Distr.
GENERAL

WP.20
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 (iv): Collaboration

Platform for International Collaboration –Preliminary thoughts from the ABS

Invited Paper

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

I. Introduction

1. The Strategic Vision[1] of the High-Level Group for Strategic Developments in Business Architecture in Statistics (HLG-BAS)[2] envisages establishment of industrial standards that will provide a foundation for developing and sharing "the means of production" among producers of statistics, and potentially create a market of commercial interest which is of benefit to producers. [Para 21]

2. For the purposes of this paper, "means of production" is assumed to refer to components that implement specific methods to perform a specific function (or a specific ***process step***[3]) within a statistical business process.

3. While, traditionally, a few large scale applications have been successfully shared (e.g. Blaise and PC Axis), the Strategic Vision suggests a more ***modular*** approach in future [Para 36b]. This paper, therefore, explores the necessary and sufficient conditions which would enable ***Assemble To Order (ATO) statistical production processes*** based on flexible assembly of reusable and sharable metadata driven components.

4. The paper proposes these conditions should be addressed via defining, agreeing and applying common architectural standards. It then suggests a way forward in order to realise the required combination of conditions in practice.

---

[1] http://www1.unece.org/stat/platform/display/hlgbas/Strategic+Vision
[2] http://www1.unece.org/stat/platform/display/hlgbas/High-Level+Group+for+Strategic+Developments+in+Business+Architecture+in+Statistics+%28HLG-BAS%29
[3] This paper uses the terms "Process Step" and "Component" in the same way as the paper *Standard Process Steps in Statistics* presented by Statistics Netherlands at MSIS in 2011.

II. Paradigms to support sharing "means of production"

A. The dominant paradigm – "Stand-alone" development

5. The Invited Paper *Co-operation Models for Software Development*[4] presented to MSIS 2011 by UNIDO confirms the most prevalent paradigm currently is for one agency to develop software and then make it available, either freely or for a fee, for other agencies to share. PC-Axis, Blaise and OECD.Stat are among the examples listed.

6. Each of these examples might be considered quite large applications (or suites of applications). They do not make it simple for another agency to choose to harness just one or two modular functions. They might loosely be described as "large footprint" applications.

7. While further collaborative development might follow, this approach allows developers to create an initial application focused on meeting a single set of practical business requirements. This is in contrast with the "co-design and co-develop" paradigm discussed in II.B. where initial requirements, specifications and designs are developed through international collaboration and negotiation.

8. Other agencies then have a tangible product, with a tangible architecture, which they can then review to assess

- how well the initial application is performing for the agency that developed it
- what would be required to allow the application to function effectively within their own business and IT environments
- whether the application is "best of breed" compared with alternatives being used by other agencies to address the same set of business needs

9. Along with these important practical advantages, a number of challenges may arise under this paradigm.

10. For example, although another agency may find it more cost effective overall to adopt and integrate the solution which has been offered for sharing, rather than building its own alternative application, the shared application will seldom be an ideal "fit" for the other agency's business and technical environments. Particularly if the shared application has a "large footprint", complex "plumbing" (numerous and complex "technical process steps" as per III.B.) will be required to integrate business use of the shared application with

- other business ***process steps***
- the components/applications which support the other business ***process steps***
- corporate stores of statistical information
- the broader IT environment and platforms used by the agency

11. This can lead to high costs in integrating new releases of the shared application. It also increases the cost and complexity associated with choosing in future to use an alternative application/component to support some or all of the business ***process steps*** previously supported by the shared application. In this way it may act as a brake on innovation and agility related to business processes and the IT environment which supports them.

12. For the agency that originally developed the application there can be complications also. Their development program for the application typically becomes constrained to some degree by the product development and support requirements of the other agencies that are now using it. Anecdotally it appears common for the agencies that have developed and shared such applications to express some concern that they are pressured to fulfil roles that software vendors would otherwise have performed.

13. The above description of challenges in no way suggests the "stand alone" development paradigm has been a poor choice. It appears to have been the best choice available historically. It may remain the best choice in many cases.

---

[4] http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.50/2011/wp.30.e.pdf

B. The emerging paradigm – "Co-design and co-develop"

14. The paper *Co-operation Models for Software Development* refers to this paradigm also. The two examples quoted are the Statistics Open Standards (SOS) Group[5] of Nordic countries and the informal CSTAT group Statistical Network[6]. The paper notes that one aim of such groups is to facilitate cooperation in the field of development of software.

15. Some ESSnet collaborations, especially where multiple agencies collaborate in regard to a specific work package, might also be seen as examples of this approach.

16. Such groups typically include a focus on development of new applications/components, which are designed to be shared, rather than collaborative evolution of applications that were initially developed on a "stand alone" basis.

17. It is notable that SOS, the Statistical Network and the ESS (European Statistical System) focus on harnessing standards to underpin designs and to facilitate shared use of the outputs.

18. Without reference to standards
- debates about design would be expected to be more protracted
    - each agency might advocate a design that would allow the product to best integrate within its own specific business and IT environment
- there would be less certainty that the results of one collaborative effort within the group would be able to integrate with the results of another collaborative effort.

19. Collaborations of this type have not been prominent for as long. It appears they have not (at least yet) produced major outputs which have been widely shared in practice. Assessment of relative strengths and weaknesses, therefore, is more speculative.

20. Having assembled requirements from more diverse sources, and designed the application/component for sharing from the outset, it should be expected the resulting product will more readily meet diverse needs and will more readily be integrated in different agencies' specific business and IT environments.

