HHS Public Access

Author manuscript

Cell Syst. Author manuscript; available in PMC 2020 November 27.

Published in final edited form as:

Cell Syst. 2019 November 27; 9(5): 417–421. doi:10.1016/j.cels.2019.09.011.

FAIRshake: Toolkit to Evaluate the FAIRness of Research Digital Resources

Daniel J.B. Clarke[1], **Lily Wang**[1], **Alex Jones**[2], **Megan L. Wojciechowicz**[1], **Denis Torre**[1], **Kathleen M. Jagodnik**[1], **Sherry L. Jenkins**[1], **Peter McQuilton**[3], **Zachary Flamholz**[1], **Moshe C. Silverstein**[1], **Brian M. Schilder**[1], **Kimberly Robasky**[4], **Claris Castillo**[4], **Ray Idaszak**[4], **Stanley C. Ahalt**[4], **Jason Williams**[5], **Stephan Schurer**[6], **Daniel J. Cooper**[6], **Ricardo de Miranda Azevedo**[7], **Juergen A. Klenk**[2], **Melissa A. Haendel**[8], **Jared Nedzel**[9], **Paul Avillach**[10], **Mary E. Shimoyama**[11], **Rayna M. Harris**[12], **Meredith Gamble**[13], **Rudy Poten**[13], **Amanda L. Charbonneau**[12], **Jennie Larkin**[14], **C. Titus Brown**[12], **Vivien R. Bonazzi**[15], **Michel J. Dumontier**[7], **Susanna-Assunta Sansone**[3], **Avi Ma'ayan**[1,*]

[1]Department of Pharmacological Sciences, Mount Sinai Center for Bioinformatics, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA [2]Deloitte Consulting, 1919 N Lynn St., Arlington, VA 22209, USA [3]Oxford e-Research Centre, Department of Engineering Science, University of Oxford, Oxford OX1 3QG, UK [4]Renaissance Computing Institute, University of North Carolina, Chapel Hill, USA [5]Cold Spring Harbor Laboratory, 1 Bungtown Rd., Cold Spring Harbor, NY 11724, USA [6]University of Miami, Pharmacology, 1600 NW 10th Ave., Miami, FL 33136, USA [7]Institute of Data Science Maastricht University, Universiteitssingel 60, 6229 ER Maastricht, the Netherlands [8]Oregon State University, 307 Linus Pauling Science Center, 2900 SW Campus Way, Corvallis, OR 9733, USA [9]The Broad Institute, 415 Main St., Cambridge, MA 02142, USA [10]Harvard Medical School, Department of Biomedical Informatics, 10 Shattuck St., Boston, MA 02115, USA [11]Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, WI 53226, USA [12]Department of Population Health and Reproduction, School of Veterinary Medicine, UC Davis, Davis, CA 95616, USA [13]Curoverse, Somerville, MA 02144, USA [14]National Institute of Diabetes, Digestive, and Kidney Diseases (NIDDK), National Institutes of Health (NIH), 6707 Democracy Blvd., Bethesda, MD 20817, USA [15]Office of the Director, National Institutes of Health, 8600 Rockville Pike, Rm. 2S-20, Bethesda, MD, 20894, USA

Abstract

As more digital resources are produced by the research community, it is becoming increasingly important to harmonize and organize them for synergistic utilization. The findable, accessible, interoperable, and reusable (FAIR) guiding principles have prompted many stakeholders to consider strategies for tackling this challenge. The FAIRshake toolkit was developed to enable the establishment of community-driven FAIR metrics and rubrics paired with manual and automated

*Correspondence: avi.maayan@mssm.edu.

Availability

The primary interface to FAIRshake is at: https://fairshake.cloud.

The FAIRshake Chrome extension is available from: https://chrome.google.com/webstore/detail/fairshake/pihohcecpiomegpagadljmdifpbkhnjn?hl=en-US.

The FAIRshake source code is available from GitHub at: https://github.com/MaayanLab/FAIRshake.

