SBGN Exchange Format Requirements Specification: Draft 1.1

Sarah Boyd

July 18, 2009

1 Process Diagram

This section outlines the requirements for an exchange format for Process Diagram language of SBGN.

1.1 Process Diagram Graph Parameters

The following could be provided as parameters defining the preferred drawing of the graph:

- Location of origin[1]: where the origin of the graph is assumed to be with respect to the specification of this graph, for example (0,0) maps to bottom left, or top right.
- Preferred size: the preferred size of this graph.
- Units: the units being used to describe the sizes in the file.

1.2 Entity Pool Node Parameters

The requirements for the exchange format for EPNs depends partly on the type of the EPN, although some requirements are universal, and some should be treated as optional.

1.2.1 Parameters for specific Entity Pool Nodes

Table 1 summarises the required and optional parameters needed by EPNs in the SBGN exchange format for Process Diagram.

[1] As discussed in Auckland (April 2009), different systems use different locations for the origin when drawing (see also discussion about specifying the node location, etc), so this is included here to ensure it is discussed.

Table 1: Summary of exchange format requirements for Entity Pool Nodes

EPN	**Required**	**Optional**
Unspecified Entity	ID Node location Label Node orientation	Colour Preferred size Label drawn as multiline
Simple Chemical	ID Node location Label Node orientation Clone marker Multimeric state Unit of information	Colour Preferred size Label drawn as multiline
Macomolecule	ID Node location Label Node orientation Clone marker Multimeric state Unit of information State variable	Colour Preferred size Label drawn as multiline
Nucleid Acid Feature	ID Node location Label Node orientation Clone marker Multimeric state Unit of information State variable	Colour Preferred size Label drawn as multiline
Source/Sink	ID Node location Node orientation	Colour Preferred size
Tag	ID Node location Label Node orientation	Colour Preferred size Label drawn as multiline
Observable	ID Node location Label Node orientation	Colour Preferred size Label drawn as multiline
Orientation	ID Node location Label Node orientation	Colour Preferred size Label drawn as multiline

The definitions of the required parameters listed in Table 1 are as follows:

- ID: a unique numerical identifier for this entity within the parent PD.
- EPN location[2]: define the centre of the node relative to the origin of the graph.
- Orientation: Whether the EPN is drawn in the horizontal or vertical alignment.
- Multimeric state: boolean defining whether this EPN has a multimeric state.
- Label: a string defining the label of this node.
- Clone marker: defined if the EPN carries a clone marker, see Section 1.3 for details.
- Unit of information: defined for each unit of information carried on the EPN, see Section 1.3 for details.
- State variable: defined for each state variable carried on the EPN, see Section 1.3 for details.

The definitions of the optional parameters listed in Table 1 are as follows:

- Colour: All EPNs can specify colour as an optional parameter.
- Label drawn as multiline: optional parameter specifying where the line is split, if it is split over multiple lines.
- Preferred size: the preferred size of this EPN.

1.3 Auxiliary Units

1.3.1 Parameters for Auxiliary Units

This section describes parameters that are used to describe Auxiliary Units.

- Label [3]: clone markers can carry a label, so should be specified. To prevent ambiguity and error checking, this should be left as an empty string if no label is defined.

[2]Defining locations is going to be complex, because different systems have different origins. This is just one suggestion for how to specify this, and also see discussion regarding the preferred graph layout in the 1.1 section. Also, defining the centre is not necessarily the best option, however defining the centre of the node will also define the position of the label.

[3]Point for discussion: the label on a unit of information includes a prefix that has a controlled vocabulary. Because the prefix is separated from the label name with the colon character ":", it is relatively straightforward for a program that wants to separate the label name and prefix to parse this, rather than requiring that they be specified separately. Similarly, for the state variable label, which is separated using the "@" character. This will reduce the amount of text required to specify the exchange document, and removes the need for programs that don't allow users to manipulate these states from having to glue them back together again.

- Parent entity [4]: the EPN, container or compartment to which this auxiliary unit belongs.

Parameters that can be optionally specified for auxiliary units are:

- Location [5]: Auxiliary Items can optionally specify the original location where it should be drawn.
- Colour: Auxiliary Items can specify colour as an optional parameter.
- Preferred size: the preferred size of the Auxiliary Item.