21. While there are overhead costs associated with collaboration processes, and with integrating the outputs in local environments, the total cost per agency should be considerably lower than if each agency independently commissioned a stand-alone development. The quality of the product may also be higher, given input from experienced designers, business experts and testers from beyond the local agency.

22. Alignment with standards notwithstanding, the element of co-design can make the requirements gathering, specification assembly and design processes considerably slower and more complicated. While engagement of multiple designers brings new perspectives and can result in creative, better rounded, syntheses, there is also some risk of the design process falling back to design by committee[7]. The latter could result in design compromises based on the lowest common denominator[8], resulting in products which lack coherence and tend toward mediocrity.

23. Also, in several cases, there is anecdotal evidence of a lack of confidence from individual agencies about relying on a particular "co-design and co-develop" collaboration to result in an output that is fit for that agency's purposes within the timeframe the agency needs it. This has resulted in some level of "stand alone" design and development continuing to occur in parallel with certain "co-design and co-develop" collaborations.

---

[5] http://www1.unece.org/stat/platform/display/msis/Statistical+Open+Standards+Group
[6] http://www1.unece.org/stat/platform/display/msis/Statistical+Network
[7] http://en.wikipedia.org/wiki/Design_by_committee
[8] http://en.wikipedia.org/wiki/Lowest_common_denominator#Non-mathematical_usage

While concerns about absolute reliance on such collaborations may be understandable, given this is a relatively unproven method, the result is that resources and focus are diverted away from the collaborative effort.

C. Platform based development

24. This paradigm is illustrated well by the boom in development, and marketing, of "apps" for smart phones. This is especially true in the case of Apple's iOS platform associated with iPhones and the case of the Google led Open Handset Alliance's Android OS platform which has been implemented by several producers of smart phones (e.g. Motorola, Samsung, Sony, LG, HTC). iOS and Android OS have become examples of "industry platforms" in an industry which is likely to have many connections with both collection and dissemination of statistics in future.

25. In order to support common "business processes" for users (such as managing contact details for friends and associates, managing multiple modes of communication, managing tasks, defining and managing lists, managing diary commitments, planning and navigating journeys, discovering and harnessing services which are available) there is typically a large choice of "apps" available on a free or charged basis. Users can choose an app which suits their current needs and preferences. Information is available to guide selection from the array of choices. Such information includes number of downloads, user ratings and documentation of features.

26. If a user has been harnessing one app for a particular purpose (e.g. Managing tasks), but then elects to use a different app for that purpose in future, the transition is typically almost seamless. They do not need to change the way they use other apps for related, but separate, purposes (e.g. calendar for due dates, contact list for task assignment).

27. In a common language sense, rather than a technical IT sense, such platforms can be seen as supporting selection and use of apps on a ***plug and play*** basis.

28. Enablers for this include

- The apps are ***modular*** in nature. Each app supports a small number of closely related, functions. Their "footprint" (see II.A.) is dainty and they are designed to "dance" with whatever other apps the user may choose to select for related purposes.

- The apps harness a common information model, defined by the platform, for basic "objects" such as current location, details of contacts, tasks, dates/calendar etc. Each app does not require input of this information in a specialised format, nor does it output such information in a specialised format. This can be viewed as a form of ***connectivity*** where information based on a common model can flow into, and out of, a series of different apps which are being used to perform a series of functions.
    - Both platforms provide a range of APIs that developers of apps can harness to interact with the common information model. Such APIs save time for developers and make the common information model easy to use in practice in a consistent and correct manner.

29. By selecting apps from established "stores" such as Apple's App Store or Google Play (which was known as "Android Market" until 6 March 2012) users are assured that some level of independent checking has been undertaken of the behaviour of the app and its suitability for the platform. Consistent with support for **plug and play**, at both the end user and the technical level, a strong emphasis is placed on consistency of interfaces, both in terms of the way the app interfaces with users and with the operating system.

30. Advantages for developers include
- Ability to design for a standard, well defined, and well supported, operating system (iOS or Android OS).

    - While active "install bases" for iOS and Android OS can be estimated different ways[9], an app developed for either OS has, at a minimum, a current user market in the order of a hundred million
- Ability to use common, well known and supported, languages (e.g. C or Java) for developing applications
- Availability of IDEs (Integrated Development Environments) including Xcode (which also supports development of applications for OS X used by Apple Macs) and Eclipse
- Availability of project templates and widget libraries that help speed development.
- Ability to use simulators of iOS or Android OS environments while developing applications using a PC or Mac
- Availability of platform specific communities to assist in testing newly developed apps

31. This paradigm is highly successful for smart phones and related devices (e.g. tablets). There are, however, a number of fundamental barriers to applying exactly the same approach to the industry of official statistics.

- iOS and Android typically run on a small range of devices which have been specifically designed to support the platform. There is currently, however, great diversity in regard to devices, operating systems and platforms used to support statistical production process internationally. Some, but not all, of this diversity is intrinsic to differences in what is required for different phases and sub-processes of the GSBPM and what is required for different forms of statistical activity (e.g. collecting data via surveying individual respondents versus via harnessing "big data" sources).

  For some purposes, for example, hand held devices such as smart phones might be perfect. For some other purposes within statistical production, however, they are manifestly not fit for purpose.

  The aim for statistical production, therefore, is to establish support for ***plug and play*** and ***connectivity*** without being so prescriptive in regard to device or platform.

