Integrating Data Collection: wins and challenges

2016 October, 4th

Paulo Saraiva
paulo.saraiva@ine.pt
Integrated Data Collection

Centralized Data Collection
Since 2005
Modular components

SIGINQ

Survey  Unit  Occur  Collect  Report  Analysis  Update  Manag.  Help

Method  SIGUA  Supplement  Errors  Specific  Consult  Batch  Table  GPap
         block prop  Launch  Tables  transfers  update  Manag.

Specific  Manage  By mode  Status  Generic  Transfer to  Survey  Register
Generic  Cross  Consult  Validations  Specific  analysis  Sample
          Reports  Reports  Consult

Specific  Open / Close  Primary Val  Generic  Consult  Common
          Close  Reports

Specific  Upload  Insert  Manage  REUSE  Survey  Specific
          entries

Data Entry
## Conceptual Variables List by Theme

<table>
<thead>
<tr>
<th>Designation</th>
<th>Validity Start</th>
<th>Metadata System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imports of goods</td>
<td>10-05-2006</td>
<td></td>
</tr>
<tr>
<td>Exports of goods</td>
<td>10-05-2006</td>
<td></td>
</tr>
<tr>
<td>Type of trade of enterprise</td>
<td>09-06-2006</td>
<td></td>
</tr>
<tr>
<td>Monthly indexes of unitary value of imports</td>
<td>17-11-2006</td>
<td></td>
</tr>
<tr>
<td>Geographic localization of good</td>
<td>28-02-2007</td>
<td></td>
</tr>
<tr>
<td>Place of source of good</td>
<td>28-02-2007</td>
<td></td>
</tr>
</tbody>
</table>
Electronic Data Collection

- Internet service
- All business surveys with webform;
- **95.5% via internet** (2015)
- Multiple associations.
Respondent management

Unit: 01.0000000000 – COMPANY TEST 1, S.A.

Related subscribers

<table>
<thead>
<tr>
<th>ID</th>
<th>Option</th>
<th>Subscriber</th>
<th>Creation</th>
<th>Survey</th>
<th>Activation</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGE000001</td>
<td></td>
<td>Peter Respondent Smith</td>
<td>2006-01-12 18:43</td>
<td>ICC</td>
<td></td>
</tr>
<tr>
<td>SHL000002</td>
<td></td>
<td>Mary Respondent Lee</td>
<td>2007-11-20 15:17</td>
<td>IECom</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>INTRA-CH</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>INTRA-EX</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IENF</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IUIX</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IVNE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IVNEI</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PTRE</td>
<td></td>
</tr>
<tr>
<td>TFA000003</td>
<td></td>
<td>Tim Respondent Amber</td>
<td>2009-04-03 17:54</td>
<td>IECEPA</td>
<td></td>
</tr>
</tbody>
</table>

GPAP information

Survey | Sample | Occurrence | Status   | Start    | End      | Mode     | Date (subscriber ID) | Type   |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ICC</td>
<td>6</td>
<td>201609</td>
<td>Registado</td>
<td>01-09-2016</td>
<td>14-09-2016</td>
<td>Webinq</td>
<td>09-09-2016 (SHL000002)</td>
<td>Resposta</td>
</tr>
<tr>
<td>ICC</td>
<td>6</td>
<td>201608</td>
<td>Registado</td>
<td>01-08-2016</td>
<td>12-08-2016</td>
<td>Webinq</td>
<td>03-08-2016 (SHL000002)</td>
<td>Resposta</td>
</tr>
</tbody>
</table>
Benefits of integration

- Process **Standardization** and **Harmonization**: By creating uniform processes across various production units (S) and eliminating differences between processes with the same scope (H) we have improved economics of scale and efficiency, as well as reducing the risk of failure.
- Production becomes **more reliable**, with more **quality** and less variations of levels;
- Less **expenses** in development of innovative new practices, and in the administration of processes;
Benefits of integration

• Speeding up new developments;
• Reduced time in delivering results;
• Freeing resources for more valued tasks;
• Accelerated learning curve; more knowledge and versatility of the technical staff;
• Intensive data sharing and integration, as well as parallel processes;
• Better relationship with data providers;
• Improved accountability and security.
Data integration

Multi Survey integration:
- Survey code
- Iteration ID
- Unit ID
- Variable ID
- Variable Value
- Value Version ID
- Data Collection Mode

Multi View integration:
- Surveys variables
- Registers
- Metadata
- Paradata
Big win!

In-house solution

• In-house solution;
• Teamwork involving multidisciplinary skills;
• High value knowledge asset.
Challenges

All that shines is not gold…

• Still remaining survey oriented;
  – “It is now too easy to set up a new survey!”

• Single point of failure perception;

• Volatility of statistical units definitions;

• Some loss of flexibility;

• New data sources…
Challenges

Improve collection designs

<table>
<thead>
<tr>
<th>Specify Needs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Identify needs</td>
</tr>
<tr>
<td>1.2 Consult &amp; confirm needs</td>
</tr>
<tr>
<td>1.3 Establish output objectives</td>
</tr>
<tr>
<td>1.4 Identify concepts</td>
</tr>
<tr>
<td>1.5 Check data availability</td>
</tr>
<tr>
<td>1.6 Prepare business case</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Design</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Design outputs</td>
</tr>
<tr>
<td>2.2 Design variable descriptions</td>
</tr>
<tr>
<td>2.3 Design collection</td>
</tr>
<tr>
<td>2.4 Design frame &amp; sample</td>
</tr>
<tr>
<td>2.5 Design processing &amp; analysis</td>
</tr>
<tr>
<td>2.6 Design production systems &amp; workflow</td>
</tr>
</tbody>
</table>

- Improve GSBPM *Specify Needs* and *Design* phases
- Strong collaborative approach across all NSI
Challenges

Full Multi Source Integration

- Surveys
- Simplified Business Information
- Administrative Data
Challenges
Unified Registers

Integration

Unified Statistical Units System

Business
Households
Agriculture
... Others

Statistical Unit Registers
Population & Samples
Challenges

Multi purpose views

Internal users
- Improved production quality;
- Coherence analysis.

External users
- Improved data access (secure and confidential) for:
  - Academia and researchers;
  - Data Providers.
Challenges

Remain Relevant

REUSE

admin
data

Motivate

Management

Respondent

PERSONAL FEEDBACK
Integrating Data Collection: wins and challenges

Paulo Saraiva
paulo.saraiva@ine.pt

“Life begins at the end of your Comfort Zone”

- Neale Walsh