

Tracking and Establishing Provenance of Earth Science Datasets: A NASA-Based Example

Hampapuram K. Ramapriyan^{a, *}, Justin C. Goldstein^{b, c}, Hook Hua^d, Robert E. Wolfe^{b, e}

a. Science Systems and Applications, Inc., Lanham, MD b. U.S. Global Change Research Program (USGCRP), Washington, DC c. ICF International, Fairfax, VA
d. Jet Propulsion Laboratory (JPL), California Institute of Technology, Pasadena, CA e. NASA Goddard Space Flight Center (GSFC), Greenbelt, MD
* corresponding author: Hampapuram.Ramapriyan@saihq.com

Abstract

Information quality is of paramount importance to science. Accurate, scientifically vetted and statistically meaningful and, ideally, reproducible information engenders scientific trust and research opportunities. Not surprisingly, federal bodies (e.g., NASA, NOAA, USGS) have very strictly affirmed the importance of information quality in their product requirements. So-called Highly Influential Scientific Assessments (HISA) such as The Third US National Climate Assessment (NCA3) published in 2014 undergo a very rigorous review process to ensure transparency and credibility. To support the transparency of such reports, the U.S. Global Change Research Program (USGCRP) has developed the Global Change Information System (GCIS). **A recent activity was performed to trace the provenance as completely as possible for all NCA3 figures that were predominantly based on NASA data.** This poster presents the mechanics of that project and the lessons learned from that activity.

“NASA requires a higher standard of quality for information that is considered influential. Influential scientific, financial, or statistical information is defined as NASA information that, when disseminated, will have or does have clear and substantial impact on important public policies or important private sector decisions.”

http://www.nasa.gov/sites/default/files/517756main_FINAL_NASA_guidelines.pdf

Provenance Information in the Global Change Information System (GCIS)

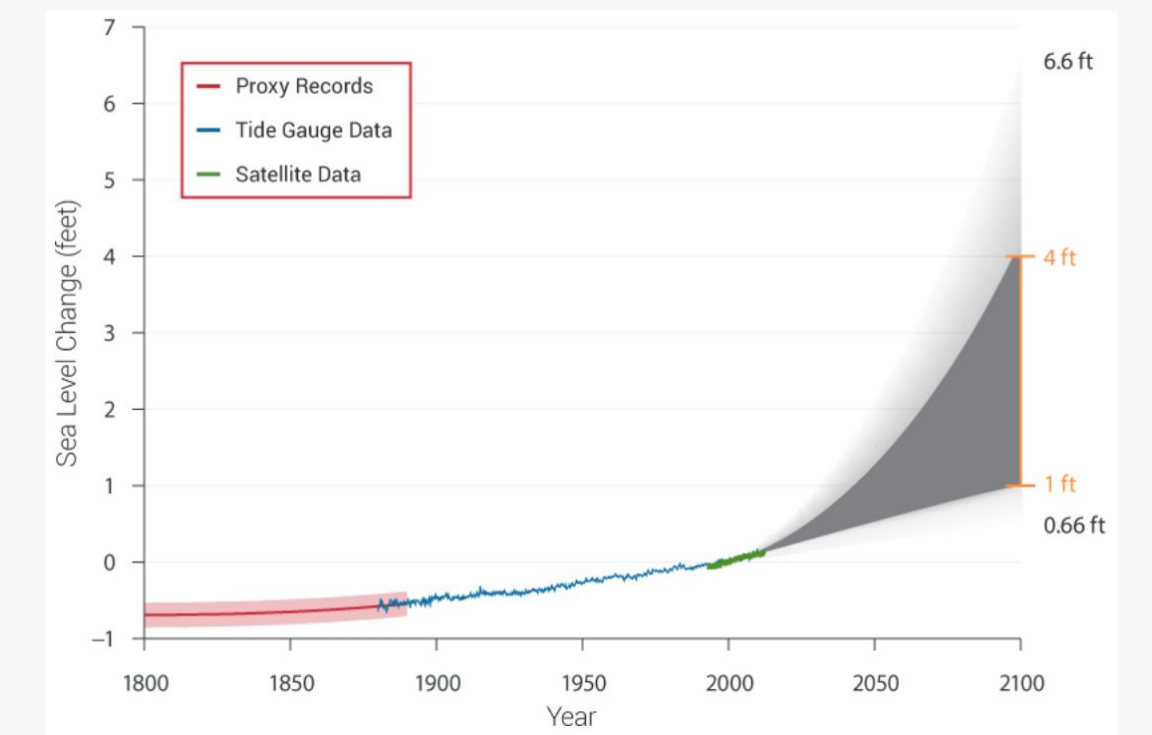


<http://data.globalchange.gov>

NASA Figure Example in NCA3

Figure 5.1 Past and Projected Changes in Global Sea Level Rise

Supports Key Message: “Global sea level has risen by about 8 inches since reliable record keeping began in 1880. It is projected to rise another 1 to 4 feet by 2100.”

