



Decision Trees Workflow Approach Proposal

2017-03-28

CONTENTS

- 1 General approach
- 2 Project plan
- 3 Responses
- 4 Company background
- 5 Communications
- 6 Budget
- A Abbreviations

1 GENERAL APPROACH

To us, the essence of commoning is sharing how, why, and what we do. To do this, we need a way to explicitly share our practice. Decision trees describe patterns of practice (or *behavior*) that can be used as an analytical tool to study and refine practices themselves, and can even be used as semi-automated workflows to guide the practice of research *going forward*. We see decision trees being valuable in describing any process that happens in the course of research, not just in a handful of cherry-picked types of situations.

This concept of decision trees carries some echoes of what Cameron Neylon has called for in his blog post [GitHub for science? Shouldn't we perhaps build TCP/IP first?](#) We need some way to describe the processes and practices of research in a way that opens that understanding to both humans as well as machines, in a uniform way, across implementations, domains, tools, and people. The approach we are taking is to describe the processes using RDF vocabularies, one of which will be created as part of this proposal.

SVG will be used for the visual representation of these decision trees for the reason that they are portable, and the SVG Tiny 1.2 specification allows [embedded RDF attributes](#) that will allow the data represented by these trees (not just the authoring metadata, but the purpose and intent of each node of the decision tree) to be accessible to machines as well as humans. This is a critical distinction that will enable these decision trees to actually be used by machines to simplify and assist in the process of research.

Of course, the idea with these decision trees is not to mandate specific research practices, but to create a way to practice and share our patterns of practice in a collaborative, peer-to-peer sort of way. If we do this right, decision trees will enable a more streamlined, even joyful experience, and people will be empowered in their efforts to reach the greatest potential for themselves and for their research.

2 PROJECT PLAN

The high-level deliverables for this proposal include:

- A **decision-tree workflow** for creating decision trees
- The creation of the **research intent ontology** to describe the research process and affordances within the practice of research
- **Guidance and documentation** on the patterns and vocabularies used in this approach
- A selection of **example decision trees**
- Put all of the above **on the Web**

2.1 Workflow

In the spirit of practicing what we preach, our intent in this proposal is to use decision trees to create a workflow for creating decision trees. This workflow will use the **research intent ontology** in conjunction with the **design intent ontology** and the w3c Provenance ontology (among others) and take someone step-by-step through the process of creating a decision tree.

This proposal will also include a proof-of-concept to convert from a textual representation of a decision tree (e.g. in Turtle syntax) to other RDF representations using RDF/XML, HTML/RDFA, JSON-LD in an automated way. This mechanism will also be able to output an SVG diagram with embedded metadata using the supplied design template, as well as convert from the SVG back to another RDF representation.

2.2 Research Intent Ontology

In order to facilitate the commoning interactions that are needed for designing

and using a decision tree, we need every piece of the tree open to machines, not just the authoring metadata. Because of this, using a generic decision tree tool will not be adequate. We are proposing to create an ontology to describe the pieces of the research process and how they fit together. This ontology will be used in addition to other ontologies and open standards, and is needed in order to appropriately model the pieces we need to talk about when we talk about scholarly commoning.

This ontology will also be used to describe affordances, or potential actions made possible by a given object or environment,¹ for research. Most of the time these affordances are domain-specific, but there are also many that are not domain-specific—that are used to accomplish such activities as annotating a document or asking a question.

2.3 Guidance and documentation

Every effort will be made to preserve the processes undertaken during the research and design of the ontology and workflows, so that this journey can be open for others to learn from or extend. The idea is that decision trees as well as other open structures will be used to guide people in developing or using other decision trees.

We will also include other documentation as is necessary.

2.4 Example decision trees

We will start with the decision trees that have been created as part of Work Product 3 of the Scholarly Commons Working Group as examples to learn from in designing. The workflow for creating decision trees will also function as an extended example of a fairly complex decision tree. Other examples may be included as seems helpful.

¹ From Wiktionary

2.5 On the Web

During the period of the work, these examples and supporting documentation will be shared under a page dedicated to this project on pentandra.com. If desired, these can also be put on scholarlycommons.org, but that would require further discussion as to how to proceed with that, and is not currently part of this proposal.

