

Steps Towards a Curated SBOL Repository

Creating and Annotating Models from Literature

Zach Zundel¹, Chris Myers²

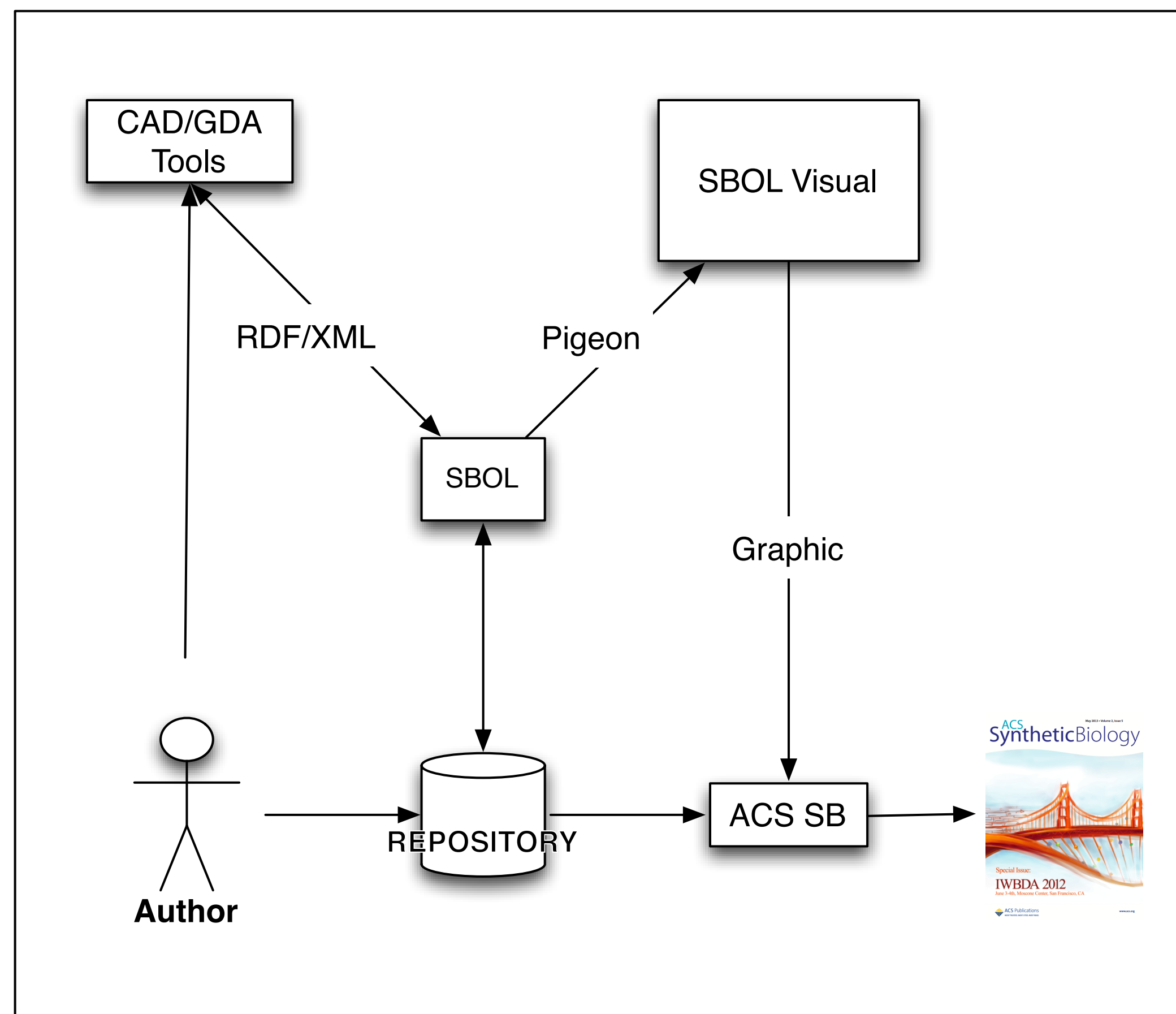
¹Departement of Bioengineering,

²Department of Electrical and Computer Engineering,
University of Utah



Introduction

- Standards are key to the success of systems and synthetic biology
- Allow for easy exchange and standard expression of data across labs and projects
- SBOL (Synthetic Biology Open Language) [1] is an evolving standard for the exchange of genetic data
- SBOL allows creation of hierarchical, modular representations of genetic structure and function
- Many different tools handle different components of the SBOL creation process
- Current tools must be examined to determine a usable workflow for creating SBOL
- A widely-applicable workflow would facilitate adoption of SBOL by the larger community
- Curated repositories provide a template for the type of workflow to be developed