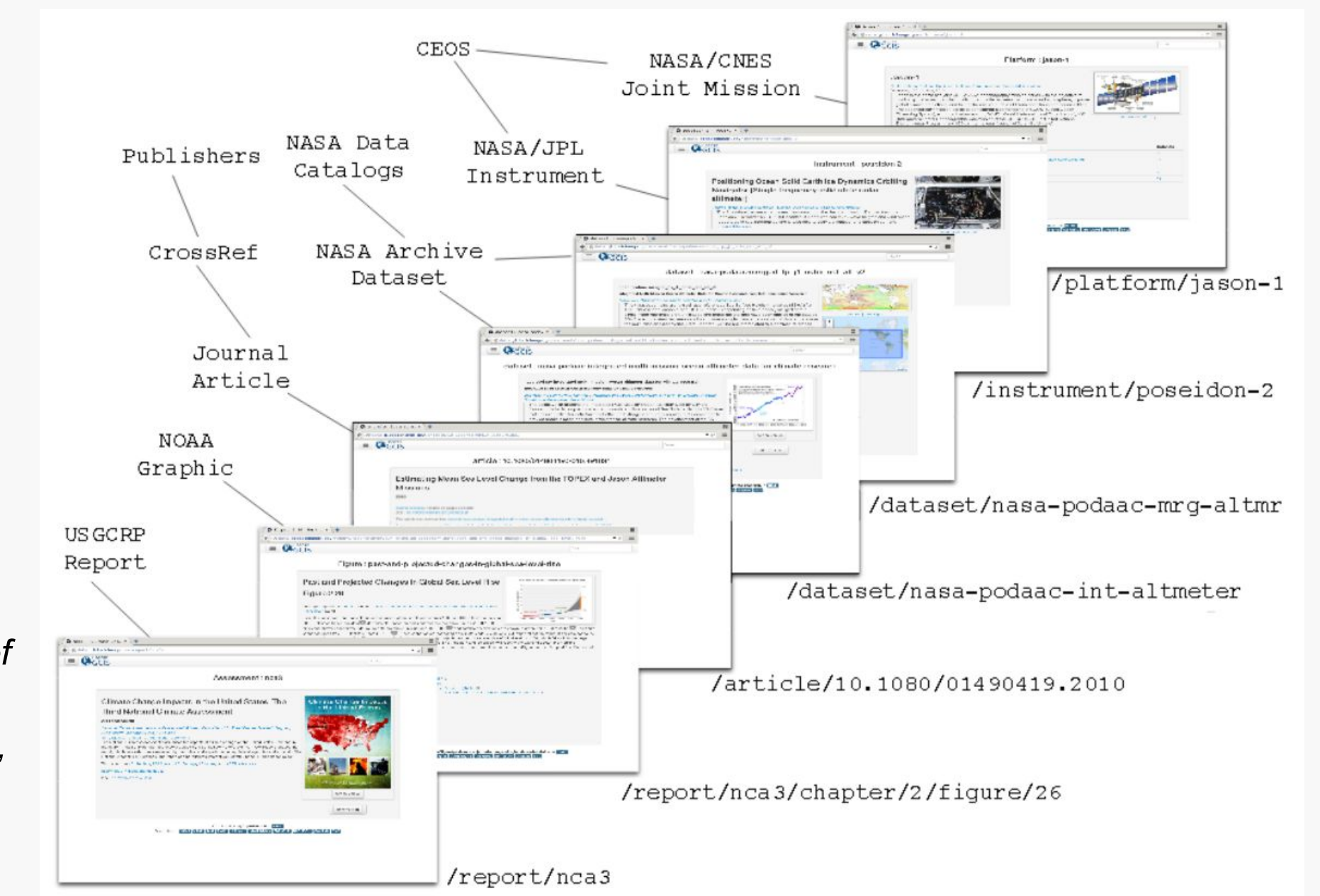


<http://nca2014.globalchange.gov/highlights/report-findings/future-climate/graphics/past-and-projected-changes-global-sea-level>

Includes NASA satellite data from 1993 to 2012.

