

Distr.  
GENERAL

WP.2  
16 April 2012

ENGLISH ONLY

UNITED NATIONS ECONOMIC COMMISSION  
FOR EUROPE (UNECE)  
CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION  
STATISTICAL OFFICE OF THE EUROPEAN  
UNION (EUROSTAT)

ORGANISATION FOR ECONOMIC COOPERATION  
AND DEVELOPMENT (OECD)  
STATISTICS DIRECTORATE

Meeting on the Management of Statistical Information Systems (MSIS 2012)  
(Washington, DC, 21-23 May 2012)

Topic (i): Architecture

## **Toward GSIM V1.0 as a cornerstone for common reference architecture**

### **Invited Paper**

Prepared by Alistair Hamilton and Siu-Ming Tam, Australian Bureau of Statistics (ABS), Australia

## **I. Introduction**

1. The Generic Statistical Information Model (GSIM)<sup>1</sup> is a cornerstone of the strategic vision<sup>2</sup> of the High Level Group for Strategic Developments in Business Architecture in Statistics (HLG-BAS)<sup>3</sup> related to industrialising process of producing official statistics.
2. GSIM is at a fairly early stage of development, especially compared to its fellow conceptual cornerstone, the Generic Statistical Business Process Model (GSBPM)<sup>4</sup>.
3. Several initiatives and agencies are currently seeking to progress industrialisation in practice. Many of these groups have identified a “critical path dependency” on GSIM being sufficiently defined and stable, so it can be harnessed as a common reference framework for their modelling statistical information. This is described further in Section II.B. of this paper.
4. HLG-BAS recognises the urgency of developing GSIM to a point where it can be applied in practice to support industrialisation. The initiation of the GSIM Sprint process by HLG-BAS toward the end of 2011 represented an innovative and strong response to this urgent priority.
5. The process of developing GSIM is itself a pioneering example of the type of collaboration across the official statistics industry which HLG-BAS seeks to foster in future.

---

<sup>1</sup> <http://www1.unece.org/stat/platform/display/metis/Generic+Statistical+Information+Model+%28GSIM%29>

<sup>2</sup> <http://www1.unece.org/stat/platform/display/hlgbas/Strategic+Vision>

<sup>3</sup> <http://www1.unece.org/stat/platform/display/hlgbas/High-Level+Group+for+Strategic+Developments+in+Business+Architecture+in+Statistics+%28HLG-BAS%29>

<sup>4</sup> <http://www1.unece.org/stat/platform/display/metis/The+Generic+Statistical+Business+Process+Model>

6. While other important development activities are currently being sponsored by HLG-BAS, such as developing and agreeing a strategy for implementing the strategic vision, the other activities are being progressed primarily by HLG-BAS members themselves. Development of GSIM, however, is the first collaboration under HLG-BAS to be progressed primarily by multidisciplinary teams following the broad directions and guidance provided by HLG-BAS members.
7. As well as exploring what GSIM is (and is not), and how it can contribute to common reference architecture; this paper also considers the process followed when developing GSIM. In particular, this paper considers lessons learned regarding approaches to collaborative development which might be adopted by future collaborations related to developing common architecture.
8. While, as described in Section II B, development of GSIM over the past two years has been led by other initiatives, it should be noted that the first reference to a Generic Statistical Business Information Model originated from a MSIS meeting in April 2010. This paper highlights the importance and value of the MSIS community's engagement in shaping further development of GSIM.

## II. Evolution of GSIM prior to Sprint 1

### A. Initial evolution of GSIM within the Statistical Network collaboration

9. A proposal to develop a Generic Statistical Business Information Model was discussed at the meeting of MSIS in April 2010<sup>5</sup>. The model would provide a common view of the business objects required to support business processes, and of the relationships between objects.
10. Two months later, the inaugural meeting of the *Informal CSTAT workgroup on stronger collaboration on Statistical Information Management Systems* was held (CSTAT is the OECD Committee on Statistics). The meeting initiated the Statistical Network<sup>6</sup>. NSIs within the Statistical Network aim to collaborate in regard to practical small steps to industrialise methods and processes, producing tangible results which quickly and effectively benefit all participating NSIs. The Statistical Network selected an initial set of five collaboration opportunities to improve statistical information management, four of which focused on specific processes, each of which was located within a particular phase of the GSBPM.
11. The Statistical Network also identified an essential role for a consistent reference model which would be used when defining the information required to drive any statistical production process and to describe the outputs from that process. Developing GSIM was seen as the highest priority strategic enabler for ensuring coherence across the practical "small step" collaborations the Statistical Network sought to undertake.
12. It was also recognised that GSIM could facilitate efficient and effective collaboration in the development and sharing of statistical information systems and statistical information management frameworks beyond the Statistical Network. GSIM could, in this way, help producers of official statistics build efficient metadata driven collection, processing, and dissemination systems
13. GSIM was envisaged as a common reference for modelling at the conceptual and logical/semantic levels. GSIM could then be "operationalized" on a consistent basis when defining the information required to drive statistical production processes, as well as to define the outputs (eg statistical data) and outcomes (eg process metrics) from those processes.
14. Operationalization, such as implementation in a specific IT system, was seen as progressing through associating GSIM with, for example, a commonly agreed to representation in XML. Rather than investing unnecessary time and money developing such a representation "from a clean sheet of paper", the approach

