

Preferred citation style for this presentation

Axhausen, K.W. (2001) The public use of travel surveys: The metadata perspective, resource paper presented at the 2nd International Conference on Survey Quality and Innovation, Krüger National Park, August 2001.

The public use of travel surveys: The metadata perspective

KW Axhausen

IVT

ETH

Zürich

August 2001



Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

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



 **TEST**

User's page

Introduction

Survey

Registration

Frequency

Statistics

Crosstables

Creation of classes

Help

Feedback


[Isabelle Reginster](#)
[FUNDP - Math - GRT](#)
Last version :
27 August 1998

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.

Example: Test - website

 TEST

[User's page](#)

[Introduction](#)

[Survey](#)

[Registration](#)

[Frequency](#)

[Statistics](#)

[✓ Crosstables](#)

[Creation of classes](#)

[Help](#)

[Feedback](#)

[Isabelle Reginster](#)
[FUNDP - Math - GRT](#)
Last version :
27 August 1998

Crosstables

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

1. Select a filter :

There is no selected filter

2. Please, choose one or more countries :

France
 United Kingdom

3. The current table is :

↓	→	Sex
Sex		

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 (cl) 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

ddi - DTD: Structure

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, wgted, 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)
```

ddi - DTD: Application example

```
<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'>
      <key>250 250k and more</key>
    </valrng>
    <invalrng>
      <range Units='INT' minExclusive='0' min='998' max='999'>
        <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

The screenshot displays the Nesstar software interface for a dataset titled "Transport of Goods via Harbours and Railway in Funen, 1865-1920". The interface is divided into several sections:

- Dataset contents:** A tree view on the left shows the dataset structure. The "stdyDscr" folder is selected. Under "dataDscr", 16 variables are listed, including "1 STUDIENUMMER I DDA", "2 SEKVENSNUMMER", "3 REGNSKABSÅR", "4 REGNSKABSÅR TYPE", "5 STATION ELLER HAVN", "6 UNDERSTED", "7 GODSOMSÆTNINGS ...", "8 MÆNGDE I TONS, HELE T", "9 MÆNGDE I TONS, DECIMA", "10 KILDEMÆNGDE, HELE T", "11 KILDEMÆNGDE, DECIMA", "12 KILDEMÆNGDE BENÆVN", "13 LOKALT GODS, HELE TAL", "14 LOKALT GODS, DECIMAL", "15 GENM. GODS, HELE TAL", and "16 GENM. GODS, DECIMAL".
- Selected variables:** A section with "Add", "Remove", and "Remove All" buttons and an empty list box.
- Analysis type:** A dropdown menu set to "Tabulate".
- Run analysis:** A button at the bottom of the left panel.
- Documentation:** A tabbed area on the right showing the "stdyDscr" documentation. It includes a "Title" section with the text "Transport of Goods via Harbours and Railway in Funen, 1865-1920" and an "Abstract" section with a detailed paragraph about the study's purpose and scope.

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

Nesstar: Tabulation

Consumer Dissatisfaction and Complaints, 1978

Documentation Table Graphics

Table options

v215: V215: ALDER
v216: V216: CIVILSTAND

	Ugift	Tidligere	Gift eller	Ønsker	Uoplyst
15-19 år	38.1	0.0	0.4	16.7	0.0
20-24 år	30.3	2.7	6.1	0.0	0.0
25-29 år	13.5	2.7	16.9	16.7	0.0
30-39 år	7.1	23.0	30.5	33.3	0.0
40-49 år	1.9	14.9	18.1	0.0	0.0
50-64 år	5.2	21.6	19.9	16.7	10.0
65 eller derover	3.2	33.8	7.7	0.0	20.0
Ønsker ikke at	0.0	0.0	0.4	16.7	0.0
Uoplyst	0.6	1.4	0.1	0.0	70.0
Total	100.0	100.0	100.0	100.0	100.0
N=	155	74	758	6	10

Study title : Consumer Dissatisfaction and Complaints, 1978

Dataset contents:

- 206 EN LOKAL FORBRUGERGRU...
- 207 REJSEBUREAUANKENÆVNE
- 208 STATENS HUSHOLDNINGSR
- 209 FORBRUGERKLAGENÆVNE
- 210 FORBRUGERRÅDETS KLAGE
- 211 FORBRUGEROMBUDSMAN...
- 212 EN LOKAL FORBRUGERGRU
- 213 REJSEBUREAUANKENÆVNE
- 214 KØN
- 215 ALDER
- 216 CIVILSTAND
- 217 ERHVERV
- 218 SKOLEUDDANNELSE
- 219 ANTAL PERS. I HUSSTAND

