

# Process metadata development and implementation under the GSBPM v5.0 at Statistics Spain (INE)

24 - Metadata Systems

June 2, 2016

**David Salgado** - Statistics Spain (INE)

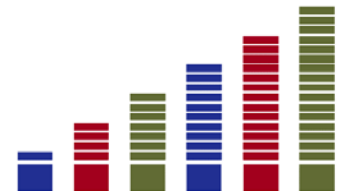
[david.salgado.fernandez@ine.es](mailto:david.salgado.fernandez@ine.es)

Ana I. Sánchez-Luengo - Statistics Spain (INE)

[anaisabel.sanchez.luengo@ine.es](mailto:anaisabel.sanchez.luengo@ine.es)

## Goals

- (i) To fulfil **Statistics Spain's commitment** made in the second round (2013-2015) of **Peer Reviews** within the European Statistical System (ESS).
  
- (ii) To assure the **institutional sustainability** of the production of Statistics Spain by documenting the current statistical production processes.
  
- (iii) To pave the way for a **deep analysis for a standardised production model**.

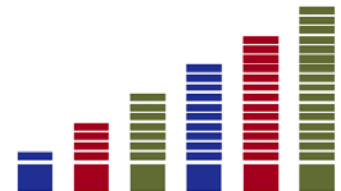


# Pilot Experience

**7** statistical operations GSBPM **phases 4-7** described at **level 2**.

1. No **detailed** information could be attained.
2. Documentation to an **extremely diverse degree of detail** (from very limited to highly condensed).
3. **Impossible** to have **comparable GSBPM level-2 processes** among different statistical operations.

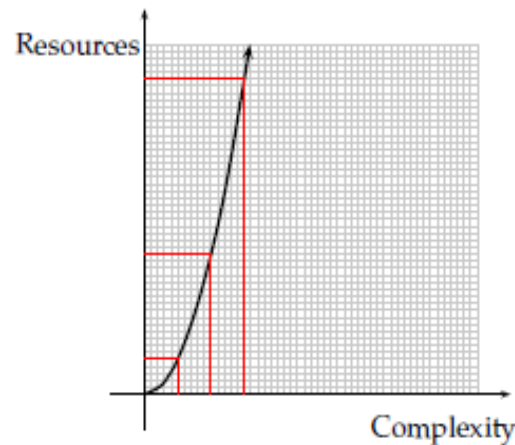
We put in place **level 3 (tasks)** for the GSBPM adapted to our needs.



# General principles

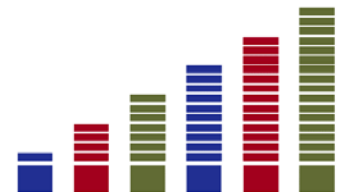
## Premise:

An official statistical production system is a **complex system**



## Approach:

Computer system design principles → **Functional modularity**



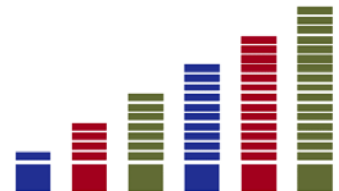
# The standard

## Tasks

- Tasks within each level-2 process: **246 tasks**, in total.
- Tasks **modularised with methodological criteria**.
- A need for **task integration**.
- Documentation elements: **input, output, throughput, tools, documentation, responsible unit(s)**.

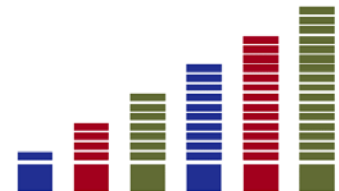
## Workflow

- Business Process Modelling and Notation language (BPMN v2.0).



# Some details

- **Task names:** *verb + name*  
Identify population; Design questionnaire; Develop IS component for error detection; Execute coding; Gather reports
- Need for a **controlled vocabulary** → thesaurus?  
Variable vs. Aggregate; Frame population vs. Directory
- **Accretion** of production elements  
Data collection management application =  
(((variables) + questionnaire) + collection instrument) + paradata
- Use of functionally modular **parameterised activities**  
`SubstituteVar(VarName, InFile, FromFile)`
- **Cross-sectional** tools documented apart



# Implementation

- Internal **working group**

Experts from data quality, data collection, sampling, dissemination, IT, 5 subject domains, statistical methodology, statistical metadata)

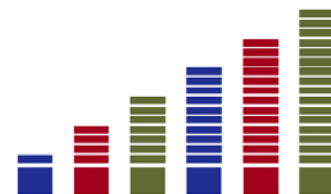
- **5 initial statistical operations to test the standard**

- **User-oriented documentation + internal briefings**

- **30 + 30 surveys currently under implementation**  
(human resource problem)

- **Computer application** developed for composing **task documentation pages**

- Only for **internal use** so far



## Some conclusions

- A modern metadata system is a **must for modernising production**
- Cultural **resistance to change**: limited conception of metadata
- **Lack of human resources** → high workload
- **Dipolar situation** between **statistics** and **computer science**: misconceptions

