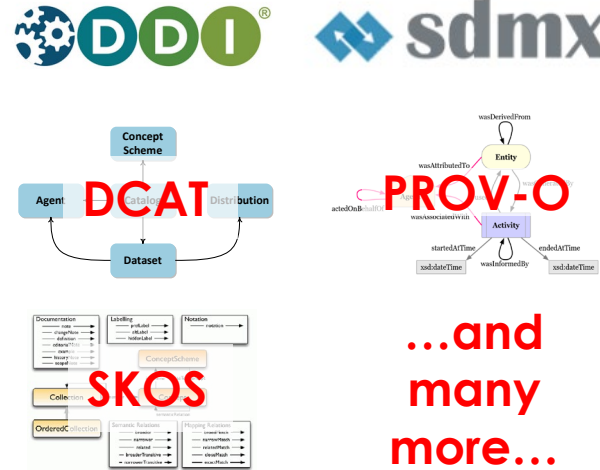
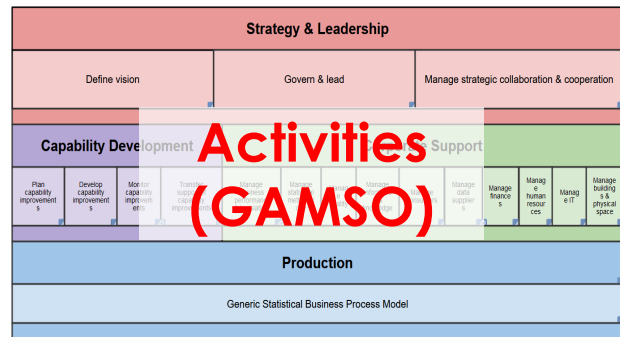
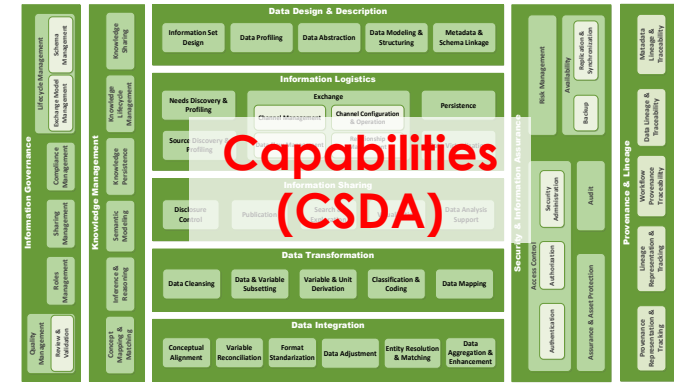
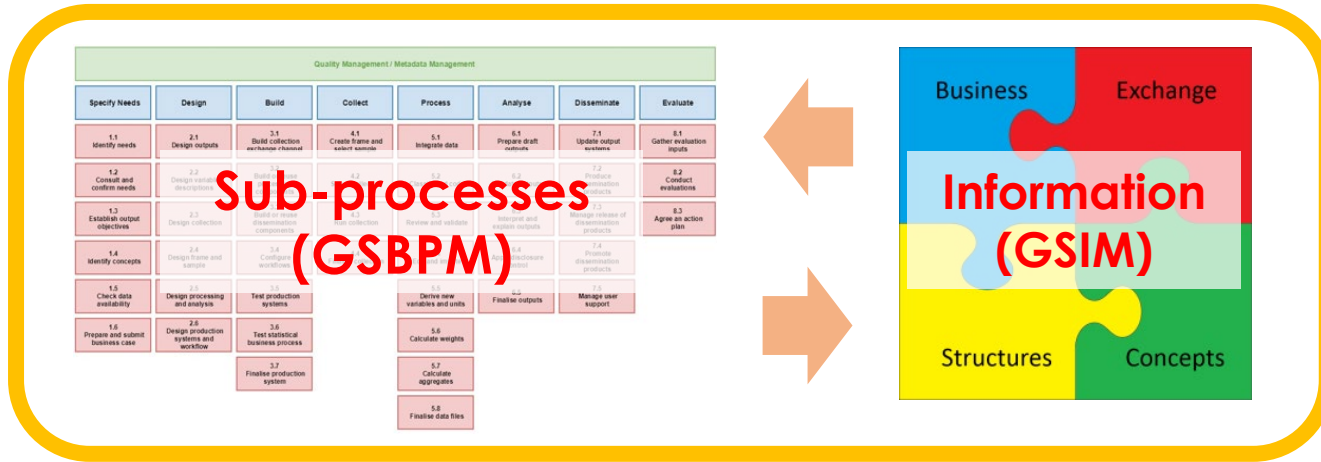


