United Nations ECE/CES/2014/3

Economic and Social Council

Distr.: General
30 January 2014

Original: English

Economic Commission for Europe

Conference of European Statisticians

Sixty-second plenary session
Paris, 9-11 April 2014
Item 5 of the provisional agenda
Work of the High-level Group for the Modernization of Statistical Production and Services

Common Statistical Production Architecture

Prepared by the High-Level Group for the Modernization of Statistical Production and Services[1]

Summary

This paper describes the Common Statistical Production Architecture, prepared under the High-level Group for the Modernization of Statistical Production and Services. The aim of the common architecture is to facilitate standardisation of the components of statistical production, regardless of where these components originate. It provides a framework, principles, processes and guidelines to help reduce the cost of developing and maintaining statistical processes and systems, and improving the responsiveness of the development cycle.

The document is submitted to the Conference of European Statisticians for information.

[1] The present document was submitted late due to resources constraints.

Contents

I. The problem statement

1. Many statistical organizations are facing common challenges. There are two major threats to the continued efficient and effective supply of core statistics that come from within statistical organizations. These are:

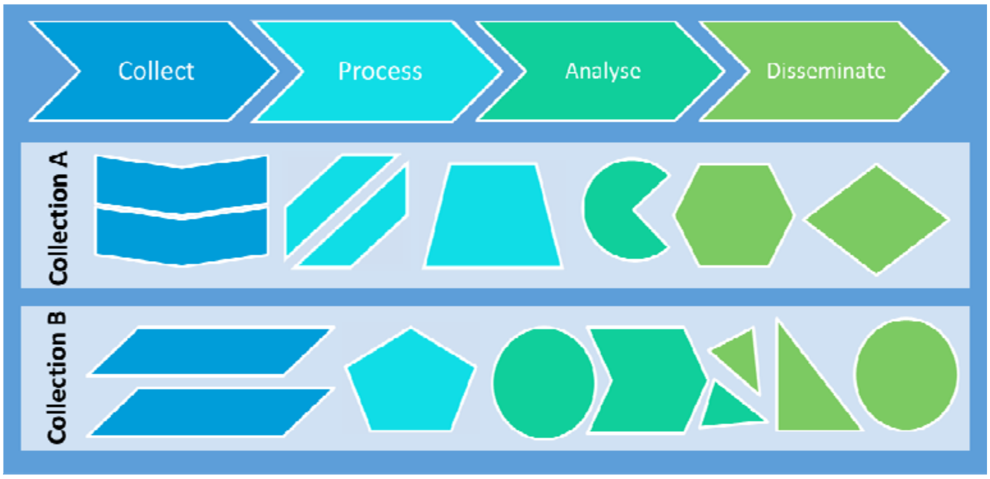
(a) rigid processes and methods, and

(b) inflexible ageing technology environments

2. Over the years, through many iterations and technology changes, statistical organizations have built up their organizational structure, production process, enabling statistical infrastructure, and technology. The cost of maintaining this business model and the associated asset bases (process, statistical, technology) is becoming insurmountable and the model of delivery is not sustainable.

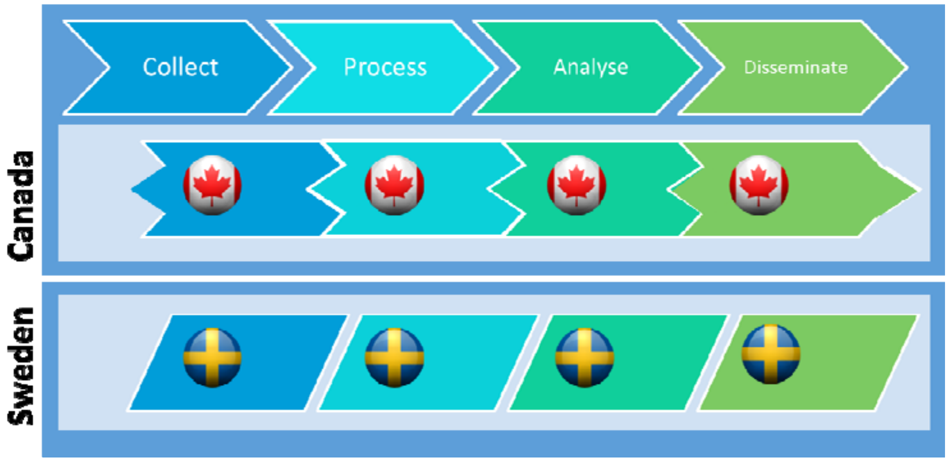
3. Historically, statistical organizations have developed their own business processes and IT-systems for producing statistical products. Therefore, although the products and the processes conceptually are very similar, the individual solutions are not (as represented by the different shapes in Figure 1). Every technical solution was built for a very specific purpose with little regard for ability to share information with other adjacent applications in the statistical cycle and with limited ability to handle similar but slightly different processes and tasks. This can be referred to as 'accidental architecture' as the process and solutions were not designed from a holistic view.

Figure 1
Accidental architectures



4. Often it is difficult to replace even one of the components supporting statistical production. Use of these processes, methods and an inflexible and aging technology environment mean that statistical organizations find it difficult to produce and share between systems data and information aligned to modern standards (for example, Data Documentation Initiative (DDI) and Statistical Data and Metadata eXchange (SDMX)). Process and methodology changes are time consuming and expensive resulting in an inflexible, unresponsive statistical organization.

Figure 2
The result of standardization within an organization

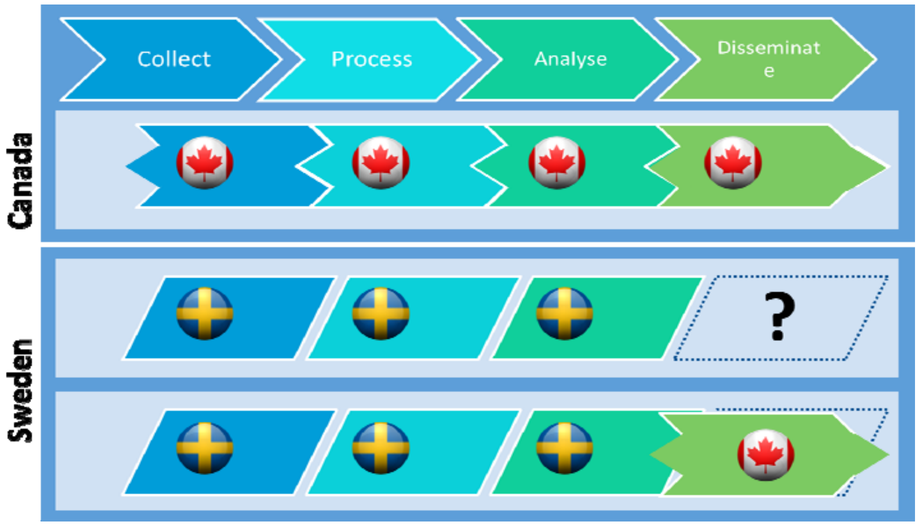


5. Many statistical organizations are modernizing and transforming their organizations using enterprise architecture to underpin their vision and change strategy. An enterprise architecture aims to create an environment which can change and support business goals. It shows what the business needs are, where the organization wants to be, and ensures that the IT strategy aligns with this. Enterprise architecture helps to remove silos, improves collaboration across an organization and ensures that the technology is aligned to the business needs. This work will enable them to standardize their organizations. This is shown in Figure 2 where, as opposed to Figure 1, the countries have standardized their components and interfaces.

6. Statistical organizations have attempted many times over the years to share their processes, methodologies and solutions, as it has long been believed that there is value in this. The mechanism for sharing has historically meant an organization taking a copy of a component and integrating it into their environment. Examples include CANCEIS (CANadian Census Edit and Imputation System) and Banff (an editing and imputation system for business surveys). However, most cases of sharing have involved significant work to integrate the component into a different processing and technology environment.