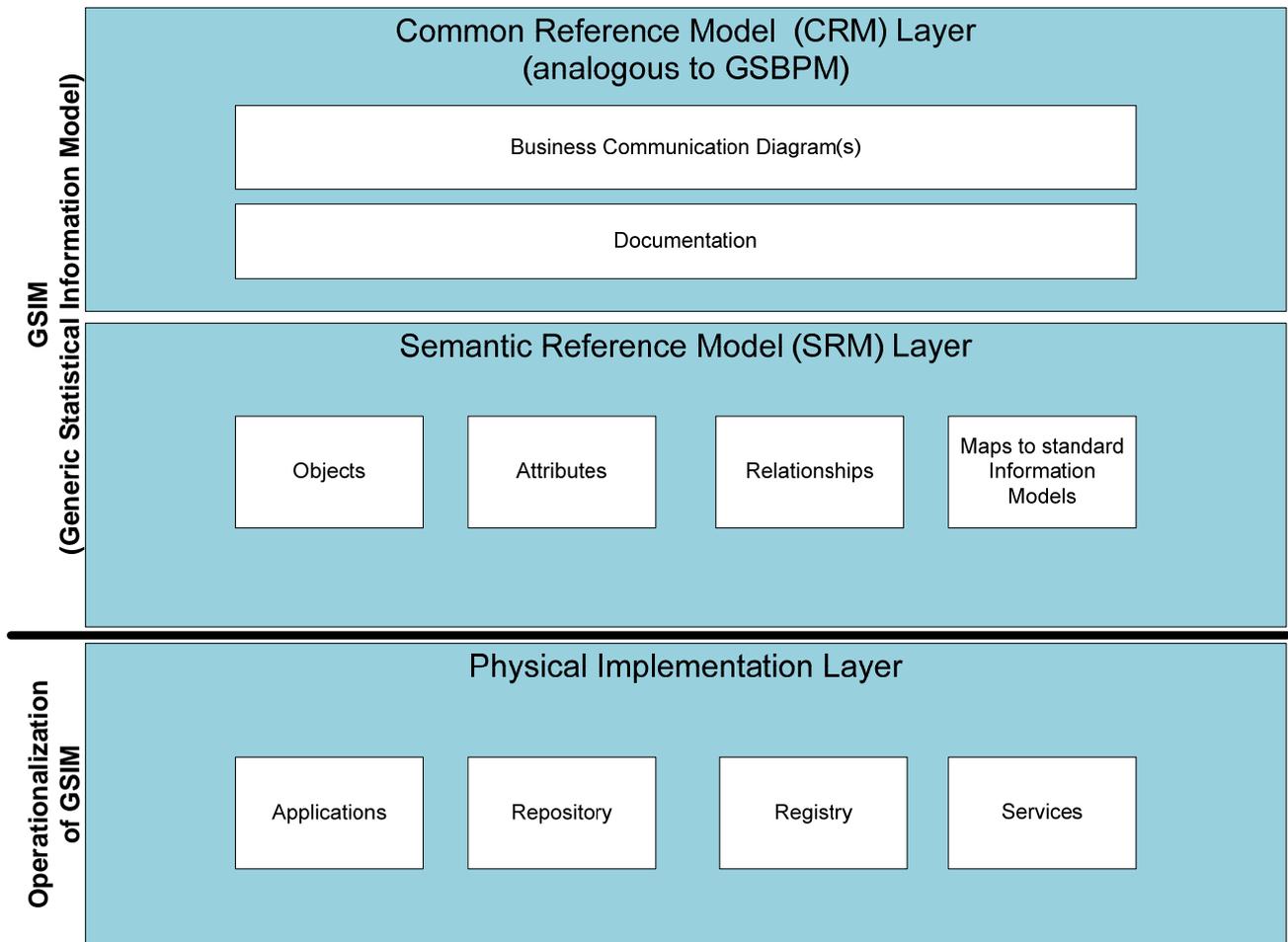
<sup>5</sup> <http://www.unece.org/stats/documents/ece/ces/ge.50/2010/crp.1.e.pdf>

<sup>6</sup> <http://www1.unece.org/stat/platform/display/msis/Statistical+Network>

would harness existing standards based representations wherever fit for purpose. SDMX and DDI-L (DDI-Lifecycle), in combination, were agreed as the best starting points in this regard.

15. It was also recognised that the development of GSIM needed to be co-ordinated with other collaborative forums and initiatives, and benefit from input from these. Examples included the UNECE Steering Group on Statistical Metadata (METIS) and the CORE ESSnet.

16. At that time, the relationship between GSIM and its operationalization was conceived as the following:



17. GSIM V0.1<sup>7</sup>, produced through the Statistical Network collaboration, was released for external review in June 2011. GSIM V0.1 was consistent with the above conceptualisation and included a detailed proposal for the design of the Common Reference Model.

18. Interested parties, such as members of METIS Steering Committee, and the CORE ESSnet collaboration, were approached directly for feedback on V0.1. In addition, GSIM was placed in the public domain via the UNECE wiki for review by other interested parties.

19. Twelve sets of comments were received, some of which were highly complementary, with suggestions for improvement. Others simply contained suggestions for improvement. Three sets of comments "re-envisioned" the approach to the top level of GSIM. The next step for the collaboration team was to engage with these three reviewers further.

