The CombineArchiveWeb application – A web based tool to handle files associated with modelling results

Martin Scharm*, Florian Wendland, Martin Peters, Markus Wolfien, Tom Theile, and Dagmar Waltemath

Department of Systems Biology and Bioinformatics, University of Rostock, Germany

Abstract. Sharing *in silico* experiments is essential for the advance of research in computational biology. Consequently, the COMBINE archive was designed as a digital container format. It eases the management of files related to a modelling result, fosters collaboration, and ultimately enables the exchange of reproducible simulation studies. However, manual handling of COMBINE archives is tedious and error prone. We therefore developed the CombineArchiveWeb application to support scientists in promoting and publishing their research by means of creating, exploring, modifying, and sharing archives. All files are equipped with meta data and can be distributed over the Web through shareable workspaces.

Introduction

Computational modelling is an indispensable tool in the life sciences. As a consequence, standardisation and exchange of models has also become essential. However, the steadily increasing size and complexity of models and derived data poses the challenge of sharing *reproducible* results. Today's results typically consist of multiple model files with semantic annotations, simulation descriptions linked to terms in a simulation algorithm ontology, structure information, result data sets with semantic annotations, and reference publications [1]. Only if all relevant files are provided, modelling results can be reliably reproduced and sharing becomes feasible.

To solve this issue, the COMBINE community [2] proposes the COMBINE archive format [3]. A COMBINE archive is a container that bundles all files related to a project into a single file. Typically, the COMBINE archive also comprises files with meta data such as people attributions and details about files inside the archive. The meta data is encoded in XML/RDF. For example, VCards refer to contributors and DCTERMS capture creation and modification dates. The meta data can be used to define relations among the files contained in the archive. It can also be evaluated to query and compare COMBINE archives, or to quickly get an overview of the content without necessarily loading the single files. An example is given in Figure 1. The files included in an archive are listed

---

* To whom correspondence should be addressed

and their meta data are displayed in a human readable format. In this example, the archive contains information about the results from Liebal *et. al* [4]: a model description in SBML format, a simulation description in SED-ML format, the reference publication in PDF format and two data tables. Viewing and modifying meta data is easily possible through the forms on the right hand side.

In summary, the COMBINE archive is a solid format for sharing reproducible models and *in silico* experiments with collaborators and public databases.

However, manual handling of COMBINE archives is tedious and error prone. Therefore, we developed a CombineArchive Toolkit (`http://sems.uni-rostock.de/cat`). It consists of a library, a desktop application, and a web based interface.

The CombineArchiveLibrary[1] is written in Java and implements the COMBINE archive specification [5]. It offers all necessary methods to handle COMBINE archives, including:

- Extracting single files or the whole archive
- Browsing through the archive
- Adding and removing files
- Renaming and reorganising files
- Attaching and retrieving meta information

The CombineArchiveLibrary is for tool developers. It has already been integrated with software such as the Functional Curation Project of Chaste [6].

We also integrated the CombineArchiveLibrary in the CombineArchiveDesktop application[2]. It is intended as a functional browser for COMBINE archives. One of the application's key features is the user-friendly handling of included meta data. The meta data is presented in a human readable way and can easily be modified. Since the desktop application only runs locally, your files do not leave your machine. This is especially important when data protection or intellectual property claims are of concern. The CombineArchiveDesktop application is implemented in Java and shipped as an executable Java Archive (JAR) file. Thus, it is portable and runs on Windows, Linux, and MacOS.

The second software tool is the CombineArchiveWeb application. It is described in the following.

The CombineArchiveWeb application

The CombineArchiveWeb application[3] also uses the CombineArchiveLibrary as a code base. It enables any researcher to work with the COMBINE archive format on the internet. Additionally, the application offers RESTful services for use by other client applications. The web interface connects to open model repositories, such as the CellML model repository [8], to easily retrieve models. User with special interest in privacy do not need to alienate their research. Since the

---

[1] sems.uni-rostock.de/trac/combinearchive

[2] sems.uni-rostock.de/trac/combinearchive-gui

[3] webcat.sems.uni-rostock.de

CombineArchiveWeb application is openly available and easy to install, anyone could host their own instances on private servers. In order to prohibit abuse, the maintainer of an installation can easily configure quotas for the users. For example, it is possible to limit the maximum file size, the maximum age of an untouched archive, the number of archives per workspace, the number of files in an archive, just to name a few options. Since the CombineArchiveWeb application is not intended as a database for long-term storage of simulation studies old archives are deleted regularly.

Fig. 1. CombineArchiveWeb application. The web interface provides means to: **create** archives, e.g. from a modelling project to be submitted to a journal or an open repository; **explore** other people's work, e.g. by downloading archives from the CellML model repository [8] and studying included files; **modify** an existing archive, e.g. by improving, extending, and correcting its content; and **share** archives with project partners.

Researchers who wish to share their results through the CombineArchiveWeb application can simply upload all relevant files belonging to a piece of modelling work. This automatically creates a workspace. Workspaces are an ideal tool to manage files. They foster collaborations and prevent inconsistencies in versions of files. Uploaded files are immediately bundled in a COMBINE archive. Such virtual experiments can be downloaded at any time and from any location. The created workspace can also be shared with collaborators to work from different physical locations. The CombineArchiveWeb application supports researchers in exploring and reproducing scientific results. For example, authors of a publication may decide to provide their code as a COMBINE archive. When consumers open the archive, the CombineArchiveWeb application automatically reads the files and their meta data. It builds the links between the various files and presents the contents in a human readable format (see again Figure 1).

Summary

The CombineArchive Toolkit implements the latest COMBINE archive specification and, thus, breathes life into the idea of sharing all files necessary to reproduce an *in silico* experiment. We present a web based, graphical user interface for interacting with COMBINE archives. The CombineArchiveWeb application grants users intuitive access to COMBINE archives. Our tools and code are openly available through our project website at sems.uni-rostock.de/cat. We invite you to use our software, to provide feedback and suggestions, and to contribute to the further development of the CombineArchive Toolkit.

References

1. Waltemath *et al.*: Possibilities for Integrating Model-related Data in Computational Biology. In: *Proceedings of the 9th International Conference on Data Integration in the Life Sciences*. CEUR Workshops, 2013.
2. Le Novère *et al.*: Meeting report from the first meetings of the Computational Modeling in Biology Network (COMBINE). *Standards in genomic sciences*, 2011.
3. Bergmann *et al.*: COMBINE archive: One File To Share Them All. *arXiv*, 2014.
4. Liebal *et al.*: Proteolysis of beta-galactosidase following SigmaB activation in *Bacillus subtilis*. *Molecular BioSystems*, 2012.
5. Bergmann *et al.*: COMBINE Archive Specification, 2014.
6. Cooper *et al.*: High-throughput functional curation of cellular electrophysiology models. *Progress in Biophysics and Molecular Biology*, 2011.
7. Li *et al.*: BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models. *BMC Systems Biology*, 2010.
8. Lloyd *et al.*: The CellML Model Repository. *Bioinformatics*, 2008.