Janelia Research Campus - Open Science Software Initiative

Proposal: Open Science Software Portfolio

Tiago Ferreira and Konrad Rokicki

BACKGROUND

Several efforts have been made to develop an online portfolio of Janelia software. The internal wiki hosts several lists of software and is conveniently editable by anyone at Janelia, but it is not externally viewable, making it irrelevant for Open Science. The Janelia.org website has a software page[1] maintained by Innovation Management, but is difficult to browse and rarely updated because it has the bottleneck of a single administrator. The newer Open Science index page[2] has more recently been updated and has a better filtering mechanism, but suffers from the same administrative bottleneck and displays outdated and unmaintained software in many cases. The Janelia.org pages are focused on cataloging Janelia's output and not on a developer's software discovery workflow.

A recent effort to inventory active open science software at Janelia was made by SciComp on GitHub[3]. This site is both externally accessible and editable by Janelians via pull-requests, but its simplistic structure does not allow for much information to be conveyed and forces an unintuitive, hierarchical categorization scheme.

PROPOSED WORK

To disseminate and celebrate Janelia's commitment to Open Science Software, we propose the creation of a new portfolio of actively maintained open source software at Janelia. This project adopts a software-driven strategy that falls under Model 3 of the enumerated OSSI project types.

The site will combine the best features of all the previous attempts:

- Public-facing
- Editable by Janelia software developers using the GItHub pull-request workflow
- Supporting full text search with faceted filtering

The site shall be technically well-executed using modern web development techniques such as *responsive web design* (so that the site works on all devices, including mobile phones) and

[1] https://www.janelia.org/tool/software
[2] https://www.janelia.org/open-science/overview
[3] https://github.com/JaneliaSciComp/awesome-janelia-software

continuous integration (to streamline deployments). Importantly, the site itself will be representative of software engineering excellence at Janelia.

The site content will be open sourced on GitHub, and subject to pull-requests, issue tickets, continuous integration, and all the other best practices we espouse for Open Science Software. Any Janelia developer will be able to contribute content or code to the site to improve it over time. In effect, the site will be an Open Science Software playground for experimentation with open source collaboration and software engineering best practices while fulfilling Janelia's needs of Open Science Software dissemination.

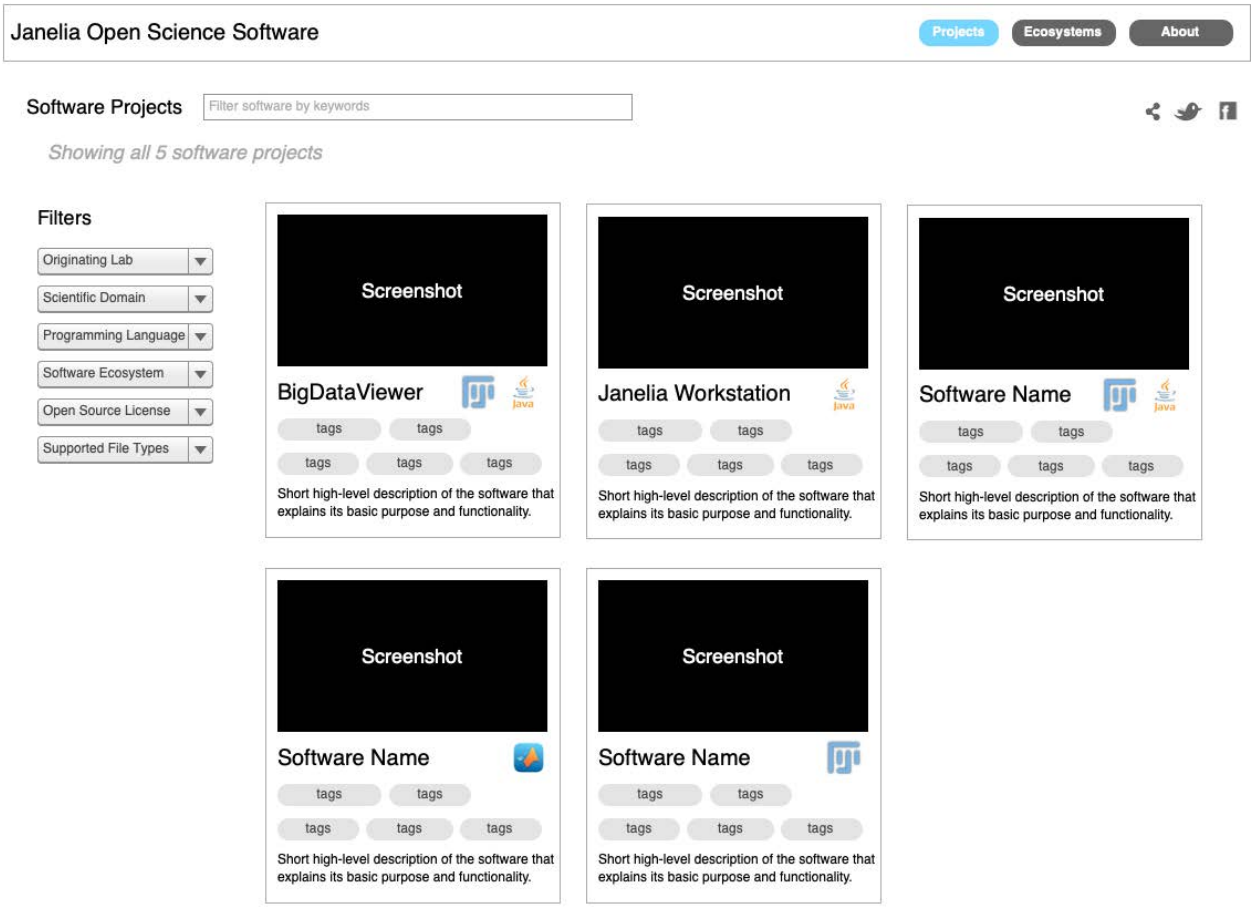

Figure 1: Wireframe example of searching for software projects

We initially envision two top-level sections: projects and ecosystems.

The **projects** page (Figure 1) will allow developers to find individual software projects. filtering on a variety of criteria and full text search of the page contents. Filters could be created on a variety of criteria, for example:

- Originating Lab/Project
- Scientific Domain (Imaging, Electrophysiology, ...)

- Associated Publications
- Model Organism (Fly, Mouse, C.Elegans, ...)
- Software Type (Application, Webapp, Service, ...)
- Programming Language(s) (Java, Python, Julia, MATLAB, ...)
- Software Ecosystem (n5/BDV, Dask, Janelia Workstation, ...)
- Open Source License (BSD-3 Clause, MIT, CC-by-0, ...)
- Supported File Types (n5, Zarr, SWC, NWB, ...)

The list of matching software will be displayed as “baseball cards” with screenshots and high-level descriptions. Clicking a card will take the user to a more detailed page dedicated to that piece of software (Figure 2).

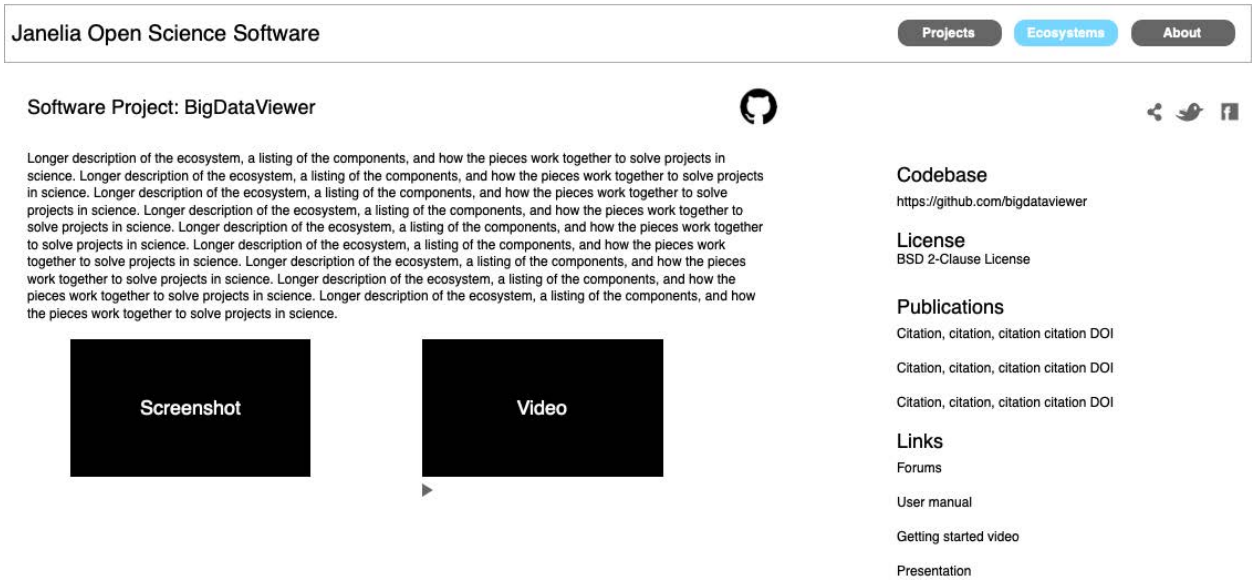

Figure 2: wireframe example of a project detail page

The **ecosystems** page will provide a higher-level view of the software projects and how they can be combined into solutions and workflows. It will provide the same filtering interface as the software projects page. However, after drilling down into an ecosystem (Figure 3), the user will see a list of projects involved in the ecosystem in addition to a description of how to get started with that ecosystem and contribute to it.

The initial site will focus on discovery of software ecosystems and projects, but over time other sections such as blogs could be added to the site.

TECHNICAL APPROACH

To exploit GitHub as a transparent, developer-accessible content management system, we suggest developing the site using a static site generator (e.g. Jekyll, Hugo, etc.) The site

deployment shall be seamlessly integrated with GitHub Pages to automatically make committed content available. Each software project page will be formatted using the Markdown language to maximize accessibility for software developers. Each software tool will be described on its own Markdown page, with “front matter” metadata for indexing and categorization.