FAIR assessments. FAIR assessments are visualized as an insignia that can be embedded within digital-resources-hosting websites. Using FAIRshake, a variety of biomedical digital resources were manually and automatically evaluated for their level of FAIRness.

Introduction

The findable, accessible, interoperable, and reusable (FAIR) guiding principles describe an urgent need to improve the infrastructure supporting scholarly data reuse and outline several existing resources that already demonstrate various aspects of FAIR and associated driving technologies (Wilkinson et al., 2016). A specific emphasis has been placed on ensuring that machines can exchange interpretable data and metadata. Following the FAIR principles, the resource description framework (RDF) is a key globally accepted framework for data and knowledge representation that is intended to be read and interpreted by machines. A critical challenge in fulfilling the goals outlined by the FAIR guiding principles is the lack of consensus with respect to agreement on using certain standards. In an effort to address this challenge, a comprehensive community-driven approach was taken to assemble descriptions of standards, repositories, and policies and make them easily accessible from one source (Sansone et al., 2019). By collecting community-accepted elements of this kind, FAIRsharing can reveal domain-relevant community standards with respect to the FAIR principles. Several initiatives have begun to develop their own understandings of FAIRness and developed some methods of assessing FAIRness by self- and peer-reviewed manual question-answer approaches (Cox and Yu, 2017; Dillo and De Leeuw, 2014). Because there are different strategies for asserting FAIRness, efforts so far have been independent of one another and as such not comparable. While the biomedical research community at large mostly embraces the FAIR guidelines, there is still some confusion about the difference between being FAIR and being open access, what it means to be FAIR, and how the FAIR principles compare with other standards (Hasnain and Rebholz-Schuhmann, 2018).

In order to bring the FAIR principles into practice and to provide more clarity about their meaning, a template was created for constructing FAIR metrics around the original FAIR guiding principles (Wilkinson et al., 2018). The publication that describes the FAIR metrics contains self-evaluations by nine organizations. While the FAIR metrics are provided on GitHub so the community can contribute to their development, the original authors of the FAIR metrics claim that these metrics are universal and aim to cover all types of digital objects for all organizations. In the publication of the universal FAIR metrics, it was envisioned that a framework for automated evaluation of FAIRness could be devised using self-describing and programmatically executable metrics. This was followed by an initial attempt to develop a system that evaluates FAIR maturity level (Wilkinson et al., 2019).

While the universal FAIR metrics developed by some of the original authors of the FAIR guiding principles provide a concrete guide on how to assess FAIRness, the universal FAIR metrics may not fit all domains and specific requirements. For example, a recent review by a group consisting of biopharma researchers and representatives suggests that the biopharma community has unique requirements, so being FAIR for them may mean a different thing compared with other digital object producers (Wise et al., 2019).

In order to facilitate digital resource producers to define, assess, and implement their own FAIRness criteria for their specialized specific projects, we developed FAIRshake. FAIRshake enables the community to develop new standards, or reuse existing standards, to define and evaluate FAIRness. Thus, FAIRshake allows the co-existence of multiple metrics and rubrics, enabling the community to develop standards more democratically. FAIRshake is a toolkit that enables the systematic assessment of the FAIRness of any digital resource. Compared with previous attempts to develop FAIRness assessment tooling, the FAIRshake toolkit has more features. It contains a database that enlists users, projects, digital resources, metrics, rubrics, and assessments (Figure 1); it is a full-stack application with a user interface; and it comes with a browser extension and a bookmarklet to enable viewing and submitting assessments from any website. The FAIR assessment results are visualized as an insignia that represents the FAIR score in a compact grid of squares colored in red, blue, and purple. Below, we briefly describe the various components of FAIRshake, how they are related to each other, and how the FAIRshake system has been already used to assess the FAIRness of thousands of digital resources for numerous projects.