Provenance provided by GCIS metadata links to instrument and platform via reference and dataset.

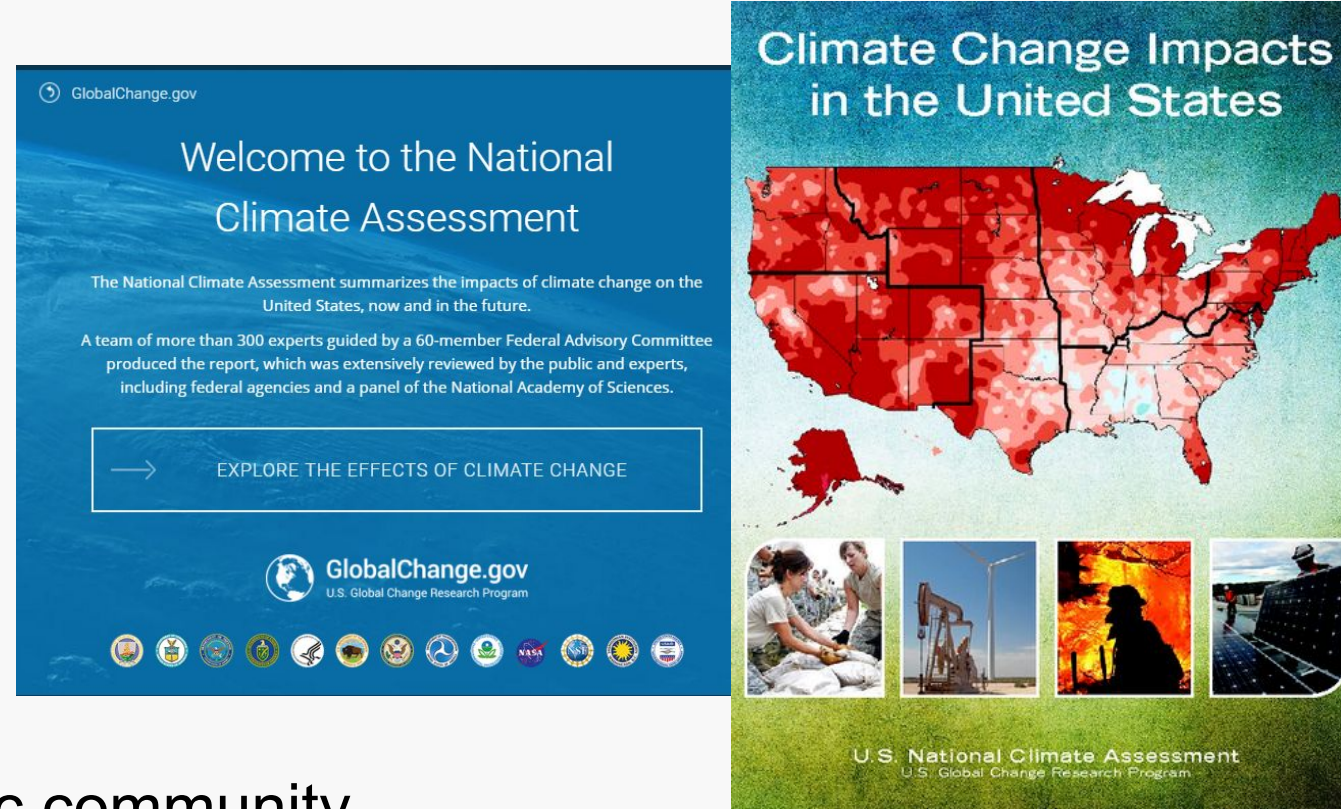
An example from one of 20 of the NCA3 figures that use NASA satellite, instrument and/or model data.



NCA3 and GCIS

The Challenge: NCA3 (<http://nca2014.globalchange.gov>)

- Summarizes the impacts of climate change on the United States, now and in the future
- Produced in compliance with the Global Change Research Act of 1990 and delivered to the U.S. Congress in May 2014
- Written with extensive contributions from the scientific community
 - 30 chapters and supplementary material, including 12 Findings, 149 Key Messages, 290 figures, 20 tables and 3395 refs.

