Archiving travel data: The Mobidrive example
ETH Transport Data Archive

Author(s):
Chalasani, V.S.; Schönfelder, S.; Axhausen, Kay W.

Publication Date:
2002-07

Permanent Link:
https://doi.org/10.3929/ethz-a-004403590

Rights / License:
In Copyright - Non-Commercial Use Permitted
ETH Transport Data Archive

Archiving travel data: The Mobidrive example

VS Chalasani, S Schönfelder, KW Axhausen
Arbeitsbericht Verkehrs- und Raumplanung

Archivierung von Verkehrsverhaltensdaten: Das Mobidrive Beispiel

VS Chalasani
IVT
ETH
CH-8093 Zürich

S Schönfelder
IVT
ETH
CH-8093 Zürich

KW Axhausen
IVT
ETH
CH-8093 Zurich

Telephone:+41-1-633 33 40
Telefax: +41-1-633 10 57
chalasani@ivt.baug.ethz.ch

Telephone: +41-1-633 30 92
Telefax: +41-1-633 10 57
schoenfelder@ivt.baug.ethz.ch

Telephone: +41-1-633 39 43
Telefax: +41-1-633 10 57
axhausen@ivt.baug.ethz.ch

Juli 2002

Kurzfassung


Schlagworte

Mobidrive, Meta – Daten, XML, DDI, Datenarchiv, Verkehrsbefragungen, NESSTAR

Zitierungsvorschlag

Archiving travel data: The Mobidrive example

Abstract

The paper discusses the necessity and importance of archiving transport data, in particular travel behaviour data. It explicitly identifies the inefficient efforts in public usage of transport data. Though there are many organisations/institutions working seriously on archiving data in social sciences sector, a specific transport data archive was not yet been.

The paper provides the methodology to archive a typical transport data. Various tools used in archiving data are described. An open methodology is provided in order to take care of the diversity of data formats.

Mobidrive- a six week travel diary was archived at Institute of Transport planning and Transport systems (IVT), ETH Zurich. The issues raised and the way they were addressed while archiving a long term travel behaviour dataset are discussed in this paper. The salient features of secondary usage of the archived data are emphasised. The tools useful for secondary analysis are described. Safe and secure data issues also are addressed from data administrator’s perspective.

Keywords

Mobidrive, Data archive, Metadata, DDI, XML, Travel surveys, NESSTAR, NSD, Data builder, XML generator, Secondary usage, database.

Preferred citation style

1 Introduction

1.1 Transport data and the necessity of archiving

The art of collecting, analysing, preserving and reanalysing data has proved to be important and necessary in all sciences. In transport planning, specifically for travel behaviour analysis, data holds the key role and is the main resource for a study. Today there are many institutions, organisations, persons involved in collecting transport data by conducting surveys. The data collected by such a diversity of surveys is diverse and difficult to understand. Fundamental reasons for this diversity in data are differences in the

- Purposes of the surveys.
- Place and the universe of the sample.
- Available resources and other constraints.
- Definitions of variables and categories.

The complexity of the data restricts the data set’s domain both in time and space. Still, the data could be used more generally if proper documentation were available. For an efficient use of the available resources, data set format standardisation is essential. Data documentation standardisation allows the successors or next generations to understand and make the data set useful for their purposes. Metadata is one of the most efficient techniques for data documentation standardisation. A metadata standard was developed by the Data Documentation Initiative (DDI)\(^1\), which basically uses the XML DTD technology. The Networked Social Science Tools And Resources (NESSTAR)\(^2\) project build books for data archiving and analysis around this standard. To archive the data set at hand-Mobidrive, the methodology developed by NESSSTAR was followed.

Mobidrive\(^3\) is a six week travel dairy, which captures the rhythms of daily life over a long period. Two cities with similar characteristics were chosen for this study: Karlsruhe and Halle,

\(^1\) www.icpsr.umich.edu/DDI
\(^2\) www.nesstar.org
\(^3\) www.ivt.baug.ethz.ch/vrp/projekte_mobidrive_d.html
both in Germany. The survey was conducted in two phases, pre-test and main study. The Mobidrive data was analysed using SAS 8.1, which implies the original data format.

The archiving of the Mobidrive data according to ddi metadata standards will be explained in forthcoming sections.

