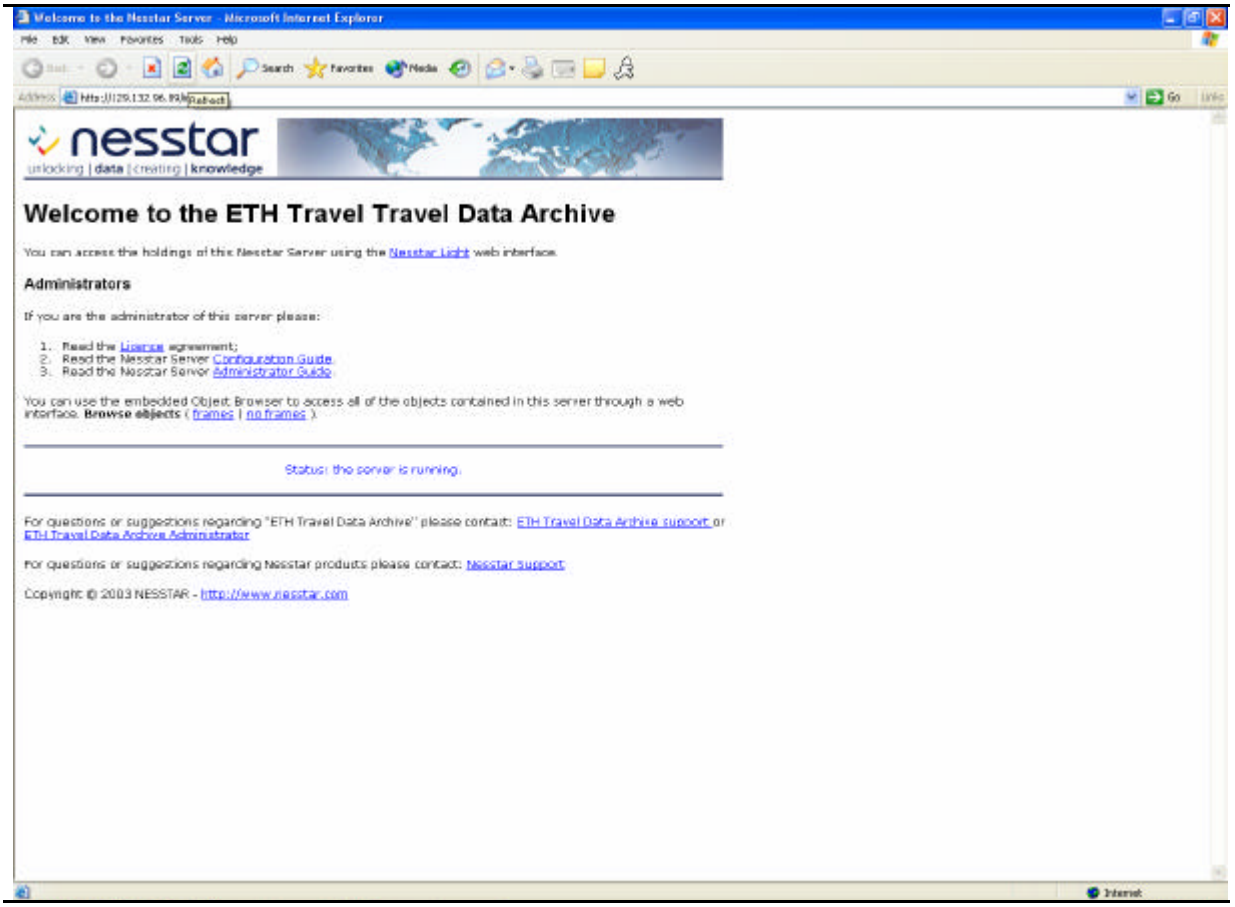


Travel data archiving: An art of presenting and preserving travel survey data

VS Chalasani, IVT ETH Zürich

Conference paper STRC 2004

STRC 4th Swiss Transport Research Conference

Monte Verità / Ascona, March 25-26, 2004

Travel data archiving: An art of presenting and preserving travel survey data

VS Chalasani
IVT
ETH Zürich
Zürich

Phone: 01-633 33 40
Fax: 01-633 10 57
e-Mail: chalasani@ivt.baug.ethz.ch

Abstract

Partial use of travel data and information in secondary analysis is a major concern in transportation planning, where understanding of the historical developments is a pivotal issue. Loss of valuable data and information could be due to improper documentation and inefficient preservation. Besides, data sharing allows understanding the relevant studies elsewhere in the world. However, sharing does not make sense without proper preservation and presentation of the data and information. By archiving the survey datasets, one can achieve the stated functions, which keep the data alive and a treasure of future. But, archiving diversified and complex travel data demands a word wide standard. This report discusses the existing metadata standard i.e. Data Documentation Initiative (DDI) and its suitability to travel data archiving. By exploring the advantages and limitations of the existing DDI, it was concluded that the existing DDI is not enough to prepare a complete travel metadata standard. An attempt is made in developing an exclusive travel DDI which inherits the existing DDI stricture but, has a few additional elements to accommodate the missing metadata. An exclusive travel data archive, ETH Travel Data Archive, was established at this institute and datasets from a wide range of travel surveys were archived. Two sets of archiving tools developed by NESSTAR were considered in this study. Taking the merits of these archiving tools, this study has come up with a viable solution to archive the travel data. This viable solution is briefly outlined in this report.

