In support of an open code policy which is inclusive of commercial technologies to accelerate reproducibility of science.

Prepared for: Space Studies Board Evaluation Committee

National Academies of Sciences, Engineering, and Medicine

505 Fifth Street NW

Washington, DC, 20001

Authored By: Tripp Corbett, NASA Account Manager, Esri, ccorbett@esri.com, 305-434-0710

Co-Authors: Dr. Dawn Wright, Esri Chief Scientist

Marten Hogeweg, Esri Sciences Services Practice Lead

12 January 2018

380 New York Street
Redlands, California 92373-8100 USA
T 909 793 2853

Short Summary

Esri welcomes the opportunity to respond to NASA's request for comment on a future open code policy. We embrace open approaches to building information systems. We believe that a policy envisioned by NASA can create a healthy business ecosystem where commercial off-the-shelf and open-source solutions complement each other. As such, the open code policy should apply to the extensible code and workflows NASA creates, and it should not exclude the use of commercial off-the-shelf (COTS) tools as part of scientists' research. COTS technology is widely used and extensible. Scientists can use COTS technology to create their own algorithms and processes, which represent the real "science" that provides value when shared.

Who We Are

Esri is a unique commercial entity operating under five core tenets, one of which is to advance geospatial science. For nearly half a century, we have pushed the boundaries of geographic science and showed the world the possibilities of powerful geospatial technologies. We have done this in close collaboration with federal, academic, and commercial geographic science practitioners. Since science is so important to us, we provide the NASA Science Mission Directorate and other governmental, non-profit, and academic institutions with tools that can assist with their missions and research.

In the fall of 2015, Esri renewed our commitment to open technology (esri.com/open). We rededicated ourselves to supporting open standards, an open system through APIs, and open data. Our ArcGIS platform (explained here) now lives symbiotically with open code. For example, we maintain a GitHub site (esri.github.io) where we share hundreds of tools, interfaces, and other code with the community. We also host codesharing.arcgis.com where our users share their code with each other.

Esri is active in the open source and open data communities through organizations like the Open Geospatial Consortium and the R Foundation. We also participate in national, regional, and global communities of interest such as the United States Geoplatform, the European INSPIRE program, and the Group on Earth Observations. Within these communities, we discuss the use of standards for data, metadata, and web service interfaces, allowing for the creation of systems of systems that connect users to providers of geospatial and earth observation data. We also sponsor and participate at conferences such as FOSS4G, LocationTech, AGU, and EGU.

From a technology perspective, we own, maintain, and share many open-source projects (via GitHub)—including Geoportal Server and Terraformer—that are used globally across many industries. We are an active and significant participant on the open-source translator library for raster geospatial data formats, called GDAL. Further, we have developed ArcGIS to support PostgreSQL and PostGIS. We make certain our products work with open-source technologies, identify and submit bugs to the open-

source community, and answer questions to help open-source developers and users. In addition, our professional service division, when acting as a system integrator, may use open-source technology to build custom solutions for customers; similarly, our software development teams may use open source within the software itself.

Esri is submitting this response to show how the National Academy of Sciences (NAS) Committee can architect a policy that is not exclusive of such symbiosis, and that is inclusive of the open code community who build on top of COTS technology.

Executive Summary

As the National Academy of Sciences (NAS) is investigating and recommending best practices for the Science Mission Directorate (SMD) around an open code policy, we believe it is beneficial to contribute our support as a commercial entity.

We write to support the vibrant community of coders, developers, and scientific users that create and share code built on top of our COTS tools every day. We believe the open code policy should apply to the extensible code and workflows NASA creates, and not exclude the use of COTS tools as part of scientists' research. In fact, in working with scientists at agencies such as NASA, we've learned that deciding between open source and COTS is not an either/or question. Instead, scientific users should choose a hybrid model, using both open-source and closed-source tools based on their needs.

Esri already recommends and supports the use of open-source technology to enhance our users' experience with our COTS software. For example, our platform supports the Red Hat Enterprise Linux and SUSE Linux operating systems. Geodatabases can be created on the open-source RDBMS PostgreSQL. ArcGIS Server can work with open-source web servers, such as Apache Tomcat, to distribute GIS services, and more.

Any COTS technology that is widely used within SMD is used because it empowers scientists to create their own algorithms and processes, which represent the real "science" from which the community will realize value when shared. If the policy specifically mentions that it does not exclude using COTS technology, the policy will reduce confusion and avoid the unnecessary recreation of tools that already exist.

Question Responses

1. What positive and negative impacts would arise for you, your workplace, your NASA-funded research, science in general, education, commerce, society, and so on, if all future NASA-funded science code were required to be open source? For example, what maintenance and support issues might arise from open source policies that would not otherwise arise? What relevant experiences have you had with science codes owing to sharing or access constraints? How might negative impacts be mitigated?

If the open code policy is not clear about what areas it covers, we believe there will be confusion in the scientific community that results in duplicated work and loss of productivity. The open code policy should explicitly state that it covers the workflows, processes, and algorithms developed by users, and that it does not exclude the use of COTS tools to support research. If the policy does not do this, it will have a negative impact by slowing the pace of science: scientists will spend more time and precious project dollars discovering, integrating, and maintaining open-source software projects (often with less functionality) before they can even begin their science.

