HLG Initiatives and SDMX role in them

Steven Vale
UNECE
steven.vale@unece.org

Juan Muñoz
INEGI
Juan.Munoz@inegi.org.mx
Contents

- The age for statistics modernization
  - Modernization and the Need for Standards
- Modernization and the High-level Group
  - HLG Initiatives
- HLG Projects and the role of SDMX in them
  - CSPA Implementation
  - Big Data
Need for Standards

• A standard defines a level of quality or attainment.
• Standards establishes principles and models and pave the way to share knowledge.
• Sharing knowledge is a way to share value and improvements.
• … and we are living in an age of fast evolution, where modernization of process, products and services are compulsory.
The age of new challenges

What Happens in an Internet Minute?

- 639,800 GB of global IP data transferred
- 20 New victims of identity theft
- 204 million Emails sent
- 47,000 App downloads
- $83,000 in sales
- 61,141 Hours of music
- 20 million Photo views
- 320+ New Twitter accounts
- 3,000 Photo uploads
- 100,000 New tweets
- 135 Botnet infections
- 1,300 New mobile users
- 100+ New LinkedIn accounts
- 277,000 Logins
- 6 million Facebook views
- 2+ million Search queries
- 30 Hours of video uploaded
- 1.3 million Video views

And Future Growth is Staggering

Source: Hongkiat.com
Data deluge

In the last 2 years more information was created than in the whole of the rest of human history!

All Global Data in Zettabytes

1ZB = 1,126,000,000,000,000,000,000 bytes (approx)
Nature of Data is Changing
The Challenges

- Increasing cost & difficulty of acquiring data
- New competitors & changing expectations
- Rapid changes in the environment
- Reducing budget
- Competition for skilled resources
- Riding the big data wave
Using common standards, statistics can be produced more efficiently

No domain is special!

Do new methods and tools support this vision, or do they reinforce a stove-pipe mentality?
Balanced evolution

Products and Services

Processes

Standards and tools

Evolution
Organizing for Modernization

- High Level Group for the Modernization of Statistical Production and Services (HLG)
  - Created by the Conference of European Statisticians in 2010
  - 10 heads of national and international statistical organisations
Standards-based Modernization
Modernization Committees

- Organization Framework and Evaluation
  - Change Management, Organizational Frameworks for Collaboration, Legal and Licensing, Building Competencies, Guidelines for Managers including best practice, Evaluation including Costs and Benefits, Communicating Modernization

- Production and Methods
  - CSPA, Big Data, Open Data, IT Methodology, Collaboration, Skills

- Products and Sources
  - Big Data, CSPA implementation, Mixed mode data collection, Mobile devices, Marketing, CES paper on the value of official statistics

- Standards
  - Governance, maintenance, support and integration of key standards, Support for CSPA implementation, Quality indicators, Semantic web, Link to geospatial standards, Big Data, Glossary
The GSBPM 5.0

Quality Management / Metadata Management

Specify Needs
- 1.1 Identify needs
- 1.2 Consult & confirm needs
- 1.3 Establish output objectives
- 1.4 Identify concepts
- 1.5 Check data availability
- 1.6 Prepare business case

Design
- 2.1 Design outputs
- 2.2 Design variable descriptions
- 2.3 Design collection
- 2.4 Design frame & sample
- 2.5 Design processing & analysis
- 2.6 Design production systems & workflow

Build
- 3.1 Build collection instrument
- 3.2 Build or enhance process components
- 3.3 Build or enhance dissemination components
- 3.4 Configure workflows
- 3.5 Test production system
- 3.6 Test statistical business process
- 3.7 Finalise production system

Collect
- 4.1 Create frame & select sample
- 4.2 Set up collection
- 4.3 Run collection
- 4.4 Finalise collection

Process
- 5.1 Integrate data
- 5.2 Classify & code
- 5.3 Review & validate
- 5.4 Edit & impute
- 5.5 Derive new variables & units
- 5.6 Calculate weights
- 5.7 Calculate aggregates
- 5.8 Finalise data files

Analyse
- 6.1 Prepare draft outputs
- 6.2 Validate outputs
- 6.3 Interpret & explain outputs
- 6.4 Apply disclosure control
- 6.5 Finalise outputs
- 6.6 Calculate aggregates

Disseminate
- 7.1 Update output systems
- 7.2 Produce dissemination products
- 7.3 Manage release of dissemination products
- 7.4 Promote dissemination products
- 7.5 Manage user support

Evaluate
- 8.1 Gather evaluation inputs
- 8.2 Conduct evaluation
- 8.3 Agree an action plan
Framework for combining standards?