Keywords

Data archiving – travel surveys – Data Documentation Initiative (DDI) – NESSTAR – ETH Travel Data Archive - 3rd Swiss Transport Research Conference – STRC 2004 – Monte Verità

1. Need for travel data archiving

Travel data is mainly collected from various surveys, with wide range of instruments starting from relatively simple roadside interviews to the complex travel dairies. Enormous amounts of resources involved (both human as well as financial) have made the travel data precious and a treasure for the future. Survey methodology contains a series of logical and connected stages. Various factors influence different stages of the travel survey. Information about these factors and their influence is vital in data interpretation. Data and information are often confused. Data is the basic material collected, while information is the process and meaningful result of establishing its meaning and characteristics in a form that can be communicated to others. Data files mostly are with numbers and depending on the stored format it may contain codes. In addition to the basic data files, associated information obtained from a survey can be classified into two categories:

- Explanatory information: Explanatory information is essential to the informed use of a dataset. Without this information, no full understanding of the dataset and its contents can be achieved. This is the minimum information that should be created and preserved. Explanatory information comprises of the following:
    - Information about data collection methods
    - Information about the structure of the dataset
    - Technical information
    - Variables and values, coding and classification schemas
    - Information about the derived variables
    - Weighting and expansion
    - Data source
    - Confidentiality, validation, and checks
- Contextual information: Contextual information explains users about the context in which the data were collected and information about the uses to which the data were put. This forms a comprehensive historical record for future researchers. Whilst not essential, inclusion of this information is strongly recommended.

Many datasets are collected with a set of well planned objectives. Once the overall aim has been achieved, there is a tendency amongst some data producers to consider the life of study to be complete. In practice, there are numerous examples where travel data has been subject to secondary analysis and has proven to be valuable. Data usability has long been important in

transportation, especially in transport planning and forecasting. By preserving the data and making it available for future analysis, the life of the data can be increased and the potential of the study can be heightened. However, data producers or initial users of the data are usually aware of many of the peculiarities and special features of the data that they collect, and the interpretations and limitations that need to be applied. This information is highly essential in secondary analysis and use. Efficient documentation of this information allow the secondary user in better understanding the survey methodology applied and preliminary analyses performed, which is essential for the secondary analysis. Thus, in addition to the data preservation, effective secondary analysis requires good documentation. Another important feature of the data is sharing. By providing efficiently documented information over the internet enables faster access to the required datasets. In fact, "Document searches" is one of travel data collection types (Richardson, Ampt and Meyburg, 1995). Combining these three issues in a single framework would certainly help both researchers and users in understanding the data quickly and efficiently.

Because of the wide range of data collection methods and various factors influencing the survey methodology, travel survey data is diversified. Fundamental reasons for this diversity in data are differences in the

- purposes of the surveys
- universe and sampling procedure
- available resources and other constraints, and
- definitions of variables and categories.

Efficient documentation of this diversified data needs a worldwide standard. An exclusive travel data documentation standard can bring the travel data under a standard structure and end-users can certainly benefit by adjusting to this standard. Thus, the travel data can be made available to the general public.

Other important issues of interest are the tools to document the travel data and a common platform to share. Primary objective in developing these tools is to generate automated data documentation under the standard travel data documentation specifications. Development of a virtual platform is essential to cater the sharing needs of documented survey datasets. In brief, the following are the essential requirements for efficient travel data archiving:

- Efficient data documentation: A general structure that accommodates the diversified travel data

- Documentation tools: Tools that bring three categorized documents under a common roof and publish the metadata
- A virtual common platform: To store the documented data for sharing
- Search engine: To search the relevant surveys and browse the documented data.

Data archiving fully meets the above requirements. Data archiving is an art of presenting and preserving the data and associated information. This report elaborates the data archiving and its need in transportation sector.

The main objective of this paper is to explain the need for presenting and preserving travel data and its advantages. However, the following issues are also discussed in this report:

- Data archiving and its benefits
- Significance of a world-wide data documentation standard
- Available tools to archive the travel data; advantages and disadvantages in travel data archiving, and
- Recommendations for an exclusive travel data documentation initiative.

This paper is organized in the following way: The state-of-the art on data archiving is briefly discussed in the next chapter (chapter 2). Data documentation and the significance of a world-wide standard are discussed in chapter 3. Chapter 4 mainly discusses the experiences in travel data archiving and the tools used in data archiving. Recommendations and possible alterations to the existing world-wide data documentation standard for the evolution of an exclusive travel data documentation standard are discussed in Chapter 5. Conclusions are specified in Chapter 6.

