

PROV-SEM overview

James Cheney

Problem #1

*When any new language design project is nearing completion, there is always a mad rush to get new features added before standardization. **The rush is mad indeed, because it leads into a trap from which there is no escape.** A feature which is omitted can always be added later, when its design and its implications are well understood. A feature which is included before it is fully understood can never be removed later.*

—C.A. R. Hoare, Turing Award lecture, 1980

Problem #2

- PROV-O mismatches with PROV-DM
 - Not clear that it's complete
 - Not clear that it's a "good fit"
 - Not clear how we measure completeness or goodness-of-fit

Problem #3

- "CISC" (PROV-DM)
 - aimed at casual use, "scruffies"?
 - wasGeneratedBy
- "RISC" (PROV-O)
 - aimed at sophisticated/pedantic users?
 - QualifiedInvolvement/explicit events

How can we avoid these problems?

- Formally specify "meaning" of PROV-DM?
 - Sanity check/justification for features or inferences
 - Hopefully alignable with semantics of PROV-O (inherited from RDF & OWL)
- Formally specify mapping from PROV-DM to PROV-O?
 - Rough measure of completeness, goodness of fit
 - Hopefully, implementable automatic translations

Questions

1. What is the goal of the formal semantics ?
2. What metric(s) will we use? (how do we know it's "not done")?
3. What process will we use to reconcile features of PROV-DM or PROV-O that complicate the semantics (or cannot be modeled cleanly)?

Plan

- Update semantics draft to fit Prov-DM 3WD
 - & format as HTML for FPWD of note
 - raise issues on PROV-DM or PROV-O?
- Solicit comments, vote by end of February
 - May be able to enlist help from Dagstuhl participants?
- Up or down:
 - if there is no clear consensus that it's helpful, & I'm the only one working on it, effort probably better spent elsewhere