Towards collaborative and automated development of resources for data standards in synthetic biology

Jake Sumner Ajibode¹, Jacob Beal², James Scott-Brown³ Thomas E. Gorochowski⁴, Chris J. Myers⁵, Goksel Misirli¹

¹Keele University, ²Raytheon BBN Technologies, ³University of Oxford, ⁴University of Bristol, ⁵University of Colorado Boulder

w 8q52 @ students.keele.ac.uk, jakebeal @ ieee.org, james @ jamesscottbrown.com thomas.gorochowski@bristol.ac.uk, chris.myers@colorado.edu, g.misirli@keele.ac.uk ac.uk ac.u

1 INTRODUCTION

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

53

Data standards in synthetic biology are becoming ever more important as the number of tools addressing different needs increases, such as designing genetic circuits and visualizing and storing the designs. The Synthetic Biology Open Language (SBOL) [4] has been developed to provide a mechanism for the electronic exchange and common understanding of these designs and related information. Moreover, SBOL Visual [1] standardizes the representation of genetic circuit designs via well-defined glyphs.

The SBOL community has adopted GitHub for online col-23 laboration, issue tracking and the development of specifica-24 tions. The community website builds on GitHub pages and is 25 accessible via HTTP. The SBOL data and visual standards fol-26 low an incremental release process, by applying common Git 27 related processes such as branching for incremental develop-28 ments and making and tagging releases with explicit version 29 numbering. Standard enhancement proposals are prepared 30 by community members, discussed, and finally incorporated 31 into minor or major SBOL versions. The changes between 32 each version can affect existing tools and additional mecha-33 nisms are needed for documentation and verification. As the 34 number and frequency of these new releases, there is a need 35 to automate related changes, prepare related documentation, 36 and verify the changes. Such an automated process can also 37 allow testing changes even before incorporating them into 38 formal releases. 39

An automated approach is becoming more urgent for SBOL Visual. Each glyph comes with an SVG file, providing information about how the glyph can be connected to the others in genetic circuit designs, with additional features such as height and width.

Metadata about each glyph is stored in a markdown file to 45 make the information more readable and editable for human 46 access. Each markdown file includes information about the 47 types of biological parts or interactions that the glyph can 48 be used for, via ontological terms from the Sequence Ontol-49 ogy [3] and the Systems Biology Ontology [2]. In addition, 50 these files iclude allowed alternative glyphs, free text notes 51 and other related information. These markdown files should 52

ideally be validated after each change in the context of all other glyph files and metadata from different files to prevent human error.

We previously developed SBOL Visual Ontology version 2 (SBOL-VO2) [6] to make the metadata about these glyphs machine accessible. Moreover, we developed the visual ontology web service (VOWS) to make the glyphs searchable and to find the most relevant glyph for a provided SO or SBO term for different types of biological parts and interactions. An ontological version of SBOL2 has also been developed [5]. However, the web service is currently not suitable for supporting an automated approach for new releases of the visual ontology or testing it against a specific snapshot of the files. Moreover, these resources need to be updated to incorporate the new SBOL 3 data and visual standards.

Here, we report the latest developments related to the automation process for SBOL Visual. This work involves developing a new SBOL Visual ontology for version 3 (SBOL-VO3) and improvements to the visual ontology web service for better user experience. We then demonstrate an automated workflow approach to auto-generate the visual ontology and related documentation to release standard specifications and glyphs more quickly and robustly with error checking.

2 SBOL VISUAL ONTOLOGY 3 (SBOL-VO3)

In the new version of the SBOL Visual ontology, ontological terms related to glyphs have been grouped into four main branches to represent sequence features, molecular species, molecular interactions for binary relationships, and biological processes with multiple inputs and outputs. Sequence feature terms are linked to respective SO terms, whilst SBO is used in all other terms, to indicate the types of all molecular interactions, processes and non-DNA based molecular terms. Another new feature is to capture network level representation of genetic circuits using directed edges. SBO terms are used to provide information about the roles of molecules in molecular interactions and processes, and hence the direction of edges. The hasHead and hasTail terms indicate the direction of binary interactions. For example, 'hasHead some

105

106

54

107 (role some SBO:0000642)' axiom for an inhibition interac-108 tion specifies that the the direction of the interaction starts from an entity with the inhibitor role (SBO:0000642). Simi-109 110 larly, hasIncoming and hasOutgoing properties indicate the 111 intputs and outputs for participating biological molecules. 112 This visual ontology refers to terms from the latest SBOL 3 113 data ontology (SBOL-OWL3).

114 Inline with the major changes in SBOL 3, we developed 115 the proof-of-concept SBOL-OWL3 ontology. This ontology 116 captures the rich SBOL 3 data model, which uses a graphbased approach to represent genetic circuit designs. Hence, 117 SBOL-VO3 and SBOL-OWL3 ontologies act as a machine-118 119 accessible source to connect the SBOL Visual glyphs with SBOL specific data entities. 120

3 SBOL VISUAL ONTOLOGY WEB SERVICE

123 In this work, we added new features to the previously devel-124 oped SBOL Visual Ontology Web Service [6]. Moreover, we 125 decoupled the web service from a particular version of an SBOL Visual ontology and other related SO and SBO ontolo-126 127 gies. These external resources are now loaded directly from 128 the web. Different versions of the visual ontology is held at a 129 GitHub repository. As default, the web service uses the latest 130 version of the SBOL Visual ontology and the latest glyph 131 files. However, users can also specify a particular release 132 number. Alternatively, users can provide HTTP URLs for the 133 ontology and the base glyph folder. These URLs can also be 134 for a specific GitHub branch to allow testing of resources 135 under development. Different media types for glyphs can also be requested. 136