- The "common information model" required to support statistical production is much more extensive, complex and specialised than the basic information model supported by the iOS and Android platforms.

- Where a well designed app for iOS or Android can lead to hundreds of thousands of installations, the market for apps to support production of official statistics is much more limited. This would be likely to act as a disincentive for many commercial and amateur developers.

  That said, existence of an Internet Statistical Apps Store (ISAS), consisting of applications that conform with relevant ***plug and play*** and ***connectivity*** standards, would offer third party developers a more extensive market than individual NSIs – with their specific IT environment requirements – do currently.

  The fact that the standards for ***plug and play*** and ***connectivity*** would not be specific to a particular device or platform (i.e. they would be "technology agnostic") means it is expected that <u>some</u> developers/vendors could make modular components available simply by providing wrappers/adapters[10] which conformed with the required architectural standards rather than needing to produce and maintain a separate version of their underlying code.

D. Common architectural standards based development

32. This is the paradigm explored in further depth in the remainder of this paper.

---

[9] http://blog.flurry.com/bid/79595/Kaboom-iOS-and-Android-International-Installed-Base-Expansion
[10] http://en.wikipedia.org/wiki/Adapter_pattern

33. It envisages architectural standards being defined for ***modularity***, ***plug and play, connectivity*** and ***metadata driven processes*** in sufficient practical detail to allow

- developers to know exactly what they must do (and must not do) in order to comply with those standards
- compliance (or non-compliance) of apps with the standards to be determined objectively

34. These architectural standards, however, would not unduly constrain the developer's opportunity to innovate in regard to the design of the app itself.

35. This paradigm can be seen as seeking to combine strengths from the previous three.

- It aims to emulate the success of platform based development (II.C.) without requiring a common platform.

- Compared with existing "co-design and co-development" initiatives (II.B.), this approach aims to be more concrete and application oriented in regard to standards. This allows design and development of each app to be undertaken on something closer to a "stand alone" basis (II.A.) (although there could still be collaboration around business requirements and functional specifications). At the same time, there is a much reduced level of this risk that the app will prove difficult and expensive to integrate with other agencies' business and IT environments.

36. This approach can be seen as very strongly aligned with the vision of HLG-BAS. In advocating collaboration and sharing related to "means of production", the vision places particularly strong emphasises on the need to align architecture to allow this to be achieved in practice. Specifically in regard to alignment of architecture the vision provides the following call to action

> *For this all to become reality, things will have to change.*

37. In regard to individual initiatives, the ESS initiatives CORA (Common Reference Architecture)[11] and CORE (Common Reference Environment)[12] can be seen as addressing enablers for this approach, particularly in regard to ***modularity*** and ***plug and play*** respectively. In regard to ***modularity,*** this can be considered both from a business perspective (e.g. definition of ***process steps*** and the business level components that support them) and from a technical perspective. Some important reusable technical services (e.g. checking whether a user is authorised to perform the requested operation, storing the results of an operation in a database) might be too low level to be apparent in the business view of ***modularity***. In this regard the CORA framework, including the service design constraints it proposes, might be seen as focusing primarily on ***modularity*** from a technical perspective.

38. Development of GSIM (Generic Statistical Information Model[13]), and its relationship with SDMX and DDI, can also be seen as addressing enablers for this approach, particularly in regard to ***connectivity*** but also in terms of how statistical information objects, including business rules, can be associated with business processes in practical terms to achieve ***metadata driven processes*** *("configurable, rule based" production processes).*

39. Architectural design requirements that embody ***modularity*** *(from business and technical perspectives)*, ***plug and play, connectivity*** and ***metadata driven processes*** are also in the process of being further defined by the ABS to allow these characteristics to underpin architectural design of the ABS 2017 business transformation program.

40. While there is a range of work underway currently to enable this approach to be applied in practice, there are no known existing large scale implementations which can be referenced as case study examples.

---

[11] http://www1.unece.org/stat/platform/display/msis/CORA
[12] http://www1.unece.org/stat/platform/display/msis/CORE
[13] http://www1.unece.org/stat/platform/display/metis/Generic+Statistical+Information+Model+(GSIM)

41. Section III of this paper, therefore, illustrates what is proposed from first principles. It illustrates the practical business value added through architectural requirements related to ***plug and play***, ***modularity***, ***connectivity*** and ***metadata driven processes***. Section III also illustrates how this supports reusable ***Assemble To Order (ATO) statistical production processes*** and an ATO business model more generally.

42. Section IV then reviews in more detail the extent to which enablers for this approach have already been delivered by the CORA and CORE ESSnets and by the development of GSIM.

43. Section V then proposes practical next steps to realise ATO statistical production processes underpinned by appropriate architectural standards.

III. Statistical Production Systems: A "first principles" perspective

A. Broad conceptualisation of statistical production systems

44. A statistical production system can be defined as a grouping (or "assembly") of relevant statistical processes for the purpose of producing specified statistical outputs.

