

# **Ontology-Vocabulary Characteristics Relevant to Suitability for **Semantic Web & Big Data** applications.**

Hackathon Report  
Ontology Summit 2014

Amanda Vizedom  
*29 April 2014*

# Participants

- Core team:
  - Amanda Vizedom,
  - Andrea Westerinen
  - Simon Spero
- Drop-in Observers:
  - Mieza Hamka
  - Lamar Henderson
  - Anatoly Levunchuk
- Drop-in Contributor:
  - Dan Brickley

# Objectives

- In selecting, creating, or effectively using an ontology or metadata vocabulary for (re)use, it is important to understand not only the requirements of the intended use, but the characteristics of any candidate ontology or vocabulary.
- Create a catalogue, represented in an ontology, of ontology and metadata vocabulary characteristics potentially relevant to suitability for some semantic web or big data application.
- Create this catalogue as an open-source, collaboratively developed, extensible, and publicly available resource.
- Create this catalogue in a form usable by both humans and machine applications, suitable for use in such contexts as:
  - manually creating characteristics metadata for an ontology or metadata vocabulary;
  - representing the outputs of evaluations of such artifacts;
  - listing use-relevant metadata for such artifacts in repositories, registries, and other places and forms in which ontologies and metadata vocabularies may be presented for use or consideration.
- Take a bite out of obstacles to useful description

# Hackathon Activities (1)

- Initiated and ramped up VOCREF project: *Vocabulary and Ontology Characteristics Relevant to Evaluation of Fitness*
- Established user-independent GitHub repository at <https://github.com/vocref/vocref> and expected to stay there as project continues. Current and future contributors can fork repository and work freely, submit pull request to initiate integration into main.
- Chose OWL2 KR language, based on existing familiarity and level of use among semantic web and big data application designers, as well as ease of integration into existing repository and other environments.
- Chose OWL Functional Syntax as the stored serialization form, due to its vastly better behavior with version control systems (compared, for example, to OWL RDF).
- Wrangled modularity considerations into initial approach (see further slide)

# Hackathon Activities (2)

- Most hackathon-day work focused on top-level module, vocref-top
  - Many representations and concepts related to ontologies and evaluation, from a variety of sources, were considered and weeded out (as, for example, out-of-scope or specific to the assumptions and focus or their original context, or of dubious quality in one way or another).
  - As a result, we arrived at a satisfyingly solid vocref-top, where work will continue.
- some additional content modules started to test and illustrate the approach.
- Review of candidate ontologies, vocabularies, and concepts from evaluation tools and methods for whole or partial reuse:
  - Gathered a good stack of resources to evaluate; stored in the repository
  - Initial dent was made ; some filtered out, some partially integrated.
- Issue store:
  - Many specific work items identified and documented as task-sized issues (bugs, enhancements, or questions)
  - Currently have 17 open, 30 closed
  - Repository was started in a user account and moved to user-neutral; in the process, a few documented issues were left behind Most had already been closed, but a few, especially related to documentation for newcomers, need to be copied over and addressed in the near term.
  - Issue store is important progress because identification & documentation of issues functions as discussion handle, specification of requirements, and coordination/collaboration mechanism for contributors.

# Hackathon Activities (3)

- One issue tagged for completion during the hackathon remains incomplete: Adding sufficient annotations to concepts that were reused from the 2013 "ontology of ontology evaluation" output. This task is expected to be completed within the next month or so.
- An additional batch of three issues were tagged for completion by the end of the Ontology Summit. Eight issues have not been tied to any schedule, though several of these are still undergoing work and may be closable soon. A number of additional points and to-do items were noted in discussion and will be added.
- Some issues could not be worked on simultaneously simply because of the lack of sufficient automated merge capabilities when working on the same module (vocref-top). We did need to take turns with substantial changes to this ontology.
- Hangout was very productive. The collaboration time enabled the team to sort through considerable material effectively, consulting one another and bouncing ideas regularly. Pending a bit more documentation for newcomers (one of the near-term issues), the base now established is good and can support ongoing, asynchronous development of vocref as an open source resource.

# VOCREF Approach

- Focus on specific characteristics of vocabularies and ontologies that make them more or less suitable for SW & BD apps
- Pay specific attention to characteristics discussed during the Summit in relation to specific SW/BD applications and experiences.
- Look for and Review existing treatments (of V/O characteristics and evaluation) and triage the characteristics identified there for reuse
- Leave out elements from existing treatments that do not meet VOCREF focus, are too vague or general to be helpful, shed more heat than light

# Not as useful as some seem to think:

- Classification of Semantic Resources by overall type, in entirety
  - e.g.: Is it an ontology or a vocabulary?
  - Especially where these classifications are themselves not standardized
  - Does not provide reliable information about nature of resources
  - Distracts attention from finer-granularity information that is needed for operational understanding, sufficient to evaluate a resource against some specific use requirements.
- Description without clear definition and grounding
  - E.g., a rating of “usability” without
    - a clear definition of what is meant by usability and
    - Grounding of this notion in some documented method of assessment
- Description of VO characteristics, possibly including evaluation methods and metrics, but without good reason for thinking that those characteristics, methods, and metrics are significant wrt suitability for (some kinds of ) use
- E.g., reviews or evaluation results that leave unstated
  - what aspects of ontologies or metadata vocabularies they assess,
  - how those aspects are modeled
  - whether they address relationships between aspects & suitability for varied uses.

