Wiesbaden 2014 – Special Session
Dissemination of EGR data on multinational enterprise groups and data quality management

D. Francoz
Eurostat
The context of the study

- Global enterprise groups are major actors of the economy
  - Which groups operating in Europe had the most important growth over the last 4 years?
  - Where did the European multinational groups realise their growth?
  - Which are the most dynamic groups in Europe?
- EGR is a rich source of economic information
  - Could it be helpful to analyse economic growth and globalisation?
The context of the study

- Prior aim of EGR is to be the backbone for statistics on globalisation
  - provide NSIs a good basis to define sampling frames
  - needs to include information characterising the units (legal units, groups, enterprises)
  - This information could be used for economic analysis under certain quality condition
- Proceed in 3 steps
  - Define economic indicators to be followed in the EGR
  - Ensure that the quality is high enough
  - Calculate and disseminate the economic indicators
The context of the study

• Analysis done on EGR V1.0 for the reference year 2011

• Check in EGR V2.0 if the same quality issues
Variables available

- **GEG level**
  - Global employment of the group
  - Global turnover of the group
  - Global assets
  - NACE code of the group
  - Country of the global decision centre (GDC)

- **Legal unit level**
  - Number of persons employed
  - NACE code
  - Country code

Information that could be used to calculate new variables at the GEG level
Calculated Variables

• NACE code of the first, second and third ... activities of the group
• Employment of the group per country (EU employment)
• Size class of the group (in terms of employment or turnover)
• Indicator of diversification of activity
• Indicator of geographical influence in Europe
Dissemination and quality

• Related matters
  • No dissemination without a certain degree of quality
    ✓ Interpretation of data not to be distorted by quality issues
  • Dissemination identifies quality issues
    ✓ Using files reveals errors

• 2 types of quality issues in the EGR
  • “primary” degree issues
    ✓ intrinsic errors on some variables (missing information, wrong codification, not updated data...)
  • “secondary” degree issues
    ✓ use of different sources which are inconsistent (duplication of legal units, inconsistent treatments across sources...)

• In any case, quality depends of the source of information
EGR sources in EGR V1.0

- NSIs
- CDPs
- LEGAL UNITS DATA
- GROUPS DATA
- EUROSTAT
## Quality Issues Encountered

<table>
<thead>
<tr>
<th>Variables</th>
<th>Issues</th>
<th>Ways of detection</th>
<th>Impact on Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group's perimeter</strong></td>
<td>Missing legal units</td>
<td>no way</td>
<td>geographic impact underestimated</td>
</tr>
<tr>
<td></td>
<td>Duplicated legal units/ status of the legal unit not updated</td>
<td>sum of LeU employment &gt; global GEG employment</td>
<td>EU size of the GEG overestimated</td>
</tr>
<tr>
<td><strong>Number of persons employed</strong></td>
<td>Double count due to duplication or CDP estimation</td>
<td>sum of LeU employment &gt; global GEG employment</td>
<td>EU size of the GEG overestimated</td>
</tr>
<tr>
<td></td>
<td>missing number of employees</td>
<td>Variable is empty</td>
<td>EU size of the GEG underestimated</td>
</tr>
<tr>
<td><strong>NACE code</strong></td>
<td>Misclassification of units</td>
<td>Difficult to detect</td>
<td>miscalculation of the GEG NACE</td>
</tr>
<tr>
<td>Variables</td>
<td>Issues</td>
<td>Ways of detection</td>
<td>Impact on statistics</td>
</tr>
<tr>
<td>----------------------------</td>
<td>--------------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Group’s perimeter</td>
<td>Split of groups (due to inconsistent information between sources)</td>
<td>Missing global employment can be an indicator</td>
<td>Underestimation of the GEGs size, over-weighting of the small GEGs in the whole population</td>
</tr>
<tr>
<td>Global employment</td>
<td>Missing data</td>
<td></td>
<td>Underestimation of the GEGs size</td>
</tr>
<tr>
<td>Country of the global</td>
<td>Confusion between the global group head and the global decision centre.</td>
<td>High % of employment outside the GDC country</td>
<td>Wrong interpretation by country</td>
</tr>
<tr>
<td>decision centre</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
How to deal with quality issues?

- Some issues can be centrally solved:
  - **By manual checking:**
    - Systematic search for the right GDC for the largest groups with GDC in fiscal heavens
    - Manual collection of economic information on GEGs by looking at the institutional information given by the GEG (annual report and/or web sites)
    - Identification and merging of groups with close names
  - **by imputation methods**
    - Estimation of the GEGs employment by summing the employment of their subsidiaries
    - Calculation of the NACE code of the GEG taking into account the NACE of their subsidiaries weighted by the employment
How to deal with quality issues?

• Some issues require NSIs intervention
• Need to guide them in their checking/correcting actions
  • Have a checking/correcting process consistent over the countries
• Central editing of issues
• To be checked by the concerned NSIs
# How to deal with quality issues?

## Preliminary list of anomalies to be edited

**At the legal unit level:**

- List of legal units with a NACE code of 6420 and an employment > 10
- List of legal units with a high employment (>5000) and a reference date < n - 3
- List of legal units with reference dates < n – 5 (for NACE, employment, and status)

**At the group level:**

- GEGs without employment figures
- GEGs with a global employment lower than the sum of the subsidiaries employment
- GEGs with a high part of employment outside the GDC country
- GEGs with no active legal unit outside the country of the GDC
- GEGs without employee outside the country of the GDC
- GEGs with an employment growth rate higher than 25%. It can be the signal of an acquisition or of an anomaly in the current process.
- GEGs with a decrease of the employment higher than 50%
Dissemination

- Compilation of economic indicators according to 2 dimensions
  - Economic impact of globalization in EU countries
    - Share of employment in enterprises under foreign control
  - Global economic impact of European multinational enterprise groups
    - Growth rate of total employment in EU GEGs
    - Growth rate of EU employment in EU GEGs
- Tables by EU-countries
- Ad-hoc economic analysis
Examples

• Be sure that the quality of data is high enough
  • Cleaning of EGR file to remove (as much as possible) suspicious groups
    ✓ Groups with no employment outside of the GDC country
    ✓ Groups active in only one country
    ✓ Groups with less than 200 persons employed
    ✓ 15655 GEGs in EGR 2011 ➔ 8184 GEGs after cleaning
    ✓ Manual identification of 250 duplicated groups
  
• Restriction of the study population to the GEGs for which global economic information has been checked
  ✓ Global data consistent with the GEG information
  ✓ Improvement of the economic figures
### Representativeness of the study population

<table>
<thead>
<tr>
<th></th>
<th>Number of GEGs</th>
<th>GEGs' Employment</th>
<th>Representativeness in EGR Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>EGR groups selected for the study (reference year 2011)</td>
<td>8184</td>
<td>92862982</td>
<td></td>
</tr>
<tr>
<td>Groups with manually checked employment data in 2012</td>
<td>3134</td>
<td>77631367</td>
<td>38%</td>
</tr>
<tr>
<td>Groups with manually checked employment data in 2010, 2011 and 2012</td>
<td>2345</td>
<td>55543524</td>
<td>29%</td>
</tr>
</tbody>
</table>

Better representativeness in terms of employment
- Small groups under-represented
- Size analysis is not possible
Examples

EU groups contribution according to the GDC country

- Global employment 2010
- Global employment 2011
- Global employment 2012
Employment evolution between 2010 and 2012 according to the GDC country

To be compared with evolution in EU countries and in GDC country
Quality of the figures less high

Employment in non-EU countries under-estimated

Reveal potential problems of GDC (LI, LU and CY)
Thank you for your attention