# Learn to Write a Geosciences Paper of the Future:

## Best Practices for Digital Scholarship, Reproducible Research, and Open Science

http://www.scientificpaperofthefuture.org/gpf/
ontosoft@gmail.org

## Session Description

This tutorial covers best practices in reproducible research, open science, and digital scholarship that help researchers increase citations for their papers, get credit for all their research products, augment their vitae with data and software that they have written, write compelling data management plans for funding proposals, comply with new funder and journal requirements, and practice open and reproducible science. We begin with a motivation for authors through an overview of why scientists, publishers, funders, and the public care about science practices. Next, we describe how to make research data accessible through publication in a public repository, including metadata, a license for reuse, and citable using a unique and persistent identifier. We then show how to make software accessible by making it available in a public repository, with a license, and a unique and citable persistent identifier. We also cover how to document in a software registry key information that helps others reuse research software. We then discuss how to document provenance and methods by explicitly describing related computations and outcomes in a workflow sketch, a formal workflow, or a provenance record, possibly with a persistent identifier. Finally, we provide a summary checklist for authors, and show how to manage their scholarly identity, reputation, and impact throughout their careers.

A Geosciences Paper of the Future (GPF) follows best practices to document all the associated digital products that result from the research reported in the paper. This means that a paper would include:

1. Data available in a public repository, including documented metadata, a clear license specifying conditions of use, and citable using a unique and persistent identifier
2. Software available in a public repository, with documentation, a license for reuse, and a unique and citable using a persistent identifier
3. Provenance of the results by explicitly describing the series of computations and their outcome in a workflow sketch, a formal workflow, or a provenance record, possibly in a shared repository and with a unique and persistent identifier

Authors of GPFs follow best practices of digital scholarship, reproducible research, and open science.

## Why Learn to Write a Geosciences Paper of the Future

- **Get credit for all your research products**: The data you generate, the software you write, and the new methods you create are all valuable research products that you can cite and get credit for.
- **Increase citations to your papers**: Studies show that papers are cited more when they document well the computational methods, data, and software.
- **Augment your vitae with data and software that you have written**: You can convey your strengths in computational methods by including citable research products in your vitae.
- **Write compelling Data Management Plans for your funders**: Proposals can be more successful if they include effective yet simple approaches to disseminating data and software.
- **Address new journal requirements**: Address the increasing demands of journal editors who ask for detailed documentation of the computational aspects of your work.
- **Practice open and reproducible science**: Adopt best practices that improve your reputation as a scientist.

## Training Topics

The training session is divided into two sessions, each 90mins with a break in the middle. Training materials are provided to all participants. For each training topic, basic concepts and best practices are introduced. A summary at the end provides specific advice and pointers to implement those best practices.

| **Sessions** | **Topics** | **Concepts Covered** |
|---|---|---|
| *I: Making data and software accessible* | Scientific publications in the future | An overview of the benefits of augmenting papers with data, software, and provenance, all properly documented and cited |
| | Making data accessible | How to publish data in a public shared repository, select and use a license, and cite it in an article |
| | Making software accessible | How to publish software in a public repository, select and use a license, and cite it in an article |
| *Break* | | |
| *II: Describing software and provenance* | Documenting software with metadata | How to describe general metadata about software so others can understand and use it |
| | Documenting the provenance of results | How to describe provenance in terms of the computations that were executed in order to obtain the results reported in a paper |
| | Documenting methods as workflows | How to describe general computational methods in a paper as data flow across software components |

## Presenter

**Yolanda Gil** is Director of Knowledge Technologies and Associate Division Director at the Information Sciences Institute of the University of Southern California, and Research Professor in the Computer Science Department. She received her M.S. and Ph. D. degrees in Computer Science from Carnegie Mellon University, with a focus on artificial intelligence. Dr. Gil collaborates with scientists in different domains on semantic workflows and metadata capture, social knowledge collection, and computer-mediated collaboration. She chaired the World Wide Web Consortium (W3C) group that led to the PROV provenance standard, and has done extensive work on reproducibility and semantic metadata. As part of her participation in the NSF EarthCube program, she and her team developed the Geoscience Paper of the Future Initiative that created training materials with best practices from digital scholarship, reproducible research, and open science. She is a Fellow of the Association for Computing Machinery (ACM), and Past Chair of its Special Interest Group in Artificial Intelligence (SIGAI). She is also Fellow of the Association for the Advancement of Artificial Intelligence (AAAI), and was elected as its 24th President in 2016.

## Acknowledgments

The Geoscience Paper of the Future (GPF) Initiative was initiated by the EarthCube OntoSoft project with funding from the National Science Foundation. The Initiative was created by the project's Early Career Advisory Board, consisting of 30 young researchers in diverse areas of geosciences. They recently published an article summarizing the recommended best practices, and many are authors of GPF articles published under a special issue of the AGU Earth and Space Science journal. They developed training materials, which are available under a Creative Commons license from http://dx.doi.org/10.5281/zenodo.159206. Pointers and more information can be found at http://www.scientificpaperofthefuture.org/gpf/.