

FAIR Vocabulary - case study: The Establishment Means GBIF Taxon

Name of vocabulary	Establishment Means GBIF Vocabulary
Scope (subject) of vocabulary	We narrowed down to this vocabulary on “Establishment means” from this list GBIF https://rs.gbif.org/vocabulary/gbif/ We also noted that in this GBIF list there are other vocabularies that are available from multiple sources (e.g., Units of Measure, Reference, etc.)
Vocabulary source	https://rs.gbif.org/vocabulary/gbif/establishment_means.xml
Who contributed to this case-study?	Natalia Atkins, Simon Wall, Doug Palmer, Lesley Wyborn, Adrian Burton
1. Who is the custodian of the vocabulary? (Community, organization, etc)	GBIF - custodian
2. Licensing: a. What is the license on the original vocabulary? b. Does it permit repurposing? c. What license will apply to the FAIR vocabulary? e.g. Creative Commons , Licentia	a) Licencing is not explicitly stated for vocabs, but other GBIF information is mostly CC0. b) Doug to confirm with GBIF

<p>3. Vocabulary content:</p> <p>a. Does every term have a unique <i>label</i>, and complete <i>definition</i>? If not, where could you get them from?</p> <p>b. Are the definitions consistent and distinct?</p> <p>c. Does the vocabulary have a <i>structure</i>: Are there well defined subsets? Is there a hierarchy of terms?</p> <p>d. Are there cross-references within the vocabulary?</p> <p>e. Are there cross-references to items in other vocabularies?</p>	<p>a) Yes - Adrian says they are pretty good compared with other vocabs. Doug says community accepted: Specifies level of hierarchy in definition.</p> <p>b) Yes</p> <p>c) The vocabulary does not have an explicit hierarchy, the hierarchy is text in the definitions</p> <p>d) There are no internal cross references (these are in text inside the definition)</p> <p>e) There no references to items in external vocabularies</p>
<p>4. Point of truth:</p> <p>a. What is the maintenance environment for the vocabulary content? (e.g. a file in a version control repository)</p> <p>b. Is the vocabulary maintenance history externally visible?</p>	<p>a) Don't know how the terms are generated or maintained. It has an issue date, but tells when converted into XML form: no version control. There is a GitHub: https://rs.gbif.org/vocabulary/gbif/ See also: https://github.com/gbif/gbif-api/blob/dev/src/main/java/org/gbif/api/vocabulary/EstablishmentMeans.java</p> <p>b) No - vocabulary maintenance is not visible, neither is the maintenance history and there is no provenance trail. This is probably a result of migration from code-based enumeration to a more vocabulary-based approach. (The ALA has a similar problem where vocabulary terms are buried in text file resources in code.)</p>
<p>5. Web identifiers:</p>	

<p>a. What is the base URI for the vocabulary? Can it be managed over the long term?</p> <p>b. <i>{is there a}</i> Design a pattern(s) to be used for each term URI</p>	<p>a) http://rs.gbif.org/vocabulary/gbif/establishment_means/ Yes</p> <p>b) base/term</p>
<p>6. Encoding:</p> <p>a. Basic hierarchical vocabulary: implement a SKOS representations of the terms (e.g using SKOS-Play! or PoolParty)</p> <p>b. Richer semantics? Can you create a OWL Ontology or RDF Shapes description?</p>	<p>a) There is no SKOS, but could be turned into one</p> <ul style="list-style-type: none"> i) To have SKOS elements we would need to go into partnership with GBIF. GBIF have not yet published a SKOS instance yet, so we need a publisher to do that kind of work with so that we can validate the SKOS version. ii) From a practical perspective, GBIF can translate URIs into standardised terms - they do plan to create a SKOS version. <p>b) You could create an OWL Ontology and RDF</p> <ul style="list-style-type: none"> i) The OWL ontology could (and maybe should) create sub-classes /sub-properties of SKOS:Concept, SKOS: broader etc. that reflect the specific domain being modelled. ii) Some of the other vocabularies, such as ranks, need domain-specific attributes, eg. isLinnaean for ranks. iii) It would be helpful to associate icons with terms on a per-application basis (eg. taxonomic status of excluded shows a no-entry sign in the NSL, TK labels have graphics associated with them)
<p>7. Complete the vocabulary metadata: some options: DCAT (W3C Data Catalog Vocabulary), LOV metadata guidelines</p>	<p>Currently do not use any of these metadata standards to Doug's knowledge - they are not RDF. The XMLs</p>

<p>(Metadata for Linked Open Vocabularies), OMV (BioPortal), MOD (Metadata for AgroPortal)</p>	<p>use Dublin Core terms.</p>
<p>8. Where will you register the vocabulary? Some options: RVA, LOV, BioPortal, Agroportal, BARTOC , OBO, CESSDA</p>	<ol style="list-style-type: none"> 1) Are already registered in RVA with pointers to GBIF (used by DAWE) See: https://vocabs.ardc.edu.au/search/#!/?activeTab=vocabularies&pp=15&q=GBIF but only raises awareness of the existence of the vocab. 2) ARDC has connections with Donald Hobern and Dave Martin: ARDC would become a secondary publication point. 3) GBIF have a performance related desire to run in house on a local server so they can crunch through billions of records to do data integration and aggregation that they do. 4) GBIF does need a public point of access. 5) GBIF does not have any connections with the other vocab registries listed.
<p>9. How will you make the vocabulary accessible? (i.e. make vocab and term URIs resolve)</p>	<p>They are accessible, but each term resolves back to a vocabulary hash term, rather than a vocabulary slash term. That is, you resolve to a fragment of a document, not to the exact term (is this acceptable?), e.g. https://rs.gbif.org/vocabulary/gbif/establishment_means/invasive maps onto https://rs.gbif.org/vocabulary/gbif/establishment_means.xml#invasive</p> <p>For increased accessibility, each term needs its own page.</p>
<p>10. How will the FAIR vocabulary be maintained?</p>	<p>Need to engage with GBIF, need a governance document and guidelines like what Anu showed if it is</p>

<p>How will the custodians be involved?</p> <p>How will the community be engaged and informed?</p>	<p>to be maintainable. These documents need to be public!</p> <p>Custodians can be involved by being maintainers</p> <p>GBIF needs to open these up for the broader community. The assumption is that other people will reuse this, but initially they were developed for internal consumption. Need to get over a big hump from internal use to external access and work out how you will use it, better documentation, formal change management procedures etc (things implicit in management and maintenance). It is in GBIFs interest to have community consensus around the terms, their definition and their upgrading.</p>
--	---