

ClearlyDefined meeting agenda and notes

October 7, 2024

- Nick Vidal (OSI)
- Jeff Luszcz (GitHub)
-

Agenda:

August 5, 2024

- Nick Vidal (OSI)
- Jeff Luszcz (GitHub)
- Lewis Jones (GitHub)
- Lisa Hoong (Bloomberg)
- Tom Bedford (Bloomberg)
- Alyssa Wright (Bloomberg)
- Jeff Wilcox (Microsoft)
- Qing Tomlinson (SAP)
- Brian Duran (SAP)
- Jeff Mendoza (Kusari/GUAC)

Agenda:

- Slides:
<https://docs.google.com/presentation/d/1ZwCv55R8O3fTjBnvTIASoe10YBMuZ3XVQsrBfBj-9k/edit>
- SOSS Fusion talk in October
- Open Compliance Summit talk in October
- Call for Case Studies
- Governance / Charter
 - Steering committee (approved)
 - Outreach committee (approved)

- Jeff Luszcz (GitHub), Alyssa Wright (Bloomberg), Brian Duran (SAP), and Nick Vidal (OSI) have been elected to lead the ClearlyDefined Outreach Committee. Respecting the charter, they'll serve for one year (starting Oct. 7, 2024). The committee structure follows the recommendation of having a maximum of one member per company. The members who attended the Community meeting reached a consensus on having an even number of members (4 currently) given that the Outreach Committee is less likely to require a tiebreaker. The new elected members will select who will serve as the Chair and may invite other active members of the community to be part of the committee. Congratulations to the first Outreach Committee of ClearlyDefined!

July 1, 2024

Attendees:

- Nick Vidal (OSI)
- Lisa Hoong (Bloomberg)
- Jeff Luszcz (GitHub)

Agenda:

- Meeting slides have agenda slide and further meeting notes in the deck:
 - 📄 ClearlyDefined Meeting 2024-07-01
- Call for blog posts / LicenseRef (Nick with help from maintainers)
- Conda support (Basit and Sam)
- Looking for case studies
- Share with other groups (e.g. TODO GRoup and OpenChain)
- Conda Support: Lisa put 700 Conda packages on the queue but hasn't seen them get processes
- New Forum: Start using Discourse for discussions.
- Community outreach
 - Scancode (Philippe)
 - GUAC / Kusari (Ben / Jeff)

- Stacklok (Craig / Stacey)
- Red Hat (Ralph Bean)

June 3, 2024

Attendees:

- Nick Vidal (OSI)
- Qing Tomlinson (SAP)
- E. Lynette Rayle (GitHub)
- Jeff Luszcz (GitHub)
- Brian Duran (SAP)

Agenda:

- Meeting slides have agenda slide and further meeting notes in the deck:
<https://docs.google.com/presentation/d/1y0Q5d229aAqTfHPKKiEjg1PKeF5XPIYU3RPIDxsirXY/edit#slide=id.p>
- Review PRs to close out
- New meeting format
- New website / documentation
- Blog / case studies request for content from GitHub, SAP, etc
 - Brian will work on SAP content
 - Lynette will look at the ORT days content for reuse
 - Nick suggests preso to TodoGroup
 - Jeff Luszcz recommends updated ClearlyDefined preso to OpenChain
- New forum
- SOSS Fusion / CDCD
- PyCon recap

May 6, 2024

Attendees:

- Nick Vidal (OSI)
- Qing Tomlinson (SAP)
- E. Lynette Rayle (GitHub)
- Jeff Luszcz (GitHub)
- Jeff Wilcox (Microsoft)
- Jamie Magee (Microsoft)

- Brian Duran (SAP)

Agenda:

- Introductions
- Communication: move to Discourse (from Discord + mailing lists)
- New documentation platform (Docusaurus)
- Adopt new [charter](#), more open governance model
 - Scope / Benefits
- Collaboration with other groups:
 - ORT (OSS Review Toolkit)
 - TODO Group (GitHub, SAP, Bloomberg case studies)
 - OpenSSF
 - GUAC
 - OpenChain
- Lynette and Qing leading Developer's meeting (every Wednesday 10:30am Eastern)
- Periodic [releases](#) using semver
- Developer roadmap 2024
- Questions from community:
 - Hi, I'm having some issues where clearlydefined isn't harvesting some things I expected it to. I told it to on the UI, but it didn't seem to change anything. An example of one I'd expect is `pypi/pypi/-/pywin32/306`. I also tried using a script and the api to tell it to harvest them, but still nothing. Version 301 is available, but 306 refuses to harvest. Is this the right place to ask for help on this or should I ask elsewhere?
 - The reason is due to lack of source packages present for `pypi/pypi/-/pywin32`. Similar reasons at [Harvesting not working · Issue #561 · clearlydefined/crawler \(github.com\)](#)

April 8, 2024

Attendees:

- Nick Vidal (OSI)
- Qing Tomlinson (SAP)
- Tom Bedford (Bloomberg)

Agenda:

- New member: Cisco (onboarding)
- Communication: move to Discourse (from Discord + mailing lists)
- New documentation platform (Docusaurus)
- Adopt new [charter](#), more open governance model
 - Scope / Benefits
- Collaboration with other groups:
 - ORT (OSS Review Toolkit)
 - TODO Group (GitHub, SAP, Bloomberg case studies)
 - OpenSSF
 - GUAC
 - OpenChain
- Lynette and Qing leading Developer's meeting (every Wednesday 10:30am Eastern)
- Periodic [releases](#) using semver
- Developer roadmap 2024
- Events: Tom to present at OSS NA

March 11, 2024

Attendees:

- Nick Vidal (OSI)
- Tom Bedford (Bloomberg)
- Qing Tomlinson (SAP)
- Brian Duran (SAP)
- Basit Ayantunde (Codethink)
- Sam Bishop (Codethink)

Agenda:

- ClearlyDefined Rate Limiting: is there scope for the rate limits to be raised generally, or whether there is a possibility for the limits to be raised for specific users of the service? (Sam)
- [FOSS Backstage](#) - March 4-5, Berlin (Qing, Nick)
- [ORT Community Days](#) - March 6-7, Berlin (Qing, Nick)
 - ORT Package Curation Data
 - Metadata interpretations (not legally relevant)
 - Factual / companies willing share

- Legal curations
 - Concluded license / companies not willing to share
 - Due mostly because of scancode false positives
- ClearlyDefined [slides](#) (Lynette, Nick)
- Questions from ClearlyDefined presentation:
 - Estimate of time to complete harvest; allows me to choose whether I should come back later or hang around (getting time estimate can be difficult)
 - Notification when a harvest I requested is done (there's a way to poll, see integration test <https://github.com/clearlydefined/operations/blob/55e35641a8d3e6861d167a32ef84b30f7f70a3c3/tools/integration/lib/harvester.js#L55>). The harvest APIs require revision to be provided.
 - SETI@home distributed type processing; main challenge is how to ensure data is not tampered with
 - Question: Should data change if tool versions (e.g. scancode) change, as in, re-run all dependencies using the later version of the tool? Some clear objections to doing this b/c what does it mean if the data changes.
 - Concerns about some of the curated data that was approved; How to improve trust of data
 - ORT has a cli command that creates PRs for CD; Need to talk with Sebastian about the process and problems
 - ORT doesn't want CD data due to old tooling and questionable curations (from Sebastian's perspective)
 - Add trust levels to curations (T=0 for NOASSERTION, T=1 for one human review, T=2 for two agreeing human reviews, T=3 for three agreeing human reviews) is proposed by Marcel from Bosch. Sebastian from ORT thinks there should not have a NOASSERTION after curation (maybe OTHER, or no SPDX license).
 - IPFS to distribute all CD data across many hosted data stores (a lot of work)
 - Look at other places curations are stored (e.g. OSSelot, ORT, etc.)

- Add Purl support (as a search). CD is now influenced by Purl.

February 5, 2024

Attendees:

- Nick Vidal (OSI)
- Brian Duran (SAP)
- Sam Bishop (Codethink)
- E. Lynette Rayle (GitHub)
- Qing Tomlinson (SAP)
- Roman Iakovlev (GitHub)
- Tom Beford (Bloomberg)
- Basit Ayantunde (Codethink)

Agenda:

- Community
 - [ORT Community Days](#) - March 6-7, Berlin (Nick)
 - FOSDEM - Compliance meeting (Tom)
 - Discourse (migrate from Discord and mailing list)
- Operations
 - Complete the Transfer to CD Azure (Lynette/Qing)
 - Move crawler from Microsoft in Azure to ClearlyDefined Azure
 - Ensure that a group of devs have permissions required for DevOps
 - Move deploys to GitHub actions
 - Permission Management (Lynette/Qing)
 - Determine the primary set of maintainers and make sure they are admins on all repos.
 - Determine the set of devs that will have super-admin permissions in Azure
 - Determine the set of devs that have access to key vault
 - What user will connect Azure to GitHub?
 - What user will connect GitHub to Azure?
 - Other places with permissions?
- PRs
 - PR approval process (Qing)

- Dev meetings every Wednesday (Lynette)
- GitHub Actions to deploy dev service app (Lynette)
- Conda Support (Basit)
- Upgrade Licensee / Tests (Yash/Qing)
- LicenseDB Mapping (Lukas/Qing)
- Upgrade Scancode / Beyond SPDX (Lukas)

January 8, 2024

Attendees:

- Nick Vidal (OSI)
- Tom Bedford (Bloomberg)
- Lisa Hoong (Bloomberg)
- Sam Bishop (Codethink)
- Qing Tomlinson (SAP)
- Brian Duran (SAP)
- Lukas Spieß (GitHub)
- Basit Ayantunde (Codethink)

Agenda:

- [Happy New Year!](#)
- [Recap 2023 / Vision 2024](#)
- [ORT Community Days](#) - March 6-7, Berlin
- Website
- Documentation
 - Local harvest
- Operations
 - Enhancement in progress
 - Best practices to merge dependabot PR. Need to have good unit test coverage.
- PRs which need attention
 - Codethink work on Conda support
 - Discussion on Conda coordinates, see [conda crawler implementation by lamarr · Pull Request #532 · clearlydefined/crawler \(github.com\)](#)

December 4, 2023

Attendees:

- Nick Vidal (OSI)
- Daniel Silverstone (Codethink)
- Sam Bishop (Codethink)
- Basit Ayantunde (Codethink)
- E. Lynette Rayle (GitHub)
- Jeff Wilcox (Microsoft)
- Lisa Hoong (Bloomberg)
- Qing Tomlinson (SAP)

Agenda:

- Goals / Prioritized backlog
 - [SAP](#)
 - [GitHub](#)
- Documentation
 - Local harvest
- Operations
- PRs which need attention
 - Codethink work on Conda support - Needs review:
<https://github.com/clearlydefined/crawler/pull/532>
<https://github.com/clearlydefined/service/pull/1019>
<https://github.com/clearlydefined/website/pull/1028>
- Proposed steps for updating release/deploy process
 - move domain name
 - rename master branches to main
 - switch to [semantic versioning](#)
 - Create a semver policy for the project (ie. what is covered by the semver guarantees)
 - create GitHub action to deploy on release
 - create GitHub release with currently deployed code under semver version (e.g. 1.0.0)
 - create release with code in main branches

November 6, 2023

Attendees:

- Nick Vidal (OSI)
- Tom (Bloomberg)

- Jeff (Microsoft)
- Jamie (Microsoft)
- Lisa (Bloomberg)
- Qing (SAP)

Agenda:

- Qing's improvement on summarizing license information
- Qing's PR: <https://github.com/clearlydefined/service/pull/999>
- Jeff: deploy changes, make it easier for new people to contribute
- Nick: discuss roadmap for 2024

October 2, 2023

Attendees:

- Nick Vidal (OSI)
- Roman Iakovlev (GitHub)
- Yaroslav (JetBrains / github.com/yrusskih)
- Tom Bedford (Bloomberg)
- Lisa Hoong (Bloomberg)
- Jeff Wilcox (Microsoft)
- E. Lynette Rayle (GitHub)
- Thomas Steenbergen (ORT/SPDX/TODO)
- Aleksandrs Volodjkins (GitHub)

Agenda:

- Thomas (co-founder of ClearlyDefined) introduces himself and tells about his vision about the future of ClearlyDefined.
- Introduction from Yaroslav from JetBrains.
- Aleksandr from GitHub to present his analysis of ClearlyDefined dataset: 25% NOASSERTION or EMPTY.
- Next meeting: Roadmap and prioritizing of changes that can be made to ClearlyDefined and its ecosystem.

September 5, 2023

Attendees:

- Nick Vidal
- Qing Tomlison (SAP)

- Roman Iakovlev (GitHub)
- Lisa Hoong (Bloomberg)

Agenda:

- Nick: mentioned potential collaboration from JetBrains and Stacklok
- Roman: more coverage, beyond SPDX, not just open source licenses
- Roman: create own local harvest, add other licenses and share it with others as well
- Roman: an API to plug in their own harvester
- Nick to open an issue for community feedback

August 7, 2023

Attendees:

- Nick Vidal

Agenda:

- Lynette has created a [document](#) to explain how to run a ClearlyDefined Harvester locally
- Nick set a [proposal](#) to the Open Infrastructure Fund.
- GitHub added over 17.5 million new package licenses to our database, expanding the license coverage for packages that appear in dependency graph, dependency insights, dependency review, and a repository's software bill of materials (SBOM). Package licenses dictate how a package can be used, making them an essential aspect of compliance when working with open source software.
- Lynette had issues trying to get `docker_dev_env_experiment` to build. Jeff suggested looking at peer dependency conflicts on the [likely older] scancode version being used.

June 5, 2023

Attendees:

- Bloomberg: Alyssa Wright, Tom Bedford, Vineet Aguiar
- SAP: Qing Tomlinson, Brian Duran, Gordon Lee
- GitHub: Ashley Wolf, Roman Iakovlev
- Microsoft: Jeff Wilcox, Manny Martinez
- Kusari/GUAC: Jeff Mendoza

Agenda:

- Introduction from Alyssa Wright, Tom Bedford, and Vineet Aguiar (Bloomberg).
- Introduction from Ashley Wolf and Roman Iakovlev (GitHub).
- Updates from Gordon, Qing Tomlinson, and Brian Duran (SAP): seeking to replace cognitive search (not cost effective).
- Updates from Manny Martinez (Microsoft): subscription cost down, need to focus on log analytics.
- Updates from Jeff Mendoza: Integration between [GUAC](#) and ClearlyDefined.
- Making progress wrt to ClearlyDefined charter from Nick Vidal (OSI) and Jeff Wilcox (Microsoft).

May 1, 2023

- Recent work to support coordinates suggestion and definition stats in database as optimization
- ClearlyDefined: proposal to change charter
 - Limit scope to licensing metadata (collaborate with ecosystem on security, accessibility, etc)
 - Add motivation and open governance section

April 3, 2023

- ClearlyDefined: OSI-driven service next steps
- Transfer the community call to Nick

March 6, 2023

- Manny went over some back-end optimizations, including to the database
- There were concerns file data could disappear as a result of these optimizations
- One potential solution could be to store the files' diff instead of listing every file
- Discussed the topic of "wildcard" / multi-version curation
- Discussed the importance of having a project roadmap

February 6, 2023

- Manny on Dragos' team is working on ClearlyDefined full-time
- Working on authentication for the API
- Qing's been working on ClearlyDefined with Manny
- Dragos is working on changes for the blob storage
 - Testing today
- Some folks would be okay without files as part of the API
- Do we still have a lot of data in 'definitions'?
- Observation of output of production system, seems to be writing to blobstorage and mongodb - does blob storage have a complete definition as well?
- Mongo db for search; definition in blobstorage
- Changing how we store to blobstorage
- Mongodb isn't saving differential; dumping the entire collection there.
- Harvesting getting queued but never completing? What's the best way to track that down?

- What is the best course of action for harvesting not working?
 - May need to deploy more crawlers
 - Contribute compute
- Will the mongodb and blob storage changes be transparent to the end user? Yes, no API changes

September 26, 2022

Attendees:

-

Organization:

-

Technology:

-

September 12, 2022

Attendees:

- Carol
- Qing
- Stefano
- Manny
- Jimmy
- Brian

Organization:

- Community Manager position at OSI now available
 - <https://blog.opensource.org/now-hiring-community-manager-for-clearlydefined-project/>
 - <https://opensource.zohorecruit.com/jobs/Careers/685709000000424749/Community-Manager-for-ClearlyDefined?source=CareerSite>
 - Could be a very high impact position at OSI (full time employee of the organization)
 - Working on partners, adoptions of ClearlyDefined amongst external projects, building funding
 - Any leads on good candidates would be welcome
 - How much experience in open source? - at least some familiarity with open source communities
 - A little bit of knowledge of license environment
 - Some understanding of package managers (or a node or pypi developer?)

Technology:

- PR ready for review. Appreciate any feedback! - Manny will review in the next day or two
- Qing can do some pair programming or set up another meeting to review if needed
 - UI
 - <https://github.com/clearlydefined/website/pull/998>
 - <https://github.com/clearlydefined/website/pull/997>
 - <https://github.com/clearlydefined/website/pull/994>
 - <https://github.com/clearlydefined/website/pull/990>
 - Crawler
 - <https://github.com/clearlydefined/crawler/pull/479>
 - Service
 - <https://github.com/clearlydefined/service/pull/951>

August 15, 2022

Attendees:

- Carol
- Ching
- Jimmy
- Dragos
- Brian
- ceN (Manny)

Organization:

- Introduce Manny
 - One dedicated developer working on ClearlyDefined at Microsoft
 - Michael has left Microsoft
- OSI Community Manager position

Technology:

- Doing some optimizations on the backend
 - Would like to put API behind authentication? (will be staged)
 - OAuth or some other authentication
 - What would be the model for signing up for access? You have to know someone in the community?
 - GitHub username?
 - Further rate limiting?
 - Maybe Azure API management platform?
 - Gradually adding user agents to see who's calling us? - Manny has added some logging (about half the calls have a user agent)
 - Manny could refuse calls that don't have a user agent

- High priority bugs list after API work (prioritize work on those high priority bugs)
 - Discuss at the next community meeting in two weeks and going forward
- Put in tests for scenarios e.g. search on the UI; continuously check if they're working
 - Monitoring on scenario basis
- Ching's PR is in; Michael pushed that PR; on app service restart in deployment is not working properly
 - Manny to followup on PR after
 - <https://github.com/clearlydefined/crawler/pull/475>

June 6, 2022

Attendees:

- Qing
- Carol
- Brian
- Jimmy
- Jeff
- Dragos

Organization:

- Reminder of ClearlyDefined engineer position open: <https://careers.microsoft.com/us/en/job/1338883/Senior-Software-Engineer>

Technology:

- Infrastructure spikes
 - Two potential issues - exceeding quota; response timeout?
 - Just recently increased quota to 2k request per hour
- How is the ClearlyDefined service working for folks?
 - SAP - some compliance processes that are very dependent on the data
 - Qing working on some optimization, they might help?
 - SAP is working with a local university and some of their data science students; want to look into the NOASSERTION cases and see if there's improvements to be made there
 - Haven't heard complaints internally
 - Consider splitting the PRs into smaller PRs to deploy more safely?
 - Maybe split the PR - but this may lose the conversation on one of the PRs?
 - Caching change - has to change each ClearlyDefined subclass
 - Queueing - scoped queueing (that's small)
 - Added unit tests for each change
 - PR <https://github.com/clearlydefined/crawler/pull/467>
 - Turnaround time for harvest after queueing a request - Dragos to follow-up
 - What exactly is the rate limiting?
 - <https://api.clearlydefined.io/api-docs/>

- Brian to update docs on the rate limiting
- Beginning of July planing more improvements for ClearlyDefined

May 23, 2022

Attendees:

- Carol
- Jeff

Organization:

- ClearlyDefined engineer position open:
<https://careers.microsoft.com/us/en/job/1338883/Senior-Software-Engineer>

Technology:

-

April 11, 2022

Attendees:

- Qing
- Carol
- Brian
- Jimmy

Organization:

- Moving curation meeting to “as needed” for now.
 - #curation channel on Discord always open
 - We can schedule a meeting whenever needed

Technology:

- Discuss pull requests to be reviewed (Qing) - both are ready for review
<https://github.com/clearlydefined/website/pull/971>
<https://github.com/clearlydefined/crawler/pull/467>
 - Carol to get someone to review and address
 - Qing is happy to meet with the reviewer if needed
- Another PR from SAP that needs review -
 - <https://github.com/clearlydefined/service/pull/899>
 - <https://github.com/clearlydefined/crawler/pull/452>
- Discuss any pressing issues to be worked on (Qing)
 - Website issues to be addressed
 - Column filters for browsing files (when you filter it only shows files within the current view). For finding a specific file you used to be able to click on the filter and select a license; that filter feature only shows the files in the

current view now. Intended functionality?

Issue submitted: <https://github.com/clearlydefined/website/issues/973>

March 14, 2022

Attendees:

- Carol
- Jeff
- Swadhin
- Brian
- Jimmy

Organization:

- February OSI blog post draft:
<https://docs.google.com/document/d/1crj2G6oD7KbUWGjr79tgwrXiSVvONv0y5buS-aolUFM/edit#heading=h.t4ctj4lciwd4>

Technology:

- Discuss <https://github.com/clearlydefined/website/issues/953> [Brian]
 - Affecting usability and drill-down capability for users
 - Qing took a look
 - <https://github.com/clearlydefined/website/pull/966> ready for review by jeffwilcox@outlook.com
 - Maybe reviews some time this week?
- Swadhin running into some curation issues - re: node components from npm - contact Ariel and Candice offline to discuss further

February 14, 2022

Attendees:

- Carol
- Jeff
- Swadhin
- Ching
- Brian
- Nell

Organization:

- Team update
 - Nell will be moving to a different team at Microsoft
 - Contacts for ClearlyDefined will be Carol and Jeff Wilcox

- Nell will be around but won't be part of ClearlyDefined
 - Likely for Microsoft to hire another engineer to work on ClearlyDefined
 - Transition starts next week
- There is a curation meeting today; agenda here:
https://docs.google.com/document/d/11fcwtu6w9Qqt61PiBbuyQzVQDWR7v17YakmPnw_uD28/edit

Technology:

- New Gradle Plugin Support (thanks to Qing!)
 - Now live in production!
 - Awesome contribution :-)
- Pushing the deploy button on the new UI (thanks to Andrew and Kyle, our fantastic contractors)
 - Getting pushed after this meeting; largely complete
 - Continue to let Nell know if you see anything after the launch
 - Exciting new launch as well!
- Issues Ching was verifying; two were fixed; some are still issues
 - Ching will add comments to the issue so we have the history

January 31, 2022

Attendees:

- Brian D.
- Jimmy Kung
- Qing Tomlinson
- Nell Shamrell-Harrington
- Swadhin

Organization:

- Swadhin - also from SAP - helping with some of the alignment on ClearlyDefined internally, has done some curations (welcome!)