2. Data archiving

Since from its evolution in the mid-1960s, when the Inter-university Consortium for Political and Social Research (ICPSR) formed, until late-1990s data archiving is mainly confined to political and social sciences sectors. Ever growing technologies have led the data archivists from a simple preservation of codes used in survey samples to a multi-phased process that archives enormous amount of data from different sectors. Most of the existing data archives at national level and belong to political and social science sectors. However, the broad standards in data archiving have motivated the archivists and researchers in other sectors. Recent years has experienced data archivists from different sectors such as health sciences, environmental sciences, etc.

Figure 1 Archiving data from a survey and interaction between archived data and user

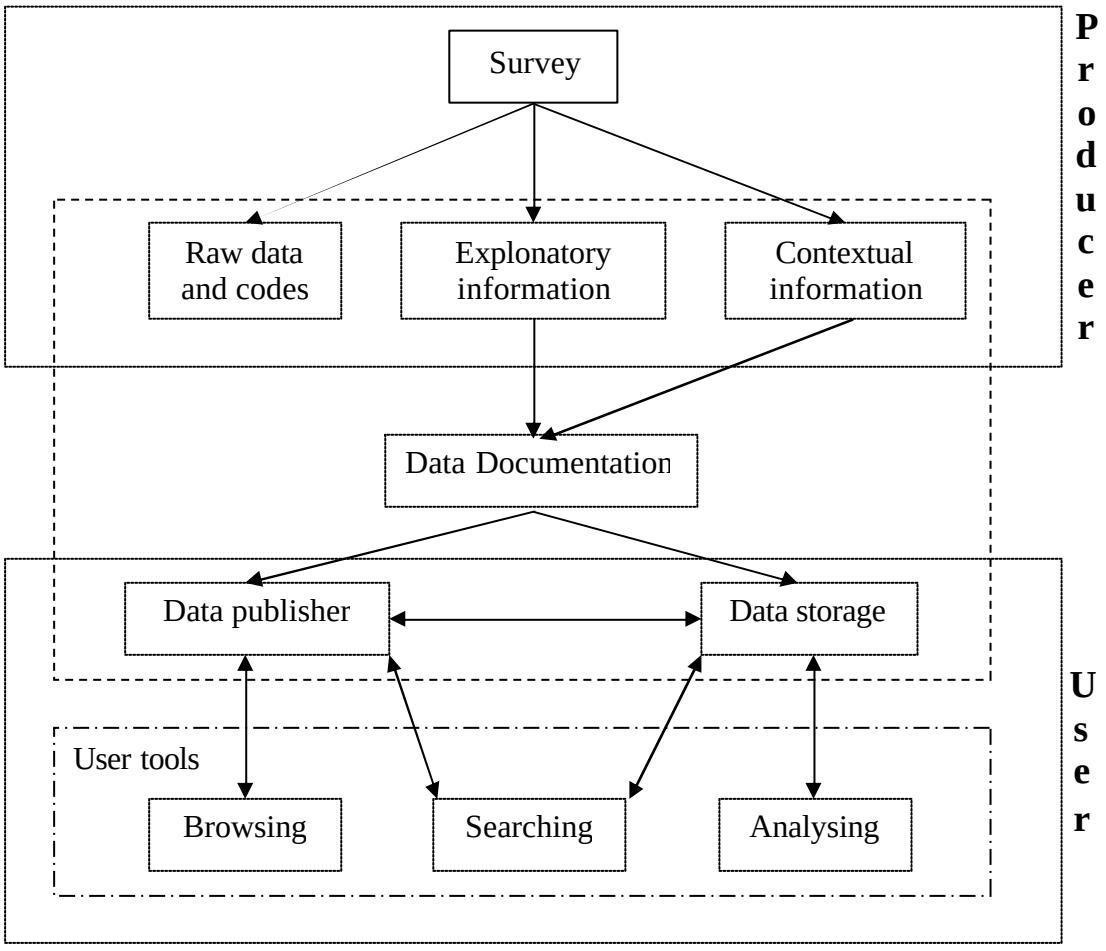


Figure 1 shows the archiving procedure and the data interaction with producer and end-user. Though the archiving principle is the same, archives differ mainly due to numerous data documentation standards or initiatives. Well known data documentation standards are XML-DTD based Data Documentation Initiative (DDI), Common warehouse Metamodel (CWM – http://www.cwmforum.org), and the cristal data object model (http://www.faster-data.org/Metadata/papers/Cristal.htm). Among these DDI is widely accepted standard and is discussed in the next chapter. Supporting the DDI standard, various archiving tools have been developed by different consortiums such as, NESSTAR, FASTER, etc.