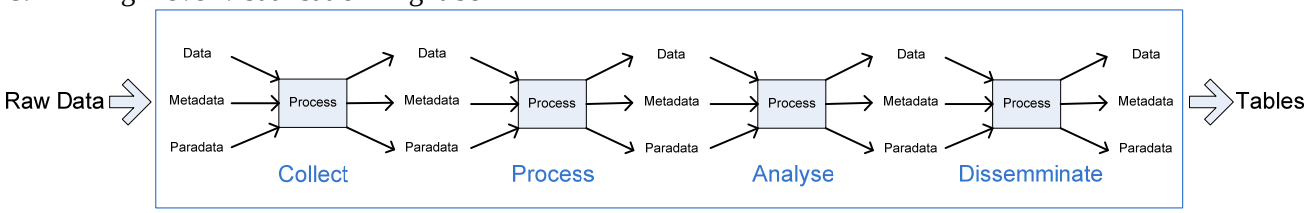
45. The emphasis here is on "system" at the level of the business processes required to transform raw material (inputs) to the required outputs. The term is therefore being used in a broader sense than simply referring to the assembly of IT solutions which support production.

46. Based on systems theory[14], a system can be defined using the definition of desired outputs to understand what inputs are necessary. The following questions can be used for this approach:

1) What essential outputs must the system produce in order to satisfy the system users' requirements?
2) What transformations are necessary to produce these outputs?
3) What inputs are necessary for these transformations to produce the desired outputs?
4) What types of information does the system need to use and retain?

47. Much of the focus in designing and performing statistical production focuses on the second point. This includes establishing what methods (methodology) will be used to achieve the required transformations and how these methods will be applied to inputs in practice, often using appropriate IT technology.

48. A high level visualisation might be



49. As illustrated in GSIM V0.3[15], the same ITO model can apply at many levels. For example, it can be applied to a particular sub-process within one of the four phases of the GSBPM illustrated above, or to a particular ***process step*** within a particular sub-process.

50. Particular ***process steps*** may be performed by staff rather than being automated through use of IT. In some of these cases staff may use

- IT generated outputs (e.g. reports) to assist them in deciding what action to take, and/or
- IT tools to apply their actions. (For example, the staff member may decide whether an edit should be applied, and what that edit should be, and then use an IT tool to update the database accordingly.)

[14] http://en.wikibooks.org/wiki/Systems_Theory/Inputs-Outputs
[15] http://www1.unece.org/stat/platform/display/metis/GSIM+Version+0.3

B. Transformations that add statistical value versus technical process steps

51. Ideally each of the individual transformation steps in a statistical production system should be included because that step adds statistical value to the raw material. The sub-processes defined in the GSBPM, such as imputing missing values, are examples of such transformations.

52. In reality, however, many steps are currently performed for technical reasons rather than statistical reasons. For example

- Translating data and/or metadata from one format to another (e.g. XML to CSV) so it can be used as an input to the software that will be used to perform the next statistical transformation.
- Updating a localised repository or configuration file with information that has already been defined elsewhere.
    - For example, the hierarchical structure which should be used for aggregation might have been established when a classification was first defined. Details of the aggregation structure might still, however, need to be entered separately in different places in different formats to drive IT systems for editing/validating, for estimation and for time series analysis.

53. Statistical production systems can be streamlined, and made less susceptible to human error, if

- the number of ***process steps*** required for "technical reasons", rather than to add statistical value, can be minimised, and
- any remaining ***process steps*** required for "technical reasons" are fully automated and ***metadata driven***.

C. Practical value added through "plug and play", "modularity" and "connectivity"

54. The term ***plug and play*** was used in a common language sense in II.C. In the context of a statistical production system, the term refers to the ability of components to be "plugged in" (and "unplugged") without requiring either of the following

- addition of numerous and/or complex technical process steps
- changes in selection or design of other statistical components,

in order for them to integrate with other statistical process steps within the production system.

55. The ability of components to be ***metadata driven*** assists in achieving ***plug and play***...Where components are not ***metadata driven*** they are likely to require separate configuration steps (on a one off or on-going basis) in order for them to integrate within existing statistical production processes.

56. ***Plug and play*** design, therefore, contributes to streamlining the operation of statistical production systems, hence reducing costs, and makes them less susceptible to human error.

57. In order for ***plug and play*** to be achieved in practice, appropriate ***modularity*** (described as "dainty footprints" in II.C.) is essential. If a component/application is only capable of provide a large set of integrated functionality, rather than being able to offer individual functions, or small sets of closely related functions, in a ***modular*** fashion then

- "plugging it in" will require "unplugging" certain other statistical components which were being used previously (possibly with high levels of fitness for purpose), and
- "unplugging it" will require finding one or more applications, or sets of components, which can replace the large, customised "footprint" of the application which is being replaced

58. As noted in II.C, in some cases ***modularity*** may be achieved simply by developing and offering appropriate wrappers/adapters[16] rather than needing to fundamentally redevelop underlying larger scale applications

59. Streamlining via ***process steps*** that ***plug and play*** can be contrasted with typical practice currently where statistical processes are not interoperable and human intervention is required.

60. One example of current problems mentioned in III.B was the need to separately define information about a hierarchical structure to be used for aggregation for the purposes of editing/validating (e.g. checking components sum to totals on forms submitted by respondents), for estimation and for time series analysis.

61. This issue is addressed through ***connectivity*.** The idea of practical application of a common information model is described in II.C. Where ***plug and play*** tends to focus on how components relate to the ***process steps*** "either side of them", ***connectivity*** enables common use, and reuse, of relevant information inputs by ***process steps*** that are positioned far apart in the statistical production system.

D. Reuse, including value added through "metadata driven processes"

62. Discussion in the previous two subsections has focused on reducing and automating technical process steps to support streamlining and quality assurance of statistical production.

63. Achieving reuse of process steps (transformations) that add statistical value is also crucial to supporting streamlining.