Technology:

- Revisions to coordinate format for Go components
 - Michael will write up a proposal for comment on GitHub
- UI/UX update
 - Will be going up on dev.clearlydefined.io this week for review by community and curators

November 8, 2021

Attendees:

- Carol
- Nell
- Brian

- Jimmy

Organization:

- OSI requested we no longer provide monthly reports and switch to monthly blog posts; we'll change over to that format on Dec 1

Technology:

- UI update
 - Hoping to launch around week of Nov 22; if that's not possible it will certainly be done by end of the year
 - Looking for input on the logo - use the same we've historically had or switch to a new look?
 - Discord is the best place for feedback on the designs
 - <https://marvelapp.com/prototype/cj36a3g/screen/82530647>
- Go update
 - Go support has launched
 - There is a bug present but it mostly works
 - Bug is for components that have slashes in coordinates
 - Nell working on fixing bug this week

October 11, 2021

Attendees:

- Nell
- Carol

Organization:

- SAP/ClearlyDefined webinar upcoming - Tuesday, October 18 at 8am PDT/5pm CEST
 - [Slides](#)
 - [Sign up to attend](#)
- UI/UX Update

Technology:

-

September 27, 2021

Attendees:

- Nell Shamrell-Harrington

Organization:

-

Technology:

- Infrastructure issues

August 30, 2021

Attendees:

- Carol
- Nell
- Brian
- Jimmy
- Raj

Organization:

- Go over last month's OSI board report:
<https://docs.google.com/document/d/1G-caLcHzWp0BBiXsmyvsl56WZos3Ew04ayJ5E8l68Mo/edit>
- Share link for all historical board reports:
https://drive.google.com/drive/folders/1fmcYf_3AARq2w-FqCBGdkV160UnXulsU
- There will be a SAP/ClearlyDefined presentation (online webinar) in October; details will be forthcoming if folks would like to attend.

Technology:

- UI/UX mockup and work: Are there going to be any performance enhancements?
 - Probably not; more focused on being more intuitive and clear

August 2, 2021

Attendees:

- Nell
- Kevin
- Carol
- meehan

Organization:

-

Technology:

- New ClearlyDefined UI
 - Prototype setup will be shared wider community for folks to click on
- Go implementation is getting iterated on; Nell working on hooking up crawler and locally harvesting

July 19, 2021

Attendees:

- Brian
- Nell
- Carol
- Raj
- gkylee

Organization:

- OSI monthly board report for July:
 - https://docs.google.com/document/d/1OqHTTRyZl6v_EveZ-vO7PbxhjhyqYLT5wlbaJTghLUY/edit
 - Heavy focus on technical infrastructure
 - Partial harvesting issue is mostly fixed; a few still too big and Nell has some ideas on fixing those
 - UI redesign continuing; designer may be able to attend the next meeting
 - Google has shut down their crawlers; AWS having some trouble and working with them to resolve

Technology:

- Go coordinate design <https://github.com/clearlydefined/service/pull/862>
 - No central package manager repository
 - Pulls component directly from source
 - Google has a proxy that resolves source
 - Proxy.golang.org will be our first provider
 - Resolved the slashes issue by replacing with dots
 - Design of Go coordinates based on Maven work
 - Initially only supporting Go modules

June 21, 2021

Attendees:

- Carol
- Nell
- Brian
- Jimmy

Organization:

- Review last month's board report:
https://docs.google.com/document/d/1UCXGLI_tmIE0y-gGCv-Pk34uqdu1M8P2fXRLsk9IZg0/edit
 - Nell to follow up with OSI about posting them more publicly
- No meeting in two weeks for July 5 holiday

Technology:

- GitLab is now a source for git components!
- Still working on troubleshooting partially harvested components
- UI/UX contractor is conducting interviews with community; will be presenting design in a month to get feedback
- SAP is tracking components and working to see if there's anything to report on partially harvested or "sat idle for a while"
-

June 7, 2021

Attendees:

-

Organization:

- "Scaling up" curation efforts within organizations (does anyone have tips or recommendations?) [Brian]
 - What can be automated?
 - How does Component governance use the CD curations api?
 - What logic triggers NOASSERTION?
 - Nell will investigate and bring back information at either the next community meeting or the one after
 - Offered help on partially harvested
- Review last month's progress report to the OSI board [Nell]
 - Are these publicly available? [Brian]

Technology:

-

May 24, 2021

Attendees:

- Carol
- Nell
- Kevin

- Jeff

Organization:

- Curation meeting today after this one:
https://docs.google.com/document/d/11fcwtu6w9Qqt61PiBbuyQzVQDWR7v17YakmPnw_uD28/edit#heading=h.kpq0ap3ril3

Technology:

- Adding gitlab as a source for git components
 - Proposal: <https://github.com/clearlydefined/service/issues/842>
 - Based on proposal for adding maven
 - No objections from the meeting
- Reworking how we verify SPDX license expressions in the clearlydefined/spdx npm packages
 - <https://github.com/clearlydefined/spdx/issues/21>
 - Can't use 2 "or"s (3 licenses)
 - What should we do to fix it? Proposals on issue above.
 - Community member has offered to show this to the SPDX tech team
 - May not be the SPDX normalization itself? Maybe the parenthesis insertion for term ordering (no tests for that)?
 - Need some test cases to reveal the (potential) problems with ANDs
 - Nell to look into further test cases
- Email on mailing list from the weekend
 - Email the person individually and let them know it's a public mailing list
 - Carol to follow-up

April 12, 2021

Attendees:

- Nell
- RoadLessTraveled (Steve)
- Jimmy
- Brian
- Raj
- Jeff

Organization:

- We will have a curation meeting today
https://docs.google.com/document/d/11fcwtu6w9Qqt61PiBbuyQzVQDWR7v17YakmPnw_uD28/edit#heading=h.kpq0ap3ril3

Technology:

- Criteria for a source being added to ClearlyDefined (Steve)
 - <https://github.com/clearlydefined/service/issues/819>
 - MSFT is very reliant on information in ClearlyDefined
 - Bunch of teams at MSFT that produce software for Android, reliant on packages on maven.google.com (not published to MavenCentral)
 - Proposed six criteria
 - Discoverability
 - Primary Source
 - Reputability
 - Security
 - Automation
 - Relationship
 - Consensus - looks good
 - Maven.google.com - in code review
 - Working on establishing contact with Google OSPO
- <https://github.com/clearlydefined/curated-data/issues/10303>
 - Scores for binaries are often not very high because license info is stripped out
 - Problem for orgs that use standard like “you can use this component if it is above this score”
 - Organizations - you could fall back to the source if you wish to and look at the score of the source
 - Judgement needed - does the source url include all of the source or too much of the source? Is there a mapping of a package to a repo? Has caused initial reluctance
 - Should we make it available and people can choose whether to use or not
 - Maven binaries - lack of license info in package - any way we could improve would be helpful
 - We have most of the data
 - Moot if we don't have Source url or easily mapped source repo
 - Do we calculate two sources and show score for whichever is higher?
 - Brain - automation - looking at scoring algorithm. Documentation around effective score vs. tool score is lacking
 - Should update documentation - might be a source of confusion
 - Going forward with issue - how would they do it? If we didn't do automation, but if they were to use binary and source scores, how would they combine them?

March 29, 2021

Attendees:

- Carol
- RoadLessTraveled (Steve)
- Brian
- Jimmy

- Erika
- Raj

Organization:

- We will have a curation meeting today:
https://docs.google.com/document/d/11fcwtu6w9Qqt61PiBbuyQzVQDWR7v17YakmPnw_uD28/edit

Technology:

- What are the standards for adding a source to ClearlyDefined? (i.e. should we add <https://maven.google.com/web/index.html> as a source for some Android packages that are not available on Maven Central?)
 - <https://github.com/clearlydefined/service/pull/601>
 - <https://github.com/clearlydefined/crawler/pull/332>
 - How do we determine what site is the primary source for a package type?
 - Should we use non-public APIs to get packages for ClearlyDefined?
 - Google is not the only company that makes private-ish Maven repositories
 - Steve will define criteria upon which we determine if a package source is trustable and good and why the GOogle
 - Steve to submit PRs for actual work for community to provide technical comments on

March 15, 2021

Attendees:

- Nell
- Michael
- Kevin
- Carol
- Brian
- Disulliv
- Jimmy
- Erika

Organization:

- We will be hosting a curation meeting after this one:
https://docs.google.com/document/d/11fcwtu6w9Qqt61PiBbuyQzVQDWR7v17YakmPnw_uD28/edit#heading=h.kpq0ap3ril3

Technology:

- Discuss multi-version curation PR: <https://github.com/clearlydefined/service/pull/810>
 - In curation submission it will look for different versions of the same component with similar metadata or license by looking at the hash
 - This will help curate multiple versions of components at the same time

- Matching versions will add amendment to PR along with the description of why the versions match
- Will also recompute the definition
- Michael will take a screencast of a demo and post in the Discord channel
- Will also present at the curation meeting today
- Nell will do a code review later today
- Please comment on issue now; with goal of merging within the next couple weeks
- Discuss these issues
 - Website
 - <https://github.com/clearlydefined/website/issues/896>
 - When you look at a component on browse page, some of the scoring elements seem to be different
 - Some missing elements
 - Screenshots in the issue
 - Seems to be inconsistent between views
 - Service
 - <https://github.com/clearlydefined/service/issues/776>
 - Discussed at community meeting; enhancement
 - <https://github.com/clearlydefined/service/issues/743>
 - Investigate if authors who don't use tags affects this
 - Potentially add tags as something separate
 - Crawler
 - <https://github.com/clearlydefined/crawler/issues/348>
 - We currently support Deb packages
 - Part of providing more support and data on containers
 - Will table it for a little while and continue a larger discussion around container scanning

March 1, 2021

Attendees:

- Nell
- Kevin
- Carol
- Jeff
- Brian
- Jimmy
- kesmith
- Raj
- SheriffManzoor

Organization:

- Curation meeting cancelled today

Technology:

- Discuss these enhancement requests
 - Website
 - <https://github.com/clearlydefined/website/issues/889>
 - UI can only update GitHub source URLs; can't update the other URLs
 - Should we let these be curated?
 - Fixed the errors in the definitions already
 - Maven issue was a systematic issue that's been fixed; we don't necessarily see a need to update these URLs at this time; we can revisit
 - Service
 - <https://github.com/clearlydefined/service/issues/750>
 - Reason for keeping them as blobs is because it's cheap and fast to serve up
 - Might be worth investigating if there are any Azure blob slowness?
 - We might examine generalizing in the future; not pursuing at this time
 - Storage is pluggable if you write an adapter for another provider
 - <https://github.com/clearlydefined/service/issues/753>
 - Some users might not trust curations?
 - Might want to understand from the users what the use case is here
 - <https://github.com/clearlydefined/service/issues/744>
 - Do we need an API for this?
 - Does the search API have this?
 - May be that we don't have it documented?
 - Check if this is present in the API and if we could surface it better
 - Crawler
 - <https://github.com/clearlydefined/crawler/issues/384>
 - Upgrade the Azure blob storage handler?
 - Reasonable to look at upgrade; quite a few breaking change versions ahead
 - May not be very high priority but will get looked into
- Cloudflare errors
 - Any timeouts still?
 - Found the API call it was related to; search related for too many lookups in database
 - No reports of timeouts lately

February 1, 2021

Attendees:

- Carol
- Kevin
- Nell
- Kesmith
- Jeff McAffer
- Brian
- Justin
- Jimmy
- Erika

Organization:

- Feb 15 is a US holiday - cancel this meeting?
 - Cancel Feb 15
 - Consider having this meeting on days that aren't Mondays
- Any topics for the curation steering committee meeting today?
 - https://docs.google.com/document/d/11fcwtu6w9Qqt61PiBbuyQzVQDWR7v17YakmPnw_uD28/edit
 - If not, should we consider cancelling this meeting going forward?
 - Cancel today's meeting but we'll keep it on the calendar for now at biweekly

Technology:

- Discuss whether to implement these enhancements
 - <https://github.com/clearlydefined/website/issues/873>
 - Where to put it?
 - Have it part of the curation page
 - Would there be a link to the curation guide?
 - Need to surface curation guide more
 - <https://github.com/clearlydefined/website/issues/613>
 - No objections
 - Implementation needs to be considered further
 - <https://github.com/clearlydefined/service/issues/660>
 - No "monetary reason" to do this as this point
 - Local installations of Mongo _do_ seem to be working
 - No urgent reason to switch to a relational database
 - Funders might rather see funds go towards working on the project; not for infrastructure if avoidable
 - Funds not a problem for now.
 - Just close - no objections
 - <https://github.com/clearlydefined/service/issues/697>
 - Should ClearlyDefined stop only accepting SPDX licenses?

- License ref proposal - normalize license text and be able to identify it as SPDX
- Consider implementing license ref functionality
- Close issue in service and keep curated data issue open
- <https://github.com/python/peps/blob/master/pep-0639.rst>
- Keep the issues of 1) how to literally indicate the license vs 2) the license itself being clear
- <https://github.com/clearlydefined/crawler/issues/418>
 - Need to update the tools
 - Do we need to re-run them on everything?
 - How big a difference will it make?
 - Talk to the Google and Amazon folks running our crawlers to make them aware when we're re-running things
 -
- [JeffMendoza] [Design proposal: Multi-version Curation](#)
 - Request having him moving it to a GitHub issue
 - Have everyone to read and comment
 - Jeff to attend next meeting to discuss

January 4, 2021

Attendees:

- Carol
- Kevin
- Jeff
- Nell
- Brian

Organization:

- [Review annual report for OSI](#) (Carol)

Technology:

- Status and next steps
 - Nell will send a draft roadmap
 - On list for future:
 - Pending harvest indicator
 - Mirroring
 - Go support
- Nell has created a Docker-ized development environment (will include link in General channel in Discord)

October 26, 2020

Attendees:

- Carol
- Kevin
- Jeff
- Sharon
- Wwahammy
- Nell
- Brian
- Jimmy
- Raj
- Laura

Organization:

- Introduce Nell, new Microsoft ClearlyDefined engineer
- Export of files and licenses
 - Can currently one discover the licenses one by one?
 - No “export” option (csv or xls?)
 - Definitions API
 - No UI affordance for it
 - Anything you can do in the UI you can do via the API
 - User trying to use ClearlyDefined and export files
 - Search for all files for a particular license
 - [api.clearlydefined.io/definitions/\[coordinates\]](https://api.clearlydefined.io/definitions/[coordinates])
- Curation meeting after this one
 - Agenda:
 - https://docs.google.com/document/d/11fcwtu6w9Qqt61PiBbuyQzVQDWR7v17YakmPnw_uD28/edit

Technology:

- 524's
 - When running the Notice file generation - about 25% of the time getting 524 errors
 - Working in chunks of 250 packages at a time
 - Open an issue - include sample list of components
- Notice me reintegration
 - Just a matter of time and resources to do the reintegration
 - Likely very little customization that will need to be removed?
 - <https://github.com/houdiniproject/noticeme>
-

September 28, 2020

Attendees:

- Carol

- Sharon
- Michael
- Jimmy
- Laura

Organization:

- Hired Microsoft engineer to work on ClearlyDefined
- Curation meeting after this one
- Talk on ClearlyDefined for State of the Source has been posted:
https://www.youtube.com/watch?v=kizaL_ZNRFM&t=1296s

Technology:

- Reports from Discord the API is slow and folks are running into errors
 - Curation API - is returning 524s because of Cloudflare
 - # of components in the Notice file may be affecting this
 - Slowness report may need to be investigated
 - Michael can investigate further
- Update on mirroring discussion
 - May need to be followed up on in a couple weeks

September 14, 2020

Attendees:

- Carol
- Laura
- Jeff M.
- Brian
- Jimmy
- Michael
- Sharon

Organization:

- State of the Source presentation
- Curation meeting after this one
 - Brief summary of curations for this meeting's agenda
- Case study of Laura and Microsoft's work on getting curations from other parts of Microsoft - TODO
 - Trends in communities or ecosystems?
 - Lessons learned?

Technology:

- Road map [jeff]
 - Talked about doing a road map
 - Call for participation in road map

- Carol to collect issues for the road map
 - See <https://github.com/clearlydefined/clearlydefined/issues/130>
 - Cloudflare 503's?
- State of hiring [jeff]
 - Any further along on the process?
 - Still working on hiring and have some candidates in the pipeline, no good news yet
- Michael planning to spend some time on some operations work for the service

August 3, 2020

Attendees:

- Carol
- Wwahammy
- Philippe
- Michael Zeng
- SheriffManzoor
- Laura M
- Jeff McAffer
- Kevin Fleming

Organization:

- Debrief from the discussion with the OSI sponsors on incubation exit
 - <https://docs.google.com/document/d/1G5UDyC6e5TbgsKYDzgiWdkknbZKkTFxKyK0tUbZZLFE/edit>
 - Have a session on follow-up for TODOs from meeting
 - Schedule a call dedicated to follow-up
- Funding allocation and usage
 - Funding at OSI available for use
 - Roadmap for work
 - Mirroring obviously important for the community and top priority
 - SPDX 3.0 and ClearlyDefined documents
 - Triaging for bugs after that and smaller issues that need attention
 - Support for relational DB over MongoDB
- New contributors
 - Michael Zeng - taking some responsibilities on in ClearlyDefined
- Curation meeting after this one - https://docs.google.com/document/d/11fcwtu6w9Qqt61PiBbuyQzVQDWR7v17YakmPnw_uD28/edit

Technology:

- Houdini noticeme presentation
 - Houdini is with Software Freedom Conservancy

- <https://github.com/houdiniproject/noticeme>
- Can be used for CI
- Features for included.json file
 - Code that was vendored or pasted in code but not a dependency; describe in json file and it gets included in notice file
- Can create difference notice filenames
- Want to add a warning for situations where package couldn't be added because there isn't enough info on ClearlyDefined
- Would be nice to include packages that aren't on npm registry
- Don't have handling for git links that aren't packages
- Development planning/roadmap
 - Start an issue and add major topics we'd like to discuss
 - Roadmap issue in GitHub to begin discussing important topics with links to issues
- Further discussion on mirroring solution
 - Two possible solutions -
 - "Just a copy of the data"
 - Mirror of ClearlyDefined the service itself
 - Technique by which data goes from clearlydefined.io to any given mirror
 - Multiple potential solutions to this
 - Page blob of changes that happened in a day; make those available to the mirror
 - Minimum desired latency?
 - Currently finding new packages within about a day
 - Latency is likely a concern for curation scenarios where a curation is merged and the user wants the updated definition
- Issue to investigate Cloudflare

July 20, 2020

Attendees:

- Carol
- Laura
- Philippe
- Sharon
- Kevin
- Jeff
- Brian
- Erika

Organization

- ClearlyDefined/OSI meeting re: moving out of incubator status on Wednesday, July 29 at 9am Pacific

- Please contact Carol if you want details on meeting
- [Curation meeting](#) after this one
- Laura: Update on PR curation for missing license cases pilot
 - Started rolling out to pilot groups
 - Continues to roll out; all PRs created have been merged
 - Continuing to roll it out
 - Will update in curation meeting as well
 - How-tos or case studies as an outcome from this for the community would be good
 - Might re-word “automation”
- Philippe: Registering the ClearlyDefined trademark at the OSI
 - To discuss in the incubator status meeting

Technology:

- Remove dependency on CosmosDB/MongoDB for a relational DB
<https://github.com/clearlydefined/service/issues/660> (Philippe)
 - Consider use cases here and consider solutions to those
 - Running locally
 - Mirroring
 - Someone to write a Postgres store?
- Store JSON files in a database, not in a filesystem tree or some "blob" store
<https://github.com/clearlydefined/service/issues/750> (Philippe)
- Cloudflare timeouts are reproducible <https://github.com/clearlydefined/service/issues/749> (Philippe)
 - Add returns of 400s for queries that aren't supported
 - Look into using Cloudflare for fewer things or not at all; may not be serving us the way we intended
 - More detailed caching discussion?

July 6, 2020

Attendees:

- Carol
- Lamac
- Kevin
- Zack
- Dennis
- Sharon
- SheriffManzoor
- Philippe

Organization:

- Philippe: What about moving to a neutral org that would control and own the data and infrastructure?
 - Folks aren't happy with the infrastructure being under Microsoft
 - Carol to set up meeting with OSI re: moving beyond incubator status; share meeting details with the group
- Laura (Lamac): Update on PR curation for missing license cases (previous presentation context-
<https://docs.google.com/document/d/1y8NP9o271BOfQn6Udz6rfRG--Q-9XhUUZigjcNHqOGQ/edit?usp=sharing>)
 - Looking to turn this on soon; working with a small number of teams and starting to submit curations
 - Treat this as an experiment
 - Curations will look like they're from a bot, not a person
 - People will be encouraged to follow-up on curations if the curators have questions
- Any topics for the curation meeting today? We'll cancel otherwise.
- Philippe: What about vulnerabilities and vulnerablecode: this is atopic to bring up for next call.

Technology:

- Philippe: Discuss next step on data mirroring with ClearCode
 - The presentation from last meeting is at
[https://github.com/nexB/clearcode-toolkit/blob/master/docs/ClearCodeIntroduction-June 2020.pdf](https://github.com/nexB/clearcode-toolkit/blob/master/docs/ClearCodeIntroduction-June%202020.pdf)
 - The ticket is at <https://github.com/clearlydefined/service/issues/386>
 - Software Heritage work to see if they can seed db of the data

June 22, 2020

Attendees:

- Carol
- Kevin
- Dennis
- Roman
- Lamac
- Brian
- Sharon
- Philippe
- zack

Organization:

- Question from Philippe: Are the project assets owned by the OSI?

