Where did this data come from and how was it made? Introducing a provenance model into ADAPT

Car, Nick; Cox, Simon; Ferreyra, R. Andres; Rhea, Stuart T.

The agricultural industry is increasingly interconnected and the volume of data being exchanged among growers and other actors is rapidly escalating. It is also becoming critically important to document the *provenance* of the digital objects being exchanged; i.e., information about the events and people involved in producing, delivering, or modifying a piece of data or thing. This provenance can be used to form assessments about an object of interest's quality, reliability or trustworthiness, and can be crucially important for mitigating liability.

The World-Wide Web Consortium (W3C) has recently published a provenance standard, PROV. This standard accommodates agent-centered provenance (e.g., Who created this data set? Who collected or tested this soil sample?), entity-centered provenance (tracing the origins of a document or data, or parts thereof, to other documents and data; e.g., From which raw data does this calculated result derive? What outputs derive from this input?) and process-centered provenance (describing what steps were taken to generate a particular piece of information; (e.g., What corrections have been applied to this yield dataset? What geometric and/or radiometric corrections have been applied to this drone-collected remote sensing dataset?)

The requirements-gathering work of AgGateway's field operations projects (SPADE, PAIL) motivated work on an open-source format-conversion toolkit, ADAPT. At the core of ADAPT is a common object model that provides an opportunity for the industry to transcend long-standing interoperability limitations and exchange data in a common format. This platform also provides a valuable opportunity to introduce a provenance framework that can be used to support traceability and principled decision-making in agriculture. This paper describes how ADAPT can use the PROV model. It describes the constraints involved (e.g., introducing PROV required a non-breaking change in the ADAPT code base) and the rationale for selecting a particular solution from a set of possible implementations that are available when a domain model seeks to adhere to PROV formalisms. We give specific examples of using ADAPT+PROV to document field operations: tracking change in reference data and variable type registries; documenting the lifecycle of machine configuration data; and expressing the causal relationships among field operation business process Core Documents (i.e., Plan, Observations and Measurements, Recommendations, Work Orders, and Work Records).

Author information:

- Car, Nick
  Geoscience Australia
  Melbourne (Australia)
  Nicholas.Car@ga.gov.au

- Cox, Simon
  CSIRO
  Melbourne (Australia)
  Simon.Cox@csiro.au

- Ferreyra, R. Andres (ASABE Member, **CONTACT AUTHOR**)
  Ag Connections, LLC
  1576 Killdeer Trail
  Murray, KY 42071 (United States)
  (270) 227-0778
  andres.ferreyra@agconnections.com

- Rhea, Stuart T.
  Ag Connections, LLC
  Murray, KY (United States)
  stuart.rhea@agconnections.com