

Core Concepts

Context

The primary focus of the ToIP Foundation is not just on technology (e.g. cryptography, DIDs, protocols, VCs, etc.), but also on governance and on business, legal and social aspects. Its mission to construct, maintain and improve a global, pervasive, scalable and interoperable infrastructure for the (international) exchange of verified and certified data is quite complex, and daunting". This not only requires technology to be provided (which is, or should be the same for everyone, i.e. an infrastructure). It also requires that different businesses with their different business models can use it for their specific, subjective purposes. And that each individual business and user is provided with capabilities that facilitate its compliance with the rules, regulations and (internal and external) policies that apply to that entity - the set of such rules, regulations and policies being different for every such entity, and dependent on the society, the legal jurisdictions and individual preferences. All this is to be realized by people and organizations from different backgrounds - different cultures, languages, expertise, jurisdictions etc., all of whom have their own mindset, objectives and interests that they would like to see served.

The aim of this WG is to enable people in the ToIP community to actually understand what someone else means, to the extent and (in-depth) precision that they need, and facilitating this by producing deliverables/results/products that are fit for the purposes that they pursue. Initially, we expected to see the development of a common glossary, that lists (and summarizes) the basic words we use in the ToIP community. It would include terms defined within as well as outside of ToIP (e.g. by [NIST](#), [Sovrin](#), W3C's [VC](#), [DID](#) standards, and others).

However, the [minutes](#) on a IIW meeting topic '[glossary effort](#)' showed that developing a common glossary is quite difficult. This is underlined by a [post of Eugene Kim \(2006\)](#). But even when an effort to establish a 'common glossary' were to be successful, that doesn't imply that the 'commonality' extends beyond the set of its creators. The idea itself of establishing a terminology and subsequently (cautiously, but nevertheless forcefully) imposing it on others, is a highly centralistic way of doing things. And it doesn't work (it never has).

The WG recognizes that different groups use (slightly or quite) different [terminologies](#), and acknowledges their 'sovereignty' in doing this. Thus, such groups will be enabled to define their own [terms](#), yet at the same time facilitated to use terms defined elsewhere. As each group [curates](#) its own terminology, they each have the ability to decide to what extent they will adopt the terms of other groups into their own terminology. We trust that the various ToIP WGs and TFs will work together and the need to harmonize terminology will arise as their cooperation takes on more solid forms.

We expect subgroups of the ToIP community (e.g. WGs, TFs, TIPs) to create their own specific terminologies that help them serve their needs as they focus on specific objectives (thus facilitating domain/objective-specific jargon). The CTWG will assist them where appropriate, and ensure that (in the midterm!) [glossaries](#) can be generated from each such terminology.

Also, we expect to include more precise (theoretical?) specifications of underlying concepts, e.g. in terms of conceptual/mental models. Such models help to obtain a more in-depth understanding of ideas that are worth and necessary to be shared within one or more community sub-groups. They may also facilitate the learning process that (new) community members go through as they try to understand what it is we're actually doing. And they may help to 'spread the word' in specifically targeted (e.g. business and legal) audiences. A specific focus of this WG is to establish relations between the concepts of the mental models and the terms defined in the various glossaries.

A model for some of the deliverables of this WG is one or more websites that would resemble the [Legal Dictionary](#). This site not only provides a definition of various terms, but also a brief description of their backgrounds, various use-cases that exemplify the relevance of (and distinctions made by) the terms, and other useful information.

Finally, we expect to see results that we haven't thought of yet, the construction of which will be initiated as the need arises, by (representatives of) those that need such results for a specific purpose. Perhaps we might produce a method for resolving terminological discussions that can be lengthy and do not always get properly resolved (e.g. as in id-core issues [#4](#), [#122](#)). Here, we leverage [a prior collaboration](#) between Daniel Hardman (Evernym) and Rieks Joosten (TNO).

Requirements

The Corpus of Terminology MUST have:

1. Source control and build processes managed in github.
2. A well defined syntax for contributing concepts/relations, and for each of them an identifier by which it can be identified within the scope of the Corpus.
3. A well defined syntax for attributing terms to such (established) concepts/relations for specific contexts/domains.
4. A well defined CI/CD process.
5. A simple process for contributing further content.
6. A simple publicly accessible website, containing at least the Corpus-identifiers and their definitions, possibly inspired by the '[Legal Dictionary](#)'.
7. A PDF document for every published version, containing at least the Corpus-identifiers and their definitions.