- There is a clear need for a standard to ensure that uniform data is expressed
- This standard should be open-source and allow for free exchange of data
- Reduces the load for researchers who may be approached by those replicating their work
- End goal is SBOL adoption by journals so all publications have supplemental annotated SBOL models



Current Issues

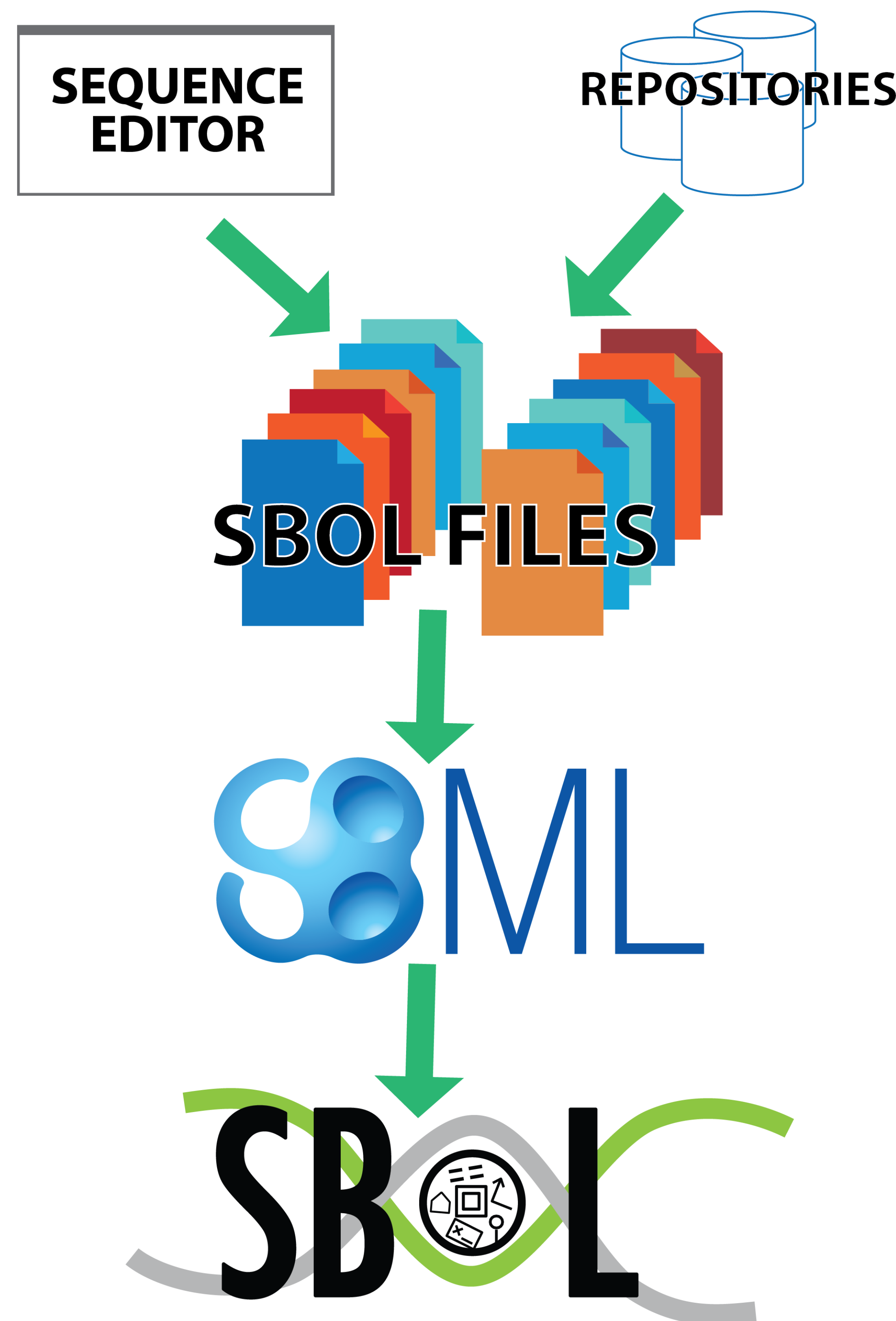
- No standard way of expressing data in synthetic biology publications
- Variation here often leads to missing valuable pieces of information
- Information that is present could be present in an obtuse or difficult-to-use form
- Oftentimes, crucial information is missing all together
- Data is often difficult to obtain from researchers after publication
- When given, data is not always formatted in a standard manner
- Even when a standard is used, it is not always possible to ensure that implied homologues are valid
- Lack of requirements for publication allows many of these issues to continue