1.2 Background

Though data archiving has existed for some time, it was restricted to mainly to the social sciences. The current work i.e. archiving Mobidrive is probably the first travel behaviour dataset, that is to be archived under metadata standards. If a database is considered as knowledge, it blossoms only by enhancing its quantity as well as its quality or by on-going use. One such way of increasing the quantity and quality of a database is data exchange, exchanges of databases of similar characteristics in particular. Data exchange has existed in the past, but not to the desirable extent possible. This is due to the lack of sophisticated communication facilities and other resources. The time taken in understanding foreign dataset(s) was enormous. In order to perform secondary analysis, one should be thoroughly familiar with the dataset(s) and the primary analysis, that has been conducted previously. The above stated difficulties and obstacles discourage the use of foreign dataset(s) and of older dataset(s). All these problems can be addressed by proper documentation of dataset(s). Major advantages of archived data are greater accessibility, uniformity in documentation and easier to perform secondary analysis. One can access the database, get to know its details and is able to perform secondary analyses in a more efficient and easier manner.

The European data archives collaborate intensively and offer a joint catalogue of their holding. Many of these are switching to the common data documentation standard Data Documentation Initiative (DDI) mentioned above, developed by the Inter-University Consortium for Political and Social Research (ICPSR), Michigan, USA (see www.icpsr.umich.edu).
2 Metadata and data archiving

Metadata is defined as the data associated with objects which relieves their potential users of having to have full advance knowledge of their existence or characteristics. In essence, metadata answer the who, what, when, where, why, and how about the every facet of the data that are being documented. Metadata can also provide evaluative information on an object. Metadata can viewed from two perspectives:

Machine understandable metadata: It provides knowledge about data to software processes (configuring interfaces, driving transformations, sub-setting, access control, disclosure control, etc.).

Human understandable metadata:

- It helps researchers identify datasets which meet certain basic criteria related to their research interests. Study level metadata facilitate the identification, evaluation and retrieval of datasets. Full and explicit documentation of the origin, contents, location and accessibility of data is provided by metadata entities and attributes and their organisation. Hence, the planner / user should be able to identify studies, which have the potential to provide needed data. This will more easily be the case if the metadata is compatible with existing means of data searching.

- It should facilitate an evaluation of the relevance, efficacy and quality of the data identified.

- It should facilitate the retrieval of identified datasets.

In order to maintain uniformity, international standards for metadata schema were set and the most popular metadata schemas are:

- Dublin core (http://dublincore.org)
- AACR2 (Anglo-American Cataloguing Rules) (www.nlc-bnc.ca/jsc/docs.html)
- ISAD(G) General International Standard of Archival Description. (www.ica.org/biblio/com/cds/isad_g_2e.pdf)

There are numerous standards or initiatives based on these metadata schemas. Current metadata standards or initiatives are:
• DDI (the Data Documentation Initiative) ([www.icpsr.umich.edu/DDI/](http://www.icpsr.umich.edu/DDI/)) is an XML-DTD developed within the world of social science data archives to describe archived data (mainly rectangular survey-files).

• The Cristal data object model ([www.faster-data.org/Metadata/papers/Cristal.htm](http://www.faster-data.org/Metadata/papers/Cristal.htm)) is under development at CBS and the Statistical Open Source group and aimed at a generalised object model for describing micro datasets and multidimensional cubes.

• GESMES([www.ecb.int/stats/gesmes/gesmes.htm](http://www.ecb.int/stats/gesmes/gesmes.htm)) is a Eurostat supported metadata interchange protocol.

• ISO11179 ([www.diffuse.org/meta.html](http://www.diffuse.org/meta.html)) is an ISO/IEC standard for description of data elements

• CWM - Common Warehouse Metamodel([www.cwmforum.org/](http://www.cwmforum.org/)) is an Object Management Group (OMG) effort to create a metamodel for easy interchange of warehouse metadata between warehouse tools, warehouse platforms and warehouse metadata repositories.

• BRIDGE ([www.unece.org/stats/documents/1998/02/metis/7.e.pdf](http://www.unece.org/stats/documents/1998/02/metis/7.e.pdf)) is a classification model developed by several European Statistical offices (within the IMIM project).

Among the above the most advanced, flexible and user friendly standard is DDI. In the current work, DDI was the chosen metadata standard and the data archiving was performed using tools developed by Networked Social Science Tools and Resources (NESSTAR).

