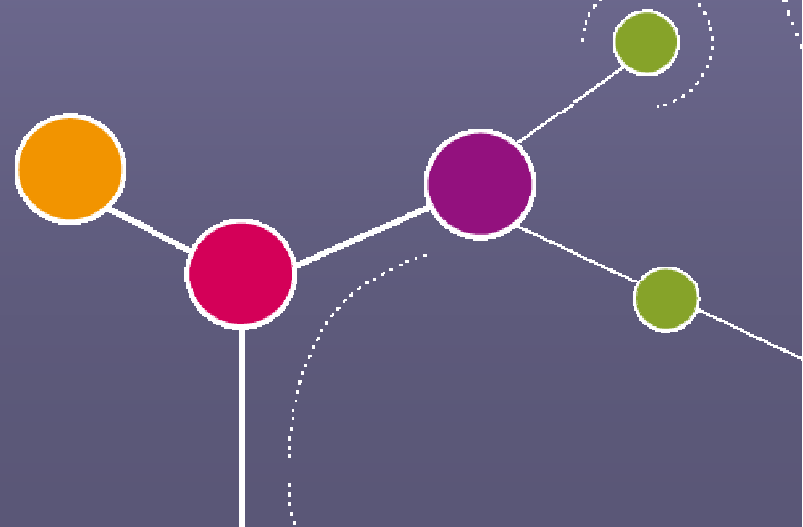




LOV Symposium: Linking and Opening Vocabularies – 2012-06-18

# Vocabulary Commons Sustainability & Responsibility

[pierre-yves.vandenbussche@mondeca.com](mailto:pierre-yves.vandenbussche@mondeca.com)  
[bernard.vatant@mondeca.com](mailto:bernard.vatant@mondeca.com)





- Vocabulary Commons
- An observatory for the Commons
  - The Linked Open Vocabularies project
  - Music Ontology example
  - Lessons learned so far
- Moving forward
  - What is missing
  - FOAF example
  - Possible actions
- To infinity ... and beyond ?

