Check for updates

DATA NOTE

A fully featured COMBINE archive of a simulation study on syncytial mitotic cycles in *Drosophila* embryos [version 1; peer review: 1 approved, 2 approved with reservations]

Martin Scharm, Dagmar Waltemath

Department of Systems Biology and Bioinformatics, Institute of Computer Science, University of Rostock, Rostock, Germany

Abstract
COMBINE archives are standardised containers for data files related to a simulation study in computational biology. This manuscript describes a fully featured archive of a previously published simulation study, including (i) the original publication, (ii) the model, (iii) the analyses, and (iv) metadata describing the files and their origin. With the archived data at hand, it is possible to reproduce the results of the original work. The archive can be used for both, educational and research purposes. Anyone may reuse, extend and update the archive to make it a valuable resource for the scientific community.

Keywords
COMBINE , data , containers

This article is included in the Data: Use and Reuse collection.

Open Peer Review

Reviewer Status ? ✔ ?

	Invited Reviewers		
	1	**2**	**3**
version 1 29 Sep 2016	? report	✔ report	? report

1 **Lars Juhl Jensen**, University of Copenhagen, Copenhagen, Denmark

2 **Laurence Calzone**, Institut Curie, Paris, France
   Mihaly Koltai, Institut Curie, Paris, France

3 **Alan Garny**, University of Auckland, Auckland, New Zealand

Any reports and responses or comments on the article can be found at the end of the article.

How to cite this article: Scharm M and Waltemath D. **A fully featured COMBINE archive of a simulation study on syncytial mitotic cycles in *Drosophila* embryos [version 1; peer review: 1 approved, 2 approved with reservations]** F1000Research 2016, **5**:2421 (https://doi.org/10.12688/f1000research.9379.1)

First published: 29 Sep 2016, **5**:2421 (https://doi.org/10.12688/f1000research.9379.1)

Introduction

In systems biology and systems medicine, the steadily increasing size and complexity of simulation studies pose additional challenges to sharing reproducible results[1]. Repeated mentions of problems with replication and reproducibility[2–4] led to new standards, tools, and methods for the transfer of reproducible simulation studies[5–9]. Several projects and initiatives already deal with reproducibility issues, such as COMBINE (co.mbine.org), FAIRDOM (fair-dom.org), and the Reproducibility Initiative (reproducibilityinitiative.org).

The *Computational Modeling in Biology Network* (COMBINE) coordinates the development of standard formats for various aspects of a simulation study: The Systems Biology Markup Language (SBML)[10] and CellML[11] encode the mathematical models; the Systems Biology Graphical Notation (SBGN)[12] encodes the visual representation of models; the Simulation Experiment Description Markup Language (SED-ML)[13] encodes the simulation recipes; and the Systems Biology Result Markup Language (SBRML)[14] encodes numerical data and simulation results.

Today's studies consist of multiple, heterogeneous, and sometimes distributed data files, leading to the challenge of exchanging complete and thus reproducible results. To close this gap, the COMBINE community developed the COMBINE archive[8]. A COMBINE archive is a single file that aggregates all data files and information necessary to reproduce a simulation study in computational biology. The skeleton of a COMBINE archive consists of a manifest and a metadata file, specified by the Open Modeling EXchange format (OMEX).

Here we describe a fully featured COMBINE archive, which encodes an investigation of the syncytial mitotic cycles in *Drosophila* embryos[15]. The study published by Calzone *et al.* proposes a dynamical model for the molecular events underlying rapid, synchronous, syncytial nuclear division cycles in *Drosophila* embryos. This particular study was chosen for several reasons. Firstly, the paper, the documentation, and the related data are openly accessible. Secondly, the model is available in two standard formats: The CellML encoding is available from the Physiome Model Repository[16] at models.cellml.org/exposure/1a3f36d015121d5596565fe7d9afb332 and the SBML encoding is available from BioModels[17] at www.ebi.ac.uk/biomodels-main/BIOMD0000000144. Thirdly, both model files are already curated, which increases the level of trust. Fourthly, the model describes a common biological system (cell cycle). Thus, the basic mechanisms of the encoded biology should be familiar to many researchers, reducing the effort of understanding the example.

This archive contains files that are openly available for download, as well as previously unpublished files that were generated using COMBINE-compliant software tools (see Section Materials and methods). When executed, it reproduces the original findings by Calzone *et al.*

Materials and methods

The fully featured COMBINE archive was created in three subsequent steps. Firstly, all available materials relating to the study were automatically retrieved from an online resource (initial archive). Secondly, the data files were organised into subdirectories, following the different aspects of a simulation study (documentation, model, experiment, result). Thirdly, missing files were manually retrieved from web resources or created using COMBINE-compliant software tools. The three steps are described in the following.