The Corpus MUST NOT have:

1. A skill requirement on programming knowledge as that will reduce contributors.

The Corpus SHOULD be:

1. Reusable and easy to leverage in ToIP repos.
2. Usable for language translation via separate self-organized language specific repos. These repos should be aggregators of the baseline glossary and any TIPs.
3. Usable for mapping its identifiers/terms to those in use in other contexts/domains.
4. Consumable at the RAW content level (.md files) by external groups who wish to render content in a different manner.

Solution Approaches

We SHOULD:

1. Use a github repo to manage the corpus.
 - a. Consider using a Creative Commons license instead of an Apache license; it may be more appropriate.
 - b. Require DCO/IPR for contributors to the repo. Anybody who complies with the DCO/IPR requirements can submit to the corpus by raising a PR.
 - c. No need to manually maintain metadata about who edited what, when. We have commit history and git praise/blame.
 - d. Use github issues to debate decisions about term statuses. Anybody can raise an issue.
2. Use existing pervasive opens source documentation tools such as [Spec-Up](#), [Docusaurus](#), and/or [GitHub Pages](#):
 - a. Each concept is described in a separate markdown doc that conforms to a simple template (see below). Concepts link to related concepts.
 - b. Each term is a separate markdown doc that conforms to a different simple template (see below again). Terms label concepts; links from concepts to terms remain implicit in the markdown version of the data, to avoid redundant editing. Having terms and concepts as separate documents that cross-link allows for synonyms, antonyms, preferred and deprecated and superseded labels for the same concept, localization, and so forth. They also allow for the peaceful co-existence of multiple terminologies (= sets of terms, namespaces, ...)
 - c. Each context glossary is a separate markdown doc that conforms to another different simple template (see below once again). A glossary is an alphabetic list of terms relating to a specific subject, or for use in a specific domain, with explanations. The markdown document specifies the scope of the glossary, and the selection criteria for terms.
 - d. Provides extendable CI/CD pipeline for the repo, and write unit tests to enforce any process rules, quality checks, and best practices the WG adopts.
 - e. CI/CD process should enable live website and refreshed PDF document after each approved and merged PR.
3. Define the criteria for giving a term the statuses. What are grounds for saying it is deprecated, superseded, etc. (Criteria are published in a doc in the repo, so debating changes to criteria means a PR and github issue.)
4. Create a release process guidelines.
 - a. Define difference between live glossary and a "blessed version". Suggest once per quarter, with names like "2019v1" (where 1 is a quarter). This format is not semver-compatible, because we have no need to wrestle issues of forward and backward compatibility--but it is easy to understand, parse, and reference in a URI.
5. Establish a ToIP website level access experience
 - a. Access to main Glossary in all language versions
 - b. Access to TIP Glossaries

We MAY:

1. Leverage existing CI/CD approaches (sample code repos) for incorporating [Spec-Up](#), [Docusaurus](#), and/or [GitHub Pages](#).
2. Suggest to the tech WG that they may write a generator tool that walks the repo, building in memory a semantic network of concepts that are cross-linked to terms, and emitting various incarnations of the content:
 - a. Browsable static html that's copied to a website, [glossary.decentralized.foundation](#). The website should be indexed by Google and have search based on elasticsearch.
 - b. A .zip file of the static html that could be copied to other web sites.
 - c. An ebook format (e.g., epub).
 - d. Possibly, occasionally, a JIT-printed SKU published on [kdp.amazon.com](#).
3. Create a crawler process that collects terminology from various sources (contexts), for the purpose of mapping terminology as is used and/or defined in that context onto the concepts/relations in our Corpus
4. Create a process for pulling new content (terms, concepts) from the MM_WG
 - a. A source is declared in a config file that's committed to the repo. This means anybody can propose a source by submitting a PR and debating its validity in a github issue.
 - b. Sources could include W3C Respec docs, IETF RFCs, Aries RFCs, DIDComm specs hosted at DIF, etc. Corporate websites wouldn't work because A) they're too partisan; B) they'd require random, browser-style web crawling, which is too hard to automate well.
 - c. Crawler pulls docs and scans them, looking for regexes that allow it to isolate term declarations, their associated definitions, and examples that demonstrate their usage.
 - d. Output from crawler is a set of candidate terms that must be either admitted to a pipeline, or rejected, by human judgment. Candidates that are already in the corpus are ignored, so this just helps us keep up to date with evolving term usage in our industry.

Content Templates: [Archived](#)