

SBOL Visual 2 Ontology

Göksel Mısırlı¹, Jacob Beal², Thomas E. Gorochoowski³, Guy-Bart Stan⁴, Anil Wipat⁵, Chris Myers⁶

¹Keele University, UK, ²Raytheon BBN Technologies, USA, ³University of Bristol, UK, ⁴Imperial College London, UK,

⁵Newcastle University, UK, ⁶University of Utah, USA

g.misirli@keele.ac.uk, jakebeal@ieee.org, thomas.gorochoowski@bristol.ac.uk, g.stan@imperial.ac.uk,

a.wipat@newcastle.ac.uk, myers@ece.utah.edu

1 INTRODUCTION

The ability to visually represent genetic circuits can aid human understanding and improve the dissemination of information. Such visual representations are especially useful in publications for helping to explain complex relationships between constituent parts of large genetic circuits. However, when these diagrams are created manually, variations in how information is presented may cause issues in interpreting the meaning of different glyphs and how they are connected. Moreover, there are an increasing number of computational tools and repositories for synthetic biology that can automatically render visual depictions of genetic circuits. The SBOL Visual [2] standard has been developed to provide guidelines on how genetic design features should map to suitable glyphs, as well as how various glyphs can be connected together. However, an actual computational implementation of mapping genetic parts to suitable glyphs has previously been left to tool developers.

The SBOL Visual community has created a set of standard glyphs for a variety of commonly used genetic parts. Glyphs are proposed by members of the community and, through discussions, decisions are made about the types of part each can be used to represent. In addition to recommended glyphs for specific types of part, SBOL Visual specifies generic and alternative glyphs for many. These annotations are available as free text. For each genetic part type, a single human-editable Markdown file is created, which includes information about the mapping between recommended and alternative glyphs, and the relevant genetic parts through biological roles and molecular interactions, which are identified via commonly used ontological terms.

Ontological terms are a powerful way to represent a large amount of information via simple URIs, which can then further point to additional properties of terms and the relationships of terms with other biological concepts. Furthermore, the meaning of a term is also derived from all its parents. As a result, an ontological representation of SBOL Visual is highly desirable for computational integration and processing of visual guidelines with other existing ontologies and tools. Previously, an ontology was created for SBOL 1 and included mappings only between DNA-based parts and SO terms [8]. Since then SBOL has grown into a richer data model, with many more glyphs defined, as well as new classes of glyphs

and relationships between glyphs. Thus, the ontology needed to be reconstructed in the light of these developments.

To address this, we have developed the SBOL Visual 2 Ontology, which we use to represent the constraints about genetic circuit glyphs and their relationships to other ontological terms. Recently, an ontology called SBOL-OWL [7] was developed to provide semantic meaning for terms from the SBOL standard [3] in a machine accessible format. Using an ontological mapping, the SBOL Visual Ontology further integrates information about standardized glyphs with the SBOL standard.

2 THE SBOL VISUAL ONTOLOGY

The SBOL Visual 2 Ontology (SBOL-VO) was programmatically constructed using Markdown files that are created and managed by the SBOL Visual community. The base class in the ontology is `Glyph`, a subclass of which corresponds to an individual glyph.

A class representing a glyph may include the following Annotation properties: `rdfs:label` (name), `rdfs:comment` (description), `defaultGlyph` (the name of the glyph file), `glyphDirectory` (the folder containing the glyph), `notes` (additional free text information), `recommended` (whether the glyph is recommended or not), `prototypicalExample` (an example use of the glyph).

SBOL-VO was directly integrated with the SBOL standard via SBOL-OWL through ontological restrictions. These restrictions are created based on ontological terms associated with a glyph. For example `AptamerGlyph` is defined to be a glyph for `sbol:ComponentDefinition` entities with the role of `SO:0000031` which is a Sequence Ontology (SO) [5] term used for aptamers. These restrictions can be defined for `ComponentDefinition` entities that represent genetic parts or `sbol:Interaction` entities that represent molecular interactions. The following rules were applied to create these restrictions:

- If a glyph is associated with an SO term, a restriction is created for the `ComponentDefinition` entity using the `role` property. For example, `AptamerGlyph isGlyphOf some ComponentDefinition with a role of SO:0000031`.
- If a glyph is associated with a BioPAX [4] term, the restriction is created for the `ComponentDefinition` entity using the `type` property. For example, `ComplexGlyph`