- [Curation meeting](#) today after this one

Technology:

- ClearCode: Philippe has developed a reasonably robust tool that can sync and keep syncing the whole ClearlyDefined dataset that he would like to present in the weekly call.
 - Philippe - can you provide a link to the slides after the meeting?
 - Tool to mirror clearlydefined definitions
 - <https://github.com/nexb/clearcode-toolkit>

June 8, 2020

Attendees:

- Carol
- Eyar
- Roman
- Shimon
- Sharon
- Lamac
- Brian
- Jeff
- Jeff
- Erika
- Kevin
- Raj
- Dennis

Organization:

- [Curation steering committee meeting](#) after this one

Technology:

- Mirroring options and potential implementation options discussion
 - <https://github.com/clearlydefined/service/issues/386>
 - Couple possible approaches:
 - replicate database; very simple database “copy”
 - Full replica; secondary or duplicate service that runs in your infrastructure and mirrors the data in a “read-only API”
 - Datree - recently integrated clearlydefined
 - Licensing information, but not for specific packages; have more complex queries
 - Use case in issue linked above
 - Currently just calling API within rate limit; have processes for keeping up-to-date and making sure no data is lost
 - Use this information combined with their own data

- Running very complex queries - e.g. “all the packages that have a specific license and other metadata”
- All packages with GPL v2 license, from node.js, and have a specific name
- Bloomberg -
 - Mirror of some portion of data to have internal service to do analysis against the data
 - Can’t use the API because of confidentiality; can’t send all the queries because it discloses more information than they can
 - Very important to know if data about package changes
 - Polling for changes to data would be inefficient; need “something like an RSS feed”
 - Won’t be running service internally; API is not a concern
- Need to discuss syncing mechanism, syncing frequency
- Latency - how fresh does the data need to be? Minutes, hours, days?
 - 24 hours - datree
 - 24 hours? A week? Bloomberg?
- Backend storage - service uses Mongo database and blob storage
 - File-based or database backend
 - Frequency and sharing/updating of data
 - Update/tarball/zip of all definitions that have changed
 - More of an event-based system including versioning
 - Replicate data and have a process for that; something on top that has read-only replica
 - Daily update and reconcile point
- Need a more concrete design and implementation solution
- Presentation and discussion: anonymous PR curations for missing license cases, from Microsoft - Laura
 - <https://docs.google.com/document/d/1y8NP9o271BOfQn6Udz6rfRG--Q-9XhUUZigjcNHqOGQ/edit?usp=sharing>

April 27, 2020

Attendees:

- Carol
- Leslie
- Brian
- Dennis
- Raj
- Erika
- Sharon
- Philippe

Organization:

- Status on ClearlyDefined Wish List
- OnBoarding Process
- [Curation steering committee](#) meeting today - please add agenda items if you have them

Technology:

- New Github Repo to host CD utilities for Python:
 - <https://github.com/clearlydefined/clearlydefined-python>

April 13, 2020

Attendees:

- Leslie
- Carol
- Brian
- Kevin
- Karan
- Raj
- Erika

Organization:

- ClearlyDefined Wish List ~ Discussion and Next Steps: Leslie
- https://docs.google.com/document/d/1bRYO2wgq4fb_jdNTCUzTT-JWGTFzM0ZoP8sM9prxfzs/edit
- Curation meeting today after this one - any agenda items?
 - https://docs.google.com/document/d/11fcwtu6w9Qqt61PiBbuyQzVQDWR7v17YakmPnw_uD28/edit#

Technology:

- Scoring topic - Kevin
 - Popup for scores has both tools and curation link to scoring formula

March 30, 2020

Attendees:

- Philippe
- Carol
- Leslie
- Sharon
- Dennis
- Brian
- Raj

- Jeff
- Jeff
- Erika

Organization:

- White Paper: ClearlyDefined 2020 Organizational Design: Leslie
<https://docs.google.com/document/d/1KlhzgoRkxnQIsZbWr8rVg7XezMfe7xZjtdpyMZmAigo/edit>
- Please circulate to your networks that there's a gsoc project to get SPDX to interface with clearlydefined. we're looking for student candidates and potential mentors from the community if you know anyone.
https://wiki.spdx.org/view/GSOC/GSOC_ProjectIdeas#ClearlyDefined_exporting_and_importing_SPDX_documents
- Incubation status @ OSI: what's next? Trademark registration @ OSI?
- Any topics for the curation steering committee meeting?
 - No topics - meeting cancelled for this week

Technology:

-

March 16, 2020

Attendees:

- Carol
- Leslie
- Dennis
- Brian
- Jeff
- Raj
- Jeff
- Erika
- Philippe

Organization:

- White Paper: ClearlyDefined Organizational Design 2020
 - <https://docs.google.com/document/d/1KlhzgoRkxnQIsZbWr8rVg7XezMfe7xZjtdpyMZmAigo/edit>
 - Review paper offline and talk more in depth in the coming weeks
- Curation steering committee meeting in an hour -
<https://docs.google.com/document/d/11fcwtu6w9Qqt61PiBbuyQzVQDWR7v17YakmPnwuD28/edit#>

Technology:

- Update on operations [Jeff Mendoza]

- <https://github.com/clearlydefined/operations>
 - <https://github.com/clearlydefined/service/blob/master/docs/rampup.md>
 - Ramping folks up with more of the day-to-day service work
 - Cataloged all the pieces in Azure Portal for running the service
 - Went through all existing errors and made issues for queries that produce errors
 - Soon moving some issues to repos they represent
- GSoC - student working on ClearlyDefined synchronization
 - Jeff Mendoza volunteering to mentor

March 2, 2020

Attendees:

- Carol
- Sharon
- Brian
- Philippe
- Antonino
- Dennis
- Jeff
- Jeff
- Jake

Organization:

- Microsoft is hiring for a full-time engineer to work on ClearlyDefined: <https://careers.microsoft.com/us/en/job/800721/Principal-Software-Engineer>
- New members from Jeff's team(at Microsoft) will be joining the project/team; will be working on bug fixes and other things
- GSoC project idea <https://github.com/nexB/aboutcode/wiki/Project-Ideas-ClearlyDefined-Sync-Tool>
 - AboutCode to co-host project for ClearlyDefined?
 - Can help as a mentor or informally fielding questions from students
 - Tool for synchronization and helping to make updates to service to help with ClearlyDefined
- Contractors not working on ClearlyDefined currently due to financial resources at OSI
 - We'd love more resources if you know of folks who want to donate!
- ClearlyDefined curation call after this one at 9am Pacific

Technology:

- ClearlySecured: presentation and discussion on vulnerability databases
 - Antonino will share slides to mailing list; we'll reconvene in two weeks to discuss more in depth
 - Sharon/Eclipse interested in participating in the project
 - Jeff McAffer to look into GitHub security API

- Questions on Cargo packages - we get throttled by crates.io
 - Haven't made any effort to follow requirements for bot on cargo
 - Anyone want to work on it?
 - Questions on status for harvest - not easy to estimate "doneness"/status of harvest
 - Running/failed/complete - could the community contribute?

February 3, 2020

Attendees:

- Philippe
- Dennis
- Brian
- Sharon
- Raj
- Kevin
- Erika

Organization:

- GSoC: How about mirroring improvements to the ClearlyDefined database?
 - Also consider the need for better reporting of dependencies.
- Brian mentioned that there is interest in being mentors. There is probably no requirement for travel, generally done by online conferencing.
- softwareheritage.org is registered as a candidate GSoC organization and will support a ClearlyDefined mirroring project. See <https://www.softwareheritage.org/>
- FOSDEM meeting. There was lots of attendance and participation from ClearlyDefined at the Friday Hackathon on January 31. Lots of requirements gathered. Notes will be published soon. Expect multiple updates to the ClearlyDefined Tickets online.
- Sharon asked about C++ support. Philippe mentioned that what package management support is available for C and C++ is in the context of a few Linux distributions. Big gap -- there should be a ticket. Sharon will get more information and enter a ticket about the requirements and a request for more support in this area. Additional resources may be available to help.

Technology:

- Different technical approaches to mirroring and dependencies are being discussed.
- Mirroring See <https://github.com/clearlydefined/service/issues/386>
- Dependencies See <https://github.com/clearlydefined/website/issues/787>
 - Note that will need coordination with scancode-toolkit.

January 20, 2020

Attendees:

- Philippe
- Dennis
- Erika

Organization:

- FOSDEM meeting. There is a hackathon that takes place on January 31st in Bruxelles with many other open source compliance-related open source projects.
 - See for details and self registration:
 - <https://docs.google.com/document/d/1UphruKKAAlsoUEidPCwTF2LCcHFnQkvQCQ9luTXfDupw/edit#>
 - At this stage, Carol S., Jeff Mc., Jeff M. and Philippe O. have signed up
- Annual Report to the OSI was posted By Carol
 - <https://docs.google.com/document/d/1rx5H7d1IM5syHB7miKpcGYmOPB-JIAerDNdeJZhu7V0/edit>
 - We discussed it at a high level
- Google Summer of Code: <https://summerofcode.withgoogle.com/>
 - It would be great to participate.
 - For that we need to submit the project @ GSoC by Feb 5th and:
 - An organization admin (Philippe O) and a backup admin volunteer TBD
 - Projects ideas: we can use the new label created on the main GitHub projects "GSoC-Candidate" to tag new and existing issues as project ideas candidates:
 - <https://github.com/clearlydefined/website/labels/GSoC-Candidate>
 - <https://github.com/clearlydefined/crawler/labels/GSoC-Candidate>
 - <https://github.com/clearlydefined/service/labels/GSoC-Candidate>
 - To suggest a new idea please create a ticket:
 - For the website: use <https://github.com/clearlydefined/website/issues>
 - For the crawler: use <https://github.com/clearlydefined/crawler/issues>
 - For the service: use <https://github.com/clearlydefined/service/issues>
 - and add the "GSoC-Candidate" label.
 - A wiki page for the GSoC participation
 - Later on we will need to sign up a few volunteer mentors.
 - As a first-year org, we will likely get one or two students assigned in the best case if we are accepted as a GSoC org.

Technology:

- See GSoC above

January 6, 2020

Attendees:

- Philippe
- Carol
- Dennis
- Sharon
- Jeff Mendoza

Organization:

- Reminder meetings are biweekly from now; the next meeting is in two weeks.
- Curation meeting after this one.
- A few FOSDEM, Copyleft Conf, and other talks coming up about compliance and/or ClearlyDefined.
 - Philippe will send details on compliance summit to Discord chat and mailing list.
 - Aiming to make compliance summit a hackathon; would like to work on interoperability topics.

Technology:

- Philippe: working on large scale scancode scans quality and accuracy review (automated)
- Got more requests over the holidays for mirroring. Following up on the bug offline.

December 16, 2019

Attendees:

- Georg Link
- Philippe
- Sharon
- Jeff
- Brian
- Erica
- The Raj

Organization:

- Georg would like to integrate CD licensing and licensing score/metrics in CHAOS
 - See the design at <https://github.com/clearlydefined/license-score/blob/master/ClearlyLicensedMetrics.md>
- Jeff will be doing a Talk on ClearlyDefined for Rust Cargo packages @ FOSDEM
- Philippe will setup an event brite for the hackathon the day before FOSDEM on January 31st

Technology:

- Jeff introduced a new tool to integrate ClearlyDefined data lookup for Rust-based projects in a Ci pipeline
- Philippe is exercising the API for retrieval while Brian and team are doing the same for posting.

December 2, 2019

Attendees:

- Carol
- Kevin
- Gene
- Philippe
- William
- Sharon
- Tom
- Brian
- Simone
- Gkylee
- Jeff

Organization:

- Last meeting of the year will be in two weeks on Dec. 16. Meetings will resume on January 6 and proceed biweekly from there.
- Curation steering committee meeting after this one.
- [Philippe] Room available and confirmed at <https://www.betacowork.com/> in Brussels on January 31st the day before FOSDEM.
 - Philippe will include event details and RSVPs for the event in Discord and mailing list.

Technology:

- [Gene] Focus on supportability of the crawler and the service.
 - Looking for help with bug fixes this week if someone's available?
- [Philippe] Some issues reported on the API
 - Philippe will post script he used for this work.
 - <https://github.com/clearlydefined/service/issues/675>
 - <https://github.com/clearlydefined/service/issues/670>
 - Some may be related to sync of definitions
 - <https://github.com/clearlydefined/service/issues/386>
 - William will follow up with Jeff on this one

November 18, 2019

Attendees:

- Carol (on mute)

- Simone
- Tom
- Brian
- Jacob
- Dennis
- Sharon
- Gene
- Philippe
- William
- Gkylee
- Erika

Organization:

- Tom - has been running legal and policy dev room at FOSDEM; got interested and involved in the project
- Reminder there's a curation steering committee meeting after this one.
- We have a room in Brussels the day before FOSDEM re: open source compliance tooling; same room as a previous meeting from a previous year

Technology:

- [Gene] Discuss [definitions mirroring](#) design.
 - Issue 386 - <https://github.com/clearlydefined/service/issues/386>
 - Gene added a lower-level design; going to discuss with Jeff and William
 - William says either solution is fine with him
- [Simone] Npm dependencies upgraded on Website
 - No more vulnerabilities
- [Simone] Client fix for curated definitions, wondering if a check is needed on the service for consistency
 - Curated label on definitions
 - Wondering if for consistency we should apply the same filter on the service?
 - Issue 848 - <https://github.com/clearlydefined/website/issues/848>
 - Return only curation related to the specific version?
 - If a user doesn't provide a version they would expect everything and if a user provides a version they expect that specific one?
 - Data returned by API from database - cached?
- [Simone] Chart for Tools ran per day implemented
 - cleardefined.io/status page
 - PR open and looking for review
- [Ignacio] CloudFlare fix will be released tomorrow
 - Caching invalidation

November 4, 2019

Attendees:

- Kevin
- Ignacio
- Carol
- Simone
- Gene
- Jeff
- Dennis
- Jeff
- Sharon

Organization:

- [Curating steering committee meeting](#) at 9am Pacific after this one.

Technology:

- [Gene] November milestone created for the website and the service. It'll be updated with more issues today and later as needed.
- [Gene] Improved declared NuGet license detection by 88% based on license URLs both for open source licenses and EULA.
- [Gene] Recomputing Maven coordinates following Ignacio's fix for source URLs.
 - Recomputed coordinates; queued up messages and being processed now
- [Gene] The crawler was refactored to include only the used ghcrawler code. It should reduce the maintenance burden and enable future changes. There are currently no more dependencies with known security vulnerabilities in the crawler!
 - Created a new folder called ghcrawler and copied the files that were used
 - Ran tools to make sure it's aligned with the project
 - Running for a week with no problems
 - Have some issues it'll become easy to resolve
- [Ignacio] Closed [item #651 related to issues with user code consuming the API](#), let them reopen when they share more input (the item is not reproducible outside their code, and so far they are not responsive).
 - Need headers of response (if it's from Cloudflare or from us).
 - Maybe something on his side?
- [Ignacio] Still considering options with the [#418 queuing of URLs to be made _dirty_ on CloudFlare](#) (maybe **making the queue LIFO** can make more sense from a use-case point of view?). As per Gene's suggestion, we're queuing; but instead of 5 minutes I suggest we queue for 1 minute (accounting for the limit of only 20 URLs that can be sent at a time). Also, tests have been created for what CloudFlare replies in the case of throttling (status code 429); but a real-life end-to-end hasn't been executed.
 - Don't have enterprise-level service from Cloudflare; have to use API to mark URL-by-URL
 - Issues:
 - Tagging

- Being robust about how we call API and so throttling
 - Research throttling library to solve this?
 - Ignacio will make a PR and get reviews (and keep it simple)
- [Jeff McAffer] Mirroring?
 - Folks would like ClearlyDefined in their own workflows; hard without mirroring
 - Proposal and then getting the work done
 - Add to November milestone?
 - Run service locally but basically read-only and can only do point queries
 - How to mirror data? Push, pull?
 - Need to talk more about which model to use
- [Jeff McAffer] Embedded package support?
 - <https://github.com/clearlydefined/service/pull/425>
 - Old PR that's out of date but adds this support; would be good to review and try to get up to date
 - Not currently doing the right thing for embedded package support
 - Need to discuss further

October 21, 2019

Attendees:

- Carol
- Simone
- Ignacio
- Kevin
- Dennis
- Gene
- Jeff M.
- Justin
- Erika

Organization:

- Jeff Mendoza and Carol are working on submission for FOSDEM and CopyleftConf for ClearlyDefined.
 - Submitting to Legal and Policy Dev Room + Dependency Management Dev Room
 - If anyone is interested in talking more about it, email Carol.
- Reminder that there a curation meeting today after this meeting.

Technology:

- [Gene] Resolved Azure search services errors. The search on the website should now work as expected. The numbers on the stats page should now be higher.
- [Gene] Improved declared license detection of Debian packages & sources. Data harvesting is in progress.

- [Gene] Working on improving NuGet declared & detected license rates based on declared license URLs.
- Simone working on some stats page stuff

October 7, 2019

Attendees:

- Dennis
- Brian
- Gene
- Gordon
- Ignacio
- Jacob
- Jeff McAffer
- Jeff Mendoza
- Kevin
- Philippe
- Justin
- Sabrina
- Sharon
- Simone
- Raj

Organization:

-

Technology:

- [JeffM] Researching Container support
- [JeffM] Building CLI-level language integrations
- [JeffM] Question: Is Swagger spec quality hindering adoption?
- [Ignacio] Apply curations to files of the same hash, will push today -no schema changes
 - [Justin] - I'm thrilled by this.
 - For further info kindly check out [#501](#)
 - We must track different changes on different files (even if they have the same hash) as different actual changes on the PR that's submitted.
 - Jeff asked to see some PRs created.
- [Philippe O.] How do we tag a package as having been curated if there is no data change?

September 23, 2019

Attendees:

- Carol

- Sharon
- Jeff
- Gene
- Brian
- Justin
- Jeff

Organization:

- Curation steering committee meeting after this one
- Change to declared and discovered license concepts and definitions.
<https://github.com/clearlydefined/clearlydefined/pull/111>
 - When files are very long, how do we determine the declared license?
 - Might be a collision with scoring algorithm?
 - For curation it's more efficient to think of declared license as overall license and subcomponents with other licenses
 - Here is the node example:
<https://github.com/nodejs/node/blob/0974fc6a25c343744235331eb20d31f6412ff7e1/LICENSE>
 - Please review the PR above for the change to the policy/procedure
 - If we agree with the PR we need to update the scoring algorithm
 - Bring back to the next meeting for discussion
 - Clearer definition of when to curate a license in the 'declared' field
 - <https://github.com/jshint/jshint/tree/07ff25d3f4379415f0b438d9fb7cc6f6cabb7f7a> (one file is JSON
<https://clearlydefined.io/definitions/npm/npmjs/-/jshint/2.10.2>)
 - <https://raw.githubusercontent.com/OData/odata.net/master/LICENSE.txt>
 (do we just curate MIT or MIT and CC)
 - <https://github.com/nodejs/node/blob/0974fc6a25c343744235331eb20d31f6412ff7e1/LICENSE> or <https://github.com/lz4/lz4/blob/dev/LICENSE> -
 what do we include in 'declared field'

Technology:

- Completed the set of Swagger-related requests fulfilling
<https://github.com/clearlydefined/service/issues/614> (there is a PR on
<https://github.com/clearlydefined/clearlydefined/pull/112> that changes the documentation).
- Commitment to empty the "todo" for <https://github.com/orgs/clearlydefined/projects/35> by the end of this week, so calling for any important issues that may be missing.
- Mirroring the data for consumers
 - Mirroring definitions to address issues: scale, performance, reliability, performance
 - Various approaches to mirroring the data
 - Like to get feedback and get work done in this space

- Specifically talking about definitions, not harvested data
- Curations still happening on the website; not locally
- Consider starting a simple, read-only version of the service?
- Jeff will open an issue
 - There already was one. I (Jeff) updated the content based on the discussion today. See <https://github.com/clearlydefined/service/issues/386>

September 9, 2019

Attendees:

- Carol
- Simone
- Ignacio
- Gene
- Dennis
- Raj
- Brian
- Justin
- Jeff

Organization:

- Curation steering committee meeting today
- OSI Board meeting on 20 Sept - Carol will make a report and put it in the Google Drive if folks want to look it over
 - https://drive.google.com/drive/folders/1fmcYf_3AARq2w-FqCBGdkV160UnXulsU
- Maven repos
 - Question in curation steering committee meeting for where source for Maven should be (mvnrepository.org?)
 - Should we have some consistency around which repository we point to for source here?
 - Should we standardize on official repositories for these package managers? - no objection
- Google Groups - visibility toggle?
 - Carol to chat with OSI about mailing lists

Technology:

- Debian integration
 - Now have Debian support
 - Gene will run some more tests and everything looks good we'll start queueing up things for harvest

- Update on <https://github.com/clearlydefined/service/issues/604> [discussion as set up at the #implementation channel]; let's not forget to decide whether we want an automated tool to revise PRs that may have resulted in merges missed.
 - Reverted original issue that caused the problem
 - Curations' casing now not matching backend storage casing
 - Proposing to still support William's request but post-processing to have a flag
 - Cases where we want matching behavior vs. don't want matching behavior?
 - Endpoint can feature a flag that allows you to *not* use the post-processing
 - Ignacio will make the tool to go assess curations and see the scope of problem
- Swagger: we have started working on Issues related to it; after we're done, ideas on how to keep it updated? I.e. ideas to incorporate on Unit Tests or process.
 - Ideas on to make Swagger so it stays updated?
 - Updating endpoints also has to update the Swagger?
- Additional information on curations that don't pass validation. Can we shoot an automated email?

August 26, 2019

Attendees:

- Carol
- Ignacio
- Simone
- Jacob
- Dennis
- William
- Brian
- Gene
- Jeff
- Erika

Organization:

- William talked to TODO Group meeting last week about ClearlyDefined
 - Presentation went well - typical interest, everyone excited
 - Call to action to get people to let us know about adoption blockers and work out how they can contribute
 - General interest in looking into any blockers to adoption
 - Brief discussion about containers
 - Michael @ FOSSology brought it up in his talk at OSSNA
 - GSoC student who did integration for consumption
 - ClearlyDefined continues to be a point of discussion
 - Tern/containers conversation to pick up
- Plan for reaching out to more package communities
 - Philippe working on Python: every page of PyPI public index has link to libraries.io - work with PyPI maintainers to get link to ClearlyDefined

- We should do the same with other package types we support in ClearlyDefined such as Maven, npm, etc...
- Reminder that we have a curation meeting at 9am PDT today

Technology:

- Website Status: (Simone)
 - Added Harvest Status Indicator
<https://github.com/clearlydefined/website/issues/742>
 - Continue work here.
 - Browse search with namespace
 - Improved URL sharing shortening method
 - Had a bug in the long URLs vs short URLs change
 - Working on PHP integration
 - Caught some errors, continuing work here
- Issues review:
 - Bring in place “Allow to set the source location to an URL pointing to an archive”
<https://github.com/clearlydefined/website/issues/801> ? (Simone)
 - #604 on service - <https://github.com/clearlydefined/service/issues/604> - my conclusion so far is that this issue is actually the CDN cache acting - previous related work on purging CDN cache by using tags (<https://github.com/clearlydefined/service/pull/503>) was stopped because our plan with CloudFlare does not allow for the use of this feature - let's also discuss other options [Ignacio]
 - Possibly related to protobuf shortened URLs?
 - Can't reproduce, need the team to report within two hours to investigate further
 - Additionally, please share PR numbers when referring to issues with PRs, that helps greatly to track down root causes. [Ignacio]
 - Endpoint for badge; a different idea would be to use the website for it? - let's discuss [Ignacio]
 - For using on other websites? Are we interested in doing that?
 - As part of reaching out to other communities

August 12, 2019

Attendees:

- Erika
- Carol
- Dennis
- Brian
- William
- Jacob
- Simone

- Philippe
- Raj

Organization:

- Reminder there's a curation steering committee meeting in 1 hour

Technology:

- Gene has made some ScanCode bugs
- Dependencies proposal from Brian - #787

August 5, 2019

Attendees:

- Carol
- Dennis
- Simone
- Ignacio
- Gene
- Jacob

Organization:

- [OSI monthly board report](#)
- Reminder Carol is presenting at the OpenChain meeting in 1 hour

Technology:

- Created August project and milestones
- Proceeding on Debian integration following the new specs on path by Gene
 - Using version, architecture and revision
- Security
 - nexB project on vuln database
 - Contribution to ClearlySecure at some point in the future

July 29, 2019

Attendees:

- Carol
- Simone
- William
- Gene
- Jacob
- Dennis
- Erika
- Justin
- Philippe

Organization:

- Carol is presenting on ClearlyDefined at the OpenChain weekly meeting next week on August 5 - can send along the call details if anyone would like to call in
- Curation steering committee meeting today

Technology:

- Fixed Source Location object error on the UI
 - Simone fixed this - PR is open for this
- Questions on Debian Architecture (see <https://github.com/clearlydefined/crawler/issues/321>)
 - we have some behaviors on the website that allows a user to drop a URL on the components list to get a specific package, by parsing the URL into a path string. Since a Debian source package URL doesn't contain a specific architecture (for example <https://sources.debian.org/src/async-http-client/2.6.0-1/>), how should we define what kind of architecture we need? Perhaps we can introduce a "*" namespace identifier?
 - In the Debian specific case, we could easily have the same curation changes on different namespaces (i guess). So probably, we should consider improving the suggestion workflow including also suggestions for different namespaces of the same component?