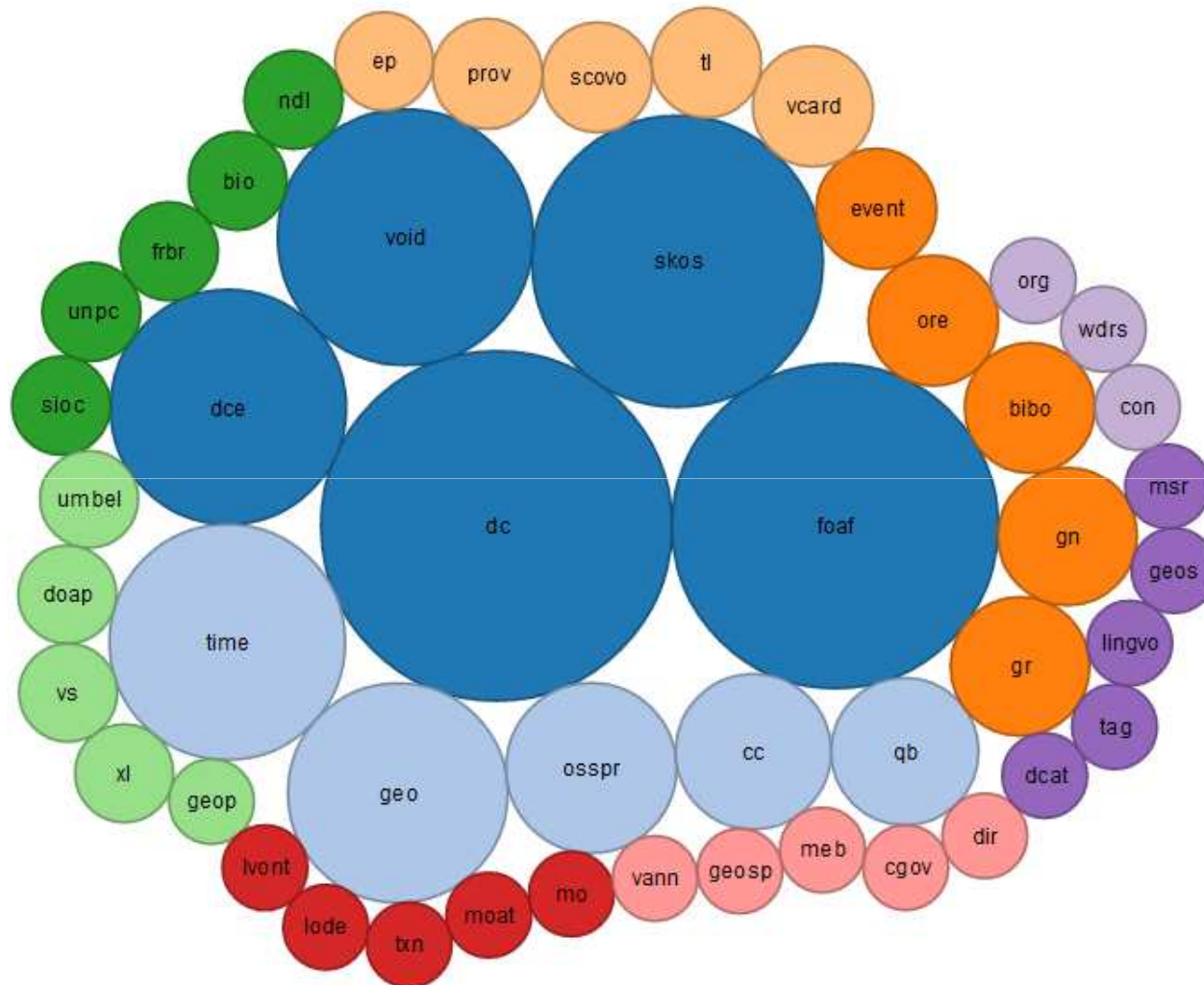


# Vocabulary Commons

- What we mean by **Vocabulary**
  - Fuzzy and arguable definition ...
  - Structured set of **types** and/or **properties** (common names)
    - ≈ Metadata Element Set
    - ≠ Value Vocabulary
  - Used to **describe resources and data on the Semantic Web**
  - Rooted in some Natural Language(s)
- The Vocabulary **Commons**
  - Published on the Web
  - Reusable under open licence
  - Interoperable and interdependent
  - **Technically sustainable** and **socially responsible**



# Vocabulary Commons



Vocabularies used in the LOD. Pierre-Yves Vandenbussche




# Vocabulary Commons

- Why reusing vocabularies?
  - **Technical** engineering
    - Take advantage of “good practices”
    - Sources of ideas or enhancements to its own vocabulary/dataset
  - **Social** engineering
    - Easing the understanding of my own vocabulary (by reusing already known ones)
    - ... Thus facilitate its reuse ... And the reuse of data relying on it.
- Why publishing vocabularies?
  - To foster the emergence of an **ecosystem** around the data (feedback, curators, users, etc.)



# Linked Open Vocabularies (LOV)

- An Observatory for the Vocabulary Commons
  - <http://labs.mondeca.com/dataset/lov> 
- Limited scope so far
  - RDFS/OWL vocabularies
    - Some values (SKOS) can be present « in sneak mode »
  - Limited size (no large thesaurus or classification)
  - Borderline cases such as schema.org, linkedgeodata.org
- Current figures (2012-06)
  - Over 250 vocabularies, 20k elements, 200k triples



# Linked Open Vocabularies (LOV)

- Actors involved
  - 32 organizations and 129 persons
    - Standard bodies (e.g. W3C or DCMI)
    - Heritage curators (e.g. IFLA, OCLC, Europeana, Library of Congress, European national libraries)
    - Press and Media groups (e.g. New York Times, BBC, IPTC)
    - Major Web companies (promoting initiatives like schema.org)
    - Governments and institutional data providers (e.g. data.gov, Ordnance Survey, INSEE)
    - Academics and research centers (e.g. DERI, INRIA)
    - SME, consulting companies, software editors (e.g. Talis, TopQuadrant, Mondeca)
    - Individual initiatives, more or less backed by one or more of the above (e.g. taxonconcept, geospecies, geonames, lingvoj, lexvo)



