RAIRD Information Model RIM Annexes v1_0

Annex A: Glossary

Model	Object	Group	Definition	Explanatory Text	Synonyms
RIM v1.0	**Access Agreement**		The terms and conditions agreed between the provider of confidential data and the **Research Institution** allowing access to data.		
RIM v1.0	**Access Request**		A request to obtain access to confidential data.		
RIM v1.0	**Accreditation Request**		An official statement asking to be approved as a **Research Institution** whose members may access confidential data in some form for research purposes.		
GSIM v1.1	Administrative Details	Base	A placeholder for extensions to the model based on an organization's administrative needs.	The *Administrative Details* object is designed to act as a 'placeholder' to allow for future extensions to the existing model. It allows for further information to be added about the administrative details required to maintain the other objects outlined by GSIM.	

GSIM v1.1	Administrative Register	Exchange	A source of administrative information which is obtained from an external organization (or sometimes from another department of the same organization)	The *Administrative Register* is a source of administrative information obtained from external organizations. The *Administrative Register* would be provided under a *Provision Agreement* with the supplying organization. This administrative information is usually collected for an organization's operational purposes, rather than for statistical purposes.	
GSIM v1.1	Agent	Base	An actor that performs a role in relation to the statistical *Business Process.*	An *Agent* may be either an *Organization* or an *Individual*. An *Organization* may be an entire organization or entities within a larger organization, such as departments or divisions. An *Organization* may have sub *Agents,* which may be either other *Organizations* within the parent *Organization* or *Individuals* that belong to that *Organization*.	
GSIM v1.1	Agent Role	Base	The function or activities of an *Agent*, in regard to their involvement in the statistical *Business Process*.	An *Agent Role* may apply to either type of *Agent* - an *Organization* or *Individual*. A common example would be to identify which individuals or departments within an organization provide administrative data.	
RIM v1.0	**Analysis Data Set**		The data as selected and extracted from the **Event History Data Store**.		

GSIM v1.1	Assessment	Business	The result of the analysis of the quality and effectiveness of any activity undertaken by a statistical organization and recommendations on how these can be improved.	An *Assessment* can be of a variety of types. One example may include a gap analysis, where a current state is determined along with what is needed to reach its target state. Alternately, an *Assessment* may compare current processes against a set of requirements, for example a new *Statistical Need* or change in the operating environment. An *Assessment* can use various information objects as inputs, whether they are the main objects that the *Assessment* is about or auxiliary information objects that help accomplish the *Assessment*.	
GSIM v1.1	Attribute Component	Structures	The role given to a *Represented Variable* in the context of a *Data Structure,* which supplies information other than identification or measures.	For example the publication status of an observation (e.g. provisional, final, revised)	
GSIM v1.1	Business Case	Business	A proposal for a body of work that will deliver outputs designed to achieve outcomes. A *Business Case* will provide the reasoning for undertaking a *Statistical Support Program* to initiate a new *Statistical Program Design* for an existing *Statistical Program,* or an entirely new *Statistical Program*, as well as the details of the change proposed.	A *Business Case* is produced as a result of a detailed consideration of a *Change Definition*. It sets out a plan for how the change described by the *Change Definition* can be achieved. A *Business Case* usually comprises various evaluations. The *Business Case* will specify the stakeholders that are impacted by the *Statistical Need* or by the different solutions that are required to implement it.	

GSIM v1.1	Business Function	Business	Something an enterprise does, or needs to do, in order to achieve its objectives.	A *Business Function* delivers added value from a business point of view. It is delivered by bringing together people, processes and technology (resources), for a specific business purpose. *Business Functions* answer in a generic sense "What business purpose does this *Business Service* or *Process Step* serve?" Through identifying the *Business Function* associated with each *Business Service or Process Step* it increases the documentation of the use of the associated *Business Services* and *Process Steps*, to enable future reuse. A *Business Function* may be defined directly with descriptive text and/or through reference to an existing catalogue of *Business* GSIM v1.1*Functions*. The phases and sub processes defined within GSBPM can be used as an internationally agreed basis for cataloguing high level *Business Functions*. A catalogue might also include *Business Functions* defined at a lower level than "sub process". For example, "Identify and address outliers" might be catalogued as a lower level *Business Function* with the "Review, validate and edit" function (5.3) defined within GSBPM.	

GSIM v1.1	Business Process	Business	The set of *Process Steps* to perform one of more *Business Functions* to deliver a *Statistical Program Cycle or Statistical Support Program*.	For example, a particular *Statistical Program Cycle* might include several data collection activities, the corresponding editing activities for each collection and the production and dissemination of final outputs. Each of these may be considered separate *Business Processes* for the *Statistical Program Cycle*.	
GSIM v1.1	Business Service	Business	A means of performing a *Business Function* (an ability that an organization possesses, typically expressed in general and high level terms and requiring a combination of organization, people, processes and technology to achieve).	A *Business Service* may provide one means of accessing a particular *Business Function*. The operation of a *Business Service* will perform one or more *Business Processes*. The explicitly defined interface of a *Business Service* can be seen as representing a "service contract". If particular inputs are provided then the service will deliver particular outputs in compliance within specific parameters (for example, within a particular period of time). Note: The interface of a *Business Service* is not necessarily IT based. For example, a typical postal service will have a number of service interfaces: • Public letter box for posting letters • Counter at post office for interacting with postal workers	

GSIM v1.1	Category	Concepts	A *Concept* whose role is to extensionally define and measure a characteristic.	*Categories* for the *Concept* of sex include: Male, Female Note: An extensional definition is a description of a *Concept* by enumerating all of its sub ordinate *Concepts* under one criterion or sub division. For example - the Noble Gases (in the periodic table) is extensionally defined by the set of elements including Helium, Neon, Argon, Krypton, Xenon, Radon. (ISO 1087-1)	class
GSIM v1.1	Category Item	Concepts	An element of a *Category Set.*	A type of *Node* particular to a *Category Set* type of *Node Set*. A *Category Item* contains the meaning of a *Category* without any associated representation.	
GSIM v1.1	Category Set	Concepts	A list of *Categories*	A *Category Set* is a type of *Node Set* which groups *Categories* through the use of *Category Items*. The *Categories* in a *Category Set* typically have no assigned *Designations* (*Codes*). For example: Male, Female	

GSIM v1.1	Change Definition	Business	A structured, well-defined specification for a proposed change.	A related object - the *Statistical Need* - is a change expression as it has been received by an organization. A *Statistical Need* is a raw expression of a proposed change, and is not necessarily well-defined. A *Change Definition* is created when a *Statistical Need* is analyzed by an organization, and expresses the raw need in well-defined, structured terms. A *Change Definition* does not assess the feasibility of the change or propose solutions to deliver the change - this role is satisfied by the *Business Case* object. The precise structure or organization of a *Change Definition* can be further specified by rules or standards local to a given organization. It also includes the specific *Concepts* to be measured and the *Population* that is under consideration. Once a *Statistical Need* has been received, the first step is to do the conceptual work to establish what it is we are trying to measure. The final output of this conceptual work is the *Change Definition.* The next step is to assess how we are going to make the measurements - to design a solution and put forward a proposal for a body of work that will deliver on the requirements of the original *Statistical Need*	

GSIM v1.1	Classification Family	Concepts	A *Classification Family* is a group of *Classification Series* related from a particular point of view. The *Classification Family* is related by being based on a common *Concept* (e.g. economic activity).	Different classification databases may use different types of *Classification Families* and have different names for the families, as no standard has been agreed upon.	
GSIM v1.1	Classification Index	Concepts	A *Classification Index* is an ordered list (alphabetical, in code order etc.) of Cla*ssification Index Entries.* A *Classification Index* can relate to one particular or to several *Statistical Classifications*.	A *Classification Index* shows the relationship between text found in statistical data sources (responses to survey questionnaires, administrative records) and one or more *Statistical Classifications*. A *Classification Index* may be used to assign the codes for Classification Items to observations in statistical collections. A *Statistical Classification* is a subtype of *Node Set*. The relationship between *Statistical Classification* and *Classification Index* can also be extended to include the other *Node Set* types - *Code List* and *Category Set*.	

GSIM v1.1	Classification Index Entry	Concepts	A *Classification Index Entry* is a word or a short text (e.g. the name of a locality, an economic activity or an occupational title) describing a type of object/unit or object property to which a *Classification Item* applies, together with the code of the corresponding *Classification Item*. Each *Classification Index Entry* typically refers to one item of the *Statistical Classification*. Although a *Classification Index Entry* may be associated with a *Classification Item* at any *Level* of a *Statistical Classification*, *Classification Index Entries* are normally associated with items at the lowest *Level*.	A *Classification Item* is a subtype of *Node*. The relationship between *Classification Item* and *Classification Index Entry* can also be extended to include the other *Node* types - *Code Item* and *Category Item*.	
GSIM v1.1	Classification Item	Concepts	A *Classification Item* represents a *Category* at a certain *Level* within a *Statistical Classification*. It defines the content and the borders of the *Category*. A *Unit* can be classified to one and only one item at each Level of a *Statistical Classification*.		
GSIM v1.1	Classification Series	Concepts	A *Classification Series* is an ensemble of one or more *Statistical Classifications*, based on the same concept, and related to each other as versions or updates. Typically, these *Statistical Classifications* have the same name (for example, ISIC or ISCO).		
GSIM v1.1	Code	Concepts	A *Designation* for a *Category.*	*Codes* are unique within their *Code List*. Example: M (Male) F (Female).	

GSIM v1.1	Code Item	Concepts	An element of a *Code List*.	A type of *Node* particular to a *Code List* type of *Node Set*. A *Code Item* combines the meaning of the included *Category* with a *Code* representation.	
GSIM v1.1	Code List	Concepts	A list of *Categories* where each *Category* has a predefined *Code* assigned to it.	A kind of *Node Set* for which the *Category* contained in each *Node* has a *Code* assigned as a *Designation*. For example:1 - Male2 - Female	
GSIM v1.1	Code Value	Concepts	An alpha-numeric string used to represent a *Code*.	A *Code Value* is a subtype of *Sign* - a way of denoting the value of a *Code*. This is a kind of *Sign* used for *Codes*.	
GSIM v1.1	Concept	Concepts	Unit of thought differentiated by characteristics.		
GSIM v1.1	Concept System	Concepts	Set of *Concepts* structured by the relations among them.	Here are 2 examples 1) Concept of Sex: Male, Female, Other 2) ISIC (the list is too long to write down)	
GSIM v1.1	Conceptual Domain	Concepts	Set of valid *Concepts*.	The *Concepts* can be described by either enumeration or by an expression.	
RIM v1.0	**Corrective Action**		An action taken to reduce the disclosure risk of a *Data Set.*	These actions may include cell suppression, aggregation, techniques for detecting differential disclosure, etc. Transformation of the data is typical of such actions.	