# Modules & Architecture

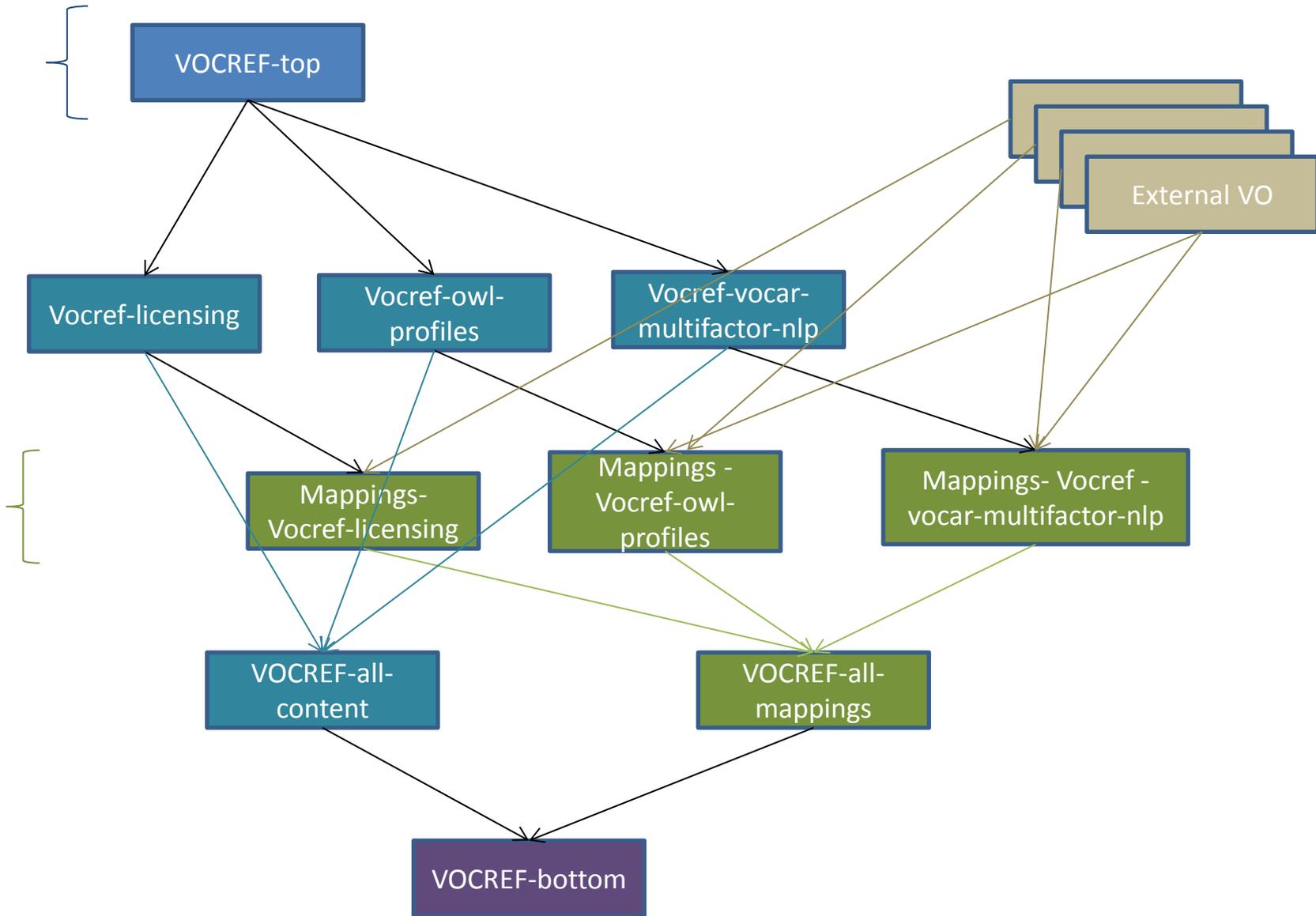
- Goal: enable both
  - (a)small ontology modules with lowest feasible import burden and
  - (b)additional modules (importing both the content ontologies modules and external vocabularies and ontologies) containing mappings between VOCREF and those external resources.
- Essential structure is spindle-style.
- Primary, content ontology levels include:
  - Top-level framework ontology (vocref-top) containing the essential concepts for representing vocabularies and ontologies, evaluation-related characteristics, and relationships between the two, as well as certain core classes of characteristics;
  - Small content ontologies inheriting from vocref-top, modeling families of characteristics;
  - An all-vocref-content ontology inheriting all vocref-internal content.
- For mapping, ontology modules exist as needed, importing from particular content modules and external VO to which they should be aligned, containing mapping assertions.
- Overall structure should enable users to use pieces of vocref appropriate to their application, with as little weight beyond the need as feasible, while maintaining semantic integrity.
- Initial, hackathon-time pass is rough.
- Refactoring is a possibility, if indicated by test application and evaluation down the road a bit.

# Modules & Architecture (2)

Minimal content ontology, with core framework concepts

Content modules

Mapping modules (btwn VOCREF and external VOs)



# What's on github.com/vocref now

- In the code repository:
  - References and Resources
    - External ontologies for potential reuse
      - E.g, ADMS, DCMI terms, DOAP, OMV ....
    - Papers And Specs
      - Paperson evaluation, metrics, ranking, what makes an ontology or vocabulary a good one,...
  - Test and Evaluation Resources
    - Beginnings or stubs for Capability Questions, Design Consistency Tests, RW examples to turn in to use cases (and then into tests)
  - VOCREF top (what all modules inherit)
  - VOCREF content modules
  - VOCREF mapping modules
  - VOCREF bottom (inherits all modules) (single grab & go when don't need to minimize)
- Issues:
  - Known bugs that need fixing, issues that need addressing in current OWL
  - Known areas needing coverage, identified as enhancement requests
  - Comments and asynchronous discussion on each issue, if/when it occurs
  - Assignment of issues, representing that someone has taken on an issue to work on.

# Status

- Will continue as an Open Source Project
- All three core members want to continue work
- Project is open to new contributors