2.6 Risks

2.6.1 Where's the user experience?

As the emphasis in this proposal is on creating a mechanism to describe the practice of research, the implementation of the decision tree workflow for this proposal will use a rather simple textual user interface. This has a number of benefits. First, it will put the focus where it needs to be at this point, to make sure that the workflow is right without the distraction of a graphical user interface. It is more important to make sure that we can openly (to humans and machines) describe, streamline, and simplify the process of research. Also given the short time frame of this solicitation, we think it would not be prudent to rush the development of these foundational structures in order to try to develop a good user experience in such a short time period. Second, it will be easier for others to take these simple ideas and implement them more easily in their own software projects and platforms.

A companion proposal will be forthcoming to take on creating a more visual user experience in a more fully-functioning open-research environment that would also serve as a prototype of scholarly commoning.

2.6.2 Implications

Yes, we are implying that every piece of scholarship is valued not just to people, but to machines as well. People and machines need to know not just what the pieces are, but how they fit together as well. By addressing each piece independently as a piece of scholarship, we are acknowledging a change and shift in what we value as scholarly contributions. This may seem very strange or unusual to some, but we feel that the benefit will be absolutely worth it and open the door to the Scholarly Commons.

3 RESPONSES

Reuse of existing, open-source components wherever possible

Yes, we will be using a variety of open-source libraries, ontologies, as much as possible

Reduction of number of steps required to make a commons-compliant object

Yes, even though we are doing a textual interface for the proof-of-concept included in this proposal, and it will require knowledge of a textual representation (e.g. Turtle or a DSL), it will take two steps to create a decision tree:

1. Write a declarative representation of the decision tree.
2. Generate the representations.

Should be both human and machine readable

Yes, see [above](#).

The final trees should be static but hyperlinked

Yes, even the proof-of-concept will create static svg diagrams that can have embedded hyperlinks and linked data.

Each should have a persistent identifier and include versioning capabilities

The persistent identifier and versioning capabilities are handled in the linked metadata. In the proof-of-concept, this will be more a manual process.

There should be a manual/template to enable adaptation by others

Yes, this is handled in the [guidance and documentation](#) section above, and this will target both developers and researchers.

Need to be able to create new trees and new versions of existing ones

You can copy a representation (textual or svg) and edit at will. We will use a common resource version pattern for versioning in the proof-of-concept.

Need to be able to provide granular feedback in a simple way

We will be using the hypothes.is client with the proof-of-concept to elicit feedback during the design process.

Need to be downloadable

Check.

Need to link them to authoring metadata (how to cite them)

We will use unique identifiers for the author and any other contributors, and these data will be embedded within the decision tree representation.

Need to be exportable into and importable from standard formats

We will be generating representations of the decision trees in all common RDF syntaxes.

Users should be able to comment on any node or decision point

Every node or decision point will have a unique identifier on which comments can be made.

Need to link with one or more external repositories

The companion proposal describes a system that functions as a repository as well.

Additional credit will be given to proposals that consider 508 compliance

The svg format is textual and consideration will be given to make these objects as accessible to those with disabilities as possible.

Interoperability with other open source components

Other than being interoperable with svg editors, the decision trees at the beginning will not be interoperable as far as the research process goes until open-source components are built to use the decision trees in the research process.

4 COMPANY BACKGROUND

<https://pentandra.com/company/>

5 COMMUNICATIONS

At least weekly updates via call, and interim communication via email to provide updates and seek feedback on the ongoing development of the ontology and workflows.

GitHub issues can also be used for communication throughout the development process.

6 BUDGET

Please see **caveat**. Quoting from the most relevant part: since this would be a fairly hefty time commitment, we'd love to be compensated, but even if we were compensated only partly or not at all, we would still love to help.

A ABBREVIATIONS

DSL	Domain Specific Language
HTML	HyperText Markup Language
IP	Internet Protocol
JSON-LD	JavaScript Object Notation for Linked Data
RDF	Resource Description Framework
RDFa	Resource Description Framework in Attributes
SVG	Scalable Vector Graphics
TCP	Transmission Control Protocol
W3C	World Wide Web Consortium
XML	eXtensible Markup Language