The FAIRshake Framework

Overall, FAIRshake provides mechanisms to associate digital objects with rubrics and metrics to perform FAIR assessments. These assessments are communicated via the FAIR insignia (Figure 1). The FAIRshake toolkit is composed of elements that include a full-stack web-server application containing a user interface with a search engine, a backend database, and an application-programming interface (API), as well as a Chrome extension and a bookmarklet (Table 1). FAIRshake also contains FAIR analytics modules that produce statistical reports about collections of assessments for a specific project. In an effort to make FAIRshake adhere to the FAIR guidelines, the FAIRshake endpoint-REST API is machine readable with documentations for SmartAPI, Swagger/OpenAPI (https://swagger.io/), and CoreAPI (https://www.coreapi.org/). The API can be accessed via the human-friendly counterparts of these specifications with the REST Framework API explorer (https://www.django-rest-framework.org/topics/browsable-api/), Swagger UI (https://swagger.io/tools/swagger-ui/), and CoreAPI UI. A Jupyter Notebook and YouTube tutorials are available to guide users through the process of using the FAIRshake interface and accessing FAIRshake programmatically.

To perform and visualize FAIR assessments with FAIRshake, users must follow several steps (Table 2). First, users are required to sign up. Sign up is available via standard registration and via OAuth implementation of GitHub, ORCID, and Globus (Foster and Kesselman, 1997). Next, users are required to create a project. Projects are a bundle of thematically relevant digital resources. Project descriptions contain minimal information for identifying, displaying, and indexing the project. Within projects, users can assess the FAIRness of an arbitrary collection of digital resources. Project analytics are available to help a user better understand the overall FAIRness of the digital resources contained within the project. Next, users need to associate the digital objects in their projects with rubrics and metrics. FAIR metrics are questions that assess whether a digital object complies with a specific aspect of FAIR. A FAIR metric is directly related to one of the FAIR guiding principles. In order to make FAIR metrics reusable, FAIRshake collects information about each metric, and when

users attempt to associate a digital resource with metrics, existing metrics are provided as a first choice. FAIR metrics represent a human-described concept that may or may not be automated; automation of such concepts can be done independently by linking actual source code to reference a persistent identifier of that metric semantic. Without linked code, metrics are simply questions that can be answered manually. FAIRshake defines several categorical answer types to FAIR metrics when manually assessed that are ultimately quantified to a value in a range between zero and one, or take the property of undefined. Programmatically, metric code can quantify the satisfaction of a given FAIR metric within this same range. The concept of a metric is supplemented with that of a FAIR rubric. A FAIR rubric is a collection of FAIR metrics. An assessment of a digital resource is performed using a specific rubric by obtaining answers to the metrics within the rubric. The use of a FAIR rubric makes it possible to establish a relevant and applicable group of metrics for a large number of digital resources, typically under the umbrella of a specific project, while enabling reuse of metrics both for comparisons across different projects and for automation. Linking rubrics to digital resources by association helps users understand the context of the FAIR metrics that are best suited to assess the digital resources in their projects.

FAIR assessments can be performed manually or automatically on a digital resource that is associated with a rubric. Leveraging RDF, FAIRshake automatically extracts RDF-expressed schema.org metadata from URLs with Extruct (Termehchy and Winslett, 2010), a library for extracting embedded metadata from HTML markup. This approach is utilized by major search engines to index websites and bind information together. Using this RDF-expressed metadata alone, some FAIR metrics are automatically resolved, including those designed with RDF in mind. As schema.org expands its vocabulary through initiatives such as Bio-schemas (Garcia et al., 2017), RDF will enable more automated assessments. Adopting other non-RDF based standards has also been accomplished with FAIRshake. In summary, any assessment of a digital resource within FAIRshake attempts to obtain answers automatically. The newly assessed digital resource will now have an associated insignia that reflects the results of the FAIR assessment.

The FAIRshake insignia uses a color gradient from blue (satisfactory) to red (unsatisfactory), visualizing how well a digital resource satisfied the FAIR metrics of the chosen rubric. Because the same digital resources can be assessed by different rubrics, composed of different metrics, the insignia dynamically expands to fit all assessments. If answers to the rubric are missing, the squares associated with these metrics will be colored in gray. Developers of data and tool portals can visualize FAIRshake insignias on their site. A standalone JavaScript library for generating the insignias at any hosting website with few lines of code is provided. Alternatively, through this library, a browser extension and bookmarklet were developed for rendering the visualization of FAIR insignias on any website without the need of the hosting site to modify their website’s source code.