### 3 DDI, NESSTAR and NSD

The Data Documentation Initiative (DDI) was established in 1995 to create a universally supported metadata standard for the social science community. DDI was initiated and organised by the Inter-University Consortium for Political and Social Research (ICPSR), at the University of Michigan. DDI was an improvement on the previous OSIRIS metadata standard, which was used to document files at the University of Michigan. The DDI consortium comprises members from social science data archives and libraries from USA, Canada and Europe and from major producers of statistical data (like the US Bureau of the Census, the US Bureau of Labour Statistics, Statistics Canada and Health Canada). The consortium developed a first version of the standard which was expressed as an SGML-DTD. It was translated into XML in 1997. The first official version of XML based DDI-DTD was published in March 2000 ([www.icpsr.umich.edu/DDI/Version1.dtd](http://www.icpsr.umich.edu/DDI/Version1.dtd)).
The European data archives in the consortium have adopted the DDI DTD. Most of the European data archives adopted the DDI DTD based on common tools. The organisation which is developing these data archiving tools is NESSTAR. NESSTAR has developed an integrated common platform to archive a dataset under DDI DTD standards. This was done with the help of the Norwegian Social sciences Data services (NSD), who developed associated tools like Data Builder, XML Generator and is now developing Web DIAS Publisher- a combination of data builder and XML generator. NESSTAR is also supporting the concept of a Data Web creation(http://www.nesstar.org/papers/NesstarOverview.ppt), which can be described as:

- To embed hyperlinks to external Web objects in every metadata element.
- Ability to describe any resource on a NESSTAR server as a hyperlink, which provides an environment for tight integration between data and other web resources.
- To bring live data into on-line texts.
- Linking external documents and objects to the data using the metadata as a bridge.

Some of the salient features in using NESSTAR are:

- A common architecture for a totally distributed virtual data library.
- Ability to locate multiple data resources across organisational boundaries.
- Ability to browse detailed information about these data sources.
- Ability to perform simple data analysis and visualisation over the net.
- Downloading an appropriate subset of data in one of a number of formats.
- Convert the data from existing source into a web friendly format.
- Manage a distributed data server with simple to use administrative tools.

### 3.1 DDI DTD

DDI is a standard for the content, presentation, transport, and preservation of metadata about datasets in the social and behavioural sciences. While DTD is a mark-up language, which maintains the integrity of document content and structure by defining all the document types and their structural elements and by defining their relationships among the elements.

All the archived data has to be described using the standard DDI DTD codebook in XML (in social sciences metadata are often called codebooks). A typical DDI DTD codebook’s information structure can be viewed in Figure 1 and sample outline code of a DDI DTD code book in Figure 2.
Figure 1  DDI codebook structure

```
Codebook

Document description
Study description
Data files description
Variable description
Other documentation

Title          Study information          Name          Name          Other material
Responsibility Abstract          Contents          Label          Notes
Production     Methodology           Structure       Type
Distribution   Processing         Dimensions       Range of data values
Series
Version
```
DDI Codebook Outline -- Version 1.02.1

* == ELEMENT IS OPTIONAL & REPEATABLE
+ == ELEMENT IS MANDATORY & REPEATABLE
? == ELEMENT IS OPTIONAL & NON-REPEATABLE
== ELEMENT IS MANDATORY & NON-REPEATABLE

0.0 codeBook (ATT == ID, xml:lang, source, version)