64. Although, given different inputs, it would be possible conceptually, it will very rarely, if ever, be the case in practice that two complete statistical production processes, required to produce two different outputs, will require exactly the same combination of ***process steps***, including using the same methods used to undertake each of the required statistical transformations.

65. Nevertheless, many statistical ***process steps*** will be in common between different production processes. This is particularly the case at the level of business processes, even where it is not the same at the level of methods used to achieve the required transformation. For example, both survey data and administrative data is likely to undergo some form of imputation and estimation/aggregation – even if in some (but not all) cases the statistical methods used for the purpose may be different in detail.

66. Achieving appropriate re-use of process steps and

- (where appropriate) the specific statistical methods used to perform those processes/transformations, and
- (where appropriate) IT applications/services used to implement the methods in practice

will allow greater concentration on maintaining and supporting (through documentation, methodology and IT) a smaller, more consistent, and more quality assured set of process steps that can be assembled and configured (e.g. on a ***metadata driven*** basis) as required

[16] http://en.wikipedia.org/wiki/Adapter_pattern

67. Greater re-use of processes which support ***plug and play*** and ***connectivity*** facilitates and promotes greater reuse of existing data and metadata also. It is no longer necessary to undertake a number of technical process steps on that existing data and metadata in order for it to be re-used by a second production process. This significantly changes the "cost versus benefit" equation when it comes to reusing existing data and metadata.

68. This form of re-use places an emphasis on ***process steps***, and the components which implement them, being ***metadata driven***. Rather than being "hard wired" in some form, the components need to be able to

- accept as inputs, and process appropriately, a more diverse range of data and of metadata objects (eg classifications, variables) which might be associated with different statistical production processes related to different subject matter domains of statistics
- be able to select the specific statistical transformation methods to be applied to the specific inputs based on
    - explicit methodological parameters fed into the process step, and/or
    - business rules which determine the methods to be applied based on characteristics of the specific data and/or metadata inputs supplied to the process.

69. A broad definition of metadata driven processes might include cases where an appropriate range of well-structured information is made available to assist consistent and effective decision making by a staff member rather than being applied systematically by an automated process. In such cases the metadata may not completely predetermine the outcome but would be expected to shape outcomes over time.

70. A further argument for processes (whether automated or not) being formally defined and metadata driven is that it makes explicit (and corporatized) knowledge and understanding that might otherwise be confined to a few experienced staff who in future will retire or move on to other work.

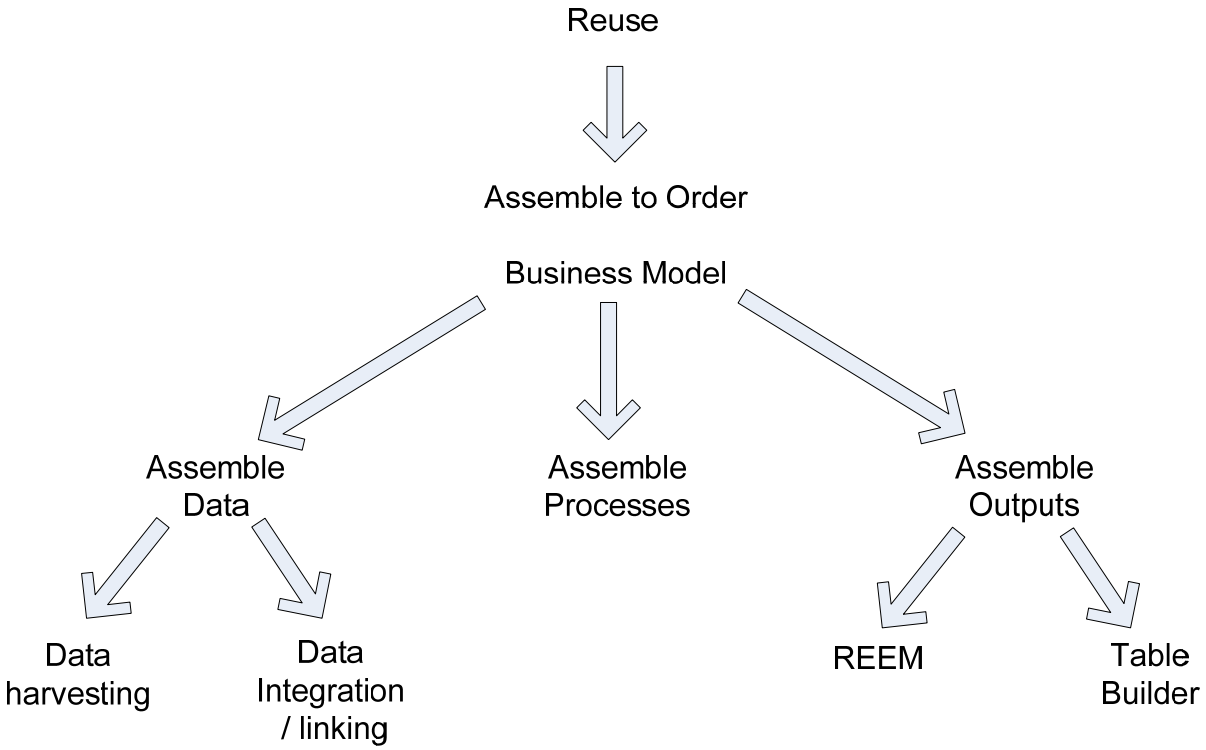
E. The Assemble To Order (ATO) Business Model

71. By assembling the right raw data and metadata inputs, and by assembling the appropriate set of metadata driven process steps, it becomes possible to "assemble outputs" to order.