FAIRshake was already applied to assess the FAIRness of many digital objects that belong to various high-profile projects (Table 3). The first use of FAIRshake involved the manual assessment of 150 tools and datasets developed by the Alliance of Genome Resources (AGR) (https://www.alliancegenome.org/).Detailed results and breakdown of these assessments were captured in an HTML table and associated Jupyter Notebooks that are

available at https://maayanlab.github.io/AGR-FAIR-Website/. Overall, we observed that the examined AGR tools and datasets scored well in regard to providing data for download, use of ontologies, and providing contact information, while most AGR tools and datasets did not provide the source code, versioning information, or API access (Figure 2).

Limitations and Challenges

The FAIRshake platform is complex. Before beginning to use FAIRshake, users must have some training about concepts like FAIR metrics and rubrics. Associating a digital object with the "right" rubric is not trivial. While the co-existence of multiple rubrics provides flexibility and freedom in the choice of how one may define FAIR, this approach has the risk of having too many different interpretations of the guidelines with undesired partial redundancy that is not consolidated into a shared standard. We hope that with increasing use of FAIRshake, users will be able to reuse metrics without the need to create new ones. This can potentially enable the development of a grassroots, eventually widely accepted standard.

Incentivizing Users with Carrots and Sticks—When community standards are developed, global adoption is needed in order to facilitate their true enabling potential. Community adoption of FAIRness-endorsed standards is challenging, because digital object producers do not always see the added benefit in spending the time, effort, and resources to FAIRify their digital products. In most cases, digital object producers will use the excuse that they do not have the required resources to spend on FAIRification. Thus, there are currently few incentives for them to make their products FAIRer. Such incentives can be nurtured. Specifically, these incentives can be divided into carrots and sticks. If more FAIRenabled resources become used by the community—for example, if researchers will begin using resources such as Google Datasets (Halevy et al., 2016) more frequently for their research—digital object producers will want to be listed there. If data citations begin to soar, digital object producers will have the incentive to participate. These are carrot incentives for FAIRification. At the same time, funders and journals can demand that published data meet certain community-accepted standards before they are accepted for publication or become eligible for funding. This is achieved, for example, when gene expression data are deposited into the Gene Expression Omnibus (GEO) or when solved protein structure coordinates are deposited into the Protein Data Bank (PDB). Funders and journals requiring researchers to take the needed steps in order to ensure the FAIRness of the digital resources they produce is a stick approach. However, convicting funders and journals to enforce new rules is often difficult due to a possible backlash from the researchers, who will simply "go somewhere else." Ultimately, FAIRification benefits all—the digital object producers, the journals, the funders, and the users who are the real consumers of these digital resources. The question, and challenge, is simply determining who is responsible for performing the FAIRification task, who will pay for it, and what it means to do it— and, perhaps, overdo it. The concept of "digital objects needs to be born FAIR" suggests that this activity needs to be done by the data producers at an early stage.

Discussion

FAIRshake was developed as a toolkit to promote the FAIRification of digital objects produced by research projects. FAIRshake is not intended to judge or penalize digital resource producers but rather to promote the awareness about standards. The purpose of FAIRshake is to guide digital object producers to implement community-accepted best practices for their own benefit of attracting, retaining, and enabling more engagement with the digital objects they produce. There is common confusion between assessing the quality of a resource and assessing its FAIRness. It should be made clear that FAIRshake was designed to assess FAIRness, and a low FAIR score does not mean that a digital resource is lacking quality, usefulness, user friendliness, or innovation. Another aspect of confusion about FAIR is the association of FAIRness with openness. Being FAIR does not entail making data, source code, tools, or any other digital resource free and openly available. Rather, the FAIR guidelines only require that access and usage policies are provided and stated clearly (Haendel et al., 2016; Mons et al., 2017).