7. Figure 3 attempts to explain why the difficulty in sharing or reuse occurs. The figure assumes that the two statistical organizations in the figure develop all their business capability and supporting components and interfaces in a standard way (i.e. they have an Enterprise Architecture as shown in Figure 2). Each organization has standardized within their own organization but not in the same way as the other organization. As shown in Figures 2 and 3 where each country has a different shape of component - Canadian components have a zig zag shape and Sweden have components with slanted edges. If Sweden needs a new component, ideally they need a component with a slanted edge. It can be seen in the third row of Figure 3 that while a component from Canada might support the same process and incorporate robust statistical methodologies, it will not be simple to integrate it into the Swedish environment.

Figure 3
Why sharing /reuse is hard now



II. Common Statistical Production Architecture

8. As part of the modernization effort, the High Level Group for the Modernization of Statistical Production and Services (HLG) wishes to take action in order to address the problems and issues described in the previous section. For this reason, HLG has put priority on the development of the Common Statistical Production Architecture (CSPA) and its implementation.

9. If the official statistical industry had greater alignment at the business, information and application levels, then sharing would be easier. CSPA will assist statistical organizations address these problems by providing a framework, including principles, processes and guidelines, to help reduce the cost of developing and maintaining processes and systems and improving the responsiveness of the development cycle. Sharing and reuse of process components will become easier - not only within organizations, but across the industry as a whole.

10. The value proposition of CSPA, in providing statistical organizations with a standard framework, is to:

- facilitate the process of modernization
- provide guidance for transformation within statistical organizations
- facilitate the reuse / sharing of solutions and services and the standardization of processes, and thus a reduction in costs of production
- encourage interoperability of systems and processes
- provide a basis for flexible information systems to accomplish their mission and to respond to new challenges and opportunities
- enable international collaboration initiatives for building common infrastructures and services
- foster alignment with existing industry standards such as the Generic Statistical Business Process Model (GSBPM) and the Generic Statistical Information Model (GSIM)

11. CSPA is the industry architecture for the official statistics industry. An industry architecture is a set of agreed common principles and standards designed to promote greater interoperability within and between the different stakeholders that make up an “industry”, where an industry is defined as a set of organizations with similar inputs, processes, outputs and goals (in this case official statistics).

12. CSPA provides a reference architecture for official statistics. It describes:

- What the official statistical industry wants to achieve : this is the goals and vision (or future state)
- How the industry can achieve this: this is the principles that guide decisions on strategic development and how statistics are produced
- What the industry will have to do: the industry will need to adopt an architecture which will require them to comply with CSPA.

13. A number of frameworks focusing on specific areas have already been developed. CSPA builds on and uses these existing frameworks, notably the GSBPM and GSIM, as the necessary shared industry vocabulary. Adoption of these frameworks by organizations in the industry will improve the common understanding and alignment necessary for joint development, sharing and reuse of components.

14. CSPA complements and uses these pre-existing frameworks by describing the mechanisms to design, build and share components with well-defined functionality that can be integrated in multiple processes easily. CSPA focuses on relating the strategic directions of the HLG to shared principles, practices and guidelines for defining, developing and deploying Statistical Services in order to produce statistics more efficiently.

15. CSPA brings together these existing frameworks and introduces the new frameworks related to Statistical Services (described in Section V. Application Architecture) to create an agreed top level description of the 'system' of producing statistics which is in alignment with the modernization initiative.

16. CSPA gives users an understanding of the different statistical production elements (i.e. processes, information, applications, services) that make up a statistical organization and how those elements relate to each other. It also provides a common vocabulary with which to discuss implementations, with the aim to stress commonality. It is an approach to enabling the vision and strategy of the statistical industry, by providing a clear, cohesive, and achievable picture of what is required to get there.

A. Scope of architecture

17. CSPA is a reference architecture for the statistical industry. The scope of CSPA is statistical production across the processes defined by the GSBPM (i.e. it does not characterize a full enterprise architecture for a statistical organization). It is understood that statistical organizations may also have a more general Enterprise Architecture (for example an Enterprise Architecture used by all government agencies in a particular country).

18. CSPA is descriptive, rather than prescriptive, its focus is to support the facilitation, sharing and reuse of Statistical Services both across and within statistical organizations. CSPA is not a static reference architecture; it is designed to evolve further over time

19. CSPA is designed for use by investment decision makers in developed statistical organizations. While developing organizations are not excluded, a reasonable level of Enterprise Architecture maturity and a modern technical environment is required for implementation. There are options for making Statistical Services developed using CSPA

available to developing statistical organizations, these will be outlined in future versions of the architecture.

20. An important concept in architecture is the “separation of concerns”. For that reason, the architecture is separated into a number of “perspectives”. These “perspectives” are:

- Business Architecture which defines what the industry does and how it is done (statistics in our case)
- Information Architecture which describes the information, its flows and uses across the industry, and how that information is managed
- Application Architecture which describes the set of practices used to select, define or design software components and their relationships, and
- Technology Architecture which describes the infrastructure technology underlying (supporting) the other architecture perspectives.

21. CSPA includes

- Motivations for constructing and using CSPA through the description of requirements
- Sufficient business and information architecture descriptions and principles as are necessary for CSPA's scope
- Application architecture and associated principles for the delivery of Statistical Services
- Technology architecture and principles - limited to the delivery of Statistical Services

22. It should be noted that CSPA does not include enterprise, business, application and technology architecture descriptions which are not directly aligned to CSPA scope, nor does it prescribe technology environments of statistical organizations.

B. Service Oriented Architecture

23. The value of the architecture is that it enables collaboration in developing and using Statistical Services which will allow statistical organizations to create flexible business processes and systems for statistical production more easily.

24. The architecture is based on an architectural style called Service Oriented Architecture (SOA). This style focuses on Services (Statistical Services in this case). A service is a representation of a real world business activity with a specified outcome. It is self-contained and can be reused by a number of business processes (either within or across statistical organizations).

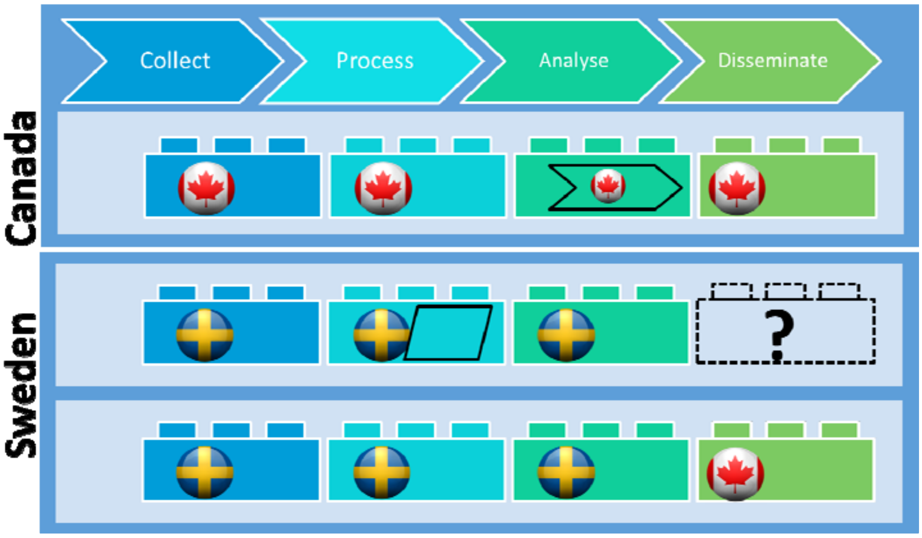
25. A Statistical Service will perform one or more tasks in the statistical process. Statistical Services will be at different levels of granularity. An atomic or fine grained Statistical Service encapsulates a small piece of functionality. An atomic service may, for example, support the application of a particular methodological option or a methodological step within a GSBPM sub process. Coarse grained or aggregate Statistical Services will encapsulate a larger piece of functionality, for example, a whole GSBPM sub process. These may be composed of a number of atomic services.