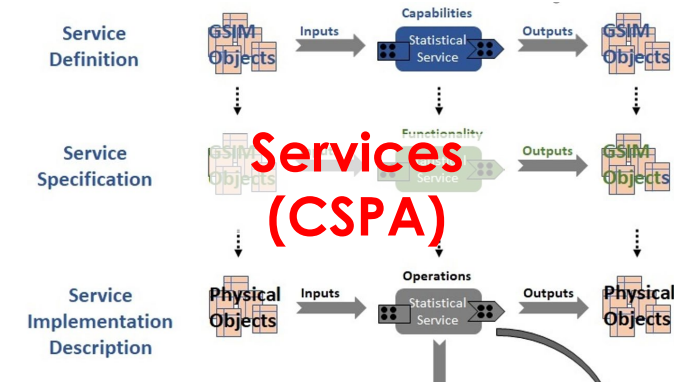
New Developments from the HLG- MOS Supporting Standards Group - Linking GSBPM and GSIM -

Flavio Rizzolo (Statistics Canada) - June 28th, 2022

Standards landscape



...and many more...



Linking GSBPM and GSIM - Task Team

20 members over 3 years from 11 participating institutions

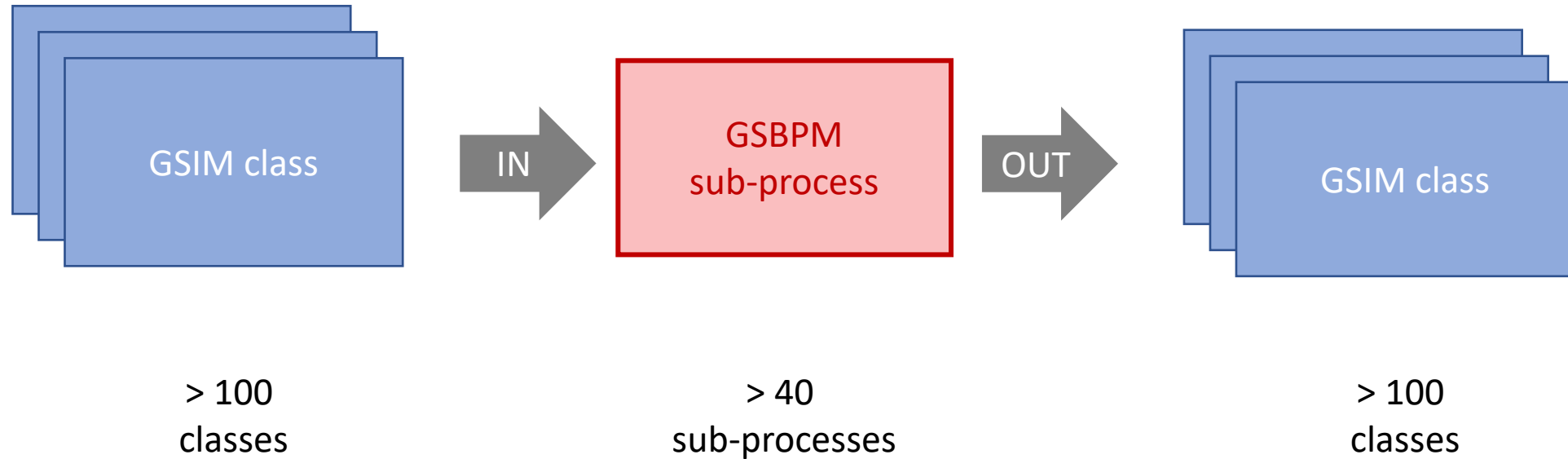
Country	Member
Italy	Marina Signore - Chair Giorgia Simeoni Mauro Scanu Claudia Brunini Carlo Vaccari
Canada	Flavio Rizzolo - Chair Francine Kalonji
Mexico	Juan Munoz Ildeliza Ramos Jesús Togno
Sweden	Patrik Wahlgren

