

Provenance Representation in the Global Change Information System (GCIS)

Curt Tilmes

U.S. Global Change Research Program,
1717 Pennsylvania Ave NW, Suite 250
Washington D.C. 20006, USA.
Curt.Tilmes@nasa.gov

Abstract

Global climate change is a topic that has become very controversial despite strong support within the scientific community. It is common for agencies releasing information about climate change to be served with Freedom of Information Act (FOIA) requests for everything that led to that conclusion. Capturing and presenting the provenance, linking to the research papers, data sets, models, analyses, observation instruments and satellites, etc. supporting key findings has the potential to mitigate skepticism in this domain.

The U.S. Global Change Research Program (USGCRP) is now coordinating the production of a National Climate Assessment (NCA) that presents our best understanding of global change. We are now developing a Global Change Information System (GCIS) that will present the content of that report and its provenance, including the scientific support for the findings of the assessment. We are using an approach that will present this information both through a human accessible web site as well as a machine readable interface for automated mining of the provenance graph. We plan to use the developing W3C PROV Data Model and Ontology for this system.

1 Background

The U.S. Global Change Research Program (USGCRP)¹ coordinates and integrates federal research on changes in the global environment and their implications for society. The USGCRP began as a presidential initiative in 1989 and was mandated by Congress in the Global Change Research Act of 1990[1] (GCRA), which called for “*a comprehensive and integrated United States research program which will assist the Nation and the world to understand, assess, predict, and respond to human-induced and natural processes of global change.*”

Thirteen U.S. federal departments and agencies participate in the USGCRP: Department of Commerce, Department of Defense, Department of Energy, Department of the Interior, Department of State, Department of Transportation,

¹ <http://globalchange.gov>

Department of Health and Human Services, Department of Agriculture, National Aeronautics and Space Administration, National Science Foundation, Smithsonian Institution, Agency for International Development and the Environmental Protection Agency.

The USGCRP is developing a Global Change Information System (GCIS) that will utilize the developing W3C PROV² recommendations to eventually represent the provenance for all of the information related to global change across the U.S. federal government. The first implementation will provide provenance for the National Climate Assessment (NCA).

2 National Climate Assessment (NCA)

The GCRA requires a report to the President and the Congress every four years that integrates, evaluates, and interprets the findings of the USGCRP; analyzes the effects of global change on the natural environment, agriculture, energy production and use, land and water resources, transportation, human health and welfare, human social systems, and biological diversity; and analyzes current trends in global change, both human-induced and natural, and projects major trends for the subsequent 25 to 100 years.

The National Climate Assessment and Development Advisory Committee (NCADAC) is a Federal Advisory Committee[2] with 60 members, including 45 non-federal members and 16 federal ex-officio representatives. that provides advice and recommendations for the NCA process. As of this writing, the NCA has defined 30 chapters and selected 62 “Convening Lead Authors” and 180 “Lead Authors.” The names and institutional affiliations of 240 contributing authors are a critical part of the provenance of the NCA we will be capturing with the process. All of that information will be of course be part of the printed and web-based text of the document, but will also be represented through machine accessible APIs.

Through an open, public process, the NCA has received over 500 distinct technical inputs, many of which are reports distilling and synthesizing even more information, coming from thousands of individuals around the federal government, non-governmental organizations, academic institutions, etc. The inputs include peer-reviewed scientific publications, model data, observational data (physical, societal, economic), historical data, sectoral and regional assessments, and data at a variety of scales and resolutions. Most original data are archived in long term agency data centers responsible for long term stewardship of the items, but some includes unconventional information collected from public health departments, states and tribes, NGOs, and data collected but not yet reviewed. Where the data are transformed into new graphics, graphs or charts, the process and methods used must be clearly and reproducibly documented.

This poses a tremendous challenge (and opportunity!) for provenance capture, archive, and presentation. We will represent that information using the

² <http://www.w3.org/TR/prov-dm/>

PROV ontology and make the complete information about the NCA itself as well as all of the inputs to the process available through a publicly accessible web site and SPARQL end point. The GCIS will provide links from the content and findings of the NCA back to all of their predecessor artifacts.

3 Provenance Representation

The GCIS assigns globally unique, persistent identifiers to all of the entities, activities and agents relevant to our discussions of provenance. These are located in the USGCRP namespace rooted under <http://globalchange.gov/id>. We are linking to existing identifiers where possible and appropriate, using journal or data center assigned DOIs for papers and datasets. NASA's Global Change Master Directory³ has also assigned reusable identifiers for many of the important datasets and services we are referencing. PROV can be extended with domain defined types and specialized agent roles like the "Convening Lead Authors."

All of the globalchange.gov URI identifiers will be resolvable through HTTP content negotiation to either human readable HTML web pages, or machine readable encodings of the metadata describing the item and linking back to the repository for that item (such as a journal site for a paper, or an agency data center for an observational dataset). Where items are derived from other items, they will link back to their predecessor "entities," and "activity" representations with sufficient detail to reproduce the activity.

As an exercise to explore alternative methods of presenting the NCA, the previous report *Global Climate Change Impacts in the United States (2009)*[3] was transformed into a web site⁴ with additional pages for each figure and footnote to more information including links back to datasets and data centers.

4 Future Plans

The GCIS is very much a work in progress. We have only begun mapping the myriad of resources into the PROV Data Model. Indeed, at this time, PROV itself is not yet complete, being only a "public working draft." Nevertheless, using PROV to describe the provenance of the NCA will have benefits for each. Beyond the NCA and other synthesis reports, the GCIS will be used to present information about global change from across the agencies of the U.S. Global Change Research Program.

References

1. U.S. Code: Global Change Research Act of 1990 (P.L. 101-606) (1990)
2. U.S. Code: The Federal Advisory Committee Act (5 U.S.C. App.) (1972)
3. U.S. Global Change Research Program: *Global Climate Change Impacts in the United States* (2009)

³ <http://gcmd.nasa.gov>

⁴ <http://nca2009.globalchange.gov>