# Music Ontology

- [musicontology.com](http://musicontology.com)
- « Social Engineering »
  - Recidivists creators (Frédéric Giasson et Yves Raymond) and well known
  - Good documentation, discussion group, license
  - Persistent URI (PURL)
  - A set of related tools on sourceforge (Audio collection management, crawler, mapping, etc.)
- Indication of each elements stability level
  - Stable level 2, stable level 1, unstable level 2, unstable level 1
- Extensions : instruments, accords, notation musicale, etc.





# Music Ontology

- [musicontology.com](http://musicontology.com)
- « Social Engineering »
  - Recidivists creators (Frédéric Giasson et Yves Raymond) and well known
  - Good documentation, discussion group, license
  - Persistent URI (PURL)
  - A set of related tools on sourceforge (Audio collection management, crawler, mapping, etc.)
- Indication of each elements stability level
  - Stable level 2, stable level 1, unstable level 2, unstable level 1
- Extensions : instruments, accords, notation musicale, etc.



# Harvesting metadata ...

## MO - Music Ontology

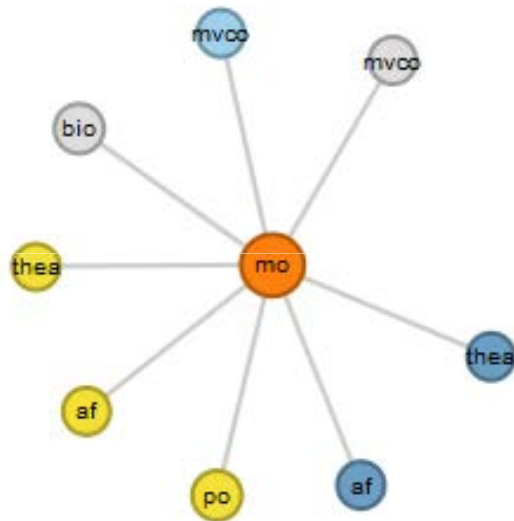
Property	Value
is part of vocabulary space	<a href="#">All &gt; Media &gt; Music</a>
Vocabulary URI	<a href="http://purl.org/ontology/mo/">http://purl.org/ontology/mo/</a>
Prefix	mo
Namespace URI	<a href="http://purl.org/ontology/mo/">http://purl.org/ontology/mo/</a>
is Defined By	<a href="http://motools.sourceforge.net/doc/musicontology.rdfs">http://motools.sourceforge.net/doc/musicontology.rdfs</a>
Description	The Music Ontology Specification provides main concepts and pr (tracks) on the Semantic Web @en
Last modified	2010-11-28
Creator	<a href="#">Bob Ferris</a> , <a href="#">Frédéric Giasson</a> , <a href="#">Yves Raimond</a>
Class number	59
Property number	163
Homepage	<a href="http://motools.sourceforge.net/doc/musicontology.html">http://motools.sourceforge.net/doc/musicontology.html</a>
See also	<a href="http://schemapedia.com/schemas/mo">http://schemapedia.com/schemas/mo</a>
Represented by	format-mo