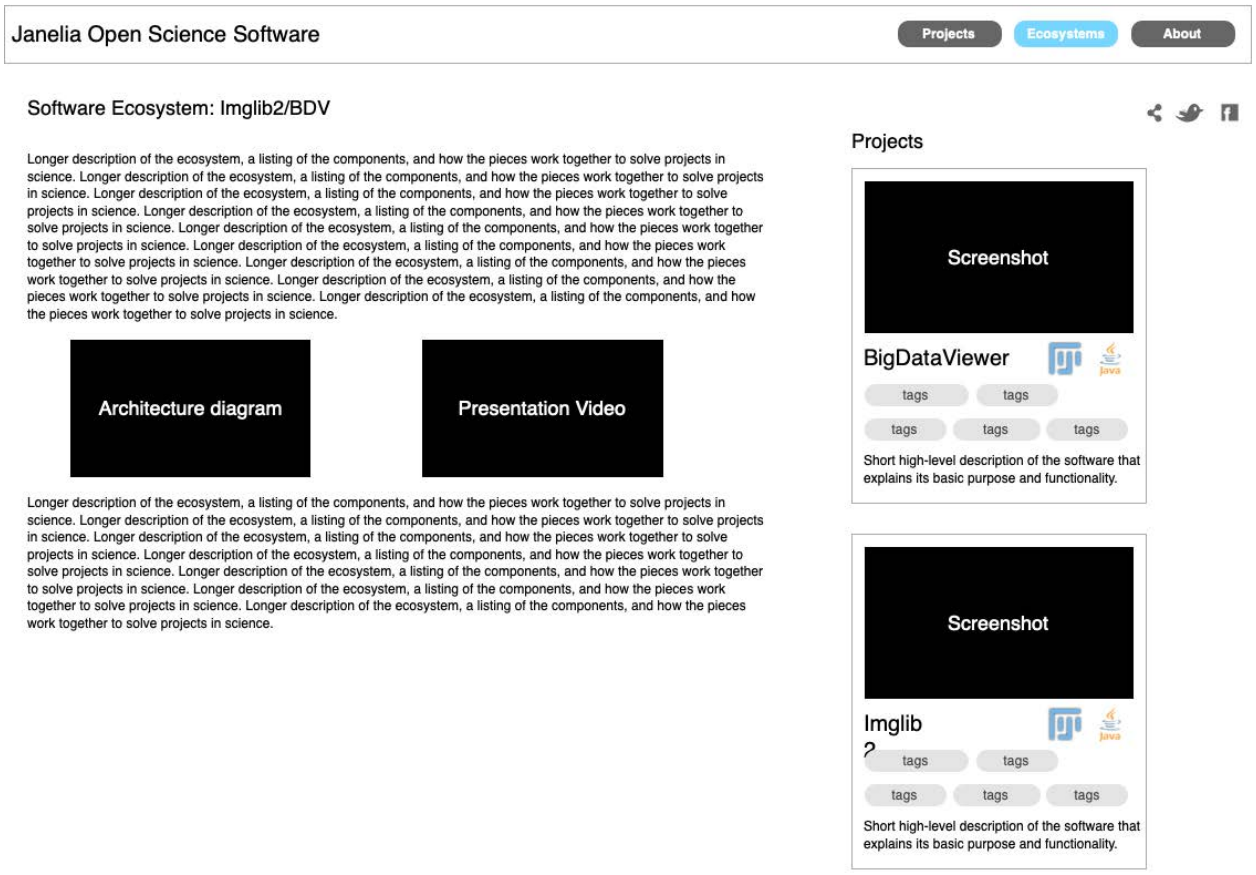

Figure 3: Wireframe example of viewing a description of a software ecosystem

Note that the approach proposed here is a first draft of the requirements and site design. If awarded, the first step in the implementation plan would be to gather feedback on this proposed design from stakeholders including the OSSI committee and Janelians producing computational software, and then revise the design and technical approach as necessary.

EVALUATION OF IMPACT

The impact of this site should be proportional to Janelia’s impact on Open Science Software. As an index and “first stop” for developers looking to reuse scientific software, we expect it to receive traffic from both external and internal users.

The site development itself could have an impact on collaboration within Janelia, if developers across labs, projects, and support teams can be persuaded to contribute to its development.

ESTIMATED EFFORT

We estimate 1.0 FTE month of development effort to complete an initial prototype and populate data for a subset of existing projects. At this point, further development of the site could be crowd-sourced from Janelia developers, but a purely crowd-sourced site is not likely to succeed without a centralized maintenance effort to guide it.

If the initial site is successfully completed and well-liked by Janelians, we recommend funding additional effort for maintaining the site codebase, integrating crowd-sourced content, writing more elaborate content, and further developing the site functionality. As a part of this effort, we could reach out to scientific software developers at Janelia and document their tools for them.