Drawing Inspiration from Extant Analogous Systems

- Many systems exist that address similar issues with other information
- The curated BioModels repository gives models where specific species are correlated to ensure clarity
- The journal adoption of a GenBank standard for expressing sequences has facilitated data exchange
- Repositories of usable data fragments facilitate the creation of novel systems
- In each of these systems, data is presented in a uniform way and curated by some entity
- Adoption of SBOL and creation of a curation pathway would aid further research

General Workflow

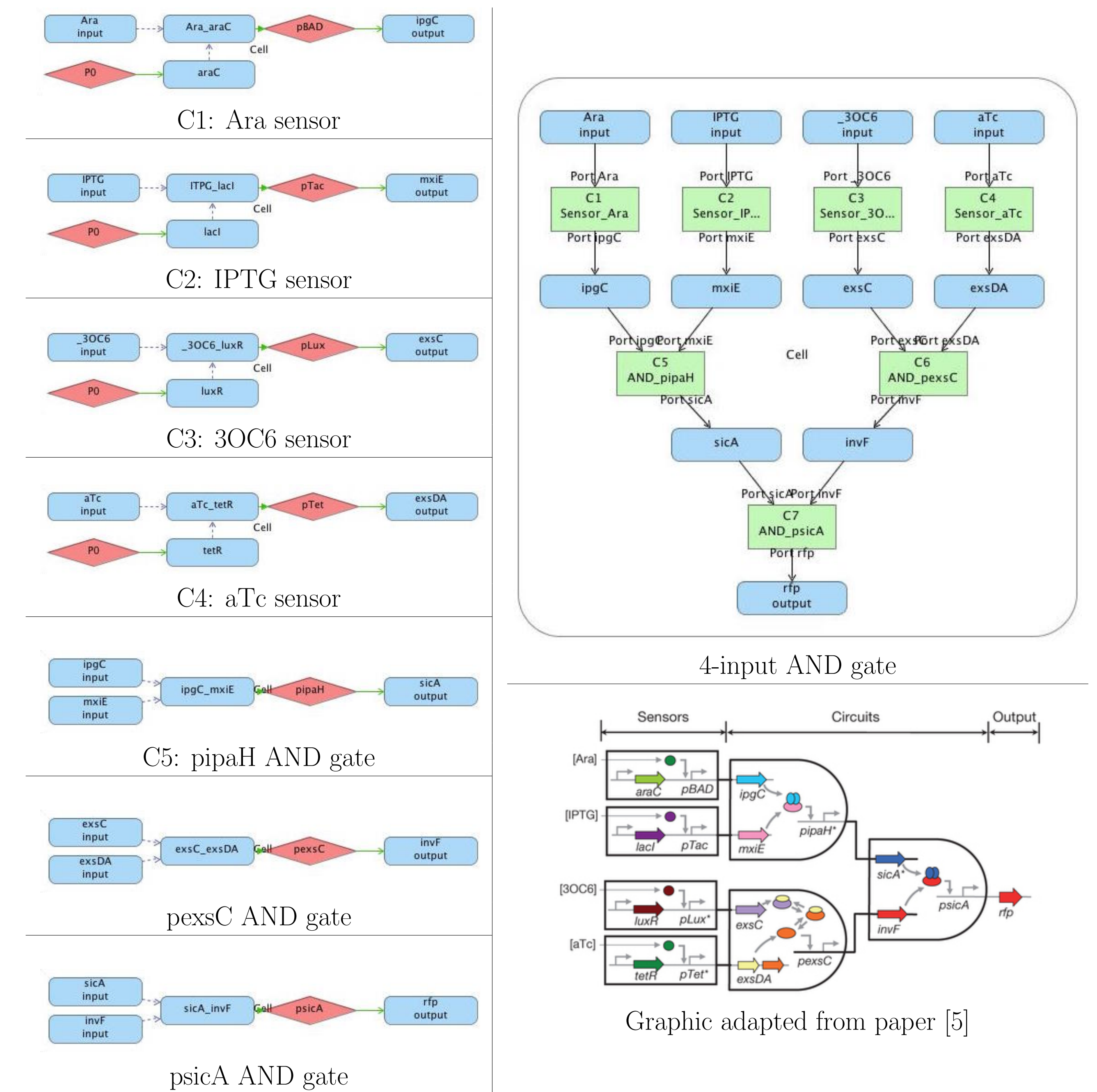
1. Create coding sequences using a coding sequence editor or download SBOL files from a repository
2. Create a model describing interactions in SBML [2] or another modeling language
3. Annotate model using the SBOL from step 1 [3]
4. Export finished model as complete SBOL
5. Annotate species using an ontology to ensure that homologous species are explicitly defined



