Representing Humanities Research Data using Complementary Provenance Models

Bridget Almas, Monica Berti, Sayeed Choudhury, David Dubin, Megan Senseney, Karen M. Wickett.

Overview

The Systematic Assertion Model (SAM) is a formal account of research data and its content - relating data identity, change over time, and varieties of scientific or scholarly equivalence directly to research transactions. Humanities use cases include a breadth of computational and interpersonal activities that call for a more inclusive understanding of provenance than SAM alone can provide. Integrating complementary views of the same research enterprise by combining SAM entities and properties with those from other models such as W3C’s PROV and Open Annotation (OA) provides a richer, more contextualized view of data encoding and use in humanities research.

Use Case Details


B’s Assertion [B2] is warranted by A’s ancient Assertion [A1].

Summary: A scholar wants to annotate what she thinks is a quotation or text re-use of a lost text attributed to Istros in an extant source text by Plutarch (a reference to the origin of the word ‘sycophants’ where Istros is not named). To substantiate her argument, the scholar must also identify corroborating material, including instances of Istros’ text in other primary sources, in this example that of Athenaeus Deipn. 3.6 (who does name Istros as the source).

Conclusion

Combining complementary data provenance models enables us to more precisely track and document shared resources, ultimately improving data quality and encouraging further sharing. Using PROV Activities we share precise details about system actions and processes leading to the extraction of text and creation of URI endpoints. We use SAM to identify the contingent aspect of the underlying resources as things which are subject to interpretation and which were in existence prior to their use as data in our analysis. We use OA to share concrete serializations of the analyses in the form of annotations.