Ontology Development for Provenance Tracing in National Climate Assessment of the US Global Change Research Program

Linyun Fu (fu20@rpi.edu), Xiaogang Ma (max7@rpi.edu), Jin Zheng (zhengj3@rpi.edu), Justin Goldstein (jgoldstein@usgcrp.gov), Brian Duggan (bduggan@usgcrp.gov), Patrick West (westp@rpi.edu), Steve Aulenbach (aulenbach@usgcrp.gov), Curt Tilmes (Curt.Tilmes@nasa.gov), Peter Fox (pfox@cs.rpi.edu)

Tetherless World Constellation, Rensselaer Polytechnic Institute, Troy, NY, United States
U.S. Global Change Research Program, Washington D.C., United States
NASA Goddard Space Flight Center, Greenbelt, MD, United States

BACKGROUND

The United States Global Change Research Program (USGCRP) is an interagency program that coordinates and integrates Federal research on changes in the global environment and their implications for society.

The National Climate Assessment (NCA) is an assessment conducted under the auspices of the Global Change Research Act of 1990, which requires a report to the President and Congress every four years that evaluates, integrates and interprets the findings of the USGCRP with the intent to advance an inclusive and sustained process for assessing and communicating scientific knowledge of the impacts, risks and vulnerabilities associated with a changing global climate in support of decision making across the United States.

The Global Change Information System (GCIS) is an information system under development through the USGCRP that establishes data interfaces and interoperable repositories of climate and global change data which can be easily and efficiently accessed, integrated with other data sets, maintained over time and expanded as needed into the future.

Provenance is information about entities, activities, and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability or trustworthiness. The goal of PROV is to enable the wide publication and interchange of metadata about the Web and other information systems. The PROV Ontology (PROV-O) defines the OWL2 Web Ontology Language encoding of the W3C Provenance Data Model.

WHAT WE DO

Our ongoing work is capturing provenance in the draft of the Third National Climate Assessment (NCA3) report of the USGCRP.

Our future work is to capture provenance of publications, datasets, models, organizations, instruments, experiments, people, etc., eventually covering the entire scope of global change.

We follow a use case driven iterative approach.

CURRENT RESULT

GCIS ontology version 1.1 and 1.2
(a) Classes and properties representing a brief structure of the draft NCA3
(b) Classes and properties related to the findings of the draft NCA3 and each chapter in it
(c) Classes and properties about sensors, instruments, platforms, and algorithms, etc. that datasets are derived from

To find the ontology: gcis ontology rpi

Glossary:
USGCRP - United States Global Change Research Program
NCA - National Climate Assessment
GCIS - Global Change Information System
PROV-O - The PROV Ontology, which defines the OWL2 Web Ontology Language encoding of the W3C Provenance Data Model

Posters: EGU2014-2360

We present a PROV-based tool for the National Climate Assessment and use cases for use in other Federal programs.

A reader wishes to identify the source of the data used to produce a particular figure in the draft NCA3. A reference to the paper in which the image contained in this figure was originally published appears in the figure caption. Clicking that reference displays a page of metadata information about the paper, including links to the datasets used in that paper. Pursuing each of those links presents a page of metadata information about the dataset, including a link back to the agency/data center web page describing the dataset in more detail and making the actual data available for order or download.

A viewer sees that Chapter 6 (Agriculture) in the draft NCA3 was written by a group of chapter authors mentioned in a list. On the title page of that chapter the reader can view the role of each author, e.g., convening lead author, lead author or contributing author, in the generation of this report chapter.

We used the PROV-O ontology to describe this use case.

We use the PROV-O ontology to describe the publication-activity-contributor relationship in GCIS.

A viewer sees that the caption of Figure 1.2 “Sea Level Rise: Past, Present and Future” of the draft NCA3 cites four data sources. Selecting the third citation displays a page of information about the cited paper and a citation to the dataset used in that paper. Information about the dataset includes a formal description of its origin, that is, the dataset is derived from data produced by the TOPEX/Poseidon, Jason-1 and Jason-2 altimeter missions funded by NASA and CNES. Clicking a link to each of these missions presents a page about the platforms, instruments and sensors in that mission.

We adapted the entity-activity-agent relationship in PROV-O to describe the publication-activity-contributor relationship in GCIS.