July 22, 2019

Attendees:

- Erika
- Philippe
- Simone
- Dennis
- Brian
- Ignacio
- gkylee
- TheRaj

Carol is off in vacation

Organization:

- No topics

Technology:

- New feature request - [Improved transparency of managed dependencies](#) [Brian]
- Final agreement on not using architecture as namespace for Debian? [Ignacio]

- We said: let's continue this conversation on Wednesday; by the way, let's open up that Implementation meeting (mid week) to others who may want to join

July 15, 2019

Attendees:

- Simone
- Carol
- Ignacio
- Brian
- Philippe
- Erika
- Jacob
- Justin
- TheRaj
- Jeff

Organization:

- Curation steering committee meeting today

Technology:

- Two types of bugs while people were curating:
 - Differences in schema for curations and definitions
 - Created a tool (has rules for changing from one schema to the next version)
 - Offer to push the tool as part of the service (it may help in the future)
 - Unknowns
 - Users would retry and then it worked
 - First we evaluated if it happened with first curations on a package, but no
 - Ask for help to users (avoiding mixing the schema issue)
- Runtime curations: there is a ticket to show all definition by pending curations
 - there are least two providers: MongoDB API and GitHub API. We do not know which is being used (missing production config files)
 - Jeff states this is the MongoDB API that's used in production
 - Simone would need to get access to some exact sample records that are stored in MongoDB
- Some issues with Fossology returning non-SPDX

July 8, 2019

Attendees:

- Dennis
- Erika

- Jacob
- Carol
- Philippe
- Ignacio
- Simone
- Brian
- Raj

Organization:

-

Technology:

- Can we get the email from the user so to attribute their commits? (Ignacio)
 - Using git as a backend; recognizes name and email of user
 - Some users don't have email publicly listed
 - Add UI where people are notified that they will not attributed to them
 - Message in UI to let the user know the commits/contribution will not be attributed to them, but let them continue to make the contribution without their information provided
- Debian integration:
 - Use Architecture as a namespace as suggested by Gene? (Simone)
 - Discussed in curation channel
 - Will we support additional sources?

July 1, 2019

Attendees:

- Carol
- William
- Gene
- Simone
- Dennis
- Jacob
- Ignacio

Organization:

- Reminder there's a curation steering committee meeting today after this meeting

Technology:

- Please refresh our decision for <https://github.com/clearlydefined/service/issues/534> we said leave undecoded, right?
 - Leave undecoded
- Sorry to bring this one back for discussion yet again
<https://github.com/clearlydefined/service/issues/501> - the decision is sound but may

have consequences (it's a kind of involved change of schema); go for it? The alternative would be to shy away and implement in two stages (start avoiding the schema change, go for it at a second stage).

- Don't want to group in the schema by hash
- Convenience for curators; don't want to require files to have paths for curation
- Keep the schema
- Created July Project and milestones: <https://github.com/orgs/clearlydefined/projects/33>
 - Left the June project with completed issues and issues under review
- Buttons on harvest page - use logos instead of names and put in a hover with their names
- [Issue 764](#) - Files that we harvest link to, but otherwise don't link
 - Hover with "to get this file, download the component"
 - Faded color for the UI
 - Always show a link to the file; on click reach out to SWH to see if they have the contents of the file; if they do, show it; if not, tell the user they need to download the whole package
 - Tell Philippe it works okay; close the issue

June 24, 2019

Attendees

- Carol
- Brian
- Ignacio
- Gene
- William
- Philippe
- Erika

Organization

- Issues - please help by labeling bugs

Technology

- Option to initialize a re-scan of a component definition - Brian
 - New versions of ScanCode can be run on demand when we add the new version
 - Would be nice to get more automated information on new versions of ScanCode when they're released

June 17, 2019

Attendees:

- Carol
- Jordi
- Ignacio

- Simone
- Brian
- Dennis
- Erika
- Gene
- Philippe
- Justin

Organization:

- Reminder there is a curation steering committee meeting today after this meeting.

Technology:

- Introduce an idea of component “state”? Just harvested, currently running tools, nothing queued..? (Carol)
- Service. Let us figure out how to handle issues related to store/search. Importance: most current issues are store or search related.
 - They are normally provider/adaptor-specific, so if an issue appears on Azure, it can't be reproduced on Mongo.
 - Can't check configuration of indices
 - We tried sharing (privately) dev credentials. This worked to some extent:
 - Success: all issues are consistently repeatable.
 - But: no interactive tool to check search indices configuration, or perform tests.
 - We also tried using readonly and query tokens. Readonly tokens work great. Query tokens let us repeat the problem but don't let us configure indices.
 - ~~Option: use Azurite~~. It doesn't support search.
 - Option: we could set up screen sharing sessions. This would require coordination and would be case-by-case. Pros: access to the real dev environment with all available tooling; two sets of eyes, and shared experience. Cons: requires coordination, depending on bandwidth the experience may not be great.
 - Option: we could set up independent copies of the store for a single developer to play with, accessible with their own credentials (how much would it cost? Maybe find those 100 in Azure credit that at some point were given away?). Pros: total independence, total equivalence to prod. Cons: cost, the need to copy from another storage.
 - William will look into it and talk with Gene

June 10, 2019

Attendees:

- Carol
- Simone
- Brian
- Erika

- Gene
- Jacob
- Justin
- Jeff

Organization:

- [Monthly board report for OSI](#)

Technology:

- June Milestone:
 - Added suggestions in License Picker Modal
 - Working on fixing Multiselect problems
 - Investigating on UI / Service strange behaviors on definition details (some data are missing from the service)
 - Service not sending info about licenses, files, or tools
 - Definition may not being recomputed (webhooks somehow getting lost?)
 - Let Gene know about a case for the exact coordinates and look in the database directly
 - Could be the caching layer
 - Open pull request for service on Cloudflare caching
- Mirroring
 - Awareness and call for interest
 - Need to put in place a mechanism for organizations that want to mirror into a local environment
 - Addresses a scenario where organizations don't want to call out to a third party service
 - How to mirror in a timely and accurate way and allow people to have their own local datastore
 - Want a stream of updates with append-only log
 - Who would be interested in helping with this and providing input?
 - Philippe & Jeff
 - Jacob will check with his team about time
- ClearlySecure
 - Call for interest
 - Philippe maybe interested
 - Mapping between component ids and CPEs to enable finding CVEs
 - Information around assessments

June 3, 2019

Attendees:

- Jeff

- Simo
- Ignacio
- Dennis
- Philippe
- Justin
- Brian
- Erika

Organization:

Technology:

- Service Update:
 - Shall we remove NOASSERTION from any ANDs or ORs? (code and tests are contradicting each other).
=> after discussion, it became clear that we need to keep NOASSERTION, since tools report that for any files that are part of a project but can't be told as a particular license kind (it's a legitimate assertion that can be legitimately combined)
 - If we don't, shall we take NOASSERTION as a special value, and sort it last, so to make expressions more readable?
=> not a good idea, since alphabetical is deterministic and we have other values that are just as special as NOASSERTION
 - Additional notice: make sure that the code will enable combinations of NOASSERTION and other things.
- Philippe mentioned LicenseExpression (nexus/license-expression) that does license reduction etc in Python. There is a mechanical transformation to JS.
 - Philippe to look at that branch and see if it is usable.
 - <https://github.com/nexus/license-expression>
- There is an ongoing discussion in the SPDX legal list about whether or not and how to be more inclusive with the license identifiers. This is pretty important for the ClearlyDefined community. See the list at // Copyright (c) Facebook, Inc. and its affiliates. All Rights Reserved.
- // Copyright (c) 2017 International Business Machines Corp.
- // All rights reserved.
- // This source code is licensed under both the GPLv2 (found in the
- // COPYING file in the root directory) and Apache 2.0 License
- // (found in the LICENSE.Apache file in the root directory).
-
- #pragma once
-
- #ifdef __cplusplus
- extern "C" {
- #endif

-
- extern uint32_t crc32c_ppc(uint32_t crc, unsigned char const *buffer,
- unsigned len);
-
-
- #ifdef __cplusplus
- }
- #endif
-
-
- New June Milestones created
 - Priorities?
- Justin and crew have created a Curation Process doc that seems to be working well.
 - <https://github.com/clearlydefined/clearlydefined/pull/103>

Addenda:

- Service update NOASSERTION background

More about NOASSERTION

Monday, June 3, 2019 4:30 PM

Given:

<https://github.com/clearlydefined/service/blob/master/lib/spdx.js#L137>

...the merging code suggests that merging X with 'NOASSERTION' will give X, in any order and in any operation (OR/AND).

Since:

Ruby OR (NOASSERTION OR NOASSERTION)

...is perceived as an issue (#461),

Then:

...parse the tree before merging; remove any branches that are equivalent to NOASSERTION (e.g. "(NOASSERTION AND NOASSERTION)" "((NOASSERTION OR NOASSERTION) OR (NOASSERTION AND NOASSERTION))")

But:

There are two tests authored about three months ago, which expect to handle pre-existing (not merged) combinations of NOASSERTION:

```

OR('MIT AND NOASSERTION', 'MIT') => 'MIT OR (MIT AND NOASSERTION)'
OR('MIT OR NOASSERTION', 'MIT') => 'MIT OR NOASSERTION'

```

Additional matter (if the original interpretation is wrong):

OR(Apache-2.0 OR NOASSERTION, Apache-2.0) produces Apache-2.0 OR NOASSERTION

-whereas

OR(Ruby OR NOASSERTION, Ruby) produces NOASSERTION OR Ruby

➤ Wouldn't it help readability, to put "NOASSERTION" at the right side of any operation (regardless of the alphabetical order)?

May 27, 2019

Attendees:

- Erika
- Brian
- Philippe
- Ignacio
- Simo
- Gordon

Organization:

-

Technology:

- Website Updates:
 - New chart on the Stats Page
 - Improved experience on the file list (show expanded folders while filtering)
 - Working on improving performances on file list
- Service Update:
 - Need <https://github.com/clearlydefined/service/pull/530> to be merged in order to fix all the issues related to the merge of data (mainly the preview API)

May 20, 2019

Attendees:

- Carol
- Ignacio
- Simone
- Dennis
- Jeff
- Justin
- Jacob
- Dan
- Philippe

Organization:

- SPDX Legal Team and the SPDX License List (suggested by Dennis Clark/nexB)
 - The SPDX Legal Group would benefit from greater community participation.
 - Information: <https://lists.spdx.org/g/Spdx-legal/>
 - Next SPDX Legal call is
 - Thursday, May 30 9:00 – 10:00am Pacific Time
 - Every 2 weeks on Thursday
 - Web conference: <http://uberconference.com/SPDXTeam>
 - Optional dial in number: 415-881-1586 No PIN needed
 - New visitors are always welcome.
 - To get an idea of the current SPDX Legal Team work that needs help, visit:
 - <https://github.com/spdx/license-list-XML/issues>

- Curation steering committee today after this meeting ([agenda](#))

Technology:

- 15 PRs under review
- Website Update:
 - Score Renderer Updates - updated license doc on <https://github.com/clearlydefined/license-score/pull/10>
 - Filter on Workspace to show only definitions with changes
 - Improved Contribute modal with list of involved definitions
 - Feedback needed in <https://github.com/clearlydefined/website/issues/691> (Include other versions of the same component in full details page)
 - Feedback needed in <https://github.com/clearlydefined/website/issues/687> (Give user feedback when things are queued for harvesting)
- Service updates:
 - Change of schema: <https://github.com/clearlydefined/service/issues/483> (I'd still suggest adding a cloudflare flush for good measure when redeploying, to avoid 2 hour possible annoyances).
 - The difference between “interactive” tools and automated tools -an override hint for merging. <https://github.com/clearlydefined/service/issues/526> Do we need to expand this to other fields? E.g. would it be useful to remove dates? Suggestion: let this be evaluated by the OP -Brian Duran.
 - May want lodash “merge”?
 - All curatable properties should be overwriting whatever is there when there's a curation
 - Multiple sort fields, but “revision” goes after “name” <https://github.com/clearlydefined/service/issues/392> ; need for handling semver? E.g. “2.2.12” making it before “2.2.9”
 - Maven URL should be good to go <https://github.com/clearlydefined/service/issues/449>
 - 2 Stalled PRs from previous week; any blockers there? -maybe it'd be good to catch up on Wednesday about this?
- Dan added OTHER PR with pictures of what it looks like

May 13, 2019

Attendees:

- Simone
- Ignacio
- Carol
- Brian
- Philippe
- Justin
- Jeff

- Dan
- Erika

Organization:

- OSI board meeting was good; everyone's happy about ClearlyDefined

Technology:

- Website updates
 - Worked on fixing some bugs
 - Upgrade Monaco Editor version to meet some recent requirements
 - Discussion on implementation channel last week
 - Discussion about Eject React project to access Webpack configuration
- Service topics
 - Store components (Azure/files/mongo) debugging and testing. (rel. Issue 312)
 - Bootstrapping Dev VMs/docker composites. Catching fresh install bugs.
 - Plea for a technical meeting mid-week.
 - Send an email to Dan/Carol to arrange on the calendar
- NOASSERTION and curations
 - Any way to tell the *reason* something is NOASSERTION?
 - Validated tag for NOASSERTION if someone verifies it's a commercial license (for example)
 - New properties for things that are public domain or commercial - OTHER?
 - Do experiment to see where would need to change to add OTHER

May 6, 2019

Attendees:

-

Organization:

-

Technology:

- New May 2019 Project and milestone available
- Website
 - Fixed SPDX Picker in FilterBar adding NOASSERTION
 - Updated Apply Facets PR
 - Fixed bug on Popover Rendering
 - Fixed bug on licenses with NOASSERTION
 - Working on using Protobuf to shorten URL strings
- Service
 - Improved suggestions adding discovered licenses
 - Improved cache management on the service with Cloudflare

- Working on suggested issues

April 29, 2019

Attendees:

- Carol
- Dennis
- Simone
- Erika
- Jacob
- Brian
- Philippe
- Dan

Organization:

- OSI in-person board meeting next week: any topics we'd like to bring to the board?
 - Philippe will finish up the license clarity report - will have a draft to review and eventually post to opensource.org

Technology:

- Code snippets discussion
 - Richard provided some points; talk about Justin providing input
 - How to handle snippet cases?
 - No way to provide notes/context on ClearlyDefined
 - Mark them NOASSERTION?
 - Talk with Richard further
 - Continue conversation on Discord; put on agenda for curation steering committee
- April Milestone Update:
 - Multiselect released
 - Added logo to the components search
 - Switched search by provider with search by type in browse page
 - Planning to add same functionality to definition page
 - Added Filter bar that show all active filters and sorts
 - Added persistence of data in the ui
 - Separate navigation store on Redux
 - Fixed UI Bug on curation tabs
 - Added search by query string functionality in the browse page
 - Added links to definition files using Software Heritage
 - Working on URL Shortener with protobufs on the UI after analyzing the requirement on the service
 - Working on Cloudflare Cache
 - Working on improving Suggestions
 - Managing some recent bugs on the UI

- SPDX picker on filter bar
 - Under review; ready to be merged?
 - Presence of/absence of
 - Nice to have other values that aren't SPDX to put on optionally
- Facets at the folder level
 - Shouldn't show core
 - Nothing they can do to put something in core

April 22, 2019

Attendees:

- Dan
- Jeff
- Dennis
- Jacob
- Erika

Organization:

-

Technology:

- Scanning progress
 - Queue is at 20M
 - Dan to look at monitoring Traversed vs Processed to get an idea of how fast the queue will empty
- Container stuff
 - Nothing new. This week hopefully
- PR/issues
 -
- Amazon status (what's up?)
 - Status quo
 - Need to look at the use of the Data.
- Service side trial work for new guy

April 15, 2019

Attendees:

- Carol
- Jeff
- Philippe
- Simone
- Dennis
- Jeffrey

Organization:

- Barcelona meetings
 - Legal network Legal and Licensing workshop this year
 - Mostly legally focused, compliance discussion as well
 - Lots of interest in ClearlyDefined
 - Lots of folks seem to have financial resources to help with compliance
 - Jeff gave a workshop on ClearlyDefined, well attended
 - Lots of great questions
 - SPDX formats for inputs/outputs was commented on

Technology:

- April Milestone
 - Improved filters section on Browse page
 - Need to understand how to integrate this with multiselect
 - Started attachments implementation
 - Reviewed pending issues / PRs
- List of PRs pending review

April 8, 2019

Attendees:

- Carol
- Dan
- Simone
- Brian
- Raj
- Philippe

Organization:

- Curation steering committee meeting today
- Jeff doing a talk on ClearlyDefined and curation workshop
- Philippe is looking for a place to publish license clarity blog
 - May put in a PR for docs.clearlydefined.io

Technology:

- April Milestone Update:
 - Updated Multiselect Behavior
 - Changing from mockup
 - Working on #641 Update UI on homepage
 - Updated fonts and colors all over the UI
 - Fixed download URL for npm packages

- Done on the service
- Prioritize #641 (keep it to homepage) and then we'll push to prod
- Harvesting another million packages
- Philippe is meeting with Jeff re: handling ScanCode
- Maven updates on unpacked jars

April 1, 2019

Attendees:

- Dan
- Carol
- Simone
- Jacob
- Dennis
- Brian
- Philippe
- Jeffrey
- Raj

Organization:

- Philippe will send a draft license clarity report to the mailing list

Technology:

- April Project Available
 - Moved all the outstanding issues from March project
- April Milestones created for Website, Service, Crawler
 - Priority for this month?
 - UI issue with changes for homepage - 641
 - Apply to both Workspace and Browse page
 - Active experience in dev and not released
 - URL/download component issue on service
 - 449 - Maven issue as well
- List of PRs to be reviewed
 - Dan to review open PRs
- Changes to service going out today for GitHub integration
 - Removing lazy harvest
 - Removing auth from harvest API
 - Much easier to harvest components
 - More explicit
- Added no files from expand on response API - trims down payload
 - Workspace we could omit files?
 - Fill in data on click on magnifying glass
 - Make drag and drop on package-lock file faster
- Testing Lerna configuration

- WIP PR open on this
- Facets editing on File List completed, needs review
- License score terms
 - https://docs.google.com/document/d/14ti1IVxb8GuMVvtd8ozggca_QJpFGYfGGGggopnwFe8/edit#

March 25, 2019

Attendees:

- Carol
- Dan
- Simone
- Brian
- Jeff
- Dennis
- Jacob
- Gianfranco

Organization:

- Curation steering committee meeting today

Technology:

- Lerna work
 - Prototype in place
 - Now want to proceed with that
- March Milestone Updates:
 - Working on Facets selection in File List, Browse Page performance and style improvements
 - Opened PRs for Multiselect, Typeahead upgrade, Browse page Filters
- Priority for this week?
- April: Data quality examination - Dan and Carol
 - NPM investigation of NOASSERTION things
- Embedded package work hasn't happened yet - need to finish that work
- Article about 5k packages (Philippe) - should publish in a wide forum, maybe opensource.com or similar
- see-file/NOASSERTION
 - FOSSology/nomos produces some see-file for some scans it produces
 - What to do with noisy scanner results
 - PR open for changes for that - <https://github.com/clearlydefined/service/pull/455>
 - Create a longer list on no-op results/cases

March 18, 2019

Attendees:

- Carol
- Simone
- Philippe
- Jeff
- Dan
- Jacob
- Erika
- Brian
- Gianfranco
- Jeffrey
- Dennis

Organization:

- OSLS
 - Unconference session on ClearlyDefined
 - Did demo at OpenChain summit
 - People seem to be excited and impressed
 - Eclipse Foundation is much more involved/using the data
 - Bosch looking to support FOSSology integration
 - Interest in using this in containers - would be good to get together a prototype
 - Tern team discussion around integration there as well
- ClearlySecure
 - Please let Carol know if you'd like to be involved in the conversation
 - Jacob from Amazon will do some research and find a POC for their interest
 - Philippe interested as well

Technology:

- March Milestone Updates:
 - Fixed e2e problem that blocked the CI
 - Updated the UI fixing some UI bugs and improving the overall experience starting from the new changes in place (Spdx Filter, Scoring)
 - Working on Multiselect
 - Working on Search experience on Browse page: questions about how to improve user experience, see <https://github.com/clearlydefined/website/pull/591>
 - Clear filters button
 - PR <https://github.com/clearlydefined/website/pull/484> needs to be reviewed
 - Typeahead needs to be upgraded? See <https://github.com/clearlydefined/website/issues/540>
 - Consider removing this?
 - Investigate other options
 - Questions about Facets editing experience: FileList editing or smart GlobPicker?
 - Also need validation on user entries
 - Just add incremental functionality to add a facet to the file list

- Only select a facet on folder
 - Note: if a non-core facet is selected, core will not be selected
- Questions about validation button
 - <https://github.com/clearlydefined/website/issues/247>
 - Who should see the validation button?
 - Could the action be reverted?
 - Should a filter be added to the browse page?
 - Service implementation required
- Questions about Curation details
 - <https://github.com/clearlydefined/website/issues/154>
 - How to show details? Should we bring back the accordion view?
- Questions about Browse page sort on refresh
 - <https://github.com/clearlydefined/website/issues/592>
 - UI or service implementation?
 - Sort by release date as a default
- Extra license conversation from OSLS (Philippe)
 - Only describe things that can be represented as SPDX expressions
 - If it doesn't have an SPDX identifier we don't talk about it in ClearlyDefined
 - Currently declining to use license refs because there is no normalization
 - Discussion around facility for other licenses
 - Volunteered ClearlyDefined to host a license registry if that ends up being needed
 - License refs only really meant to be used in SPDX docs; may need to make a new thing for this case
- Package URL (Philippe)

March 11, 2019

Attendees:

- Carol
- Dan
- Brian
- Jeffrey
- Erika
- Philippe

Organization:

- ClearlySecure?
- Curation steering committee meeting today

Technology:

- Website (Simone, Gianfranco):
 - Browse page is now the new Home page

- New score field of definition from the service, website to remove `computeScore()`?
 - Probably extra code that can get removed at some point
- Clarification about Facets editing in the file list (issue [#478](#))
 - Carol to flesh out the comments on the facet editing
- Priorities for this week?
- License filter is different between the Browse and Workspace page (from issue [#496](#))
- Service
 - New score aggregation fields
- Philippe replying about embedded packages on <https://github.com/clearlydefined/service/pull/425>
- SPDX and Philippe are in conversation about providing a way to list licenses that aren't on the official license list

March 4, 2019

Attendees:

- Carol
- Dan
- Jacob
- Simone
- Erika
- Brian
- Jeff
- Gianfranco
- Alex

Organization:

- Communication and Engagement Templates for Project Teams [Brian/Erika]
 - Sent an email template to Jeff and Carol
 - Carol made a stub doc for the comm for folks to look over: <https://docs.google.com/document/d/1tJ7yflph8XsUb1ZgXiq8cN6qdiGZEInCHhkq53L1vaw/edit>

Technology:

- Handling and discovering embedded packages using ScanCode
 - <https://github.com/clearlydefined/service/pull/425>
 - Follow-up with Philippe
 - Also want to think about the curation experience for embedded components
- Website Updates (Simone, Gianfranco):
 - New FileList with ant-design released, PR open with UI improvements
 - Fixed UI layout issues

- Working on Browse Page issues
- Service Updates (Simone):
 - Download actual component implemented, needs improvements
 - Expanded details about Curation in place (list of merged and pending curations)
- Define March Milestone prioritized activities
 - By end of this week main page should be Browse page
 - All of proposed website features should be done
- Issue in FOSSology repo: <https://github.com/fossology/fossology/issues/1292>
 - Copyright tool is producing bad data - may need to turn this off for now

February 25, 2019

Attendees:

- Simone
- Carol
- Gianfranco
- Brian
- Dan
- Jeff
- Erika
- Jeffrey

Organization:

- Curation meeting today

Technology:

- February Milestone update (Simone, Gianfranco):
 - UI Responsive Fixes
 - Full Details responsive view: which items should be shown?
 - Probably don't need to include the file list
 - Don't need harvested data tabbed view included
 - Drop the detailed information in file list and below for now; we'll add it later if needed
 - Multiselect for components: needed users contribution on this
 - Probably 10's of things - may not want to limit that in any way
 - Format Painter or copy/paste
 - Make change to one definition and then apply to all others users choose
 - Only for license for now
 - Switched FileList from react-table to ant-design Table
 - PR open; fixes bugs with react table
 - Some small issues with filters
 - E2E integrated in the CI
 - UI fixes (Maven icon, Facets background)

- Download component: needs service implementation
 - Add section to Described and have three URLs
 - Simone and Gianfranco can work on the service
- Curations details: should we contribute on service development?
 - Queries over Mongo curation store
 - Curations API endpoint functionality should be moved off Definitions endpoint
 - Dan can take a look at API
 - Definitions view also needs a different API
 - Dan will put route in that
<https://github.com/clearlydefined/website/pull/320> is pointing at
 - Simone to simplify curations at bottom to just be summary
- Dan is rolling out monk and copyright
 - Have new Scancode out as well
 - Monk is being broken out into chunks of 500 at a time
 - Break them out for copyright at well (?)
 - Going to reharvest and rescan everything aggressively after all these all are in

February 18, 2019

Attendees:

- Jacob
- Philippe
- Alex
- Simone
- Gianfranco

Organization:

-

Technology:

- February Milestone: (SIMONE)
 - Service:
 - Issues in the project backlog
 - Available to contribute and help on any outstanding issue
 - Website:
 - Merged E2E
 - Working on FileList bug
 - Stats page, is the requirement fulfilled?
 - Question about structure for Standalone components
 - Given the investigation already made on this topic (custom implementation, Lerna pattern), should we put this requirement in standby or proceed on a specific way?

- Need updates on Issue 462 - downloading the component
 - We should understand if this behavior will be implemented on the service or on the ui side
- Reviewing website outstanding issues
- New scancode release 3.0.2 ready for use.

February 11, 2019

Attendees:

- Carol
- Jeff
- Brian
- Dan
- Philippe
- Erika
- Jeffrey

Organization:

- [OSI board monthly report for February](#)
- Curation steering committee meeting today
- FOSDEM

Technology:

- ScanCode 3.0?
 - Philippe - out today
 - Philippe and Dan to meet to discuss tomorrow
 - Won't have plugin for license text until later; won't queue up the rest of the things that need rescanning until that plugin is available
- Issue 462 - downloading the component
 - Put it in the definition
 - 3 URLs - component/web UI, version, download URL
- Simone updates:
 - February 2019 project has been created and available here: <https://github.com/orgs/clearlydefined/projects/28>
 - We've also created February 2019 milestone for each of the repos
 - Thanks to the new /find API, last week we've finalized the work on the Suggestions, together with @dabutvin, and the behavior is finally in place (i saw someone already used it to contribute)
 - @dabutvin has started a work on the new Browse Page based on the same /find API, which has been integrated by us, the PR is still open but i think it's ready to go
 - during the weekend i saw some issues coming, i've already worked on some of them today on the UI

February 4, 2019

Attendees:

-

Organizational agenda:

-

Technology agenda:

- January Milestone
 - Defined E2E strategy (Simone, Alex)
 - Added configuration to run Unit and Integration tests separately
 - Fixed UI Issues with Tests coverage (Simone)
- February Milestone
 - Need informations to complete Suggestions API (Simone, Dan)

January 28, 2019

Attendees:

- Carol
- Simone
- Brian
- Dan
- Erika
- Jacob
- Jeff
- Jeffrey

Organizational agenda:

- Curation steering committee meeting today
- FOSDEM presentation & meetings

Technology agenda:

- January Milestone:
 - E2E tests
 - Ongoing work on naming for identifiers to assist in UI tests
 - Name test things with a prefix: t- or t_
 - Pull request #416 - Dan, Jacob, Simone to work together
 - Test setup configuration
 - Just a few small things to clean up
 - UI Bug Fixed
 - A few open issues fixed this week
 - Status of Suggestions from a service point of view?

- May not want to compute suggestions up front, could end up being quite costly
- List of existing curations and list of suggestions as separate
- Want to do benchmarking on database for search vs. mongo queries
- Jeff and Dan will share code with Simone for work on the service

January 21, 2019

Attendees:

- Philippe
- Alex
- Simone

Organizational agenda:

Technology agenda:

- January Milestone:
 - E2E Tests have been put in place and merged, needs to improve them with newly behaviors
 - License Picker has been released, we should expect feedback from users. Philippe thinks it could be better an inline solution with an assistant on completion. Most of the licenses are plain and not complex.
 - Test coverage has been improved and we should do more work on this direction

January 14, 2019

Attendees:

- Carol
- Thomas
- Dan
- Simone
- Jeffrey
- Brian
- Erika
- Jeff

Organizational agenda:

- Curation meeting today - we meet every other week for ½ hour
- Quick summary of CD presentation
 - SAP had an internal meeting re: ClearlyDefined
 - Colleagues had some input
 - Project sounded interesting overall to them
 - Focus turned to security - lots of interest there
 - Would love to talk with folks about how to scope that out

Technology agenda:

- January Milestone:
 - License Picker
 - Finalized work and published a video on its behavior; integrated comments from Jeff
 - A lot of methods shared with the service - take a direction on how to share those methods between the service and website?
 - Don't block this PR on that
 - Would like feedback on how that works for curators
 - Stylesheet questions outstanding
 - What's the right standard mechanism for injecting styles into components? - Expose users to custom classes and use that to set style
 - Yay tests
 - Standalone Components
 - Give Lerna a test and see if it works better?
 - New items?
 - Suggestions? - let's try to get that done this week
 - Data for curations should be in document store; need to write an end point that gives us that
 - Dan and Jeff will check it out
 - Suggestions should be in for the end of the month
 - Details page has a place already for suggestions
 - Outstanding PRs
 - E2E Tests and Tests coverage - using local data
 - Figure out if there's prior work here we can replicate
 - Update Scancode - Philippe has said that 2.9.8 and 3.0 are no different except for documentation
 - Need a real story from Philippe on when 3.0 is going to be ready
 - Roll out and not rescan quickly
 - Status code checks PR - make sure error codes we're giving and provide the user the right information and errors
 - Jeff will take a look at this PR today
 - Ping FOSSology folks about issue with monk SQL query - Steffan, Max, Michael

January 7, 2019

Attendees:

- Carol
- Dan
- Jacob
- Brian

- Simone
- Erika

Organizational agenda:

- Open Source Leadership Summit CFP

Technology agenda:

- January milestone
 - Open items for the month
 - Plan a meeting with Alex - afternoon Pacific time?
 - SPDX license picker - finalized the work, small UI render issues
 - Suggestions API on service - Simone
- 500k more things got crawled over the holidays - npm, nuget, maven(?) - we'll have a total of 400k definitions now
- Cocoa Pods integration? - need to followup
- Mods to the crawler for testing
- Hashes etc in the crawler
- Run Licensee everywhere

December 17, 2018

Attendees:

- Carol
- Simone
- Alex
- Dan
- Erika
- Jeff
- Jeffrey

Organizational agenda:

- Last meeting of the year; first meeting of new year will be Jan 7

Technology agenda:

- Curation syncing issue still ongoing?
 - Syncing not running regularly
 - Logic app - Dan will look into this
- License SPDX Picker (Simone)
 - Integrated with UI
 - PR has a video of the behavior
 - Simone would like feedback on implementation
- Suggestions API (Simone)
 - Updated PR recently
 - Error in ScanCode - failing if a definition is missing a license

- Will open a PR to fix that
- Check Curation Exists (Simone)
 - Completed work; merged PR to service
- Check Status code of response (Simone)
 - Starting work on this
 - Have some comments
 - Should submit an example of what they're going to do and we'll see how it goes
 - Need to keep it really simple
 - Should have a clean separation between service and routes
 - How to verify if app insights is already logging a particular error - unless there's a log message the only thing app insights will do is logging outbound calls as dependencies
- Jeff did some stuff on the weekend about adding hashes of each file and surfacing those
- Jeff added some things with licensee and part of the scoring algorithm that needs to know whether things are the license text
- Suggestions PR may be ready to merge? - Simone to make a video of the user experience
 - Do a demo video
- Push to prod today

December 10, 2018

Attendees:

- Carol
- Simone
- Brian
- Dan
- Thomas
- Philippe
- Jacob
- Jeff
- Alex

Organizational agenda:

- [OSI board report for December](#)
- Next week is our last meeting of the year, no meeting on Dec 24 or Dec 31

Technology agenda:

- Browse page additions
 - Missing APIs
 - Need some work from Jeff or Dan
- SPDX License Picker (Simone)
 - Finish implementation on the next few days
 - Working for plain licenses for now

- Go towards a monorepo and have separate npms but just one repo
 - We can carve out into a separate npm at some point if needed
 - Share a movie/mockup
- Tests Coverage (Simone)
 - Improving tests on website
 - Tests for 4 or 5 components and currently writing more
- Standalone Component question (Simone)
 - Monorepo
- Suggestions API
 - Implemented a new method for suggestions API and close to finished
 - Need to fix some test errors; original and new implementations are a little different
 - Built all methods for all the suggestions fields
 - WIP PR
- Service PR #288 Question (check that curations actually exist)
 - Thomas has fixed some of the comments, pushed some changes on this
 - In a list of curations if one is missing the whole PR should fail?
 - Fail the whole submission if there's a problem with one piece; show them there was a problem and explain in error message how to fix it
- Snyk Authorization
 - Problems with a fail during a PR last week
 - Dan to add Simone, Alessandro, Thomas to Snyk authorization - needs email addresses
- Scoring algorithm
 - Implemented scoring algorithm
 - Scoring element around "uses spdx" - we only use SPDX; if you have a license you get points and if you use SPDX you get points when really it should only be one thing
 - Won't have anything that doesn't have a representation in an SPDX identifier
- ScanCode v3.0
 - Philippe will be working on this this week
 - Preparing a release that will be available tomorrow - 2.9.8 (that will be the frozen code essentially)
 - Documentation is the only thing left
- Philippe to create list of SPDX license expressions for Dan
 - Mapping from SPDX identifier to ScanCode identifier - we could put that mapping into the code
 - Table/json file from Philippe or in ScanCode 3

December 3, 2018

Attendees:

- Brian

- Carol
- Simone
- Jacob
- Jeff
- Alex
- Jeffrey
- Erika

Organizational agenda:

- Meeting on Dec 10 and 17; no meetings on Dec 24 or 31; first meeting of new year on Jan 7, 2019

Technology agenda:

- November Milestone Update (Simone, Alex)
 - Browse Page
 - finalized
 - Folder Structure
 - Did some bug fixes
 - Working on suggestions and complete in the next few days
- Testing and puppeteer
 - Open pull request - 286
 - What is the status?
 - Implemented unit tests on website
 - Blocked on react scripts?
 - Write documentation the tests for folks to use
 - Want to implement puppeteer to get end-to-end tests
 - Continue work on this month
 - User story for developers to be able to write end-to-end tests
- Pull request 313 (website)
 - Workflow for publishing npm (when does it get published?)
 - Don't want people to get out of sync
 - Let's assume Daniel isn't going to do any further work on this
 - Enable developers who are doing other work to integrate ClearlyDefined into their workflows
 - @clearlydefined on npmjs
- Scoring algorithm details should be surfaced on website
 - As a user you should be able to see why the score is the way it is
 - Possibly just static text that links to webpage about making it better
 - Guide users directly to specific files and remedies later
 - Decide on roadmap for the work
- License expression editor and license picker
 - Need to understand what the UI is

November 26, 2018

Attendees:

- Carol
- Jeff
- Alex
- Simone
- Brian
- Thomas
- Erika

Organizational agenda:

- CFPs all submitted? - yes
 - Carol and Jeff to confirm on pentabarf for FOSDEM
- New folks from SAP on curation meeting, some other new attendees

Technology agenda:

- Go over outstanding PRs
 - Suggestion implementation for website nearly done
 - Simone talked to codescoop folks re: extract components
 - Decided on mono repo approach - Simone will investigate CI build for open PR
 - PR build just needs to run tests; CI build needs to publish docker image
 - Could do npm link for dev? Prod build should publish the npm
 - Explore fully the full 'cycle' for developer experience or matching of components and things running on website and not matching what's published in npm
 - Not super priority, but let's continue work on this
 - Browse page - separating logic; once done with the refactor can build the new browse page
 - Need an API, use an existing API to return all the required data?
 - Probably need a new API that returns the list of components (will need to be paginated)
 - Issue 265; PR 287 - small cosmetic change (Jeff to merge today)
 - PR 288 - checking components in registry with provider; working from Dan's auto-harvest work
 - Lint errors on this PR
 - Starting on suggestions API next
 - Goal for today to get mergeable PRs merged and pushed to prod
- PR build infrastructure
 - Over the holiday Jeff moved the PR builds off Travis and to Azure Pipelines; in a new spot keep an eye on it
 - Update CI builds over the next few days
- Scoring algorithm

- Implemented
- Need to talk to Philippe about some tweaks

November 19, 2018

Attendees:

- Philippe Ombredanne
- Brian Duran (SAP)
- Jeff
- Simone
- Thomas Moran
- Jacob
- Erika Cruz (SAP)

Organizational agenda:

- LLW and FOSDEM conference submissions are due soon

Technology agenda:

- November Milestone Update (Simone, Thomas)
 - Suggested data
 - Browse page
 - Refactoring Page definition
 - Service PRs
- PR and curation database queries
- Google: crawler on GCP with crawler id? -- Jeff to follow up

November 12, 2018

Attendees:

- Jeff
- Thomas Moran
- Jacob
- Philippe

Organizational agenda:

-

Technology agenda:

- Problem running Scancode
 - Thomas is unable to run the crawler
 - Philippe will look at any issues
 - Suggest that Thomas just hack the crawler
- Service PRs
- Website update

- DocDB to be deployed prod
- Amazon crawlers
- TPN generation

November 5, 2018

Attendees:

- Carol
- Brian
- Jacob
- Philippe
- Thomas
- Simone
- Alex
- Jeff

Organizational agenda:

- OSI board report - [November, 2018](#)
- Curation steering committee meeting this week after this meeting
- Spring conferences - FOSDEM and OSLS and FSFE legal network
 - Philippe speaking at Paris open source summit - he'll provide slides
 - Philippe will submit for license scoring - legal devroom

Technology agenda:

- October Milestone Updates (Alex, Simone, Thomas)
 - Outstanding questions on Website and Service Issues
 - Website - mainly work related to DnD
 - Some refactoring work
 - Questions on some issues -
 - 170: adding a custom links into checks? Not checks, would be the status API; need a deep link to that pull request; check if this is already implemented?
 - 339: need feedback on solution implementation
 - Service
 - 260: Storage of curations in datastore - implementation of storage and updating based on PR changes is implemented in that PR; might be already implemented
 - Not working on any of the queries for any of the suggestions
 - Defining relationships for suggestion candidates - temporal, version, or source/package relationships
 - Given schemas of definitions and YAML from curations, we need to determine what relationships would be interesting there

- 165: Logging - is there another instance of Winston being used somewhere? Create a logger based on Winston and pass that in, assume it is the Winston API
 - <https://github.com/Microsoft/ghcrawler/blob/jm/apicrawler/crawlerFactory.js#L351>
 - How to make logger(s) available to the different elements in the service?
 - Middleware to tack logger onto requests as they come in?
 - Put it in the options object?
 - Don't use a global
- DocDB - implemented MongoDB support for it
 - Monitoring pull requests; use webhook to GH for open, close, or synchronize events
 - Pull requests getting stored in DB on open and get files for PR (curations) and also put them in a separate collection
 - No config checked into source control; once it's working it will be part of dev service deployment and connected to dev Cosmos DB
- Go through and close September and October and opening November - separate meeting for Carol to set up
- Status report on license score (Philippe)
 - Completed run on top 5,000 packages for license score
 - 150 out of 5,000 packages that have a score above 80
 - RubyGems may have a bug?
 - Separate meeting to discuss metrics
 - Scoring and our conversations should be based on ClearlyDefined
 - Post on legal network mailing list introducing ClearlyDefined
 - Want to implement scoring computation in ClearlyDefined
 - Eventually we'll want to expose how the score is computed in the UI?
 - Top level key files may not only have to do with licensing; may eventually have to do with security
 - Want to know which licenses we have full text for
 - Array with list of *which* license keys have the full text
 - Adding a property (e.g. ContainsLicenseText) to schema that's an array of SPDX license identifiers; if it's present the corresponding file has the full license text
 - Recompute 5,000 packages for scoring once it's implemented
 - Canonicalize list of 5,000 - "ClearlyDefined Index" - benchmark/bellweather list; encode directly in service and page on clearlydefined.io with "state of the world" summary statistics about different package managers and associated scores
 - Philippe to provide data and list
 - Article on scoring topic

October 29, 2018

Attendees:

- Carol
- Brian
- Jeffrey
- Jacob
- Philippe
- Simone

Organizational agenda:

- Curation steering committee meeting next week
 - Review [documentation PR](#)
 - Review [docs](#)

Technology agenda:

- Philippe: what is the process to update the definition schema? (to support license scoring data)

October 22, 2018

Attendees:

- Carol
- Simone
- Dan
- Jacob
- Brian
- Thomas
- Alex
- Philippe

Organizational agenda:

- Curation steering committee meeting again today after this meeting
- ClearlyDefined and little attribution builder at OSLS

Technology agenda:

- October Milestone update (Simone)
 - Tests delivered for UI
 - Drag and drop behavior
 - Service side work - questions that need resolution - have json data coming back in reference; trying to create a static root but doesn't look like it's working
 - Submit all changes on definitions
 - Increase overall data visibility - merged or pending PRs - waiting for docdb implementation
 - Updated documentation on UI
- Weird data showing up in definitions - #319

- May be from FOSSology?
- Will recompute all our definitions and that will finish up the docdb work

October 15, 2018

Attendees:

- Carol
- Simone
- Ricardo
- Thomas
- Brian
- Jacob
- Jeffrey
- Philippe
- Jeff

Organizational agenda:

- Curation steering committee meeting today and next week at 9am Pacific ([doc](#))

Technology agenda:

- Swagger UI (Thomas)
 - Yaml file and the json schema are very different
 - Need input from Jeff
 - Separate API doc?
 - Losing a lot of what's in the yaml, but all you get with the curations.json when run through swagger is models
 - Continue down that route or not?
 - Jeff would like to see the code to review
- October Milestone Update (Simone)
 - Addressed some issues, currently working on some of it and plan to fix up the start of this week
 - A lot in progress, but some are open PRs
- October 2018 Work - Issues Prioritization - <https://github.com/orgs/clearlydefined/projects/24>
 - Jeff will take a look at changes and merge today
- Testing on website - Simone has started to bring in the tests on implementation; planning to bring in tests on React scripts as well
 - Implement some examples for testing
- Logging - Thomas is working on logging
 - couldn't find reference to Winston for crawler
 - Ghcrawler uses Winston, instantiate Winston at beginning and pass along as logger everywhere
 - Crawlerfactory.js - in API branch (link in implementation channel)
 - Service: go in app.js

- Middleware that sticks logger onto requests
 - Jeff will put more info in issue
- Curation contribution
 - Make consistent; doesn't have a contribute button on modal popup
 - Leave code and behavior as it is and change buttons
 - User flow is dig deep into component, then come back to list, pop up details, etc and then contribute is from list
 - Details view should be more "modally" - should have margin at top and not look so much like a new page
 - Full details page that's been deep linked should have contribute on it
- Drag and drop support for other things than files
 - Did an implementation for testing behavior and URLs
 - Tests - maybe remove React drop zone; currently only using it for accepting or rejecting specific files and can do that in pure javascript
- Jeff has PRs open that need review, hopefully for today
- Purl - not a substitute for direct url to download; doesn't distinguish between type and provider
 - Two URLs is the recommended behavior and relate the two
 - Making it a query parameter makes it a second class notion; should be a first class notion
 - Coordinates surface to users in multiple ways and it's consistent but might not be powerful enough

October 8, 2018

Attendees:

- Alex
- Philippe
- Simone
- Riccardo
- Thomas
- Jeff

Organizational agenda:

- [OSI board report](#) for October
- FOSDEM - Oct 13 deadline for Main track proposals
<https://fosdem.org/2019/news/2018-08-10-call-for-participation/>
- OSS Summit Europe in two weeks
- Open Compliance Summit Japan

Technology agenda:

- Doc DB status

- October 2018 Work - Issues Prioritization - <https://github.com/orgs/clearlydefined/projects/24>
- Environment Configuration
 - Should be completed today. Jeff to review PR.
- Website Test approach investigation
 - Simone to update to react-script 2.0 and test in separate branch
- Pending PRs
 - Jeff to ensure Issue open and fully described
 - Alex et al to implement
- Low score/things that need curation view
 - Look to add a Browse capability
 - Need an Easy button workflow to guide people of the things that need attention
- Related values
 - Discussed the shift from related curations to related data
 - Jeff to add issues to GitHub to ensure this topic is covered
- A new reviewer view for curations
- Harvesting top N components