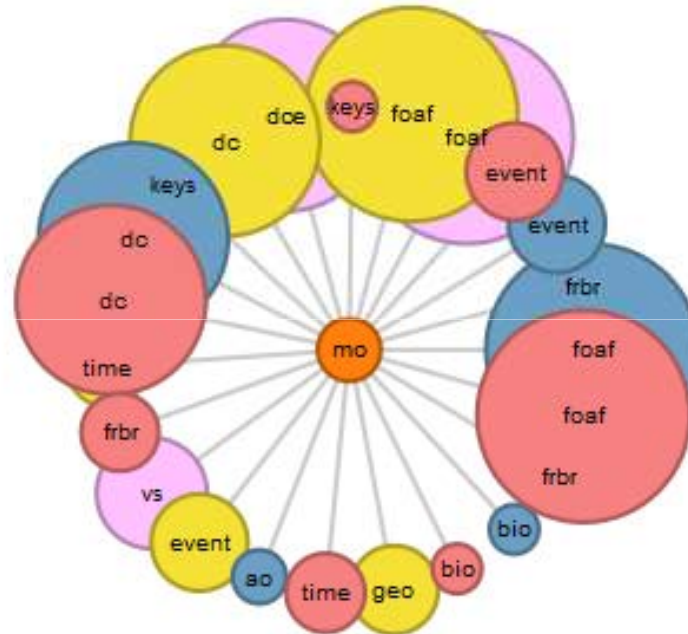
# Adding « social » metadata ...

## Vocabulary links:

Vocabularies referencing "mo" (8)



Vocabularies referenced by "mo" (23)



### Link types

- Imports
- Similar to
- Used by
- Relies on
- Metadata vocabulary
- Extends
- Specializes
- Generalizes
- Has equivalences with
- Has disjunctions with

