

Appendix

README Auto-generation

A common practice with data management and data publishing is providing proper documentation often in the form of a README.txt. This is a step that we outline in guides and tutorials that we provide to users of ReDATA. However, through our data curation experience, we noticed that users struggle to adopt this best practice, often requiring a request during the data curation review process. To ensure higher adoption and more consistent formatting, the construction of the README.txt was shifted away from an expectation on the ReDATA user to include a readme file at the time of deposit (whether self-created or based on a supplied template) to a task that a data curator conducted as part of the data curation process. In the construction of the README.txt we decided to treat the customized content for each deposit as metadata records and to develop software to enable automation. Here, a plain text template is used in combination with jinja2, a Python template engine library. By treating the contents as metadata, the software can populate the template with the README metadata to construct the README.txt. These metadata include general ones that are included as part of the data deposit submission (e.g. title, author list, citation, license, DOI link), which is retrievable from the Figshare API, along with metadata to better describe the data deposit. For the latter we use a Qualtrics survey to gather the metadata, prompting the depositor/user to specify information about the files and folder contents, the methods and software use, and contributing roles of authors. This README.txt is then uploaded by the data curator during the data curation process. One advantage of a metadata-templating approach is the easy ability to update the README.txt with changes to the deposit or the description. For example, a common correction needed involves changing the title of the data. These changes are easily consumed from the Figshare API

and are incorporated into a revised README.txt. Our methodology is further discussed in Ly et al. (2021) with illustrations.

Authentication with ReQUIAM

ReQUIAM is a pure Python application running on a dedicated virtual private server (VPS). Users log in to ReDATA using the university's single sign-on. In order to control access, quotas based on the kind of user (faculty, student, etc.), and to assign users to groups based on their departmental affiliation, ReQUIAM leverages the university's enterprise directory service (EDS) and associated identity and access management services such as Grouper (InCommon 2021). On a nightly basis, ReQUIAM reads the information from EDS for all active users at the institution and assigns users into Grouper groups based on several rules. There are groups for controlling who can log in to ReDATA (only users with library privileges), assigning quotas (undergraduate students, graduate students, faculty, and staff have different initial quotas), and assigning individuals to groups for organization purposes. Information about the Figshare groups functionality is available elsewhere (Reed 2016; Figshare 2021). The Grouper group information is written back to EDS automatically via systems operated by university IT. ORCID ids are attached to individual EDS records when a user visits a separate page (not operated by the ReDATA team) which links their ORCID account to their university account. Upon logging in to ReDATA, the EDS information is read by Figshare and the appropriate quotas, groups, are set and the ORCID id is linked to the user's profile and attached to future deposits.

Appendix References

- Figshare. 2021. “Figshare for Institutions Admin User Guide: Home.” March 1, 2021. <https://figshare.libguides.com/figshare-for-institutions-admin-user-guide>.
- InCommon. 2021. “Grouper.” InCommon. 2021. <https://www.incommon.org/software/grouper/>.
- Ly, C., Fernando Rios, and Megan Hardeman. (2021). *Data Curation in the Cloud with the Figshare API: Presented by the University of Arizona* (Version 1). figshare. <https://doi.org/10.6084/m9.figshare.14730420.v1>
- Reed, Robyn B. 2016. “Figshare for Institutions.” *Journal of the Medical Library Association* : JMLA 104 (4): 376–78. <https://doi.org/10.3163/1536-5050.104.4.031>