Country	Member
Hungary	Zoltán Vereczkei Csaba Ábry Eva Hajosn Ender
Poland	Anna Dlugosz
South Korea	Yulla Choi
Australia	Annette Tyler
ILO	Edgardo Greising
DDI	Jay Greenfield
Egypt	Ayman Hathoot

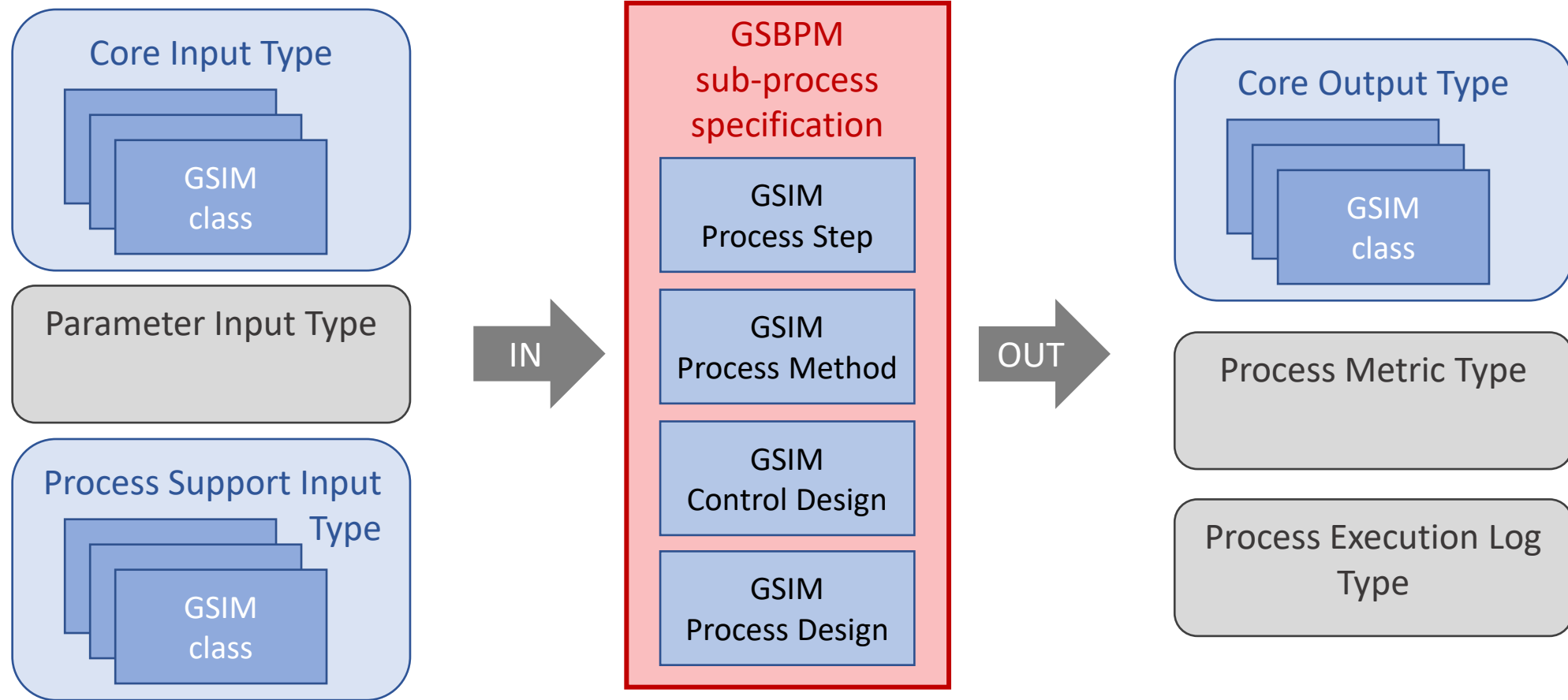
Linking GSBPM and GSIM - Goals

- Derive a robust set of GSIM classes that could be used as inputs and outputs based on use cases commonly taking place in many statistical organisations.
- Elicit non-obvious relationships between the two models, improving usability and supporting a wider adoption;
- Make it easier to design systems to track information flow through statistical business processes;
- Contribute to building a “de facto” integrated view of the ModernStats models (now expanded by the [Core Ontology for Official Statistics](#) work);
- Make it easier to create implementations, in particular CSPA services (now continued by the just-started [SDMX-DDI-GSBPM mapping](#) work)