|---- 1.0 docDscr* (ATT == ID, xml:lang, source)
|     |---- 1.1 citation? (ATT == ID, xml:lang, source, MARCURI)
|     |     |---- 1.1.1 titlStmt (ATT == ID, xml:lang, source)
|     |     |     |---- 1.1.1.1 titl (ATT == ID, xml:lang, source)
|     |     |     |---- 1.1.1.2 subTitl* (ATT == ID, xml:lang, source)
|     |     |     |---- 1.1.1.3 altTitl* (ATT == ID, xml:lang, source)
|     |     |     |---- 1.1.1.4 parTitl* (ATT == ID, xml:lang, source)
|     |     |     +---- 1.1.1.5 IDNo* (ATT == ID, xml:lang, source, agency, level)
|     |     |---- 1.1.2 rspStmt? (ATT == ID, xml:lang, source)
|     |     |     |---- 1.1.2.1 AuthEnty* (ATT == ID, xml:lang, source, affiliation)
|     |     |     +---- 1.1.2.2 othId* (ATT == ID, xml:lang, source, type, role, affiliation)
|     |     |---- 1.1.3 prodStmt? (ATT == ID, xml:lang, source)
|     |     |     |---- 1.1.3.1 producer* (ATT == ID, xml:lang, source, abbr, affiliation, role)
|     |     |     |---- 1.1.3.2 copyright? (ATT == ID, xml:lang, source)
|     |     |     |---- 1.1.3.3 prodDate* (ATT == ID, xml:lang, source, date)
|     |     |     |---- 1.1.3.4 prodPlac* (ATT == ID, xml:lang, source)
|     |     |     |---- 1.1.3.5 software* (ATT == ID, xml:lang, source, date, version)
|     |     |     |---- 1.1.3.6 fundAg* (ATT == ID, xml:lang, source, abbr, role)
|     |     |     +---- 1.1.3.7 grantNo* (ATT == ID, xml:lang, source, agency, role)
|     |     |---- 1.1.4 distStmt? (ATT == ID, xml:lang, source)
|     |     |     |---- 1.1.4.1 distrbtr* (ATT == ID, xml:lang, source, abbr, affiliation, URI)
|     |     |     |---- 1.1.4.2 contact* (ATT == ID, xml:lang, source, affiliation, URI, email)
|     |     |     |---- 1.1.4.3 depositr* (ATT == ID, xml:lang, source, abbr, affiliation)
|     |     |     |---- 1.1.4.4 depDate* (ATT == ID, xml:lang, source, date)
|     |     |     +---- 1.1.4.5 distDate? (ATT == ID, xml:lang, source, date)

Source: [http://www.icpsr.umich.edu/DDI/codebook.html](http://www.icpsr.umich.edu/DDI/codebook.html)
4 Archiving tools

The various archiving tools developed by NESSTAR and NSD are:

- Data Builder.
- XML Generator.
- NESSTAR Publisher.
- NESSTAR Light.
- NESSTAR Explorer.
- Web DAIS Publisher.

All of the above will be discussed in detail below with respect to developer, input and processing steps.

4.1 Data Builder

Data Builder was developed by NSD. Data Builder allows to construct a dataset or import a pre-formatted dataset. One of the major advantages of Data Builder is enabling the import of different formats, such as SPSS, txt, dbf, etc. Data Builder contains a DDI DTD codebook transformer and it automatically transform any dataset that has been imported into it. Main functions of Data Builder are:

- Reading the dataset and develop a DDI DTD codebook for the same.
- Reading the each entry in the dataset and creating the appropriate slots in the DDI DTD codebook.
- Produces four files as output
  - *.nsf
  - *.dkm
  - *.ind
  - *.xml
- Facilitating the data entry for the first three parts (i.e. Data Documentation, Study Description and File Description) of the DDI DTD codebook (see Figure 2).

Data Builder is the most important archiving tool. In case of a pre-formatted dataset, data builder will immediately provide DDI DTD codebook source code of high quality and precision. Data Builder also facilitates label editing, adding or editing labels for the respective category values.
4.2 XML Generator

XML Generator was developed by NSD. It is mainly useful to edit additional information about 16 categorical descriptions of the variables, such as:

- Literal / Pre / Post question text.
- Response unit / Universe.
- Weighting.
- Interviewer instructions, etc.

XML Generator needs a *.nsf file from the Data Builder. It will not affect the original data, but adds additional information into the *.xml file. XML Generator only makes changes in the *.xml file, and does not corrupt the other three files which have been produced by Data Builder. However it produces a *.xgs file, which is useful in reopening changed or edited *.xml files.

XML Generator also provides a facility to import a pre-formatted dataset, but it has no provision to construct a new dataset. One can import a pre formatted dataset directly into XML Generator, if there is no label editing. As the first three parts of the DDI DTD codebook can be filled only in Data Builder, it is better to use XML Generator after Data Builder.

One of the most important features of XML Generator is to define or modify the data structure that is going to be used. Before going into details, the concepts “variable grouping” and “hierarchical datasets” will be briefly introduced, i.e. clustering variables under a common heading.