26. The granularity of Statistical Services should be based on a balanced consideration between the efficiency of the Statistical Service and the flexibility required for sharing purposes - larger Statistical Services will usually enable greater efficiency, whereas a finer granularity will allow greater flexibility for supporting sharing and reuse. Services, regardless of their granularity, must meet the architectural requirements and be aligned with CSPA principles.

27. By adopting this common reference architecture, it will be easier for each organization to standardize and combine the components of statistical production, regardless of where the Statistical Services are built. As shown in Figure 4, Sweden could reuse a Statistical Service from Canada because they both use the same component.

28. CSPA will facilitate the sharing and reuse of Statistical Services both across and within statistical organizations. The Statistical Services that are shared or reused across statistical organizations might be new Statistical Services that are built to comply with CSPA or legacy/existing tools wrapped to be Statistical Services which comply with the architecture. This is shown in Figure 4 by the shapes inside the building blocks.

Figure 4
Making sharing and reuse easier



C. Using CSPA

29. CSPA also provides a starting point for concerted developments of statistical infrastructure and shared investment across statistical organizations. CSPA is sometimes referred to as a “plug and play” architecture. The idea is that replacing Statistical Services should be as easy as pulling out a component and plugging another one in. There are a number of ways in which CSPA may be used by statistical organizations. These are outlined in the sections below.

1. Strategic Planning

30. If statistical organizations are creating and using an industry strategy (“Industry Architecture”), this leads to projects/work programs. An example is the Proof of Concept for CSPA. They could also integrate/streamline their investment strategies. Where a statistical organization plans to contribute to and or use CSPA in the future, they should modify and integrate their road maps to align with the CSPA framework. Each statistical organization needs to define a strategy to move from its current state to the common future state defined in their roadmap.

2. Development within statistical organizations

31. When a statistical organization identifies the need for a new Statistical Service, there are a number of options they can pursue. In order to fill the gap, the statistical

organization can look for Statistical Services that are available in the collaborative space (that is, in the Global Artefact Catalogue).

32. If an appropriate Statistical Service is not found in the CSPA Global Artefact Catalogue, the statistical organization can either:

- start designing and developing a new Statistical Service internally, or
- modify an existing Statistical Service to meet new functional and/or non-functional requirements.

33. This could be done independently or in collaboration with other statistical organizations. This development work should be done in alignment with CSPA to ensure that the new Statistical Services can be added to the CSPA Global Artefact Catalogue for sharing and reuse by other statistical organizations.

34. Sharing means exchanging concepts, designs or software, where each user of a service creates and operates its own implementation of that service. There are levels of sharing. A limited form of sharing would be to provide another participant with the means to replicate (make a copy) the asset (for example give the source code) (i.e. they share an aspect of the asset only). A more involved form of sharing would entail that asset is actually been made entirely common (in this case the asset is also reused).

35. Reuse means common use of a single implementation of a service, with only one organization acting as the service provider (the one who runs the service).

36. It is thought that in the current environment, it is more likely that statistical services will be shared across organizations rather than reused. Sharing between statistical organizations still leaves the option of reusing the various implementations internally within the environment of the individual statistical organization.

3. Vendors

37. A statistical organization may choose to have a vendor develop a Statistical Service. A vendor in this case means either a third party commercial vendor or a statistical organization that is selling a product. In the case of a new Statistical Service, the statistical organization should request that it is built in accordance with CSPA. When the product already exists, statistical organizations should verify together if the product meets relevant community requirements. If it does not, statistical organizations can try to influence the vendor to meet requirements. If it meets the community requirements, statistical organizations would ask the vendor to register the Statistical Service implementation to the Global Artefact Catalogue.

D. Impact on organizations

38. There will be a number of required changes for an organization implementing CSPA. Adoption of CSPA will require investment with a view to generating the long term benefits identified in the value proposition (see paragraph 10).

39. The main changes required at the organization level can be grouped as:

(a) People changes

- Openness to international cooperation
- Building trust in international partners (especially as they may be building services for your organization

- Sense of compromise (acceptance that nothing will be optimized for local use, rather it will be optimized for international or corporate use)
- Development of new functional roles to support use of the architecture (e.g. Assembler, Builder

(b) Process changes

- Adoption of an industry wide perspective
- Different approach to business process management and design
- Commitment to service (contract between different functional units)

(c) Technology changes

- setting up an adequate middleware infrastructure (messaging, repositories)
- uplift of physical network capabilities (bandwidth, etc.)
- management of security features

40. In addition to the costs and the targeted benefits, an organization adopting CSPA will benefit from:

- a sustainable and efficient strategy to cope with legacy and phasing out of existing applications
- a cycle that enables cost saving from reduction in production costs to be reinvested in further infrastructure transformation
- a positive image both on national and international/industry scene

III. Business architecture

41. The Statistical Network[2] is currently undertaking a project on Business Architecture. CSPA has adopted the outputs of this work.

42. The definition of Business Architecture being used by the Statistical Network Business Architecture is given below.

Business Architecture covers all the activities undertaken by a statistical organization, including those undertaken to conceptualize, design, build and maintain information and application assets used in the production of statistical outputs. Business Architecture drives the Information, Application and Technology architectures for a statistical organization.

43. CSPA focuses on architectural considerations associated with statistical production as bounded by GSBPM. Business concerns such as:

- ensuring that the corporate work program for a statistical organization best addresses the needs of its external stakeholders, or
- recruiting, retaining and developing staff with relevant skills are not central to CSPA. Such concerns are, however, very important considerations in an organization specific business architecture.

[2] The Statistical Network is a collaboration group involving the National Statistics Organisations of Australia, Canada, Italy, New Zealand, Norway, Sweden and the United Kingdom.

44. Organizations that have formally defined business architecture can reference CSPA when describing aspects of their business architecture which are fundamentally in common with other producers of official statistics.

A. Describing statistical production

45. To enable efficient and consistent documentation and understanding of CSPA three related concepts have been adopted which are relevant for all readers in relation to “Statistical Production”. These are Business Function, Business Process and Business Service.

46. The terms and definitions used in CSPA for these concepts are drawn from GSIM. The terminology and modelling of GSIM aligns with The Open Group Architectural Framework (TOGAF). TOGAF is widely known for defining architectural frameworks.

47. Following is a brief overview of these three concepts[3], a more detailed discussion of these concepts and statistical production is contained in Annex 1.

1. Business Function

48. GSIM defines “Business Function” as *something an enterprise does, or needs to do, in order to achieve its objectives.* This represents a simpler expression of the definition used in TOGAF.

49. When identifying Business Functions, the emphasis is on an enterprise level (“whole of business”) perspective, recognizing that different parts of the business may have different detailed requirements in regard to a particular function. At the level of the Business Function, there is no implementation detail.

2. Business Process

50. A Business Process is a set of process steps to perform on or more Business Functions to deliver a Statistical Program. Key aspects include:

- a process consists of a series of steps (activities/tasks)
- there is sequencing (or “flow”) between steps
- a business process is undertaken for a particular purpose
- what is represented (for the sake of simplicity and clarity) as a single step in a high level depiction of a process might – when viewed in more detail - comprise a lower level (sub)process consisting of multiple steps

3. Business Service

51. A Business Service is the means of accessing a Business Function. It will perform one or more Business Processes. It is the who - or what - will undertake the work associated with each function. Business services should be scoped to support flexible sequencing and configuration of Business Functions within different Business Processes. A Business Service has an explicitly defined interface that requires the knowledge of what the service will deliver (including in what time frame) given a particular set of inputs. A Statistical Service is a kind of Business Service.

