

Adaptation of the GSBPM Model to Statistics Production Based on an Integrated System of Statistical Registers

ModernStats World Workshop 2022

27- 29 June 2022

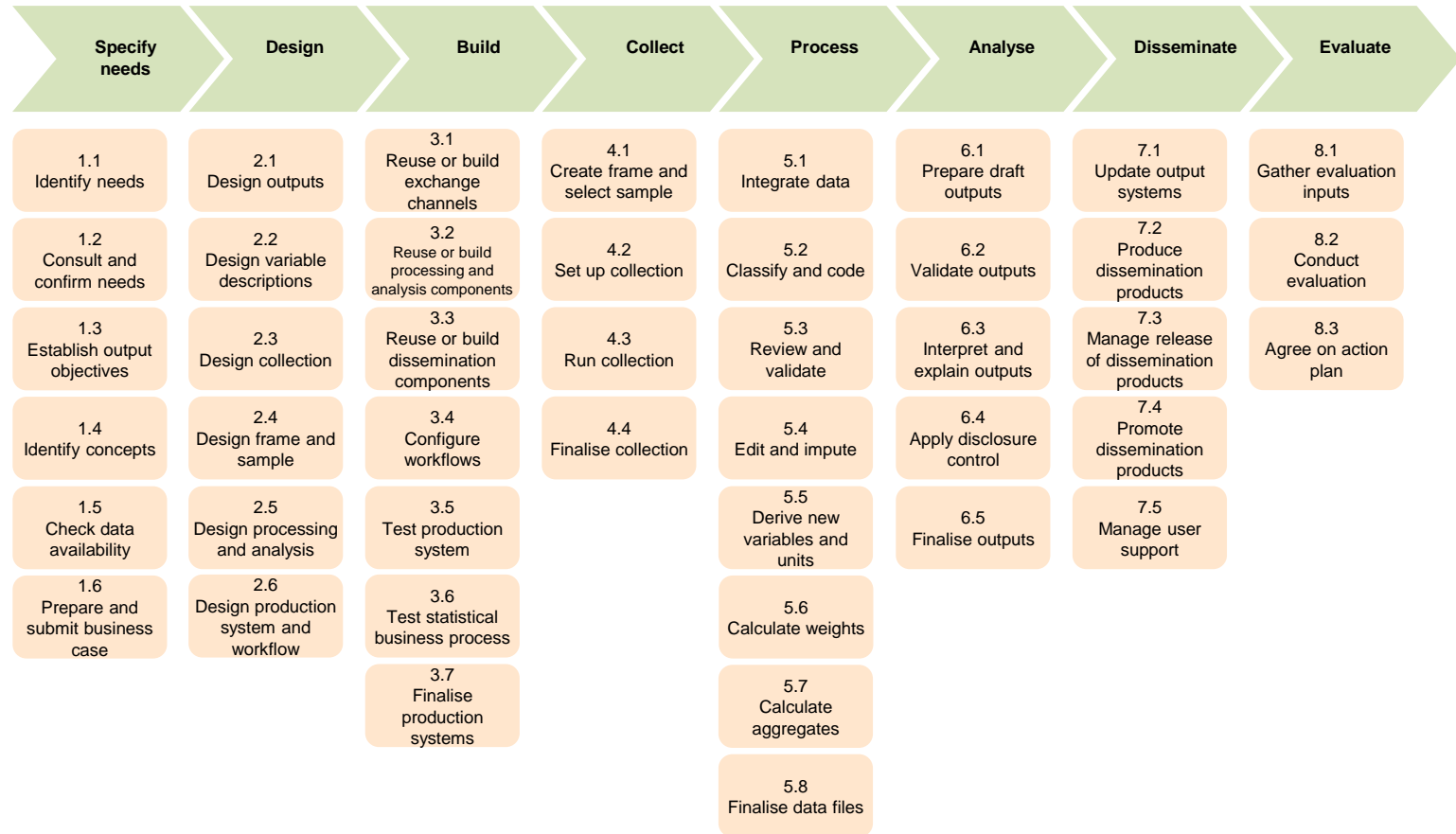
Belgrade, Serbia

Federico Segui – Deputy Director General, NSI of Uruguay

Edgardo Greising – Head of Knowledge Management Solutions Unit – STATISTICS, ILO

GSBPM 5.1

Overarching Processes



Facing the problem...