<sup>7</sup> <http://www1.unece.org/stat/platform/display/metis/GSIM+Common+Reference+Model>

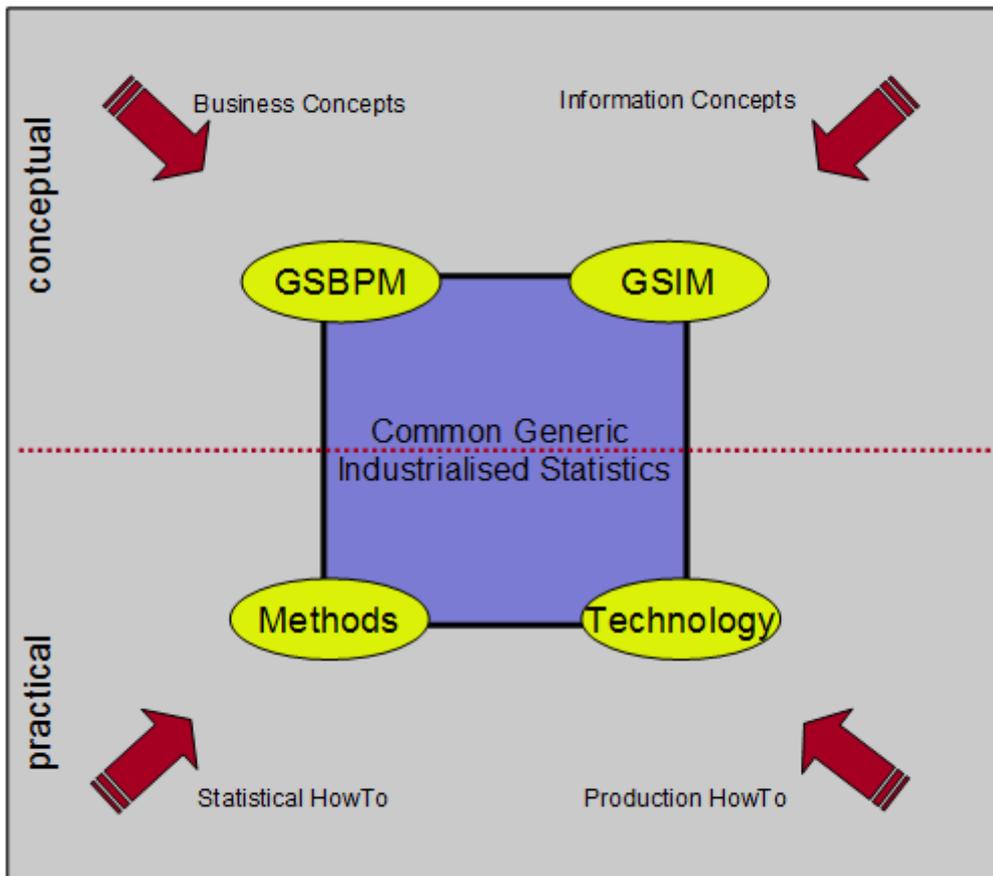
20. Several elements of feedback highlighted that it would broaden input, understanding and agreement regarding GSIM if active engagement during development extended beyond the six NSIs within the Statistical Network.

## B. Broadening interest in GSIM

21. In March 2011, during the early drafting of GSIM V0.1, the Strategic Vision from HLG-BAS was released.

22. The Vision referenced GSIM as a cornerstone for industrialising statistics, including through the following well known diagram.

Figure 1  
Industrialising statistics



23. In addition to being a cornerstone of the HLG-BAS strategic vision, by June 2010, when V0.1 from the Statistical Network was released for review, GSIM had also been referred to:

- as an element of the European Statistical System (ESS) Reference Enterprise Architecture<sup>8</sup> as enunciated by Eurostat; and
- by the SDMX/DDI Dialogue<sup>9</sup> as a common model onto which DDI and SDMX could be mapped, to identify overlaps, gaps and interfaces between these standard

24. While not direct participants in the Statistical Network collaboration, these additional stakeholders identified themselves as dependent on GSIM, and wished to ensure it was fit for their purposes. There was

<sup>8</sup> <http://www.cros-portal.eu/page/ess-reference-enterprise-architecture>

<sup>9</sup> <http://www1.unece.org/stat/platform/display/metis/SDMX+DDI+Dialogue+-+Overview+Page>

strong stakeholder interest, therefore, in a collaborative way forward for developing GSIM which engaged stakeholders beyond the Statistical Network more actively.

25. It was not surprising, also, that the various initiatives represented at the inaugural Workshop on Strategic Developments in Business Architecture<sup>10</sup>, which was convened by HLG-BAS in October/November 2011, emphasised the importance of developing GSIM quickly.

26. While there was uniform agreement at the Workshop that development of GSIM should be a top priority, particularly as other initiatives depended on it, participants did not have a consistent view of the expected scope, purpose and nature of GSIM.

### C. HLG-BAS initiates GSIM Sprints

27. HLG-BAS met immediately after the Workshop and requested that the Secretariat team *prepare proposals for a "sprint" session to accelerate the development of the GSIM*".