2.1 NESSTAR

Networked Social Science Tools and Resources (NESSTAR) is a Europe based data archive's consortium. The NESSTAR consortium has been formed at the UK data archive in the University of Essex and the Norwegian Social Science Data Services in Bergen. NESSTAR has developed an integrated suite of software tools to archive the data in general and this software suite is DDI complaint. NESSTAR tools also facilitate to locate and use the archived data over web. In archiving the data, the NESSTAR tools hyperlinks the information from the following inter-linked sectors:

- Data: Raw data and associated information about
  - Survey: Data files, codes used
  - Indicators: Relevant indicators derived from preliminary analysis
  - Administrative: Survey administrative details
  - Geographical: Spatial information such as geo-codes, regions, etc.
- Tools: Various tools used in
  - Finding and sorting: Search engines
  - Browsing: Compatible tools to browse the archived data on data severs
  - Analysing: To perform either primary or secondary analysis
  - Publishing: Tools to publish the metadata on data server
- Text: Information about
  - Journal articles: All the journal articles
  - User guides: user guides of the various tools used

    - Methodology instructions: Instruction followed either in survey or in analysis
- People: Information about the persons involved
    - Email: Email of the contact persons (data producer, document producer, etc.)
    - Conferences: List of conferences organized
    - Experts: Information about experts in the field of interest
    - Discussion lists: Discussion lists such as forums, mail groups, etc.

Various archiving tools developed at NESSTAR are discussed in the forthcoming section. Some of the salient features in using NESSTAR software package 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. Data Documentation

Data documentation is the rich and full technical documentation of a dataset. Early documentation in 1960s was called a "codebook" because it documented the codes used in sample surveys. Few data archives developed the first electronic standard documentation in the form of OSIRIS codebook. For individual survey datasets OSIRIS provided documentation at the variable level, such as variable names, variable labels, and missing values. Recognizing that much more is being documented than codes, the word "codebook" has been superceded by the general term "metadata", meaning data about data. Metadata is defined as the data associated with objects which relieve their potential users of having to have full advance knowledge of their existence or characteristics. In essence, metadata answer 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 be viewed from two perspectives:

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

Human understandable metadata:

- 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.

- Facilitates an evaluation of the relevance, efficacy and quality of the data identified.

- Facilitates the retrieval of identified datasets

Earlier studies have partially discussed the travel related metadata in various contexts (Wigan, 2001; Harvey 2003).

Dublin core (http://dublincore.org) is one of the well known standard metadata schemas. Data documentation Initiative (DDI) is the most advanced, flexible, and user-friendly metadata standard that is developed based on the Dublin core metadata schema.

3.1 Data Documentation Initiative (DDI)

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 is 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. Since then, without disturbing its structure, several amendments have been made to the DDI specifications. After the formation of a new alliance (DDI alliance: http://www.ddialliance.org) in late 2002, the latest DDI specification (version 2.0) was released in July 2003. The noteworthy enhancement in the latest version of the DDI specification is an aggregate/tabular data extension that permits the mark-up of aggregate data. The structure of the DDI specification, which is driven by the XML DTD, is described in the subsequent section.

3.1.1 DDI structure

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. The DDI-tree contains five main sections:

1. Document Description: This section consists of bibliographic information describing the metadata document and the sources that have been used to create it.
2. Study Description: This section contains information about the data collection.
3. Data Files Description: This section describes each single file of the data collection.
4. Variable Description: This section describes each single variable in the dataset. The detailed description includes description of format variable and value labels, definitions, question texts, imputations, etc. In contrast to the previous section (Data files description) which provides physical description of the data, this section provides logical description.

5. Other study related material: This section can include references to reports and publications, other documentation (referenced by URI's) that are relevant to the user of the study.

The latest DDI specifications (version 2.0) are outlined in Appendix I. Benefits of the DDI approach can be explained by summarizing the DDI facilities:

- Marked-up codebooks using the DDI specification supports interoperability i.e. codebooks can be exchanged and transported seamlessly.
- DDI structure was designed to encourage the use of a comprehensive set of elements to describe social science datasets as completely and as thoroughly as possible. Thus, the documented information through DDI is rich in content.
- Single DDI document can be used for multiple purposes. A DDI document contains all of the information necessary to produce several different types of output, including, for example, a traditional social science codebook, a bibliographic record, or SAS/SPSS/Stata data definition statements. Thus, the document may be repurposed for different needs and applications. Changes made to the core document will be passed along to any output generated.
- On-line subsetting and analysis: Because the DDI markup extends down to the variable level and provides a standard uniform structure and content for variables, DDI documents are easily imported into on-line analysis systems, rendering datasets more readily usable for a wider audience.
- DDI provides a high precision in searching. Since each of the elements in a DDI-compliant document is tagged in a specific way, field-specific searches across documents and studies are enabled. For example, a library of DDI documents could be searched to identify datasets covering protest demonstrations during the 1960s in specific states or countries.

Need for an exclusive DDI to support the travel data is discussed in chapter 5.

Existing DDI does not fully support exclusive DDI for the travel data does not exists. And little progress has been made in the past in regards to the development of travel DDI.