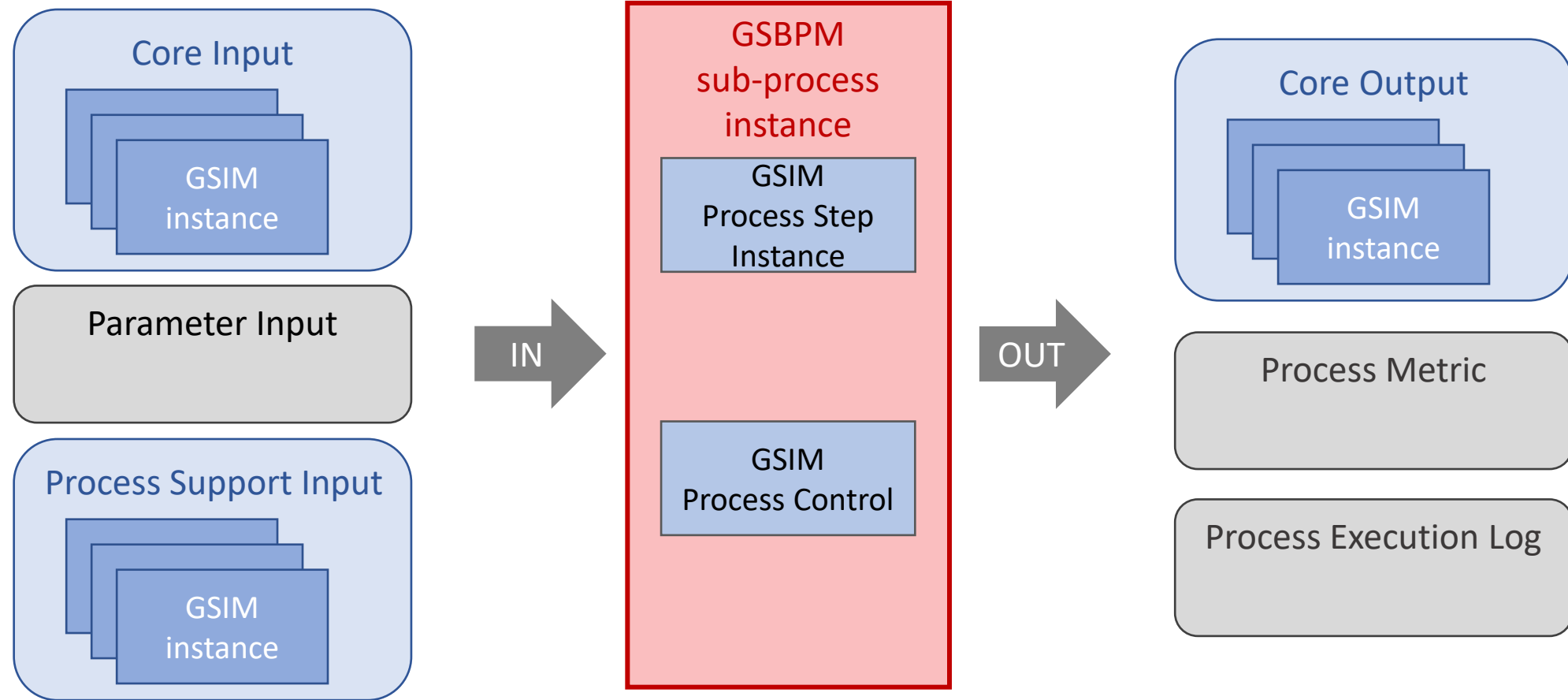
Linking GSBPM and GSIM - Context



Linking GSBPM and GSIM - Specification



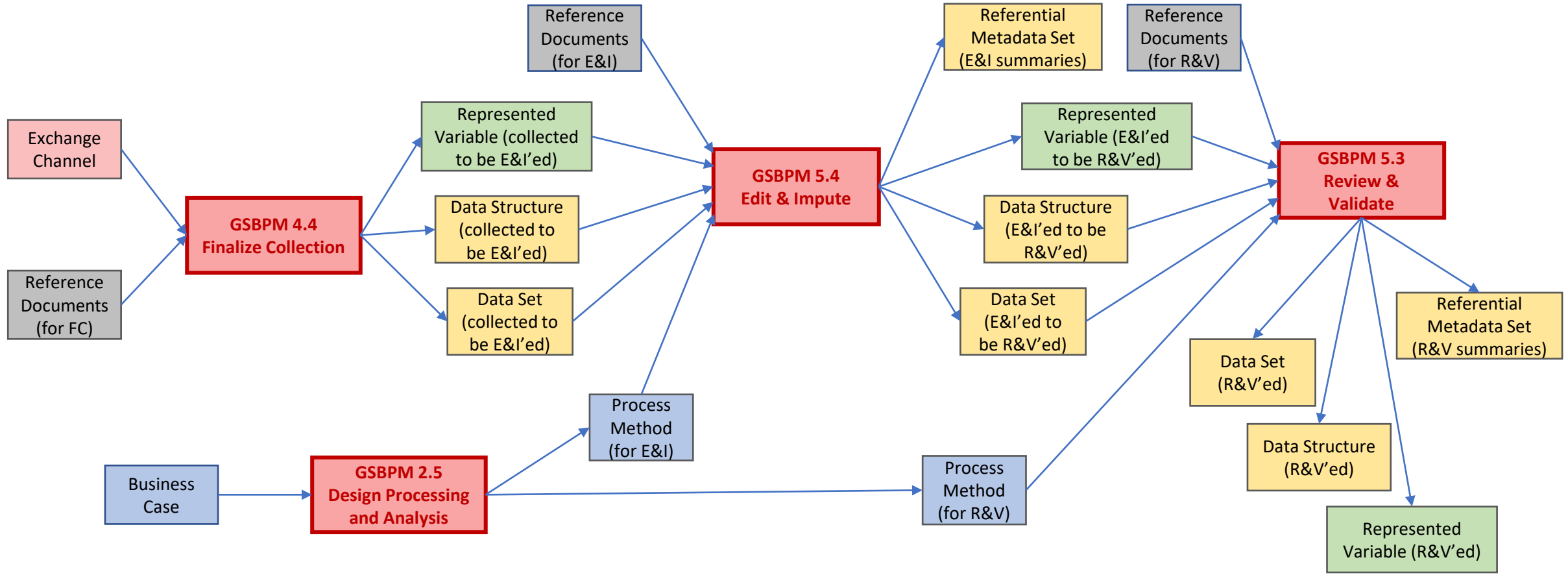
Linking GSBPM and GSIM - Instance



Specification template example: 5.3 Review and Validate

Process Input Specification	Process Design	Process Output Specification
<p><u>Core Input type</u></p> <p>Data Sets (Unit Data Sets) to be reviewed and validated</p> <p>Data Structures associated with Data Sets to understand Data Sets</p> <p>Represented Variables to be reviewed and validated</p> <p>Process Methods that specifies methodology for review and validation (e.g., calculating plausibility or validity) which can be represented as Rules, as designed in Phase 2</p> <p><u>Parameter Input type</u></p> <p>Parameter values to be used for review and validation methodologies as specified in Process Method such as:</p> <ul style="list-style-type: none"> • Limit value for edit Rule (interval for valid values) • Threshold for checking outlier <p><u>Process Support Input type</u></p> <p>Auxiliary Data Sets or any Information Resource to be used for review and validation, e.g., historic comparison, macro-level comparison</p> <p>Technical / methodological handbooks, policies or guidelines to be followed regarding data validation as well as quality management</p>	<p><u>Process Method</u></p> <p>Review Data Sets and Process Methods</p> <p>Apply Process Methods and Rules to review Data Sets</p> <p>Apply