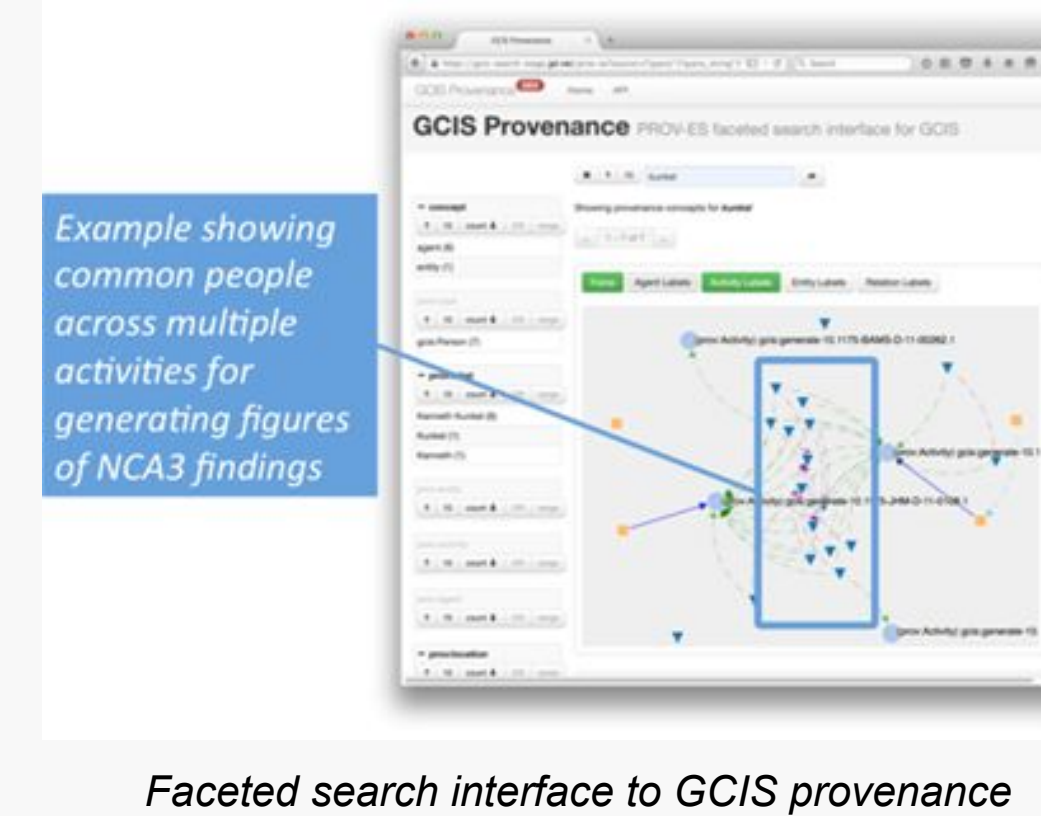
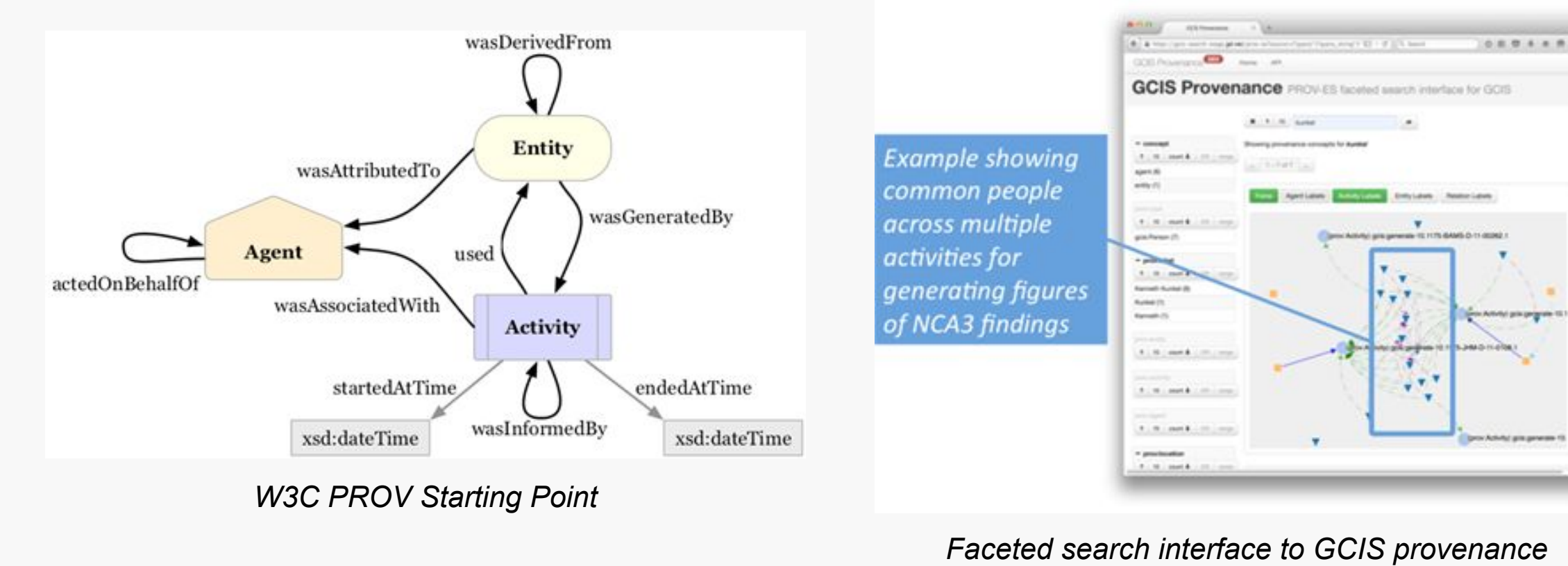