GSIM v1.1	Correspondence Table	Concepts	A Correspondence Table expresses the relationship between two Statistical Classifications. These are typically: two versions from the same Classification Series; Statistical Classifications from different Classification Series; a variant and the version on which it is based; or, different versions of a variant. In the first and last examples, the Correspondence Table facilitates comparability over time. Correspondence relationships are shown in both directions.	A Statistical Classification is a subtype of Node Set. The relationship between Statistical Classification and Correspondence Table can also be extended to include the other Node Sets - Code List and Category Set.	
RIM v1.0	**Data Catalogue**		The set of metadata describing the data contained in the **Event History Data Store** as presented to researchers so they can browse the content of the store.		
GSIM v1.1	Data Point	Structures	A placeholder (or cell) for the value of an *Instance Variable*	Field in a *Data Structure* which corresponds to a cell in a table. The *Data Point* is structural and distinct from the value (the *Datum*) that it holds.	
GSIM v1.1	Data Resource	Structures	An organized collection of stored information made of one or more *Data Sets.*	*Data Resources* are collections of data that are used by a statistical activity to produce information. *Data Resource* is a specialization of an *Information Resource* .	data source

GSIM v1.1	Data Set	Structures	An organized collection of data.	Examples of *Data Sets* could be observation registers, time series, longitudinal data, survey data, rectangular data sets, event-history data, tables, data tables, cubes, registers, hypercubes, and matrixes. A broader term for *Data Set* co uld be data. A narrower term for *Da ta Set* could be data element, data record, cell, field.	database, data file, file, table
GSIM v1.1	Data Structure	Structures	Defines the structure of an organized collection of data (*D ata Set*).	The structure is described using *Dat a Structure Components* that can be either *Attribut e Components, Identifier Components* or *Mea sure Components*. Examples for unit data include social security number, country of residence, age, citizenship, country of birth, where the social security number and the country of residence are both identifying components and the others are measured variables obtained directly or indirectly from the person (*Uni t*).	

GSIM v1.1	Data Structure Component	Structures	The role of the *Represented Variable* in the context of a *Data Structure.*	A *Data Structure Component* can be an *Attribute Component, Measure Component* or an *Identifier Component*. Example of *Attribute Component*: The publication status of an observation such as provisional, revised. Example of *Measure Component*: age and height of a person in a *Unit Data Set* or number of citizens and number of households in a country in a *Data Set* for multiple countries (*Dimensional Data Set*). Example of *Identifier Component*: The personal identification number of a Swedish citizen for unit data or the name of a country in the European Union for dimensional data.	
GSIM v1.1	Datum	Concepts	A value.	A *Datum* is the actual instance of data that was collected or derived. It is the value which populates a *Data Point*. A *Datum* is the value found in a cell of a table.	value
RIM v1.0	**Datum-Based Data Set**		A *Data Set* which records a set of observations specific to time-bound events, for a range of variables, structured according to a **Datum-Based Data Structure**.	Each observation could become a single row in the Data Set. Each row would then have a value for which a *Unit* identifier, a variable reference (indicating the type of event), and an **Event Period** are recorded.	
RIM v1.0	**Datum-Based Data Structure**		The description of how a *Data Set* is arranged, in which a set of time-bound events or statuses relative to events are recorded, for a known set of *Units*.	For each *Unit*, one or more event-statuses will be recorded. Each record in the *Data Set* could identify a *Unit*, make reference to a variable indicating the event, hold the value of a status, and indicate the point or period in time for which the value is relevant.	

GSIM v1.1	Described Conceptual Domain	Concepts	A *Conceptual Domain* defined by an expression.	For example: All real numbers between 0 and 1.	Non-enumerated conceptual domain
GSIM v1.1	Described Value Domain	Concepts	A *Value Domain* defi ned by an expression.	For example: All real decimal numbers between 0 and 1.	Non-enumerated value domai
GSIM v1.1	Designation	Concepts	The name given to an object for identification.	The association of a *Concept* with a *Sign* that denotes it.	
GSIM v1.1	Dimensional Data Point	Structures	A placeholder (or cell) for the value of an *Instance Variable* with respect to either a *Unit* or *Population*.	A *Dimensional Data Point* is uniquely identified by the combination of exactly one value for each of the dimensions (*Identifie r Component*) and one measure (*Meas ure Component*). There may be multiple values for the same *Dimension al Data Point* that is for the same combination of dimension values and the same measure. The different values represent different versions of the data in the *Data Point*. Values are only distinguished on the basis of quality, date/time of measurement or calculation, status, etc. This is handled through the mechanisms provided by the *Datu m* information object.	cell
GSIM v1.1	Dimensional Data Set	Structures	A collection of dimensional data that conforms to a known structure.		
GSIM v1.1	Dimensional Data Structure	Structures	Describes the structure of a *Dimen sional Data Set*.	For example (city, average income, total population) where the city is the *Identifier Component* and the others are measured variables.	
RIM v1.0	**Disclosive Data Point**		A *Data Point* which is potentially disclosive on its own.	Such aspects as the discoverability or sensitivity of the value may make it potentially disclosive.	
RIM v1.0	**Disclosive Data Point Combination**		A combination of *Dat a Points* which, when taken together, pose a potential disclosure risk.	The *Data Points* by themselves could be disclosive or non-disclosive.	

RIM v1.0	**Disclosure Control Rule**		A *Rule* based on a *Process Method* which determines if a **Provisional Output** is disclosive.		
RIM v1.0	**Disclosure Risk Assessment**		The level of disclosure risk, determined by an assessment of both the assigned risk factors and those arising from the data taken in context.	Assigned risk factors are attributes of the data such as discoverability or sensitivity; contextual factors are attributes such as skewness or granularity.	
GSIM v1.1	Enumerated Conceptual Domain	Concepts	A *Conceptual Domain* expressed as a list of *Categories*	For instance, the Sex *Categories*: 'Male' and 'Female'	
GSIM v1.1	Enumerated Value Domain	Concepts	A *Value Domain* expressed as a list of *Categories* and associated *Codes.*	Example - Sex Codes <m, male>; <f, female>; <o, other>.	
GSIM v1.1	Environment Change	Business	A requirement for change that originates from a change in the operating environment of the statistical organization.	An *Environment Change* reflects change in the context in which a statistical organization operates. *Environment Changes* can be of different origins and also take different forms. They can result from a precise event (budget cut, new legislation enforced) or from a progressive process (technical or methodological progress, application or tool obsolescence). Other examples of *Environment Changes* include the availability of a new *Information Resource*, the opportunity for new collaboration between organizations, etc.	
RIM v1.0	**Event History Data Store**		A system which provides access to a set of Event History data, structured according to a **Datum-Based Data Structure**.		

RIM v1.0	**Event History Input Data Set**		The data which is loaded into an **Event History Data Store**.	Typically, this will be structured according to a **Datum-Based Data Structure**; it is accompanied by an **Input Metadata Set** for load into the **Event History Data Store**.	
RIM v1.0	**Event Period**		The point of time at which an event occurs, or the period of time during which the status of a *Unit* persists in regards to that event.	If an event or status is observed at a single point in time, then that point in time is the event period. When a change in status occurs, an end-time is recorded, and a new event status begins.	
GSIM v1.1	Exchange Channel	Exchange	A means of exchanging data.	An abstract object that describes the means to receive (data collection) or send (dissemination) information. Different *Exchange Channels* are used for collection and dissemination. Examples of collection *Exchange Channel* include *Questionnaire*, *Web Scraper Channel* and *Administrative Register*. The only example of a dissemination *Exchange Channel* currently contained in GSIM is *Product*. Additional *Exchange Channels* can be added to the model as needed by individual organizations.	
RIM v1.0	**Final Output**		Result of analysis on the **Analysis Data Set** that has been subject to disclosure control, and is deemed safe to be viewed and published by a researcher.		
GSIM v1.1	Identifiable Artefact	Base	An abstract class that comprises the basic attributes and associations needed for identification, naming and other documentation.	An instance of any GSIM information object is an *Identifiable Artifact*.	

GSIM v1.1	Identifier Component	Structures	The role given to a *Represented Variable* in the context of a *Data Structure* to identify the unit in an organized collection of data.	An *Identifier Component* is a sub-type of *Data Structure Component*. The personal identification number of a Swedish citizen for unit data or the name of a country in the European Union for dimensional data.	
GSIM v1.1	Individual	Base	A person who acts, or is designated to act towards a specific purpose.		
GSIM v1.1	Information Consumer	Exchange	A person or organization that consumes disseminated data.	The *Information Consumer* accesses a set of information via a *Product* (or potentially via another Exchange Channel), which contains one or more *Presentations.* The *Information Consumer's* access to the information is subject to a *Provision Agreement*, which sets out conditions of access.	
GSIM v1.1	Information Provider	Exchange	An *Individual* or *Organization* that provides collected information.	An *Information Provider* possesses sets of information (that it has generated, collected, produced, bought or otherwise acquired) and is willing to supply that information (data or referential metadata) to the statistical office. The two parties use a *Provision Agreement* to agree the *Data Structure* and *Referential Metadata Structure* of the data to be exchanged via an *Exchange Channel*.	information supplier, data supplier

GSIM v1.1	Information Request	Business	An outline of a need for new information required for a particular purpose.	An *Information Request* is a special case of *Statistical Need* that may come in an organized form, for example by specifying on which *Subject Field* the information is required. It may also be a more general request and require refinement by the statistical agency and formalized in a *Change Definition*.	
GSIM v1.1	Information Resource	Structures	An abstract notion that is any organized collection of information.	There currently are only two concrete sub classes: *Data Resource* and *Referential Metadata Resource*. The *Information Resource* allows the model to be extended to other types of resource.	
GSIM v1.1	Information Set	Structures	Organized collections of statistical content.	Statistical organizations collect, process, analyze and disseminate *Information Sets*, which contain data (*Data Sets*), referential metadata (*Referential Metadata Sets*), or potentially other types of statistical content, which could be included in addition types of *Information Set*.	
RIM v1.0	**Input Metadata Set**		The set of metadata needed to fully document and describe the structure of an **Event History Input Data Set.**	The metadata consists of *Concepts*, *Codelists* and *Statistical Classifications*, variable descriptions, and higher-level metadata regarding how the accompanying data can be used by researchers.	
GSIM v1.1	Instance Question	Exchange	The use of a *Question* in a particular *Questionnaire*.	The Instance Question is the use of a *Question* in a particular *Questionnaire Component*. This also includes the use of the *Question* in a *Question Block*, which is a particular type of *Questionnaire Component*.	

GSIM v1.1	Instance Question Block	Exchange	The use of a *Question Block* in a particular *Questionnaire*.	The *Instance Question Block* is the use of a *Question Block* in a particular *Questionnaire Component*. This also includes the use of a *Question Block* in another *Question Block*, as it is a particular type of *Questionnaire Component*.	
GSIM v1.1	Instance Statement	Exchange	The use of a *Statement* in a particular *Questionnaire*.	The *Instance Statement* is the use of a *Statement* in a particular *Questionnaire Component*. This also includes the use of the *Statement* in a *Question Block*, which is a particular type of *Questionnaire Component*.	
GSIM v1.1	Instance Variable	Concepts	The use of a *Represented Variable* within a *Data Set*. It may include information about the source of the data.	The *Instance Variable* is used to describe actual instances of data that have been collected. Here are 3 examples:1) Gender: Dan Gillman has gender <m, male>, Arofan Gregory has gender<m, male>, etc.2) Number of employees: Microsoft has 90,000 employees; IBM has 433,000 employees, etc.3) Endowment: Johns Hopkins has endowment of <3, $1,000,000 and above>,Yale has endowment of <3, $1,000,000 and above>, etc.	
GSIM v1.1	Level	Concepts	A *Statistical Classification* has a structure which is composed of one or several *Levels*. A *Level* often is associated with a *Concept*, which defines it. In a hierarchical classification the *Classification Items* of each *Level* but the highest are aggregated to the nearest higher *Level*. A linear classification has only one *Level*.	A *Statistical Classification* is a subtype of *Node Set*. The relationship between *Statistical Classification* and *Level* can also be extended to include the other *Node Set* types - *Code List* and *Category Set*.	

