Preferred citation style for this presentation

The public use of travel surveys: The metadata perspective

KW Axhausen

IVT
ETH
Zürich

August 2001
A look back

The action list arising from Eibsee and elsewhere was:

- Systematic archiving of transport data sets
- Stipulation of archiving as part of contracts and grants
- Development of uniform data set descriptors
- Implementation of web-based on-line tabulation software
- Development of specialised transport data archives
What has been achieved?

- Systematic archiving: No progress, but see www.itdb.bts.gov
- Required archiving: No progress
- Uniform data set descriptions: Substantial progress outside transport
- On-line tabulation: Limited progress
- Specialised data archives: No progress
On-line tabulation

No wave of implementations in spite of the web growing popularity:

- Substantial set-up costs
- No standardisation
- No customer pressure

Current examples:

- www-cta.ornl.gov/npts
- sturm.math.fundp.ac.be/~test and related sites for the UK NTS and the Belgian NTS
Specialised on-line tabulation

Capabilities:

• Tabulation (one, two and three-way)
• Some graphing
• Proper weighting and suppression of cells with unreliable estimates

Missing:

• Aggregation functions
• Mapping
Example: Test - website

Technologies for European Surveys of Travel behaviour

Welcome to the TEST Web Site

Click on the "Help" item in the left window for additional explanations.
### Crosstables

This page allows you to cross some variables. [(Help)]

1. Select a filter: [Filter]  [Delete]

   There is no selected filter

2. Please, choose one or more countries:
   - [ ] France
   - [ ] United Kingdom

3. The current table is:

<table>
<thead>
<tr>
<th>Sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
</tr>
</tbody>
</table>

   To choose another row variable, click on the first radiobutton beside the variable. To choose another column variable, click on the second radiobutton beside the variable.

A (*) beside the variable means that the counting is done on pre-defined classes. A (cf) beside the variable means that the counting is done on user's created classes.
Metadata: data about data

Why:

• Better searching
• Higher quality documentation
• Base for data analysis tools

What:

• Standardised set of descriptors
• Joint vocabulary for the descriptors
Metadata: What?

In the context of survey research:
- Description of the archived files
- Description of the survey and its conduct
- Description of the responsibilities for the survey and its documentation
- Supplementary data

Examples:
- Any standardised data base or data exchange protocol
- “Dublin core” (bibliography and responsibilities)
- ddi - DTD
Metadata: How?

Development:
- Suitable standard development process
- Ad-hoc developments
- Competitive market-driven development

Technologies:
- Any suitable database description language
- Edifact
- SGML
- XML
XML: extended mark-up language

What is it?

• Subset of SGML
• Language to describe contents (and rendering)

What does it offer?

• Easy definition of languages to describe data
• Imposes grammar checks on files through “document type definitions”
ddi and Nesstar: Progress through standardisation

ddi: Data documentation initiative (University of Michigan)

- XML - DTD to describe data sets for archiving

Nesstar: ddi - DTD based software to publish, search and interrogate data on the web (EU - funded, British and Norwegian data archives)

- Client software to search and tabulation
- Server software for publication
Structured description:

- docDscr (Document)
- stdyDscr (Study)
- fileDscr (File(s))
- dataDscr (Variables)
- othMat (Other materials)
ddi - DTD: Example

-- 4.0 dataDscr* (ATT == ID, xml:lang, source)

-- 4.2 var* (ATT == ID, xml:lang, source, name, ...)

  -- 4.2.1 location* (ATT == ID, xml:lang, source, StartPos, EndPos, width, RecSegNo, fileid)
  -- 4.2.2 labl* (ATT == ID, xml:lang, source, level, vendor)
  -- 4.2.3 imputation? (ATT == ID, xml:lang, source)
  -- 4.2.6 respUnit? (ATT == ID, xml:lang, source)
  -- 4.2.7 anlysUnit? (ATT == ID, xml:lang, source)
  -- 4.2.9 valrng* (ATT == ID, xml:lang, source)
  -- 4.2.10 invalrng* (ATT == ID, xml:lang, source)
  -- 4.2.13 TotlResp? (ATT == ID, xml:lang, source)
  -- 4.2.14 sumStat* (ATT == ID, xml:lang, source, wgtd, weight, type)
  -- 4.2.15 txt* (ATT == ID, xml:lang, source, level)
  -- 4.2.17 catgryGrp* (ATT == ID, xml:lang, source, missing, missType, catgry, catGrp)
  -- 4.2.18 catgry* (ATT == ID, xml:lang, source, missing, missType, country, sdatrefs)
  -- 4.2.23 varFormat? (ATT == ID, xml:lang, source, type, formatname, schema, category, URI)
<fileName ID='Household'>Household file</fileName>

<var files='Household' name='hhinc' qstn='A-54' >
  <labl>Household Income</labl>
  <location StartPos='55' EndPos='57' width='3'></location>
  <imputation>Hotdesk imputation using size, hours worked, car owned</imputation>
  <respUnit>Head of household</respUnit>
  <anlysUnit>Household</anlysUnit>
  <qstn>What is your total household income [kUS$] after taxes and social security</qstn>
  <valrng>
    <range Units='INT' maxExclusive='250' min='5' max='240'></range>
    <key>250 250k and more</key>
  </valrng>
  <invalrng>
    <range Units='INT' minExclusive='0' min='998' max='999'></range>
    <key>0 Refused 998 Dont know 999 Not applicable</key>
  </invalrng>
  <TotlResp>450 valid responses</TotlResp>
  <sumStat type='Min'>5</sumStat>
  <sumStat type='Max'>220</sumStat>
  <sumStat type='Min'>65</sumStat>
</var>
Nesstar: Browsing

**Transport of Goods via Harbours and Railway in Funen, 1865-1920**

**Title**

Transport of Goods via Harbours and Railway in Funen, 1865-1920

**Abstract**

Transportation is an important prerequisite for the production of a society, and it contributes to economic growth by making a still larger division of labour possible. However, transportation of goods has only been dealt with to a limited degree in previous research literature. The purpose of this survey is, in a well-defined limited area, to remedy this situation. This will be in the form of a regional quantitative analysis of the transportation of goods via harbours and railway in Funen in the period 1865-1920. An attempt will be made at throwing light on two aspects in particular: 1) The relative importance of these two means of transportation for the turnover as such as well as for selected items (of goods); and 2) How the trade of goods influenced the city system in Funen. Who were the winners and the losers in the battle of attracting as much trade from the surrounding area as possible? The period 1865-1920 was chosen as an analysis of the harbour and railway trade with only a limited degree of uncertainty will be the same as Funen's total wholesale trade with the surrounding world. The other means of conveyance at that time could not cope with wholesale trade. The capacity andtractive power of the horse-drawn carriage was limited, and

Study title: Transport of Goods via Harbours and Railway in Funen, 1865-1920
Nesstar: Tabulation

Study title: Consumer Dissatisfaction and Complaints, 1978
Nesstar: Graphics
Summary and outlook

What has happened?

- Too little within transport
- ddi - DTD is a substantial start
- Nesstar shows the possibilities of standard-based software

What should we do?

- Organisation of a standardisation process
- Develop our domain - specific extensions
- Using the ddi - DTD to get going