28. The GSIM Sprint Proposal<sup>11</sup> was subsequently agreed by HLG-BAS members in December 2011.

- GSIM Sprint 1 was to be hosted by the Statistical Office of the Republic of Slovenia (SORS) in Ljubljana, Slovenia from 20 February to 2 March 2012.
- GSIM Sprint 2 was to be hosted by the Statistical Office of the Republic of Korea (KOSTAT) in Daejeon, Korea from 16 April to 27 April 2012.

29. The term "Sprint" draws from the agile development process and has the following characteristics, including:

- Collaboration of multi-disciplinary experts
- A "time-boxed" period of work (i.e. to be undertaken within a set period of time), and
- A closely defined and agreed output, that is "potentially shippable" (i.e. stands in its own right, as a tested and useable output, even if it is recognised further work may be warranted to improve the product)

## III. GSIM Sprint 1

### A. Design of Sprint 1

30. The team for GSIM Sprint 1<sup>12</sup> comprised invited participants from national and international agencies. Their work was supported by a professional facilitator and secretariat staff.

31. Participants were selected to ensure a balance of backgrounds within the multidisciplinary team including

- business ("survey statistician") perspectives
- methodology perspectives
- informatics and enterprise architecture specialists who had a strategic, business oriented perspective (rather than overly technical or overly theoretical perspectives)

32. In setting the GSIM Sprint Objectives<sup>13</sup>, HLG-BAS tasked Sprint 1 participants with:

- ensuring strong agreement on the fundamental scope and purpose for GSIM;
- ensuring strong agreement on the value and use of GSIM;
- designing a readily understood high level view

<sup>10</sup> <http://www.unece.org/stats/documents/2011.10.hlgbas.html>

<sup>11</sup> <http://www1.unece.org/stat/platform/download/attachments/64880986/21+GSIM+Sprint+Proposal.doc?version=1>

<sup>12</sup> <http://www1.unece.org/stat/platform/download/attachments/64880941/Participants.docx?version=8>

<sup>13</sup>

<http://www1.unece.org/stat/platform/download/attachments/64880941/GSIM+Sprint+Participant+Information.docx?version=2>

- proposing a business case; and
- work planning for the further development of GSIM.

33. The program for Sprint 1 consisted of three broad phases:

Phase 1 (Days 1 to 3)	Reach an agreed definition of the problem, the requirements and the outputs Sprint 1 will deliver. This phase included ensuring a common understanding of the context, background and parameters for the task of developing the GSIM.
Phase 2 (Days 4 to 8)	Develop the agreed outputs
Phase 3 (Days 9 to 10)	"Package and present" the outputs to facilitate input from a wider and more diverse audience over subsequent weeks

34. Phase 2 of the Sprint was conducted utilising open space techniques<sup>14</sup>, designed to promote team cohesion, self-organisation and encourage creativity in designing outputs.

35. While Phase 1 was programmed in detail in advance, Phase 2 was not. It did, however, include some regular features such as “morning reflections” as well as “soap box” sessions which allowed two minutes for participants to introduce new ideas and proposed directions. Reflections and soap box sessions were often used to introduce the outcomes of discussions which had been held with colleagues “out of hours” (often over dinner and/or drinks) following the previous day’s formal conclusion.

36. A common design feature was working in smaller syndicate groups for prolonged periods to focus on different issues. From time to time the Team would convene as a whole to review the progress and proposals from syndicate groups and to ensure co-ordination of efforts and harmonisation of ideas across the Sprint Team as a whole.

37. A fundamental rule was that all participants needed to participate in all sessions. This meant that:

- all participants needed to stay in Slovenia for the full duration of the Sprint
- all participants needed to stay engaged
  - participants were advised in advance that other pressing work program items for their own agency would need to be deferred or delegated
    - where absolutely necessary, other pressing work was progressed out of hours rather than by “opting out of” certain sessions of the Sprint

38. All of the design features outlined above proved very valuable in practice. Participants became so enthusiastically engrossed in their work that a common role for the facilitator was to remind them to take breaks.

39. There were also three teleconferences during Sprint 1 with leading experts (“practitioners”) who had been unable to attend Sprint 1 in person. These proved extremely valuable in ensuring constructive questioning and challenging of key points of consensus that had been reached within the group. In some cases, this questioning led the Sprint Team to reconsider its conclusions, and undertake further analysis and testing. In other cases, the questions forced the Team to better explain its proposals, and the rationales behind them. For example, where the original explanation might have omitted points that had become assumed knowledge within the Team additional detail was added. Both sets of outcomes were invaluable in improving the quality of outputs from the Sprint.

40. Extending the set of disciplines within the team, perhaps even more than the extended number of agencies participating, led to early fundamental challenges to the design directions proposed previously within

<sup>14</sup> [http://en.wikipedia.org/wiki/Open-space\\_technology](http://en.wikipedia.org/wiki/Open-space_technology)

the Statistical Network collaboration related to GSIM. Strong perspectives from a number of methodologists, in particular, ensured speculative thinking and valuable divergence of opinions during the early days of the Sprint. The fact the Sprint launched with such diversity of opinion and experience made the extent of consensus reached at the end even more remarkable.

## B. Key GSIM related outputs from Sprint 1

41. This section summarises the outputs from Sprint 1 rather than reviewing and discussing the proposed model for GSIM in detail. The main output from GSIM Sprint 1 was GSIM V0.3<sup>15</sup>.