GSIM v1.1	Logical Record	Structures	Describes a type of *Unit Data Record* for one *Unit Type* within a *Unit Data Set*.	Examples: household, person or dwelling record.	
GSIM v1.1	Map	Concepts	A *Map* is an expression of the relation between a *Classification Item* in a source *Statistical Classification* and a corresponding *Classification Item* in the target *Statistical Classification*. The *Map* should specify whether the relationship between the two *Classification Items* is partial or complete. Depending on the relationship type of the *Correspondence Table*, there may be several *Maps* for a single source or target item.	The use of *Correspondence Tables* and *Maps* can be extended to include all types of *Node* and *Node Set*. This means that a *Correspondence Table* could map between the items of *Statistical Classifications*, *Code Lists* or *Category Sets*.	
GSIM v1.1	Measure Component	Structures	The role given to a *Represented Variable* in the context of a *Data Structure* to hold the observed/derived values for a particular *Unit* in an organized collection of data.	A *Measure Component* is a sub-type of *Data Structure* Component. For example age and height of a person in a *Unit Data Set* or number of citizens and number of households in a country in a *Data Set* for multiple countries (*Dimensional Data Set*).	
GSIM v1.1	Node	Concepts	A combination of a *Category* and related attributes.	A *Node* is created as a *Category*, *Code* or *Classification Item* for the purpose of defining the situation in which the *Category* is being used.	
GSIM v1.1	Node Set	Concepts	A set of *Nodes.*	*Node Set* is a kind of *Concept System*. Here are 2 examples: 1) Sex *Categories* • Male • Female • Other 2) Sex *Codes* • <m, male> • <f, female> • <o, other>	
RIM v1.0	**Nondisclosive Data Point**		A *Data Point* which is not by itself potentially disclosive.		

GSIM v1.1	Organization	Base	A unique framework of authority within which a person or persons act, or are designated to act, towards some purpose.		
GSIM v1.1	Output Specification	Exchange	Defines how *Information Sets* consumed by a *Product* are presented to *Information Consumers*.	The *Output Specification* specifies *Products* and defines the *Presentations* they contain. The *Output Specification* may be fully defined during the design process (such as in a paper publication or a predefined web report), or may be a combination of designed specification supplemented by user selections (such as in an online data query tool).	
GSIM v1.1	Population	Concepts	The total membership of a defined class of people, objects or events.	A population is used to describe the total membership of a group of people, objects or events based on characteristics, e.g. time and geographic boundaries. Here are 3 examples –1. US adult persons on 13 November 19562. US computer companies at the end of 20123. Universities in the US 1 January 2011.	

GSIM v1.1	Presentation	Exchange	The way data and referential metadata are presented in a *Product*.	A *Product* has one or more *Presentations,* which present data and referential metadata from *Information Sets*. A *Presentation* is defined by an *Output Specification*. *Presentation* can be in different forms; e.g. tables, graphs, structured data files.Examples: • A table of data. Based on a *Data Set*, the related *Data Structure* is used to label the column and row headings for the table. The *Data Set* is used to populate the cells in the table. Reference metadata is used to populate footnotes and cell notes on the table. Confidentiality rules are applied to the *Data Set* to suppress any disclosive cells. • A data file based on a standard (e.g. SDMX). • A PDF document describing a *Classification*. • Any structural metadata object expressed in a standard format (e.g. DDI 3.1 XML). • A list of *Products* or services (e.g. a product catalogue or a web services description language (WSDL) file). • A web page containing *Classifications*, descriptions of *Variables*, etc.	

GSIM v1.1	Process Control	Business	A set of decision points which determine the flow between the *Process Steps* used to perform a *Business Process*.	The typical use of *Process Control* is to determine what happens next after a *Process Step* is executed. The possible paths, and the decision criteria, associated with a *Process Control* are specified as part of designing a production process, captured in a *Process Control Design*. There is typically a very close relationship between the design of a process and the design of a *Process Control*.	

GSIM v1.1	Process Control Design	Business	The specification of the decision points required during the execution of a *Business Process*.	The design of a *Process Control* typically takes place as part of the design of the process itself. This involves determining the conditional routing between the various sub-processes and services used by the executing process associated with the *Process* Control and specified by the *Process Control Design.* It is possible to define a *Process Control* where the next step in the *Process Step* that will be executed is a fixed value rather than a "choice" between two or more possibilities. Where such a design would be appropriate, this feature allows, for example, initiation of a step in the Process *Step* representing the GSBPM Process Phase (5) to always lead to initiation of GSBPM sub-process Integrate Data (5.1) as the next step. This allows a process designer to divide a *Business Process* into logical steps (for example, where each step performs a specific *Business Function* through re-use of a *Business Service*) even if these process GSIM v1.1steps will always follow each other in the same order. In all cases, the *Process Control Design* defines and the *Process Control* manages the flow between *Process Steps*, even where the flow is "trivial". *Process Design* is left to focus entirely on the design of the process itself, not sequencing between steps.	

GSIM v1.1	Process Design	Business	The specification of how a *Process Step* will be performed. This includes specifying the types of *Process Inputs* required and the type of *Process Outputs* that will be produced.	A *Process Design* is the design time specification of a *Process Step* that is performed as part of a run-time *Business Service*. A *Process Step* can be as big or small as the designer of a particular *Business Service* chooses. From a design perspective, one *Process Step* can contain "sub-steps", each of which is conceptualized as a (smaller) *Process Step* in its own right. Each of those "sub-steps" may contain "sub-steps" within them and so on indefinitely. It is a decision for the process designer to what extent to subdivide steps. At some level it will be appropriate to consider a *Process Step* to be a discrete task (unit of work) without warranting further subdivision. At that level the *Process Step* is designed to process particular *Process Inputs,* according to a particular *Process Method*, to produce particular *Process Output*s. The flow between a *Process Step* and any sub steps is managed via *Process Control*.	
GSIM v1.1	Process Input	Business	Any instance of an information object which is supplied to a *Process Step Instance* at the time its execution is initiated.	*Process Input* might include information that is used as an input that will be transformed (e.g. a *Data Set*), information that is used to control specific parameters of the process (e.g. a *Rule*), and information that is used as reference to guide the process (e.g. a *Code List*).	

GSIM v1.1	Process Input Specification	Business	A record of the types of inputs required for a *Process Design*.	The *Process Input Specification* enumerates the *Process Inputs* required at the time a *Process Design* is executed. For example, if five different *Process Inputs* are required, the *Process Input Specification* will describe each of the five inputs. For each required *Process Input* the *Process Input Specification* will record the type of information object (based on GSIM) which will be used as the *Process Input* (example types might be a *Dimensional Data Set* or a *Classification*). The *Process Input* to be provided at the time of *Process Step* execution will then be a specific instance of the type of information object specified by the *Process Input Specification*. For example, if a *Process Input Specification* requires a *Dimensional Data Set* then the corresponding *Process Input* provided at the time of *Process Step* execution will be a particular *Dimensional Data Set.*	
GSIM v1.1	Process Method	Business	A specification of the technique which will be used to perform the unit of work.	The technique specified by a *Process Method* is independent from any choice of technologies and/or other tools which will be used to apply that technique in a particular instance. The definition of the technique may, however, intrinsically require the application of specific *Rules* (for example, mathematical or logical formulas). A *Process Method* describes a particular method for performing a *Process Step.*	

GSIM v1.1	Process Output	Business	Any instance of an information object which is produced by a *Process Step* as a result of its execution.	*Process Outputs* have an attribute of Process Output Type, which has two possible values: • *Transformed Output* is the result which provides the "reason for existence" of the *Process Step*. If that output were no longer required then there would be no need for the *Process Step* in its current form. Typically a *Transformed Output* is either a *Process Input* to a subsequent *Process Step* or it represents the final product from a statistical business process. • A *Process Metric* records information about the execution of a *Process Step*. For example, how long it took to complete execution of the *Process Step* and what percentage of records in the *Process Input* was updated by the *Process Step* to produce the *Transformed Output*.	

GSIM v1.1	Process Output Specification	Business	A record of the types of outputs required for a Process Design.	The *Process Output Specification* enumerates the *Process Outputs* that are expected to be produced at the time a *Process Design* is executed. For example, if five different *Process Outputs* expected, the *Process Output Specification* will describe each of the five outputs. For each expected *Process Output* the *Process Output Specification* will record the type of information object (based on GSIM) which will be used as the *Process Output* (Example types might be a *Dimensional Data Set* or a *Classification*). The *Process Output* to be provided at the time of *Process Step* execution will then be a specific instance of the type of information object specified by the *Process Output Specification*. For example, if a *Process Output Specification* expects a *Dimensional Data Set* then the corresponding *Process Output* provided at the time of *Process Step* execution will be a particular *Dimensional Data Set*.	

GSIM v1.1	Process Pattern	Business	A nominated set of *Process Designs*, and associated *Process Control Designs* (flow), which have been highlighted for possible reuse.	In a particular *Business Process*, some *Process Steps* may be unique to that *Business Process* while others may be applicable to other *Business Processes*. A *Process Pattern* can be seen as a reusable template. It is a means to accelerate design processes and to achieve sharing and reuse of design patterns which have proved effective. Reuse of *Process Patterns* can indicate the possibility to reuse related *Business Services.* By deciding to reuse a *Process Pattern*, a designer is actually reusing the pattern of *Process Designs* and *Process Control Designs* associated with that *Process Pattern*. They will receive a new instance of the Process *Designs* and *Process Control Designs*. If they then tailor their "instance" of the *Process Designs* and *Process Control Designs* to better meet their needs they will not change the definition of the reusable *Process Pattern*.	
GSIM v1.1	Process Step	Business	A *Process Step* is a work package that performs a *Business Process*. A *Process Step* implements the *Process Step Design* specified in order to produce the outputs for which the *Process Step* was designed.	Each *Process Step* is the use of a *Process Step Design* in a particular context (e.g. within a specific *Business Process*). At the time of execution a *Process Step Instance* specifies the actual instances of input objects (for example, specific *Data Sets*, specific *Variables*) to be supplied.	