By facilitating the creation of both manual and automated FAIR assessments and enabling FAIR metric findability, FAIRshake promotes the involvement of more stakeholders. Starting with the process of manual FAIR assessments, the capacity for automation is expected to further expand as more adoption is realized. The findability of FAIR metrics within FAIRshake makes it possible to design community-adopted metrics that can be customized for specific purposes but at the same time, for general and generic uses. FAIRshake strives to evolve with the community, adding new features to accommodate community demands as they arise while facilitating more assessments. With its feature of enabling the development of FAIR metrics and rubrics by any user, the assessment of digital resources can happen before the community agrees on the definition of what it means to be FAIR. FAIRshake facilitates dynamic metric re-use, and it provides analytical tools to understand the global and relative performance of resources and metrics. With transparency, FAIRshake enables the community to study the FAIRness of the resources they produce and use.

FAIRshake was developed to meet the demands of the biomedical research community. With integration of a number of community-accepted standards, including RDF, DATS, SmartAPI, and schema.org, FAIRshake is already capable of facilitating FAIR assessments of a diverse set of digital objects, including datasets, tools, repositories, and APIs. Throughout our initial assessments, it has become clear that many established community standards are not being employed within the biomedical research community, largely due to a lack of awareness. As the community continues to evolve toward better defining FAIRness, the FAIR metrics are expected to converge, and the FAIR assessments are likely to become more automated.

FAIRshake will continue to evolve with community demand. Continued improvements to the clarity, usability, and FAIRness of FAIRshake are planned. Similarly, through integration with existing FAIR-embracing resources such as FAIRSharing, FAIRshake will enable the display of assessments on digital resource landing pages so that a broader community of users will become more aware of FAIRshake. The FAIRshake platform codebase can be

reused for the assessment of other digital and physical products, such as publications, events, books, and courses. However, such assessments may not be relevant to the FAIR guiding principles. Nevertheless, the FAIRshake platform is flexible enough that it can facilitate other related applications, even potentially repurposing FAIRshake as a platform for scientific peer review.

ACKNOWLEDGMENTS

This work was partially supported by the National Institutes of Health, United States, grant numbers OT3-OD025467, OT3-OD025459, and U54-HL127624.

REFERENCES

Cox S and Yu J. (2017) OzNome 5-star Tool: A Rating System for making data FAIR and Trustable. eResearch Australasia 2017.

Dillo I, and De Leeuw L. (2014). Data Seal of Approval: Certification for sustainable and trusted data repositories (The Hague: Data Archiving and Networked Services (DANS)), p. 20.

Foster I, and Kesselman C. (1997). Globus: A metacomputing infrastructure toolkit. The International Journal of Supercomputer Applications and High Performance Computing 11, 115–128.

Garcia L, Giraldo O, Garcia A and Dumontier M. (2017) Bioschemas: schema. org for the Life Sciences. Proceedings of SWAT4LS.

Haendel M, Su A, McMurry J, Chute C, Mungall C, Good B, Wu C, McWeeney S, Hochheiser H, and Robinson P. (2016). Metrics to assess value of biomedical digital repositories: response to RFI NOT-OD-16–133 (Zenodo).

Halevy A, Korn F, Noy NF, Olston C, Polyzotis N, Roy S, and Whang SE (2016), Proceedings of the 2016 International Conference on Management of Data ACM, San Francisco, California, USA, pp. 795–806.

Hasnain A, and Rebholz-Schuhmann D. (2018). Assessing FAIR Data Principles Against the 5-Star Open Data Principles In The Semantic Web: ESWC 2018 Satellite Events (Springer International Publishing), pp. 469–477.

Mons B, Neylon C, Velterop J, Dumontier M, da Silva Santos LOB, and Wilkinson MD (2017). Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. Inf. Serv. Use 37, 49–56.

Sansone S-A, McQuilton P, Rocca-Serra P, Gonzalez-Beltran A, Izzo M, Lister AL, and Thurston M; FAIRsharing Community (2019). FAIRsharing as a community approach to standards, repositories and policies. Nat. Biotechnol. 37, 358–367. [PubMed: 30940948]

Termehchy A, and Winslett M. (2010). EXTRUCT: using deep structural information in XML keyword search. Proceedings VLDB Endowment 3, 1593–1596.

Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da Silva Santos LB, Bourne PE, et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3, 160018.

Wilkinson MD, Sansone S-A, Schultes E, Doorn P, Bonino da Silva Santos, L.O., and Dumontier, M. (2018). A design framework and exemplar metrics for FAIRness. Sci. Data 5, 180118.

Wilkinson MD, Dumontier M, Sansone S-A, Bonino da Silva Santos, L.O., Prieto, M., Batista, D., McQuilton, P., Kuhn, T., Rocca-Serra, P., Crosas, M., and Schultes, E. (2019). Evaluating FAIR maturity through a scalable, automated, community-governed framework. Sci. Data 6, 174. [PubMed: 31541130]

Wise J, de Barron AG, Splendiani A, Balali-Mood B, Vasant D, Little E, Mellino G, Harrow I, Smith I, Taubert J, et al. (2019). Implementation and relevance of FAIR data principles in biopharmaceutical R&D. Drug Discov. Today 24, 933–938. [PubMed: 30690198]

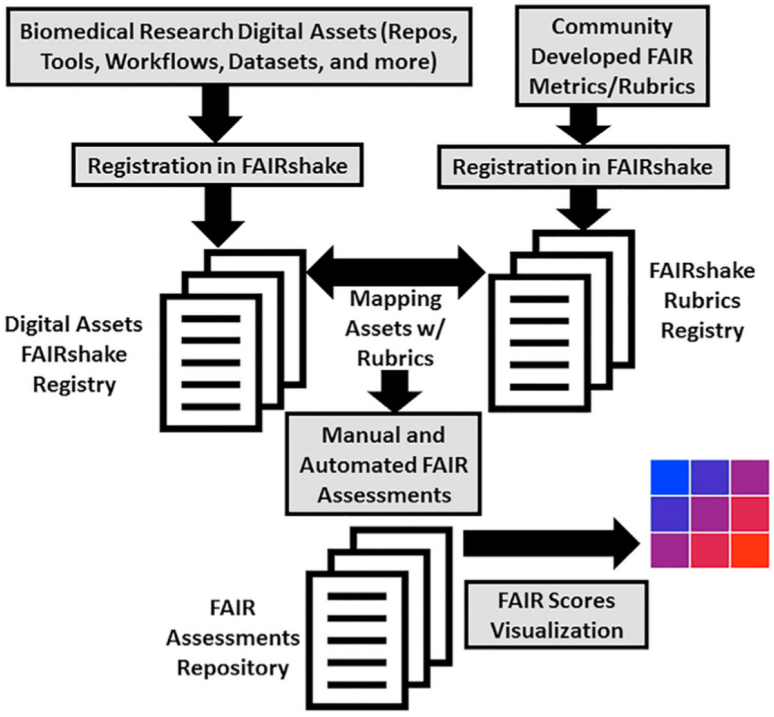

Figure 1. A Diagram Illustrating FAIRshake's Workflow

Digital resources from various projects are paired with FAIR metrics and rubrics to perform assessments that are visualized with the FAIR insignia.