42. A key point emphasised by participants in Sprint 1 is that the model delivered was a work in progress, with many details still to be determined. By the time of the MSIS meeting on 21-23 May 2012, the work of Sprint 2 will have addressed many of these issues, partly based on feedback from external reviewers of the outputs from Sprint 1. This means it would be moot to discuss unresolved issues in detail in this paper. Key issues yet to be resolved at the end of Sprint 1 are set out in the documentation of GSIM V0.3 on the web.

43. The Executive Summary for GSIM V0.3 provides the following one sentence description of GSIM.

The Generic Statistical Information Model (GSIM) is a reference framework of information objects, which enables generic descriptions of data and metadata definition, management, and use throughout the statistical production process.

44. The documentation for V0.3 provides the following rationales for GSIM

*As a common reference framework for information objects, the GSIM will facilitate the modernisation of statistical production by improving communication at different levels:*

- *Between the different roles in statistical production (statisticians, methodologists and information technology experts);*
- *Between the statistical subject matter domains;*
- *Between statistical organisations at the national and international levels.*

*Implementation of the GSIM in combination with the GSBPM and common methods integrated in standardised components in a more modular production system will:*

- *Create an environment prepared for reuse and sharing of methods, components and processes*
- *Offer the opportunity to implement rule based process control, thus minimising human intervention in the production process.*
- *Generate economies of scale within and between statistical organisations through common development of tools and methods.*

45. The following points were agreed regarding the scope of GSIM:

- GSIM provides the information object framework for the complete GSBPM
- GSIM models information objects used in statistical production, but not at the technology level
- GSIM is aware of Environmental information (e.g. Finance, Legal, HR) but does not model it

46. The following purposes were agreed for GSIM

- Increase the standardization of information objects and flows
- Improve communication within the statistical community and with users and providers of official statistics
- Increase intra- and inter-agency reuse of data and ways of making statistics

<sup>15</sup>GSIM V0.2, as drafted during Sprint 1, was subject to some minor editorial corrections before being “published” for external review as GSIM V0.3. <http://www1.unece.org/stat/platform/display/metis/GSIM+Version+0.3>

- Improve collaboration
- Enable configurable, rule-based ways of making statistics
- Provide a basis for innovation

47. The success criteria agreed for GSIM were to establish a model which is:

- Communicable
- Stable
- Applicable
- Complete
- Inspirational
- Fit for use (Effective)
- Sustainable

48. Definitions of the terms used for criteria are provided in the V0.3 documentation.

49. Sprint 1 confirmed that two key principles in the development of the GSIM are to:

1. re-use previous work wherever relevant, and
2. ensure GSIM integrates with other relevant standards and international initiatives including GSBPM, CORE, SDMX and DDI

### **C. Key advances for GSIM V0.3 compared with GSIM V0.1**

50. The rationale, scope and purpose for GSIM arising from Sprint 1 appear highly consistent with those underpinning the earlier Statistical Network collaboration regarding GSIM. These outputs from Sprint 1, however, are more broadly agreed and more clearly expressed.

51. V0.3 provides greater clarity and detail in the following regards:

1. V0.3 simplifies the top level to four Groups, where V0.1 had 14 elements at Level 1. Each of the Groups in V0.3 is clearly differentiated from the others in terms of its character. This makes it easier for readers to achieve an overview of the model and its key elements.
2. V0.3 provides a stronger and clearer emphasis on the “Production” Group which provides the primary connection between information inputs (e.g. data and metadata, including business rules) and the statistical business processes they drive. This firm and practical connection between information and process is one of the critical requirements when it comes to industrialising statistical production. V0.1 referenced Events, Methods and Business Processes (symbolised by the GSBPM) but was not as precise and detailed in this regard.
3. V0.3 more clearly separates and organises the “Structural” Group, which includes datasets and data elements. This is, once again, intended to focus on the practical. Automated business processes interact primarily with structural content.
4. V0.3 describes interactions between the external environment and internal production processes through the Statistical Service Group. V0.1 provided a more elaborate Level 0 for context/environment. During the early development of GSIM, the fact that V0.3 provides less detail on context/environment may help maximise the attention focused on the definition and structuring of information required to industrialise internal production processes.

It is vitally important that internal production processes are responsive to external needs (of both users and providers) and to opportunities (e.g. to develop and promote appropriate statistical standards for application by others within a national statistical system), but this connection might be most productively explored once there is a firm proposal for the “core” of GSIM.

5. While still spanning fundamentally the combined scope, as noted in II.A. V0.3 no longer makes a distinction between a “Common Reference Model” (CRM) component and a “Semantic Reference Model” (SRM). It is intended to instead offer “views” on GSIM for specific audiences. For example, the “view” recommended for general business staff may be similar in content to the CRM proposed previously, where the view recommended for informatics experts and IT developers may be similar in content to the SRM proposed previously.

This change emphasises that the task is to create a single common model, not two separate components, even though different aspects of the underlying common model may be of interest to different audiences. The changed approach also makes it easy to define, from a common base, as many views as are required to meet the needs of different applications of the model, and of different audiences for the model.