Retrieving an initial COMBINE archive

The initial version of the COMBINE archive was generated using the web-based software tool M2CAT[18] Version 0.1 (m2cat.sems.uni-rostock.de). Among the suggested archives for the work by Calzone *et al.*, we chose the simulation study containing a CellML model and a visualisation of the model in three different formats (PNG, SVG, AI). M2CAT automatically generated the initial COMBINE archive from these files. It also added metadata to the archive, such as annotations to creators, contributors, and modification times. M2CAT retrieved this metadata from the corresponding GIT project in the Physiome Model Repository (`git log`).

Organising the COMBINE archive

For convenience, the files inside the COMBINE archive were structured in subfolders. The initial archive was therefor imported into the CombineArchiveWeb application (WebCAT,[9]) Version 0.4.13 (webcat.sems.uni-rostock.de). WebCAT is a web interface to display and modify the files contained in an archive, together with metadata and file structures. The files inside the archive were organised in four directories, which reflect the different aspects of a simulation study:

- `documentation/`: files that describe and document the model and/or experiment (*empty*)

- `model/`: files that encode and visualise the biological system (*4 files*)

- `experiment/`: files that encode the *in silico* setup of the experiment (*empty*)

- `result/`: files that result from running the experiment (*empty*)

All files in the initial archive were stored in the `model/` directory. However, these files alone are not sufficient to reproduce the study.

Extending the COMBINE archive

To make the encoded study reproducible, the COMBINE archive needs to be extended with additional files.

The article is typically the central object of a research study. For this study, the original publication by Calzone *et al.*, together with available supplementary information, was retrieved from the homepage of the journal *Molecular Systems Biology* (msb.embopress.org/content/3/1/131). Using WebCAT, the files were uploaded to the `documentation/`directory of the archive. The automatically added metadata was adjusted to attribute the authors of the publication and to state when and where the files were downloaded. In the background, WebCAT encoded the metadata in RDF/XML and added it to the archive.

The model is not only available in CellML format, but also in SBML format. The SBML file was retrieved from BioModels (www.ebi.ac.uk/biomodels-main/download?mid=BIOMD0000 000144, SBML Level 2 Version 1) and uploaded to the model/ directory. Again, the metadata was corrected to attribute the original authors, curators, and contributors, as stated on the BioModels website (www.ebi.ac.uk/biomodels-main/BIOMD0000000144) and in the model document.

The simulation description is essential to run the experiment. It defines the simulation environment and the output of the *in silico* execution. As no simulation description was found in any of the