4.2.1 Variable groups

Often there is an enormous amount of data and a large number of variables. It is difficult for a new user to get know about each variable and its content. One way of simplifying the complexity of a dataset is variable grouping.

XML Generator has a provision for variable grouping at the individual dataset level. XML Generator keeps track of variable groups created in its output file *.xgs.
4.2.2 Hierarchical dataset

Most of travel behaviour surveys consist of a series of surveys. Series of periodical surveys, for example the Swiss Micro Census will be taken for every 5 years or a series of surveys, during the same time period but at different locations. As the theme behind these series is same, the individual datasets can be combined. If one considers detailed surveys like travel diaries, different elements of these surveys will be conducted for different analysis units. These specific datasets address different topics. It is often difficult to connect two variables from two different datasets. A hierarchical dataset structure guides the user through these complexities. A hierarchical dataset is one which contains two or more datasets. To link these datasets, basic key variables are needed.

XML Generator facilitates the definition of hierarchical datasets. Basic key variable selection plays vital role in hierarchical dataset formation. A variable whose values are unique is a valid basic key variable. Each file should contain at least one such valid basic key variable. In hierarchical dataset formation, only the DDI DTD codebook source code will be added and the data files (.nsf, .dkm and .ind) will remain separate. However, when uploading a hierarchical dataset, all the relevant individual data files must be uploaded according to their file number. The hierarchical data structure will be saved in an *.hxgs output file. To reopen a hierarchical dataset, the *.hxgs file is essential.

4.3 NESSTAR Publisher

NESSTAR publisher was developed by NESSTAR. Once an *.xml file is generated by XML Generator or Data Builder, literally it is ready to be published i.e. uploading it to the server is possible. To ensure that everything is correct and to avoid unwanted errors like typing mistakes, etc., the *.xml file should be parsed (checked for errors) and one should observe the visualisation of the dataset. Using NESSTAR Publisher, these functions can be performed. NESSTAR Publisher will parse the file and then it publishes the DDI DTD codebook source code written in xml. Thus NESSTAR Publisher is a useful tool to avoid errors. Only the data publisher or administrator can access the NESSTAR Publisher software.

4.4 NESSTAR Light

NESSTAR Light was developed by NESSTAR. A file that is available as a virtual data resource cannot be browsed by an ordinary browser (like Internet Explorer, Netscape Navigator,
etc.). NESSTAR Light is a web based browser, which will interact with the NESSTAR server and allows to query and visualise the data. NESSTAR Light is driven by the NESSTAR Light engine, which comes with the server. Unless one knows the server page (i.e. IP address), it is not possible to get access to NESSTAR Light. Despite this drawback, the advantages of NESSTAR Light usage are:

- No need to download of a dedicated browser.
- Web based browser, which allows for greater accessibility.
- Secondary analysis and sub-setting is possible with greater ease then with the alternative NESSTAR Explorer.

4.5 NESSTAR Explorer

NESSTAR Explorer is a special web based browser with NESSTAR Light functionality. The major difference between NESSTAR Light and NESSTAR Explorer is that, NESSTAR Explorer is an independent software and needs to be downloaded on a personal computer. It also interacts with the server to perform secondary analysis or to download a subset of the dataset. NESSTAR Explorer is an ideal browser for a remote user, while NESSTAR Light useful in case of known users. NESSTAR Explorer is downloadable free of charge at www.nesstar.org/explorer/.

4.6 Web DIAS Publisher

This is an improved and sophisticated tool, which is currently under development. Web DIAS Publisher is the combination of Data Builder and XML Generator.
5 Steps required

The steps required in archiving a travel behaviour dataset are:

- Generating and modifying the DDI DTD codebook source code: This task can be performed either with Data Builder or with XML Generator. As it is essential to fill the first three parts (administrative details) of the DDI DTD codebook, it is suggested to start always with Data Builder.

- Specifying additional information on categories: XML Generator allows to add this information.

- Defining the data structure: Before adding any categorical description, it is highly recommended to define or modify the data structure.

- Parsing and visualising the DDI DTD codebook with the NESSTAR Publisher.

- Dataset uploading: NESSTAR Server should be used to perform this task.

- Browsing the dataset: Either NESSTAR Light or NESSTAR Explorer can be used to browse, visualise, subset and performing secondary analysis on the archived dataset.

For a complete picture of the flow of data from one tool to another, see Figure 3., which details the data flow and the interaction among the tools.
Figure 3  Data flow in archiving.

