

The Synthetic Biology Open Language

Goksel Misirli, James McLaughlin, Matthew Pocock, Anil Wipat

School of Computing Science, Newcastle University, UK

Synthetic biology has great potential to develop novel biological systems. As in many engineering disciplines, synthetic biologists also use techniques such as standardisation, modularity and abstraction, in order to develop predictable applications and to scale up designs. However, engineering biological systems is not a trivial task. These systems can be complex and they can have very large design spaces. Identified possible solutions are verified through design-build-test cycles, which may involve several tasks and require inputs from different researchers. Computational tools are especially useful to identify these solutions and to facilitate the execution of these tasks. However, tools often focus upon particular aspects of design processes and work in isolated manner. Data standards are needed to unambiguously exchange information between these tools in order to facilitate the execution of complex workflows, and to increase the reproducibility of designs across different labs.

The Synthetic Biology Open Language (SBOL) is a data standard that has been specifically developed to exchange biological system designs. Using this language, biological systems can be defined with individual design components, which can represent DNA, RNA, proteins or metabolites. SBOL facilitates the reuse of these components and existing complex designs. Desired roles of each component and possible interactions can also be specified. Moreover, SBOL allows incorporating application specific information.

SBOL utilises existing Semantic Web resources to represent, store and retrieve data. Designs are serialized using the RDF/XML format. Although SBOL defines its own terms, it reuse existing controlled vocabularies and ontologies, where possible. For example, terms from the BioPAX ontology are used to indicate types of design components and Sequence Ontology terms are used to indicate roles of DNA-based components. Application specific data can either be embedded as RDF triples within SBOL entities or defined as RDF resources at the top level. Although, such custom data may not be understood directly by SBOL libraries, they are retained during read and write processes.

SBOL is aimed directly for tool developers; however, through the development of SBOL compliant tools wet-lab biologists can take the advantage of many useful tools in order to create and refine their designs. The development of SBOL is carried out openly by the SBOL Developers group¹. This development is coordinated by elected editors and an SBOL chair, and a steering committee. Currently, SBOL Developers include more than 120 members from 52 organisations in 15 countries. Libraries in different languages are being developed openly and they are available freely².

¹<http://sbolstandard.org>

²<https://github.com/SynBioDex>