open repositories, an initial version was created using the SED-ML Web Tools (SWT) Version 2.1 (bqfbergmann.dyndns.org/SED-ML_Web_Tools). SWT takes the model files and creates a default simulation description with standard settings. For this study, a default SED-ML file encodes instructions to generate 66 plots and a data table. Each plot describes the change of concentration in one species of the model. The data table contains all numerical values. Based on the default script, a second SED-ML file (Calzone2007-simulation-figure-1B.xml) was generated to reassemble Figure 1B of the original publication. Using WebCAT, both SED-ML scripts were added to the experiment/ directory of the archive. The metadata for the new files was added.
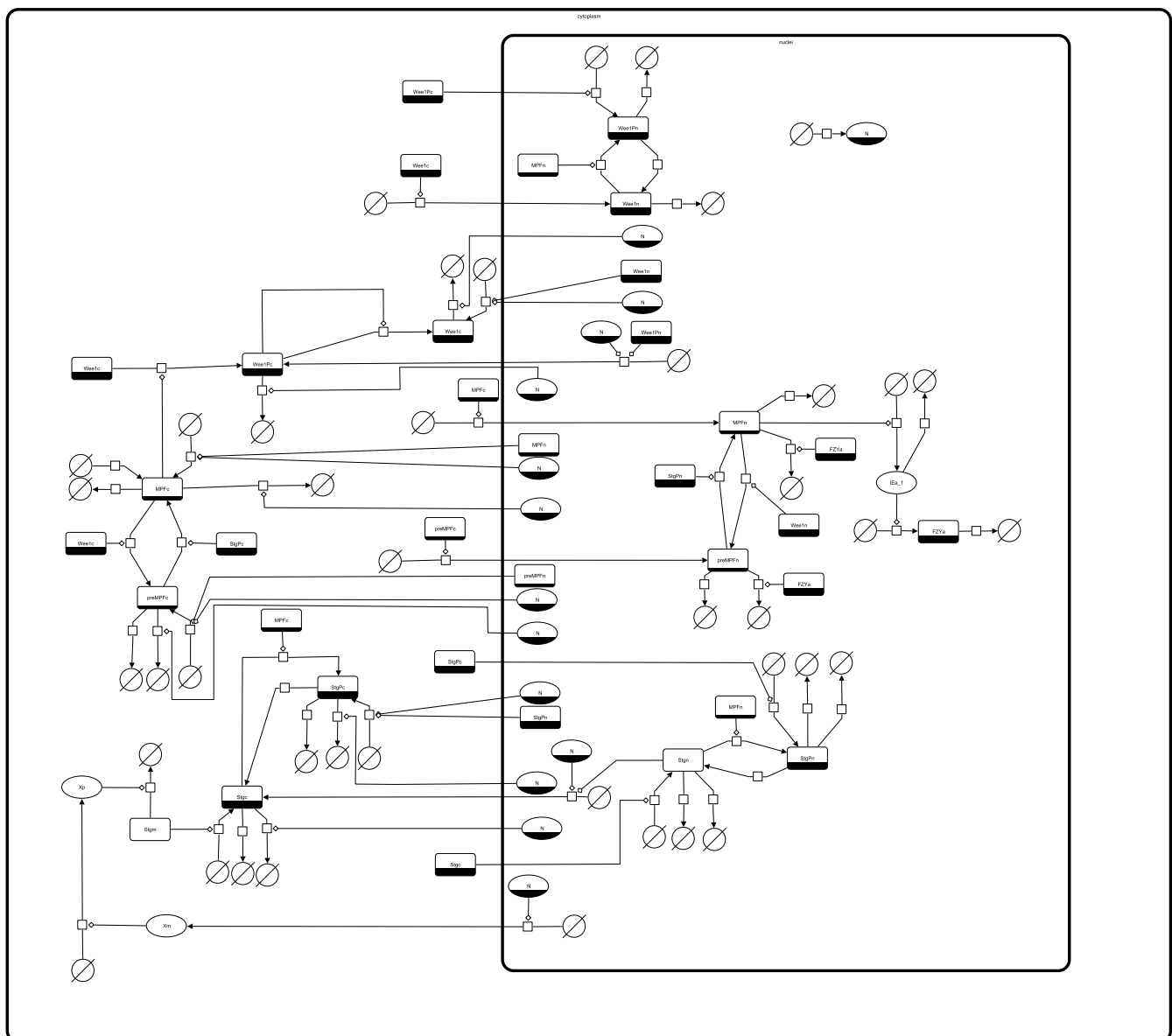

Figure 1. Visualisation of the model. This figure shows the SBGN-PD compliant reaction network, as encoded in the SBML model obtained from BioModels. The figure was generated and modified using SBGN-ED.

The simulation results reflect the behaviour of a model under certain conditions. The script defined in `Calzone2007-simulation-figure-1B.xml` was loaded into SWT and into the stand-alone software program COPASI Version 4.15 Build 95[19]. The plots generated by both tools show that the developed *in silico* experiment reproduces the results from the paper. Using WebCAT, the figures produced by SWT and COPASI were uploaded and added to the `result/` directory of the archive. Metadata, such as the versions of the software tools, was added accordingly.

The visualisation of a model helps to understand the encoded biological system. For this study, an SBGN-compliant visualisation of the model was created using SBGN-ED Version 1.5.1[20] together with VANTED Version 2.1.0[21]. SBGN-ED generated an automatic layout of the uploaded SBML model, which was then improved manually. The resulting Figure 1 was exported in different formats (GraphML[22], GML (www.fim.uni-passau.de/index.php?id=17297&L=1), PNG image, PDF, and SBGN-ML[23]). Using WebCAT the files were uploaded to the `model/sbgn` directory and metadata was provided.

Data description

The archive consists of 21 files (Table 1). Among these files are the `manifest.xml` and the `metadata.rdf`, which form

Table 1. Content of the fully featured COMBINE archive. The table lists all files included in the presented COMBINE archive together with formats and descriptions. The indentation indicates the directory structure used to organise the files in the archive.