October 1, 2018

Attendees:

- Carol
- Brian
- Alex
- Simone
- Jacob
- Philippe
- Jeff

Organizational agenda:

-

Technology agenda:

- Update on September Milestone (Simone and Alessandro)
 - General updates
 - Some outstanding issues from September - drag and drop
 - Addressed reverting changes
 - Showing users specific curations and all info from definitions
 - Jeff reviewing WIP PRs
 - Website Issues and PR
 - Service implementation
- WIP refactor of service
 - Changing this week; move definition to documentdb

- Amazon crawler in prod?
 - Jacob will discuss with James
- Comments on everything that needed comments; things that could be merged were merged
- Linux support from Craig (close?) - Jeff will ping Craig
- Codescoop folks are doing work around refactoring for pluggability; need more discussion
- Would like to get db changes in as soon as possible
- Review PR for documentdb change; can focus mock API work on this change
- Need to make choice on attachment storage
 - Now have harvest and definition store separated
 - Do we want a separate attachment store as well?
- Need more tests as well
- Alex - how does deployment in dev work? Happens as soon as you merge into master
 - Some changes not appearing on dev website?
 - Historically we've had trouble with create react job agent that doesn't reload pages
 - Jeff will check
 - Possible it's deployed but failing to restart?
- Working on switching deployment pipeline to be a public account to give everyone access; that should happen this week
- Philippe created a repo for tracking scores: license-score
 - Working on that repo this week
 - Recompute scores on existing scans
 - Data elements missing? Because of version of scancode and some of the options on running the scans
 - Scoring is based on definition that's already been computed
 - Need to follow up on scoring discussion
 - Also going to make a list of 5,000 popular packages, compute their score, and monitor score over time (1,000 RubyGems; 1,000 PyPI; 1,000 NuGet; 1,000 maven; 1,000 npm) - can use `top` route in crawler
 - Philippe will cc Jeff on email about maven popular packages

September 24, 2018

Attendees:

- Alex
- Simone
- Carol
- Jacob
- Brian
- Jeffrey
- Thomas
- Philippe

- Jeff

Organizational agenda:

- Carol sending an email about starting a curation committee to mailing list today - last chance to express interest
 - Reach out to Rashmi about curation steering committee
 - Jacob may be interested in participating
 - Jeffrey/SAP may be interested in curation

Technology agenda:

- Updates on September Milestone Work(Simone and Alex)
 - Notes about website PR #301 and service PR #236
 - PRs are waiting for review
 - #301: needs review and approval with #236; changes structure of data
 - Carol to talk to Jeff about reviewing
 - Waiting on database development
- Discussion about issues #290 - #291 and related PR and deploy process (Simone and Alex)
 - Needs backend development; use mock APIs
 - Shouldn't go to production
 - Can make a WIP/Do Not Merge PR for review to look over
- Jeff spent time looking over documentdb/store exploration (MongoDB)
 - Bigger refactoring than expected
 - Aiming to get a PR out for people to look at
- Philippe will submit a PR for scoring computation and code changes in ClearlyDefined for scoring changes
 - Scoring algorithm should not rely on newer version of ScanCode; have 100,000 definitions from current version

September 17, 2018

Attendees:

- Carol
- Brian
- Philippe
- Jeffrey
- Simone
- Gianfranco
- Alex
- Jeff

Organizational agenda:

- Philippe made submission to Paris Open Source Summit - talk on ClearlyDefined

- Eole - European open source legal event - submitting a talk for that as well, curation and other legal issues

Technology agenda:

- License scoring update:
 - https://docs.google.com/document/d/1o_VuVnP1PMoYOXZHSm0TyDX4CsqDXM3zqESAE0AoUnQ/edit#
 - Presence or absence of license text is binary
 - Working implementation in branch in ScanCode of scoring: <https://github.com/nexB/scancode-toolkit/tree/license-scoring>
 - Implemented focusing on core facet (makes a big difference)
 - For gifs do we care if they have copyright/license text files?
 - Philippe will provide examples
 - We'll table the question of images for now
 - New release of scancode and Philippe will push to branch with scoring computation (and then move to Javascript code as opposed to Python code)
 - Would like to see some more samples
 - Should have something working/in place by end of month
- Push notifications proposal
 - Spec - Simone
 - <https://docs.google.com/document/d/1BeFUFa25nau3gW3nOhBZ3jv1YL-BjyrBthIVyLeznhg/edit#>
 - Web sockets for real-time communication with service
 - For any kind of data we add in the future as well
 - Added some assumptions:
 - Push notifications for approved and merged curations
 - Does it matter that we specifically focus on those two events? - not very much
 - When we put web sockets in place we can use it for anything
 - Already gets webhooks, we shouldn't have to poll
 - Consider SPIKE concluded
 - Would like to think about external service approach; want to generally implement less
 - Use redis session so it's kept external to node?
 - Consider a database that has a notification channel
 - Push to database all changes we want to push and clients can subscribe to channel and get data
 - Firebase, couchdb?
- Alex/Simone update for September progress
 - Three issues under review
 - Currently working on enhancing contribution description (in website and service)
 - Have an updated mockup of how this will work
 - Working on reverting changes (need feedback)

- Structure for standalone components and source location GitHub selector - no update from CodeScoop (Carol will ping them)
- Apply related curations (schedule time to talk more about that)
- Hope to get PRs through quickly
- Jeff will try to get them reviewed soon
- Some of the issues in progress have been completed they just need review
- Service-side things on data storage for supporting the September milestone
 - Issue 290
 - Seeing existing or pending curations on a definition
 - We don't currently have that data
 - We need a database that talks about curations
 - One PR could potentially affect multiple definitions and components
 - Need a multi-dimensional way of querying contributions
 - Need a database with list of all contributions, curations, PRs, etc
 - There will be more scenarios where people want to query our definitions
 - Currently have paths and with partial paths you can find definitions
 - Cross-type and provider queries are hard
 - Jeff wrote a doc
 - Need a database (can be a side thing and keep blob storage or we can throw away blob storage and switch to database entirely)
 - Need something that's fast for point accesses
 - Document database (docdb, mongo, etc) as one and only golden store for definitions
 - Should be reasonably fast for point queries
 - Can create lots of different kinds of queries
 - More costly
 - Don't need to worry about scale *too* much
 - Firebase has real-time capabilities out of the box
 - Pricing is a concern
 - Low implementation cost but we need to look at on-going cost
 - Need a list of requirements:
 - <https://docs.google.com/document/d/1pB5ue8It7mRO1K4tesLGMTXcbZWVg7Q8N0ibTlft1fA/edit>
 - Another scenario we need to support is mirroring the data and being able to not worry about hitting the API
 - Database with replication capabilities or streaming from feeds would be really great
 - Please add content and comments
 - Would like to get some conclusion this week

September 10, 2018

Attendees:

- Carol

- Simone
- Brian
- Jeffrey
- Gianfranco
- Jacob

Organizational agenda:

- [September OSI board report](#)
- Curation steering committee starting up, please let Carol know if you want to participate

Technology agenda:

- Update on September Milestone (Simone, Gianfranco)
 - Working on issues and fixed some bugs
 - Keeping an eye on issues coming up
 - Need to talk with Jeff, but have some specific questions
 - Ask Jeff about dev website - does it autobuild? How to check what the latest build is for dev?

September 3, 2018

Attendees:

- Alex
- Gianfranco
- Jeff
- Philippe
- Simo
- Thomas
- Valtteri

Organizational agenda:

-

Technical agenda:

- Alex et al fixing bugs...
 - Copyrights #284 <https://github.com/clearlydefined/website/issues/284> many copyrights per file. Make the UI sticky so users do not lose edits
 - Update doc to reflect the new UI terms
- September UI work
 - Remove stale code around curations etc
 - <https://github.com/clearlydefined/website/issues/287>
 - “Push notifications” for changes while you are on the pages a la GitHub
 - Think about this and see how much work etc.
 - Spend a day to get the pros and cons

- Robustness and tests
 - <https://github.com/clearlydefined/website/issues/24>
 - <https://github.com/clearlydefined/website/issues/108>
- Source location picker #179 <https://github.com/clearlydefined/website/issues/179>
 - Generalise from harvester picker code. Suggest a way for packaging and independent picker component re-usable in any 3rd party app
 - Daniel will work on this and make a proposal
- Embeddable stand-alone contribution widget for single component field value
 - <https://github.com/clearlydefined/website/issues/289>
 - Encourage contributions by enabling minimum friction contribution widget for 3rd party apps
 - E.g. Whenever license value of a particular component (and version) is displayed in an application there is an affordance to submit a contribution of fixed value on the spot.
 - Daniel will work on this and make a proposal
- Improve curation visibility story (show which components have curations, find pending curations, ...)
 - <https://github.com/clearlydefined/website/issues/290>
 - <https://github.com/clearlydefined/website/issues/187>
 - <https://github.com/clearlydefined/website/issues/149>
 - Pull forward older curations workflow
<https://github.com/clearlydefined/website/issues/291>
- Better description for curations
 - <https://github.com/clearlydefined/website/issues/282>
- Handle limit on the number of components per request
 - <https://github.com/clearlydefined/website/issues/243>
- September Service work
 - Scoring
 - <https://github.com/clearlydefined/service/issues/225>
 - Revisit Coordinates, purl, ...
 - Big picture data source – which components have defs, curation, pending curations, ...
 - Jeff to kick off with a doc summarizing usecases
 - <https://github.com/clearlydefined/service/issues/226>
 - Harvest permissions story
 - Model for who should be able to submit
 - Implementation with tokens? Or?
 - Limit on how many defs can be requested at once (save the service from DOS).
 - <https://github.com/clearlydefined/service/issues/192>
 - Needs the UI to “page”
- September Crawler work
 - Support more package managers (Go, Rust, ...)

- Validate PyPi and Gems
- Crawl lots of stuff

August 27, 2018

Attendees:

- Carol
- Jeff
- Alex
- Brian
- Jeffrey
- Jacob
- Simone
- Gianfranco

Organizational agenda:

- OSS NA
 - Thursday we're talking about ClearlyDefined
 - We'll be demoing and showing the UI
 - Please focus on bug reporting and things in the UI that need fixes
 - Demoing on dev for showing curations

Technology agenda:

- Case sensitivity work
 - We've done a bunch of work on case sensitivity on package names and other parts of the crawler and service
 - We'll have a doc about how we treat everything
 - We had some package managers that were case insensitive but case preserving
 - Trying to preserve case all the way through the workflow but we store all the blobs in lowercase
 - You can never have two packages differing by case and having the same versions
 - Only a few historical cases in npm that have the same version but different casing
 - NPM does not allow anymore packages that are differently cased but same version
 - User should be able to enter anything they like and the repositories provide the canonical name and that's what we use from then on
- Weekend deployment work
 - Source location changes
 - Different situations that use source location
 - Still has a URL as a helper but makes it so people don't have to worry about it

- Client writers (ORT, SW360, etc) don't have to worry about how to construct URL; it's now computed
- Compute the URL knowing that each of the repositories work a certain way
- Pushed some changes to the UI that change the way we handle user input on the curation for source location
- Everything has been deployed
- Would like some more tests :-)
- Crawler on AWS
 - Put together a CI pipeline to get the crawler updating
 - All seems to be working
 - Want everyone on the same version of the image; don't want crawlers running for example without changes
 - Jacob putting in webhooks to trigger rebuild after updates

August 20, 2018

Attendees:

- Jeff
- Carol
- Philippe
- Mark
- Jacob
- Alex
- Brian
- Simone
- Oscar
- Thomas

Organizational agenda:

-

Technology agenda:

- UI updates
 - File list and full details view
 - This sounds like it will get merged soon
 - Copyright editing
 - License picker
- Case sensitivity
 - Couple PRs open around case sensitivity
 - Due to quirk of how files were being stored we were losing casing on files
 - Been in the process of figuring out how that should work
 - Looking to roll that out this week
 - Going to have to do data migration and need to restore case on some pieces

- Coordinates should be case preserving but case insensitive (preserve what package manager gives)
- Data store itself is all case insensitive
- This will cause a problem for package managers that are actually case sensitive and allow duplicates: UpperCase Foo and lowercase foo
- Exceptional enough we can ignore it?
- If we're looking at unzipping packages elsewhere we should look at extractcode or something else
- Jeff is rolling this out, so there may be some quirkiness in the data over the next few days
- We also need to fix a few curations, those may have wrong casing
- Amz prod crawling
 - We'd like to point to Google and Amazon as running prod versions of the crawler
 - Prod setup for the crawler on aws
 - Jacob will followup with James and dig into autoupdates
- Vancouver
 - Next week is the conference OSS NA
 - Jeff is giving a talk on ClearlyDefined
 - Would like to have things implemented for showing at the conference by Wed
 - James will be at OSS NA
 - Brian from SAP will be there as well
- API for hackathon
 - API for checking if packages are in ClearlyDefined and then providing the ability to curate on ORT
 - Plan to do it on Friday during hackathon
- Scoring
 - https://docs.google.com/document/d/1o_VuVnP1PMoYOXZHSm0TyDX4CsqDXM3zqESAE0AoUnQ/edit#
 - Proposal #4 is the current proposal for implementation
 - Curated vs "inherent" score

August 13, 2018

Attendees:

- Jeffrey
- Carol
- Craig
- Simone
- Gianfranco
- Alex
- Mark
- Jeff
- Thomas

- Philippe
- Valtteri

Organizational agenda:

- Thomas will be visiting and we'll be doing a hackathon on ClearlyDefined and ORT in Redmond on August 24. If anyone wants to join and is around let us know and we'll send you the details.
- Philippe wants to organize a call about scoring

Technology agenda:

- Update on scoring
 - Updated scoring document with a new proposal
 - https://docs.google.com/document/d/1o_VuVnP1PMoYOXZHSm0TyDX4CsqDXM3zqESAE0AoUnQ/edit#
 - Progressive proportional elements consistently
 - Doing scoring directly on service side?
 - May not be able to use scancode computed score?
 - Jeff refactored code because we need to compute score prior
 - We'll follow up on call
 - Philippe is putting together a list of popular packages to compute scores on and to check about top popular maven packages
- Update on full details page (Simo)
 - Friday: released the read only version
 - Noticed some issues and are working on fixes
 - Working on editing experience
 - Content name of schema changing: look for both of them, there will only ever be one
 - To get content run the crawler; we should rerun the crawler on harvested data after Jeff's changes are merged
 - Should be able to delete content of harvested data or point to a different file and that should give you content
 - Editing experience is most important; prioritize that
 - Fixing issues with read only mode and then moving onto editing experience
- Discussion on how to improve the Developer Experience (Gianfranco)
 - Ideal experience for someone wanting to contribute
 - Helping people understand how the repos work together and find something to contribute
 - People not understanding how to get started?
 - Thomas will connect folks with Thomas Graff from Siemens
 - Folks need help on understanding schema?
 - <https://docs.clearlydefined.io/clearly>
 - <https://docs.clearlydefined.io/code/architecture>
- Including raw file content in harvesting

- Jeff spent some time refactoring code so we can correctly harvest
- For licenses, etc we want the actual files for multiple use cases
- Jeff has two PRs, one in crawler and one in service
- Fixed things along the way; if folks could look at PRs that would be great
- ScanCode will have tagging for things it's detected
- "What's an interesting file?" LICENSE, NOTICE, NOTICES...? Contributors? Authors?
 - Philippe will follow-up with files for ScanCode
<https://github.com/nexB/scancode-toolkit/blob/develop/src/summarycode/classify.py#L67>
 - Copying, *.rst
 - CHANGELOG - readme and license or copying
 - THIRD_PARTY, 3RDPARTY
- Source could potentially link back to the commit
- Proposal for ORT/ClearlyDefined hackathon
 - Tooling or redirect API on ClearlyDefined for curation
 - Handshake for whether the package exists on CD, if it needs to be harvested, and then if it's available for curation, a link to the page for contributing
 - Harvesting is not really meant for responsiveness - no telling when it will be done and hard to tell *when* it's done
- Linux packages
 - How to construct coordinates for Debian/Ubuntu?

August 6, 2018

Attendees:

- Carol
- Simone
- Alex
- Jacob
- Philippe
- Gianfranco
- Jeff
- Thomas

Organizational agenda:

- [Monthly board report for OSI](#)

Technology agenda:

- Full Details View - Update (Alex, Simone)
 - <https://xd.adobe.com/view/1fef18d7-16f1-4264-57f3-75cb6d1cf9ac-69de/>
 (Prototype Mockup)

- Should we have the contribute button on deep link page? (currently on the browse page)
- Full Details Page should have contribute (not View)
- Popup for copyright editing shouldn't have line wrap
- Undo dialogue?
- Session storage currently (closing window won't produce the same data)
- Verify issue 99 - do a change, push to master, let it go to dev and try on dev.clearlydefined (had problems getting rid of that agent) - Jeff is closing
- When we have actual content for one of the files: not done yet, but open content in new link/url (file has a "content" property)
- Each file has a list of facets it's in; licensed section has a list of facets; in each facet it has the number of files and aggregate set of discovered
- Schema:
<https://github.com/clearlydefined/service/blob/master/schemas/definition.json>
- Each facet will end up having two scores: "inherent" and "effective"; need to come with a high-level way to convey "state" on component as a whole
- License, copyright, and facet editing is the priority - we just need something that works end-to-end
- Jacob and James have been working on getting a deployment in aws
- Philippe working on ClearlyDefined score document
 - We are implementing inherent and effective score by Wednesday
- Thomas is building in ORT an exclude filter
 - Scans have license false positives
 - "Exclude with reasons" for test code, etc
 - Line up with ClearlyDefined
 - Thomas will post to mailing list
- Philippe is adding auto-generated facets from ScanCode scans

July 30, 2018

Attendees:

- Jeff
- Carol
- Teju
- Philippe
- Alex
- Simone
- Thomas
- Jacob

Organizational agenda:

- Conferences coming up for the fall
 - We'll be at OSS Vancouver - will anyone else be attending?

- Submitted for OSS Edinburgh
 - Other conferences to consider?
- ClearlyDefined funds update from OSI

Technology agenda:

- ClearlyDefined score review with Philippe
 - Two scores that are probably interesting - effective score (inherent score + curations) and inherent score
 - Different scores for Described, Licensed, Secure, Localized, Accessible (maybe aggregate score?)
 - Philippe will update on the second iteration of the score
 - Steering away from judgement about components
 - Thomas and Philippe will connect
- Infrastructure steering committee
 - Jacob and James working on getting it running in their cloud
 - Investigating how to handle secrets
 - Jacob and James to follow-up after investigation
 - By this call next week we'll have a plan
- Full contribution workflow
 - Goal for the next month is getting the full round tripping of data that goes into a definition and could be contributed
 - Need full details page
 - File list view
 - Need it all together in a flow to define facets and get through a flow
 - Demo of the flow at OSS Vancouver
 - Simone is working on mockup branch
 - Come up with some mockups and share with community
 - All the APIs should be available
 - Use ant-design component for new UI?
 - File list *and* full view in ant-design?
 - Prioritize but likely both so we aren't using too many libraries
 - Iterate daily on mockups and have something running in very near future
 - High frequency
- Jacob is working on splitting documentation generation stuff off the attribution builder

July 23, 2018

Attendees:

- Tyler
- Carol
- Alex
- Jeff

- Thomas
- Mark
- Philippe
- Simone

Organizational agenda:

- ClearlyDefined update from Thomas
 - All day workshop at Siemens
 - Presented about ClearlyDefined
 - Many folks at Siemens working on open source
 - Feedback on too much documentation - don't know how to get started
 - Thomas will work with them
 - Siemens is looking for another data source too
 - Showed web interface for ORT (got feedback and needs to make some changes)
 - No integration with ClearlyDefined and ORT yet
 - Forming TODO Group Europe?
 - Lots of folks building tooling and being active
- OSCON ClearlyDefined update
 - Monday hackathon largely on ClearlyDefined
 - Launchpad ecosystem integration
 - Google folks working on getting crawler running in Google Cloud
 - Amazon folks working on AWS crawler
 - ClearlyDefined will soon be running in all the major clouds
 - Storage all in one place, centralized
 - Enable companies to donate compute
 - Bug fixes, doc fixes
 - Interest in having TODO Group chapters

Technology agenda:

- <https://github.com/clearlydefined/website/issues/194>
 - Where to put implementation?
 - How much work has already been done?
 - Emphasis is on full detail, no so much popup
 - Parts of full detail page are in process
 - Carve up into discrete chunks
 - Defer to the end
- File list
 - Completed for read-only view
 - Problem with the unexpanded feature of the component
 - Sorting or filtering data?
 - Main functionality has been completed
 - Jeff to test today or tomorrow
 - Initial impressions are good

- Thomas and other folks are using ant-design (bundle of other npms)
- Simone will look at ant-design and see if it's helpful
- Tree component (tree table? Not sure)
- React table is the one we're using and has tree table capabilities
- Open an issue on React table repo and ask about sorting?
- In discussion with React table developer
- Expanding and sorting seems to be the big issue
- More useful widgets than bootstrap - do we have general interest in making the UI as a whole like this
- Google material design?
- Thomas looking at a bunch of React frameworks - went with ant-design
- Thomas will share list of other frameworks he looked at
- Ant-design is styleable (CodeScoop is using this as well and has been making customizations)
- Localization already built in to ant-design
- Simone and Alex will look at ant-design and will provide input
- We may have other UI challenges that are addressed by ant-design?
- <https://github.com/ant-design>
- Details page
 - Users see something in the list and want to get full details
 - Provide a page with all this data - quite a bit of information
 - Several chunks of data to be partitioned up
 - File list
 - Graphic - layout
 - Facet definitions
 - Other widgets that show simple things/raw data and curations
 - Inspect page right now has three chunks of this data - Curations, Definition, Raw Data
 - How to better present this data? Some of them maybe not important?
 - Tyler expressed interest: open source Fridays every other week; 3-4 hours of work (ping Discord and see how things are going)
 - Simone and Alex - expand/collapse; once the sorting issue is resolved this bit of work is done; look at ant-design and make a recommendation
 - Carol and Jeff have other things to follow up about as well
- FOSSology update
 - Dong Ma has been doing integration for FOSSology
 - Got nomos running independent
 - FOSSology folks seem to think the database isn't needed (code expects a database)
 - Work in progress?
 - Won't put a lot of effort into making FOSSology into working stand alone, we don't plan to manage a database
- Jacob working on making oss-attribution-builder independent

