EU-SILC in Austria

2003: cross-sectional survey
since 2004: integrated cross-sectional and longitudinal survey with rotational design

minimum effective sample size
4.500 households per year
3.250 households between two waves

CAPI survey
fieldwork outsourced
A complex survey

Household context
Personal Register
Person1
Person2
Person3 (Child)
Household quest.
Housing, etc.

Personal quest. 2005
Person1
Income, Edu., Job ...
Person2
Income, Edu., Job ...

Consistency is required!

Longitudinal component
Information from other waves

Personal quest. 2004
Person1
Income, Edu., Job ...
Person2
Income, Edu., Job ...

Consistency also in the longitudinal component is required!

Changes in the survey
in the social transfers system
to the questionnaire

Changes to variables must be met!

Data management...

... is the whole process from the survey to final data and indicators

during the field work
checks during the CAPI-interview
early data checks

after the field work
data editing, imputation, weighting
computation of target variables and indicators
validation
Structure

- Checking raw data and feeding back inconsistencies to field work
- Editing income-variables
- Editing non-income variables
- Imputations
- Weighting and non-response analysis
- Calculating Eurostat target variables
- Calculating Laeken indicators and other key indicators
- Creating tables for publication
- Creating a data set for external data users

Basic principles

**Standardization**

of all relevant tasks (checking, editing, imputation, analysis)

generalization vs. single case solutions

**Transparency**

program code and documentation

files saved at various stages

flags

**Flexibility**

amendments

changes
Example of how to generalize

Problem: No receipt of income from employment/self-employment for people with months worked in previous year >0:

- Check during field work
- Decisions on the basis of case listing

Definition of scenarios:
Verification of plausibility for different groups
e.g. apprentices, unpaid family workers, voluntary workers etc.

Formulating general rules and programming different solutions

```
compute anz_monate=0.
compute anz_monate=sum(alm, pensm, p075012, p080002).
```

Number of cases read: 12   Number of cases listed: 12

```
***Keine Einkommensbezüge bei aktiven Personen (ohne mithelfende), die auch jetzt aktiv sind.
*sollten ein Erwerbseinkommen haben, außer 12 Monate eine Pension oder eine Arbeitslosenleistung, Krankengeld, Wochengeld oder Mutterschaftsgeld erhalten.
```

```
miss val p075012 p080002 (lo thru 0).
```

```
temp.
```

```
step 1
```

```
step 2
```

```
step 3
```

```
step 4
```

```
*Personen, die im Bezugsjahr 20 und jünger waren, sich als erwerbstätig einstufen, werden auf
ja beim Unselbständigeneinkommen gesetzt.
*Lehrlingsproblem.
```

Editing procedures for income data

- Income component received?

  - n.a. = no answer
  - N/G = net/gross
  - G/N = gross/net

  step 1

  step 2

  step 3

  step 4
Technical implementation

Programming
in SPSS code
main-control-syntax with sub-tasks in single syntax files
common programming rules
automation and standardization with macros

Variable names
variable name convention helps to standardize

Flags
meta-information for income variables useful for standardization and documentation

Variable name convention

<table>
<thead>
<tr>
<th>Meaning of the digits</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>digit 1: identification of the data file (D, R, H, K, oder P)</td>
<td>P</td>
<td>0</td>
<td>5</td>
<td>7</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>F</td>
</tr>
<tr>
<td>digits 2-4: number of the question corresponding to the questionnaire</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>digits 5-6: number of the sub-item</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>digit 7: information on variable type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0=no income variable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-5=income variable:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1=component received yes/no</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2=frequency of receit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3=gross amount</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4=net amount</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5=category</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>digit 8: to be filled optionally, e.g. with &quot;F&quot; for flag variable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Example

<table>
<thead>
<tr>
<th>digit variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>P........</td>
<td>P</td>
<td>0</td>
<td>5</td>
<td>7</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>F</td>
</tr>
<tr>
<td>P057....</td>
<td>It is question number 57 from the personal questionnaire.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P05702</td>
<td>It is the second item of question 57.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P057024</td>
<td>It is the question for the net amount.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P057024F</td>
<td>It is the flag variable of variable p057024.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Flags for income variables

-2 not applicable
-1 no answer and not (yet) imputed
1 value according to survey
2 value from category imputation
3 value from net-gross or gross-net conversion
4 value logically deduced
5 value statistically imputed with longitudinal method
6 value statistically imputed with cross-sectional method
7 value from survey was corrected

- value computed from a monthly income (this code applies only to variables of yearly income)

Example: Income distribution by flag for survivors’ pensions (EU-SILC 2005)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Total distr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid n</td>
<td>519</td>
<td>10</td>
<td>12</td>
<td>9</td>
<td>15</td>
<td>566</td>
<td></td>
</tr>
<tr>
<td>Valid n in %</td>
<td>92</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Minimum</td>
<td>50</td>
<td>17</td>
<td>99</td>
<td>301</td>
<td>126</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Maximum</td>
<td>4,000</td>
<td>960</td>
<td>1,221</td>
<td>1,933</td>
<td>1,056</td>
<td>4,000</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>600</td>
<td>440</td>
<td>549</td>
<td>906</td>
<td>568</td>
<td>600</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>623</td>
<td>527</td>
<td>591</td>
<td>1,141</td>
<td>617</td>
<td>627</td>
<td></td>
</tr>
<tr>
<td>Stddev</td>
<td>352</td>
<td>340</td>
<td>331</td>
<td>546</td>
<td>285</td>
<td>359</td>
<td></td>
</tr>
</tbody>
</table>

Finally…

…calculation of target variables, Laeken indicators and Gender Pay Gap

Analysis and quality assessment

- Comparison to external data and previous SILC waves
- Eurostat data checks
- Tables

Reports and publications

- Quality reporting
- Publication of results, news items, articles
- Data set for external users
## Challenges

- Longitudinal checks and analysis, longitudinal target variables!
- Improve documentation
- Integration of external comparisons and quality control
- Editing of analysis variables and providing them for external users
- Integration of questionnaire