4 AUTOMATION

121

122

137

138

150

139 Our goal is to develop a fully automated workflow to test 140 incremental or major updates and use them directly via the SBOL visual web service (Figure 1). To facilitate this ap-141 142 proach, we aim to use GitHub actions. As a result, specific GitHub-related actions such as new releases or commits to 143 144 a particular branch can result in the auto-generation of the 145 SBOL Visual ontology, related documentation, and the use of 146 the visual glyphs via the web service directly. We developed 147 scripts that can be linked to GitHub actions so that the SBOL 148 community to take advantage of these resources in a timely 149 fashion.

5 CONCLUSION 151

152 Ontologies are ideal to capture domain knowledge. Here, we used this approach to provide machine accessible represen-153 tation of the SBOL 3 data and visual standards in the form of 154 ontologies. In order to help the SBOL community deal with 155 156 incremental and major changes, we adopt an automated ap-157 proach based on GitHub actions. This automation enables 158 testing changes rapidly and releasing them to the community 159

Jake S. Ajibode et al.

quickly. The improvements enhance the user experience via different additional options when using the web service. The SBOL community is working on parametric glyphs that can be customised for different properties such as colour, width, and height. Our goal is to incorporate additional metadata about these parameters into the SBOL Visual ontology and use the web service to serve parametric glyphs for design and layout tools in the future.

6 ACKNOWLEDGEMENTS

We thank SBOL Industrial Consortium for funding this work, and the SBOL community for continuous support sand feedback. 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] BAIG, H., FONTANAROSSA, P., KULKARNI, V., MCLAUGHLIN, J., VAIDYANATHAN, P., BARTLEY, B., BHAKTA, S., BHATIA, S., BISSELL, M., CLANCY, K., COX, R. S., MORENO, A. G., GOROCHOWSKI, T., GRUNBERG, R., Lee, J., Luna, A., Madsen, C., Misirli, G., Nguyen, T., Novere, N. L., PALCHICK, Z., POCOCK, M., ROEHNER, N., SAURO, H., SCOTT-BROWN, J., SEXTON, J. T., STAN, G.-B., TABOR, J. J., TERRY, L., VILAR, M. V., VOIGT, C. A., WIPAT, A., ZONG, D., ZUNDEL, Z., BEAL, J., AND MYERS, C. Synthetic biology open language visual (sbol visual) version 2.3. J. Integr. Bioinform. (2021), 20200045.
- COURTOT, M., JUTY, N., KNÜPFER, C., WALTEMATH, D., ZHUKOVA, A., [2] DRÄGER, A., DUMONTIER, M., FINNEY, A., GOLEBIEWSKI, M., HASTINGS, J., HOOPS, S., KEATING, S., KELL, D. B., KERRIEN, S., LAWSON, J., LISTER, A., LU, J., MACHNE, R., MENDES, P., POCOCK, M., RODRIGUEZ, N., VILLEGER, A., Wilkinson, D. J., Wimalaratne, S., Laibe, C., Hucka, M., and LE NOVÈRE, N. Controlled vocabularies and semantics in systems biology. Mol. Syst. Biol. 7, 1 (2011), 543.
- [3] EILBECK, K., LEWIS, S. E., MUNGALL, C. J., YANDELL, M., STEIN, L., DURBIN, R., AND ASHBURNER, M. The sequence ontology: a tool for the unification of genome annotations. Genome Biol. 6, 5 (2005), R44.
- [4] MCLAUGHLIN, J. A., BEAL, J., MISIRLI, G., GRÜNBERG, R., BARTLEY, B. A., Scott-Brown, J., Vaidyanathan, P., Fontanarrosa, P., Oberortner, E., WIPAT, A., GOROCHOWSKI, T. E., AND MYERS, C. J. The synthetic biology open language (sbol) version 3: Simplified data exchange for bioengineering. Front Bioeng. Biotechnol. 8 (2020), 1009.
- [5] MISIRLI, G., TAYLOR, R., GOñI-MORENO, A., MCLAUGHLIN, J. A., MYERS, C., Gennari, J. H., Lord, P., and Wipat, A. Sbol-owl: An ontological approach for formal and semantic representation of synthetic biology information. ACS Synth. Biol. 8, 7 (2019), 1498-1514.
- MISIRLI, G., BEAL, J., GOROCHOWSKI, T. E., STAN, G.-B., WIPAT, A., AND [6] MYERS, C. J. Sbol visual 2 ontology. ACS Synth. Biol. 9, 4 (04 2020), 972-977.

195

196

197

198

199

200

201

202

203

204

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

209 210

211 212

Automation for data standards in synthetic biology

IWBDA 2021, September 20-24, 2021, Virtual



Figure 1: A use case involving the use the automated approach. A user makes a new release due to a change in metadata about an SBOL Visual glyph. As a result, the SBOL Visual ontology and the online documentation is auto-generated. Tools can start using this change immediately. Alternatively, they can start working on a specific version of the ontology and glyphs. Tools can either query the ontology file or utilise the web service to find the most suitable glyphs for given SO or SBO terms.

	299
	300
	301
	302
	303
	304
:	305
1	306
1	307
1	308
1	309
1	310
1	311
1	312
1	313
	314
	315
1	316
:	317
3	318