The Approach: GCIS (<http://data.globalchange.gov>)

- Developed to better coordinate and integrate the use of federal information products on changes in the global environment and the implications of those changes for society
- Open-source, web-based resource with API for traceable, sound global change data, information, and products
- Key access point to assessments, reports and tools produced by USGCRP
- Each GCIS item has a unique, persistent identifier in the form of Uniform Resource Identifier (URI) that may be related to other identifiers (e.g. DOIs)
- Maintains two concurrent models of the data: a **relational** model, and a **semantic** model (supports SPARQL queries)

Provenance Tracking in GCIS

- Approach:** The NASA Earth Science Data Systems Working Group (ESDSWG) is developing a Provenance for Earth Science (PROV-ES) extension, which is used by GCIS for provenance codification. PROV-ES can also be used by systems outside of NASA.
 - Key GCIS concepts and attributes are mapped onto W3C PROV types
- Provenance information collected for NASA figures and images prior to publishing NCA3 was sparse
 - After publication, NASA provenance was collected for representation in GCIS by contacting report authors or through other NASA connections



- GCIS discovery service includes a **provenance faceted search capability** enabling users to facet navigate GCIS resources in the context of provenance
 - Enables users to “drill-down” by applying a sequential set of selection criteria across different facets (values) of the GCIS content

Lessons Learned and Recommendations

- It is difficult to trace back to derive provenance after reports are completed and delivered.
 - Authors are busy individuals who have spent a considerable effort into research and writing the report
 - Authors typically have not maintained complete documentation
 - Means more work for the authors or an independent effort
- Attempts to follow up with authors on provenance is sometimes misinterpreted as questioning their research
- It is useful to provide the authors with detailed instructions and templates before they start writing
 - The information needed is the inputs, outputs and methods (descriptive and/or mathematical) for each dataset used, images or figures generated, and key messages
- An independent team should be involved during the generation of a report to check for completeness of traceability from a non-expert reader’s point of view
 - This helps minimize the impact on the report publication schedule
- Underlying information should be held in a long-lived repository and be easily accessible to users for at least as long as the reports are deemed to be of interest to the community
 - Especially for a Highly Influential Scientific Assessment
- These lessons are being applied to future USGCRP assessments
 - Future work is needed to classify these lessons (and other lessons real-world uses of provenance) and associated them with high-level topics

USGCRP assists the U.S. and the world to understand, assess, predict and respond to human-induced and natural processes of global change



USGCRP - Thirteen Agencies, One Vision: Empower the Nation with Global Change Science

Acknowledgements

We acknowledge contributions to GCIS from many current and former colleagues at NASA, NOAA, USGS, OSTP and other agency partners; at USGCRP National Coordination Office and Technical Support Unit; as well as JPL and RPI.