The Journal of Open Source Software

Bacting: a next generation, command line version of Bioclipse

Egon Willighagen[1]

1 Dept of Bioinformatics - BiGCaT, NUTRIM, Maastricht University

DOI: 10.21105/joss.02558

Software

- Review
- Repository
- Archive

Editor: Mark A. Jensen

Reviewers:

- @Zethson
- @arcuri82

Submitted: 22 July 2020
Published: 23 June 2021

Summary

Bioclipse (https://github.com/bioclipse) was originally developed as an interactive user interface (UI) based on Eclipse for research in the fields of biology and chemistry (Spjuth et al., 2007). It was later extended with scripting functionality and scripts could be written in JavaScript, Python, and Groovy (Spjuth et al., 2009). An innovative aspect of the second Bioclipse version was that Bioclipse plugins could inject domain specific functionality into the scripting language. This was done using OSGi (https://www.osgi.org/) and the Spring Framework (https://spring.io/), making so-called *managers* accessible in scripts. However, there have not been any recent Bioclipse releases. Bacting is a next generation, command line version of Bioclipse, that is more easily updated, built, released, and used. A subset of the original functionality is available, and some managers have already been updated to use more recent versions of dependencies.

Statement of Need

While Bioclipse has served our research for many years, a number of limitations has made this increasingly hard. For example, the dependency of Bioclipse on the Eclipse UI requires the scripts to be run inside a running Bioclipse application. This makes repeatedly running of a script needlessly hard and use in continuous integration systems or use on computing platforms impossible. A second problem was that the build and release system of Bioclipse was complex, making it hard for others to repeat creating new releases. This is reflected in the lack of recent releases and complicates the process for external developers wishing to make patches.

These needs triggered a next generation design of Bioclipse: 1. the managers providing the domain-specific functionality would need to be usable on the command line; 2. building the Bioclipse managers should be possible on the command line, ideally with continuous build systems; 3. Bacting should be easy to install and reuse.

Implementation

To keep the option open to backport new functionality to Bioclipse, the API is copied as precisely as possible. However, there are some differences. For example, there is only a single manager class, and no longer interfaces for both the scripting language or for running Bioclipse user interface. This means that the *IFile* to *String* translations in the API do not exist in Bioclipse. Furthermore, there are currently no progress monitors. That said, the source code implementing the method is otherwise identical and easily translated back to the original Bioclipse source code.

This is done by separating the Bioclipse code from the Bacting manager implementations. The latter is mainly described in this paper, and the former is found as much more stable code in the GitHub https://github.com/egonw/bacting-bioclipse repository. The code is identical to the original Bioclipse code, but mavenized in this repository, allowing deployment on Maven Central.

The Bacting manager are found in the https://github.com/egonw/bacting repository and while the managers in this repository share most of the code with the original Bioclipse implementations, they are still considered new implementations and therefore are tested using JUnit. A second important difference is that Bioclipse documentation was found on the manager interfaces, but in Bacting the JavaDoc is found is found in the implementations of the managers. A final difference is how the managers are used: because they are not injected into the scripting language, each manager needs to be created manually, which requires one extra line of code for each manager.

Continuous integration and releases

Bacting is hosted on GitHub and takes advantage of the integrations with Zenodo for automatic archiving of releases (see https://github.com/egonw/bacting/releases) and with GitHub Actions for continuous integration (see https://github.com/egonw/bacting/actions). Maven is used as a build system and automatically downloads the dependencies when compiling the source code. GitHub Actions compiles the source code regularly with Java 8, 11, and 14. During the process the JUnit 5 unit tests are run and the compilation aborted when there are testing failures. The extend to which the tests execute code in the managers is tested with JaCoCo (https://www.jacoco.org/jacoco/) and reported online with Codecov at (https://codecov.io/gh/egonw/bacting).

Releases are made at irregular intervals, but often triggered by downstream uses that need additional Bioclipse functionality to be ported. Releases are created with the `mvn releas e:prepare` and `mvn release:perform` process that tags the commit, updates the version numbers, and uploads the release to Maven Central. Second, a changelog is written for the GitHub releases page, which triggers the archiving on Zenodo (see https://doi.org/10.5281/zenodo.2638709). Finally, at that moment the JavaDoc is also generated and uploaded to another GitHub repository (see https://github.com/egonw/bacting-api) making it available online with GitHub pages at https://egonw.github.io/bacting-api/.

Updated dependencies of managers

The *cdk* manager wrapping Chemistry Development Kit functionality was updated to version 2.3, released in 2017 (Mayfield et al., 2019; E. L. Willighagen et al., 2017). The *opsin* manager was updated to use OPSIN version 2.5.0, released in 2020 (Lowe et al., 2011). The *bridgedb* manager was updated to BridgeDb version 2.3.10, released in 2020 (Brenninkmeijer et al., 2020; Iersel et al., 2010).

