GitLab as a tool for Research Data Management: Candidate for a Base4NFDI general service

Marcel Tschöpe[1], Dominik Brilhaus[2], Jonathan Bauer[1], Dirk von Suchodoletz[1]

[1] eScience Department, University of Freiburg, Germany
[2] Cluster of Excellence on Plant Sciences (CEPLAS), Heinrich-Heine-University Düsseldorf, Germany

1. GitLab As DataHUB and Science Gateway

The DataPLANT team found that Git and the GitLab framework met key needs for a data hub in RDM. These needs included versioning, group collaboration, and easy access management. For this, Git provides versioning, which makes it possible to track and undo changes within a Git repository. To store the repository data, GitLFS is used, especially for large files.

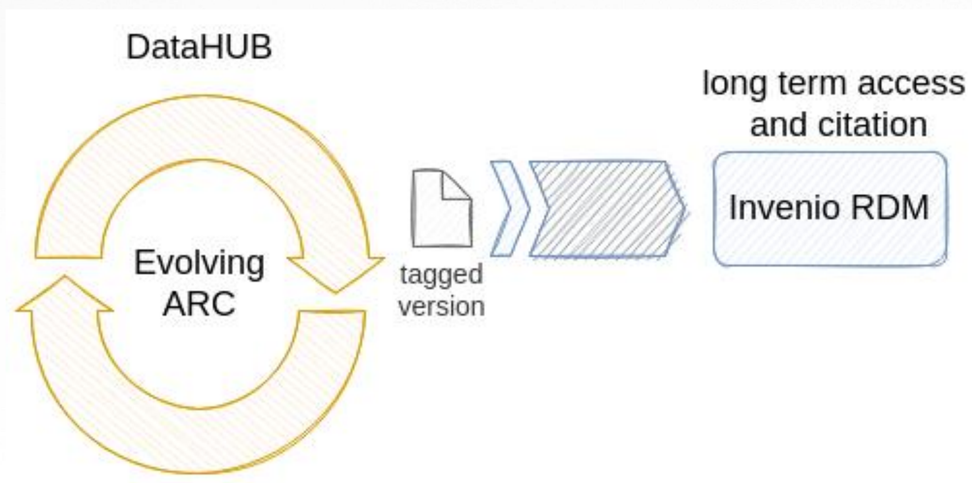


Figure 1: Evolving ARC Workflow

2. User Management

GitLab itself adds features for fine-grained access management, allowing users to form collaborative groups that they can manage themselves, fostering easy and flexible collaboration across institutions.

GitLab gives detailed access control, allowing users to create and manage their own collaboration groups, which helps collaboration across organizations happen easily and flexibly.

3. User-friendliness

Git provides functions such as branching, forking, and merging, which are highly beneficial for the Annotated Research Context's RDM lifecycle.

DataPLANT offers an user-friendly tool called "ARC Commander" that simplifies Git commands, making the workflow easier for users.