GSIM v1.1	Process Step Instance	Business	An executed step in a *Business Process*. A *Process Step Instance* specifies the actual inputs to and outputs from for an occurrence of a *Process Step*.	Each *Process Step* is the use of a *Process Step Design* in a particular context (e.g. within a specific Business Process). At the time of execution a *Process Step Instance* specifies the actual instances of input objects (for example, specific Data Sets, specific Variables) to be supplied. Each *Process Step Instance* may produce unique results even though the *Process Step* remains constant. Even when the inputs remain the same, metrics such as the elapsed time to complete execution of process step may vary from execution to execution. For this reason, each *Process Step Instance* details of inputs and outputs for that instance of implementing the *Process Step*. In this way it is possible to trace the flow of execution of a *Business Process* through all the *Process Steps* which were involved.	
GSIM v1.1	Product	Exchange	A package of content that can be disseminated as a whole.	A Product is the only defined type of Exchange Channel for outgoing information. A Product packages Presentations of Information Sets for an Information Consumer. The Product and its Presentations are generated according to Output Specifications, which define how the information from the Information Sets it consumes are presented to the Information Consumer. The Protocol for a Product determines the mechanism by which the Product is disseminated (e.g. website, SDMX web	

				service, paper publication). A Provision Agreement between the statistics office and the Information Consumer governs the use of a Product by the Information Consumer. The Provision Agreement, which may be explicitly or implicitly agreed, provides the legal or other basis by which the two parties agree to exchange data. In many cases, dissemination Provision Agreements are implicit in the terms of use published by the statistics office. For static Products (e.g. paper publications), specifications are predetermined. For dynamic products, aspects of specification could be determined by the Information Consumer at run time. Both cases result in Output Specifications specifying Information Set data or referential metadata that will be	

				included in each Presentation within the Product.	
RIM v1.0	**Project**		A collaborative enterprise, involving research, which is carefully planned to achieve a particular aim.		
RIM v1.0	**Project Profile**		A set of information used to administer a **Project**.		
GSIM v1.1	Protocol	Exchange	The mechanism for exchanging information through an *Exchange Channel.*	A Protocol specifies the mechanism (e.g. SDMX web service, data file exchange, web robot, face to face interview, mailed paper form) of exchanging information through an *Exchange Channel*.	
GSIM v1.1	Provision Agreement	Exchange	The legal or other basis by which two parties agree to exchange data.	A *Provision Agreement* between the statistical organization and the *Information Provider* (collection) or the *Inf ormation Consumer* (dissemination) governs the use of *E xchange Channels*. The *Provision Agreement*, which may be explicitly or implicitly agreed, provides the legal or other basis by which the two parties agree to exchange data. The parties also use the *Provision Agreement* to agree the *Data Structure* a nd *Referential Metadata Structure* o f the information to be exchanged.	
RIM v1.0	**Provisional Output**		The product of the analysis of an **Analy sis Data Set** by a researcher, which is submitted for finalization.	If the **Provisional Output** passes disclosure control, it becomes a **Final Output** and may be published. Otherwise, it may be subjected to **Correct ive Actions** to reduce the risk of disclosure.	

GSIM v1.1	Question	Exchange	Describes the text used to elicit a response for the *Concept* to be measured.	A *Question* may be a single question used to obtain a response, or may be a multiple question, a construct which links multiple sub-questions, each with their own response. A *Question* also includes a relationship to the *Value Domain* to document the associated response criteria for the question. A single response question will have one *Value Domain* associated with it, while a 'multiple question' may have more than one *Value Domain*. A *Question* should be designed with re-use in mind, as it can be used in multiple *Questionnaires.*	Multiple Question
GSIM v1.1	Question Block	Exchange	A set of *Questions, Statements* or instructions which are used together.	A *Question Block* should be designed for reuse, as it can be used in multiple *Questionnaires*. The *Question Block* is a type of *Questionnaire Component*. A statistical organization will often have a number of *Question Blocks* which they reuse in a number of *Questionnaires*. Examples of *Question Blocks* include: • Household *Question Block* • Income *Question Block* • Employment *Question Block*	Question Module
GSIM v1.1	Questionnaire	Exchange	A concrete and usable tool to elicit information from observation *units*.	This is an example of a way statistical organizations collect information (an exchange channel). Each mode should be interpreted as a new *Questionnaire* derived from the *Questionnaire Specification*. The *Questionnaire* is a subtype of *Exchange Channel*, as it is a way in which data is obtained.	

GSIM v1.1	Questionnaire Component	Exchange	A record of the flow of a *Questionnaire Specification* and its use of *Questions, Question Blocks* and *Statements*	Defines the structure of the *Questionnaire Specification*, as a combination of *Questions, Question Blocks* and *Statements.* It is the object which groups together all the components of a *Questionnaire.* *A Questionnaire Component* is recursive, in that it can refer to other *Questionnaire Components* and accompanying *Questionnaire Logic* objects at a lower level. It is only at the top level where the *Questionnaire Component* links to the *Questionnaire Specification,*	Question Block
GSIM v1.1	Questionnaire Logic	Exchange	Governs the sequence of *Questions*, *Question Blocks* and *Statements* based on factors such as the current location, the response to the previous questions etc., invoking navigation and validation rules to apply.		Routing
GSIM v1.1	Questionnaire Specification	Exchange	The tool designed to elicit information from observation *Units*.	This represents the complete questionnaire design, with a relationship to the top level *Questionnaire Component*. There may be many different *Questionnaire Specifications*, for the same surveys, or tailored to individual observation *Units* (respondents) so that there would be a different *Questionnaire Specification* for each respondent. The design would also differ depending upon the specific mode of collection the *Questionnaire* is designed for.	

GSIM v1.1	Record Relationship	Structures	Describes relationships between *Logical Records* within a *Unit Data Structure*. It must have both a source *Logical Record* and a target *Logical Record* in order to define the relationship.	Example: Relationship between person and household *Logical Records* within a *Unit Data Set*.	
GSIM v1.1	Referential Metadata Attribute	Structures	The role given to a *Represented Variable* to supply information in the context of a *Referential Metadata Structure*.		
GSIM v1.1	Referential Metadata Content Item	Structures	The content describing a particular characteristic of a *Referential Metadata Subject*.	A *Referential Metadata Content Item* contains the actual content describing a particular characteristic of a *Referential Metadata Subject*.	
GSIM v1.1	Referential Metadata Resource	Structures	An organized collection of stored information consisting of one or more *Referential Metadata Sets*.	*Referential Metadata Resources* are collections of structured information that may be used by a statistical activity to produce information. This information object is a specialization of an *Information Resource*.	
GSIM v1.1	Referential Metadata Set	Structures	An organized collection of referential metadata for a given *Referential Metadata Subject*.	*Referential Metadata Sets* organize referential metadata. Each *Referential Metadata Set* uses a *Referential Metadata Structure* to define a structured list of *Referential Metadata Attributes* for a given *Referential Metadata Subject*.	

GSIM v1.1	Referential Metadata Structure	Structures	Defines the structure of an organized collection of referential metadata (*Referential Metadata Set*).	*A Referential Metadata Structure defines* a structured list of *Referential Metadata Attributes* f or a given *Referentia l Metadata Subject*. Examples of *Referen tial Metadata Attributes* are those that describe quality information and methodologies. Examples of subject are: objects like a *Q uestionnaire* or a *Cla ssification*, or collections of data like a *Data Set*, or any *Data Point or* set of *Data Points* creat ed from a specific *D ata Structure.*	Metadata Structure Definition
GSIM v1.1	Referential Metadata Subject	Structures	Identifies the subject of an organized collection of referential metadata.	The *Referential Metadata Subject* id entifies the subject of the metadata that can be reported using this Referential Metadata Structure. These subjects may be any GSIM object type, or any *Data Point or* set of *Data Points* created from a specific *Data Structure.* Examples: The GSIM object type may be Product for which there is a list specified in a *Value Domain*. The *Value Domain* specifies the list of actual Products for which reference metadata can be reported or authored using this *Referential Metadata Structure.*	
GSIM v1.1	Referential Metadata Subject Item	Structures	Identifies the actual subject for which referential metadata is reported.	Examples are an actual *Product* such as Balance of Payments and International Investment Position, Australia, June 2013, or a collection of *Data Points* such as the *Data Points* fo r a single region within a *Data Set* co vering all regions for a country.	

RIM v1.0	**Refusal Rationale**		The official reason provided for denying a **Research Institution** accreditation or for denying a **Project** access to confidential data.		
RIM v1.0	**Relationship**		The connection between two Units, of a specified type.	The Relationship object represents the connection of two units, for example marriage, employment, etc. It is sometimes seen in a real form (a civil contract in the case of marriage, or an employment contract in the case of an employer-employee relationship), but this is not always the case.	
GSIM v1.1	Represented Variable	Concepts	A combination of a characteristic of a population to be measured and how that measure will be represented.	Example: The pair (Number of Employees, Integer), where "Number of Employees" is the characteristic of the population (*Variable)* and "Integer" is how that measure will be represented (*Value Domain).*	
RIM v1.0	**Represented Variable**			This is a trivial extension of *Represented Variable* created for the purpose of adding disclosure control attributes, in particular sensitivity	
RIM v1.0	**Research Institution**		An institution which conducts research.		
RIM v1.0	**Research Publication**		A published document which answers a research question based on an analysis of data.		
RIM v1.0	**Researcher**		A post holder with a research title or other academic/scientific personnel at approved research units, and students studying for a masters degree or PhD, and post graduates provided that these are under the supervision of a qualified researcher at such a research unit.		

GSIM v1.1	Rule	Business	A specific mathematical or logical expression which can be evaluated to determine specific behavior.	*Rules* are of several types: they may be derived from methods to determine the control flow of a process when it is being designed and executed; they may be used as the input parameters of processes (e.g., imputation rules, edit rules); and they may be used to drive the logical flow of a questionnaire. There are many forms of *R ules* and their purpose, character and expression can vary greatly.	
GSIM v1.1	Scraping Process Map	Exchange	Maps a web scraping process to a specific website.	*Scraping Process Map* is an essential element of the *Web Scraper Channel*. The process being mapped can be a *Bu siness Service* or a *Process Step*.	
GSIM v1.1	Sign	Concepts	Something that suggests the presence or existence of a fact, condition, or quality.	It is a perceivable object. This object is used to denote a *Co ncept* as a *Designati on*.	
GSIM v1.1	Statement	Exchange	A report of facts in a *Questionnaire*	*Statements* are often included to provide further explanation to respondents. Example: "The following questions are about your health". The object is also used to represent completion instructions for the interviewer or respondent. *Statement* should be designed with re-use in mind as it can be used in numerous *Q uestionnaires*.	Interviewer Instruction Instruction

GSIM v1.1	Statistical Classification	Concepts	A *Statistical Classification* is a set of *Categories* which may be assigned to one or more variables registered in statistical surveys or administrative files, and used in the production and dissemination of statistics. The *Categories* at each *Level* of the classification structure must be mutually exclusive and jointly exhaustive of all objects/units in the population of interest.	The *Categories* are defined with reference to one or more characteristics of a particular population of units of observation. A *Statistical Classification* may have a flat, linear structure or may be hierarchically structured, such that all *Categories* at lower *Levels* are sub-*Categories* of *Categories* at the next *Level* up. *Categories* in *Statistical Classifications* are represented in the information model as *Classification Items*.	
GSIM v1.1	Statistical Need	Business	A requirement, request or other notification that will be considered by an organization. A *Statistical Need* does not necessarily have structure or format - it is a 'raw' need as received by the organization. A *Statistical Need* may be of a variety of types including *Environmental Change* or *Information Request*.	The *Statistical Need* is a proposed or imposed requirement, request or other notification as it has been received by an organization. A *Statistical Need* is a raw expression of a requirement, and is not necessarily well-defined. A related object - *Change Definition* - is created when a *Statistical Need* is analyzed by an organization. *Change Definition* expresses the raw need in well-defined, structured terms.Once a *Statistical Need* has been received, the first step is to do the conceptual work to establish what it is we are trying to measure. The final output of this conceptual work is the *Change Definition*. In some cases, the *Statistical Need* can result from the *Assessment* of the quality, efficiency, etc. of an existing process.	
GSIM v1.1	Statistical Program	Business			