4. Travel data archiving: experiences from “ETH Travel Data Archive”

4.1 Archiving tools

NESSTAR has come up with its first software suite comprises of a set of archiving tools. Following tools were used in archiving the survey datasets:

- Data Builder1.1.3
- XML Generator 1.2
- NESSTAR Publisher 1.1
- NESSTAR Light 1.1
- NESSTAR Explorer 1.0
- NESSTAR Server 1.1.2

Detailed description of the data archiving procedure was explained in the earlier reports (Chalasani, Schoenfelder, and Axhausen, 2002). Datasets from two surveys (Mobidrive and Microcensus travel 2000) were archived under this server. In March 2003, NESSTAR has come up with an improved software suite. Tools in this software suite are:

- NESSTAR Server v2.16
- NESSTAR Publisher v2.1
- NESSTAR Hierarchical Builder v2.1
- NESSTAR Cube Builder v2.1
- NESSTAR Light v2.0

Basic functions of all the above tools are briefly explained in the subsequent sections.

4.1.1 NESSTAR Server

Built as an extension to a normal web server, NESSTAR Server provide all the usual facilities for publishing web content and has the ability to publish statistical information that can be

searched, browsed, analysed, and downloaded by the end-users. With a refined access control policy, controlled access to all the datasets sorted in a specific catalogue can be managed. Additional feature introduced in the latest NESSTAR server (version2.16) is a simple online user management tool that enables the server administrator for a better user management with controlled access.

4.1.2 NESSTAR Publisher

Essential and frequently used distinct features from Data Builder, XML Generator, and previous NESSTAR Publisher were structurally edited in designing the latest NESSTAR Publisher. NESSTAR Publisher consists of data and metadata conversion and editing tools, enabling the archivists to prepare these materials for publishing on a web server. This software tool allows datasets enhancement by combining a wide range of catalogue and contextual information.

4.1.3 NESSTAR Cube Builder

This recently developed tool is an extension to the NESSTAR Publisher. NESSTAR Cube Builder is designed to prepare metadata for multi-dimensional data. A cube is a multidimensional table consists of a number of dimensions (variables) and at least one measure variable. Both the NESSTAR Publisher and NESSTAR Cube Builder are interlinked, i.e. while working with any dataset users can shift from one to the other. Users have to have a thorough knowledge on both the data and Cube Builder to prepare metadata on cubicles.

4.1.4 NESSTAR Hierarchical Builder

NESSTAR Hierarchical Builder allows to bridge different datasets obtained from a survey and datasets from a series of surveys. Hierarchical bridging enables secondary analysis among variables from different datasets. Identification of one or a set of basic key variables, whose combination can produce a unique observation of the dataset of interest, is the fundamental concept in dataset hierarchy. Chosen basic key variables act as pillars to bridge with other datasets. A real-time example on datasets hierarchy is explained in Appendix II.

4.1.5 NESSTAR Light

NESSTAR Light is a tool to interact with different virtual libraries around the world. NEESTAR Light allows users to search for, locate, browse, sub-set, and download a wide variety of statistical and related archived data published over web. NESSTAR Light consists of a web based search engine and an embedded statistical analysis tool.

Though the latest NESSTAR software suite has simplified the archivists job, it suffers with severe violation of the data documentation standards. However, a viable solution has been implemented to overcome these drawbacks. While explaining the data archiving procedure in the next section, this solution is well explained.

4.2 Archiving travel data: A viable solution

Taking the merits and limitations of the two sets of archiving tools, a viable solution has been framed to archive the travel survey data at this institute. This working experience based viable solution has been formulated because of the following reasons:

- Serious limitations in the latest archiving tools (NESSTAR Server 2.16 compatible) were found in metadata generation.
- It was understood that these limitations can be overcame by timely utilisation of the earlier archiving tools (NESSTAR Server 1.1.2 compatible)

