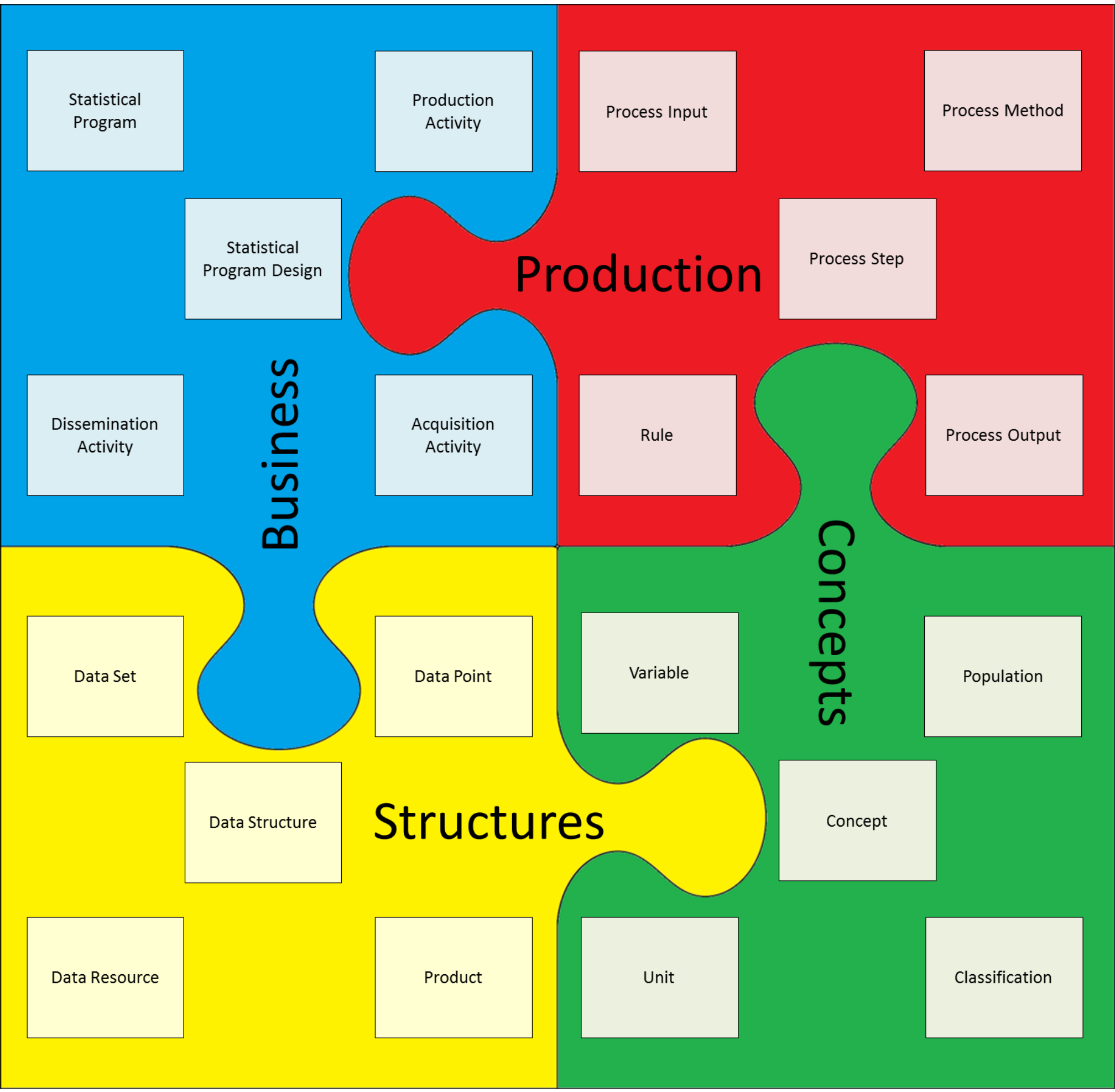
Generic Statistical Information Model (GSIM)



GSIM is a reference framework of internationally agreed definitions, attributes and relationships that describe the pieces of information that are used in the production of official statistics (information objects).

This framework enables generic descriptions of the definition, management and use of data and metadata throughout the statistical production process.

GSIMPROJECT

BUSINESS

The GSIM covers the whole statistical process. It includes information objects used in the Specify Needs phase (such as Statistical Need and Business Case) right through to information objects used in the Evaluation phase (such as Evaluation and Assessment).

Statistical Need

may change program design

Statistical Program Design

may initiate

has

Statistical Program

The model identifies different types of activities that represent the major steps in the statistical production process . These types have been specifically identified in the model but other types could be defined.

A Statistical Program is the overarching, ongoing activity that an organization undertakes to produce statistics. A Statistical Program has an associated set of Statistical Program Designs that identify the methodology used for the Statistical Program.

is comprised of

Statistical Activity

has subtype

Acquisition, Production, Dissemination Activity

includes

uses

Data Channel

The model supports collecting data via surveys or registers. A Data Channel identifies the Instrument used to collect data and the Data Resource from which the data is collected. An Instrument is the description of the tool that will be used to collect data (for example a survey instrument).

The GSIM supports both static and dynamic dissemination outputs. This includes Products (for example publications, press releases, etc) and Representations (which may contain data, structural metadata or conceptual metadata).