File	Format	Description
manifest.xml	Omex	Skeleton, automatically generated by WebCAT
metadata.rdf	Omex	Skeleton, automatically generated by WebCAT
README.md	Markdown	Human readable information for users stumbling upon the archive
model/		
BIOMD0000000144.xml	SBML L2V1	origin: www.ebi.ac.uk/biomodels-main/download?mid=BIOMD0000000144
calzone_2007.svg	SVG	origin: models.cellml.org/workspace/calzone_thieffry_tyson_novak_2007
calzone_2007.ai	Illustrator	origin: models.cellml.org/workspace/calzone_thieffry_tyson_novak_2007
calzone_2007.png	PNG	origin: models.cellml.org/workspace/calzone_thieffry_tyson_novak_2007
calzone_thieffry_tyson_novak_2007.cellml	CellML 1.0	origin: models.cellml.org/workspace/calzone_thieffry_tyson_novak_2007
sbgn/Calzone2007.gml	GML	SBGN compliant figure generated using SBGN-ED
sbgn/Calzone2007.graphml	GraphML	SBGN compliant figure generated using SBGN-ED
sbgn/Calzone2007.pdf	PDF	SBGN compliant figure generated using SBGN-ED
sbgn/Calzone2007.png	PNG	SBGN compliant figure generated using SBGN-ED
sbgn/Calzone2007.sbgn	SBGN-ML	SBGN-ML encoded figure generated using SBGN-ED
experiment/		
Calzone2007-default-simulation.xml	SED-ML L1V1	Simulation description generated using SED-ML Web Tools
Calzone2007-simulation-figure-1B.xml	SED-ML L1V1	Simulation description generated using SED-ML Web Tools based on Calzone2007-default-simulation.xml
documentation/		
Calzone2007.pdf	PDF	Scientific publication *"Dynamical modeling of syncytial mitotic cycles in Drosophila embryos"* obtained from msb.embopress.org/content/3/1/131
Calzone2007-supplementary-material.pdf	PDF	Supplementary information for the publication obtained from msb.embopress.org/content/3/1/131
result/		
Fig1B-bottom-COPASI.svg	SVG	Image generated by executing Calzone2007-simulation-figure-1B.xml on BIOMD0000000144.xml in COPASI
Fig1B-top-COPASI.svg	SVG	Image generated by executing Calzone2007-simulation-figure-1B.xml on BIOMD0000000144.xml in COPASI
Fig1B-bottom-webtools.png	PNG	Image generated by executing Calzone2007-simulation-figure-1B.xml on BIOMD0000000144.xml in SED-ML Web Tools
Fig1B-top-webtools.png	PNG	Image generated by executing Calzone2007-simulation-figure-1B.xml on BIOMD0000000144.xml in SED-ML Web Tools

the skeleton of the archive. The manifest lists the files included in the archive. The metadata file contains additional information about the files in the archive, such as creators and descriptions. A third file, README.md, contains a description for visitors of the GitHub repository, where the archive is being developed (github.com/SemsProject/CombineArchiveShowCase). The remaining 18 files are organised in four directories, cmp. Section Organising the COMBINE archive. The original publication (PDF) is stored in the documentation/ directory. The encodings of the model (CellML, SBML, graph formats) are stored in the model/ directory. The simulation descriptions (SED-ML) are stored in the experiment/ directory. The simulation results (SVG, PNG) are stored in the result/ directory.

The latest version of the compiled COMBINE archive can be accessed through our web server at scripts.sems.uni-rostock.de/getshowcase.php.

Data validation
The COMBINE archive described in this data note reproduces the results of the study published by Calzone *et al*. To validate the reproducibility, we executed the archive in different simulation tools. For example, the encoded simulation study can be executed in COPASI, cmp. Figure 2(b). The archive can also be loaded to the SWT by opening a specific URL (bqfbergmann.dyndns.org/SED-ML_Web_Tools/Home/SimulateUrl?url=http://scripts.sems.uni-rostock.de/getshowcase.php). The simulation results will immediately be shown in the web browser, cmp. Figure 2(c).

Moreover, users reported a successful reproduction of the simulation results using Tellurium[24] (github.com/SemsProject/CombineArchiveShowCase/pull/2).

Conclusions
The presented COMBINE archive provides a reproducible simulation study for a previously published model on syncytial mitotic cycles in *Drosophila* embryos[15]. The archive contains several files that were collected from online resources, e. g. the CellML model from the Physiome Model Repository or the scientific publication from the publisher's website. It also provides new files that did not exist previously, e. g. a SED-ML file to encode the simulation setup for Figure 1B of the original publication.

This fully featured archive allows scientists to reproduce the results obtained by Calzone *et al*. in software tools that can read COMBINE archives. For example, the archive was successfully executed in the SED-ML Web Tools and Tellurium. Figure 2 shows that the developed study is able to reproduce the original results.