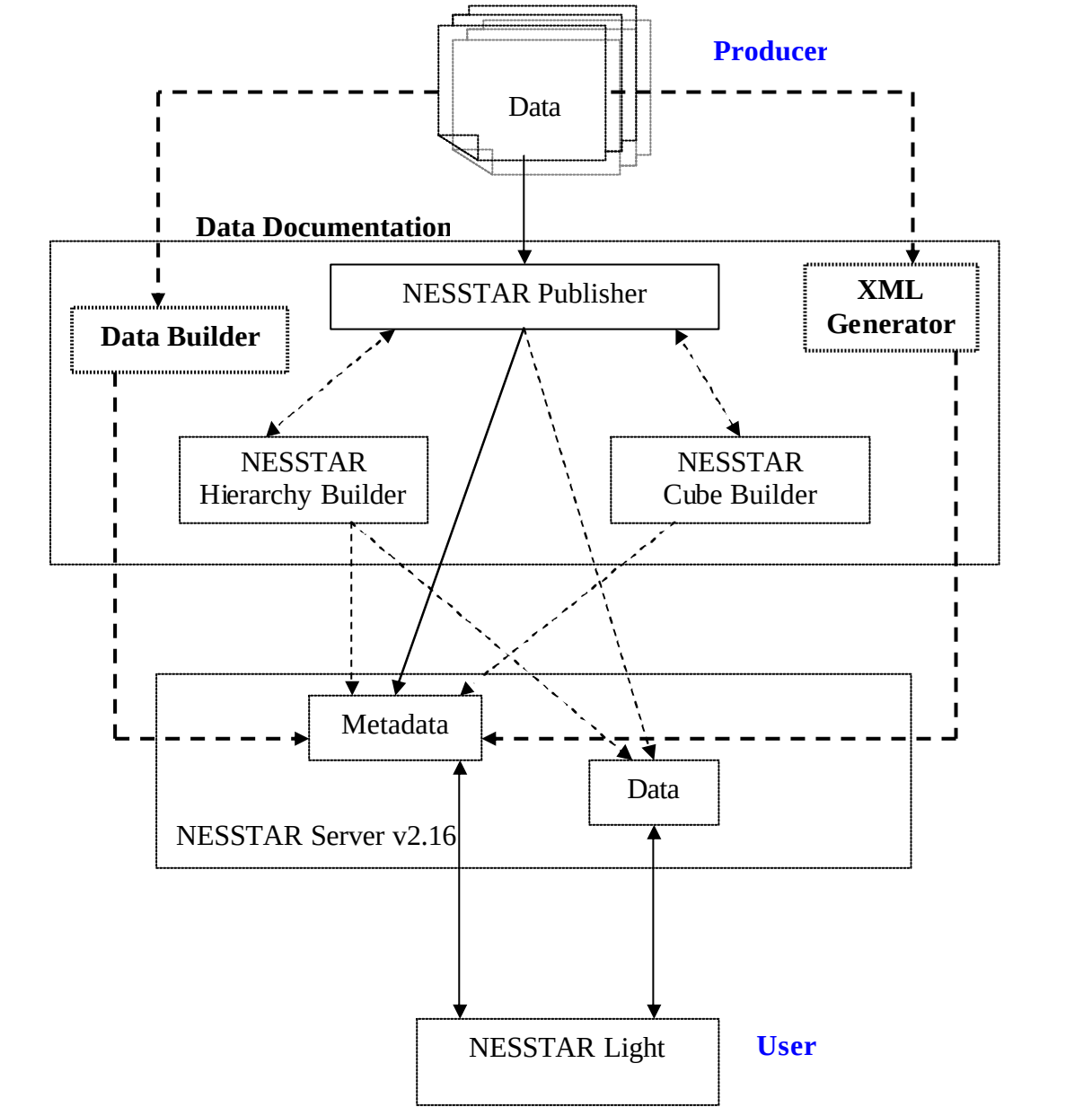
Figure 2 shows the schematic representation of the data archiving procedure. Sequential stages of the formulated travel data archiving procedure are:

- Producing metadata: Using NESSTAR Publisher, this stage produces the metadata for the entire codebook.
- Setting up hierarchy and cubes: NESSTAR Hierarchy builder is used to derive the hierarchy among the archived datasets. Pre-defined cubes can be built in the NESSTAR Cube Builder.
- Metadata enhancements: With the help of Data Builder and XML Generator, essential metadata that was ignored in NESSTAR Publisher could be derived.
- Publishing data and metadata: Archived data (with or without data) will be published either from NESSTAR Publisher or NESSTA Hierarchy Builder.
- Amending the metadata: Metadata extracted and added in the Data Builder and/or XML Generator is combined with the metadata from NESSTAR Publisher. This

elaborated metadata in XML format replaces the XML file in the server data directory

- Testing with NESSTAR Light: Archived and uploaded data undergoes few test trails for error checking and functionality.

Figure 2 Data flow and archiving procedure of the viable solution



____: Mandatory data flow _ _ _ _ :Optional data flow

Amendments in the Viable solution **Persons involved in data archiving and data usage**

: Archiving tool : Data documentation phase

4.2.1 The travel data archiving procedure

NESSTAR Publisher facilitates the dataset development and also allows datasets to be imported from the most common statistical formats:

- SPSS System (*.sav)
- SPSS Portable (*.por)
- SPSS Syntax (*.sps)
- Stata (*.dta)
- Statistica (*.sta)
- NSDstat (*.nsf), the old NSDstat format used by NSDstat Pro.
- dBase (*.dbf)
- DIF (*.dif)
- Fixed Format ASCII (*.dat)
- Delimited Text (*.txt)