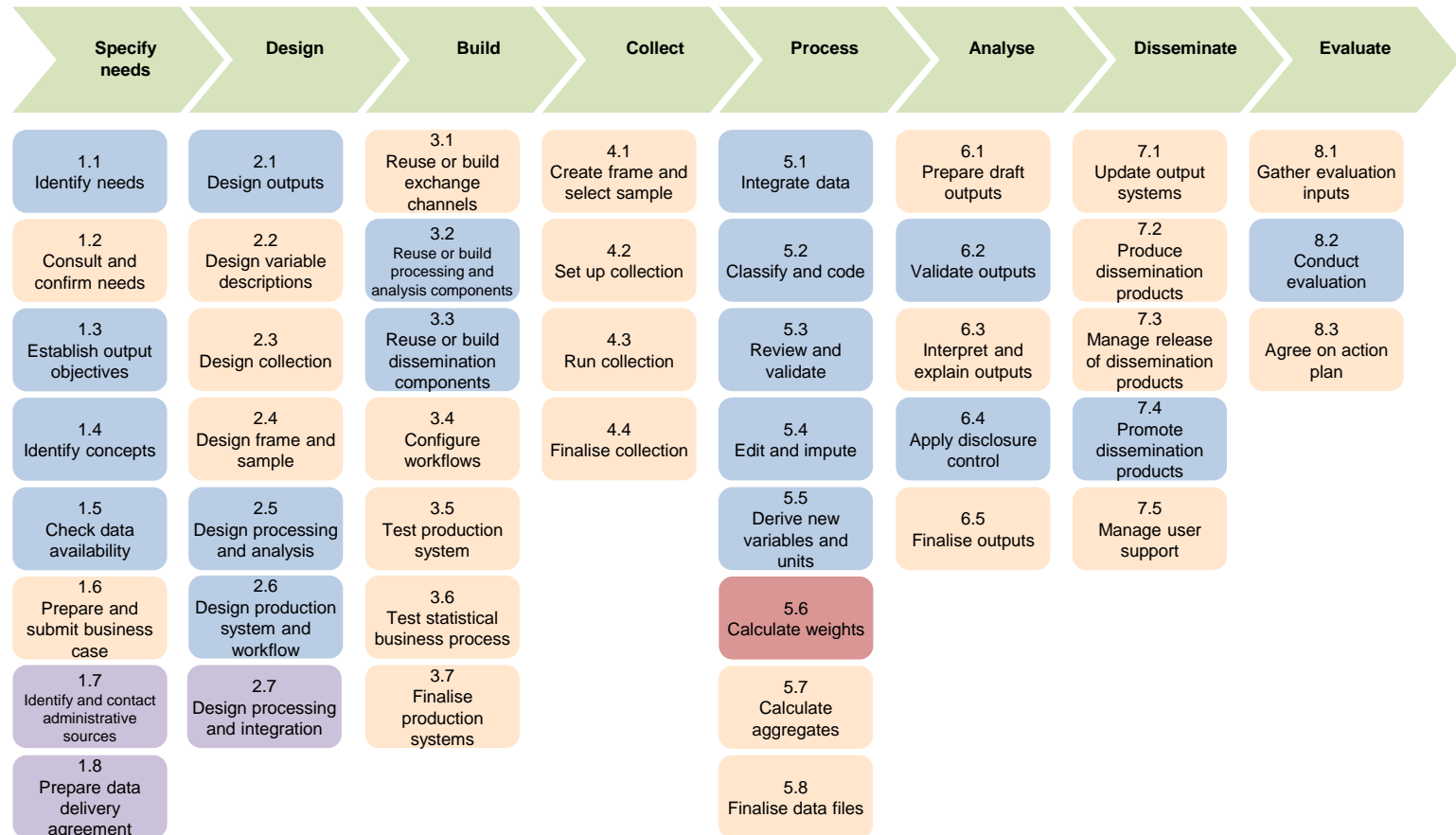
- The GSBPM model is mostly oriented to the processes of generating statistics based on censuses, surveys and statistical registers but does not map properly the production processes (creation and updating) of statistical registers that are part of an integrated system of registers.
- These are continuous “transversal” processes, independent from the several “vertical” statistical projects or studies that make use of them, which are obviously linked to the phases and sub-processes of the GSBPM.