**Specialisation** : `mo:MusicArtist rdfs:subclassOf foaf:Agent`

**Extension** : `mo:media_type rdfs:range dcterms:MediaTypa`



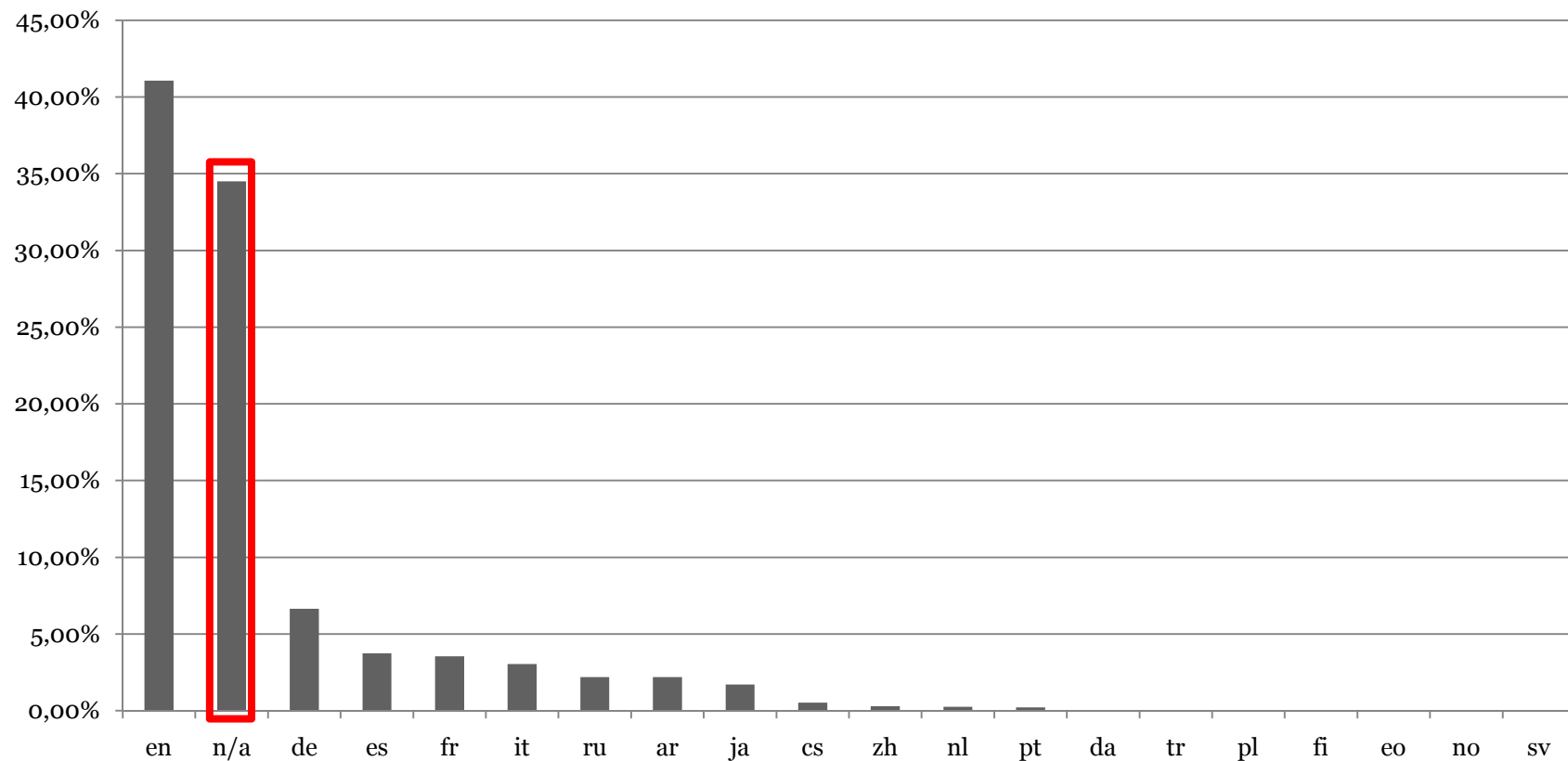
# Lessons learned

- Harvesting metadata is not obvious ...
  - Still 3% of LOV vocabularies are not dereferenceable
    - Semantic can not be known
  - Even *basic* metadata are often not available (18% explicitly declare a creation date. In the LOV dataset we have raised this figure to 85%)
    - As simple as title publisher, creator, creation date, description
  - Often difficult to identify *responsible* agents
    - Either organizations or individuals
  - Various neither sustainable nor responsible practices
    - In particular for vocabularies developed inside projects

# Lessons learned

- 12 % of vocabularies are multilingual ...

**Percentage of rdfs:labels per language among all vocabularies**





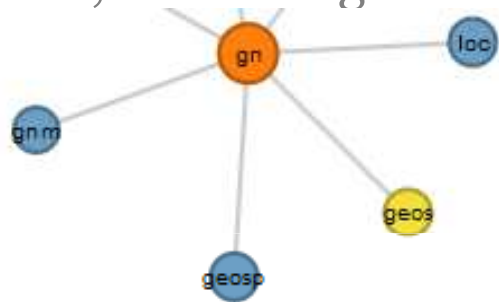
# Lessons learned

- **Positive feedback from responsible actors**
  - The project has fostered communication and cooperation
  - Started to raise awareness on sustainability issues
    - E.g. vocabs under W3C namespace and bound to recommandations,
    - Vocabs developed by individuals / 3 year research project
  - Started to make explicit **who cares** and **who doesn't**
  - More than 20 suggestions
  - Vocabulary visibility

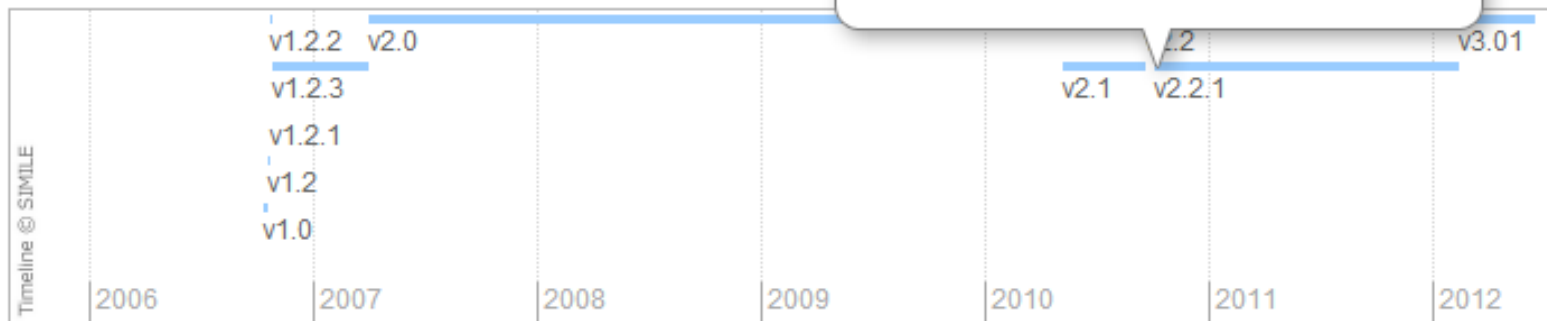
# ... so we need *yet more* metadata ...

