsd.uib.no, Homepage: http://www.nsd.uib.no

Winter 1997
The International Association for Social Science Information Services and Technology (IASSIST) is an international association of individuals who are engaged in the acquisition, processing, maintenance, and distribution of machine readable text and/or numeric social science data. The membership includes information system specialists, data base librarians or administrators, archivists, researchers, programmers, and managers. Their range of interests encompasses hard copy as well as machine readable data.

Paid-up members enjoy voting rights and receive the IASSIST QUARTERLY. They also benefit from reduced fees for attendance at regional and international conferences sponsored by IASSIST.

Membership fees are:
- Regular Membership: $40.00 per calendar year.
- Student Membership: $20.00 per calendar year.

Institutional subscriptions to the quarterly are available, but do not confer voting rights or other membership benefits.

Institutional Subscription: $70.00 per calendar year (includes one volume of the Quarterly)

Membership form

I would like to become a member of IASSIST. Please see my choice below:
Options for payment in Canadian Dollars and by Major Credit Card are available. See the following web site for details:
http://data.lib.library.ualberta.ca/iassist/mbrship2.html

- $40 (US) Regular Member
- $20 Student Member
- $70 Subscription (payment must be made in US$)
- List me in the membership directory
- Add me to the IASSIST listserv

Please make checks payable, in US funds, to IASSIST and Mail to:
IASSIST,
Assistant Treasurer
JoAnn Dionne
50360 Warren Road
Canton, MI 48187
USA

Name:
Job Title:
Organization:
Address:
City: State/Province:
Postal Code: Country:
Phone: FAX:
E-mail: URL: