Contribution ID: **46** Type: **Oral Presentation**

Provenance as a Building Block for an Open Science Infrastructure

Friday, 23 March 2018 09:00 (30 minutes)

In science, results that are not reproducible by peer scientists are valueless and of no significance. Good practices for reproducible science are to publish used codes under Open Source licenses, perform code reviews, save the computational environments with containers (e.g., Docker), use open data formats, use a data management system, and record the provenance of all actions.

This talk focuses on provenance of scientific processes as a foundation of open reproducible science and a building block of a distributed trustful open science infrastructure.

The concept of provenance is introduced and the W3C standard model PROV is presented. PROV gives an ontology, a data model, and specifications for provenance notations, for accessing and querying, and for mapping to other standards. For practical use of provenance in science processes, the talk gives strategies for recording and storing provenance from scientific workflow systems. Also recording provenance from scripts and from the generation of documents are presented.

For storing provenance, the talk shows how to facilitate graph databases (such as Neo4j), since the provenance of processes is a directed acyclic graph (DAG). Based on graph query languages such as Cypher or GraphQL, these provenance information can be analyzed. For example, the provenance can be used to proof the compliance of a scientific process, to collect all data that contributed to to scientific results such a journal paper, or detect any issues regarding privacy, security, and trust of the scientific data.

To assure trust in the scientific process, we present a technology for storing provenance and related data in blockchains and blockchain-like databases. Using blockchains for storing (provenance) graphs, one can detect if the graphs have been manipulated. This gives a higher degree of confidence that the scientific data has been produced using the described processes. We describe how provenance graphs are mapped to blockchains and show how to used blockchains implementations for storing this provenance graphs practically.

Primary author: Mr SCHREIBER, Andreas (German Aerospace Center)

Presenter: Mr SCHREIBER, Andreas (German Aerospace Center)

Session Classification: Networking, Security, Infrastructure & Operation Session

Track Classification: Networking, Security, Infrastructure & Operations