- Time-bound metadata

- Evolution of vocabulary over time / past and current status, versioning ...



Vocabulary history:





## ... so we need *yet more* metadata ...

- **Time-bound metadata**
  - Evolution of vocabulary over time / past and current status, versioning ...
- **Responsibility metadata**
  - Who is responsible for this vocabulary today?
  - What is the level of responsibility of this agent?
    - Answers to feedback, care about user community ...
- **Sustainability metadata**
  - Is there a long-term policy for this vocabulary?
  - Who will care tomorrow, 20 years from now?





# Sustainability and Responsibility concerns

- The FOAF example



- **Agreement** between DCMI and the FOAF Project (2011-05-02)
  - <http://dublincore.org/documents/dcmi-foaf/>
- Among the commitments:
  - “DCMI [has a] full technical and administrative **access to the Domain Name** (xmlns.com).”
  - “This agreement is not a legally binding contract but the public expression of a **collaborative partnership**.”
  - “Promote active collaboration between the two vocabulary maintenance organizations”
  - “For the **longer term**, if DCMI should find itself as the publisher of the final results of a FOAF Project that has ceased activity, DCMI will **maintain the documentation** of FOAF in **conformance with the surrounding technical infrastructure** (e.g., in response to a revision of W3C Resource Description Framework).”



# Further action needed !

---

- **A Social Survey of the Commons**
  - Pushed to all identified vocabulary stakeholders
  - With simple questions about their vocabulary status
    - Who cares now?
    - What is the policy for tomorrow?
    - What level of support would you engage in?
- **Best practice guidelines**
  - Vocabulary design, evaluation, selection and reuse.
- **Vocabulary recipes**
  - E.g. Linked Universities (<http://linkeduniversities.org/lu/index.php/vocabularies/>)
  - People vocabulary (<http://www.w3.org/TR/vocab-people>)



# LOV sustainability

- Who will support all this?
  - LOV supported so far by a 3-year research project
  - Needs more human resources
  - OKFN: our new home !
- Various stakeholders identified
  - Standard bodies : DCMI, W3C
  - Libraries and Library Associations (IFLA, OCLC...)
  - Institutional data publishers (Europeana, data.gov ...)
  - Press and media companies and associations (NYT, IPTC ...)
  - ... and many more
- How can they cooperate?





# Should Libraries lead the way?

---

- Libraries are sustainable ...
  - They've been there for long and are here to stay
  - Maybe longer than ad hoc entities born with the Web
  - They know how to handle long-term legacy
- Libraries are responsible ...
  - Hopefully ☺
- Libraries should have the main role in the long-term preservation of the Vocabulary Commons !



- **Library Linked Data Incubator Group**
  - <http://www.w3.org/2005/Incubator/ld/XGR-ld-vocabdataset-20111025/>
- **Linked Open Vocabularies**
  - <http://labs.mondeca.com/dataset/lov/index.html>
- **DCMI Vocabulary Management Community**
  - <http://dublincore.org/groups/vocabulary-management/>



Thanks for your attention !



[pierre-yves.vandenbussche@mondeca.com](mailto:pierre-yves.vandenbussche@mondeca.com)  
[bernard.vatant@mondeca.com](mailto:bernard.vatant@mondeca.com)