Automated workflow composition in mass spectrometry-based proteomics
Numerous software utilities operating on mass spectrometry (MS) data are described in the literature and provide specific operations as building blocks for the assembly of on-purpose workflows. Working out which tools and combinations are applicable or optimal in practice is often hard. Thus researchers face difficulties in selecting practical and effective data analysis pipelines for a specific experimental design. Results: We provide a toolkit to support researchers in identifying, comparing and benchmarking multiple workflows from individual bioinformatics tools. Automated workflow composition is enabled by the tools' semantic annotation in terms of the EDAM ontology. To demonstrate the practical use of our framework, we created and evaluated a number of logically and semantically equivalent workflows for four use cases representing frequent tasks in MS-based proteomics. Indeed we found that the results computed by the workflows could vary considerably, emphasizing the benefits of a framework that facilitates their systematic exploration.

General information
State: Published
Organisations: IT Service, Leiden University, Utrecht University, University of Southern Denmark
Contributors: Palmblad, M., Lamprecht, A. L., Ison, J., Schwämmle, V.
Pages: 656-664
Publication date: 15 Feb 2019
Peer-reviewed: Yes

Publication information
Journal: Bioinformatics
Volume: 35
Issue number: 4
ISSN (Print): 1367-4803
Ratings:
Web of Science (2019): Indexed yes
Web of Science (2018): Indexed yes
Scopus rating (2017): CiteScore 7.84
Web of Science (2017): Impact factor 5.481
Four simple recommendations to encourage best practices in research software

Scientific research relies on computer software, yet software is not always developed following practices that ensure its quality and sustainability. This manuscript does not aim to propose new software development best practices, but rather to provide simple recommendations that encourage the adoption of existing best practices. Software development best practices promote better quality software, and better quality software improves the reproducibility and reusability of research. These recommendations are designed around Open Source values, and provide practical suggestions that contribute to making research software and its source code more discoverable, reusable and transparent. This manuscript is aimed at developers, but also at organisations, projects, journals and funders that can increase the quality and sustainability of research software by encouraging the adoption of these recommendations.

General information

State: Published
Organisations: Department of Bio and Health Informatics, IT Service, High Performance Computing, ELIXIR Hub, Netherlands eScience Center, CSL Limited, National eResearch Collaboration Tools and Resources, University of Freiburg, Stockholm University, Spanish Institute of Bioinformatics, University of Edinburgh, Repositive Ltd, University of
Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data

In many disciplines, data are highly decentralized across thousands of online databases (repositories, registries, and knowledgebases). Wringing value from such databases depends on the discipline of data science and on the humble bricks and mortar that make integration possible; identifiers are a core component of this integration infrastructure. Drawing on our experience and on work by other groups, we outline 10 lessons we have learned about the identifier qualities and best practices that facilitate large-scale data integration. Specifically, we propose actions that identifier practitioners (database providers) should take in the design, provision and reuse of identifiers. We also outline the important considerations for those referencing identifiers in various circumstances, including by authors and data generators. While the importance and relevance of each lesson will vary by context, there is a need for increased awareness about how to avoid and manage common identifier problems, especially those related to persistence and web-accessibility/resolvability. We focus strongly on web-based identifiers in the life sciences; however, the principles are broadly relevant to other disciplines.
ReGaTE: Registration of Galaxy Tools in Elixir

Background: Bioinformaticians routinely use multiple software tools and data sources in their day-to-day work and have been guided in their choices by a number of cataloguing initiatives. The ELIXIR Tools and Data Services Registry (bio.tools) aims to provide a central information point, independent of any specific scientific scope within bioinformatics or technological implementation. Meanwhile, efforts to integrate bioinformatics software in workbench and workflow environments have accelerated to enable the design, automation, and reproducibility of bioinformatics experiments. One such popular environment is the Galaxy framework, with currently more than 80 publicly available Galaxy servers around the world. In the context of a generic registry for bioinformatics software, such as bio.tools, Galaxy instances constitute a major source of valuable content. Yet there has been, to date, no convenient mechanism to register such services en masse. Findings: We present ReGaTE (Registration of Galaxy Tools in Elixir), a software utility that automates the process of registering the services available in a Galaxy instance. This utility uses the BioBlend application program interface to extract service metadata from a Galaxy server, enhance the metadata with the scientific information required by bio.tools, and push it to the registry. Conclusions: ReGaTE provides a fast and convenient way to publish Galaxy services in bio.tools. By doing so, service providers may increase the visibility of their services while enriching the software discovery function that bio.tools provides for its users. The source code of ReGaTE is freely available on Github at https://github.com/C3BI-pasteur-fr/ReGaTE.

