Describing engineered biological systems with SBOL3 and ShortBOL2

Matthew Crowther¹, Lewis Grozinger¹, James McLaughlin¹, Göksel Mısırlı², Jacob Beal³, Bryan A. Bartley¹, Angel Goñi-Moreno¹, Anil Wipat¹

> ¹Newcastle University, ²Keele University, ³Raytheon BBN Technologies angel.goni-moreno@ncl.ac.uk,anil.wipat@ncl.ac.uk

1 INTRODUCTION

Data standards are essential to exchange information about the engineering of biological systems. The Synthetic Biology Open Language (SBOL) is a community-driven standard that facilitates the exchange of data relating to the design, implementation, testing and refinement of engineered biological systems [4]. Versions 1 and 2 of SBOL have gained widespread adoption, with over 170 developers, 29 SBOL supporting software tools and 42 institutions involved in their development and deployment (as of June 2020). Recently, SBOL was refactored to simplify its data model, resulting in the release of the SBOL3 specification [1].

SBOL data is created using both bespoke software tools and through the use of the SBOL specific software libraries such as pySBOL [2] and libSBOLj [6]. However, the creation of SBOL data using these libraries is often limited to tool developers with programming skills. To address this issue, numerous graphical computer aided design (CAD) tools have been developed to allow non-programmers to capture data from the design, build, test and learn (DBTL) cycle in SBOL format. However, visual design tools often support only a subset of features of the SBOL data model. Visual editing can be a slow and rather manual process that does not scale well for large designs such as genomes.

Previously, we have described a language, ShortBOL, which provides a human readable/writable short-hand for describing biological designs in SBOL [3]. ShortBOL was developed for synthetic biologists familiar with the SBOL data model who wish to rapidly describe synthetic biology designs using a text based scripting language instead of using a traditional programming language. Here, we describe a new release of ShortBOL, version 2.0. This new version sees the introduction of two new modes of use, as well as support for SBOL 3.0. The new modes of use of the language alter the levels of abstraction to consider two major categories of synthetic biologists with different expertise and backgrounds: (i) User mode for synthetic biologists with little to no understanding of the SBOL data model, and (ii) Developer mode for synthetic biology developers, familiar with programming and the SBOL data model. The new ShortBOL user-mode syntax simplifies the generation of SBOL data for the first category of users. The latter category of users

are supported through full access to the terms of the data model. SBOL 3.0 support is provided for both these different levels of abstraction. We illustrate these two modes using examples developed in SBOL3. ShortBOL as a service can be accessed from http://shortbol.org and the code is available at https://github.com/intbio-ncl/shortbol.

2 EXPLORING SBOL3 USING SHORTBOL2

ShortBOL was originally designed to be easy to use by synthetic biologists who understand the fundamentals of the SBOL data model. A standard template library is incorporated within ShortBOL, allowing the introduction of new ShortBOL language terms and different aspects of genetic designs to be generated. SBOL3 support is provided via additional templates within the ShortBOL2 application.

We have also developed a tutorial and set of examples that illustrate the use of SBOL3, and its comparison to SBOL2, to aid developers in transitioning to the new version of the standard. For example, Figure 1 shows a side by side comparison of the TetR inverter module of the classical genetic toggle switch example [5] written using ShortBOL2 in developer mode for both SBOL2 and SBOL3 (Fig.1.). The ShortBOL2 application features support for generating both SBOL2.0 and SBOL3, the choice of which is selected by the user from the application menu. Depicting SBOL as ShortBOL allows the structure of SBOL3 to be viewed and also compared to the equivalent representation in SBOL2.0. SBOL is a data exchange format and SBOL data produced by different tools are all compatible. However, with the introduction of SBOL3 there is not currently compatibility between the SBOL2 and SBOL3 data models. As tooling for the data model is developed converters will become available.

3 ABSTRACTING SBOL3 USING SHORTBOL -USER MODE

The SBOL3 data model can still be unwieldy for designers unfamiliar with computational data representations. Short-BOL2 also introduces further templates that provide a more abstract version of the ShortBOL language, aimed at the average user who does not wish to work with the SBOL3 data model at a detailed level. This mode makes designs shorter and easier to understand. As an example, Figure 2 shows a

IWBDA 2020, Aug 2020, Online



Figure 1: A TetRInverter device shown in ShortBOL for SBOL 2.0 and SBOL3. The change from the use of ModuleDefinition to a single FunctionalEntity, introduced in SBOL3, is illustrated by a side-by-side comparison of the designs.

