OSI Board Report on ClearlyDefined

Monthly report in preparation for board meeting September 13, 2018

Summary

Our general story is ClearlyDefined is now “officially open for business.” The data, data structure, number of package manager integrations, and software itself are all settled for the time being and we are confident about their quality and resiliency. We still anticipate many changes and new features, but the system can support growth and user feedback well now.

Our focus is now moving more to building the community of users and curators.

Resources

Microsoft continues to donate the blob storage and Azure compute time. Google and Amazon have now also successfully gotten the crawler running in their own cloud environments, and so requests to ClearlyDefined that are queued up could now potentially be run in any one of the three clouds.

The contractors at WebYourMind have been delivering and successful so far in what we’ve asked them to do. We also had a contractor working on our FOSSology integration, which continues.

Technology

We have redesigned some parts of the crawler and service after discovering some issues with the code’s original design, primarily around how we handle casing. Those changes have all been deployed along with some underlying changes that were needed to support the robustness we now expect of the system.

The contractors from WebYourMind have implemented the “full details page” which enables curations on a file level of copyright statements and facets. With this complete the full end-to-end user experience of discovering components on ClearlyDefined, submitting proposed curations for them, and eventually having them merged is now complete.

Community

The community will be our largest focus for the next few months. We want to expand the number of people using and relying on the data as well as those contributing curations to missing or ambiguous data. In order to reveal the places where the software or the data are lacking, we need to grow the number of people interacting with it. Our goals for the coming months are largely around getting more users and growing the community.

Upcoming goals

1. Complete the package manager integrations for PyPi and Go. Harvest and scan the top 100,000 components for npm, maven, NuGet, Git/GitHub, PyPi, and Go. We’re also targeting specific users who we want to start taking dependencies on the data to identify ~10,000 components they’d like us to harvest and scan, send us the list of the components, and we will queue them up.
2. Write a preliminary document outlining a curation steering committee and build the group of people to start participating regularly in it.
3. Design and implement the system with OpenCollective to allow small dollar amount funding of curations by outside contributors.
4. Build a marketing and communications plan for the project to start to grow ClearlyDefined’s social media and community presence.
5. Continue adding features and fixing user issues as they arise with the web UI.