

# 2020-11-09 CTWG Meeting Notes

## Date

09 Nov 2020

## Attendees

- [Drummond Reed](#)
- [Dan Gisolfi](#)
- [@Rieks Joosten](#)
- [@David Luchuk](#)
- [Scott Whitmire](#)
- [@Steven Milstein](#)

## Goals

- Establish the GitHub foundation and tooling for our CTWG work

## Agenda

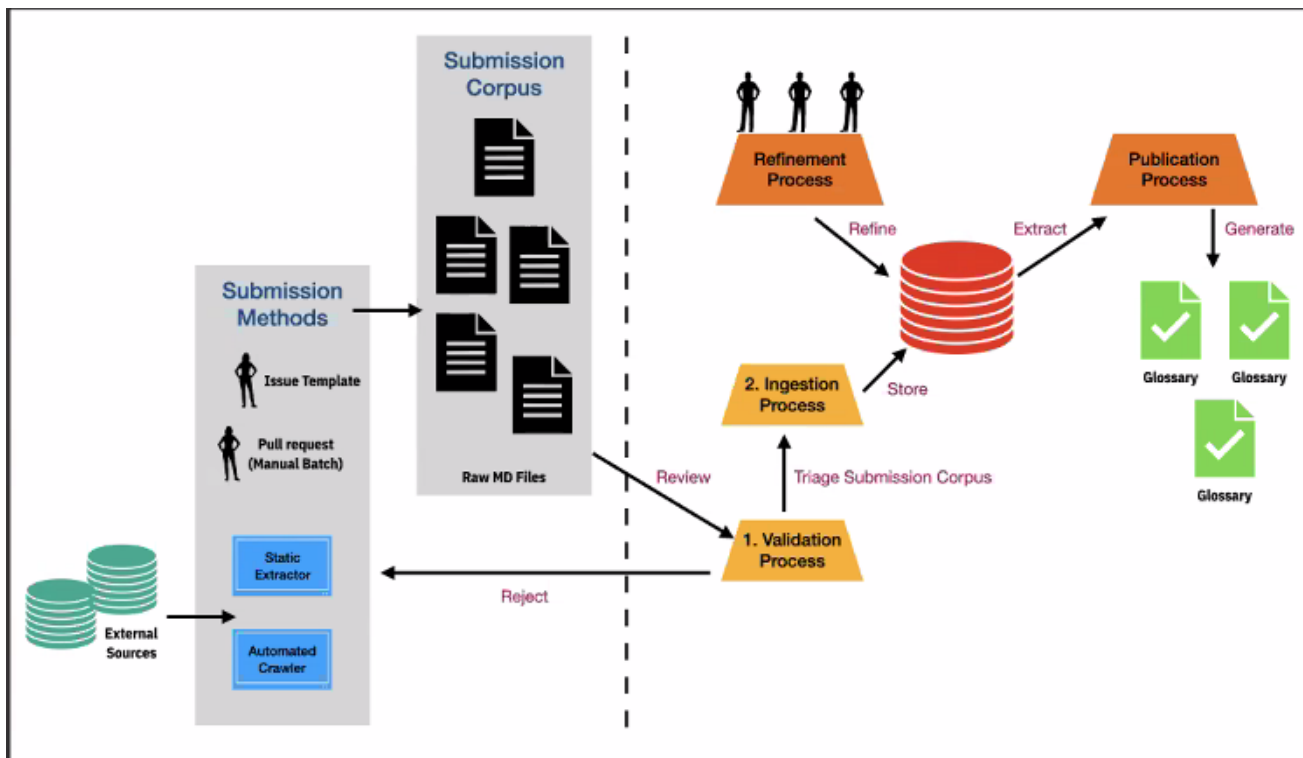
Time	Item	Who	Notes
5 mins	Welcome & Introductions	Chairs	
5 mins	Review of Action Items	Chairs	
5 mins	Update on eSSIF-Lab terminology work	<a href="#">Rieks Joosten</a>	
15 mins	GitHub strategy & coordination with Operations Team	<a href="#">Dan Gisolfi</a> <a href="#">Daniel Hardman</a>  <a href="#">David Luchuk</a>	
15 mins	Ingestion of external glossaries/vocabularies	<a href="#">Rieks Joosten</a> <a href="#">Dan Gisolfi</a> <a href="#">Daniel Hardman</a>	
5 mins	Next meeting	Chairs	

## Recording

- [Link](#)

## Notes

- Action items from last meeting
  - All PRs were merged
  - [David Luchuk](#) and [Steven Milstein](#) are coordinating with the Ops Team
  - [Rieks Joosten](#) wrote an analysis of Glossarist and posted
    - It looks like a good tool, however as a tool it is fit for purpose specifically for ISO's work
    - It is designed to do glossaries in multiple languages
- [Rieks Joosten](#) reported on the eSSIF-Lab terminology work
  - There has been an adaptation of the glossary work
  - The terminology pop-ups have been improved and are being further redesigned
  - The next step is that the subfunders are being asked to write up a functional description of the components they will be developing such that it can be added to the corpus
  - The idea is that they will be adding terminology and definitions as needed
  - This is just starting; Rieks will appraise us in future meetings
  - Rieks confirmed that we have two terminology tracks: one here in ToIP and one in eSSIF-Labs
    - They will proceed in parallel for now
    - Then we will look at merging them when the time is right
- Dan shared this picture of the overall process that we are trying to develop and integrate with Github.



- Dan ran through the high-level process. The key open questions are:
  - After ingestion (left side of the diagram), what is the refinement and publication process?
  - Are we going to use the GitHub file system as the storage system?
  - Scott noted: "If we agree on the definition another org has for a term AND they have published it in their glossary, why would we talk about "investing" a definition (and thus create an on-going maintenance task) and instead not just reference it? If we disagree with their definition, we can certainly create our own, and we can also add to it in our glossary. That should be "injesting"". As an example, we could refer to IEEE's extensive glossary work.
  - Open Issue: references to external terms.
    - First decision: do we include one?
    - Second decision: do we need to modify/refine it? Or just reference it?
- Drummond proposed to discuss next steps to get to a "minimum viable system"
  - Dan asked if we can just start with "flat-file" Markdown files in GitHub.
  - Scott: terms are going to have a state, and so at a minimum we need to track that state
    - We are also going to need to track language
    - We are going to need tooling to do that
  - Rieks: there are so many terminology tools to choose from, and it's hard to get agreement
    - This applies both to tools and terms
    - General terms can lead to endless discussions
    - So we should let terms be decided by specific scopes/stakeholders
    - The value that we can add is to provide tools that the stakeholders can use
    - We can also define criteria for this process for submitting, defining, and maintaining these terms
    - So our job can be to find out about the tooling we can use and recommend to these groups
    - Then we can add value by going over these terminology stores to see if we can help map and harmonize terms to achieve "terminology interoperability"
    - We could start out with one particular group to see if we can get it working for them
  - Dan agreed that it would good to start with one group, and that we've already started it with the Utility Foundry WG
    - But the open question is how to handle that group
  - Rieks: what we haven't agreed on is the validation process
    - Suggests can ask the group what states they want to see
  - Dan disagrees and suggests that we need to define a process that all WGs can use
  - Rieks suggests that we start very simply
    - Take in the terms as GitHub submissions
    - Make sure the submission is complete and valid by the template
    - Tag and link the Markdown file
    - Publish it
  - Dan: is there an existing tool we can use?
  - Rieks: let's get it running with the simplest process we can
  - Dan: will take the action item to research such tools
    - Feels that we need to find the simplest tool that we can start working with
    - Showed an example of a glossary produced by Mkdocs from Markdown docs in GitHub
    - But we need to have something that adds the next level of functionality: state, tags, links
  - Rieks: We need to define a data model that can be used by the other WGs
  - Dan: if the tooling we end out selecting has a data model, can we live with it
  - Drummond: Conclusions

- Our next step is to identify the tooling we need to establish a "minimum viable process" for terminology ingestion, refinement, and publication
- Dan has taken the action item to identify candidate open source tools for this
- We will discuss via our Slack channel between now and the next meeting
- At the next meeting we will (hopefully) make a decision about the initial tooling to use

## Action items

- ☒ Rieks Joosten will copy his Glossarist report to this Notes page for retention
- ☐ [Dan Gisolfi](#) will investigate open source tools available for terminology and will talk to [Daniel Hardman](#) about such tooling
- ☐ [Drummond Reed](#) will notify [Daniel Hardman](#) and also publish an update to our Slack channel

## ## What Can `Glossarist` Do For ToIP?

`[Glossarist](#)`, the tool, seems to be specifically designed for the purpose to support ISO/TC211 as it operates geo-terminology according to the ISO 19135 procedures, the result of which is [ISO/TC211 Geolexica](#). In order to determine its use for ToIP, we can first compare this result with e.g. the [Sovrin glossary](#), the [NIST glossary](#), the [eSSIF-Lab glossary](#), or the [Legal Dictionary](#) (disregarding the actual content, just looking at the kinds of data that are there), determine what it is we would need and actually use, and take it from there.

You need to make a github repo for your work (as easy as copying a template). It should be possible to bulk-import terms, but that usually requires some conversion. You can ask for help in an issue, but the dev team is < 2 people... Alternatively, you type in your terms in a desktop application. I don't know how the generation process works, but ISO/TC211 seems to have done it.

ISO/TC211 also seems to be its only user. They seem to be in line with our thinking on terms/concepts, and acknowledge that different contexts may use different terms for (almost) the same concept. They are pondering about extending the tool with relations between concepts, but it hasn't materialized in the produced result. Possibly because of the small dev team and that there's work to be done: the desktop tool loses my edits, but perhaps that is because I do not know how it works and the documentation doesn't help me there.

Conclusion: I do not think this is a tool that we should start using, primarily because it doesn't produce results that I think ToIP users would be looking for, and also because I think it takes some experimentation on CTWG side to create the specifications of the artifacts we would like to generate.