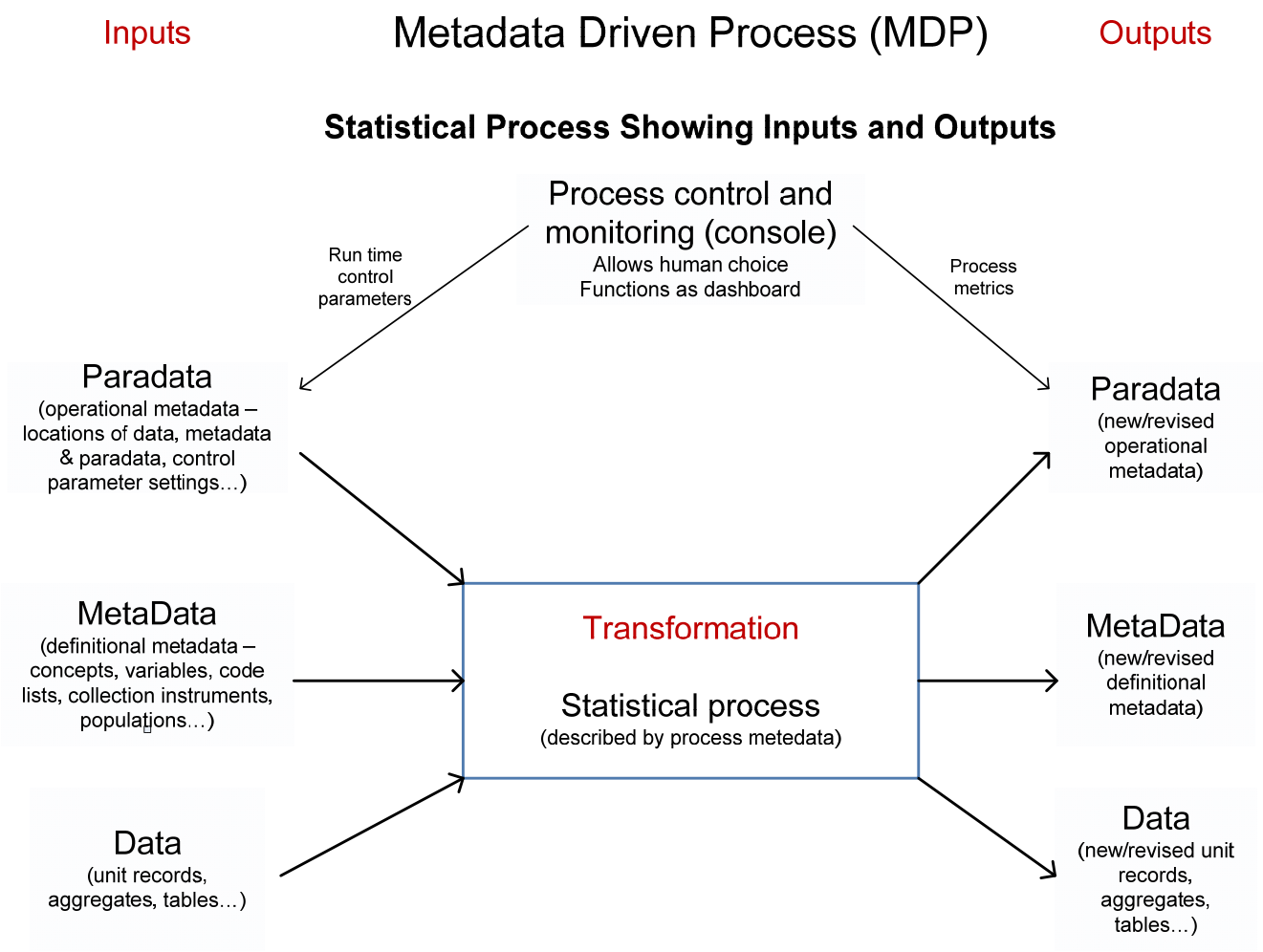
72. For example, there is a global growth in expectations and demand from researchers in various statistical subject matter domains to be able to undertake analysis based on microdata, while ensuring appropriate protection of privacy and confidentiality for individual respondents is guaranteed. Developing Confidentialised Unit Record Files (CURFs), as products to address this demand, is a quite manually customised and expensive process within the ABS currently.

73. Development work related to processes, statistical methodologies and IT solutions is underway presently which harnesses the Assemble To Order paradigm, including the design principles of ***modularity***, ***plug and play***, ***connectivity*** and ***metadata driven***. This development work aims to allow "assembly to order" of products, related to various statistical subject matter domains, which support analysis of microdata. The new products are expected to provide greater flexibility for end users while requiring less manual customisation (and expense) during production of each product.

