Abstract:

When metadata is distributed, combined, and enriched as Linked Data, the tracking of its provenance becomes a hard issue. Using data encumbered with licenses that require attribution of authorship may eventually become impractical as more and more data sets are aggregated—one of the main motivations for the call to open data under permissive licenses like CC0. Nonetheless, there are important scenarios where keeping track of provenance information becomes a necessity. A typical example is the enrichment of existing data with automatically obtained data, for instance as a result of automatic indexing. Ideally, the origins, conditions, rules and other means of production of every statement are known and can be used to put it into the right context.

In RDF, the mere representation of provenance — i.e., statements about statements — is challenging. We explore the possibilities, from the unloved reification and other proposed alternative Linked Data practices through to named graphs and recent developments regarding the upcoming next version of RDF. The session closes with a brief overview of vocabularies that can be used to actually express the provenance. This lays the ground for the PROV tutorial in the afternoon, where the two most interesting and at the same time most diverse approaches, W3C PROV and Dublin Core as a provenance vocabulary, will be introduced in detail.

There will be time to discuss use cases and open challenges
contributed by the participants. Please contact the organizer for details, if you would like to contribute a case.

**Who Should Attend:** The tutorial is intended for Linked Data practitioners who know the basic concepts of RDF and Linked Data and are interested in possible ways to publish data about the Linked Data.

**Learning Outcomes:** Participants will understand the general problems that arise if provenance information for Linked Data is to be represented and get an overview on existing solutions and best practices with their respective advantages and disadvantages.