Provenance of Figures in the Global Change Information System

Justin Goldstein (jgoldstein@usgcrp.gov)\textsuperscript{1,2}; Xiaogang Ma (max7@rpi.edu)\textsuperscript{3}; Jin Zheng (zhengj3@rpi.edu)\textsuperscript{3}; Robert David (rdavid@usgcrp.gov)\textsuperscript{1,2}; Curt Tilmes (ctilmes@usgcrp.gov)\textsuperscript{1,4}; Ana Pinheiro Privette (ana.privette@noaa.gov)\textsuperscript{3}; Steven Aulenbach (saulenbach@usgcrp.gov)\textsuperscript{1,2}; Megan McVey (mmcvney@usgcrp.gov)\textsuperscript{1,2}; Peter Fox (foxp@rpi.edu)\textsuperscript{3}

\textsuperscript{1}US Global Change Research Program, \textsuperscript{2}University Corporation for Atmospheric Research, \textsuperscript{3}Rensselaer Polytechnic Institute, \textsuperscript{4}NASA-Goddard Space Flight Center, \textsuperscript{5}North Carolina State University, Cooperative Institute for Climate and Satellites – NC

Introduction

The Draft of the 2013 National Climate Assessment (NCA), developed by the US Global Change Research Program, is a US government document which thoroughly describes the impact of climate change on the United States. It will serve as the base of the Global Change Information System (GCIS), which is a portal allowing users to interact with the NCA and to trace the provenance of figures and data sources used in the NCA using the ISO 19115: 2003 standards. The goal of provenance tracking within the GCIS is to provide information to allow a user to reproduce an image. However, the tracking of provenance is a complex task due to the vast amount of information for which metadata needs to be captured and modeled (Tilmes et al. 2013), as well as problems with the availability of data sources, especially non-archived outputs from scientific investigations which need to be tracked down individually. Here, we present a sample process of lineage tracing for a particular NCA figure lacking a complete set of metadata. The approach of lineage tracing is described here in three ways: (a) a graphical, information representation of the provenance scenario, (b) a formal provenance diagram using terminology from the W3C PROV Data Model and Ontology, and (c) a RDF description serialized in Turtle format.

(a) Graphical, Informal Representation of Provenance

![Diagram of provenance](Image)

(1) The Cddv2 precipitation and temperature dataset is clipped to the domain of the Great Plains region defined in the NCA. The characteristics of the original dataset (light green) will be provided with IDs and URLs for use in the GCIS.

(2) This dataset is used in the production of an image in a document written by Ken Kunkel.

(3) After undergoing some aesthetic changes made by Mike Squires and Jessica Griffin, the image in (2), presented in the informal illustration on the left-hand side of the poster, is displayed in the NCA. Metadata is attached to all items.

Acknowledgements

We thank Stephan Zednik (Rensselaer Polytechnic Institute) for his contributions to the GCIS provenance modeling.

(b) Formal W3C PROV Data Model and Ontology

```
<http://data.globalchange.gov/paper/10>
  a prov:Entity;
  dct:terms:title "Climate of the U.S. Great Plains";
  prov:wasAttributedTo
    <http://data.globalchange.gov/person/Kenneth_E_Kunkel>;
  prov:wasGeneratedBy
    <http://data.globalchange.gov/activity/writing/paper/10>;
  .

<http://data.globalchange.gov/activity/writing/paper/10>
  a prov:Activity;
  prov:wasAssociatedWith
    <http://data.globalchange.gov/person/Kenneth_E_Kunkel>;
  prov:used <http://data.globalchange.gov/dataset/103>;
  .

<http://data.globalchange.gov/dataset/103>
  a prov:Entity;
  prov:wasGeneratedBy
    <http://data.globalchange.gov/activity/dataset_generating/dataset/103>
```  

(c) Turtle Representation (portion)

```
<http://data.globalchange.gov/paper/10> a prov:Entity;
dcterms:title "Climate of the U.S. Great Plains";
prov:wasAttributedTo <http://data.globalchange.gov/person/Kenneth_E_Kunkel>;
prov:wasGeneratedBy <http://data.globalchange.gov/activity/writing/paper/10>;
.

<http://data.globalchange.gov/activity/writing/paper/10> a prov:Activity;
prov:wasAssociatedWith <http://data.globalchange.gov/person/Kenneth_E_Kunkel>;
prov:used <http://data.globalchange.gov/dataset/103>;
.

<http://data.globalchange.gov/dataset/103> a prov:Entity;
prov:wasGeneratedBy <http://data.globalchange.gov/activity/dataset_generating/dataset/103>
```

Summary

A sample approach to tracking the provenance of an figure in the NCA Draft is presented using three different representations: (a) a graphical, informal diagram, (b) a formal PROV data model and ontology representation, and (c) a Turtle representation. Simplified versions of these representations are presented due to the multiple layers of complexity and the multifaceted nature of various images as multiple data sources and figures may be included in one illustration. Not only does the process of provenance tracing require the locating of metadata, it often involves the development of approaches to handle instances of non-archived data.

Reference