## COMPLEMENTARY

### Quality Management / Metadata Management

<table>
<thead>
<tr>
<th>Specify Needs</th>
<th>Design</th>
<th>Build</th>
<th>Collect</th>
<th>Process</th>
<th>Analyze</th>
<th>Disseminate</th>
<th>Evaluate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Identify needs</td>
<td>2.1 Design outputs</td>
<td>3.1 Build collection instrument</td>
<td>4.1 Create frame &amp; select sample</td>
<td>5.1 Integrate data</td>
<td>7.1 Update output systems</td>
<td>8.1 Gather evaluation inputs</td>
<td></td>
</tr>
<tr>
<td>1.2 Consult &amp; confirm needs</td>
<td>2.2 Design variable descriptions</td>
<td>3.2 Build or enhance process components</td>
<td>4.2 Set up collection</td>
<td>5.2 Classify &amp; code</td>
<td>7.2 Produce dissemination products</td>
<td>8.2 Conduct evaluation</td>
<td></td>
</tr>
<tr>
<td>1.3 Establish output objectives</td>
<td>2.3 Design collection</td>
<td>3.3 Build or enhance dissemination components</td>
<td>4.3 Run collection</td>
<td>5.3 Review &amp; validate</td>
<td>7.3 Manage release of dissemination products</td>
<td>8.3 Agree an action plan</td>
<td></td>
</tr>
<tr>
<td>1.4 Identify concepts</td>
<td>2.4 Design frame &amp; sample</td>
<td>3.4 Configure workflows</td>
<td>4.4 Finalise collection</td>
<td>5.4 Edit &amp; impute</td>
<td>7.4 Promote dissemination products</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5 Check data availability</td>
<td>2.5 Design processing &amp; analysis</td>
<td>3.5 Test production system</td>
<td>4.5 Finalise outputs</td>
<td>5.5 Finalise outputs</td>
<td>7.5 Manage user support</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.6 Prepare business case</td>
<td>2.6 Design production systems &amp; workflow</td>
<td>3.6 Test statistical business process</td>
<td>4.6 Finalise data files</td>
<td>5.6 Calculate weights</td>
<td>7.6 Calculate weights</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### SDMX

- Define metadata schema
- Define metadata elements
- Define metadata rules

### DDI

- Define data dictionary
- Define data elements
- Define data rules
GSIM and GSBPM

- GSIM describes the information objects and flows within the statistical business process.
GSIM 1.1

Used to capture the designs and plans of statistical programs, and the processes undertaken to deliver those programs.

Used to catalogue the information that comes in and out of a statistical organization via Exchange Channels.

Used to define the meaning of data, providing an understanding of what the data are measuring.

Used to define the terms used in relation to information and its structure.

- Business
- Exchange
- Structures
- Concepts
Common effort

WHO WAS INVOLVED?

Australia
Canada
New Zealand
Sweden
United Kingdom
United States
Eurostat
Norway
IMF
Republic of Korea
Slovenia
UNECE
Mexico
Italy
Netherlands
Denmark
France
Conceptualization

Information Objects

Representation

Data Structures
HLG Projects 2014
Frameworks and Standards for Statistical Modernization
Layers of Architecture

- Business Layer
- Information Layer
- Implementation (Applications + Technology) Layer
Fostering Interoperability in Official Statistics: Common Statistical Production Architecture
Historically, statistical organizations have produced specialized business processes, methods and IT systems for each survey / output.
Enterprise Architecture: Common Statistical Production Architecture
## Implementing CSPA

<table>
<thead>
<tr>
<th>SERVICE</th>
<th>ORGANIZATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confidentialized analysis of microdata</td>
<td>Australian Bureau of Statistics Statistics Canada</td>
</tr>
<tr>
<td>Error correction</td>
<td>Istat</td>
</tr>
<tr>
<td>Linear error localisation</td>
<td>Statistics Netherlands</td>
</tr>
<tr>
<td>Linear rule checking</td>
<td>Statistics Netherlands</td>
</tr>
<tr>
<td>List statistical classifications</td>
<td>Statistics Norway</td>
</tr>
<tr>
<td>Retrieve statistical classification</td>
<td>Statistics Norway</td>
</tr>
<tr>
<td><strong>SDMX transform</strong></td>
<td>OECD</td>
</tr>
<tr>
<td>Sample selection</td>
<td>Statistics Netherlands</td>
</tr>
<tr>
<td>Seasonal adjustment</td>
<td>Australian Bureau of Statistics INSEE Statistics New Zealand</td>
</tr>
<tr>
<td>Statistical chart generator</td>
<td>OECD</td>
</tr>
</tbody>
</table>
Big Data
HLG and Big Data

- Paper: “What does Big Data mean for official statistics?”
- Project proposal from global task team:
  - Work package 1: Strategy and methodology
  - Work package 2: Shared computing environment (“sandbox”), practical application of methods and tools
  - Work package 3: Training and dissemination
Objectives

- To identify main strategic and methodological issues for official statistics industry
- To demonstrate the feasibility of efficient production of both novel products and ‘mainstream’ official statistics using Big Data sources
- To facilitate the sharing across organizations of knowledge, expertise, tools and methods
Task teams

- Partnerships
  - New providers and sources of data
- Privacy
  - Confidentiality and legal issues
- Quality
  - Quality framework
- Sandbox
  - IT and Methodology
  - Make experiments
SDMX and Big Data?

- Is there a role for SDMX with Big Data?
  - SDMX “too heavy”?
  - Data too big to exchange?
  - Fundamental paradigm shift: Process data at source (or in the cloud) rather than in house?
  - Just transfer aggregates back to statistical organisations? Could SDMX have a role here?

More research needed!
Some ideas are coming…

- BIG DATA could be important but at the end it’s just another source of information…
- SDMX can play a BIG ROLE:
  - Integrating different sources (results from pre-processing)
  - Transferring results of processed big data
  - Visualization of results
  - Mobility…
Get involved!

Anyone is welcome to contribute!

More Information

• HLG Wiki:  
  http://www1.unece.org/stat/platform/display/hlgbas

• LinkedIn group “Business architecture in statistics”
QUESTIONS?

Juan.Munoz@inegi.org.mx