52. Key drivers for these advances include:

1. The work on V0.3 was directly driven by the role that the HLG-BAS strategic vision foresees GSIM fulfilling in supporting industrialisation of statistical production.
  - While work on V0.1 recognised the vision as one driver, the work was initiated to address the particular need for a common information model which emerged from the Statistical Network. The Statistical Network was initiated to focus on collaboration *in practical small steps to industrialise methods and processes to quickly and effectively benefit all participating NSIs*. It was not primarily focused on developing an enduring, generic, framework for the international community of producers of official statistics as a whole.
2. A more diverse team for Sprint 1 (in terms of disciplines, in terms of countries participating and in terms of including international agencies as well as NSI).
3. Being informed by the experience with GSIM V0.1, and other initiatives, and the feedback from the international community on the outputs from those initiatives.

#### **D. External Feedback on GSIM V0.3**

53. HLG-BAS endorsed the outputs from GSIM Sprint 1 and the proposed way forward, including proceeding with Sprint 2 to be hosted by the Statistical Office of the Republic of Korea (KOSTAT) from 16 April to 27 April 2012.

54. After the outputs from Sprint 1 had been reviewed by HLG-BAS, they were released for wider external review on 9 March. By the closing date of 31 March, 15 sets of comments had been received. Many of these sets of comments were received from groups of interested experts who met to discuss V0.3. In a few cases, more than a dozen individuals contributed to a single set of comments.

55. GSIM V0.3 clearly reached a more extensive set of reviewers than V0.1 did when it was released in June 2011.

56. The majority of reviewers (there was one notable exception) supported the broad direction taken by V0.3

57. A few reviewers expressed disappointment that Sprint 1 had not progressed further in defining the proposed model.

58. All reviewers provided constructive proposals for improvement. Many suggestions related to improving the quality of the documentation including improving the choice and definition of terms. There were also many suggestions related more fundamentally to the design of the model overall, and to the design of certain groups and categories within it.

## IV. Future directions for GSIM

### A. GSIM Sprint 2

59. HLG-BAS have set the following objectives for GSIM Sprint 2:

- Review and address stakeholder feedback on GSIM v0.3
- Identify defining characteristics of objects (i.e. definitions, attributes and relationships)
- Elaborate and demonstrate business driven use cases to illustrate application
- Identify relationships with GSBPM and CORE
- Map GSIM to DDI and SDMX and identify gaps in the standards
- Develop approach to gaining broad stakeholder endorsement of GSIM post Sprint 2

60. The focus for Sprint 2 will be somewhat less “open ended” than Sprint 1, although innovative thinking relating to modelling statistical information based on business needs will remain essential. For this reason, and others, assembling exactly the same set of participants as for Sprint 1 would be inappropriate.

61. On the other hand, some continuity is required so that the basis for decisions and thinking in Sprint 1 can be conveyed accurately and consistently during Sprint 2. Slightly less than half of the participants selected for Sprint 2 participated in Sprint 1.

62. Various means, primarily teleconferences, will also be used to ensure those participants in Sprint 1 who will not be attending in Sprint 2 know the directions being considered, and are able to influence them. While it is not yet known what further development steps may be required beyond Sprint 2, the intent is to establish and maintain a growing set of “Sprinters” who have been engaged in the development of GSIM and will remain engaged in its development (including consultations and testing); even if they do not attend every Sprint session.

63. A balance of disciplines (methodologists, informatics experts, enterprise architects, business/“statistics” experts) is being maintained for Sprint 2.

64. A similar structural design and set of facilitation techniques will be applied for Sprint 2. The approach will incorporate a number of refinements based on experience and based on some differences in objectives and deliverables between the two sprints.

### B. Mapping GSIM to implementation standards (primarily SDMX and DDI)

65. When HLG-BAS, met as a group to review outcomes from Sprint 1, they were particularly keen for Sprint 2 to start looking at how well existing technical standards (particularly SDMX and DDI) might be capable of supporting implementation of GSIM in practice. HLG-BAS requested an estimate of whether the existing technical standards cover, eg, closer to 95% or 50% of what implementation of GSIM in practice might require.

66. This estimate will assist in determining how much work would be required to establish a recommended, standards aligned, means of implementing the conceptual model in practice. For example, in the unlikely event that Sprint 2 concluded that currently SDMX and DDI would cover only around 50% of GSIM implementation, then after specification of GSIM was complete there would need to be much more work to allow the whole reference model to be implemented in practice on a standards aligned basis.

67. HLG-BAS will report progress on GSIM, and the outlook ahead, to the Conference of European Statisticians in June 2012. An estimate of the extent of work required before GSIM can be readily applied by agencies in practice is an important element of that reporting.

68. HLG-BAS also wanted to be reassured that any differences between the GSIM conceptual model and the models underpinning DDI and SDMX were based on proven business needs. Where there are potential areas of connection between GSIM and either of the two standards it is important to ensure incompatibilities do not arise through “accident of separate design”.