NESSTAR Publisher extracts all the metadata incubated in the imported dataset. Among the five sections of the DDI codebook, four sections are visible in the NESSTAR Publisher. An automatically generates metadata for the files-by-files description section is hindered in the NESSTAR Publisher. Except the study title, the sections document description and study description are generally empty. Only the important and essential elements from these sections are made available for manual editing. This is a serious limitation in developing a full-fledged travel metadata, because travel data collected from travel survey with complex methodologies carry large amounts of associated information. Manual editing is the only way to overcome this limitation, but is practically discarded due to difficulties and frequent errors in editing the XML code for each element. Sensing the Data Builder's role in enhancing only required elements of the DDI codebook and content editing as handy, this outdated tool is drafted in travel data archiving. Similarly, the XML Generator allows 16 variable descriptives against 5 in the NESSTAR Publisher. Thereby inclusion of the outdated tools Data Builder and XML Generator certainly improves the travel data archiving efficiency. The concepts variable grouping and other study materials have taken enough attention of the NESSTAR consortium. Separate sub-windows for data view, variable groups, and other study variables are provided in the NESSTAR Publisher. All the variables whose characteristics fall in a short range are clubbed together under a common name, called variable group name. As the concepts variable grouping and hierarchical dataset are unchanged, it is recommended to refer

the previous reports for explanation (Chalasani, Schoenfelder, and Axhausen, 2002). After all the datasets are edited and metadata is developed, using NESSTAR Hierarchy Builder a hierarchical dataset that bridges all the included datasets can be developed. The NESSTAR Cube Builder facilitates in developing the cubes, if any. Due to lack of working experience, cubes development is purposely excluded in this report. Either metadata or both data and metadata can be published on NESSTAR Server 2.16 from one of the following:

- NESSTAR Publisher: In case of uploading individual datasets
- NESSTAR Hierarchical Builder: In case of uploading a hierarchical dataset
- NESSTAR Cube Builder: In case of uploading cubes from an individual dataset.

4.3 ETH Travel Data Archive

As stated earlier, the first travel data archive catalogue "ETHTDA" was established on the NESSTAR Server 1.1.2 in May 2002. It was then renamed as ETH Travel Data Archive and the supporting server was updated to version 2.16 (NESSTAR Server 2.16). Since its establishment, eleven datasets from different sources have been archived. Following survey datasets from different travel surveys were archived at "ETH Travel Data Archive":

- DATELINE 2002 (log-distance travel survey)
- Mobidrive (A six-week travel dairy)
- Microcensus travel 2000 (National household travel survey on daily mobility)
- Travel module of Household Income and Consumption (National household travel survey)
- Forecast based on different data types: A before and after study (Data types are - stated preference mode choice data, stated preference route choice data, and revealed preference data)
- Railway passenger traffic 1999/2001 - average working day traffic (traffic counts)
- Road passenger traffic – average cars per working day (traffic counts)
- 12 weeks of leisure travel (Activity based leisure travel)
- Travel market Switzerland 2001 (National household leisure travel survey)

Among the above datasets, Mobidrive and Microcensus travel 2000 datasets were archived using first set of archiving tools and then uploaded on the latest server version with little modifications.

5. Travel DDI

Current ICPSR enhanced DDI itself is not comprehensive enough to ensure a complete description of the standard travel survey process (see for example checklist in Richrdson, Ampt and Meyburg, 1995). The supporting DDI tools or editors to develop a comprehensive metadata are lacking in complete coverage of free form and structured DDI elements. Travel data archiving has become less attractive due to the above drawbacks i.e. inefficient metadata standard along with the limited archiving tools. Keeping the preservation aside, presentation of the travel data can be separated into two stages:

- Travel metadata: Necessary modifications to the existing DDI specification to produce a complete travel metadata.
- Publishing the travel metadata: Publishing styles based on the travel data type and context for a better understanding of the published content.

This report describes the principle of travel metadata development. Establishing an exclusive DDI standard (Travel DDI) is a unique solution for the travel metadata development. Travel Data Documentation Initiative (Travel DDI) is a modified version of the existing DDI developed by DDI alliance at ICPSR. Inheriting the structure of the present DDI, Travel DDI contains added metadata standard specifications that accommodate additional information about the travel data. Previous studies have outlined few documentation requirements to meet the special needs of transport surveys (Axhausen, 2003; Wigan 2002). However, the following requirements are chosen for an in-depth analysis on feasible additions to the Travel DDI in metadata development (Chalasani and Axhausen, 2004, forthcoming):

- Survey objectives and purpose: Brief description of the purpose in conducting the survey and its planned objectives
- Sample details: A complete description of the sample design, including sample size, sampling frame, information on eligibility criteria and screening procedures, methods by which respondents were selected and details of how the sample was drawn, including any exclusions or ineligible units
- Response rates: Response rates in sampling and for each data collection efforts including the formula and definition of the elements used, final response rate
- Data collection methods and survey protocols: Details of the data collection method including data collection mode and survey procedures. Survey protocol with detailed time frame, and detailed reporting period and survey period of each data collection effort.

- Description of the post-data collection process: Detailed information about data editing, data adjustment or imputing procedures, and weighting process.
- Interviewer characteristics and precision of estimates: Number and background of the field staff, especially if special staff employed for the survey. Sampling error with references to other possible sources of error to inform user of accuracy or precision.