PRODUCTION

has

Process Step Design

applies

Process Method

has

Rule

The GSIM was designed to enable an explicit separation between the design and execution of statistical processes. The Production Group assumes that each process step will be designed during a design phase. During the production phase, Process Steps are "executed" (performed) in accordance with their design. For each Process Step that is executed, there will be a Process Step Execution Record.

has

Process Step

has

Process Step Execution Record

records

Process Input, Process Output

The execution of a Process Step will supply Process Inputs and result in Process Outputs. In many cases, these Process Inputs and Outputs are the conceptual and structural information objects that are described in the GSIM Concepts and Structures groups.

may include objects from this group

may include objects from this group

STRUCTURES

The GSIM makes a distinction between unit and dimensional data. There are a number of data structure components defined by the model. These include identifier, measure and attribute components.

Data Structure

has

Data Structure Components

has

Data Resource

includes

Data Set

contains

Data Point

observation of

describes

CONCEPTS

The GSIM model includes rich detail on classifications. Classification management is important to statistical agencies who have complicated representations of concepts within their statistical information systems

Classification

represents

Instance Variable

uses

Represented Variable

uses

Code List

represents

The model differentiates between the objects used to conceptually describe information, and those that are representational. For example, a Variable does not contain any information on how the resulting value may be represented. This information is put into the Represented Variable. This enables the Variable definition to be reused

uses

Unit

comprises

Population

measures

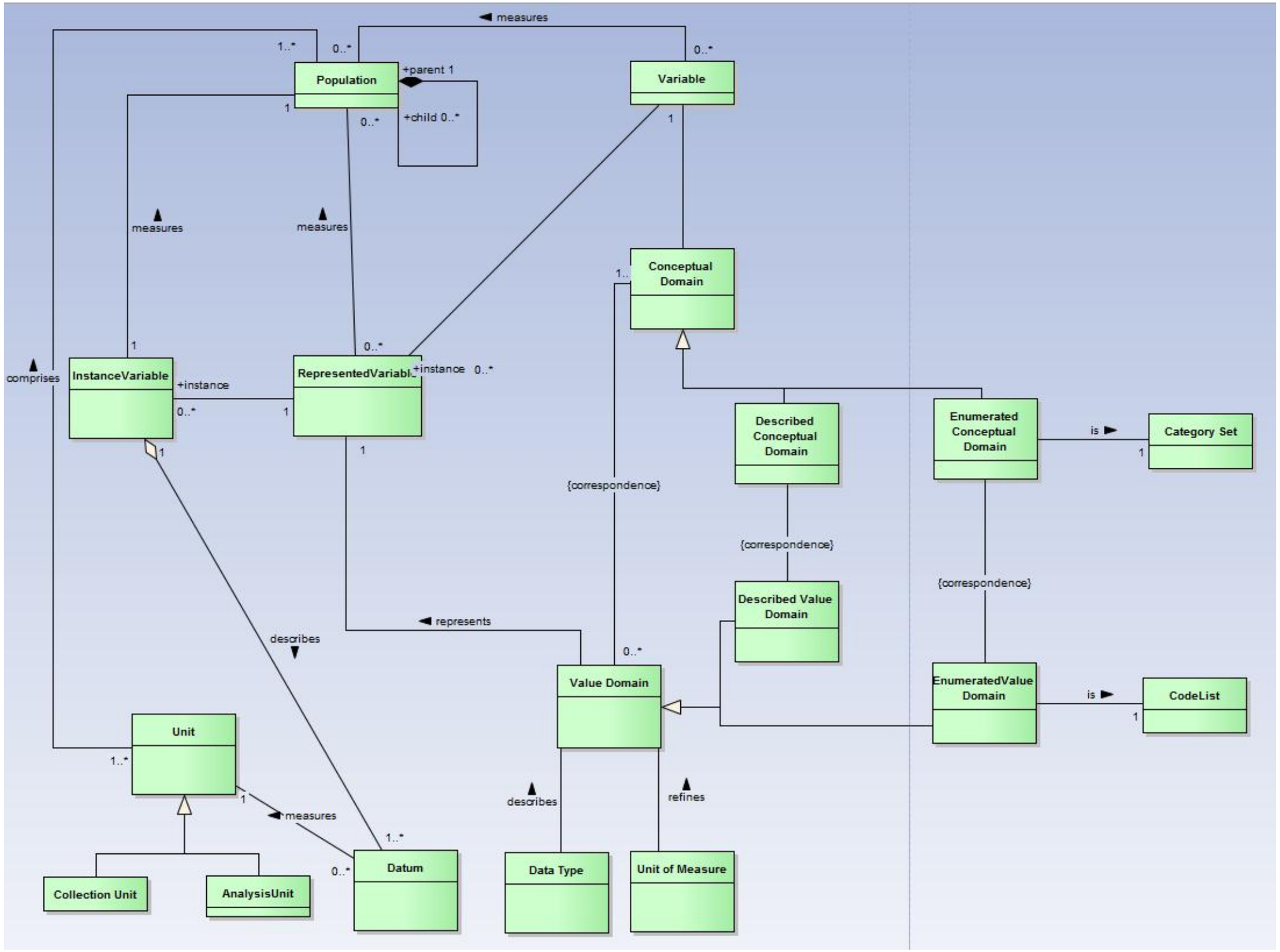
Variable

defines

Concept

groups

Concept System



Variable relationship

GSIMPROJECT