Sustainable governance for long-term preservation of RDF Vocabularies

Special session, DC-2013 conference
Jointly sponsored by W3C and DCMI
2013-09-03
Links

• Twitter hash tag “#dcmi13”
• IRC channel “#dcmi13am” and “#dcmi13pm” at:
  – http://irc.w3.org
• Session description
• Discussion paper
“RDF Vocabularies”

• Sets of properties and classes ("terms") used in Linked Data
  – Terms are identified using URIs and labelled with natural language
  – Terms are defined in machine-readable schemas and (usually) also human-readable documents
Problem: “Long-term preservation of RDF vocabularies”

• Value of a URI vocabulary depends on perceived certainty that it remain reliably accessible over time.

• As givers of meaning to datasets, vocabularies are vital to cultural memory.

• Preservation must be planned for a timeframe beyond the horizon of any institution today.

Requirements for preservation

1. Persistence of URIs (institutional guarantees)
2. Persistence of documentation
3. Appropriate versioning
Req 1. Institutional guarantees for persistence of URIs

- Owners of URI domains should publish commitment:
  - URIs will refer to the same terms in perpetuity
  - URIs will not be repurposed.
Req 2. Persistent documentation

• Term URIs should remain resolvable to description "namespace documents" (e.g., RDF schemas).

• Links between URIs and documents must be "persistent" (i.e., maintained).
Req 3. Historical versioning

• Versioning vocabularies:
  – URIs both time-specific (“frozen”) and time-generic (“latest”) snapshots
Towards a solution

• Cooperation among vocabulary maintainers
  – e.g., DCMI/FOAF cooperation
• Cooperation of maintainers with memory institutions
• Flexible cooperation among memory institutions
• "Safety through redundancy" principle
• Access to historical versions
Cooperation of maintainers with memory institutions

• Standard contracts covering rapid interventions, transfer of domain names, and long-term responsibility?

• Major institutions ensuring that key vocabularies are preserved?
Flexible cooperation among memory institutions

• Can any institution really guarantee a preservation commitment of decades?
• Redundancy strategies imply flexible arrangements among memory institutions?
• Cover institutions large and small, from the “guaranteed” to the fragile, in countries rich and poor?
"Safety through redundancy" principle

- Repositories mirroring caches of each others' data
  - "Lots Of Copies Keep Stuff Safe" (see LOCKSS)
  - Like how Internet DNS is cached. How organisms ensure survival of genes.
Access to historical versions

- Publication of time-stamped snapshots
- Memento protocol for accessing time-specific URIs given a time-generic URI and a datetime
Agenda

• Part 1 -- 11:30 to 13:00
  – Introduction
  – Lightning talks
    • DCMI perspective (Tom Baker, DCMI)
    • W3C perspective (Ivan Herman, W3C)
    • PURL.org (Richard Wallis, OCLC)
    • European Publications Office: Metadata Registry (Michael Duero, OPOCE)
    • Linked Open Vocabularies (Bernard Vatant, Mondeca)
  – Discussion

• Part 2 -- 14:30 to 16:00
  – Lightning talk
    • LOCKSS (David Rosenthal, LOCKSS)
  – Discussion