This data note describes the fully featured COMBINE Archive as published on Figshare[25]. However, we expect the archive to evolve further. The latest version of the archive is available from GitHub at github.com/SemsProject/CombineArchiveShowCase. It can also be downloaded from our website at scripts.sems.uni-rostock.de/getshowcase.php. Extensions, refinements, and comments are very welcome. Please fork the project on GitHub and contribute pull requests.
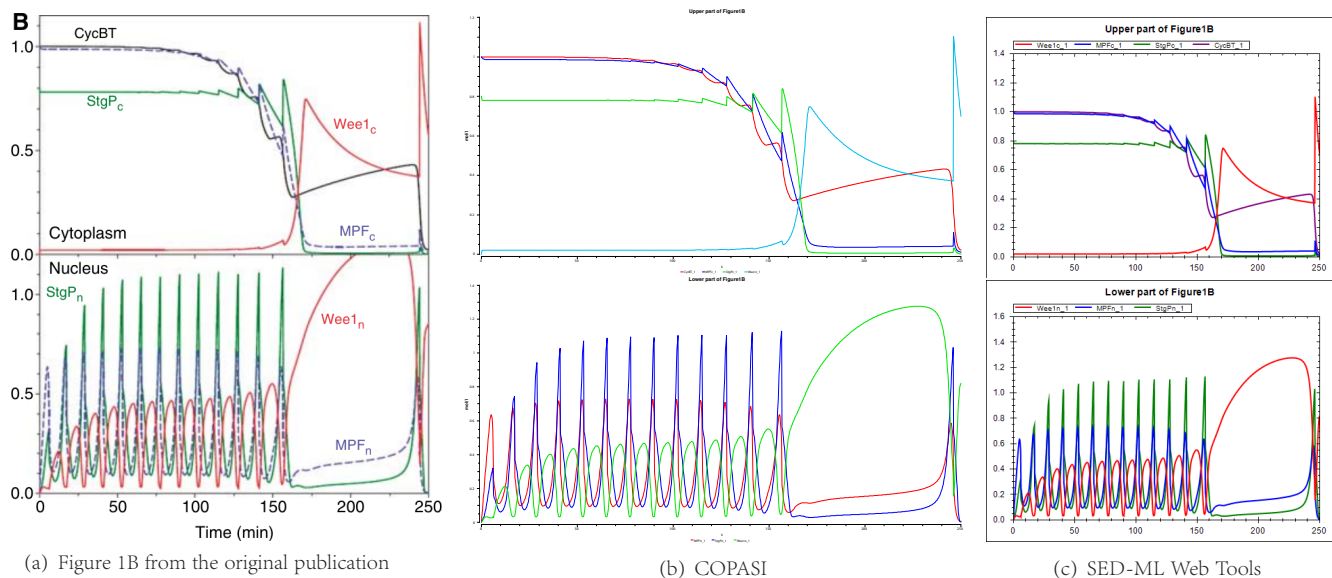

(a) Figure 1B from the original publication

(b) COPASI

(c) SED-ML Web Tools

Figure 2. Comparison of simulation results. The figure shows the simulation results included in the original publication (**2(a)**). Furthermore, the results generated by COPASI (**2(b)**) and the SWT (**2(c)**), using the SED-ML script Calzone2007-simulation-figure-1B.xml, are shown.

Data availability

The latest version of the COMBINE archive: github.com/SemsProject/CombineArchiveShowCase/ (latest commit at the time of submission: a469197)

The fully featured COMBINE archive as at the time of publication: *Figshare*: COMBINE Archive Show Case, 10.6084/m9.figshare.3427271.v1[24]

Author contributions

MS generated the data files for the archive and designed the initial version. DW and MS wrote the manuscript.

Competing interests

No competing interests were disclosed.

Grant information

Acknowledgements

References

1. Waltemath D, Wolkenhauer O: **How modeling standards, software, and initiatives support reproducibility in systems biology and systems medicine.** *IEEE Trans Biomed Eng.* 2016; (99).
   PubMed Abstract | **Publisher Full Text**

2. Prinz F, Schlange T, Asadullah K: **Believe it or not: how much can we rely on published data on potential drug targets?** *Nat Rev Drug Discov.* 2011; **10**(9): 712.
   PubMed Abstract | **Publisher Full Text**

3. Ioannidis JP, Allison DB, Ball CA, *et al.*: **Repeatability of published microarray gene expression analyses.** *Nat Genet.* 2009; **41**(2): 149–155.
   PubMed Abstract | **Publisher Full Text**

4. Begley CG, Ellis LM: **Drug development: Raise standards for preclinical cancer research.** *Nature.* 2012; **483**(7391): 531–533.
   PubMed Abstract | **Publisher Full Text**

5. Sandve GK, Nekrutenko A, Taylor J, *et al.*: **Ten simple rules for reproducible computational research.** *PLoS Comput Biol.* 2013; **9**(10): e1003285.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

6. Corcho Ó, Garijo Verdejo D, Belhajjame K, *et al.*: **Workflow-centric research objects: First class citizens in scholarly discourse.** In *2nd Workshop on Semantic Publishing.* Informatica, 2012.
   Reference Source

7. Bechhofer S, De Roure D, Gamble M, *et al.*: **Research objects: Towards exchange and reuse of digital knowledge.** *The Future of the Web for Collaborative Science.* 2010.
   Publisher Full Text