Legend:

*.bnd: NSDstat data entry
*.btd: NSDstat data entry[DOS]
*.dbf: Dbase
*.dif: DIF
*.dta: Stata
*.por: SPSS portable file
*.sav: SPSS system file
*.sta: Statistica
*.txt: ASCII
*.xml1: XML file with additional or corrected information.
5.1 Data formatting

Following file formats are recognised by Data Builder for automatic importing.

- DIF (*.dif)
- Dbase (*.dbf)
- ASCII (*.txt)
- Stata (*.dta)
- Statistica (*.sta)
- SPSS system file (*.sav)
- SPSS portable file (*.por)
- NSDstat data entry (*.bnd)
- NSDstat data entry[DOS] (*.btd)

Data importing: In case of a dataset is not in one of the above formats, it has to be converted into one of the above formats. In Data Builder, while importing the dataset, a window appears with two options. One of the options is for alphanumeric variables, either to convert the character variables into numeric variables or dropping them. Opting for the first choice will convert each character variable value into a numeric value (Data Builder allocates values sequentially) and the character variable value becomes the label of the value. The other option is about NSDstat variable labels. This option allows the Data Builder to import the variable name or label or both. This is a drawback of the Data Builder, while it converts all the character variables into a numeric format and it publishes them as numeric ones. Because of this all the variables that are published are declared as numeric though some of them are character variables.

Data Builder allows the original and direct entry of a dataset, but this will be unusual in most cases. Data Builder also permits to edit or alter the imported dataset.

Editing category labels: As the major part of travel data is likely to be numeric in nature, appropriate labelling for each variable is essential. For each of the variables and for each unique observation of a particular variable, appropriate labels will be defined. Data Builder is an ideal platform for labelling. By tracking the minimum and maximum values of a particular variable, labelling becomes also an error check.

Filling the first three parts of codebook: While importing the dataset, Data Builder will retrieve all the available information in the dataset and prepares the codebook. The codebook
comprises of five parts, document description, study description, data file description, variable description and other material. Up to this point, emphasis was given to variables and their description, which is the fourth part of the codebook. The last and fifth part i.e. other material is an optional one and most codebooks will be without this.

Among the first three parts of the codebook, the administrative details of the current documentation will be written in the first part of the codebook i.e. Document description. The current study details will be added in the second part of the codebook. The third part of the codebook contains information about the data file description.

The codebook that appears in Data Builder shows only the filled slots. In order to provide the comprehensive information, additional items have to be added to codebook. This task can be achieved by expanding the codebook and filling in the extended topics. A right click on the topic will display all the subtopics under the same. For each topic or subtopic, detailed description will be provided at the bottom, which makes the publisher’s job easy. A comprehensive knowledge of the survey and the data is essential in filling the first three parts of the codebook and it is always suggested that data administrator and publisher to be the same. Any useful information or material can be stored in fifth part of the codebook. This move will be the last in Data Builder.

Data structure (Hierarchical dataset and variable grouping): XML Generator allows to open a dataset with an *.nsf extension file. It is always suggested to finalise the data structure before adding the categorical description to the variables. A typical hierarchical dataset consists of one or more levels and one or more datasets at each of these levels. These levels also further hierarchical, i.e. under a level there may be one or more further levels. Usually a level will be used to represent a particular study or survey in a series of studies or surveys. After selecting the level where a particular dataset is to be added, a right click on the selected level will display all the possible methods that are relevant to the level. All the relevant datasets are to be added at appropriate levels as above. The most important part of the hierarchical dataset is defining the identification of the basic key variables. Each file should contain at least one basic key variable. For each of the datasets a valid basic key variable has to be specified. By clicking the ‘plus’ sign in the basic key variables section, all the variables in the particular dataset will be displayed. By selecting the variable which strictly comes under the basic key
definition, basic key variables will be added. After completing the datasets hierarchy, variable grouping at each dataset level has to be carried out. Select a particular file and click on variable groups window, variables groups can be created like in dataset hierarchy, i.e. by right clicking the mouse. After creating a new variable group, the variables whose characteristics fall in the range can be identified. Though XML Generator allows hierarchy in variable grouping, it is recommended to limit the sub-grouping to two, as more than that may create unnecessary complexity. This procedure has to performed for all the datasets.

