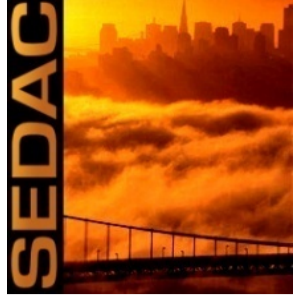# Disseminating Open Source Software with Open Data:
A Case Study from a Scientific Data Center

Robert R. Downs[1] and Pinki Mondal[2]
NASA Socioeconomic Data and Applications Center (SEDAC)
Center for International Earth Science Information Network (CIESIN), Columbia University

[1] rdowns@ciesin.columbia.edu, [2] pmondal@ciesin.columbia.edu

Prepared for Presentation to the Research Data Alliance 10th Plenary
19-21 September 2017, Montreal, Canada
Poster Session: 19-20 September 2017, Poster Number 17

Enabling the reuse of the software components (source code or software scripts) used to create or access scientific data can potentially facilitate reuse of the data by:
- fostering opportunities for replication and increasing the potential for the reproducibility of the data
- inspiring a decision to disseminate the software with the data if reusing the data requires particular software
- reducing potential barriers for reuse of the data

However, the decision to disseminate scientific software with data includes many issues including:
- additional decisions for the preparation and dissemination of the software that are similar to those associated with data dissemination but with additional complexity

Here we offer a decision tree derived from a case-study to support decision-making for software dissemination by:
- reviewing decisions to be considered in the dissemination of open source software with open data
- providing perspectives for scientific data centers and other archives that are considering open source dissemination of scientific software with scientific data products and services

Decisions Considered for Disseminating Open Source Scientific Software with Open Scientific Data

Disseminate Source Code
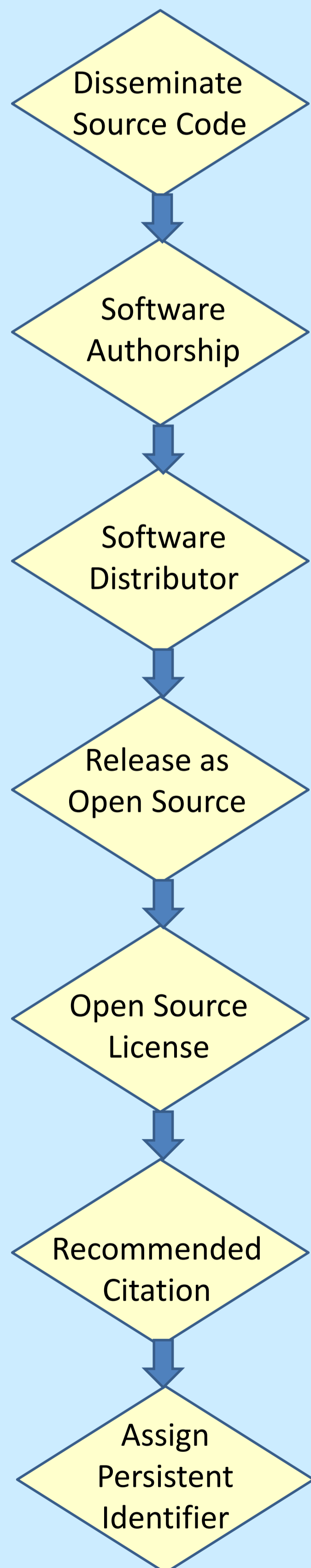
Disseminate the source code of the software to enable transparency of data processing and analysis, replication of the methods used for creating or processing the data, and possible reproducibility of the data. **The decision to disseminate the software should be made early** during the project and should be described within the data management plan and the software development plan.

Software Authorship

Identify the authorship of the software, considering the software development team members, such as the software code writers, algorithm creators, and study designers as potential candidates. The software authors are responsible for the software and are recognized for their roles as authors and providers of the software.

Software Distributor

Select a distributor to disseminate the scientific software with the scientific data. Some considerations include whether to preserve, manage, and disseminate the software along with the data and the provision of continuing services to the disciplines and communities designated to use the data.

Release as Open Source

Release as Open Source software would provide free access for others to obtain the source code and documentation for any scripts, workflows, or libraries included as part of the software package and would allow others to use the scientific software for their own purposes.

Open Source License

Apply an open source license to the software as a legal instrument that specifies the rights for use and includes a statement describing any liability associated with the use of the data. The license that is selected should be consistent with any agreements with funders and collaborators and their institutional policies and should be compatible with any software language and libraries utilized to create or use the software.

Recommended Citation

Provide a recommended citation for the software that is different than the recommended data citation and serves as an example for citing the use of the software. Note that norms for scientific disciplines and some open source software licenses may require the citation of software that is used.

Assign Persistent Identifier

Assign a persistent identifier that serves as a **link for accessing the scientific software** and resolves to the current location where the software can be accessed. Scientific data centers, repositories, and archives often assign persistent identifiers to disseminated objects.

References
Open Source Initiative: https://opensource.org/licenses
R Licenses: https://www.r-project.org/Licenses/

Center for International Earth Science Information Network
EARTH INSTITUTE | COLUMBIA UNIVERSITY