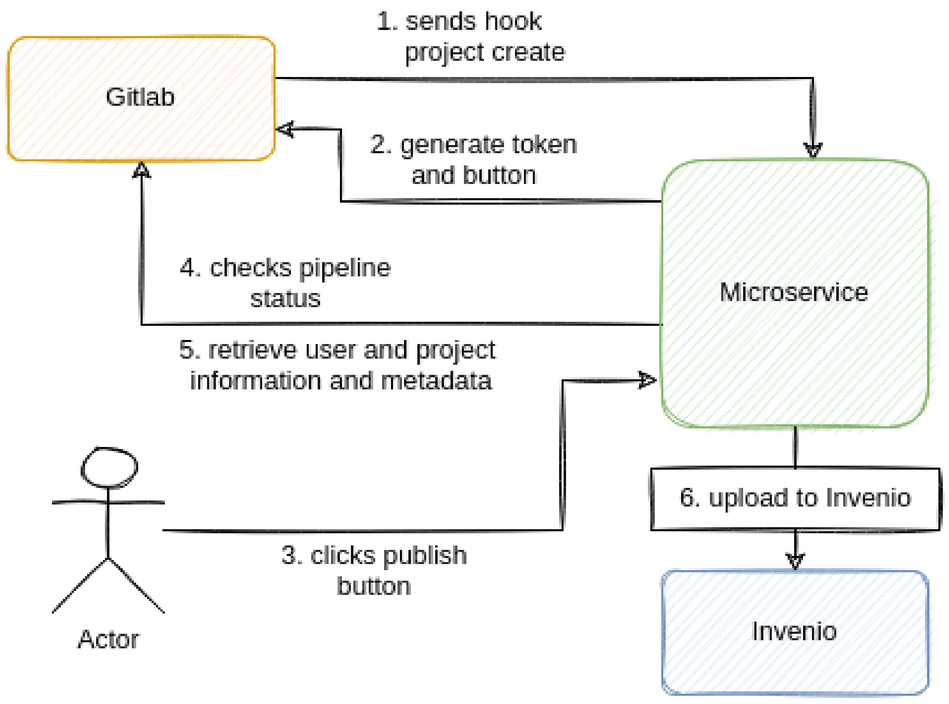


Figure 3: Microservice Workflow

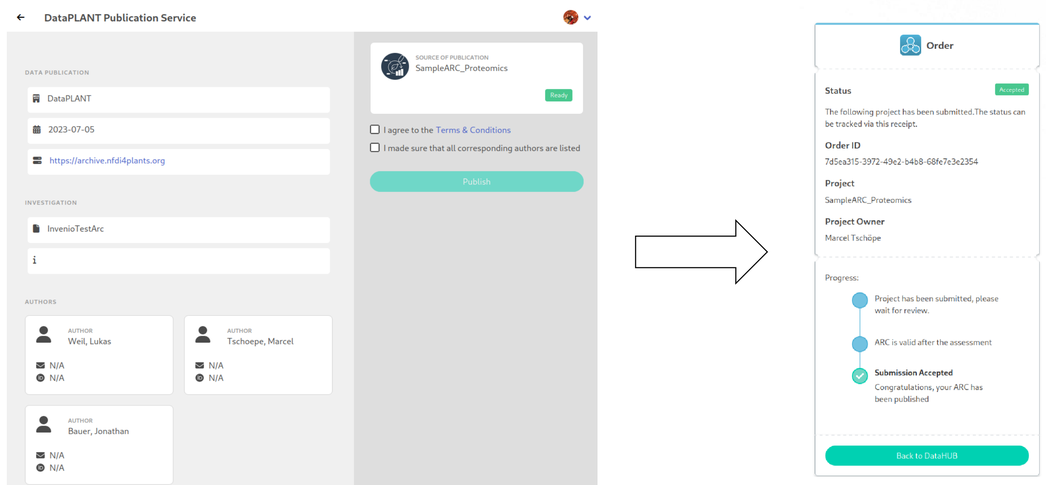
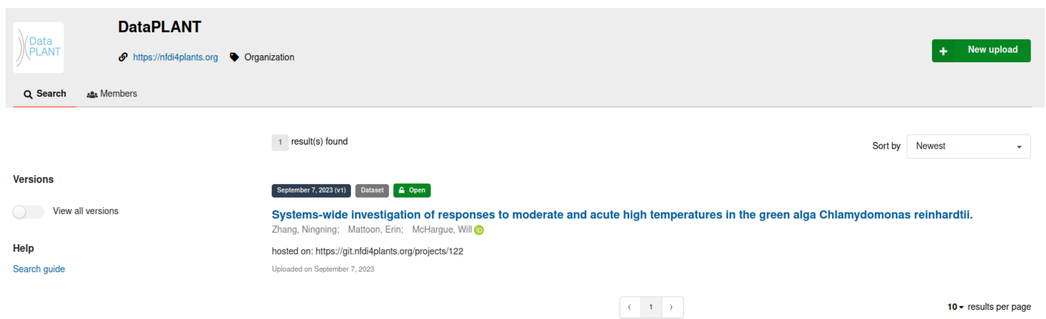


Figure 2: OAuth Application

4. Publication Workflow with API

Microservices were developed for the ARC publishing workflow due to the limitations of GitLab. These services use the GitLab REST API. In addition, with GitLab's CI/CD pipeline workflows, it is easy for a user to publish their ARCs.
This microservice is integrated into GitLab as an OAuth application (Figure 2, 3).



5. Future Proof AAI

In DataPLANT, user management relies on established authentication and authorization infrastructures (AAIs), such as Life Sciences AAI and ORCID, combined with local authentication within the central DataPLANT authentication service. The underlying infrastructure uses KeyCloak, developed by Red Hat, which supports modern authentication protocols such as OpenID-Connect and SAML. It allows for the integration of multiple AAIs and handles the complex issue of identity brokering.

By implementing an AAI identity management system that connects to GitLab and other services through these protocols, user management is simplified. KeyCloak allows the community to use their existing accounts, such as their home institution, Life Sciences AAI, ORCID iD, or a future NFDI AAI.

Funded by
Deutsche Forschungsgemeinschaft
German Research Foundation
Project no. 442077441