[3] Descriptions are also available via the CSPA Glossary

52. A central aim for CSPA is to enable more efficient and flexible support for statistical production as described by the GSBPM. Future versions of CSPA will provide more guidance on how CSPA can be applied when designing, managing and performing statistical business processes. However, the initial focus in the development of CSPA has been to clearly and consistently define (both conceptually and in practice) the Statistical Services which support statistical production. The aim is that equivalent business functions (such as imputation) within many different statistical business processes[4] will be able to reuse (or share) the same Statistical Service in the implementation of the business service.

53. Key reasons for the initial focus on Statistical Services include:

- Common Statistical Services provide the greatest opportunity for realizing savings through collaborative development, sharing and reuse
- While common statistical standards (or reference frameworks) have been agreed for business processes (GSBPM) and for statistical information (GSIM), there is not yet a common framework for Statistical Services, CSPA is filling this gap
- Once a common approach to the definition and implementation of Statistical Services is agreed, this will support defining business processes which make use of the common Statistical Services in a manner that supports the delivery of a specific business service, within a given context and purpose (business process) aligned to the business function (GSBPM).

B. Business Architecture Principles

54. Principles are high level decisions or guidelines that influence the way processes and systems are to be designed, built and governed. Principles are derived from the mission and values of the organization, taking into account the opportunities and threats that the organization faces. In CSPA, principles are used to express the high level design decisions that will shape the future statistical processes and systems.

1. Decision principles

55. Decision principles are guidelines to help decide on strategic development. They provide a basis for decision making and informing how the mission is fulfilled. They help enable sound investment decisions. The following decision principles support the outcomes sought by the High Level Group and key elements of the United Nations Principles for Official Statistics. These principles provide a basis for decision making and inform how a statistical organization sets about fulfilling its mission through strategic development.

56. A number of principles which are common to most organization’s business architecture (whether formally defined or not) are being identified through other initiatives such as work within the Statistical Network. The following business architecture decision principles are being jointly developed via the Statistical Network business architecture project[5] and the CSPA project:

(a) Principle**:** Capitalize on and influence national and international developments

[4] Whether these business processes relate to different subject matter domains in one statistical agency and/or equivalent subject matter domains in different agencies.

[5] The Statistical Network Business Architecture Project intends to expand these principles to include rationales and implications. These will be incorporated when developed.

Statement**:** Collaborate nationally and internationally to leverage and influence statistical and technological developments which support the development of shared statistical services.

(b) Principle**:** Deliver enterprise-wide benefits

Statement**:** Design and implement new or improved statistical business processes in a way that maximizes their value at an enterprise level.

(c) Principle**:** Increase the value of our statistical assets

Statement**:** Add value to the statistical organization's statistical assets (either directly or indirectly) through improved accessibility and clarity, relevance, coherence and comparability, timeliness and punctuality, accuracy and reliability and interpretability.

(d) Principle**:** Maintain community trust and information security

Statement**:** Conduct all levels of business in a manner which build the community's trust. This includes the community's trust and confidence in the statistical organization's decision making and practices and their ability to preserve the integrity, quality, security and confidentiality of the information provided.

(e) Principle**:** Maximize the use of existing data/Minimize respondent load

Statement**:** Leverage existing data from all sources (e.g. statistical surveys or administrative records) before collecting it again. Statistical organizations are to choose the source considering quality, timeliness, cost and burden on respondents. Statistical authorities monitor the respondent burden and aim to reduce it over time

(f) Principle**:** Sustain and grow the business

Statement**:** Focus investment and planning on long term sustainability and growth, both in terms of the organization's role and position within its own community as well as internationally.

(g) Principle: Take a holistic and integrated view

Statement**:** Ensure data, skills, knowledge, methods, processes, standards, frameworks, systems and other resources are consistent, reusable and interoperable across multiple business lines within a statistical organisation.

2. Design principles

57. CSPA aims to support organizations in realizing these decision principles in practice. The Business Architecture design principles have been identified for CSPA in conjunction with the Statistical Network Business Architecture project.

(a) Principle**:** Consider all capability elements

Statement: Consider all capability elements (e.g. methods, standards, processes, skills, and IT) to ensure the end result is well-integrated, measurable, and operationally effective.

(b) Principle**:** Re-use existing before designing new

Statement**:** Re-use and leverage existing data, metadata, products and capability elements wherever possible before designing new.

(c) Principle**:** Design new for re-use and easy assembly

Statement: Design and standardize all new data, metadata, products and capability elements for re-use, so they can be easily assembled and modified to accommodate changing user demands.

(d) Principle**:** Processes are metadata driven

Statement**:** Ensure the design, composition, operation and management of business processes, including all input and output interactions, are metadata driven and automated wherever possible.

(e) Principle: Adopt available standards

Statement**:** Aim to adopt open, industry recognized, and international standards where available. Statistical industry standards such as the Generic Statistics Business Process Model (GSBPM) and the Generic Statistical Information Model (GSIM) are examples of the standards to be used.

(f) Principle: Designs are output driven

Statement: Ensure the whole statistical process is output-driven. Output is the reference starting point; the statistical production process starts from the output desired, that is from required products, and goes backwards, defining the various aspects of the process.

(g) Principle**:** Enable discoverability and accessibility

Statement: Ensure data, metadata, products and capability elements are discoverable and accessible to achieve the benefits from sharing and reuse.

58. The Information and Application architecture design aspects of CSPA are directed by the CSPA Business Architecture design principles.

IV. Information Architecture

59. The Statistical Network Business Architecture provides the following definition of Information Architecture:[6]

Information Architecture (IA) classifies the information and knowledge assets gathered, produced and used within the Business Architecture. It also describes the information standards and frameworks that underpin the statistical information. IA facilitates discoverability and accessibility, leading to greater reuse and sharing.

60. In other words, Information Architecture connects information assets to the business processes that need them and the IT systems that use and manage them.

61. It includes relating the coherent and consistent definition of information assets at an enterprise level to the information needs of specific business processes and IT systems in practice.

62. As an industry architecture, the Information Architecture set out by CSPA must provide an agreed and actionable (rather than purely conceptual) connection between

- the common information frameworks and implementation standards agreed within the industry (e.g. GSIM, SDMX, DDI), and
- the practical business goals and needs to be supported under CSPA, such as the ability to share and reuse Statistical Services

63. It must support the needs of:

[6] While the standards body responsible for TOGAF recognises the term “Information Architecture”, the formal model underlying TOGAF refers to “Data Architecture”. .

- business leaders, planners and process designers who are seeking to apply the Business Architecture from CSPA and who need to understand the connection between processes and information at a business level
- application architects and developers who are seeking to apply the Application Architecture from CSPA and who need to understand how Statistical Services interact with information

A. Reference frameworks and their use

64. The Information Architecture will identify common reference frameworks to be used for aligning communication and high level (conceptual) designs.

- GSBPM will be used as a common reference when recording information in regard to business processes.
- GSIM will be used as a common reference when defining the information input to, and output from, business processes.
- A common reference framework for recording information in regard to the definition of Statistical Services is being developed as part of CSPA (this is described in Section V. Application Architecture).
- A common reference framework to use when describing statistical methods is a gap at this stage.

65. The completed Information Architecture will not only identify the reference frameworks which apply but also provide guidance on how they are applied, in combination, within CSPA.

B. CSPA implementation specifications

66. A major barrier to effective collaboration within and between statistical organizations has been the lack of common terminology. Using GSIM as a common language will increase the ability to compare information within and between statistical organizations. It allows all processes that lead to the production of statistics to be described in one integrated information model.

67. Although GSIM can be used independently, it has been designed to work in conjunction with the GSBPM. It supports GSBPM and covers the whole statistical process. It is assumed in this document that an organization either uses the GSBPM or uses another business process model, which can be mapped to the GSBPM.

