

Synthetic Biology Open Language Visual: an ontological use case

Jacqueline Quinn¹, Michal Galdzicki², Robert Sidney Cox III³, Jacob Beal⁴, Kevin Clancy⁵, Nathan Hillson⁶, Larisa Soldatova⁷

¹Autodesk, Inc., San Francisco, CA; ²University of Washington, Seattle, WA; ³RIKEN Yokohama, Japan; ⁴BBN Technologies, Cambridge, MA; ⁵Life Technologies, Carlsbad, CA; ⁶Joint BioEnergy Institute, Emeryville, CA; ⁷Brunel University, London, UK

1 SBOL AND SBOL VISUAL

Synthetic Biology Open Language (SBOL) is a data exchange standard for the specification of forward engineered genetic designs (Galdzicki 2012). SBOL Visual is the graphical counterpart to SBOL, used to represent designs in a human readable manner. The central element in the SBOL data model is the DNA Component, which represents the design of a contiguous piece of DNA. DNA Components have an assigned functional role, generally referred to as ‘part type’ among synthetic biologists and which is analogous to the feature keys in annotated DNA sequences. SBOL exploits this similarity by requiring that part types be a term from the Sequence Ontology (SO) a pre-existing and widely used project for mapping DNA sequence annotation features onto an ontology framework (Eilbeck 2005).

SBOL Visual consists of a set of symbols that can be used to graphically represent genetic designs (Quinn 2013). Each symbol represents one or more terms in the SO. In SBOL Visual, a DNA Component is represented visually by the symbol that represents its SO term. SBOL Visual symbols are an open-source and community-agreed collection of pictograms, intended to cover that subset of the SO relevant to genetic parts frequently used by synthetic biologists---at present 20 community-agreed pictograms. As the community and technology continue to develop, significant growth of the symbol set is expected. Ongoing efforts to enable such growth include development of a process for community members to propose new symbols and have them incorporated into the standard.

2 USE CASES IN SYNTHETIC BIOLOGY

Standard graphical notation is commonplace in many engineering fields, but only emerging in biological engineering. The design abstraction provided by such notation increases human readability of design diagrams. In SBOL Visual, this is an abstraction away from nucleotide based sequence views with text annotations. From an SBOL Visual diagram, the viewer should be able to identify the important components of the design and their function in the design.

Having standardized graphical notation also helps to unify software tools. Many software tools import SBOL files, and

by using SBOL Visual, they can automate visualization of such designs in a way that is compatible with other software tools and common use. Several software tools for synthetic biologists also use visual icons to allow the user to design constructs by piecing together parts de-novo, rather than starting from a single sequence of origin. In these tools, SBOL Visual allows the user to design construct functionality using part types prior to assigning a specific sequence.

3 IMPLEMENTATION OF SBOL VISUAL

The initial implementation of SBOL Visual was a simple table in Microsoft Word associating a normative image with a name and SO term. While this was effective for a small community and small initial symbol set, it is neither scalable nor machine-readable. To address these limitations, SBOL Visual has been reformulated as an ontology where the classes are represented by pictograms rather than labels. This way, relationships to symbols and terms from SBOL, SO, and other ontologies can be unambiguously specified in a scalable and machine-readable manner.

SBOL Visual has thus been expressed as an ontology in OBO, a format chosen for its simplicity and ability to display images alongside terms within the application. The ontology has also been converted to OWL format, though this resulted in the loss of image information, a preferred format of many software developers interested in SBOL Visual. In future work, implementation of a way to associate images with terms must be implemented for OWL to appropriately specify SBOL Visual.

SBOL Visual is available at BioPortal:
<http://bioportal.bioontology.org/ontologies/3226>. More information can be found at <http://sbolstandard.org/visual>.

REFERENCES

- Eilbeck, K., et al. (2005). The Sequence Ontology: a tool for the unification of genome annotations. *Genome biology*, 6(5), R44. doi:10.1186/gb-2005-6-5-r44
- Galdzicki, M. et al. (2012). Synthetic Biology Open Language (SBOL) version 1.1.0. BBF RFC #87. doi:1721.1/73909
- Quinn, J., et al. (2013). Synthetic Biology Open Language Visual (SBOL Visual), version 1.0.0 BBF RFC#93. doi:1721.1/78249
- The GenBank/EMBL/DBJ feature Table Definition http://www.insdc.org/files/feature_table.html