- Their attribution generator is meant for describing your product and generating notices files
- Breaking out notice generation part
- Have ClearlyDefined integration for that
- AboutCode toolkit is an attribution notice generator
- Gets data from a bunch of json files which are ScanCode-looking or yaml files
- Collect license text in ScanCode files?
- Currently have work items in crawler - some files we want to collect are license-related; have work items for identifying and collecting files (if some of them come from ScanCode that's great)
- License text feature is available in ScanCode
- Philippe will point Jeff to the docs and info about it
- Update on License Score definitions/computation from Philippe
 - <https://github.com/clearlydefined/clearlydefined/issues/61>
 - https://docs.google.com/document/d/1o_VuVnP1PMoYOXZHSm0TyDX4CsqDXM3zqESA0AoUnQ/edit#heading=h.I716vckoluoc
 - Want to have a separate focused meeting on ClearlyDefined scoring
 - Review the ticket and doc offline and comment on it
 - Goal is for the ClearlyDefined scoring algorithm
 - Jeff will look it over and provide feedback
 - Computation of score is moved into service, part of definition when you call API
 - We have a central place where computation is done
 - Call this week about it? - will schedule
 - Philippe, Thomas, Jeff
 - Luis Villa
- Update on ScanCode summarization (copyright, holders, license expressions)
 - Many updates to summarization
 - "More straightforward summarization" for copyright holder, statements
 - Need to update the code on ClearlyDefined side
 - When it's available Philippe can point out changes - open an issue in service when it's public describing changes
- Maven package popularity
 - Philippe has reached out to the company that has the keys to maven central
 - Provide a list of top 10,000 most downloaded packages
 - No available, reliable way to determine this
 - Philippe in the meanwhile has been using proxies for getting the most popular ones "unscientifically"
 - Should have a reply soon

July 16, 2018

Attendees:

- Philippe Ombredanne

Technology agenda:

- ClearlyDefined Score - Philippe posted a proposal for scoring. Looking for feedback on this. See https://docs.google.com/document/d/1o_VuVnP1PMoYOXZHSm0TyDX4CsqDXM3zqESAE0AoUnQ/edit#

July 9, 2018

Attendees:

- Carol
- Jeff
- Thomas
- Valtteri
- Mark
- Jacob
- Teju

Organizational agenda:

- OSCON is next week

Technology agenda:

- UI email from Valtteri
 - New a separate UI for curations workflow
 - Listing for items you need to look at (open, what others have looked at) with filter and search
 - See the details of the contribution, mark okay, rejected, approved, comment, etc.
 - Inspect tab: replace or augment? Supposed to become the “full details view”
 - Balance between adding onto GitHub vs. using what GitHub is already offering?
 - For lists of PRs, when we get to scaling issues we will need to enable views on clearlydefined.io because GitHub doesn’t offer us that
 - Reviewer can look at a contribution on clearlydefined.io and see all the details of the contribution and see the changes highlighted
 - GitHub has filtering already
 - Status checks on PR + use Checks API for a richer experience
 - May at some point want “merge” button on clearlydefined.io for Curators eventually
 - Validations: people can very easily enter bad data unintentionally, we should pre-load licenses and other data fields

- Valtteri: working on pickers for facets, licenses, URL for source location (get to a commit), etc
 - Have a lot of this functionality in the Harvest UI
- More crawler instances going: Jacob and James
 - Exploring getting an image and set of secrets and we'll let you run yourself
 - Talk next week at the hackathon
- OSS Attribution Tool backed with ClearlyDefined data
- Thomas: SPDX and JSON formats
 - Meeting this week with Kate, would like a "minimal representation of ClearlyDefined with SPDX"
 - Jeff to send Thomas an email with what he needs
 - At least the ability to output a definition in SPDX format
- FOSSology: want to find out how to run without Postgres server running, etc
 - Want to run nomos, ninka, copyright too as individual things in ClearlyDefined. Just looking for scanners/tools
 - Don't want the database integration
 - Need to resolve over Discord and in the issues
 - Michael/Oliver to enable FOSSology to run standalone?

July 2, 2018

Attendees:

- Valtteri
- Thomas
- Jeff
- Carol
- Daniel
- Teju
- Jacob
- Rashmi

Organizational agenda:

- Funding for ClearlyDefined at OSI

Technology agenda:

- Developer work for UI projects
 - File list
 - Engaged a developer to work on this in the next few weeks
 - Building on the work Jacob has done
 - Encouraging them to work in the open as well
 - UI
 - Sorting and filtering component list

- Collapse all/remove all button
 - Small UI tweaks for robustness
 - Validation and date picker
 - Still possible for people to add invalid information; data cleanliness work would be good for us to focus on in the next couple weeks (higher priority)
 - Would like someone other than Teju to work on this
- New package managers
 - Teju will start working on new package manager support
 - RubyGems and PyPi
 - Starting this week
- Full details page
 - We currently have an inspect page with ALL the details of the components
 - We want a details page that is more attractive and approachable
 - File list will show up on this details page
 - Wireframes from STRV will go into the full details page (Carol will check to make sure these are there)
 - Once the current file list work is underway we will look for some collaboration with the community
 - Valtteri and Daniel to coordinate on work on this
- Daniel will be working on this with us offline - some limited time to figure out a task that's doable in 4-5 days. Valtteri has another colleague who should be available later this week
- Would like Valtteri and Daniel to take ownership of the validation/rich entry stuff; submitting a source URL, declared notice, copyright statements, lists of license expressions we need rich components for that contribution
- New harvesting and scanning
 - Top 10,000 npms, top 10,000 nuget
- Reminder about GitHub projects and roadmap
 - <https://github.com/orgs/clearlydefined/projects>
- FOSSology
 - JSON output - Daniel has made an integration for this already
 - We have someone looking at integrating FOSSology as a scanner - Dong Ma working on this
 - Contribute the code Daniel has already written?
 - Also tried to spin up some conversations with Oliver and Michael re: integration
 - Daniel and Valtteri will collaborate with us on this
 - JSON format - not yet part of the standard way the FOSSology works
 - Do output some kind of JSON?
 - Need to run multiple scanners
- SPDX and JSON format
 - Wanted to understand the current state of SPDX and JSON
 - We'd like to have support for outputting SPDX forms of our definitions

- Understand how our definition data schema maps onto SPDX (changed our schema, that needs to be remapped)
- Output JSON would be easiest
- SPDX support for input is also interesting but we may have to visit RDF SPDX files
- Would love input from Thomas on this offline
- Jeff will create some issues for the SPDX project that captures this and we can collaborate on this

June 25, 2018

Attendees:

- Carol
- Jeffrey
- Jeff
- Teju
- Michael

Organizational agenda:

- [Roadmap](#) and [GitHub Projects](#)
 - Work items to go over with the group
 - Now have a backlog
 - Have a couple potential developers to work on ClearlyDefined
- OSCON/TODO Group hackathon - will anyone from this group be there?

Technology agenda:

- Jeffrey has a list of about a dozen components he will send over later to see if ClearlyDefined has that data

June 18, 2018

Attendees:

- Jeff
- Teju
- Jacob
- Philippe
- Valtteri
- Mark Anderson

Organizational agenda:

-

Technology agenda:

- ClearlyDefined Score - Philippe has a proposal for the score. Looking for feedback on this.
 - Keep it as simple as possible
 - How much curation is needed for the package
 - Aggregated stats on multiple packages to see if we have made any impact
 - As a consumer, how much vetting do I need to do to get the confidence that I need
- UI updates -
 - Teju talked about the recent progress in tweaking the UI for contribution workflows etc.
- Philippe evolving scancode license summarization and copyright handling
 - Should we use the copyright statement or the holders
 - Breaking down the holders is hard in general (and gives poor results)
 - Consider using just the "holder statement"
- Valteri and the UI work
 - Looking to pick up some work this work
 - Goal for this week is to get docs and a UI plan together so other folks can execute.
 - @Carol to send invites for the UI design team meeting

June 11, 2018

Attendees:

- Jeff
- Philippe
- Carol
- Teju
- Jacob
- Jeffrey
- Virginia

Organizational agenda:

- [Monthly OSI board report for June](#)
- OSSNA summit
 - Rashmi and Jeff accepted for the conference
 - New topics: harvesting and workflow topics. Show off we have lots of stuff harvested and show the contribution workflow
- OSCON TODO meeting and hackathon
 - Quite a bit of interest in a ClearlyDefined hackathon

Technology agenda:

- STRV UI work update
 - Started thinking about the curation workflow
 - Want a button to get to a view on clearlydefined.io for what the contributor saw (“What this will look like with this PR applied.”)
 - With a pull request, you could go to the equivalent of the Browse page and then be able to see the three level of details page
 - Want to start getting a curation workflow set up soon and want an easier way for people to do curation
- Data dump questions - Jeffrey
 - Wanted to see all the data per components in .xls or .json format
 - It's arbitrary json and someone would have to figure out how to put that in tabular form
 - Alternative is to figure out of blobs and give you 100,000 json files
 - We have 50-60,000 components harvested, a lot of top lists, etc
 - We're happy to have a mechanism for a data dump, we're just not there yet
 - Trying to determine if they can use the data
 - If you have components you're interested in comparing just send the list and we'll queue them up to harvest them
 - Is the list of components available some place? Via the API, but no we don't have a list you can just “browse”
 - Generate read-only SAS token to look at the data store?
 - Jeff will look at generating read-only token for looking at the data
- Contribution workflow update
 - Still evolving down the path we talked about
 - Have three levels of details
 - 2 line summary
 - 6-8 line summary
 - All the detail
 - Where we expose the per-file information
- Schema update
 - A lot of discussions about the schema for definitions
 - Final set of PRs Jeff put in
 - Brings us to have per-file copyright and license information in the schema
 - Lays groundwork for having independent editing of file and facet editing
 - Had a call with UI designers and we have another meeting this week
 - Progressing this week
 - PR also has some changes to implement everything in a complete way
 - Still on the table: whether attribution data stored per file is copyright holders or copyright statements
- Scancode and latest version
 - What's the deal with the latest version?
 - Working on summarization and faceting

- Philippe has a branch with copyright holder, copyright summarization, not really happy with the way it's reported in the scan
- Ping Jeff on the ticket? Sure
- Summary data at the top level to provide counts, copyright holders, licensing, and programming language
- Counts broken down by facets if available
- Difference between declared and discovered: the license file found at the root of the repo is the one that will be considered "declared." If you have your license file in an src file you'll be less clearly defined.
- Reported as license expression and no longer as just licenses
- Would like to run another round of harvesting
- Harvesting
 - We'd like to do the top couple thousand again and the 10 most recent versions, for example
 - If we did it we'd like to make sure we are using the most current version that's available now
 - Most recent would be 2.9.2
 - Plans to make a new intermediate release by end of this week
 - Full release maybe by end of month?
 - We'll likely pick up whatever is the most recent whenever we decide to do this
- Scoring and metrics doc update from Philippe
 - Suggest we put it in a Google Doc for discussion

June 4, 2018

Attendees:

- Carol
- Jacob
- Valtteri
- Jeff
- Philippe
- Jeffrey
- Teju

Technology agenda:

- [Data definition spec](#) and [schema update](#) PR
 - Please look over and comment as applicable
- Contribution UI work: we need to figure out how to curate lists of files (potentially thousands)
 - Sorting, filtering, etc needs to be enabled
 - AboutCode manager - some elements exist in ScanCode

- Gives you summaries and lets you drill down how you want
 - Tabular views with filters and search
 - Probably ideas that could be taken from there
- Valtteri and Jacob to coordinate on Discord about moving forward on UI work
- Jacob to put together a component that uses React table and lets the user apply tags (facest) and work with the component
- Curation workflow/community is still a green field, but next on our list of things to work on
- We would like to start thinking about the ClearlyDefined score
 - What goes into the score needs to be clearly explained, so we need to start thinking about it and what defines it
 - Presence or absence of licenses
 - Curations add or detract from the score?
 - Number of files that have licenses, etc.
 - UI elements for sorting and filtering on the score
 - If anyone's interested in driving a steering committee and then writing a spec - Philippe volunteered
 - Philippe to put together a proposal for objectives and algorithm

May 21, 2018

Attendees:

- Jacob
- Jeff
- Teju
- Jeffrey
- Valtteri
- Carol
- Thomas
- Philippe

Organizational agenda:

- We have an intern, Teju, that will be working on ClearlyDefined for the summer.
 - Looking at contribution workflow and trying to get it close to being ready to ship

Technology agenda:

- Plot some paths forward on the UI work
 - For contribution workflow, tokens for GitHub had to be worked out; got those worked out after troubleshooting that
 - Mostly on Jeff's plate to do reviews and he hopes against hope to get that done today
 - Scheduled to have a longer meeting on Friday to dig into where we are and how to go forward

- Aiming to have a contribution workflow done in ~2 weeks when our curator is back, trying to give her the ability to start to dig in
- Thomas: things have been extremely stressful; making progress on ORT.
 - Almost done with adding php support, almost done with Bower support, then into bitbake (Yocto). Had an acquisition. Making sure curation stuff works, can accept and process curation.
 - Building an excel output and html output; Thomas is learning React, building their webapp with a button that will link to ClearlyDefined. By end of a couple months to have it done.
 - Taking some of the internal stuff and making it external.
- JSON format for SPDX - delayed by Thomas; working on the spec
 - Working on this now, ongoing meetings on that
 - Gitbook - on their slack channel; under the radar but marked everything as legacy and haven't had any changes in a while
 - Jeff would like to follow-up on that separately
 - Using markdown for now
- Philippe - Scancode summarization on license and copyright, toward version 3 of Scancode
 - Introducing faceting
 - The ones that are in ClearlyDefined will be exactly the same
 - Ideally we'd really like to be exactly the same
 - Two things - being able to properly categorize code and then the notion of facets as they exist in ClearlyDefined
 - ClearlyDefined has gotten feedback that we don't know if folks are going to categorize their code or not
 - Interested in maybe making a simple binary distinction between "ships" and "doesn't ship"
 - Want to come to some understanding of what a facet is, what it's used for, and what the specific cases are, it would be good to get on the same page with this group for ClearlyDefined
 - Thomas can have Sebastian to reach out offline
- Jeffrey
 - Took a look at data
 - Asking about using the UI

May 14, 2018

Attendees:

- Carol
- Jeff
- Jacob
- Philippe

Organizational agenda:

- Data steering committee meetings: Monday, May 14, 2pm PDT and Thursday, May 24, 10am PDT

Technology agenda:

- Contribution workflow work update
 - WIP PRs seem mostly complete as far as getting something out there
 - Jacob should do code review of WIP PRs
 - Should we do multiple PRs per set of changes?
 - Jacob: I'd shoot for one pull request per change
 - Small API change to make it one commit, just have to manipulate the git tree
 - We'll get our intern starting and we'll get her to review PRs that are already written and we'll get the pull request merged
 - Want to enable some our curators to start working on that

May 7, 2018

Attendees:

- Carol
- Jeff
- Philippe
- Rashmi

Organizational agenda:

- This call moving to Discord video call from Hangouts next week

Technology agenda:

- What data we are showing or maintaining, particularly around the notice files
- We talked with Max and Will last week about the definitions and they want to extend the conversation out to this community
 - Attribution parties, something else?
- ScanCode: when you scan something with a notices file, do you call out the content of the notices file and include that in the input?
 - There's an option called "license text" which captures all the notice and license text in every file (raw data as captured)
 - Doesn't include copyright; copyright comes from copyright, license comes from license

- There is code not wired in which calls out any file that is known to be a top-level important “legal” file; code is there but not used
- ScanCode: update to copyright summary reported based on number of occurrences (copyright holders and copyright statements) - pushed those options to a branch
 - License expression summary
 - Need to mix with facets and apply to license expressions and copyright summaries
- Would be nice to capture conventions from communities and codify it, maybe not in CD
- Attempts to deduplicate license text? Not right now. Pending ticket on that from another request.
- Max @ Google is particularly keen on this and is interested in driving and the scenario we’re expecting to support
- We’ll look at turning on -license text

April 30, 2018

Attendees:

- Jacob
- Jeffrey
- Philippe
- Tyler

Organizational agenda:

- Philippe provided a quick update of the FSFE Legal Network workshop: there were a lot of discussion on ClearlyDefined. And we missed Jeff!

Technology agenda:

- Jacob: Some UI work coming up, provided feedback from discussions with Jeff
- Jeffrey: evaluating the tool:
 - How is the auto generated proposed curation?
 - There is no generation taking place
 - Also trying to look into how own data could be compared through API calls with curated data

April 23, 2018

Attendees:

- Carol
- Jeff

Organizational agenda:

- OSI in-person board meeting on Monday, 4/30. Carol will be in attendance and making a verbal report in addition to the [monthly written report](#) on ClearlyDefined. Does anyone have any specific points they want mentioned at the meeting?

Technology agenda:

-

April 9, 2018

Attendees:

- Carol
- Thomas
- Jacob
- Jeff

Organizational agenda:

- Monthly [OSI board report](#) for April due April 12
- Jacob talked to a coworker about possible getting involved in front end or maybe back end devs
- We may want to talk to the Facebook folks again - sounds like they may be interested to talk more at some point?
- CFPs - Open Source Summit North America - Thomas will submit on ORT; Jeff and Carol to talk about presenting
 - Open Source Summit Europe - due in June
 - Open Compliance Summit Tokyo
 - FOSS Backstage - Thomas is presenting in Berlin
 - Thomas is talking this Thursday at German tooling workshop; may or may not be able to talk about ClearlyDefined (cut and paste from the OSLS presentation)
- Jeff is going to legal workshop next week and will be talking about curation principles and processes - one of the goals of the workshop is to get feedback on what the legal folks think is important
 - If anyone wants to jump in and
- Carol will cancel next week's meeting

Technology agenda:

- Jacob has some time for the workflow for contribution
- Thomas working on "report of components" - finishing drafting mockups and adding a button for ClearlyDefined
 - Planning to have built by end of May but was somewhat delayed
 - Need to look into logistics and workflow and how it would work
- Contribution workflow discussion

- Confusing as a user to know what changes I've made and what happens to them
- New search - what happens to the old changes?
- Cart or basket of changes - as someone makes changes it queues up pending changes and displays to user
- Once you make changes we track them immediately in the browser
- How to show the differences when you make lots of changes?
- Showing specifically the items you've changed
- In Browse you have the affordances for adding to the list and then show "only the changed things"
- The ones you've changed we can display a UI element to show which ones you've changed
- Components will highlight themselves if they've changed
- Prototype? Jeff may have some time to work on this this week
- Needs on server side:
 - Searching - when you search right now it's by name, now we should implement some ways to search by things other than name
 - Jeff just implemented Azure Search functionality for server side search
 - License, copyright holders, release date, and coordinates (name, etc.)
 - We're indexing definitions, which are the results of curation and tools all aggregated
- Plan to talk more later this week - Jacob will let us know when he thinks he'll be ready with some more changes

April 2, 2018

Attendees:

- Carol
- Rashmi
- Philippe
- Jacob
- Michael
- Jeffrey

Organizational agenda:

-

Technology agenda:

- Update on crawled components queued up and harvested in the last couple weeks
- Contribution workflow work this week - Jacob, Carol, and Jeff
- Philippe having a sideline conversation about scanner with Gene and purl questions with Jeff

- Philippe will review scans with Gene for any bugs or issues
- Jeffrey would like to look into the harvested data on the website; still running into bug with website not displaying data. Interested in comparing data with their “pedigree data” at SAP.
 - Maybe going to talk with ORT/Thomas about their dependency work? Probably going to talk with them offline re: dependencies

March 26, 2018

Attendees:

- Carol
- Jacob
- Rashmi
- Philippe
- Jeff
- Michael
- Jeffrey

Organizational agenda:

- Legal and licensing workshop in Barcelona - one proposal was accepted; about curation and the process

Technology agenda:

- [Sprint board proposals](#)
- Contribution workflow work this sprint (Carol and ?)
- Integrated new version of Scancode - queued up a bunch of random NuGets, mavn, npms, and GitHub repos. Starting to gear up to harvest a whole bunch more stuff.
 - Would love to see more implementation for other protocols
- Jacob - can have him review more PRs, going to try to give the website more love
 - Talk more about design for contribution workflow UI design-y stuff
- Rashmi - Friday is open source Friday and they have a little more time to work on things if we want them to
 - Good first issue or help wanted issues they can work on on Friday
- Philippe - Carry over for Scancode and other items from previous week; traveling this week and will have a little less time
 - Check any issues with Scancode detection on the harvesting Gene has been doing
 - Any case where there isn't 100% conclusive detection of the licenses - could review for curation
 - There's a bug on Scancode open that Philippe could look at this week?
 - Errors in package manifest will not be treated as errors

- Philippe will create a ticket on facets to discuss some corner cases (e.g. generated code). Jeff wants to keep the set of facets as simple as possible which makes sense.
- Jeffrey is installing CD locally to see real curated data to compare with actual curated data to better understand formatting, depth, direction on curation.

March 19, 2018

Attendees:

- Carol
- Jeffrey
- Jeff
- Thomas
- Philippe
- Michael

Organizational agenda:

- Submitting for CFP for Open Source Summit North America. Other summer conferences we want to submit CD to?

Technology agenda:

- ORT? - Everyone was sick; tons of minor improvements to overall detection.
 - Fixed some Scancode issues, but mostly around dependency combinations.
 - Problems with maven modules.
 - Trying to move to technology agnostic representation.
 - Started working on adding Go support.
 - Fixing existing things and adding Go to see if they can get the model to work.
 - Also trying to improve documentation
- Philippe has done quite a bit of review on top npm scans on Scancode. Will push that and new tests in a new branch.
- Pushing semver changes
- Copyright summarization - Philippe is pulling a csv with two columns, copyright and corresponding copyright holders.
 - 100 of the most salient examples.
 - Helpful for ClearlyDefined to have a way for people to understand how we came to the value.
 - "Paragraph of text with 5 or 10 examples."

- When will the next non-beta version of Scancode will be available? - Philippe to commit to having it by Wednesday (still beta)
 - Missing some small tests, documentation, as good as production release for ClearlyDefined
- Purl? - Philippe will reply about purl ticket and will reply. No issue, not controversial.
 - Jeff will look at making that happen in CD.
 - Anyone interested in a javascript version of purl? No one who's shown up to say they'll do it. Philippe could do a transpile from python to javascript. Jeff: that would be useful to save us time on doing it manually.
- Jeff: we've been doing some work on the server side, fixing some bugs, running through the 1000s of maven and npm and checking things that go to a dead letter queue.
 - Gene on our team has been making progress with that
- Curation/contribution - Carol has been working on a contributor workflow/process
- Philippe has talked with Mike - has a todo to talk with Wayne and Sharon; need to loop in Carol and Jeff about the data

March 12, 2018

Attendees:

- Philippe
- Rashmi
- Jeff
- Andreas
- Jacob
- Carol
- Jeffrey

Organizational agenda:

- Starting a new conversation with interested folks about the curation community. Who would like to be included in that?
 - Rashmi
 - Philippe
- [Monthly OSI board report](#)
 - (Philippe) OSLS presentation by Jeff and Rashmi was superb

Technology agenda:

- [Sprint review/planning](#)
- (Jeff) Implementation:
 - Getting the service more solid
 - "Contribute" workflow is not very friendly. Start thinking about what a contribute workflow would look like.