Facing the problem...

- An empirical study was carried out applying the GSBPM model to the production of statistics based on an integrated system of statistical registers:
 - During the IADB project “Population and real estate statistics based on the use of official administrative registers in the Andean Community (2015-2018)”.
 - That empirical work was extended in Uruguay, reaching to the same conclusions.
- Anders and Britt Wallgren¹ also have some doubts about the GSBPM model as is in terms of its application in a statistics production model based on an integrated system of statistical registers.

Proposal of adaptation of GSBPM(-SRS)

Overarching Processes



Adapted

New Sub-process

Not apply

Specify needs

- User needs include requirements for updating and improving the base registers and the integrated system of statistical registers as a whole.
- The administrative registers variables necessary to meet the statistical requirements of the users and/or the integrated system of statistical registers are identified.
- Assess the usefulness of administrative data in terms of coverage, relevance, timeliness and quality or classify for further use.
- Creates or updates the Inventory of Administrative Registers (and its variables) and data owners.
- **(new) 1.7. Identify and contact administrative sources**
- **(new) 1.8. Prepare data delivery agreement**

Design

- Document the criteria for the selection of the variables of administrative registers.
- A cross-reference register should be kept with all changes of unit's identification over time.
- Preliminary discovery of administrative register data to know the data beforehand in order to elaborate a *data profile* that facilitates the design of procedures.
- Establish how to update the files or tables of the statistical production database corresponding to the statistical registers.
- **(new) 2.7 Design processing and integration**
 - Design the statistical processing methodology to be applied to the administrative registers during the 'Process' phase.

Build

- Develop new or improve existing components and services necessary for the "Transformation" of administrative registers into statistical registers and their "integration".
 - Development of routines to standardise variables, detect duplicates, code, cleanse, process, derive new variables and units, de-identify registers (protect data) and validate, update and finalise datasets.
- Communicate the updating of the system of integrated statistical registers, new statistical registers within the system and updating of base registers.

Process

- Transformation of the administrative register into a statistical register and its preparation for integration with other registers.
 - Standardise or harmonise variables, validate, clean and transform the input data, so that they can be used for statistical purposes.
 - Create/update statistical registers.
- Record linkage.
 - Prepare variables: normalisation of texts, transformation of numerical variables or date to text, concatenation of variables (key identification variables are also used).
 - Deterministic and probabilistic methods.
- Complete data discovery (data profiling).
- Detect duplicates: deterministic and probabilistic methods. Duplicates are marked as deleted (logical deletion).
- New variables and units are created after record linkage.

Analyse, Disseminate & Evaluate

Analyse

- Evaluate integration (record linkage) results.
- Statistical register is de-identified.
 - i.e. the key identification variables are replaced by random keys.
 - Contact variables or explicit identifiers (names, telephone, address, email, geographical coordinates) are eliminated.

Disseminate

- Communicate changes and data updates made to the system of statistical registers.
 - Make users aware of the new scope and possibilities of the system in terms of statistical production.

Evaluate

- Quality assessment of administrative register once it is collected.
 - Determine the usefulness of the data and its restrictions in terms of statistical use.

Thank you