Specific Workflow

1. Sequences are prepared in DNAPlotLib
2. SBML models created in iBioSim [4]
3. iBioSim is used to annotate SBML models with SBOL
4. Finally, iBioSim exports SBOL 2.0 models containing annotated genetic data and interaction descriptions

Creating a Test Case



Discussion

- The test case above included a few omissions which made recreating the model difficult
- At the least, the publication should include sequence and schematic data
- Further research would allow for experimentation with additional workflows
- SBOL must expand to include data about growth conditions and host organisms

Acknowledgements

Supported by National Science Foundation Grants CCF-1218095 and DBI-1356041.

References

- [1] B. Bartley, J. Beal, K. Clancy, G. Misirli, N. Roehmer, E. Oberortner, M. Pocock, M. Bissell, C. Madsen, T. Nguyen, Z. Zhang, J. H. Gennari, C. Myers, A. Wipat, and H. Sauro, "Synthetic biology open language (sbol) version 2.0.0," *Journal of Integrative Bioinformatics*, vol. 12, Aug. 2015.
- [2] M. Hucka, A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, and et al., "The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models," *Bioinformatics*, vol. 19, pp. 524–531, Mar. 2003.
- [3] N. Roehmer and C. Myers, "A methodology to annotate systems biology markup language models with the synthetic biology open language," *ACS Synthetic Biology*, vol. 3, pp. 57–66, Sept. 2013.
- [4] C. Madsen, C. Myers, T. Patterson, N. Roehmer, J. Stevens, and C. Winstead, "Design and test of genetic circuits using iBiosim," *Design and Test of Computers, IEEE*, vol. 29, no. 3, pp. 32–39, 2012.
- [5] T. Moon, C. Louw, A. Tamsir, B. Stanton, and C. Voigt, "Genetic programs constructed from layered logic gates in single cells," *Nature*, vol. 491, no. 7423, pp. 249–253, 2012.