- People should not see a “file of stuff” - gives them a workflow in order to make a contribution easily
 - Rashmi will talk to Mark and see if he has interest
 - Maybe another tab called “contribute” with UI for someone to contribute. Scenario where we don’t already have anything about a component so that we don’t have to *have* harvested a component in order to add things.
- (Philippe) Review in details SPDX data concepts and how they align with CD.
 - Take some of our definitions and map them onto SPDX documents. SPDX as JSON would be good to compare. Lack of easy access to the definition of data elements.
 - Jeff has a thread with Kate re: some of the mapping
- (Philippe) Definition of facets. What would be the few essential facets we want to carry around? Planned supported facets?
 - Facets are in docs - Jeff will find them and make sure folks are aware of them.
- (Philippe) Pushed Scancode copyright holders change - probably some need for some refining. Also working on summarization for licenses.
 - Docs on “what would be chosen” - longest name.
 - Algorithm from Google developed for freebase - openrefine, tool to clean up data
 - Recognize named entities, normalizing the names with some rules
 - Philippe to send some pointers to other examples of that
- Capacity:
 - A little less of Jeff and more of some other folks on the team
 - Jacob - a little bit more time, possible work on website
 - Philippe - a bit more time.
 - Working on package manifests for Scancode and if this is interesting for CD.
 - Consider using package URL in CD? Take that offline.
 - Jeffrey - a day or two; SAP has interest in contributing curated data. Want time to touch base on contribution. Set up a meeting time.

February 26, 2018

Attendees:

- Carol
- Tyler
- Max
- Rashmi
- Jeffrey

Organizational agenda:

- Current charter language [here](#). Proposed change to charter language for voting process:
 - *A topic for voting is tabled by a curator by notifying all other curators.*
 - *Once tabled, curators may vote during an open voting period lasting no less than one working week. Voting will occur on an agreed to, mutually convenient, and open medium (e.g., email, GitHub issue, etc.)*
- Adding a contributor ToS: “*Whenever you make a contribution to a ClearlyDefined repository containing notice of a license, you license your contribution under the same terms, and you agree that you have the right to license your contribution under those terms.*”
- Logo and typography are in the repo

Technology agenda:

- One week to OSLS - what's everyone working on this week?

February 19, 2018

Attendees:

-

Organizational agenda:

- Current charter language [here](#). Proposed change to charter language for voting process:
 - *A topic for voting is tabled by a curator by notifying all other curators.*
 - *Once tabled, curators may vote during an open voting period lasting no less than one working week. Voting will occur on an agreed to, mutually convenient, and open medium (e.g., email, GitHub issue, etc.)*
- Adding a contributor ToS: “*Whenever you make a contribution to a ClearlyDefined repository containing notice of a license, you license your contribution under the same terms, and you agree that you have the right to license your contribution under those terms.*”

Technology agenda:

- Sprint review/planning - last sprint before OSLS!
 - We have marked the issues with the milestone osls-demo in the [crawler](#), [service](#), [website](#), and [organizational](#) repos that must be done in the next two weeks.

February 12, 2018

Attendees:

- Carol
- Michael
- Rashmi
- Thomas

- Jeff
- Jacob
- Philippe

Organizational agenda:

- 3 weeks until Open Source Leadership Summit - need [presentation](#) done likely by end of this week so we can review and begin practice. Scheduled time with Jeff and Rashmi, but we'd love input on the content and structure in the meanwhile.
 - Thomas will be adding some content and got a lot of feedback
 - Workshop with BMW, Bosch, Siemens, Canon. Tooling workshop about tools available. People liked the idea of CD and are really concerned with giving back to the community ("not creating a database for a database").
 - How does this work with Software Heritage? Are we working with Software Heritage? Came up twice about how we relate with them. How will the data be available? Liked the ORT integration. (Call from ORT and schedule into CD)
 - Would like to see clear diagram page with blobs and scanners. How the various scanners can be integrated. Thomas will update the slide on that.
- [Press release](#) will be formally from OSI, needs to be sent to them this week for review and sign-off.
- FOSDEM trip report
 - General reception was positive. Some skepticism, but overall general interest and reception was good.
 - Discussion with Jonas and John (FSFE and FSF) - want to proceed and hear more.
 - Item from Philippe: (from Richard with Red Hat and Bradley from SFC) we should debrief and reassert our agreement on the importance of upstreaming. This has eventually a profound implication not only on our communications but also on any tools we create.
 - Max Sills from Google was also positive.
 - Philippe has also talked with PM from GitHub around license clarity.
 - Leverage and contact and work with package maintainers for package management tools and package repositories - npm, pypy, ruby gems, maven. Likely making pushing upstream vastly simplified. Philippe to present ideas on how to drive that and to talk with them about that.
- We got a request to have a public discussion on a topic when we were at FOSDEM. Should we have a public mailing list for public discussions? Make clearlydefined@googlegroups.com public?
 - Review topics in the current mailing list and review if people are comfortable with that becoming public (since people didn't have the expectation it would be public when it was done).
 - Start a discussion through an issue in GitHub for this topic? Is the issues space open for anyone? People may naturally gravitate towards that.

- Item from Philippe: we should IMHO move/create to a public Gitter chat and make the mailing groups all public too. And be all public forward. (+1)
 - Walled garden - have to sign up and can't do that without someone moderating your membership; can't make a channel public
 - None of the content is public
 - Have to be accepted as a user to participate
 - Gitter is open source

Technology agenda:

- Dev time for the next 3 weeks in order to get to demo/announcement; critical path items need to be assigned
 - Dive into ORT and ClearlyDefined integration - Thomas/Jeff/Martin (demo). Martin got something working for npm. Thomas to ask him about the integration. Jeff to email Thomas and Martin.
 - Got some contributions from Qualcomm.
 - Otherwise "the functionality we have now, just more robust."
 - Website work - buttons hooked up, etc.
 - Minimal service, use case we can support? We have the one-pagers to attract people. From there, curation? Not a lot of effort put into the curation process to date. People who want to engage on curation are the people who will define how the experience works.
 - Capture curation requests? (Enter a URL, gets into a queue,...) Do we want to set up some expectation that people entering URLs that "anything will happen."
 - Set up a way for people to express interest in a package that they'd like curated. (Or have them enter issues, which may or may not be a way to manage the work)
 - Jeff: What's the goal in asking people what package they want to have curated? Philippe: It's a way to drive and direct the project. Jeff: instead we should let the people who show up to do curation drive the packages that get curated instead.
 - Sponsorship of resources from other companies/places - maybe addresses the compute time and managing requests.
 - For the March timeframe - top 1k maven packages and npms - run ScanCode on all of those and try to curate some of them to give people a sense of what the project is about. Being able to address whole ecosystems to help people understand.
 - Philippe to make the repo for curation requests and let people register interest and make it clear in the wording that it's just a way to register to interest
 - Node and React devs! We need them! We need to get some of this done.
 - We're going to have to deploy production version of this, which will take a few hours. We just need some dev time to get these things done.
 - Go through the issues and mark them as required for OSLS - Rashmi can try to get more devs to help through March - Jeff and Carol will take some time to do that.
- Logo/typography/website update

- Jeff will email his scribbles for a potential logo out

February 5, 2018

Attendees:

- Jeffrey
- Jacob
- Michael
- Philippe

Organizational agenda:

- FOSDEM
 - Meeting/talk on Friday; plenty of discussion with individuals there about CD. Planned many meetings (Jeff/Carol to debrief?)
 - Upstreaming should be a primary goal (via Richard Fontana, Bradley Kuhn). Perhaps upstreaming was not being displayed clearly (pun) via ClearlyDefined goals. For launch, should we try out a workflow with one or two projects to see how we can upstream curated data.
 - Philippe chatted with Jonas Oberg, FSFE director and he is interested; they have “software.reuse” site with a related goal (Jonas wants to join!)
 - Philippe chatted with John Sullivan, FSF director: he wants to hear more. his pres at FOSDEM was about clear usage of licensing. Make it clearly available, don't fall into the trap of not using free software by misreading licensing
 - Libraries.io catalog of packages, dependencies, licensing
 - GitHub product manager on code review. Leverage: can we work things out with GH to easily present this info to authors or public? Establish relationships with PyPI, NPM, RubyGems, etc.

Technology agenda:

- ORT integration
- Have a look at: <https://github.com/clearlydefined/clearlydefined/issues/24> - revisit issues & pick tasks for current sprint

January 29, 2018

Attendees:

- Carol Smith
- Jeffrey Shi
- Rashmi Chitrakar
- Philippe Ombredanne
- Michael Herzog

- Jacob Peddicord
- Jeff McAffer

Organizational agenda:

- [OSI board report](#) for February 8 meeting
 - Might be worthwhile to provide a one-page update if we have enough content
- I've been building out a generic [overview document](#) to facilitate conversations about ClearlyDefined with people who are new to it. Please take a look over it and add comments and changes. We'll put this on the wiki once it's finished.
- I took the deck Jeff created for the project in December and replicated it for start of the [OSLS deck](#). It definitely needs to be updated, so please take a look and add in changes so we can get a start on the presentation. We can use this as the base for a generic deck anyone can use once it's done.

Technology agenda:

- We should update the [architectural diagrams](#) to reflect the current technology for folks to use when they're talking about the project. I've also like to use this diagram in our overview document.
- Review logo/design options from design team
 - We will share options in an offline format.
- Confirm Brussels hackathon [technology goals](#) for Friday
 - Getting ORT and ClearlyDefined onboarded. Jeffrey and Martin are working on getting that together.
 - What particular library to use - requires RESTful client.
 - Don't think we're going to web sockets right now.
 - Jeffrey to talk with Martin and Thomas to align.
- Re: SPDX
 - We want to get and store SPDX
 - Important for SPDX to be a first among equals and we can interop with other tools
 - Kate is interested in working with us to see how this can all go forward

January 22, 2018

Attendees:

- Carol Smith
- Jeff McAffer
- Philippe Ombredanne
- Rashmi Chitrakar
- Jacob Peddicord
- Jeffrey Shi

Organizational agenda:

- OSCON CFP is also coming up and due January 30. Take a look over that [CFP draft](#) as well and let's talk about OSCON.
- Do we want to participate in GSOC?
 - No - organization isn't mature enough and proposals are due today without enough time to prepare
- Hackathon - please send Carol a confirmation of their availability for dinner

Technology agenda:

- [Sprint review/planning](#)
- [Roadmap](#)

Thomas Steenbergen update: can join meeting due to scheduling conflict but hereby my update for Jan 22 following up on open items:

- Improve copyright detection ScanCode: Philippe is performing a major refactor of ScanCode to make it more modular which is awesome but at this moment codebase is not stable for us to start implementation. Will sync with Philippe
@Jeff: can you provide some examples so we have test case for when we start implementation?
- Curations format: Martin Nonnenmacher on our side will write email to Jeff and William today with how we in ORT have current implemented curations. We would like to see how we can align on single format between ClearlyDefined and ORT.
- Tomorrow Jan 23 there a call between Jeff, Jeffrey, Sebastian and myself to discuss ORT integration to ClearlyDefined
- Personally working on SPDX 2.1.1 release e.g. first release from GitHub so afterwards we can start with 2.2 in which we plan to add support for more formats like JSON and YAML. This will help ClearlyDefined as it makes merging the results from multiple scanners easier (as all follow the same standard)

January 11, 2018

Attendees:

- Carol Smith
- Thomas Steenbergen
- Rashmi Chitrakar
- Jeff McAffer
- Jacob Peddicord
- Philippe Ombredanne
- Michael Herzog

Organizational agenda:

- FOSDEM hackathon on Friday, February 2? Who will be in town?

- Anyone interested in dinner in Brussels? We have reservations at [Classico La Brasserie](#)
- Contacting TODO Group [email](#)
- Update on other people interested in the project (SAP, Bosch, Apache, Eclipse, ...)
 - Philippe can reach out to Oliver Fendt at Siemens; Thomas is meeting with Bosch, Siemens for a workshop mid February
- Built out the wiki on the organization repo a bit. [Take a look](#).
- Open Source Leadership Summit CFP is due in 10 days. Can everyone look over the draft of the [proposal](#) and please edit as you see fit?
 - Rashmi and Jeff as presenters?
 - Potentially business/technology split as two sessions?
 - Jeff/Philippe high-level presentation?
 - Thomas on the technology and/or curation flows?
 - BoF or working session with people in tandem with the announcement talk - Rashmi
- Update of OSI incubator proposal graduation language

Technology agenda:

- January hackathon is next week. Agenda is [here](#). Let's make sure scope for the hackathon is settled. Anyone else attending in person?
 - Potentially set up some parallel hacking on organization and technology to make progress on both
- Discuss logo/typography design for website?
- Slack for communications?
 - Have the conversation over email - proposal of gitter
- ScanCode perf. How best to run ScanCode frequently on small things infrequently on large things
 - "Shouldn't take 20 seconds for 3 files"
 - Scan speed is roughly linear to the size of the files and the number of scan options, speeds up pretty linearly with the number of processes
 - Multi-process it caches on disk (using block storage vs. laptop?)
 - What would you expect time to be for something small? 1-2 seconds.
 - Scan the whole linux kernel in ~20 minutes
 - Try running on the typescript repo (we got 30 minutes?)
 - Philippe can give us baselines on typescript or react and some smaller ones
 - Optimizations coming up to skip certain types of files e.g. binaries
 - Email for the implementation team
- Architecture proposal around running scans (using builds)
 - Forking the actual run of the scan into a build; input is the source to scan and the output is the scan results
 - Queuing a build isn't a bad idea, the people from Travis might not be happy about it
 - Queuing a scan as part of a build is a non-issue

- On Travis specifically the issue is retrieving the results of something that took place during the build; time limit on a given build for a job
- Difficulty is in knowing the size of the repo
- Queueing the build but won't ask Travis to do the clone; docker will do the clone.
- Need a trigger which is typically commit; but you can programmatically trigger a build
- How to handle "donated" data (partially technical, partially organization)
 - Some folks have large volumes of data that they've curated. How can we bring that data in?
 - How do we situate that data relative to the data going forward
 - Scan results or curation data? Both.
 - Two problems: technical; a lot of the scans aren't structured. Extracting that data may be difficult. Also perform scans with context of the organization; don't think the legal team may be comfortable releasing that.
 - Need a validator for curated data that can check the curated data is syntactically correct and then if they are "actually correct" with some sanity checks.
 - For automated scans this probably doesn't matter as much.
- Architecture discussion from Tuesday status
 - See how we can adjust the analyzer for some of the use cases for Jeff
 - Currently set up for depth first as part of a product review
 - Jeff's use case is breadth first and scan, for example, 1,000 npm packages.
 - Talking about using more of a queueing system and a cloud infrastructure system. Some changes to ORT to demonstrate how they envision this use case.
 - Need a queue/worked system. Talk about in the FOSDEM hackathon.

January 4, 2018

Attendees:

- Jeff McAffer
- Carol Smith
- Rashmi Chitrakar
- William Bartholomew
- Thomas Steenbergen
- Sebastian Schuberth
- Michael Herzog
- Jacob Peddicord
- Mark Matyas
- Philippe Ombredanne

Organizational agenda:

- Review [updated roadmap](#) for Jan - Mar, adjust if needed
- Confirm hackathon for Wednesday, January 17. Anyone who can attend in person?

- No strong objections
- [Draft of proposal for CFP](#) for open source leadership summit (due Jan 21)
- Review target partner adopters: Eclipse, Apache, and .NET Foundation (any other potential targets we want to reach out to and meet with?)
 - Sebastian: approach companies that run package repositories? Jeff: absolutely, likely going to have to go back to the original projects. Where are we going to have the biggest bang for our buck for March?
 - Sebastian: Maybe it makes sense to start from a language community? Jeff: no reason not to.
 - Thomas: have to be able to see what we can support with the tooling. Jeff: we're just starting the conversations. Thomas: need to work on the definition of "clearlydefined" mean - to work on in the hackathon. What is the MVP of ClearlyDefined?
 - Jacob: trying to focus on the breadth of projects or depth within projects? Jeff: thought we would be focused on a deeper amount in a narrower spectrum. E.g. the top 1k NPMs. Focus on the R community re: licensing? Overall count of packages in the community could benefit.
- Review [OSI incubator project proposal](#)
- Other docs in the [Organization](#) project need review; assign reviewers

Technology agenda:

- Demos of what's currently in place
 - Over the holiday break some folks spent some time hacking on things and looking at stuff. We've put some stakes in the ground around what this technology could look like. We got to a point where we have a very preliminary e2e story. In the previous hackathon we put a stake in the ground around what would be the core/base data. All this code is currently in the clearlydefined repos.
- Review of website
- Review [sprint board](#); get volunteers for projects assigned

November 28, 2017

Agenda:

- Demo/tools session
 - William Bartholomew, Microsoft: Current metadata service and curation proxy
 - Philippe Ombredanne
 - Sebastian Schuberth
- Naming and license

November 15, 2017

Attendees:

- Jeff McAffer
- Carol Smith
- Sebastian Schuberth
- Gil Yehuda
- Rashmi Chitrakar
- William Bartholomew

Agenda:

- Scope and intent

Jeff: Folks in open source are struggling with a lot of the same issues. There's a lot of process and tools around this, but in pooling our resources we might be able to get more traction and make our jobs easier as a collective.

This is a big, rich area, and so firstly we're hoping to scope it down to some relatively simple goals to make progress on incrementally. To make this a little more tangible, we're proposing a structure similar to DefinitelyTyped but for metadata legal data (licenses, e.g.). Finding this data (and making sure it's the right data) is difficult. Pooling our resources could make this much easier.

The challenge for us now comes with identifying, selecting, and bringing together in a workflow all the tools that currently exist along with the notion of crowdsourcing and curating the data.

With the data a lot of companies can provide we can pool our resources and

Gil: Can you differentiate this from OpenHub or something blackduck has done in the past? In what ways does this differentiate?

Jeff: A lot of the data made by OpenHub for example can be very spurious; there were difficult data quality issues and it was trying to be too all-encompassing. By focusing down we can make the quality much higher; the goal is also not open source discovery or browsing, for example, this is very much a machine-to-machine interaction. The purpose of the data is not to drive community, it's to drive compliance.

Sebastian: OpenHub only lists the declared license; in this case we would list all licenses as per scans.

Jeff: OpenHub and others are driven by commercial entities for commercial purposes. One of the goals of this project is to drive the data back upstream into upstream repos. "We don't want to have forks"; we want to drive all this data back upstream, which will make it more attractive to the upstream projects.

Gil: Want to find the proper way to scope this so that it succeeds. "What isn't in npm that you would want in npm for your node developers?"

Jeff: The accuracy issues between licenses. There's also issues with npm on source location (missing, points to the wrong place, tag information for the hash is missing).

Gil: Should the repair go into npm to fix those things? (and by extension, going to other communities to do the same)

Jeff: I don't believe they're mutually exclusive. Fixing npm to ensure the license is accurate is unclear how to do that. We as a community should work to make npm and maven better, but there will still be gaps in the data.

Rashmi: We may want to deliberately plan to outline the differentiating items about this effort. The data this project is providing is much closer to what these companies need to do business and ship products "it's a much more ardent need." This is a more compelling reason that is directly related to shipping products. Motivation for companies like ours and motivation for these projects and why they would want to participate.

Jeff: To a certain degree some of us have already built something like this. The tooling in this is another element that feeds into this. What's missing from all these is data itself.

- Organization umbrella

Option 1: Software Heritage is curating the source for all the world. In this case, we would be curating the metadata. Here OSI is also involved with the curation of the license process along with the curation of the data.

SHP is in the middle of becoming a foundation. It may be important later for an entity that can accept money in the event we want to put money into hiring people for the project, for example.

Option 2: Linux Foundation have a lot of tooling or process, again, without the data. Within the Linux Foundation there is the TODO Group, which might be another option with a more specialized focus.

Rashmi: A vote for the TODO Group as well. There are other tooling options that also could potentially provide metrics, too. Placing ourselves under the LF specifically in the TODO Group would be natural.

Gil: TODO group is great organization structurally, not as active as a working group. TODO has some structural problems, and it's a little out of scope for what TODO currently does. Would need to talk to the steering committee.

Jeff: Probably don't need to get caught up on being in the TODO Group specifically, the current TODO structure may not quite fit. The question is whether it's in the LF or not.

Gil: To what extent is there a need to focus on the commercial aspect of this? There are other commercial orgs that rely on proprietary data sets and their money is made on the scanning. Is there a thought of working with one of the commercial vendors and agreeing it's mutually beneficial?

Jeff: I've had conversations with a few of those vendors about getting access to the data, and they seem to indicate interest but it hasn't gone anywhere. In some other cases there may be some data we could get access.

Organizationally, the LF is a trade association whereas the SHP is actually more vendor-neutral. Not sure how much it matters that the LF is structured differently than the SHP in the long-term. In any event we want the data to stand alone and is interesting no matter where it comes from.

Gil: Is there a way to focus the data to particular area of technology? Mobile application data, for example, is more interesting than web application data. There may be natural alliances with organizations that have the same particular interest in the same particular parts of the data/technology. Android or iPhone, for example. Do we agree: not all data is equal and mobile is more important than others?

Jeff: Not more interested in mobile more than web. Source vs. Packaged is the more interesting bifurcation of the data

Sebastian: Not interested in distinguishing between mobile vs. Web, more interested in build system, for example.

Jeff: However, we do want to focus on groups of people who are interested in allying with us around this. We can build something, put it out there, and use it as a way to build momentum and bring it together and collaborate.

- Roadmap

Project structure is one thing we want to build out, and in parallel we want to "get something working." There are parallel process and technological questions we need to work on answering. We're looking for who wants to get involved and answer these questions.

Rashmi: Going to spend some time looking at the execution plan and give some notes back to Jeff/Carol on what participation looks like. Like the hackathon idea. Maybe we should do some brainstorming on tools we want for scanning (FOSSology, Scan Code, e.g.). Some are more promising than others, there's a few tools Qualcomm can offer in. Particularly interested in working and talking about tools.

Sebastian: Open sourced the beginning of their toolkit. Use a bunch of package managers and build systems to get the metadata and then pull it into Scan Code.

Jeff: We need pluggability and record where the curated data came from and let the upstream projects make decisions on that.

What the set of tools are now and how we would structure a system that would evolve and be transparent about the tools being used.

Rashmi will be in Seattle next week and could potentially meet us.

We want to have email exchanges about the tools. We want to hear back about your area or level of interest by Wednesday the 22nd.

Follow-on call for the week of the 27th.

Hackathon? Some folks may need time to identify the correct people in their organization to participate. Sebastian: online demo session for everyone to demonstrate what they have already. Next meeting will be demo session. Survey Of The Landscape (tool you want to demo or other technology you want to talk about).

- Review deck
- Naming
- Answer questions and set next meeting date

Questions to be answered:

- Does this make sense?
- Does this design meet with everyone's expectations of the project?
- Where will you contribute?
- How aggressively?
- Hackathon?

Expected outcomes:

- Answers to questions above
- Decision on Hackathon
- Next meeting date planned