Travel data is highly correlated to data from few scholarly sectors such as, socio-economic, geographical, weather, etc. Thus, travel data and its collection is complex when compared to social sciences data. Understanding the socio-economic background of the respondents is highly essential to analyse the travel patterns and collection of socio-economic characteristics has become a common practice in all travel surveys. Travel i.e. change of location over time, can be geographically explained and travel data always contain the geographical information in one or other form. Weather is another important factor that influences travel. Thus, weather data is also equally important in travel data. Presentation of travel data coupled with these sectors needs special presentation style. Another innovative thought that needs more attention on publishing styles is, questionnaires based publishing the data, i.e. publish the data similar to the physical appearance of the questionnaire.

The travel data archiving potential can be substantially in creased by effectively presenting the travel data through an efficient travel metadata.

6. Summary and conclusions

The travel data preservation facilitates for secondary use and analysis, and improve its survey potential. Effective documentation of associated information of the travel data dictates the extent of the secondary analysis and use. Sharing of the travel data and its associated information broadens the knowledge and gives more exposure to researchers or data producers. Data archiving concept is a shell of these issues. As travel data carry enormous amounts of associated and hindered information, travel data archiving might reduce the loss of information and adds value to the precious travel data.

A worldwide data documentation standard can be effective in documenting the associated information around the diversified travel data. Well known and established metadata standard, data documentation standard (DDI), was considered to document the travel data. DDI properties were studied and weighted for its compatibility. Analysis has revealed that the selected DDI is not enough to prepare a complete travel metadata. Considering its proven efficiency, existing DDI structure was considered as a significant worldwide travel metadata standard, possible specifications that are lacking were identified. An exclusive travel metadata standard "Travel DDI" is recommended with the addition of the missing specifications in the existing general DDI. Important issues that need special attention in Travel DDI are: Survey protocol, sample details, response rates, and data collection methods. An in-depth analysis on these issues is recommended for future research.

Supporting tools to archive the travel data under existing DDI plays vital role in minimising archivists difficulties. With the working experience in archiving travel data and uploading on the first travel data archive "ETH Travel Data Archive", advantages and drawbacks of these supporting tools were analysed. A viable solution has been framed to maximise the travel data archiving efficiency. Though the available archiving tools are severely handicapped in some sections, accepting the fact that the travel data archiving is emerging, no recommendations are suggested in this report.

7. References

Axhausen K.W. (2000) Presenting and Preserving Travel Data, in Jones, P.M. and P.R. Stropher (eds.) *Transport Surveys: Raising the Standards*, Transportation research Circular, E-C008, II-F-1 – II-F-19.

Axhausen, K.W. and M.R. Wigan (2003) Public Use of Travel Surveys: The Metadata Perspective, in Stropher, P.R. and P.M.Jones, *Transport Survey Quality and Innovation*, Elsevier, England.

Chalasani, V.S., S. Schönfelder and K.W. Axhausen.(2002) Archiving travel data: The Mobidrive example, *Arbeitsbericht Verkehrs- und Raumplanung*, **129**, Institut für Verkehrsplanung und Transportsysteme (IVT), ETH Zürich, Zürich.

Richardson, A.J., E.S. Ampt and A.H. Meyburg (1995) *Survey Methods for Transport Planning*; Eucalyptus Press, Melbourne.

Wigan, M.R. (2001) Enabling and managing greater access to transport data through metadata, paper presented at the *European Transport Conference*, Cambridge, September.

Harvey, A.S. (2003) Time-Use Metadata, *Transportation Research Record*, 1804,

8. Universal resource locators

ETH Travel Data Archive (ETHTDA)(2003), http://129.132.96.89/index.jsp

Network Social Sciences Tools and Resources (NESSTAR)(2003), Nesstar Light, http://www.nesstar.com/products/light.

Network Social Sciences Tools and Resources (NESSTAR)(2003), Nesstar Publisher v2.1, http://www.nesstar.com/products/publisher.

Network Social Sciences Tools and Resources (NESSTAR)(2003), Nesstar Server v2.16, http://www.nesstar.com/products/server.

Norwegian Social Sciences Data Services (1999), Data Builder 1.1.3, http://www.nsd.uib.no/nesstarweb/DataBuilder/

Norwegian Social Sciences Data Services (2001), XML Generator v1.6.3, http://www.nsd.uib.no/nesstarweb/xmlgen/XMLGenerator.cfm

Appendix A: Outline of the DDI codebook

The DTD view of the DDI codebook is:

```
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)
```

Simple and structured outline for the above part of the codebook is the following:

1.0 Document description

- 1.1 Citation
  - 1.1.1 Title statement
    - 1.1.1.1 Title
    - 1.1.1.2 Sub-title
    - 1.1.1.3 Alternative title
    - 1.1.1.4 Parallel title
    - 1.1.1.5 ID number