			A set of activities, which may be repeated, that describes the purpose and context of a set of *Business Process* within the context of the relevant *Statistical Program Cycles.*	The *Statistical Program* is one of a family of objects that provide the environmental context in which activities to produce statistics within a statistical organization are conducted. *Statistical Program* is the top level object that describes the purpose and objectives of a set of activities. *Statistical Program* will usually correspond to an ongoing activity such as a survey or output series. Some examples of *Statistical Program* are: • Labour Force Survey - Multipurpose Household Survey - National Accounts - Demography - Overseas Arrivals and Departures Related to the *Statistical Program* object there are *Statistical Program Design* and *Statistical Program Cycle* objects that hold the detailed information about the design and conduct of the *Business Process*. In the case of the traditional approach, an organization has received a *Statistical Need* and produced a *Change Definition* and an approved *Business Case*. The *Business Case* will specify either a change to the design or methodology of an existing *Statistical Program*, which will result	

				in a new *Statistical Program Design*; or a change to one or more existing *Statistical Programs* (for example, to add an additional objective to the *Statistical Program*); or result in a new *Statistical Program* being created. This does not include statistical support functions such as metadata management, data management (and other overarching GSBPM processes) and design functions. These activities are conducted		

				as part of *Statistical Support Programs*.	
GSIM v1.1	Statistical Program Cycle	Business	A set of activities to investigate characteristics of a given *Population* for a particular reference period.	A *Statistical Program Cycle* documents the execution of an iteration of a *Statistical Program* according to the associated *Statistical Program Design* for a certain reference period. It identifies the activities that are undertaken as a part of the cycle and the specific resources required and processes used and description of relevant methodological information used in this cycle defined by the *Statistical Program Design.*	
GSIM v1.1	Statistical Program Design	Business	The specification of the resources required, processes used and description of relevant methodological information about the set of activities undertaken to investigate characteristics of a given *Population*.	The *Statistical Program Design* is an objects that provide the operational context in which a set of *Business Processes* is conducted. A simple example is where a *Statistical Program* relates to a single survey, for example, the Labour Force Survey. The *Statistical Program* will have a series of *Statistical Program Design* objects that describe the methodology and design used throughout the life of the survey. When a methodological change is made to the survey, a new *Statistical Program Design* is created to record the details of the new design.	

GSIM v1.1	Statistical Support Program	Business	A program which is not related to the post-design cyclic production of statistical products, but is necessary to support cyclical production.	This type of program will include such functions as metadata management, data management, methodological research, and design functions. These programs correspond to the horizontal functions shown in the GSBPM, as well as programs to create new or change existing *Statistical Programs*.	
GSIM v1.1	Subject Field	Concepts	One or more *Concept Systems* used for the grouping of *Concepts* and *Categories* for the production of statistics.	A *Subject Field* is a field of special knowledge under which a set of *Concepts* and their *Designations* is used. For example, labour market, environmental expenditure, tourism, etc.	subject area, theme
GSIM v1.1	Unit	Concepts	The object of interest in a *Business Process*	Here are 3 examples - 1. Individual US person (i.e., Arofan Gregory, Dan Gillman, Barack Obama, etc.) 2. Individual US computer companies (i.e., Microsoft, Apple, IBM, etc.) 3. Individual US universities (i.e., Johns Hopkins, University of Maryland, Yale, etc.)	
GSIM v1.1	Unit Data Point	Structures	A placeholder (or cell) for the value of an *Instance Variable* with respect to a *Unit* .	This placeholder may point to multiple values representing different versions of the data. Values are only distinguished on the basis of quality, date/time of measurement or calculation, status, etc. This is handled through the mechanisms provided by the *Datum* information object.	cell

GSIM v1.1	Unit Data Record	Structures	Contains the specific values (as a collection of *Unit Data Points*) related to a given *Unit* as defined in a *Logical Record*.	For example (1212123, 48, American, United Kingdom) specifies the age (48) in years on the 1st of January 2012 in years, the current citizenship (American), and the country of birth (United Kingdom) for a person with social security number 1212123.	
GSIM v1.1	Unit Data Set	Structures	A collection of data that conforms to a known structure and describes aspects of one or more *Units*.	Example: A synthetic unit record file is a collection of artificially constructed *Unit Data Records*, combined in a file to create a *Unit Data Set*.	micro data, unit data, synthetic unit record file
GSIM v1.1	Unit Data Structure	Structures	Describes the structure of a *Unit Data Set*.	For example (social security number, country of residence, age, citizenship, country of birth) where the social security number and the country of residence are the identifying components (*Identifier Component*) and the others are measured variables obtained directly or indirectly from the person (*Unit*) and are *Measure Components* of the *Logical Record*.	file description, dataset description
GSIM v1.1	Unit Type	Concepts	A *Unit Type* is a class of objects of interest	A *Unit Type* is used to describe a class or group of *Units* based on a single characteristic, but with no specification of time and geography. For example, the *Unit Type* of "Person" groups together a set of *Units* based on the characteristic that they are 'Persons'. It concerns not only *Unit Types* used in dissemination, but anywhere in the statistical process. E.g. using administrative data might involve the use of a fiscal unit.	Object class (ISO 11179)

GSIM v0.9	**User Profile**		A set of information for administering a user within a system supporting research.		
GSIM v1.1	Value Domain	Concepts	The permitted range of values for a characteristic of a variable	The values can be described by enumeration or by an expression	
GSIM v1.1	Variable	Concepts	The use of a *Concept* as a characteristic of a *Population* intended to be measured	The Variable combines the meaning of a Concept with a Unit Type, to define the characteristic that is to be measured. Here are 3 examples -1. Sex of person2. Number of employees3. Value of endowment	
GSIM v1.1	Web Scraper Channel	Exchange	A concrete and usable tool to gather information from the Internet.	This is an example of a way statistical organizations collect information (an *Exchange Channel*). The *Web Scraper Channel* contains *Scraping Process Maps*, which map the channel to each website targeted for scraping.	

Annex B: RIM, Statistics Norway's Business Process Model (SNBPM), and the Generic Statistical Business Process Model (GSBPM)

I. Introduction

This section addresses how Statistics Norway's Business Process Model (SNBPM - an implementation of the Generic Statistical Business Process Model [GSBPM]) can be used to describe RAIRD processes. This is a difficult subject, as RAIRD is a non-traditional type of statistical product, where what is delivered to the user is not the collected microdata, nor a set of statistical tables produced by the statistical agency. Instead, the tables provided to users are those which the users themselves produce, using the RAIRD data and tools.
There are two ways in which we can understand RAIRD in the context of the SNBPM. In one sense, the RIM supports many of the functions of the statistical production process, even though many of the functions are performed by researchers and not by the developers or maintainers of RAIRD itself. The diagrams below show how RAIRD can be mapped onto the SNBPM processes and sub-processes.
The second way in which we can use GSBPM to describe RAIRD is to use the process model to highlight what the functions of RAIRD itself are. This is distinct from the production processes which users might carry out in RAIRD. It is a more limited set of functions, which will be described separately.

II. RAIRD and the Statistical Production Process

This document describes how RAIRD changes the traditional process of statistical production, by bringing the consumer of statistical products into the production flow, offering them the ability to better meet their own needs by assuming some of the tasks traditionally carried out within SSB. This change can be seen by showing where, in the traditional process as described by Statistics Norway's Business Process Model, the end-user either participates with or replaces the SSB in performing some functions.
The end-users in question are more statistically literate than typical consumers of statistical products – because they are researchers associated

with accredited research institutions, they have a level of skill in terms of working with input data that goes beyond the capabilities of most journalists or policy makers. This is a class of users that also has the most specific demands in terms of the statistical products they wish to see. RAIRD – by offering the creation of specific statistical products as a *service* to end-users – represents a different paradigm, in which both SSB and the end users benefit. SSB benefits because the end users assume some of the tasks performed today by SSB staff. Thus, the resources needed to meet the end-users demands are less. For the end users, they now are able to get what they want quicker, based on real-time interactions on a larger set of source data.

III. "Statistics as a Service" – a Different Paradigm

The traditional process of producing statistics at SSB is described by Statistics Norway's Business Process Model. When we look at this model, it shows the process of identifying what statistical products are needed, determining what input data is needed and how it will be collected, performing the data collection, processing the data, analyzing it, and creating and disseminating the desired statistical products. This is an important function for the creation of policy, the conduct of business, for informed political discourse, and for use by citizens.
It is also an expensive process, and one in which some compromises must be made. A statistical office cannot predict exactly the demand for information for every possible end user, nor can it afford to provide information on this scale. Especially for more sophisticated users – researchers – there are likely to be specific statistical products which are not produced. To meet this need, SSB has a process of allowing qualified researchers access to the input data for research purposes. Because the input data are often highly confidential, there is a long process through which researchers prove that they are qualified to see the data, and negotiate with SSB to determine exactly what data they can be provided with. In some cases, a subset of the input data is made available to answer a specific research question. In other cases, special statistical products – customized tables – are provided to researchers.
RAIRD promises to change this situation, by providing access to an important source of data for researchers – the FD-Trygd. This is integrated data taken from five different registers. It contains data about the entire population of Norway, and covers topics which are of great interest to researchers in the social and economic domains. It is also very sensitive, because of the type of data it contains.
The level of access provided to researchers by RAIRD is not what they will get if they are willing to go through the lengthy process of applying for direct access to some subset of the FD-Trygd data. In RAIRD, the researchers will never see the actual data – only statistical products created from it. This is made possible because SSB – instead of performing the work of creating the statistical products – is providing a service to researchers so that they can create their own specialized statistical products. This lowers the risk of disclosure, and thus allows for increased user access to the data. Instead of working with just a subset of the data, users can operate on the entire set of variables, using the RAIRD analysis services.
To understand how this is possible, we can examine who now performs each step of the statistical production process: in order to preserve the confidentiality of the data, some functions are permitted to be performed by users, while others are not.

IV. The Researcher in Statistics Norway's Business Process Model

In the diagram below we see the high-level view of Statistics Norway's Business Process Model. It has been color-coded such that those process steps and sub-steps colored blue are performed entirely by SSB – the steps and sub-steps colored green are either completely performed by the researcher, or are performed by the researcher working together with SSB.