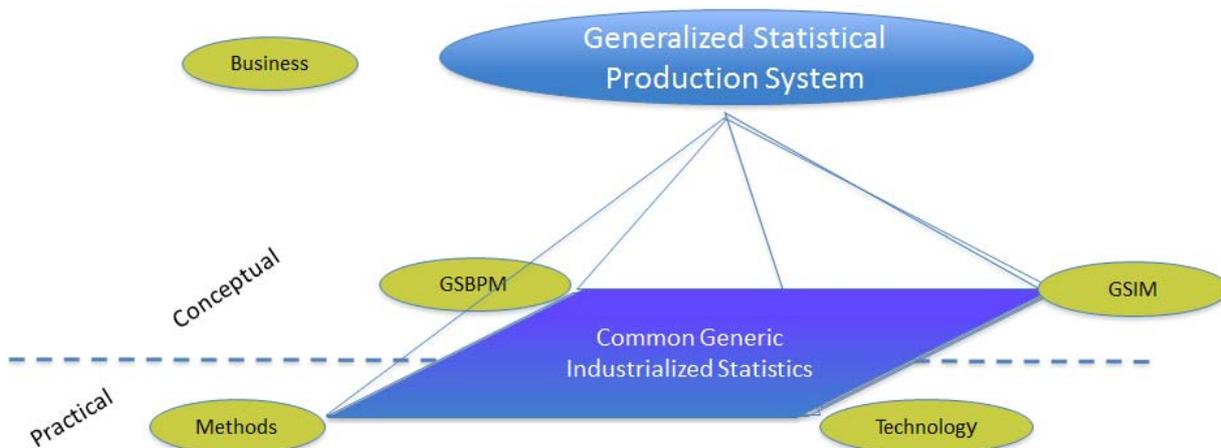
69. Finally, if there are gaps and misalignments identified in regard to the existing standards, then another purpose for having Sprint 2 broadly identify these is that it allows the SDMX and DDI development and governance processes to start addressing relevant issues at the earliest possible opportunity.

## V. “Grand Unification”

### A. How the concept of “Grand Unification” arose during Sprint 1

70. During Sprint 1, when assessing the benefits to be delivered by GSIM and how it would deliver them, it was recognised that many of the most important benefits would arise more generally through “Industrializing Statistics” rather than solely from delivery of GSIM. In these instances agreed definition of GSIM was seen as a very necessary, but not sufficient, step toward realising these benefits.

71. Participants noted that in current statistical practice, statistical subject-matter experts (“business”), methodologists and information technologists often do not work together as effectively as they could to create production solutions. This situation results in cost-demanding as well as time-consuming processes in design, implementation and maintenance of statistical systems. Grand Unification envisages a new approach that brings together these three disciplines, using standards from the GSBPM and the GSIM to make statistics.



72. GSBPM and GSIM will provide reference frameworks from the business perspective, and enable cross-disciplinary communication and understanding through common terminology. In practice, this facilitates standardised designs and solutions that are reusable to a much greater extent. However, GSBPM and GSIM will not be sufficient on their own. By nature of reference frameworks, the GSBPM and the GSIM will not provide detailed guidelines to users on how to apply them when seeking to define reusable “plug and play” components.

73. A reusable “plug and play” component might bring together:

- a commonly defined business process (possibly corresponding to a Level 2 sub-process within GSBPM or maybe at even lower level)
- a commonly defined method (or selection of methods) suitable for implementing that process
- a reusable IT solution/service suitable for implementing the selected method

74. To achieve the Grand Unification goals, it will be necessary for organisations to:

- assemble a set of necessary production components and a library of associated information objects

- the set should include internationally shared components
- put in place the appropriate infrastructure by which statistical solutions can be designed and created based on reusable components
- educate and train the staff all across the organisation to foster a culture of co-operation that is necessary to the new approach

75. As visualised in the diagram, Grand Unification will allow Business (statistical subject matter experts) to be at the apex of a pyramid founded on harmonised methods and standardised technology, as well as on GSBPM and GSIM as conceptual frameworks. By working with methodologists to select and re-use the appropriate components (“assembling to order” in accordance with the particular statistics they need to produce) statistical subject matter statisticians will be working with a Generalized Statistical Production System that is capable of meeting their specific needs.

76. “Grand Unification” can be seen as one example of how GSIM might be harnessed to contribute more broadly to common reference architecture within, and across, statistical agencies.

## **B. Response to the concept of “Grand Unification”**

77. When HLG-BAS reviewed the outputs from Sprint 1 they agreed the concept was very relevant in regard to how GSBPM and GSIM would be applied in practice to support realisation of the business benefits from industrializing statistical production. They also agreed that deciding how to progress the “Grand Unification” concept should be distinct from the further work on GSIM itself that would be undertaken by Sprint 2.

78. A second Workshop on Strategic Developments in Business Architecture, once again bringing together representatives from each member (including MSIS) of the Inventory of International Groups<sup>16</sup>, is planned for 7-8 November 2012. HLG-BAS suggested that Workshop might be well placed to review how to progress realisation of the “Grand Unification” concept and the contribution that different groups might make to this.

79. Feedback from several other external reviewers of the Sprint 1 outputs suggested that, while the concept may be important, the “branding” of the concept as “Grand Unification” was inappropriate and potentially counterproductive.

## **VI. Conclusions (so far)**

### **A. Some lessons learned**

80. As noted in the Introduction, the GSIM Sprint process, with a multidisciplinary, multiagency team of experts collocated to work intensively for two weeks, represents an innovative approach to developing architectural artefacts. Such artefacts are designed to guide, and to be harnessed across, the “industry” of producers of official statistics.

