

The CombineArchive Toolkit – facilitating the transfer of research 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 numerous files, fosters collaboration, and ultimately enables the exchange of reproducible research results. However, manual handling of COMBINE archives is tedious and error prone. We therefore developed the CombineArchive Toolkit. It supports scientists in promoting and publishing their work by means of creating, exploring, modifying, and sharing archives.

Introduction

Computational modelling has become an indispensable tool in the life sciences. Consequently, standardising and exchanging developed models has become essential. However, the steadily increasing size and complexity of models and derived data poses the challenge of sharing *reproducible* results. Today, these results typically consist of multiple model files, simulation descriptions, publications, and meta data. Only if all relevant files are provided, modelling results can be reliably reproduced and sharing becomes feasible.

To overcome this issue, the COMBINE community [1] proposes the COMBINE archive format [2]. A COMBINE archive is a container that bundles all files related to a project into a single file. Typically, it also comprises files with meta information, such as people attributions and details about files inside the archive. The COMBINE archive is a solid method 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 the CombineArchive Toolkit. It consists of a core library, a desktop application, and a web based interface. Taken together, our tools support researchers in creating and exploring COMBINE archives.

* To whom correspondence should be addressed

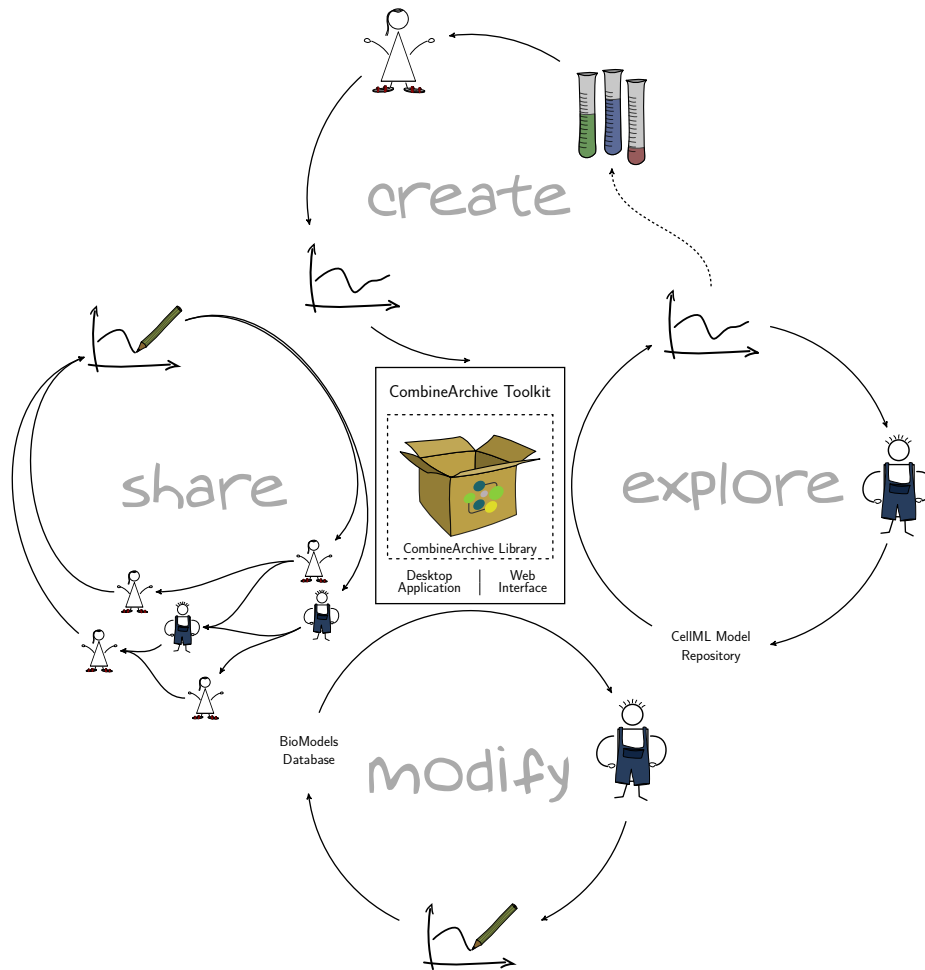


Fig. 1. Main workflows facilitated by the CombineArchive Toolkit. Our toolkit consists of a library, a desktop application, and a web interface. It provides means to: **create** archives, e.g. from a modelling project to be submitted to a journal or an open repository; **explore** other people's work, e.g. by downloading archives from BioModels Database[4] and studying included files on one's own computer; **modify** an existing archive, e.g. by improving, extending, and correcting its content; and **share** archives with project partners. Our tools are written in Java and openly available from our website sems.uni-rostock.de.

Results

Here we introduce the CombineArchive Toolkit. It consists of three software tools for developers and modellers who wish to create, explore, modify, and share COMBINE archives, cf. Figure 1. The core of our toolkit is the CombineArchive library, an implementation of the COMBINE archive specification [3]. This library can be used by other software projects to support COMBINE archives. In addition, we developed two graphical user interfaces: First, a desktop application designed for offline usage; second, a web interface adding the functionality of sharing archives online.

The CombineArchive library

The CombineArchive library¹ was implemented using latest Java technologies. 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 CombineArchive library parses and translates the meta information into an internal representation. Currently, we just support the meta data format as proposed in the COMBINE archive specification. However, the parser can be extended to also understand and interpret other RDF/XML. Given this layer of abstraction, other tools can easily implement support for COMBINE archives. Example code on how to use our library is available from our website.

The CombineArchive desktop application

The CombineArchive desktop application² is a functional browser for COMBINE archives. It complies with latest design guidelines and, thus, is intuitive to use. The desktop application supports the workflows as illustrated in Figure 1: You can easily create, modify, and export 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.

Moreover, with our desktop application you retain full control over your data. 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 CombineArchive desktop 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.

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

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

The CombineArchive web interface

The CombineArchive web interface³ also uses the CombineArchive library as a common code base. It is a centralized cloud data management system that provides basic support for collaborative work. As such, it allows you to share your workspaces with partners and to work on archives from different physical locations. Additionally, the CombineArchive web interface offers RESTful services which can also be used from other client applications. The web interface connects to open model repositories, such as BioModels Database or the CellML model repository [5], to submit and retrieve models.

Since the CombineArchive web interface is openly available and easy to install, users with special interest in privacy may host their own instance and do not need to alienate their research.

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 library and two graphical user interfaces for interacting with COMBINE archives. The CombineArchive library was build for developers and has already been integrated with other software tools, such as the Functional Curation Project of Chaste [6]. The CombineArchive desktop application and the CombineArchive web interface grant 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 CombineArchive Toolkit.

References

1. Le Novère *et al.*: Meeting report from the first meetings of the Computational Modeling in Biology Network (COMBINE). *Standards in genomic sciences*, 2011.
2. Bergmann *et al.*: COMBINE archive: One File To Share Them All. *arXiv*, 2014.
3. Bergmann *et al.*: COMBINE Archive Specification, 2014.
4. Li *et al.*: BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models. *BMC Systems Biology*, 2010.
5. Lloyd *et al.*: The CellML Model Repository. *Bioinformatics*, 2008.
6. Cooper *et al.*: High-throughput functional curation of cellular electrophysiology models. *Progress in Biophysics and Molecular Biology*, 2011.

³ sems.uni-rostock.de/trac/combinearchive-web