1.4 Process Nodes

1.4.1 Parameters for Process Nodes

All Process Nodes (PNs) require definition of the following parameters:

- ID: a unique numerical identifier for this entity within the parent PD.
- Position [6]: define the centre of the PN relative to the origin (x=0, y=0) of the graph.
- Consumption inputs: a list of the entities that are the inputs to this PN. Restricted to one input for the PN "Dissociation".
- Production outputs: a list of the entities that are the outputs for this PN. Restricted to one output for the PN "Association".

Process Nodes can also optionally define the following parameters:

- Colour: PNs can be optionally coloured.
- Preferred size: the preferred size of this PN.
- Orientation [7]: Whether the PN is drawn in the horizontal or vertical alignment.

[4] This actually doesn't have to be defined if the auxiliary unit is defined as part of the object to which it relates, which is probably a much more sensible arrangement. For example, if a state variable belongs to a macromolecule, and is included within the definition of the macromolecule, then this becomes redundant. This would also mean that an auxiliary unit can simply be declared within the definition of the "owner" entity, with just a label, which makes defining, parsing, and error checking much simpler.

[5] Points for discussion: If the location of the auxiliary item is specified, should this be relative to the graph or the parent entity? Specifying the location of the Auxiliary Unit should probably be optional, because it is something of a stylistic specification.

[6] See also discussion of EPN position in Section 1.2

[7] Since orientation of PNs is currently under discussion, it seems that this should be specified as optional, and may be calculated according to the specific style decisions of individual programs

1.5 Connecting Arcs

1.5.1 Parameters for Connecting Arcs

Table 1 summarises the required and optional parameters needed for Connecting Arcs in the SBGN exchange format for Process Diagram.

Table 2: Summary of exchange format requirements for Connecting Arcs

Connecting Arc	**Required**	**Optional**
Consumption	Source EPN Target PN Cardinality	Colour Preferred path
Production	Source PN Target EPN Cardinality	Colour Preferred path
Modulation	Source EPN Target PN ID	Colour Preferred path
Stimulation	Source EPN Target PN ID	Colour Preferred path
Catalysis	Source EPN Target PN ID	Colour Preferred path
Inhibition	Source EPN Target PN ID	Colour Preferred path
Trigger	Source EPN Target PN ID	Colour Preferred path
Logic Arc	Source EPN Target Logical Operator	Colour Preferred path
Equivalence Arc	Source EPN Target Tag	Colour Preferred path

The definition of the required parameters for Connecting Arcs are:

- ID: a unique numerical identifier for this entity within the parent PD.
- Source PN: the ID of the PN that is the source to the Connecting Arc.
- Target PN: the ID of the PN that is the target of this Connecting Arc.
- Source EPN: the ID of the EPN that is the source to the Connecting Arc.
- Target EPN: the ID of the EPN that is the target of this Connecting Arc.
- Target Logical Operator: the ID of the Logical Operator that is the target of this Connecting Arc.
- Target Tag: the ID of the Tag that is the target of this Connecting Arc.

Connecting Arcs can also optionally define the following parameters:

- Colour: Connecting Arcs can be optionally coloured.
- Preferred Path: this can be used to specify the path along which the Connecting Arc should be drawn.

1.6 Logical Operators

1.6.1 Parameters for Logical Operators

Logical Operators require the following parameters:

- Input source: ID of an EPN or logical operator. Single input for "NOT" operator, multiple inputs for "AND" and "OR" operators.
- Output target: the ID of a modulation, stimulation, catalysis, inhibition or trigger arc.
- ID: a unique numerical identifier for this entity within the parent PD.

Process Nodes can also optionally define the following parameters:

- Colour: Logical Operators can be optionally coloured.
- Preferred size: the preferred size of this Logical Operator.
- Orientation [8]: Whether the Logical Operator is drawn in the horizontal or vertical alignment.

[8] See footnote relating to discussion about orientation of PNs

1.7 Compartments

Table 3: Summary of exchange format requirements for Compartments

Connecting Arc	**Required**	**Optional**
Complex	Components Location Label State variables Units of information ID	Colour Preferred size Label drawn as multiline
Compartment	Components Location Label Units of information ID	Colour Preferred size Label drawn as multiline
Submap	Terminals Label ID	Colour Label drawn as multiline

2 Abbreviations

Abbreviations used in this document:

- EPN: Entity Pool Node
- PD: Process Diagram
- PN: Process Node
- SBGN: Systems Biology Graphical Notation