81. A secondary output that HLG-BAS sought from Sprint 1 was the impression participants had of how effective the “sprint” approach had been, and whether they thought it might be useful applied to other collaborative developments.

82. The report to HLG-BAS from Sprint 1 noted a consensus from participants that the process was very useful in accelerating progress compared to traditional approaches of remote collaboration and occasional meetings. Due to work commitments and/or personal commitments, however, it was often very difficult for participants to be available for the two week period.

---

<sup>16</sup> <http://www1.unece.org/stat/platform/display/msis/Inventory+of+International+Groups>

83. A number of Sprint 1 participants who had also participated in the earlier work on GSIM within the Statistical Network or in the CORE initiative, suggested those collaborations would have progressed with a clearer shared understanding had they commenced with a similarly intensive content related workshop. This would have improved planning and team work, While more expensive in the short term, had these collaborations commenced with a “sprint” then they may then have progressed more rapidly, and produced higher quality, better integrated deliverables, once they reverted to the usual mode of operating on a distributed basis. It was suggested, therefore, that appropriately timed and designed “sprints” might also lead to overall improved returns on investment for other collaborations.

84. The effectiveness of the design of the Sprint 1 process, based on the outcomes achieved by putting it into practice, was highly regarded by Sprint participants in their evaluations. There was strong support for such processes being professionally facilitated. This allows participants to focus their efforts on content, while someone else is monitoring and managing the sprint process to ensure it remains on target to deliver the outputs requested by the sponsors. This can be seen as analogous to the role of a “ScumMaster”<sup>17</sup> in agile software development.

85. The following criteria for selecting a facilitator appeared to work well.

- Well proven facilitation skills, including the flexibility required to facilitate open ended processes where it is not possible to plan, and prepare for, every session in detail in advance
- Some understanding and experience, at a strategic rather than technical level, related to subject matter content to be addressed by the Sprint
- No vested interest in leading the group toward one particular answer

86. One issue identified was that some potential participants were not able to join the sprint for financial reasons, or due to pressure of existing work. Whilst this was not felt to be a major issue for Sprint 1, there is a risk that sprint approaches could exclude experts from organisations facing resource restrictions. A partial mitigation for this risk would be the use of some sort of funding mechanism supported by a benefactor, or a group of contributing organisations.

87. A further issue became apparent when the outputs from the Sprint were circulated for external review. GSIM Sprint 1 spent most of its time brainstorming, debating and harmonising the content that should go into the deliverables rather than “crafting” the documentation itself. This is somewhat counter to common practice in other collaborations. Other approaches to collaboration may not be so conducive for fundamental brainstorming, debating and harmonising, but more time may be committed to documentation of conclusions and proposals. The draft documentation may undergo several rounds of review, mark-up and refinement from members of the collaboration team before it is released for external review. One focus of the refinement process is to ensure the final documentation is easier to understand for a broader audience.

88. Some reviewers, who were used to the traditional approach to preparing such documentation, considered the documentation of the major deliverable from Sprint 1 (GSIM V0.3) appeared “rushed” and unclear. It was certainly the case that, in retrospect, some important rationales and clarifications which were discussed within Sprint 1 were not clearly expressed in the final documentation.

89. Lessons in this regard include:

- Where possible, allow a little more time for reviewing and refining documentation before it is “published” for wider review. Rather than consuming too much time during the sprint, review might be done very shortly after the sprint, using more traditional methods such as circulating draft documentation for prompt editorial input and possibly holding a teleconference to review comments.
  - There was pressure to produce and circulate the Sprint 1 documentation particularly quickly in early March so that it could be reviewed externally prior to Sprint 2, which commences in mid-April.

---

<sup>17</sup> [http://en.wikipedia.org/wiki/Scrum\\_\(development\)](http://en.wikipedia.org/wiki/Scrum_(development))

- If a particular deliverable is not yet far enough through the development process, or otherwise not meaningful for review by a general business audience then be very clear in determining which audience should review the deliverable and in stating which aspects they are being asked to review.
- If agile collaborations become a norm, there may need to be evolution in the expectations of reviewers regarding what they will receive to review, and what aspects they should focus on. For example, in early drafts the focus for reviewer should be the fundamental design and proposed directions rather than details of expression.

## **B. Outlook**

90. Work on development of GSIM has been gaining momentum; at the same time as it has been gaining urgency. More and more agencies and initiatives have identified themselves as being dependent on this cornerstone.

91. HLG-BAS members are fully aware of these dependencies, and since November 2011 have allocated a great deal of their focus to guiding and accelerating progress in developing GSIM.

92. GSIM V0.3, delivered in March 2012, represents a significant advance on V0.1 but should still be considered a work in progress. Given the level of on-going sponsorship and focus from HLG-BAS, it appears likely a version of GSIM that could be applied in practice (perhaps designated “GSIM V1.0”) will be available before the end of 2012.

## **C. Engagement by MSIS community**

93. A number of members of the MSIS community, including members of the MSIS Steering Committee, have already had important roles in shaping the high level definition of GSIM.

94. Many of the future uses of GSIM as a conceptual framework, and as an architectural artefact, will be led by members of the MSIS community. It will be crucial to ensure the design and documentation of GSIM is fit for these purposes. Active engagement from the MSIS community in the further development and review of GSIM will have an important role in ensuring this.