Process Methods and Rules to validate Data Sets</p> <p>Calculate quality measures specified by Process Methods</p> <p>Update Data Sets and associated element in Data Structure with results from review and validation</p>	<p><u>Core Output type</u></p> <p>Data Sets (Unit Data Sets): updated Data Set</p> <p>Data Structure associated with Data Set</p> <p>Referential Metadata Set: descriptions of the Process Methods used, quality information summarising Process Metrics or any other relevant information to be passed along with Data Sets</p> <p><u>Process Metric type</u></p> <p>Quality measures related to review and validation such as:</p> <ul style="list-style-type: none"> • Number of validations conducted • Number of outliers detected <p>Quality measures of Process Step such as:</p> <ul style="list-style-type: none"> • Time spent to complete the Process Step (derived from Process Execution Log) • Cost spent to complete the Process Step <p><u>Process Execution Log type</u></p> <p>Execution log such as</p> <ul style="list-style-type: none"> • Time that Process Step started • Time that Process Step ended • Any message or event log generated from software used for review and validation

Linking GSBPM and GSIM - Information Flows



Outcomes and lessons learned

- The two templates –specification and instance (running examples)– provide a useful way of eliciting use cases from participating NSOs
- The new information flow diagrams (<https://bit.ly/3OH2pwb>) show which GSIM classes are most used across the statistical business process... and the gaps that exist
- Detailed report <https://bit.ly/3QPSSVp>
- Useful feedback to the [GSIM revision](#) and [Core Ontology for Official Statistics](#) task teams
- Mapping exercise proved to be more challenging for early GSBPM phases
 - Processes around requirements and design are less clear
 - Terminological issues between the two models became more problematic
 - Subject matter expertise was harder to find
- A mapping to implementation standards, e.g. DDI and SDMX, will help making the models more precise and interoperable -> new task team on [SDMX-DDI-GSBPM mapping](#)

Related Links

- Generic Statistical Information Model (GSIM): <https://bit.ly/3I8SA8x>
- New GSIM Revision: <https://bit.ly/3OEDnxO>
- Generic Statistical Business Process Model (GSBPM): <https://bit.ly/3A6GPNR>
- GSBPM-GSIM information flows and report: <https://bit.ly/3Ov1JdW>
- Core Ontology for Official Statistics (COOS): <https://github.com/linked-statistics/COOS>
- Common Statistical Production Architecture (CSPA): <https://bit.ly/3bvBuW0>
- Common Statistical Data Architecture (CSDA): <https://bit.ly/3NhCCcY>
- Generic Activity Model for Statistical Organizations (GAMSO): <https://bit.ly/3HTn79X>

ModernStats World Workshop 2022

27-29 June @Belgrade, Serbia



Twitter: @unecestat #ModernStats

Wiki page for timetable and files: <https://statswiki.unece.org/x/sYAqF>