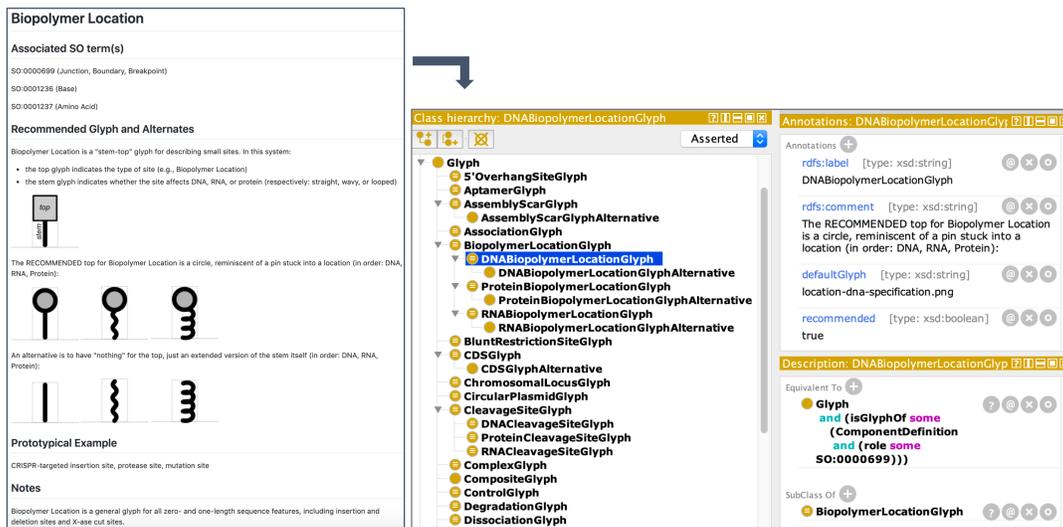


Figure 1: Free text description of recommended and alternative glyphs are used to create ontology terms and restrictions.

isGlyphOf some ComponentDefinition with a type of biopax:Complex.

- If a glyph is associated with a Systems Biology Ontology (SBO) [1] term, a restriction is created for the Interaction entity using the type property, only if the SBO term is a subclass of the biological activities and processes (i.e., SBO:0000231). For example, DegradationGlyph isGlyphOf some Interaction with a type of SBO:0000179.

Based on the representation of information in the Markdown files, hierarchical relationships between SBOL Visual terms were also created. The following rules were applied to create parent-child relationships.

- If a single glyph is included, the corresponding term is created, e.g. AptamerGlyph.
- If both a recommended glyph and an alternative glyph are included, the mapping restriction is created for the recommended term. The alternative glyph term is created as a subclass of the former and linked to the former via the isAlternativeOf property, e.g. AssemblyScarGlyph and AssemblyScarGlyphAlternative terms.
- If one generic glyph and a set of its instances (n glyphs) are included, the base class is created for the former and one recommended term is created for each instance, e.g. CleavageSiteGlyph is the parent for terms about DNA, Protein and RNA cleavage sites. If alternatives are included, they are created as subterms of the recommended terms, e.g. BiopolymerLocationGlyph (Figure 1) and its recommended and alternative terms.

The programmatic conversion was carried out using the Python programming language and using the OWLready API [6]. The ontology is available at <https://dissys.github.io/sbol-visual-ontology>.

3 CONCLUSION

SBOL-VO makes standard glyphs used for genetic circuit diagrams available to computational tools in the form of an ontology. The SBOL community heavily uses ontological terms to map genetic parts and their roles. Here, the creation of SBOL-VO and its mapping with the SBOL ontology facilitates further data integration for querying and retrieval of appropriate glyphs for genetic parts and their interactions.

4 ACKNOWLEDGEMENTS

Funded in part by NSF Expeditions in Computing Program Award #1522074 as part of the Living Computing Project. This document does not contain technology or technical data controlled under either U.S. International Traffic in Arms Regulation or U.S. Export Administration Regulations.

REFERENCES

- [1] COURTOT, M., ET AL. Controlled vocabularies and semantics in systems biology. *Molecular systems biology* 7, 1 (2011), 543.
- [2] COX, R. S., , ET AL. Synthetic biology open language visual (SBOL visual) version 2.0. *Journal of integrative bioinformatics* 15, 1 (2018).
- [3] COX, R. S., ET AL. Synthetic biology open language (SBOL) version 2.2.0. *Journal of integrative bioinformatics* 15, 1 (2018).
- [4] DEMIR, E., ET AL. The BioPAX community standard for pathway data sharing. *Nature biotechnology* 28, 9 (2010), 935.
- [5] EILBECK, K., ET AL. The sequence ontology: a tool for the unification of genome annotations. *Genome biology* 6, 5 (2005), R44.
- [6] LAMY, J.-B. Owlready: Ontology-oriented programming in Python with automatic classification and high level constructs for biomedical ontologies. *Artificial intelligence in medicine* 80 (2017), 11–28.
- [7] MISIRLI, G., ET AL. SBOL-OWL: An ontological approach for formal and semantic representation of synthetic biology information. *ACS Synthetic Biology* (2019).
- [8] QUINN, J., ET AL. Synthetic biology open language visual: an ontological use case. *Extended abstract at Bio-Ontologies* (2013).