Recording Provenance of Distributed Applications

Peter Buneman, Adria Gascon and Luc Moreau

- A research group (A) communicates with a clinical data provider (B).
- Neither side wants to reveal their provenance to the other – e.g because of patient confidentiality and confidentiality of research findings
- They want to record enough about the communication that a provenance record of the entire process could be constructed if needed – e.g. because the research had discovered the existence of a patient at risk.

PROV-AQ offers elements of solution but too much is left unspecified.

Requirements:

- Record this pairing in PROV.
- It should be possible to use PROV for reasoning about combined provenance.
- The parties should disclose as little information as possible about their local provenance, in particular URI naming schemes should be kept private.
- A third party examining A's provenance should find enough information to enable it to find the provenance graph of the recipient of A's message to B. (And vice-versa)
- We should not have to rely on a new "authority" to generate URIs
- Information regarding attribution should be preserved in the combined provenance graph, e.g. it should be easy to check whether a given entity in the joint provenance graph was generated by A or B.
- The combined provenance graph should support forward and backward reachability queries.