8. Bergmann FT, Adams R, Moodie S, *et al.*: **COMBINE archive and OMEX format: one file to share all information to reproduce a modeling project.** *BMC Bioinformatics.* 2014; **15**(1): 369.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

9. Scharm M, Wendland F, Peters M, *et al.*: **The CombineArchiveWeb application – A web-based tool to handle files associated with modelling results.** In *Proceedings of the 7th International Workshop on Semantic Web Applications and Tools for Life Sciences, Berlin, Germany, December 9–11, 2014.* 2014.
   Publisher Full Text

10. Hucka M, Finney A, Sauro HM, *et al.*: **The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models.** *Bioinformatics.* 2003; **19**(4): 524–531.
    PubMed Abstract | **Publisher Full Text**

11. Cuellar AA, Lloyd CM, Nielsen PF, *et al.*: **An overview of CellML 1.1, a biological model description language.** *SIMULATION.* 2003; **79**(12): 740–747.
    Publisher Full Text

12. Le Novère N, Hucka M, Mi H, *et al.*: **The Systems Biology Graphical Notation.** *Nat Biotechnol.* 2009; **27**(8): 735–741.
    PubMed Abstract | **Publisher Full Text**

13. Waltemath D, Adams R, Bergmann FT, *et al.*: **Reproducible computational biology experiments with SED-ML--the Simulation Experiment Description Markup Language.** *BMC Syst Biol.* 2011; **5**(1): 198.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

14. Dada JO, Spasić I, Paton NW, *et al.*: **SBRML: a markup language for associating systems biology data with models.** *Bioinformatics.* 2010; **26**(7): 932–938.
    PubMed Abstract | **Publisher Full Text**

15. Calzone L, Thieffry D, Tyson JJ, *et al.*: **Dynamical modeling of syncytial mitotic cycles in** *Drosophila* **embryos.** *Mol Syst Biol.* 2007; **3**: 131.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

16. Yu T, Lloyd CM, Nickerson DP, *et al.*: **The Physiome Model Repository 2.** *Bioinformatics.* 2011; **27**(5): 743–44.
    PubMed Abstract | **Publisher Full Text**

17. Li C, Donizelli M, Rodriguez N, *et al.*: **BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models.** *BMC Syst Biol.* 2010; **4**: 92.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

18. Scharm M, Waltemath D: **Extracting reproducible simulation studies from model repositories using the CombineArchive Toolkit.** In *Datenbanksysteme für Business, Technologie und Web (BTW 2015) – Workshopband.* Gesellschaft für Informatik, 2015; **P-242**: 137–142.
    Publisher Full Text

19. Hoops S, Sahle S, Gauges R, *et al.*: **COPASI--a COmplex PAthway SImulator.** *Bioinformatics.* 2006; **22**(24): 3067–3074.
    PubMed Abstract | **Publisher Full Text**

20. Czauderna T, Klukas C, Schreiber F: **Editing, validating and translating of SBGN maps.** *Bioinformatics.* 2010; **26**(18): 2340–2341.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

21. Rohn H, Junker A, Hartmann A, *et al.*: **VANTED v2: a framework for systems biology applications.** *BMC Syst Biol.* 2012; **6**(1): 139.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

22. Brandes U, Eiglsperger M, Herman I, *et al.*: **GraphML progress report structural layer proposal.** In *Graph Drawing.* Springer; 2002; **2265**: 501–512.
    Publisher Full Text

23. van Iersel MP, Villéger AC, Czauderna T, *et al.*: **Software support for SBGN maps: SBGN-ML and LibSBGN.** *Bioinformatics.* 2012; **28**(15): 2016–2021.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

24. Sauro HM, Choi K, Medley JK, *et al.*: **Tellurium: A Python Based Modeling and Reproducibility Platform for Systems Biology.** *bioRxiv.* 2016; 054601.
    Publisher Full Text

25. Scharm M, Touré V: **COMBINE Archive Show Case.** *Figshare.* 2016.
    Data Source

Open Peer Review

Current Peer Review Status: ❓ ✔ ❓

Version 1

Reviewer Report 11 October 2016

❓ **Alan Garny**
University of Auckland, Auckland, New Zealand

A major issue in computational biology is that of reproducibility. Here, the authors provide us with a concrete example of a reproducible simulation experiment using the Calzone *et al.* model (2007) of syncytial mitotic cycles in *Drosophila* embryos. To achieve their aim, they use established mark-up languages such as CellML, SBML and SED-ML, and package everything in a COMBINE archive. This COMBINE archive can then be used by anyone to reproduce the simulation experiment, as well as have access to the original paper, SBGN diagram, etc. This data note therefore shows us one possible, and viable, way to address the issue of reproducibility in computational biology and should, as such, be considered for indexing.