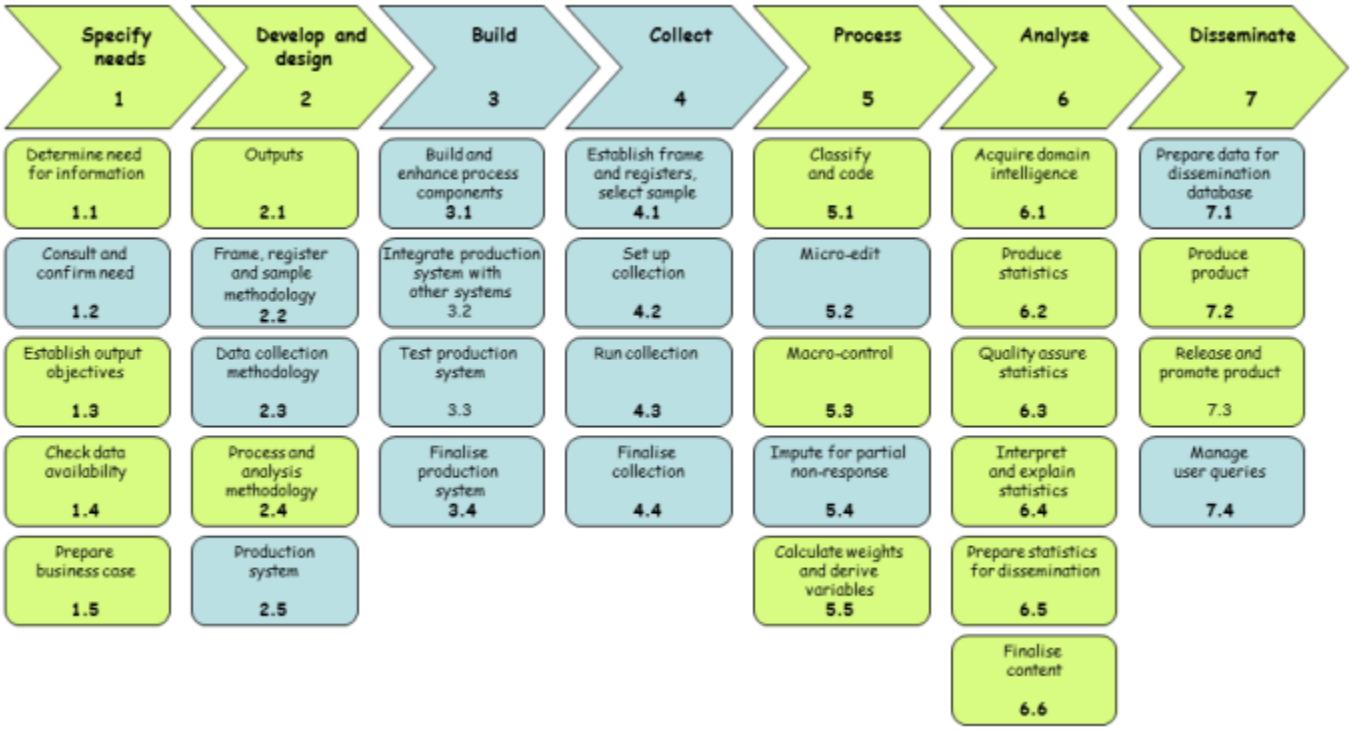

Figure 1: RAIRD support for the high-level SNBPM steps.
Each process step will be described in more detail below, but some general comments can be made. In the early phases of the process, where

the specific statistical products are designed, the researcher performs much of this work. They are the ones organizing their own research projects, identifying their research questions, etc. Steps 3 and 4 – "Build" and "Collect" have been performed by SSB in developing and deploying RAIRD itself, and the register data is loaded into the system by SSB – the researchers do not perform any system development or data collection. When it comes to steps 5 and 6, "Process" and "Analyse", almost all of the work after data editing is performed by the researcher. RAIRD does not provide a capacity for researchers to edit the data, although they can do recoding for the purposes of creating their own statistical products. Step 7, "Dissemination" is mostly handled by the researcher, since this activity involves taking the output of the RAIRD system and using it in a research paper or other publication.

Step 1 -Specify Needs

Figure 2: The SNBPM Specify Needs Step

Researchers will perform most of this work in organizing their research project. The only part of this that researchers do not perform is sub-step 1.2, since there is no requirement for it. The researchers are the ones who have identified RAIRD as containing the data they need for their research.

Step 2 – Develop and Design

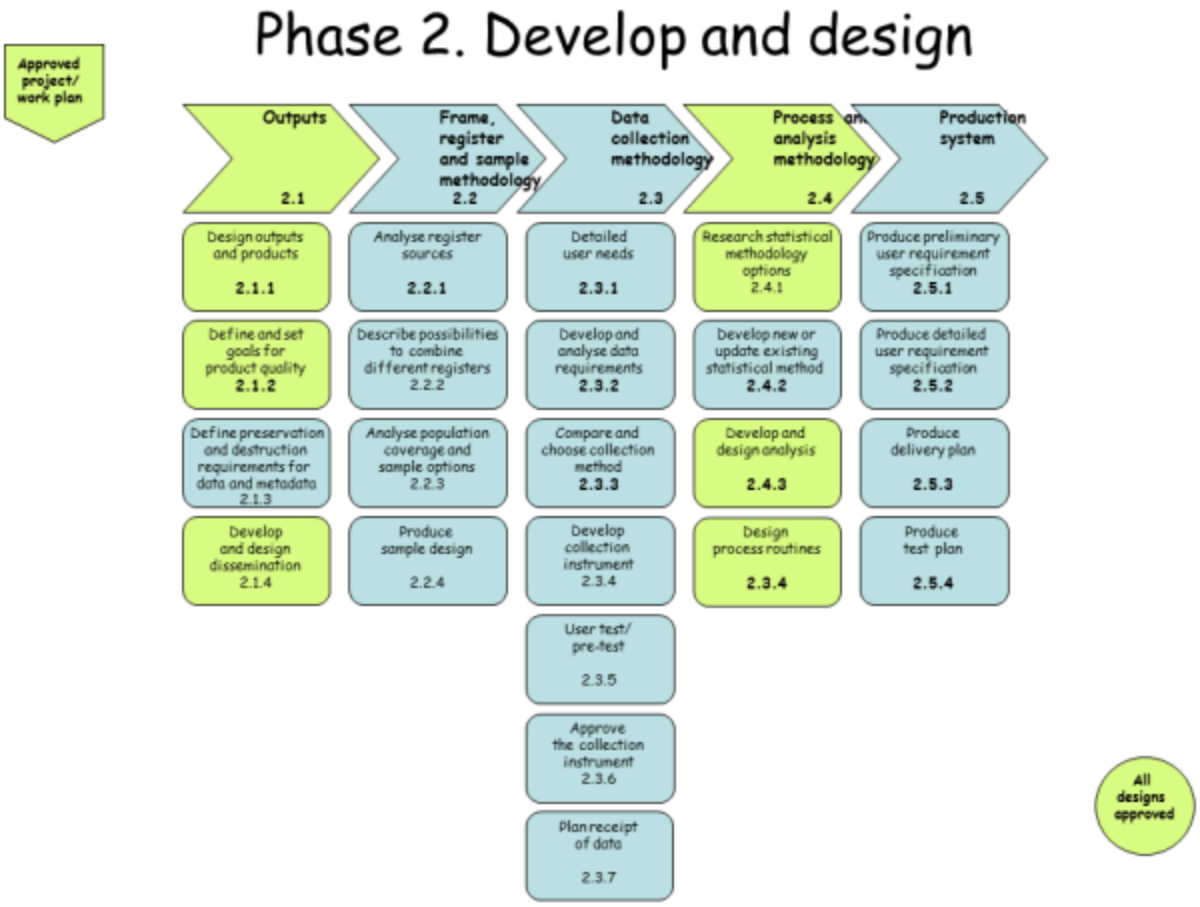

Figure 3: SNBPM Develop and Design Step.
The only sub-steps which are performed by the researcher here are those related to the design of their analysis and the statistical product desired. They are given no control over how the data is collected, or the methodology used. This is not a major issue, given that the data are coming from administrative registers.

Step 3 – Build/Step 4 – Collect

In neither of the areas does the researcher participate. The RAIRD system is developed and deployed by SSB, and the ongoing collection of the register data is conducted by them. The confidential nature of the input data is such that this cannot be performed by anyone outside of SSB.

Step 5 – Process

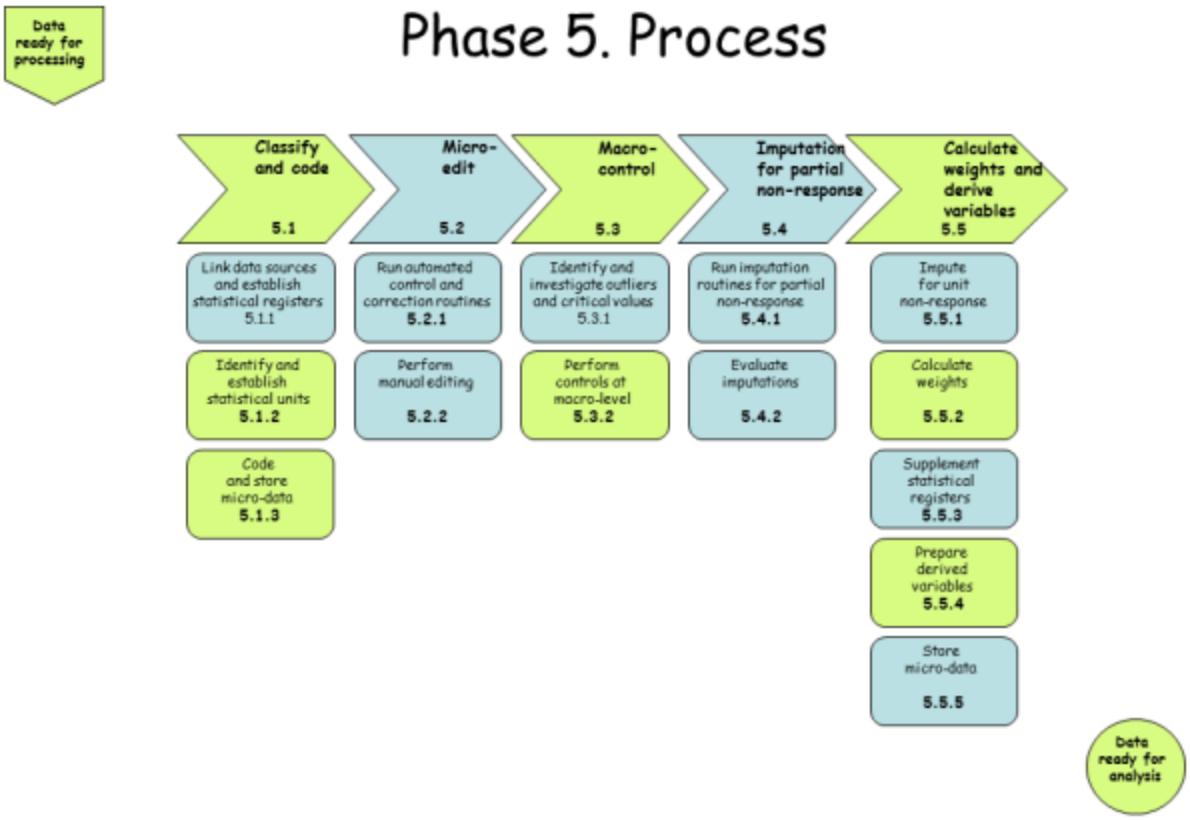

Figure 5: The SNBPM Process Step.

