A new GSDEM:
Multisource data for multiple statistic

Li-Chun Zhang\textsuperscript{1,2} and Susie Jentoft \textsuperscript{2}

\textsuperscript{1}University of Southampton (L.Zhang@soton.ac.uk)
\textsuperscript{2}Statistics Norway (lcz@ssb.no; susie.jentoft@ssb.no)
Scenario (e) of GSDEM (UNECE, 2015)

**Generic Statistical Data Editing Model (GSDEM) V. 1.0**

**Statistical Data Editing (SDE) flow models**

... Scenario (e): SDE of multisource statistics.

... *multiple* inputs $\Rightarrow$ *single* output

Two examples in Norway

- register-based Employment

- register-based Household income

NB. with common inputs, e.g. Wage Payment Data
Scenario (e) of GSDEM (UNECE, 2015)
Scenario (e): *single* output
Scenario (e) $\rightarrow$ **Parallel editing** of *multiple* statistics
Coordinated editing of *multiple* statistics
Coordinated editing of *multiple* statistics

**Steady states** (Rennsen & van Delden, 2009)

... In Norway, referred to as the *4 major states*:

1. Raw/input data → 2. Micro data
   → 3. Statistical data
   → 4. Public/disseminated data

Parallel editing leads to incoherent statistical data, e.g.

- Same input Wage Payment data: by transactions
- Employment statistics: re-arranged by persons
- Wage statistics: re-arranged by jobs
Coordinated editing of multiple statistics

Introducing of **BaSE** (in Norwegian)

- **BaSE** = Base Statistical Units
  
e.g. work relation for Employment and Wage statistics

- **common process up to a point**
  
  ... intermediate statistical data

- **coordinated process necessary afterwards**
  
  ... towards domain-specific statistical datasets

  ... and **macro accounts** as expressions of coherence
An example: A-ordning in Norway (since 2015)

Single entry for previously 5 different admin datasets

... important for alignment of reference time points

Monthly data on employment, wage and other payments

$U_{11}$: included in both Employment and Wage statistics

$U_{10}$: only included in Employment statistics

$U_{01}$: only included in Wage statistics

$U_{00}$: not included in Employment or Wage statistics

<table>
<thead>
<tr>
<th></th>
<th>September 2016</th>
<th>September 2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common to both ($U_{11}$)</td>
<td>2 678 586</td>
<td>2 653 513</td>
</tr>
<tr>
<td>Only in Employment statistics ($U_{10}$)</td>
<td>235 468</td>
<td>132 229</td>
</tr>
<tr>
<td>Only in Wage statistics ($U_{01}$)</td>
<td>21 234</td>
<td>17 818</td>
</tr>
<tr>
<td>Excluded from both ($U_{00}$)</td>
<td>2 754 146</td>
<td>2 839 966</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>5 689 434</strong></td>
<td><strong>5 643 526</strong></td>
</tr>
</tbody>
</table>

*Table 1. Partition of base statistical units for Employment and Wage Statistics.*
Towards a new **Labour Force Account (LFA)**

**Part I.** For BaSE sub-population (U11), use accounts to connect two *observable* totals: Employed (Y) and Wage payment (W)

- Account (1): \( Y \rightarrow T \) where \( T = \text{total FTEs} \)
- Account (2): \( T \rightarrow W \) via Wage per FTE

**Part II.** For Employment-only BaSE sub-population (U10), one or several *normative* accounts

- e.g. would-have-been Wage cost for persons on leave, in contrast to actual benefit payments they receive
- gross flows between observable and normative parts of LFA, e.g. in terms of person, Wage cost, FTE, etc.