Major comments:
- How do you envisage tackling the issue of provenance? Say that someone modifies your archive and makes it available to the community, how are we then supposed to know which one to use and what the differences are between the two versions? Considering that your COMBINE archive is on GitHub, I guess someone could always fork it, but maybe a better approach would be to take advantage of existing repositories in the community, such as BioModels.net and the Physiome Model Repository (PMR)? I am not sure about BioModels.net, but PMR inherently addresses the issue of provenance.

- In their supplementary material, Calzone *et al.* provide some files that can be used to reproduce their different figures. So, in effect, they allow for their results to be reproduced (using XPPAUT). It would therefore make sense to have a comparison of their 'approach' to reproducibility compared with yours.

- The two SED-ML files in your COMBINE archive refer to the SBML version of the Calzone model. Now, because you are also providing a CellML version of that model, it would be nice to also have a CellML-based version of your two SED-ML files.

Minor comments:
- **Abstract:**

- You might want to replace "previously published simulation study" with "model of syncytial mitotic cycles in *Drosophila* embryos".

- **Keywords:**
  - Apparently, authors should supply up to eight keywords.

- **Introduction:**
  - **1st paragraph:** COmputational Modeling in BIology NEtwork (i.e. "O" instead of "o" and "E" instead of "e").

  - **2nd paragraph:** "the Systems Biology…" and not "The Systems Biology…"?

  - **2nd paragraph:** wouldn't it be better to reference NuML rather than SBRML? (On http://co.mbine.org/Standards, SBRML is only mentioned as part of the NuML description.)

  - **4th paragraph:** "Here, we describe…" (i.e. missing comma).

- **Materials and methods:**
  - **Retrieving an initial COMBINE archive:**
    - It would be nice to know exactly what kind of search was done using M2CAT (probably one using the term "calzone"?) and, then, which COMBINE archive was used as an initial COMBINE archive.

  - **Organising the COMBINE archive:**
    - "The initial archive was therefor[e] imported…" (therefore).
    - It would be nice to know what those 4 initial files were (those in the *model* folder).

  - **Extending the COMBINE archive:**
    - **The article:** "…to state when and [from] where…" (from).
    - **The simulation description:**
      - What is the point of having that default simulation experiment? To use it as a starting point for reproducing Figure 1B of the Calzone paper is fine, but I don't see the point of including that default simulation experiment in the COMBINE archive.
      - It would be nice to know what simulation parameters and solver (incl. its parameters) were used to reproduce Figure 1B of the Calzone paper. (I imagine they are the same as the ones used by the authors with XPPAUT?)
      - You might want to use sysbioapps.dyndns.org/SED-ML_Web_Tools as the URL for SED-ML Web Tools (rather than bqfbergmann.dyndns.org/SED-ML_Web_Tools)?
      - It would be nice to have all the other figures of the Calzone paper also encoded in SED-ML, just so that your COMBINE archive is not only fully featured, but also complete.
    - The visualisation of a model:
      - Are you sure about the version of VANTED you used? Version 2.1.0 is somewhat old compared with the latest version available (version 2.6.3). Actually, looking at the contents of your COMBINE archive, I can see that Calzone2007.gml was generated using VANTED 2.6.2. (FWIW, SBGN-ED 1.6 has just been released.)

- The link for GML is broken (
  http://www.fim.uni-passau.de/index.php?id=17297&L=1).

- **Data description:**
  - "…in four directories, cmp. Section…"?

  - Rather than referring to scripts.sems.uni-rostock.de/getshowcase.php (in this section and elsewhere in the manuscript), you might want to refer to your GitHub repository (github.com/SemsProject/CombineArchiveShowCase) and make use of GitHub's release feature? It might be safer in the long term.

- **Data validation:**
  - "…can be executed in COPASI, cmp. Figure 2(b)."?

  - To click on the bqfbergmann.dyndns.org/SED-ML_Web_Tools/Home/SimulateUrl?url=http://scripts.sems.u link takes me to a page that reads "No model uploaded. You need to upload a model first prior to attempting to simulate it!"

  - "… shown in the web browser, cmp. Figure 2(c)."?

  - You might want to provide a URL for COPASI and Tellurium?

- **Data availability:**
  - You might want to reference your very latest commit (i.e. 6c34cc4) rather than commit a469197?

- **Figure 1:**
  - I am not sure how useful this figure is. To me, it doesn't bring anything to your data note. Not only that, but at its original size, one cannot read anything (I personally had to view it at 400% to be able to start reading the different labels).

- **Figure 2:**
  - You might want to remove the "B" in panel (a). It's confusing.

  - You might want to mention that the range of the X and Y axes, as well as the colour of the different plots, cannot currently be specified in SED-ML, hence panels (b) and (c) don't perfectly match panel (a)?