This step is largely performed by the researcher, but their ability to interact with the input data set must be carefully controlled. All of the data editing, and integration is performed by SSB, so that the input data is ready for analysis by the researcher using the RAIRD service. There are some exceptions to this: RAIRD will allow researchers to perform re-codes of data – these will not be stored - so that their statistical products can be compared with other data outside of the RAIRD environment. Further, macro-control is possible by performing tabulations and other operations; the derivation of variables will also be permitted, although these derived variables will only be used for the purpose of creating the specific statistical product wanted by the research project – they will not be stored and shared.

Step 6 – Analyse

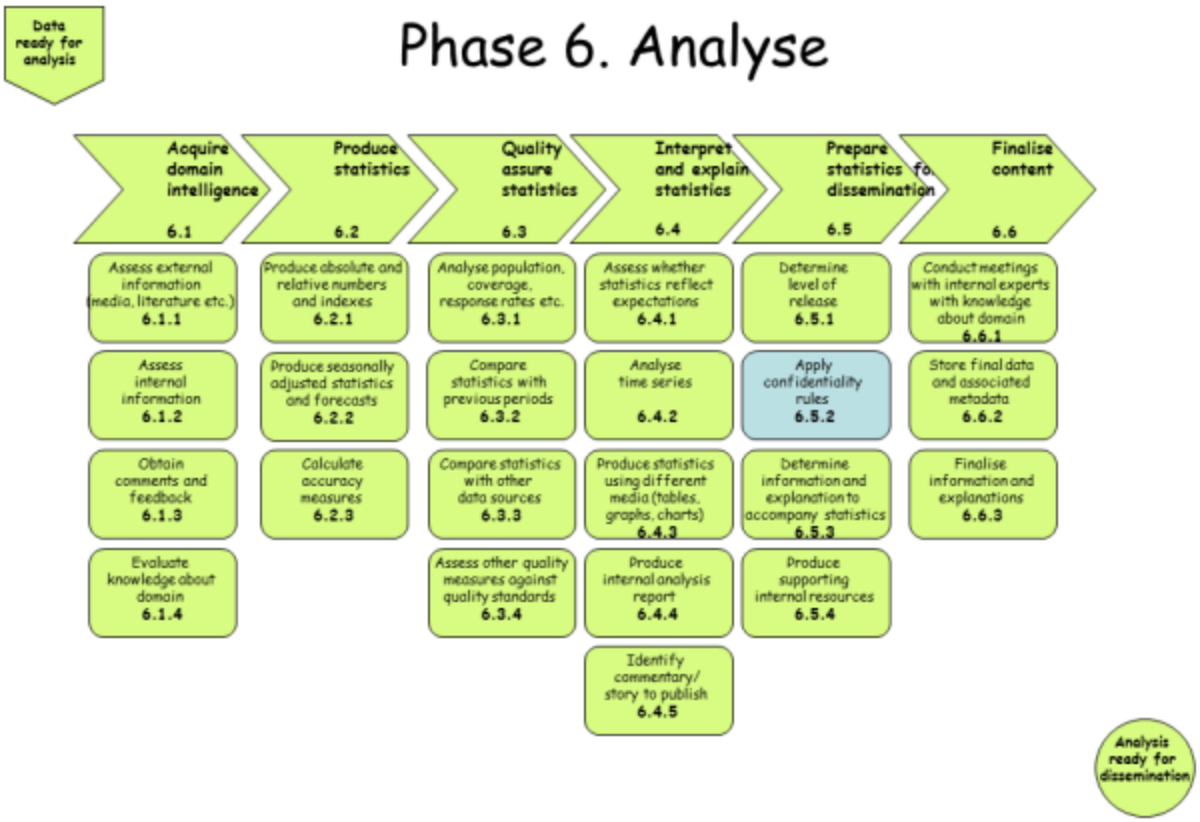

Figure 6: The SNBPM Analyze Step.

This step is – with one exception – entirely performed by the researchers. That exception is the exercise of disclosure control on the statistical products. RAIRD must control this feature, as it is responsible for guaranteeing the confidentiality of the data. Otherwise, the researchers are going to perform all the "production" functions – the tabulation, quality control, etc. and – as they begin to create their research publications – they will interpret and explain their findings.

Step 7 - Disseminate

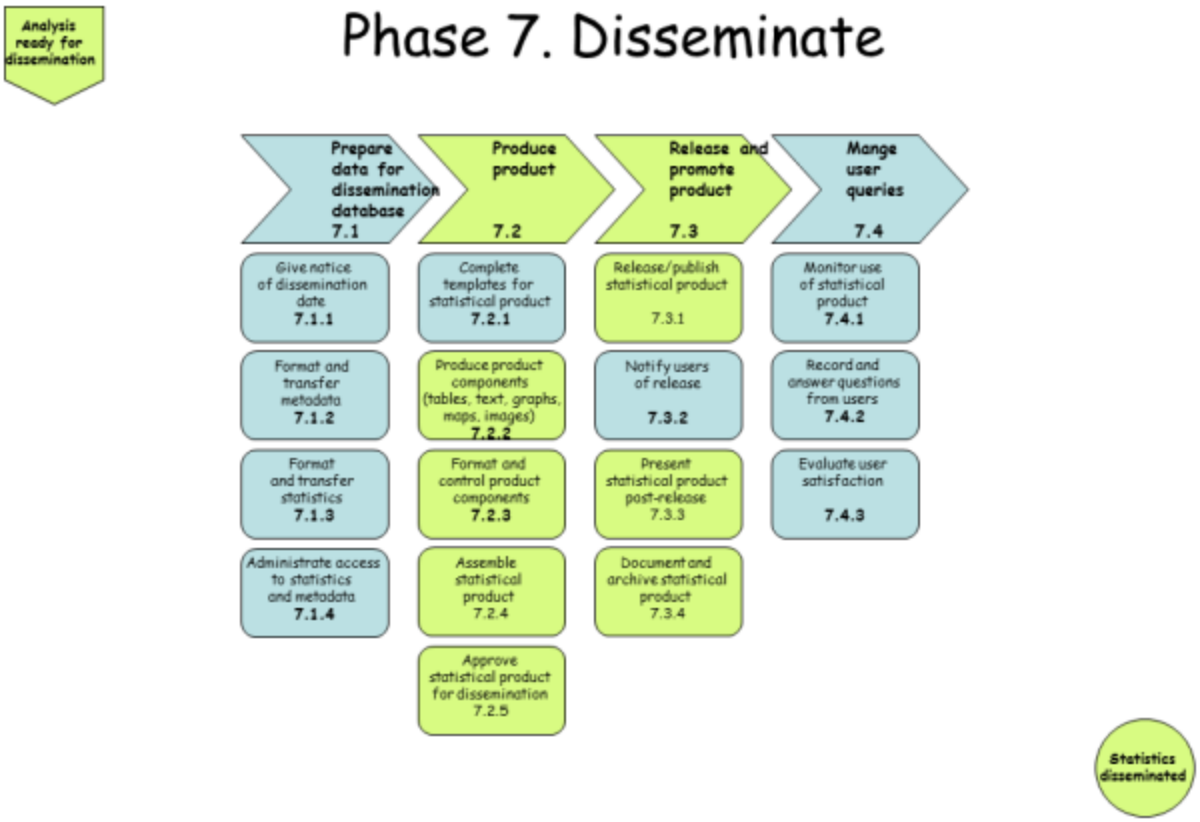

Figure 6: The SNBPM Disseminate Step.
Dissemination is the act of publication, so SSB has nothing to do here – the researchers use their own channels to publish their own research. The two tasks which the researcher does not perform – in fact, no one has to perform in the RAIRD scenario – are the preparation of the statistical products for insertion into the Dissemination Database (these statistical products do not use this dissemination platform) and the management of user queries, as SSB will not receive any.

V. Describing RAIRD Functions with the NSBPM

Of the parts of the production process which are directly offered to researchers through RAIRD, we have a sub-set of the functions which they will need to perform. Many aspects of statistical production do not require support through RAIRD, as these are functions which researchers will do on their own, through existing channels (finalizing the statistical product, for example, equates to publishing a research paper in a peer-reviewed journal or elsewhere – not something RAIRD needs to help researchers do).
From the researcher's perspective, RAIRD functions are limited to the high-level NSBPM steps as shown in the list below:
Step 5: Process
5.1 Classify and code
5.3 Macro-control
5.5 Calculate Weights and derive variables
Step 6: Analyze
6.2 Produce statistics
6.3 Quality assure statistics
6.5 Prepare statistics for dissemination
6.6 Finalize content

Annex C: Technical Overview of RIM Information Flows and GSIM

This annex provides an overview of all the information flows in RIM, showing how these are modeled strictly as information objects, independent of the systems within RAIRD. This includes the connection points with the GSIM model, and how RIM information flows specialize some of the GSIM *Data Set* and *Information Resource* objects. (See Figure 1).

Here, we can see that the ***Input Data Set*** is a specialized type of GSIM's *Referential Metadata Set*, and that the accompanying ***Event History Input Data Set*** is a RIM ***Datum-Based Data Set.*** The ***Event History Data Store*** is – from a strict information perspective – both a GSIM *Data Resource* , as it groups all of the event history data which has been loaded, and a GSIM *Referential Metadata Resource.* In order to represent this in RIM, we have identified it as a GSIM *Information Resource*, which is the superclass of both *Data Resource* and *Referential Metadata Resource.* In RIM, the ***Event History Data Store*** should be understood as having the properties of both of the two sub-classes of the GSIM *Information Resource.*

The data in the ***Event History Data Store*** forms the basis of the ***Analysis Data Sets***, which are expressed for processing as a specialization of the GSIM *Unit Data Set****. Analysis Data Sets*** are transformed as the researcher operates on them, so we have a process where one ***Analysis Data Set*** potentially goes through several iterations.

Ultimately, the ***Analysis Data Set*** will form the basis for the creation of a ***Provisional Output.*** We see that the ***Provisional Output*** is a specialization of the GSIM *Dimensional Data Set.* The process of assessing and managing disclosure risk is potentially transformative as well, so we also have the ***Final Output***, which, like the ***Provisional Output***, is a specialization of the GSIM *Dimensional Data Set.*

For all of the data sets, we can see that they correspond to GSIM *Data Structure* objects of different types (***Datum-Based Data Structure***, *Unit Data Structure*, and *Dimensional Data Structure*).