General information
State: Published
Organisations: Department of Bio and Health Informatics, High Performance Computing, Institut Pasteur Paris, University of Bergen, The Earlham Institute, Universite Pierre et Marie Curie, University of Freiburg
Contributors: Doppelt-Azeroual, O., Mareuil, F., Deveaud, E., Kalas, M., Soranzo, N., van den Beek, M., Gruening, B., Ison, J., Ménager, H.
Pages: 1-4
Publication date: 2017
Peer-reviewed: Yes

Publication information
Journal: GigaScience
Volume: 6
Issue number: 6
ISSN (Print): 2047-217X
Ratings:
Web of Science (2019): Indexed yes
Web of Science (2018): Indexed yes
Scopus rating (2017): CiteScore 6.81 SJR 5.022 SNIP 1.784
Web of Science (2017): Impact factor 7.267
Web of Science (2017): Indexed yes
Scopus rating (2016): CiteScore 4.87 SJR 5.068 SNIP 1.738
Web of Science (2016): Impact factor 6.871
Scopus rating (2015): CiteScore 8.64 SJR 4.727 SNIP 1.679
Web of Science (2015): Impact factor 7.463
Using bio.tools to generate and annotate workbench tool descriptions

Workbench and workflow systems such as Galaxy, Taverna, Chipster, or Common Workflow Language (CWL)-based frameworks, facilitate the access to bioinformatics tools in a user-friendly, scalable and reproducible way. Still, the integration of tools in such environments remains a cumbersome, time consuming and error-prone process. A major consequence is the incomplete or outdated description of tools that are often missing important information, including parameters and metadata such as publication or links to documentation. ToolDog (Tool DescriptiOn Generator) facilitates the integration of tools - which have been registered in the ELIXIR tools registry (https://bio.tools) - into workbench environments by generating tool description templates. ToolDog includes two modules. The first module analyses the source code of the bioinformatics software with language-specific plugins, and generates a skeleton for a Galaxy XML or CWL tool description. The second module is dedicated to the enrichment of the generated tool description, using metadata provided by bio.tools. This last module can also be used on its own to complete or correct existing tool descriptions with missing metadata.

General information
State: Published
Organisations: Department of Bio and Health Informatics, IT Service, Common Workflow Language Project, Institut Pasteur, University of Tartu, National Technical University of Ukraine, Kiev Polytechnic Institute, University of Freiburg
Contributors: Hillion, K., Kuzmin, I., Khodak, A., Rasche, E., Crusoe, M., Peterson, H., Ison, J., Ménager, H.
Number of pages: 14
Publication date: 2017
Peer-reviewed: Yes

Publication Information
Journal: F1000Research
Volume: 6
Article number: 2074
ISSN (Print): 2046-1402
Ratings:
Scopus rating (2017): CiteScore 1.59 SJR 0.926 SNIP 0.5
Scopus rating (2016): CiteScore 1.2 SJR 0.813 SNIP 0.423
Scopus rating (2015): CiteScore 0.87 SJR 0.62 SNIP 0.341
Scopus rating (2014): CiteScore 0.64 SJR 0.545 SNIP 0.294
Scopus rating (2013): CiteScore 0.4 SJR 0.224 SNIP 0.077
ISI indexed (2013): ISI indexed no
Original language: English
Electronic versions:
7badfa4e_5cfe_436c_9f0b_99a70bc52910_12974_Herve_Menager.pdf
DOIs:
10.12688/f1000research.12974.1