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 10 October 2016

https://doi.org/10.5256/f1000research.10101.r16691

Laurence Calzone
Institut Curie, Paris, France
Mihaly Koltai
Computational Systems Biology of Cancer group, Institut Curie, Paris, France

The article describes a COMBINE archive of a previously published ODE model (Calzone et al) of the oscillatory dynamics of cell cycle regulatory proteins in Drosophila embryos undergoing rapid syncytial nuclear division cycles.

The files are stored on GitHub in 4 subfolders, containing the original article, model files, the figures generated from the simulations, as well as metafiles. The XML and CellML files that were originally published along with the article were modified, by adding metadata on the authors.

Issues:
- There is a clear advantage to such an archive for publication of such models, reproducibility etc., but the authors do not insist on it. This type of archive could be added to publications, and provided by the authors, with a full documentation. This more general aim of the article should be more forcefully and clearly stated.

- We tried to open the model file available on github (XML) using COPASI (4.16) but some errors appeared. We downloaded the initial model from Biomodels and it worked. Same issues with CellML files.

- The link pointing to the website "bqfbergmann.dyndns.org/SED-ML_Web_Tools/" is broken, but if copy-pasted the figures are correctly displayed.

Competing Interests: No competing interests were disclosed.

We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 03 October 2016

https://doi.org/10.5256/f1000research.10101.r16695

Lars Juhl Jensen
NNF Center for Protein Research, University of Copenhagen, Copenhagen, Denmark

General comments:

The data note describes how COMBINE is used to make available a mathematical model for simulation of syncytial mitotic cycles in Drosophila. Although well written, the aim of the article is somewhat unclear to me.

The way the introduction is written and the archive is named suggest that the purpose of making the archive is to serve as a showcase for how to use COMBINE archives. However, the way the rest of the article is written focuses very much on how this specific archive was made and how it is organized. If the goal is to showcase COMBINE archives, I feel the manuscript lacks a description of not only how the specific archive was made but also why this would be a good way to make such archives in general. If, on the other hand, the aim is to make the specific model available, it lacks some more background on the specific model and possibly suggestions or examples of how it can be used.

Major comments:

I am confused about what is to be considered the primary repository / access point. The article refers to both a GitHub repository, an omex file on figshare, and a link to a php script on their own server where one can retrieve the latest version of the omex file. This leads to a number of questions:
1. Am I right that GitHub is the primary place where development is done and improvements will made?
2. Is the omex file provided by the php script always up-to-date with the GitHub repository, or may the newest version on GitHub be even newer than the latest version made available as an omex file?
3. Will the omex file on figshare be updated, or will there only be a v1 there?

Considering that figshare has both versioning and API for submission, I think the best solution would be to abandon the php script. This would eliminate redundancy, as the omex file would only be on figshare. Since figshare has both versioned and version-less DOIs for datafiles, this would allow the authors to provide a stable DOI that always points to the latest version of the omex file and at the same time allow users to always cite the specific versioned DOI of the omex file they used in their work. Given that figshare has a submission API, it should be possible to make sure that the omex file is automatically updated based on the GitHub repository whenever needed.

Technical issues:

Accessing the URL "bqfbergmann.dyndns.org/SED-ML_Web_Tools/Home/SimulateUrl?url=http://scripts.sems.uni-rostock.de by copy and paste works correctly. However, clicking the link within the article instead opens "http://bqfbergmann.dyndns.org/SED-ML_Web_Tools/Home/Simulate" and leads to an error page instead of the promised results.

The link "www.ebi.ac.uk/biomodels-main/download?mid=BIOMD0000000144" is similarly broken. When clicked, it instead tries to open "https://f1000research.com/articles/5-2421/www.ebi.ac.uk/biomodels-main/download?mid=BIOMD000000 which leads to an error.

The URL "www.fim.uni-passau.de/index.php?id=17297&L=1" leads me to an error page irrespective of whether I copy it or click it.

Minor corrections:

The authors several times use the abbreviation "cmp." to refer to other parts/figures in the article. I have never encountered this abbreviation before and was unable to find any other articles that use it. Given the context, I suspect "cf." may be what is meant.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Comments on this article

Version 1

Reader Comment 27 Sep 2017

Matthias König, Humboldt-University Berlin, Germany

The following link is not working anymore
bqfbergmann.dyndns.org/SED-ML_Web_Tools/Home/SimulateUrl?url=http://scripts.sems.uni-rostock.de/gets
.

Best
Matthias

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias

- You can publish traditional articles, null/negative results, case reports, data notes and more

- The peer review process is transparent and collaborative

- Your article is indexed in PubMed after passing peer review

- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research