68. In order for interoperability and reuse to be supported in practice when applying CSPA, the industry needs to do more than align conceptual designs using common frameworks. While GSIM is a conceptual framework for describing Statistical Information, when it comes to describing information objects in the real world we need to describe them in terms of standards for representing those objects physically (i.e. in practice) in a manner which is consistent with GSIM.

69. It is necessary to specify how a conceptual design is to be translated into an implementation which is consistent and readily sharable in practice.

70. This “standard” means of operationalizing conceptual designs can be referred to as an implementation specification (in this case, GSIM Implementation). To this end, the following has been agreed:

- No firm recommendation has yet been made on implementation specification for business processes.
- Depending on what information is being represented in practice, DDI and SDMX are expected to provide the primary basis for CSPA implementation specification in regard to statistical information (e.g. data and metadata).
- An implementation specification for Statistical Services is being developed as part of CSPA.

71. There is a need to do more than simply refer to relevant existing standards such as SDMX and DDI. The CSPA implementation specification for Statistical Services will specify:

- whether SDMX, DDI or a custom schema should be used for representing a particular GSIM information object, and
- exactly how the chosen schema will be applied for the particular purpose. In many instances there are multiple technically compliant means of achieving the same business purpose, the implementation specification will specify which should be used.

72. Implementation specifications mean CSPA is prescriptive in regard to some practical details. While it would be simpler to align with CSPA if it was less prescriptive, the practical value from alignment would be much less. It is often the case that two developments which have a “common conceptual basis”, but were implemented using completely unrelated approaches, are difficult and expensive to make interoperable and/or sharable (if it is possible at all).

73. In addition, an organization which has already implemented a different standard, or a local specification, can “map” their existing approach to the relevant implementation specification – they are not required to “rebuild” from first principles.

74. CSPA implementation specifications specify approaches which will support maximum interoperability/sharability on a cost effective basis. In particular cases it may be difficult for an organization to fully comply with a CSPA implementation specification (due to operational constraints). In these cases, compliance to the extent practical will, generally, still realize significant benefits. In other words, while CSPA implementation specifications set the bar reasonably (but not unreasonably!) high, it is recognized not all implementations may be able to achieve it fully in practice.

C. Information architecture principles

75. A number of principles which are common to most organization's information architecture (whether formally defined or not) have been agreed. These are outlined below.

(a) Principle**:** Manage information as an asset

Statement**:** Information is an asset that has value to the organization and must be managed accordingly.

(b) Principle: Manage the information lifecycle

Statement**:** All information has a lifecycle and should be managed to provide reliable identification, versioning and all information should be managed independently and beyond the scope of a single service.

(c) Principle: Protect information appropriately

Statement: All personal, confidential and classified data should be protected and the data should be treated accordingly.

(d) Principle: Use agreed models and standards

Statement: All information used as inputs and outputs to Statistical Services should be described using a common, business-oriented, reference model. A single standard should be used to define the encoding of each type of information.

(e) Principle: Capture information as early as possible

Statement: Information should be captured in a standard structured manner at the earliest possible point in the statistical business process to ensure it can be used by all subsequent services.

(f) Principle**:** Describe to ensure reuse

Statement**:** All information should be described in a manner that ensures information is reusable between services. Reuse is intended to reduce duplication, additional human intervention and reduce errors.

(g) Principle**:** Ensure there is an authoritative source

Statement**:** Information consumed and produced by services should be sourced and updated from a single authoritative source. Information should be consistent across all relevant services.

(h) Principle**:** Preserve information input into Statistical Services

Statement: Information that is input into services must be preserved in the service output to ensure no information loss.

(i) Principle: Described by metadata

Statement: All information consumed and produced by services must be described by sufficient metadata.

V. Application Architecture

76. The Statistical Network Business Architecture provides the following definition of Application Architecture:

Application Architecture (AA) classifies and hosts the individual applications describing their deployment, interactions, and relationships with the business processes of the organization (e.g. estimation, editing and seasonal adjustment tools, etc.). AA facilitates discoverability and accessibility, leading to greater reuse and sharing.

77. The CSPA Application Architecture is based on an architectural style called Service Oriented Architecture (SOA). This style focuses on Services (or Statistical Services in this case). A service is a representation of a real world business activity with a specified outcome. It is self-contained and can be reused by a number of business processes (either within or across statistical organizations).

78. Statistical Services are defined and have invokable interfaces that are called to perform business processes. SOA emphasizes the importance of loose coupling. Interactions between Statistical Services are independent, that is, they do not talk directly to each other. Organizations will need a technology solution to support communication between Statistical Services. This solution (for example a communication platform) will not affect the interfaces. It should be noted that SOA is not the same as Web Services, although they are often used in SOA.

A. Statistical service definitions, specifications and implementation descriptions

79. The level of reusability promised by the adoption of a SOA is dependent on standard definitions of the services. CSPA has three layers to the description of any service. These layers are described in the following paragraphs and Figure 5.

1. Statistical Service Definition

80. The Statistical Service Definition is at a conceptual level. In this document, the capabilities of a Statistical Service are described in terms of the GSBPM sub process that it relates to, the business function that it performs and GSIM information objects which are the inputs and outputs.

2. Statistical Service Specification

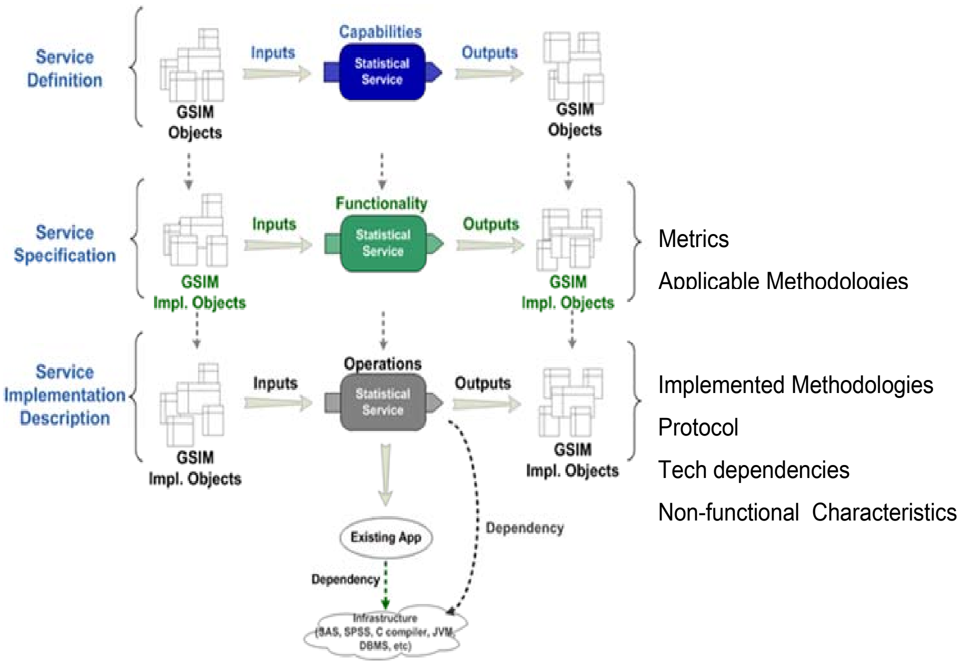
81. The Statistical Service Specification is at a logical level. In this layer, the capabilities of a Statistical Service are fleshed out into business *functions* that have GSIM implementation level objects as inputs and outputs. This document also includes metrics and methodologies.

3. Statistical Service Implementation Description

82. The Statistical Service Implementation Description is at an implementation (or physical) level. In this layer, the functions of the Statistical Service are refined into detailed operations whose inputs and outputs are GSIM implementation level objects.

