

Science Demonstrator 9: PROV-Template Registry and Expansion Service (Use Case IC_10)

Please provide your feedback on this Science Demonstrator using the questionnaire at <https://survey2.icos-cp.eu/ENVRplus-evaluator/>

Overview

PROV-Template is a proposed standard for converting existing process output such as log files into representations following the PROV Data Model (PROV-DM) specification for describing provenance of electronic resources in machine readable, structured form. Besides the potential advantage that existing process implementations can be enabled to generate PROV-DM conforming data without the need to change their underlying codebase, the general notion of using templates for describing provenance traces for recurring workflows can be used to foster interoperability and best practices for provenance data generation across individual communities. Following this motivation, the ENVRplus PROV-Template registry and expansion service prototype has been designed as public platform for describing, storing and sharing PROV-Templates across members of different RI, including a dedicated Web API for instantiating stored templates with individual data.

Scientific Objectives

- Provide convenient means for communities to start experimenting with PROV-DM
- Enable the sharing and re-use of registered PROV-Templates in order to foster interoperable provenance traces across communities.
- Provide a Python based implementation of the PROV-Template expansion mechanism.

Description

The PROV-Template Registry and Expansion Service consists of two main components. A Web based User Interface provides means to register and describe PROV-Template conforming documents in order to publish and share them online. Users can browse the available templates based on metadata and visual graph representations and re-use them either by copying and modifying their code, ideally sharing back the results via the same platform, or by instantiating them directly with own data. The latter is made possible by the second component, the expansion service which encapsulates a custom Python based implementation of the PROV-Template expansion mechanism behind an easy to use Web API. A broad overview on the system architecture is provided in Figure 1.

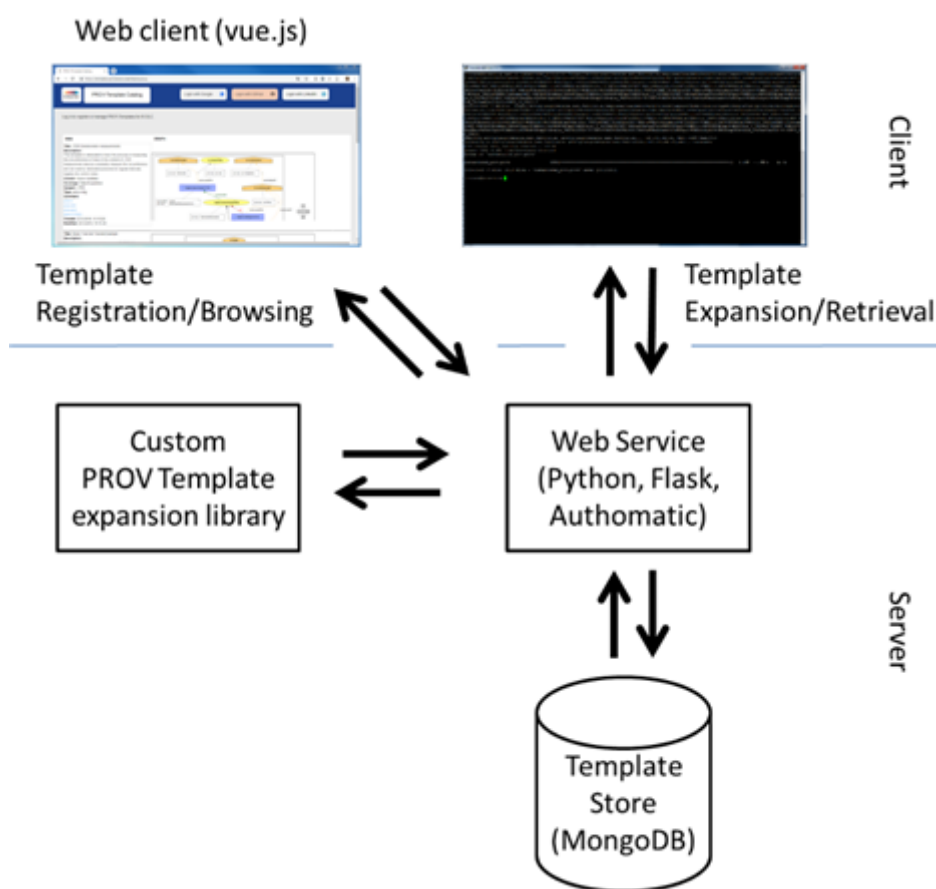


Figure 1 System Architecture

The User Interface for registering and browsing Templates is shown in Figure 2. Registered templates are provided in dedicated rows of a scrollable list, each entry featuring metadata such as title, creator, coverage, subject, etc. alongside a visualization of the template's PROV graph. Dedicated links allow the immediate download of the template in different PROV-DM serializations.

While all registered templates are accessible for everyone, the registration and subsequent modification of a specific template requires authentication. The current version of the prototype allows users to log in by choosing one of three popular social media profiles, delegating authentication to the respective identity providers and thus not requiring the collection and management of sensitive user information.