Bibliographical note
© 2017 Hillion KH et al. This is an open access article distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Source: FindIt
Tools and data services registry: a community effort to document bioinformatics resources

Life sciences are yielding huge data sets that underpin scientific discoveries fundamental to improvement in human health, agriculture and the environment. In support of these discoveries, a plethora of databases and tools are deployed, in technically complex and diverse implementations, across a spectrum of scientific disciplines. The corpus of documentation of these resources is fragmented across the Web, with much redundancy, and has lacked a common standard of information. The outcome is that scientists must often struggle to find, understand, compare and use the best resources for the task at hand. Here we present a community-driven curation effort, supported by ELIXIR—the European infrastructure for biological information—that aspires to a comprehensive and consistent registry of information about bioinformatics resources. The sustainable upkeep of this Tools and Data Services Registry is assured by a curation effort driven by and tailored to local needs, and shared amongst a network of engaged partners. As of September 2015, the registry includes 1633 resources, with depositions from 91 individual registrations including 40 institutional providers and 51 individuals. With community support, the registry can become a standard for dissemination of information about bioinformatics resources: we welcome everyone to join us in this common endeavour. The registry is freely available at https://bio.tools.
Scopus rating (2015): CiteScore 9.48 SJR 7.358 SNIP 2.631
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): CiteScore 8.74 SJR 6.64 SNIP 2.552
Web of Science (2014): Impact factor 9.112
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
Scopus rating (2013): CiteScore 8.46 SJR 6.801 SNIP 2.284
Web of Science (2013): Impact factor 8.808
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 2
Web of Science (2012): Impact factor 8.278
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 2
Scopus rating (2011): CiteScore 7.86 SJR 5.976 SNIP 2.19
Web of Science (2011): Impact factor 8.026
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 2
Scopus rating (2010): SJR 5.381 SNIP 2.034
Web of Science (2010): Impact factor 7.836
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 2
Scopus rating (2009): SJR 5.669 SNIP 1.874
BFI (2008): BFI-level 2
Scopus rating (2008): SJR 4.912 SNIP 1.578
Web of Science (2008): Indexed yes
Scopus rating (2007): SJR 5.1 SNIP 1.807
Web of Science (2007): Indexed yes
Scopus rating (2006): SJR 4.776 SNIP 2.051
Web of Science (2006): Indexed yes
Scopus rating (2005): SJR 5.092 SNIP 2.147
Web of Science (2005): Indexed yes
Scopus rating (2004): SJR 4.912 SNIP 1.971
Web of Science (2004): Indexed yes
Web of Science (2003): Indexed yes
Scopus rating (2002): SJR 3.895 SNIP 1.529
Scopus rating (2001): SJR 3.852 SNIP 1.393
Scopus rating (2000): SJR 4.936 SNIP 1.661
Web of Science (2000): Indexed yes
Scopus rating (1999): SJR 4.69 SNIP 1.548

Electronic versions:

Nucl._Acids_Res._2016_Ison_D38_47.pdf

DOIs:
10.1093/nar/gkv1116

Bibliographical note
© The Author(s) 2015. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License
Using registries to integrate bioinformatics tools and services into workbench environments

The diversity and complexity of bioinformatics resources presents significant challenges to their localisation, deployment and use, creating a need for reliable systems that address these issues. Meanwhile, users demand increasingly usable and integrated ways to access and analyse data, especially within convenient, integrated “workbench” environments. Resource descriptions are the core element of registry and workbench systems, which are used to both help the user find and comprehend available software tools, data resources, and Web Services, and to localise, execute and combine them. The descriptions are, however, hard and expensive to create and maintain, because they are volatile and require an exhaustive knowledge of the described resource, its applicability to biological research, and the data model and syntax used to describe it. We present here the Workbench Integration Enabler, a software component that will ease the integration of bioinformatics resources in a workbench environment, using their description provided by the existing ELIXIR Tools and Data Services Registry.