83. This layer fully defines the service contract, including communications protocols, by means of the Service Implementation Description. It includes a precise description of all dependencies to the underlying infrastructure, non-functional characteristics and any relevant information about the configuration of the application being wrapped, when applicable.

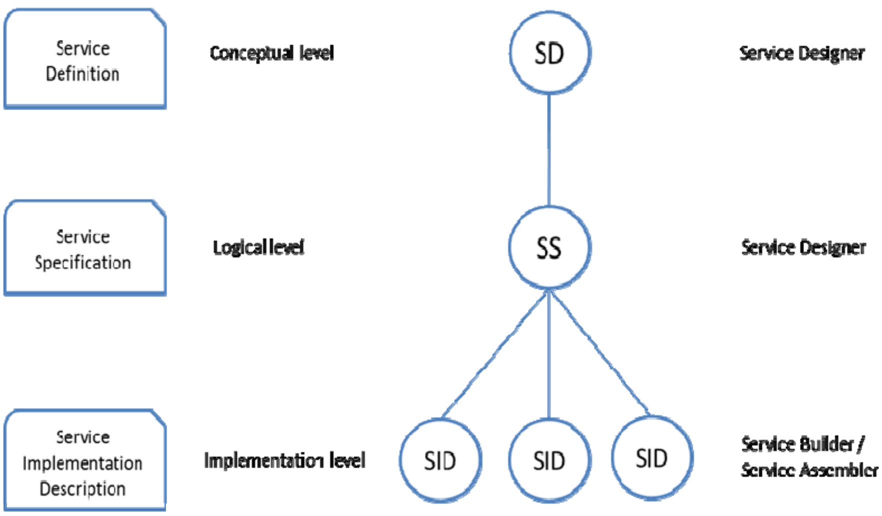
Figure 5
Service interfaces at different levels of abstraction



84. In general, there will be one Service Specification corresponding to a Service Definition, to ensure that standard data exchange can occur. At the implementation level, services may have different implementations (software dependencies, protocols, supported methodologies) reflecting the environment of the supplying organization. Each implementation must rigidly adhere to the data format specified in the Service Specification

85. There are a number of roles identified in CSPA (see Section VII) who are involved in the definition, specification, and implementation of Statistical Services. Figure 6 illustrates the relationship between these levels.

Figure 6
Linkages between Statistical Service Definition, Specification, and Implementation



B. Architecture patterns

86. In simple terms, architecture patterns describe a re-usable solution to certain classes of problems. They explain how, when and why Statistical Services can be used, as well as the impact of using them in that way. They help a Service Assembler to identify combinations that have been used successfully in the past. Although not identified in this document, there are also anti-patterns which are examples of what should not be done.

87. The benefits of using architecture patterns can be described by using the analogy of an expert chess player. To play chess, you must learn the rules and the principles (for example the value of different pieces). However, to improve and become a really good player, you need to learn the patterns used by more experienced players and apply them to your game. In the same way, you can use the principles and non-functional requirements of CSPA, but to get the maximum benefit, Service Assemblers will need to learn the architecture patterns.

88. CSPA will incorporate the Request/Response and Publish/Subscribe patterns.

1. Request/Response pattern

89. The Request/Response pattern for activating services implies a rather fixed routing of messages between services. The integration infrastructural platform implementing the process “orchestrates” the routing and executing of services. This pattern leads to less flexibility and tighter coupling between services than the Publish/Subscribe pattern described below.

90. An example of how this pattern could be applied in relation to collection is each questionnaire is stored in an entity service. The entity service exposes the operation to get the questionnaire through a service call. The indicators are computed and stored and made available using an entity service call.

91. The Request/Response pattern can be used if:

- A functional style and sequential flow is required
- It is known precisely which service interface should be called

2. Publish/Subscribe pattern

92. The Publish/Subscribe pattern could be considered an asynchronous version of the Request/Response pattern.

93. As an event is generated by an event source and is sent to the processing middleware. It is not known which functionality is triggered next. In the Request/Response pattern, the concrete service call would have been made, but this is not the case for the Publish/Subscribe pattern. For this reason, the Publish/Subscribe pattern talks about "decoupling" rather than loose coupling.

94. An example of how this pattern could be applied in relation to collection is when each questionnaire completed publishes an event that is available for subscribers downstream. Early indicators can be produced by processing collection events straight through aggregation

95. The Publish/Subscribe pattern can be used if:

- All recipients that may be interested in the event should be notified
- It is not exactly known which and how many recipients are interested in the event

- It is not known how recipients respond to this event
- Different recipients respond differently to the same event
- Only one-way communication from the sender to the recipient is possible

C. Non-functional requirements

96. In the context of CSPA, a non-functional requirement is a requirement that relates to the operation of a system. While functional requirements define what the services does (for example, error localization), the non-functional requirements describe a performance characteristic of a system (for example, authorization of who can access the resources and functions of the service). That is, non-functional requirements determine how a service behaves rather than what it should do.

97. Non-functional requirements are important to be captured in the design of the services. They have a significant influence on the software architecture of a service. The Designer[7] of a Statistical Service should identify the non-functional requirements that are relevant to that service when they are designing it. The implementation of a Statistical Service provides some functional value when assembled into a value chain within an organization. The non-functional requirements of a Statistical Service address other concerns or behaviours of the service such as performance, security, process metrics and error handling. This section provides some guidance on these concerns.

1. Multilingual Support

98. It is necessary to provide multilingual support to enhance the usability and the capacity to share Statistical Services. All services must be documented at least in English in addition to the local language(s) of the organization developing the Statistical Service. It is highly recommended that organizations that have made translations of the documentation of a Statistical Services in additional languages make them available to the community.

2. Security

99. For the purpose of this document, the security concern relates to controls that are put in place to mitigate the risk that a Statistical Service or the data it controls is misused. This section provides some basic guidance on some of these controls. However, in general it is strongly advised that each Statistical Service implementation complete a Risk Assessment and document a Risk Mitigation Plan for high and extreme risks identified in the assessment.

A. Authentication and Authorization

100. For users interacting with a Statistical Service, the process of identifying who they are (authentication) and working out what resources and functions of the service they can use (authorization) will need to be managed as part of the service. Given that the security architecture and services is a local organization concern, our goal is to avoid excess complexity in these interactions.

101. Authentication should be accomplished through interaction with the communication platform's authentication function

[7] NB: A description of all CSPA roles can be found in Section VII of this document

102. Authorization controls are the concern of the service implementation. Authorization will be controlled by either administrative interfaces on the service or via a GUI-based client for the specific service.

103. After the release of v1.0, future iterations of CSPA will look at options to 'design in' support for single sign on. Consideration will also be given in future CSPA versions to how the architecture can describe a common approach for the communication platform to pass authorization information along with other service context information.

B. Data at rest

104. Data at rest is of particular interest when a Statistical Service needs to defer state (see discussion in “Service Statelessness” in section V D). Under this circumstance, the security (e.g. encryption requirements or access control) of the data are entirely the responsibility of the Statistical Service. Where a Statistical Service already has a functional dependency on underlying technologies or platforms, it would be reasonable to make use of security functions available in those technologies.

C. Data in transit

105. Security of data in transit (e.g. contained within a message flow as part of a service invocation) will be considered in future iterations of CSPA (post version 1.0).

106. Sensitivity of statistical data varies amongst organizations and at this stage the architecture does not attempt to converge on a standard definition or treatment.

D. Machine to machine certification

107. Guidance for this will come in a future iteration of CSPA (post version 1.0). Organization specific implementations based on assembly time infrastructure can assure security for service communication (for example, use of a VLAN).

E. Performance

108. No specific guidance is provided on the performance characteristics. However, they should be declared in the Statistical Service Implementation Description and it is recommended that examples of performance level are included

F. Process Metrics

109. A Statistical Service will generally capture metrics related to the function that it performs. To all intents and purposes, these process metrics are treated by the Statistical Service as just one of its outputs and should be reflected as such in the Statistical Service Specification.