Ported Functionality

Bioclipse has a long list of managers and so far only a subset has been ported, which is briefly described in this table:

Bacting Manager	Functionality
bioclipse	Bioclipse manager with common functionality
ui	Bioclipse manager with user interface functionality

Bacting Manager	Functionality
report	Manager that provides an API to create HTML reports
cdk	Chemistry Development Kit for cheminformatics functionality (E. L. Willighagen et al., 2017)
inchi	Methods for generating and validating InChIs and InChIKeys (Spjuth et al., 2013)
pubchem	Methods to interact with the PubChem databases
chemspider	Methods to interact with the Chemspider databases
rdf	Resource Description Framework (RDF) functionality, using Apache Jena
opsin	Access to the OPSIN library for parsing IUPAC names (Lowe et al., 2011)
bridgedb	Access to the BridgeDb library for identifier mapping (Iersel et al., 2010)
biojava	Access to the Biojava library for sequence functionality (Holland et al., 2008)

The functionality of the Bioclipse managers is partly documented in the A lot of Bioclipse Scripting Language examples booklet, of which several scripts are available as Bacting examples. For example, the FullPathWikiPathways.groovy page from this booklet shows both the Bioclipse version of the script as well as the Bacting version.

Grabbing Bacting from Groovy

Use of Bacting in the Groovy language takes advantage of the fact that it is available from Maven Central, allowing `@Grab` to be used to dynamically download the code as in this example for the *cdk* manager:

```
@Grab(
  group='io.github.egonw.bacting',
  module='managers-cdk', version='0.0.15'
)

def cdk = new net.bioclipse.managers.CDKManager(".");

println cdk.fromSMILES("COC")
```

Similarly, Bacting can be used in Python using scyjava.

Use cases

Bioclipse scripts have been in use in our group in various research lines to automate repetitive work. Various scripts have now been ported to Bacting and several are now available as open notebook science repositories at https://github.com/egonw/ons-wikidata, https://github.com/egonw/ons-chebi, and https://github.com/egonw/ons-wikipathways. The scripts in these repositories are used to populate Wikidata with chemical structures to support the Scholia project (Nielsen et al., 2017; E. Willighagen et al., 2018), the WikiPathways project (Slenter et al., 2018), and feed additional metabolite identifiers into Wikidata for creation of BridgeDb identifier mapping databases in an implementation study of the ELIXIR Metabolomics Community (Rijswijk et al., 2017; E. Willighagen, 2020). Furthermore, Bacting is used to populate Wikidata with OECD Testing Guidelines in the NanoCommons project and extend the eNanoMapper ontology (see https://github.com/egonw/ons-wikidata/blob/master/OECD/convertToOWL.groovy) (Hastings et al., 2015), to generate RDF for a public data set in

the NanoSolveIT project (see https://github.com/NanoSolveIT/10.1021-acsnano.8b07562) (Afantitis et al., 2020), to create a booklet with data about the SARS-CoV-2 and related coronavirusses (see https://github.com/egonw/SARS-CoV-2-Queries), and to support Various of these use cases are ongoing and are not yet published, which is planned.

Acknowledgements

We acknowledge the contributions of the Bioclipse developers which have been ported here into the Bacting software.

References

Afantitis, A., Melagraki, G., Isigonis, P., Tsoumanis, A., Danai Varsou, D., Valsami-Jones, E., Papadiamantis, A., Ellis, L.-Jayne. A., Sarimveis, H., Doganis, P., Karatzas, P., Tsiros, P., Liampa, I., Lobaskin, V., Greco, D., Serra, A., Anneli Sofia Kinaret, P., Aliisa Saarimäki, L., Grafström, R., … Lynch, I. (2020). NanoSolveIT Project: Driving Nanoinformatics research to develop innovative and integrated tools for in silico nanosafety assessment. *Computational and Structural Biotechnology Journal*, S2001037019305112. https://doi.org/10.1016/j.csbj.2020.02.023

Brenninkmeijer, C. Y., Willighagen, E., Awasthi, M., Soiland-Reyes, S., Riutta, A., Pico, A., Dunlop, I., Nunogit, Kerber, R., De, Server, M. J. B., Summer-Kutmon, M., Mélius, J., Martens, M., Gray, A., & Ehrhart, F. (2020). *Bridgedb/BridgeDb: BridgeDb 2.3.8*. https://doi.org/10.5281/ZENODO.3855084

Hastings, J., Jeliazkova, N., Owen, G., Tsiliki, G., Munteanu, C. R., Steinbeck, C., & Willighagen, E. (2015). eNanoMapper: Harnessing ontologies to enable data integration for nanomaterial risk assessment. *Journal of Biomedical Semantics*, *6*(1). https://doi.org/10.1186/s13326-015-0005-5

Holland, R. C. G., Down, T. A., Pocock, M., Prlić, A., Huen, D., James, K., Foisy, S., Dräger, A., Yates, A., Heuer, M., & Schreiber, M. J. (2008). BioJava: An open-source framework for bioinformatics. *Bioinformatics*, *24*(18), 2096–2097. https://doi.org/10.1093/bioinformatics/btn397