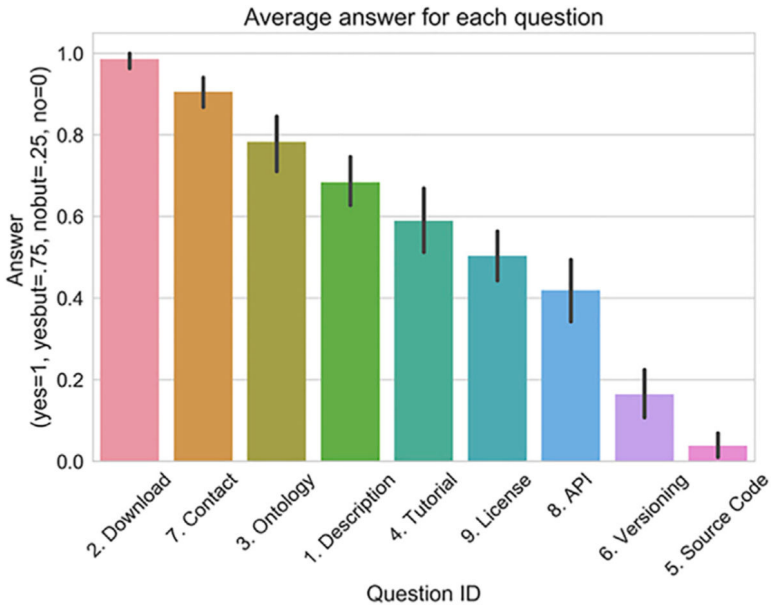

Figure 2. FAIR Assessment of AGR Tools

Distribution of average FAIR scores for 132 AGR tools assessed with an initial set of 9 FAIR metrics.

Table 1.

The Major Components of the FAIRshake Toolkit

Feature	Description	URL
Search engine	The FAIRshake search engine can be used to identify project, digital objects, rubric, and metrics.	https://fairshake.cloud/
Open source code	The FAIRshake project is open source and available from GitHub.	https://github.com/MaayanLab/FAIRshake
Swagger API	The FAIRshake API is documented in Swagger.	https://fairshake.cloud/swagger/
YouTube tutorials	There are several video tutorials on YouTube that describe how to use FAIRshake.	https://www.youtube.com/watch?v=7u0c4-yzXgA&list
FAIR analytics	Example FAIR analytics stats applied to AGR resources.	https://fairshake.cloud/project/10/stats/
Jupyter notebook tutorial	There is a Jupyter notebook tutorial that guides users on how to use FAIRshake programmatically.	https://fairshake.cloud/documentation/
Bookmarklet	Users can install a bookmarklet that enables FAIR evaluations of digital objects listed on any website.	https://fairshake.cloud/bookmarklet/
Browser extension	Users can install a browser that enables FAIR evaluations of digital objects listed on any website.	https://fairshake.cloud/chrome_extension/

Table 2.

Steps to Perform and Visualize FAIR Assessments with FAIRshake

Step	Instructions
Sign up	Fill in a registration form.
Log in	Enter user name and password.
Start a project	Fill out a form that describes the project.
Register digital objects	Register digital objects in FAIRshake and associate them with the project.
Add a FAIR metric	Fill out a form to set up the FAIR metric question and possible answers.
Add a FAIR rubric	Associate a collection of FAIR metrics with a new rubric.
Associate rubrics with digital objects	Associate each registered digital object from the project with a registered rubric.
Perform assessments	Answer each FAIR metric question to fill in the FAIR evaluation questionnaire.
Visualize the FAIR results with an insignia	Hosting websites can use a JavaScript library to visualize FAIR assessments of the digital objects they host. Alternatively, the insignia can be visualized via a browser extension or a bookmarklet.

Table 3.

Case Studies where FAIRshake Was Utilized to Perform FAIR Assessment to Evaluate Various Collections of Digital Objects

Case Study	Resource	Number of Digital Objects	URL
Manual assessment of AGR Datasets and bioinformatics tools	AGR https://www.alliancegenome.org/	150	https://fairshake.cloud/project/10/
Automated assessment of the resources listed on FAIRsharing	FAIRsharing https://fairsharing.org/	1,176	https://fairshake.cloud/project/14/
Automated assessment of TOPMed studies on dbGAP	dbGAP ftp://ftp.ncbi.nlm.nih.gov/dbgap/studies/	27	https://fairshake.cloud/project/61/
Automated assessment of APIs listed on SmartAPI	SmartAPI https://smart-api.info/	35	https://fairshake.cloud/project/53/
Automated assessment of NCBI tool and databases	NCBI https://www.ncbi.nlm.nih.gov/	227	https://fairshake.cloud/project/71/
Automated assessment of Common Fund programs datasets	NIH Common Fund https://commonfund.nih.gov/	31282	https://fairshake.cloud/project/87/