G. Error Handling

110. Error handling, in this case, relates to situations where the service fails. Error handling is left to the communication platform to handle as required. Generally there will be protocol specific requirements for flagging errors. The error codes and their meanings need to be documented in the Statistical Service Implementation Description.

D. Implementing Protocols in a Statistical Service

1. Service Statelessness

111. The principle of service statelessness is, in essence, that - An individual service should be called with all the information it needs to complete, the service shouldn't rely on previous execution. However there are cases where a Statistical Service needs to defer or persist some form of state. This principle trades off flexibility of the service to fit into a particular process with scalability of the service.

112. For CSPA, this means that there are certain situations when the Statistical Service needs to be able to keep information within the Service until a later time.

113. When designing and building a Statistical Service, the capability of deferring state information is important in two specific situations:

- When the Statistical Service needs to be used in a publish\subscribe design pattern.
- When the Statistical Service involves human-interaction and can therefore be considered long-running.

114. Statistical Services with the capability of deferring state needs to provide an endpoint to support enquiries about the deferred state. If the Statistical Service is invoked but it lacks some or all of the required information to perform the service invocation, the Statistical Service should handle this using error handling.

2. Event Messaging and Request Response

115. Communication to and from a Statistical Service could be done using several communication patterns. The two main patterns of communication that CSPA supports are Request Response and Event Messaging. These patterns differ mainly in when the information is being transferred.

116. In the *Request Response* communication pattern, information is requested whenever a Statistical Service needs it. The information need could be triggered by either a human-interaction with a Statistical Service or when an automated Statistical Service is called from a communication platform.

117. In *Event Messaging*, information is transferred as it is created. This means a Statistical Service does not need to request information as information is provided to all relevant Statistical Services at the time of creation.

118. The decision as to whether a Statistical Service is built to support both of these communication patterns is related to the environments in which the Statistical Service will work in. Supporting both communication patterns is preferred as this allows the Statistical Service to function both in Statistical Organizations using request-response as the main pattern as well as Statistical Organizations using Event Messaging as the main pattern.

119. From the point of view of a Statistical Service, there is no difference between the Request-Response and Event Messaging communication patterns, as the provision of information is done in the same manner. When a communication platform supplies the information that the Statistical Service needs, it does so by calling the relevant endpoint provided by the Statistical Service.

120. To support a *Request Response* communication pattern, the Statistical Service must provide some endpoint where an information request can be made. One possibility is that the Statistical Service is required to have an endpoint that can be supplied with parameters. These parameters would describe what information should be part of the response. For example, the parameters could include being able to specify a timeframe or a context.

121. To support *Event Messaging*, the Statistical Service needs to be able to push information out from the Statistical Service when it is created. This could be done using a configurable endpoint or message queue provided by the communication platform. Configuring the endpoint should be done by the Service Assembler.

3. How to invoke a service

122. A protocol is the technical implementation of a communication mechanism. It is used to invoke the Statistical Services deployed in a CSPA implementation.

123. A Statistical Service Implementation Description must specify one or more protocols. These protocols are associated with the following aspects of a Statistical Service.

- making the Statistical Service reachable as an endpoint for invocation
- accessing data that is declared to be passed by reference in the Statistical Service Specification

124. In the following, we provide a list of protocols that are accepted within Service implementations conforming to the CSPA specification. Protocols tagged as "recommended" should be considered in the first place because they represent established industry standards and as such they are likely to be supported in most organizations. Protocols that are accepted yet not recommended are listed for supporting legacy requirements of some organizations.

A. Protocols for invoking service endpoints

125. The protocols for invoking service endpoints which are recommended by CSPA are:

- SOAP Web Services – Service exposes a WSDL interface and is addressed by a http URI
- REST Web Services - Service exposes a REST interface and is addressed by a http URI

126. There are also a number of other protocols which are acceptable. These are:

- Microsoft Message Queue - Service is a MSMQ consumer
- Java Messaging Service - Service is a JMS consumer
- File-based invocation – the service is "invoked" when a file is placed at a known location which results in an OS-level trigger to the service; alternatively, the service can poll the location for arrival of "message" files and treat them as service invocations
- Command line interface - Service is invoked by specifying a command line to be executed on operating system runtime accessible by the platform.

127. CSPA recognized that there are other future protocols. However, these will require further exploration:

- Possible use of "stream control transfer protocol" (sctp), an efficient guaranteed delivery, connectionless, lightweight transfer protocol

128. In some instances, existing tools support database access. If the database is involved in transfer (and not merely as a local state storage for the service), we recommend that the database access be mediated through the http: protocol access above.

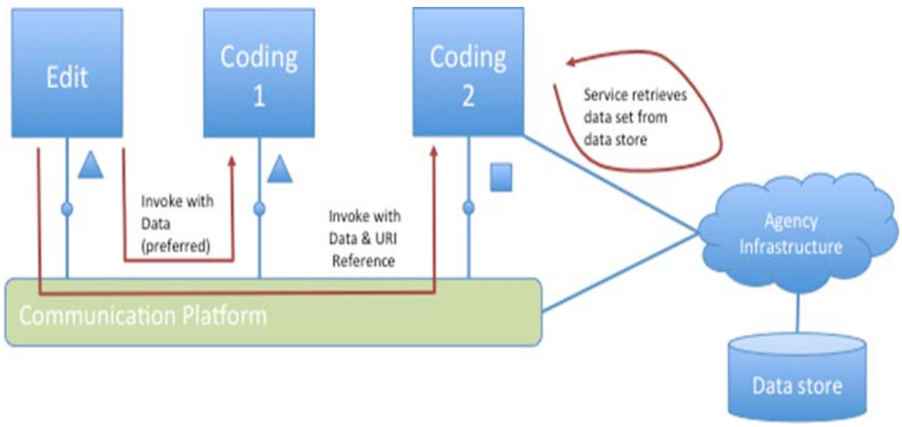
129. In general, the use of an "out of band" data transfer mechanism should be avoided wherever possible, and used only in circumstances involving the need to transfer large

volumes of data. Its addition adds increased coupling between the architecture and services, so its use must be managed carefully.

B. Protocols for passing data by reference

130. A data reference must guarantee, in the context of a specific protocol, that it can uniquely identify a dataset, file, etc. Passing data by reference requires the communication platform to support the resolution of an identifier through the specific protocol it refers to.

Figure 7
Managing large data volume transfers using "pass by reference



131. As a general rule, service invocation will involve the Statistical Service receiving a message via the organization-provided communication platform. This message will contain the necessary information objects as well as the requested service.

132. In certain circumstances, the service requires large data sets as inputs. Examples of this could include administrative data files or large survey response files. The problem is similar to a "pass by value" situation in that the input data is passed to the service via in-message approaches.

C. Problem

133. There are a number of problems that can arise if we attempt to send these data sets via the messaging interface:

- Dataset transfer time can be slow due to messaging overhead (packing / unpacking of data, message segmentation and reassembly, etc.)
- Communications platform performance may degrade due to the load of transporting the messages between services
- Service memory requirements can increase before required use (see State Deferral discussion

D. Solution

134. In order to address this problem, we provide a "pass by reference" mechanism (see Figure 7) that avoids the need to use the communication platform messaging layer to transport these large data sets.

135. The approach is as follows

- The data set being sent to the service is stored in a source location in a manner local to each organization – the location name is associated with a Uniform Resource Identifier (URI)
- The service consumer invokes the requested service by sending it a message containing the URI for the dataset
- The service provider receives the URI reference and when ready attempts to retrieve the dataset from repository or cache. If successful, it executes its service actions
- Upon completion, it may update or place a resulting dataset (if relevant) in the repository or cache

136. The implementation of a data source is local to each organization and may be implemented as part of the communications platform. Organizations may choose to implement a utility service, a repository, a file cache, or some other mechanism. URI management is also a part of local operation.