Iersel, M. P. van, Pico, A. R., Kelder, T., Gao, J., Ho, I., Hanspers, K., Conklin, B. R., & Evelo, C. T. (2010). The BridgeDb framework: Standardized access to gene, protein and metabolite identifier mapping services. *BMC Bioinformatics*, *11*(1), 5+. https://doi.org/10.1186/1471-2105-11-5

Lowe, D. M., Corbett, P. T., Murray-Rust, P., & Glen, R. C. (2011). Chemical name to structure: OPSIN, an open source solution. *Journal of Chemical Information and Modeling*, *51*(3), 739–753. https://doi.org/10.1021/ci100384d

Mayfield, J., Willighagen, E., Guha, R., Torrance, G., Ujihara, K., Rahman, S. A., Alvarsson, J., Vine, M. B., Gražulis, S., Pluskal, T., Wei, Y. C., Szisz, D., Williamson, M. J., Kochev, N., Jeliazkova, N., Bach, E., Berg, A., Clark, A., Stephan, R., … Harmon, C. (2019). *Cdk/cdk: CDK 2.3*. https://doi.org/10.5281/ZENODO.3364510

Nielsen, F. Å., Mietchen, D., & Willighagen, E. (2017). Scholia, Scientometrics and Wikidata. In E. Blomqvist, K. Hose, H. Paulheim, A. Ławrynowicz, F. Ciravegna, & O. Hartig (Eds.), *The Semantic Web: ESWC 2017 Satellite Events* (Vol. 10577, pp. 237–259). Springer International Publishing. https://doi.org/10.1007/978-3-319-70407-4_36

Rijswijk, M. van, Beirnaert, C., Caron, C., Cascante, M., Dominguez, V., Dunn, W. B., Ebbels, T. M. D., Giacomoni, F., Gonzalez-Beltran, A., Hankemeier, T., Haug, K., Izquierdo-Garcia, J. L., Jimenez, R. C., Jourdan, F., Kale, N., Klapa, M. I., Kohlbacher, O., Koort, K., Kultima, K., … Steinbeck, C. (2017). The future of metabolomics in ELIXIR. *F1000Research*, *6*, 1649. https://doi.org/10.12688/f1000research.12342.1

Slenter, D. N., Kutmon, M., Hanspers, K., Riutta, A., Windsor, J., Nunes, N., Mélius, J., Cirillo, E., Coort, S. L., Digles, D., Ehrhart, F., Giesbertz, P., Kalafati, M., Martens, M., Miller, R., Nishida, K., Rieswijk, L., Waagmeester, A., Eijssen, L. M. T., … Willighagen, E. L. (2018). WikiPathways: A multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Research*, *46*(D1), D661–D667. https://doi.org/10.1093/nar/gkx1064

Spjuth, O., Alvarsson, J., Berg, A., Eklund, M., Kuhn, S., Mäsak, C., Torrance, G., Wagener, J., Willighagen, E. L., Steinbeck, C., & Wikberg, J. E. (2009). Bioclipse 2: A scriptable integration platform for the life sciences. *BMC Bioinformatics*, *10*(1), 397. https://doi.org/10.1186/1471-2105-10-397

Spjuth, O., Berg, A., Adams, S., & Willighagen, E. L. (2013). Applications of the InChI in cheminformatics with the CDK and Bioclipse. *Journal of Cheminformatics*, *5*(1), 14. https://doi.org/10.1186/1758-2946-5-14

Spjuth, O., Helmus, T., Willighagen, E. L., Kuhn, S., Eklund, M., Wagener, J., Murray-Rust, P., Steinbeck, C., & Wikberg, J. E. (2007). Bioclipse: An open source workbench for chemo- and bioinformatics. *BMC Bioinformatics*, *8*(1), 59+. https://doi.org/10.1186/1471-2105-8-59

Willighagen, E. (2020). *Metabolite BridgeDb ID Mapping Database (20200610)*. figshare. https://doi.org/10.6084/M9.FIGSHARE.12458612

Willighagen, E. L., Mayfield, J. W., Alvarsson, J., Berg, A., Carlsson, L., Jeliazkova, N., Kuhn, S., Pluskal, T., Rojas-Chertó, M., Spjuth, O., Torrance, G., Evelo, C. T., Guha, R., & Steinbeck, C. (2017). The Chemistry Development Kit (CDK) v2.0: Atom typing, depiction, molecular formulas, and substructure searching. *Journal of Cheminformatics*, *9*(1). https://doi.org/10.1186/s13321-017-0220-4

Willighagen, E., Slenter, D., Mietchen, D., Evelo, C., & Nielsen, F. (2018). Wikidata and Scholia as a hub linking chemical knowledge. *Figshare*. https://doi.org/10.6084/m9.figshare.6356027.v1