Variable description: In total there are 16 variable descriptives, which XML Generator allows to be added to the codebook. Variable description will be performed comfortably by variable. While carrying out variable description, emphasis should be given to the relevant topics. Before generating the xml code, it is essential to check the following options. The option ‘Exclude variables with decimals from frequencies’ is on by default. It should be deselected. By generating the new enhanced XML based DDI DTD codebook, the major part of the data archiving have been performed.

Publishing the archived data in NESSTAR publisher: After opening the respective xml file in the NESSTAR publisher, go to ‘NESSTAR Preview’ to visualise the archived data. Browsing the dataset will result in identifying the mistakes that were committed in the earlier steps. These mistakes (such as typing mistakes, missing or additional information, etc.) have to be corrected at their origin.

Uploading into NESSTAR server: Once the dataset is free of errors, it can be uploaded on server. Uploading the archived data into server requires permission. With the help of the server administrator, the publisher will be able to get this permission. NESSTAR Server allows a valid publisher to upload the dataset from a remote system. The following steps illustrates the procedure to upload a typical dataset on a NESSTAR Server.

Step 1: Open the welcome page of the server (with its IP address) in a general navigator such as Internet Explorer, Netscape Navigator and click on frames version (IE5 & Netscape6 optimised) or no frames version.

Step 2: Among the properties of the catalogue, select property ‘catalogs’ and click on its value (a hyperlink to the list of catalogs hosted by the server: http://<serverIPadress>/obj/catalogs/).
Step 3: Among the available catalogs, select the relevant catalog in which the current dataset is to be uploaded by clicking the hyperlink
http://<serverIPadress>/obj/catalogs/<catalogname>

Step 4: Select the method ‘AddDataset’ to add a dataset to the parent catalog. Uploading the *.xml file allows the user only to explore the dataset. To upload an *.xml file, the data publisher should be a valid user of the server.

Step 5: To add the data files, select the property ‘datasets’ which displays all the datasets that are hosted by the server, select the newly uploaded dataset.

Step 6: By selecting the method ‘AddDataFile’, all the data files can be uploaded on to the server. When uploading a specific data file, the associated *.ind, *.dkm and *.nsf files will be uploaded. Successful uploading of data files allows to perform secondary analysis.

Browsing the archived data: Using two browsers, NESSTAR Light and NESSTAR Explorer, a valid user will be able to access, visualise, subset, download and perform secondary analysis of the archived datasets.

6 Summary and outlook

Eidgenössische Technische Hochschule Transport Data Archive(ETHTDA) is a new data archive that has been established as part of the current work, archiving of its initial travel behavioural diary dataset, the Mobidrive data (Axhausen, K. W., Schönfelder, S., .

While exploring the archived data with a NESSTAR browser, in the detailed description of a variable, categorical and numerical data is represented corresponding in a tabular form. This can be confusing, if the number of values is large. While archiving, a substantial part of the data needs to be reorganised into new variables with a limited number of categories. Neither Data Builder nor XML Generator allow to perform such re-coding. NESSTAR is developing tools to fill this gap in an upcoming version.

At Institute of Transport Planning and Systems (IVT), ETH Zurich, a NESSTAR server has been installed at http://129.132.96.89. All interested individuals or organisations are requested to contact axhausen@ivt.baug.ethz.ch or chalasani@ivt.baug.ethz.ch in order to obtain access to ETHTDA.
7 Acknowledgements

We are thankful to the colleagues at NESSTAR and NSD for their constant help throughout the project. We express our special thank to Mr. Simon W Musgrave, Marketing Director, Nesstar.org for his special interest.

8 Literature


9 Uniform Resource Locators


Eine vollständige Liste der Berichte kann vom Institut angefordert werden:

IVT ETHZ
ETH Hönggerberg (HIL)
CH - 8093 Zürich

Telephon: +41 1 633 31 05
Telefax: +41 1 633 10 57
E-Mail: hotz@ivt.baug.ethz.ch
WWW: www.ivt.baug.ethz.ch

The Working Papers Traffic and Spatial Planning are intended for the quick dissemination of the results of the members and guests of the Institute. Their content is the sole responsibility of the authors.

A complete catalogue of the papers can be obtained from:

http://www.ivt.baug.ethz.ch/veroeffent_arbeitsbericht.html