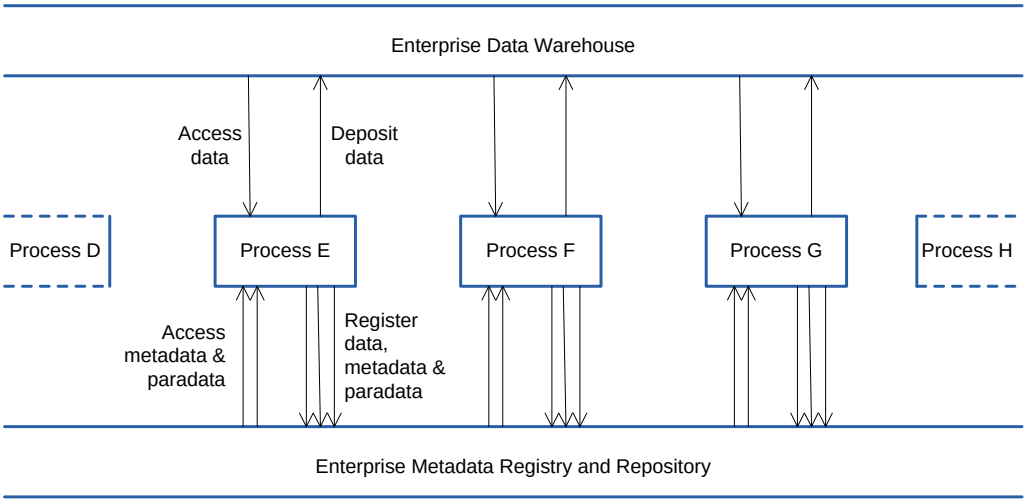
74. The Assemble To Order business model can be represented as follows.



E. Proposed high level architectural design for metadata driven processes



75. This approach makes it possible for outputs from upstream ***process steps*** to be used as inputs for downstream ***process steps*** without needing to "hardwire" the ***process steps*** together.



Sequence of Metadata Driven Processes with Inputs and Outputs

Key points: Processes do not interact directly with one another.
Data are located using paradata

76. The above approach also ensures that the life cycle of statistical data and metadata can be managed appropriately, in accordance with corporate requirements, through a set of enterprise level registries, repositories and content (e.g. data and metadata) management services.

77. A Business Process Management System (termed Statistical Workflow Management within the ABS 2017 transformation program) can be used to sequence and "orchestrate" the assembled processes without requiring the processes themselves to be "wired" together.

IV. Relationship with other architectural initiatives

A. CORA Common Layered Model for the Production of Statistics

78. A range of consultations have been undertaken by the ABS at various times with participants in the CORA and CORE ESSnet collaborations.

79. Based on these discussions, it is understood that the approach proposed in Section III. is fully consistent with the objectives of the CORA and CORE collaborations. In particular, the approach is consistent with the *CORA Common Layered Model for the Production of Statistics* as well as with the *CORE Information Model* which was developed subsequently.

80. Development of *A Common Layered Model for the Production of Statistics,* including *Definition of the Layered Common Reference Architecture,* was the target of a major work package within CORA.

81. The CORA layered model is premised on the same concept of *a process chain in which every step contributes added value to the information* and aims to *stimulate reuse of both human and automated solutions.* It shares the aspiration to reduce "platform dependency",

82. The model defines a **statistical process** as a *collection of ordered operations designed to transform observational data into statistical products.* (The elementary statistical operations within a process have been termed ***process steps*** in the current paper.)

83. CORA then defines a **service** as a *reusable generic solution for an elementary statistical operation.*

84. In CORA the process of statistics is described solely in terms of services calling services. Services are divided into layers with the intent that a service in a certain layer can access services of the first lower layer.

85. The rules applying to services calling other services, and being related to phases within the GSBPM, create a form of structured ***modularity***. Whether these rules should be considered absolute for practical business application remains questioned by some, but the layered model provides promising consistent guidance in this regard. ***Modular*** development is listed as intended benefit from applying CORA's layered model.

86. A complementary emphasis in the architectural model for ***Assemble To Order*** is that it maintains focus, for business purposes, on process and method related requirements. The CORA layered model, in contrast, focuses almost immediately on service oriented "solutions" for these requirements. This may be one reason why, at least anecdotally, the CORA layered model tended to be more accessible to those with an IT or architectural perspective than those with a business perspective.

B. CORE Information Model

87. CORE continued the work of CORA. In one work package, CORE extended the architectural model by defining a complete information model that takes into account process modelling, and specifically definition of sub-processes and communication interfaces.

88. It is noted that CORE goes in the direction of fostering the sharing of tools among NSIs. A tool developed by a specific NSI can be wrapped according to the extended CORA model and thus easily integrated within a statistical process of another NSI.

89. The CORE Information Model defines an interface communication protocol for the exchange of information between a CORA service and its environment. More specifically, the protocol describes the information elements a service receives in order to configure it, subsequently the elements a service receives as input upon execution, and finally the elements a service offers as output after execution. In this way, the communication interface between two services, or between a service and a service execution mechanism (eg a business process management system's run time execution engine) is established.

90. CORE models channels for communication regarding the following
- datasets (of various kinds)
- rules (of various kinds)
- parameters
- business objects
- free style arguments (e.g. scripts)

91. In regard to business objects, the CORE Model notes

> *modelling business objects does not belong to the scope of this model. The model knows of the existence of business objects, but knows nothing else about them. Such objects should be modelled under a mechanism specifically developed for them, such as the GSIM (Generic Statistical Information Model)*

92. This means that while the CORE Information Model provides, from a rather technical perspective, valuable architectural guidance in regard to practical design of ***plug and play*** capabilities, it deliberately avoids providing much guidance in regard to ***connectivity***.

93. In documenting possible future extensions, the model also notes that it currently incorporates a number of constraints and simplifications which may or may not prove tenable when it comes to practical application.

94. Anecdotal feedback from testing of the CORE framework suggests it has been demonstrated in practice that well described and well-designed communication interfaces for services allow them to be readily invoked

on an almost entirely platform independent basis (eg regardless of whether they are hosted in Java or .NET based environments).