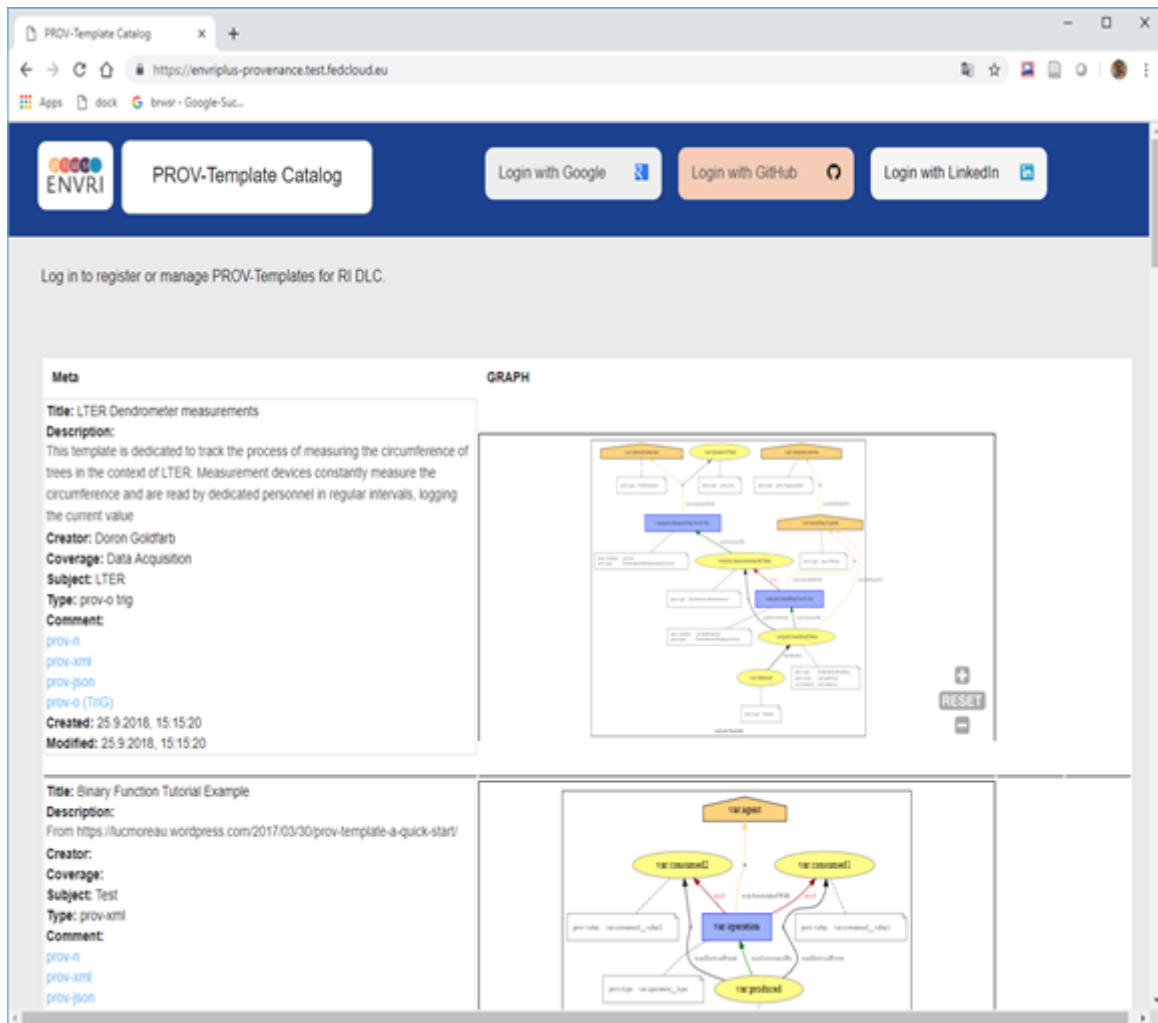


Figure 2 User Interface

The description of the process of registering templates and the API for their expansion is provided in a dedicated manual[1] and is thus omitted here. Additional information, including the source code for the Web service[2] and the expansion library[3], is provided on Github. The current content of the registry reflects the results of community experiments conducted in the context of ENVRIplus task T8.3 and includes templates for various aspects of the data life cycles of the involved communities. Besides experiments conducted throughout ENVRIplus, the prototype has already been taken up by other projects such as DARE[4], contributing significant system level extensions such as the possibility to deploy the service via Docker.

Advantages

The establishment of a central registry for PROV-Templates enables users from different communities to exchange re-usable templates for describing provenance information about common aspects of their data life-cycles. The potential benefit of this approach is twofold, on the one hand saving users valuable time and resources for not always having to start from scratch, providing them with existing blocks to modify and/or build their own provenance infrastructure on. Increasing re-use is on the other hand expected to result in more homogeneous representations of provenance information, potentially allowing to trace the genesis of datasets especially in anticipated cases where data products are passed on between different RIs.

Link to the Demonstrator

ENVRIplus PROV-Template Registry and Expansion Service



- A YouTube video describing the service is accessible at <https://youtu.be/dS58qTfscSM>
- The service is accessible via <https://envri.eu/provenancetemplates>

Contributors

Doron Goldfarb, EAA, doron.goldfarb@umweltbundesamt.at

Dr. Stephan Kindermann, DKRZ, kindermann@dkrz.de

Acknowledgement

This service is running on cloud services provided by national e-infrastructures of the EGI federation.

References

- [1] https://envriplus-provenance.test.fedcloud.eu/static/EnvriProvTemplateCatalog_Manual_v2.pdf
- [2] <https://github.com/EnvriPlus-PROV/ProvTemplateCatalog>
- [3] <https://github.com/EnvriPlus-PROV/EnvriProvTemplates>
- [4] <http://project-dare.eu/>