Selected variables:

ANTAL PERS. I HUSSTAND

215 ALDER

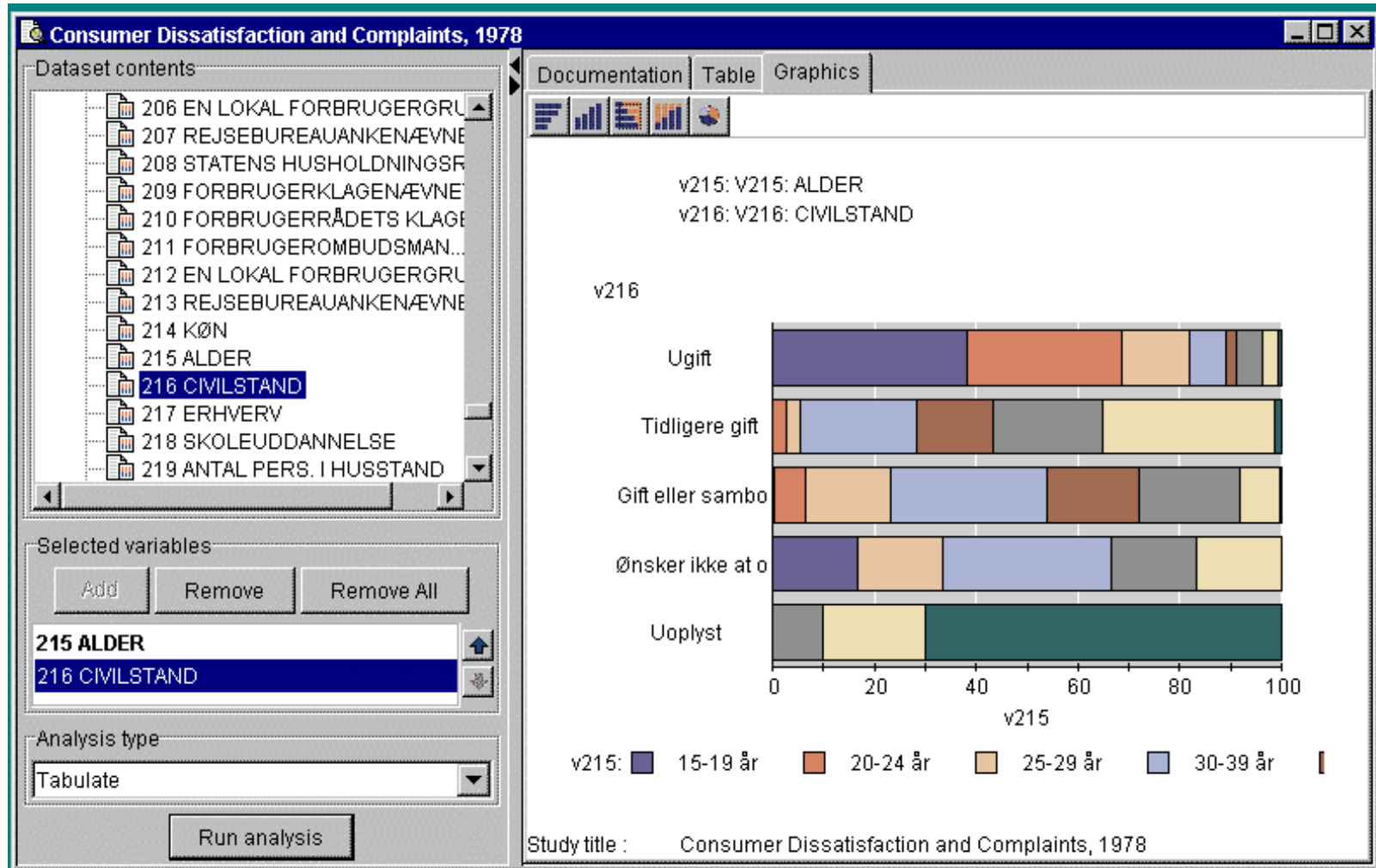
216 CIVILSTAND

Analysis type:

Tabulate

Run analysis

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