ShortBOL2 representation of the SBOL3 approach to defining the LacI inverter module of the genetic toggle switch. These ShortBOL templates allow for a more succinct representation of the design, which is then expanded out into the full representation on conversion into SBOL. Developers are still free to use the more expansive representation, if they wish, which is closer to the SBOL3 data model and allows expression of details that are not part of typical use patterns. The ShortBOL user mode has slight constrictions and is achieved using composition to create some SBOL3 objects behind the scenes. However, these constrictions are only present when working with very niche aspects of the data model.



Figure 2: A ShortBOL2 design for the LacI inverter module demonstrating the inhibition of TetR protein production by LacI protein in Developer mode and the more abstract version in User mode. Crowther and Grozinger, et al.

4 CONCLUSIONS AND FUTURE WORK

We have expanded ShortBOL1.0 to support SBOL3. Short-BOL2 aims to provide an easy to use tool for the composition of SBOL3 related data. This latest version introduces a user mode providing a more abstract version of ShortBOL where the SBOL data model can be used in a less granular fashion. ShortBOL2 allows the structure of SBOL3 to be explored succinctly by developers wishing to become familiar with the SBOL data model, who can gain exposure to the terminology and approach without having to work with the SBOL code libraries. Furthermore, users can also produce SBOL data without being exposed to the full details of the SBOL3 data model. ShortBOL makes it easier to prototype SBOL3 designs, and in the future it may be possible to simplify the ShortBOL syntax even further to provide a language that shields a user entirely from the SBOL3 data model.

5 ACKNOWLEDGEMENTS

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. The authors of this work are supported by The Engineering and Physical Sciences Research Council grants EP/J02175X/1, EP/R003629/1, EP/N031962/1 (J.M. and A.W.), and EP/R019002/1 (A.G.-M.), EPSRC studentship 34000024085 (M.C.), and the European CSA 820699 (A.G.-M. and A.W.). M.C. is supported by Doulix Ltd. J.B. and B.B. are supported in part by the Air Force Research Laboratory (AFRL) and DARPA under contract FA875017CO184.

REFERENCES

- [1] BAIG, H., FONTANARROSA, P., KULKARNI, V., MCLAUGHLIN, J., VAIDYANATHAN, P., MYERS, C., BARTLEY, B., BEAL, J., CROWTHER, M., GOROCHOWSKI, T., ET AL. Synthetic biology open language (sbol) version 3.0.0. *Journal of Integrative Bioinformatics* 17, 2 (2020), 17.
- [2] BARTLEY, B. A., CHOI, K., SAMINENI, M., ZUNDEL, Z., NGUYEN, T., MYERS, C. J., AND SAURO, H. M. pysbol: a python package for genetic design automation and standardization. ACS synthetic biology 8, 7 (2018), 1515– 1518.
- [3] CROWTHER, M., GROZINGER, L., POCOCK, M., TAYLOR, C. P., MCLAUGHLIN, J. A., MISIRLI, G., BARTLEY, B. A., BEAL, J., GONI-MORENO, A., AND WIPAT, A. Shortbol: A language for scripting designs for engineered biological systems using synthetic biology open language (sbol). ACS Synthetic Biology 9, 4 (2020), 962–966.
- [4] GALDZICKI, M., CLANCY, K. P., OBERORTNER, E., POCOCK, M., QUINN, J. Y., RODRIGUEZ, C. A., ROEHNER, N., WILSON, M. L., ADAM, L., ANDERSON, J. C., ET AL. The synthetic biology open language (sbol) provides a community standard for communicating designs in synthetic biology. *Nature biotechnology 32*, 6 (2014), 545–550.
- [5] GARDNER, T. S., CANTOR, C. R., AND COLLINS, J. J. Construction of a genetic toggle switch in escherichia coli. *Nature 403*, 6767 (2000), 339–342.
- [6] ZHANG, Z., NGUYEN, T., ROEHNER, N., MISIRLI, G., POCOCK, M., OBERORT-NER, E., SAMINENI, M., ZUNDEL, Z., BEAL, J., ET AL. libsbolj 2.0: a java library to support sbol 2.0. *IEEE life sciences letters* 1, 4 (2015), 34–37.