137. CSPA provides the following guidance for service input dataset retrieval protocols.

(a) Recommended protocols:

- Simple http: file transfer from data source to the service logic (without additional protocols such as REST)

(b) Acceptable protocols:

- ftp: file transfer from the data source to the service logic
- Use of network file system services (such as SMB, NFS) with appropriate file reference

(c) Not Recommended:

- Database retrieval using queries

E. Application Design Principles

138. The design principles have been selected to maximize the flexibility of the Statistical Services wrapped or developed in the context of CSPA. The flexibility of the Statistical Service directly impacts the level of reuse, the flexibility required of the industry vision and the ease with which a statistical organization can implement a Statistical Service.

(a) Principle**:** Maintain independence between design and implementation

Statement**:** The descriptions of Statistical Services are layered in conceptual (Statistical Service Definition), logical (Statistical Service Specification) and implementation (Statistical Service Implementation Description).

(b) Principle**:** Use available standards

Statement: The design of Statistical Services should align with, and harness, relevant existing standards and frameworks wherever possible.

(c) Principle: Use architecture patterns

Statement: Follow architecture patterns depending on best fit to requirements.

(d) Principle: Implement using GSIM

Statement**:** Manage standardized service contracts based on GSIM objects

(e) Principle: Minimize coupling .

Statement**:** Enable services to be loosely coupled externally and be aware of internal coupling.

(f) Principle: Maximize Service Autonomy

Statement: Maximize service autonomy (completeness) to enable share-ability and reusability (External & Internal).

(g) Principle: Include non-functional requirements

Statement**:** Non-functional requirements form a key input in design decisions.

VI. Technology Architecture

139. The Statistical Network Business Architecture provides the following definition of Technology Architecture:

Technology Architecture (TA) describes the IT infrastructure required to support the deployment of business services, data services and applications services, including hardware, middleware, networks, platforms, etc.

140. Within each statistical organization, there needs to be an infrastructural environment in which the generic services can be combined and configured to run as element of organization specific processes. This environment is not part of CSPA. CSPA assumes that each statistical organization has such an environment and makes statements about the characteristics and capabilities that such a platform must have in order to be able to accept and run Statistical Services that comply with CSPA.

141. Platform for Service Communication**:** A communication platform provides the capability for communication between Statistical Services. It enables inter-service communication while allowing Statistical Services to remain autonomous and adds additional capabilities for monitoring and orchestrating the information flow. To assemble a built Statistical Service, the communication platform is updated to integrate with new services. There are multiple ways of establishing a communication platform. Examples of architectural components could be BPMS, ESB, Workflow Engines, Orchestration Engines, Message Queuing and Routing.

142. Platform for Configuring and Controlling Services and Processes**:** The Platform for Controlling Service and Process execution encompasses the functionalities and tools to support the management and maintenance of services metadata, artifacts and policies. Examples of how this mechanism could be achieved include Business Process Modelling System, Lifecycle Management, Service Monitoring and Management.

143. Platform for Reporting on Services and Processes**:** The Platform for reporting is responsible for enabling real-time monitoring and near-real-time presentation of user defined business key performance indicators (KPIs). Examples of how this mechanism could be achieved are Static Dashboard or Business Activity Monitoring (also generates alerts and notifications to user when these KPIs cross specified thresholds).

A. Communication Platform

144. CSPA provides guidance on the way that organizations should go about building new or wrapping existing Statistical Services. When the time comes for an organization to use a Statistical Service that conforms to CSPA there are some organization specific technology approaches that also need consideration.

145. CSPA does not specify how organizations will coordinate the use of Statistical Services to implement a wider business process. Organizations will need a technology solution to support communication between Statistical Services since the Statistical Services are not to talk directly to each other.

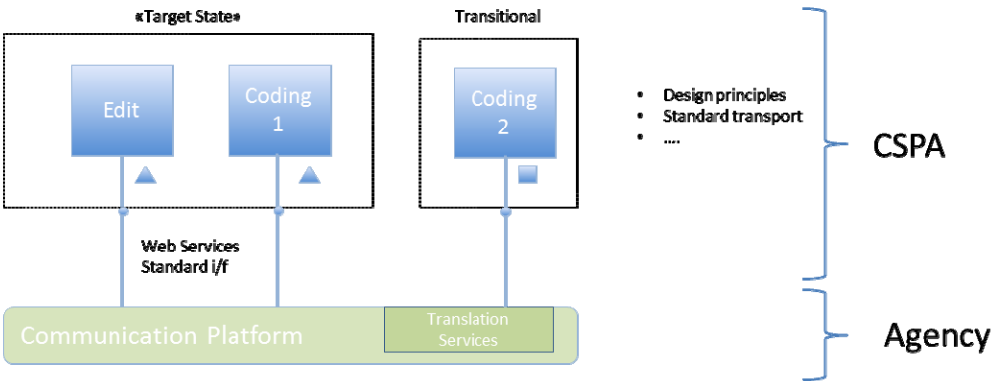
146. Where the Statistical Service being used is largely independent and interfaces between that Statistical Service and others it maybe manually managed by a person. There may also be other relatively trivial uses of Statistical Service where a bespoke solution to integrate them is developed. These are sub-optimal yet pragmatic ways of achieving reuse of Statistical Services

147. Where the integration of Statistical Services is non-trivial a communications platform of some sort will usually be required. The key functions of the communication platform are:

- Orchestration - managing the sequence of flow of invocations of the Statistical Services.
- Error handling - where Statistical Service fail or where the output of services contain erroneous cases that require a different treatment;
- Message payload translation - in particular where a Statistical Service does not support standard GSIM implementation objects - It is possible to offload this function to specialized Statistical Service;
- Auditing, Logging, Activity Monitoring;
- Performance Management;
- Security

148. Figure 8 illustrates the relationship between the elements that are specified in CSPA and the underlying Communications Platform that is local to an Organization.

Figure 8
Statistical Service Components and communications platform



149. In this diagram, two Statistical Services (Edit, Coding 1) have been defined and specified in compliance with CSPA, and their implementations are communicating with each other within the environment of a statistical organization. The Statistical Service instances communicate with each other through the organization’s communication platform – this may be a full SOA implementation (bus or broker), a CORE implementation, or some other more rudimentary platform (or no platform at all).

150. It is important to state that CSPA does not prescribe the capabilities and architecture of the underlying Communications Platform – it instead assumes that an organization’s Assemblers and Configurers will be responsible for addressing how the platform supports

the use of CSPA-compliant Statistical Services. This allows CSPA and its Statistical Services to be used by the widest possible community amongst statistical organizations, all of who may be in different stages of development and modernization.

151. In the diagram one can see that there is a second Coding Statistical Service (Coding 2) that doesn’t (yet) implement the complete CSPA Statistical Service Specification due to a transitional state. An organization may optionally make use of some form of translation service to address differences between their interfaces (typically at the information encoding level). This is seen as a transitional state – the goal is to ensure that all services adhere to the Statistical Service Specification, while allowing for differences at the Statistical Service implementation level at the protocol level (and underlying platforms).

Annex

List of Abbreviations

BPMS:	Business Process Management System
CORE:	Common Reference Environment
CSPA:	Common Statistical Production Architecture
DDI:	Data Documentation Initiative
GSBPM:	Generic Statistical Business Process Model
GSIM:	Generic Statistical Information Model
HLG:	High Level Group for the Modernisation of Statistical Production and Services
JMS:	Java Messaging Service
MSMQ:	Microsoft Message Queue
OS:	Operating System
SDMX:	Standard Data and Metadata eXchange
SOA:	Service Orient Architecture
TOGAF:	The Open Group Architectural Framework
VLAN:	Virtual Local Area Network