December 6, 2023

Dr Peter Bajcsy
Project Lead, Software and Systems Division
Information Technology Laboratory
National Institute of Standards and Technology
100 Bureau Drive MS 2201
Gaithersburg, MD 20899

RE: Request for Information Regarding File Specification for Findable, Accessible, Interoperable, and Reusable (FAIR) Containerized Computational Software (FAIR-CCS)

Dear Dr Peter Bajcsy,

Thank you for the opportunity to participate in this RFI for FAIR Containerized Computational Software. NIST is an important leader in the US for setting standards that inform best practices for scientific research and innovation. AGU's response to the RFI is based on our experience working with Earth, space, and environmental science researchers as well as computer and information scientists and data and software experts in the work that we are doing in promoting FAIR and open software sharing in AGU journals and in the wider community.

Sincerely,

Kristina Vrouwenvelder
Program Manager, Open Science Leadership
American Geophysical Union

Shelley Stall
Vice President, Open Science Leadership
American Geophysical Union

(2) Important characteristics of sets of containerized computational software for reuse.

We would like to draw attention to two important characteristics of sets of containerized computational software that play a role in reuse: attribution and licensing.

Attribution:
In the process of creating and containerizing tools for complex data analysis, it is common to build off of or reuse previously-developed models and software tools; moreover, multiple software and code components may be included in the container. Proper attribution in software reuse provides a route to acknowledge the contributions of creators of previous software versions; creators of software that has been built upon; and the creators of the current software version and container. Proper attribution also enables provenance of software to be detailed and should include consideration for how versioning of software may change contributor lists. Standards for attribution should consider both how the human-readable attribution in the software manifest file will be created and how machine-operable attribution, particularly citation, will be handled. Consider including a fully-formatted citation for the software container and information about when and where to cite software container use in the manifest file. Separate attribution for software, data, and other digital objects should be standard.

If software tools and containers reuse and adapt a large number of other, previously-developed tools, it may become difficult to properly cite each component. This case has parallels to the case of complex data citations in the Earth, space, and environmental sciences. We are members of a Research Data Alliance working group (https://www.rd-alliance.org/groups/complex-citations-working-group) aiming to address the use case of citing a large number of existing objects (e.g. datasets, software, or physical samples) in a way that allows credit for individual objects to be properly assigned. These collections of existing objects, which may comprise tens to millions+ of elements, require a 'complex' citation that could encompass subsets of elements from multiple other collections. We have used the working term of 'reliquary' for such citations; a number of alternative terms (data collection, package, crate, etc.) are also in use. Finding a solution to enable these 'reliquary' citations will enhance the reproducibility of science relying on these collections of digital objects and enable researchers and funders to trace the impact of their work, whether dataset creation or software package creation. Our Research Data Alliance effort will coordinate work towards an international recommendation in 2024.

Licensing:
Similarly to the use case of attribution, licensing questions may arise in the process of containerizing sets of software tools, which may or may not all be licensed alike, particularly if commercial or proprietary code is involved. Negotiating any discrepancies in licensing should be incorporated into the containerization workflow and licensing notes should be included in the manifest file.