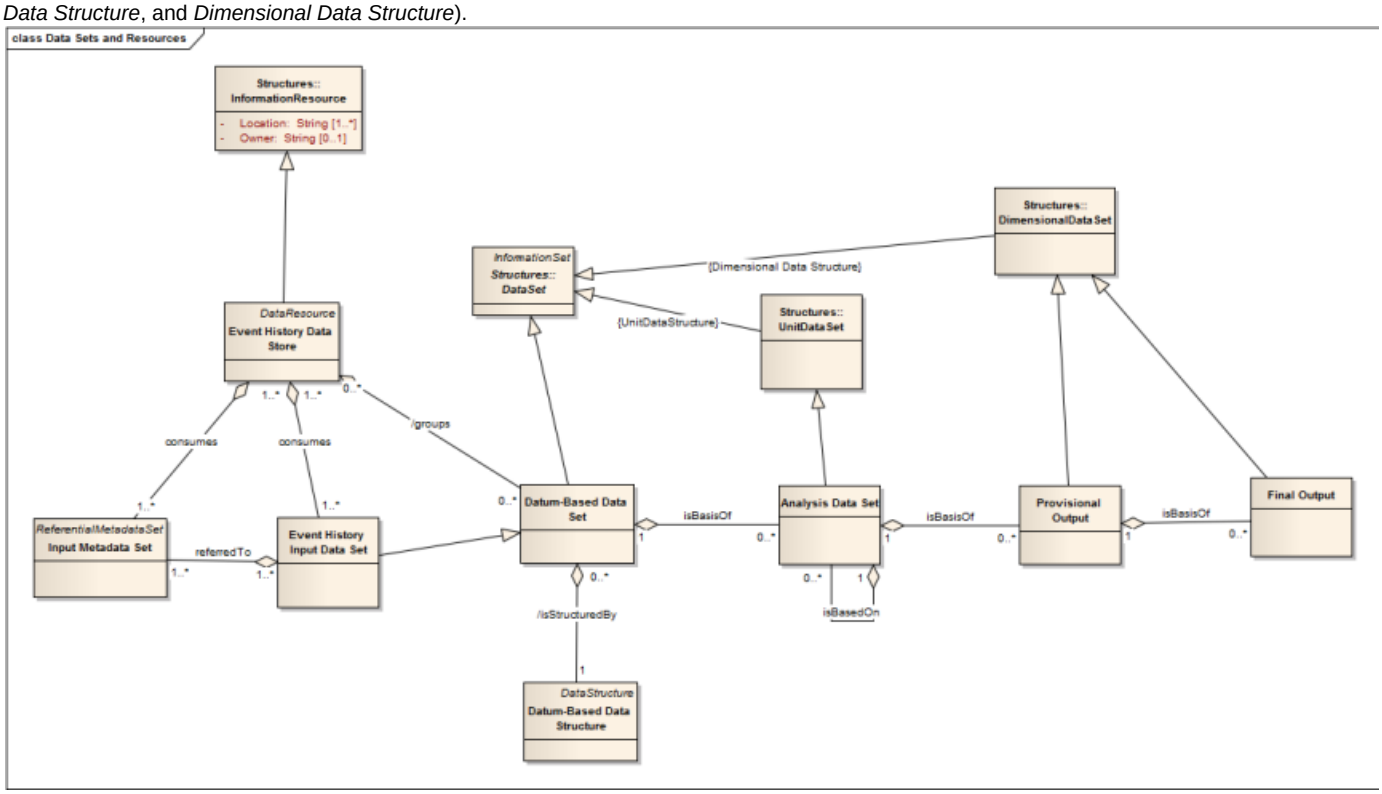

Figure 1: RIM Information Flows and Resources

There is another important information flow in the RIM which does not appear in Figure 1 – the ***Data Catalogue***. Figure 2 shows how this is represented as a specialized GSIM *Product.*

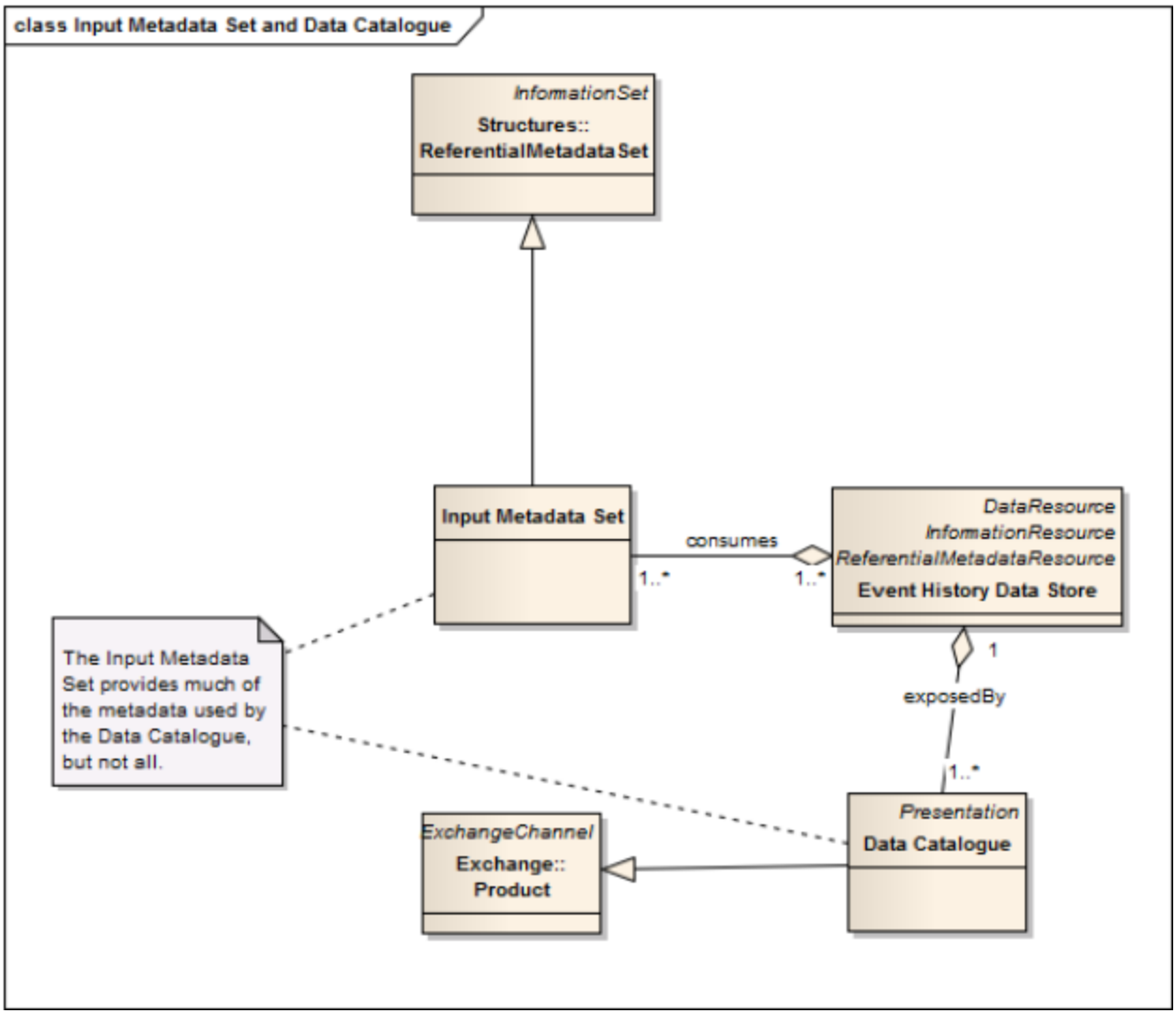

Figure 2: Information Flows in the Data Catalogue

The ***Data Catalogue*** information is supplied by the ***Event History Data Store,*** which is as mentioned above has several functions, but here is functioning as a *Referential Metadata Resource*. The ***Data Catalogue*** itself is a *Product*, which would have the other relevant GSIM objects as well (eg, *Presentation).* The ***Input Metadata Set*** is seen as a specialization of the *Referential Metadata Set.*

Annex D: Technical Considerations of Mapping RIM to DDI

In order to understand how the Data Documentation Initiative (DDI) standard can be used to implement RIM, we must first ask "Which parts of RIM require the functionality offered by DDI?" Typically, a standard such as DDI is most relevant where metadata is required in a structured, machine-actionable form. This occurs where metadata (and often the data it describes) are passed from one institution to another.

In the case of RAIRD, this occurs during the load process, where the data collected and cleaned within SSB is passed into the ***Event History Data Store***. DDI documentation might also be useful for describing the ***Final Outputs*** delivered to the researchers. Everywhere else, the data and metadata within RAIRD exists within a single implementation environment, where a serialization in XML or similar implementation syntax are generally not needed.

When we consider the requirements for the loading of metadata into the ***Event History Data Store*** (the ***Input Metadata Set***) we find that several parts of the RIM could usefully be mapped against DDI:

- Concepts and concept systems, with links to variables and external documentation
- Variable descriptions with labels and definitions
- Pre-assessment values for disclosure control, assigned to variables
- Codelists and classifications
- Links to external documentation
- Unit types and associated logical record structures (which variables are associated with which units)
- Relationships between Unit types

Given this set of information, in the XML syntax of the DDI Lifecycle standard we find the ability to describe almost all of the needed information. The only requirement we see here that DDI Lifecycle does not support is the ability to capture pre-assessment values for disclosure control at the level of the variable – this could be handled with some simple extensions to the standard, however, using the extension mechanism

recommended in DDI Lifecycle. (DDI Codebook cannot describe the relationships between unit types, nor the pre-assessment values for disclosure control.)
In most cases, the use of DDI Lifecycle for describing this metadata is obvious: the DDI ConceptScheme can describe concept systems, and links with variables and external documentation are supported (DDI Variables reference Concepts, and DDI OtherMaterials can be linked to DDI Concepts as well). DDI offers a rich description for Variables – for the RIM, this will cover all of the different types of variables (GSIM *Variable. Represented Variable,* and *Instance Variable* map to DDI's ConceptualVariable, RepresentedVariable, and Variable constructs, respectively; to support variable-level pre-assessment values, the Data Element construct would need to be extended.)
Codelists and classifications are both present in DDI Lifecycle, and map cleanly to GSIM. Unit Types are associated with Variables in DDI, and the Variables are assembled into Logical Records – these constructs can be used to support the needed description of unit types and record structures. Further, the linkages between unit types can be described in DDI Lifecycle using the RecordRelationship construct.
It should be mentioned that the variable and record descriptions are needed only to the logical level – the load process for the ***Event History Data Store*** does not require physical descriptions of the data.
There have been some standard DDI Lifecycle profiles created by UN/ECE for representing GSIM information objects, and these standard prfiles cover almost all of the metadata required for the load process. These can be found at: http://www1.unece.org/stat/platform/display/gsim/DDI+Profiles. These do not cover everything which is needed (especially when it comes to concept systems) but do cover most of the needed constructs from DDI Lifecycle.
For describing the ***Final Outputs*** using DDI, we would need to add the physical description of the data and use the DDI Lifecycle NCube construct. This is probably not of value to the researchers, although it might be desireable from the perspective of the organizations archiving the researcher's work (such as NSD).

References

RAIRD: http://www.raird.no
GSIM: http://www1.unece.org/stat/platform/display/gsim/Generic+Statistical+Information+Model.
GSBPM: http://www1.unece.org/stat/platform/display/GSBPM/Generic+Statistical+Business+Process+Model
DDI Specification: http://www.ddialliance.org/Specification/
Statistics Norway's Business Process Model (SNBPM):
http://www.ssb.no/en/omssb/om-oss/vaar-virksomhet/planer-og-meldinger/statistics-norways-business-process-model