C. GSIM

95. In contrast to the CORE Information Model, the focus of GSIM is to provide a common information model for statistical "business objects". According to its scope

GSIM models information objects used in statistical production, but not at the technology level

96. Its stated purposes include the following

- *Increase intra- and inter-agency reuse of data and ways of making statistics*
- *Enable configurable, rule-based ways of making statistics*

97. GSIM therefore intends to provide support for ***connectivity*** and, at least to some degree, for ***metadata driven processes.*** Neither GSIM nor GSBPM, however, provide the same level of detailed guidance at the technical level as the CORE Information Model when it comes to the connection between information (including metadata) and the business processes that the information drives.

98. GSIM is expected to undertake the conceptual modelling of business objects necessary to support ***connectivity.*** It is expected that the conceptual modelling undertaken by GSIM will then be tested by practical use cases. The architectural model for ***Assemble To Order***, and the CORE Information Model for supporting **plug and play**, should both provide good bases for testing and refining GSIM in this regard.

99. The planned directions for both GSIM and for the CORE Information Model consider how existing implementation standards (e.g. SDMX) could contribute to realising in practice, on a consistent basis, the proposed architectural design principles. Appropriate harnessing of existing standards wherever fit for purpose, in preference to defining new standards, can be seen as an element of the ***Assemble To Order*** paradigm also.

V. Conclusions

A. Reasons for progressing the approach proposed in this paper

100. Section II identifies that ensuring developments adhere to wisely selected common architectural standards, which are clearly defined in practice and measurable in regard to compliance, has many advantages compared with other paradigms for ensuring "means of production" are able to be shared.

101. Developing, proving and agreeing the necessary common architectural standards, to the necessary level of practical detail, will – however – take considerable time and effort. Reaching this detailed agreement will, in its own right, require considerable practical collaboration between agencies. If this vitally important, but difficult, goal is to be achieved then it is essential that consistently aligned and active support is mobilised within agencies by senior managers who understand the strategic imperative and what needs to be done in practice to address it.

102. A focus in this regard is also important because it will guide initiatives such as CORE and GSIM toward supporting a well-rounded set of architectural requirements in a coherent manner. The outputs from such initiatives, and the implementations of these outputs, might otherwise not provide coherent support for realisation of the strategic vision of HLG-BAS. There is no known existing initiative which has a high level focus across ***plug and play, modularity, connectivity*** and ***metadata driven processes*** all at the same time

103. Explaining the ***Assemble To Order Business Model*** within the ABS – along the lines set out in Section III. – has been instrumental in allowing many senior managers, and others from statistical subject management background, to understand the nature of its architectural underpinnings and why these are important from a

business perspective. Many of the target audience had previously heard the case for one or more of the underpinning architectural standards but had not appreciated the business meaning and importance of such standards, seeing them as primarily technical considerations.

104. Similarly the CORE and GSIM initiatives, together with discussions at MSIS and similar forums (e.g. SISAI – Statistical Information, Systems Architecture and Integration – within the ESS), have been important in securing "buy in" from, and aligning the thinking of, important experts. Capturing the attention, imagination and support of many key decision makers and influencers in regard to statistical business directions and investments, however, might be facilitated by starting first with a clear definition of the proposed business model and then, in that context, explaining from a business perspective which architectural standards/directions are required in order to realise it in practice.

105. It should be noted that while that proposed approach to common architectural standards is intended to go as far as facilitating the sharing in future of "built" components between agencies (including the IT implementations), it will also greatly facilitate sharing at a more general level. One example might be sharing the high level design of a component between agencies, even though different agencies may then use a different IT implementation for applying that shared design. The anecdotal feedback (IV.B.) from testing CORE is instructive in this regard because it suggests that a well modelled standard "technology agnostic" framework for defining information and processes may make the existence of different technical environments within NSIs less of an impediment to sharing being achieved in practice.

B. Proposed next steps

106. It is planned to engage with the MSIS community to test whether the case for

- the ***Assemble To Order Business Model***
- the architectural standards/directions required to realise it in practice
- the benefits it entails in regard to being able to share "means of production"

is expressed completely, clearly and compellingly enough to satisfy this well informed expert community.

107. This engagement should also lead to identification of any additional initiatives beyond CORA, CORE and GSIM whose existing work should be considered and integrated.

108. Volunteers would then be sought from within the MSIS community (or beyond) to collaborate in progressing this analysis and proposal further. This would include analysing CORA, CORE and GSIM in more depth from this business oriented, over-arching architectural perspective to identify

- which aspects from the existing models should be accepted as recommended bases achieving for ***plug and play, modularity, connectivity*** and ***metadata driven processes***
- which aspects from the existing models should undergo further review from a business perspective, possibly including practical testing against use case, before either confirming their acceptance or proposing refinements/modifications
- which, if any, high level architectural design features and guidance that appear to be required from a top down perspective are not yet reflected in CORA, CORE, GSIM or in other relevant standards and guides.

109. It is anticipated that the volunteer team which was progressing design and expression of the ***Assemble To Order Business Model***, and its underpinning architectural standards/directions, would also explore their thinking with senior managers and other business oriented experts in their local agency.

110. This approach would allow the collaboration partners to develop and present a well-articulated business model, a set of proposed architectural standards/directions and a proposed program of next steps to HLG-BAS, and to representatives from the Inventory of International Groups associated with HLG-BAS, at the next Business Architecture in Statistics workshop which is expected to take place from 7-8 November 2012.