COTS software offers clear value to users of all kinds, including scientists. For example, many software companies today are creating software platforms. A platform provides an ecosystem of developers, partners, users, and other collaborators who contribute towards enriching the platform's capabilities. Because these platforms allow scientists to plug into existing, dynamic, and growing ecosystems, scientists can focus on innovating in their domains, rather than laboring over fundamental core functionality. Further, scientists can rest assured the platforms are being maintained to meet federal security standards such as FISMA, FedRAMP, TRUSTe, & Privacy Shield (Trust.ArcGIS).

Not everyone can create a platform technology that provides these ecosystems. Platform technologies require extensive R&D efforts, strong maintenance, and trusted extensibility mechanisms to withstand continuous updates. They are best served by companies with a cohesive technology and self-sustaining business model, resulting in COTS software. Scientists can leverage these COTS platforms for innovation.

With respect to support and maintenance, relying on extensible COTS technology allows scientists to outsource code maintenance. COTS software versions are frozen and archived in central locations with defined support timelines; they are easy to access and deploy when needed. On the other hand, finding and implementing all the open-source components used by another scientist would be difficult and time consuming, complicating future projects.

Some scientists may prefer to use open-source tools over COTS tools because existing COTS tools cannot accomplish all they need. But some scientists prefer the opposite, for similar reasons. Scientists should be able to choose the tools that work for them, whether open source or COTS. The Committee can make this clear in the policy.

A policy that permits the use of COTS tools will also have positive impacts to society. For example, many civic, nonprofit, governmental, academic, and commercial entities outside of the immediate research sphere rely on COTS technology to accomplish their missions. When these entities use open code built on top of COTS technology, scientists can accelerate the adoption of the science yielded from their research.

Along these lines, to avoid any uncertainty about adoption by the broader community, it is important for all open code to have clear usage constraints. This includes calling out whether code can be used for commercial purposes.

2. What would be the consequences, positive or negative, if NASA exercised any rights it may have to require that existing codes previously developed under NASA funding be made open source?

Esri has always maintained that code developed by our users belongs to them, and Esri has no claim to it. Our proprietary object libraries are protected, and we do reserve the rights to them (esri.com/legal). However, there is nothing precluding users who have built algorithms, processes, or workflows on top of these libraries from sharing those new items as open code. Consequently, from our perspective, there would not be any negative impact from requiring existing code developed by NASA to be made open source. The positive impact would potentially be the acceleration of science.

3. If a future policy is in place which would require all NASA-funded science codes to be made available under an open source license, what exceptions, if any, might be made to this policy? What principles might be applied in granting and then overseeing such exceptions, and what parallel measures could be taken to mitigate any detrimental effects an exception might have on code availability and re-use?

It would be important to include exclusions around proprietary/COTS tools used inside the code, rather than the code itself. Esri's GitHub site (hyperlinked earlier) provides a great example of how code itself can be open sourced under licenses like Apache 2.0, while the intellectual property in the underlying objects remain reserved. A similar policy for all extensible COTS technology will reduce bureaucratic obstacles for exception approval requests for each vendor's platform.

4. What lessons can be drawn from your experience with open data policies that might help inform future open source policies?

Here are two examples of federal agencies that work very well with the commercial sector in the field of open data:

- NOAA: NOAA generates a significant amount of data every day from satellites, radars, ships, weather models, and more. While publicly available, this data can be difficult to download and work with in high volumes. NOAA created the Big Data Project to work with industry partners to store data, foster innovation, and make NOAA data more accessible. NOAA now works with five commercial infrastructure-as-a-service (IaaS) providers to broaden access to NOAA's data resources. More information is available at www.noaa.gov/big-data-project.

- NTIS: The National Technical Information Service (NTIS) uses the Joint Venture Partnership (JVP) program to quickly connect leading technology partners to government agency customers using pre-established partnership agreements. Joint Venture partners are leaders in areas like data discovery, interoperability, analytics, security, and privacy. NTIS works with the nation's top private-sector companies, nonprofits, NGOs, and academic institutions. More information about NTIS and the JVP program is available at www.ntis.gov/about/.

5. What policy differences, if any, might be considered for NASA-funded science codes produced as part of a research grant versus those produced under other NASA funding mechanisms, such as contracts….

We do not have value-add input on this question to contribute.

6. What special (non-obvious) considerations might exist for codes with multiple funding sources or codes that incorporate proprietary libraries or other restricted information, such as International Traffic in Arms Regulations (ITAR)-regulated code?

The policy should consider ways to protect proprietary/COTS libraries. But the situation described in the question actually points to an important way that COTS software helps an open code policy. In this context, building open code on top of COTS technology helps absolve users from part of their ITAR responsibilities, since the commercial company ensures and restricts obtainment of its code by entities in restricted countries. While users are not completely absolved (their code could be reverse engineered to work with other tools), it does help prevent the advanced science created by NASA from getting passed to foreign entities deemed as enemies.

7. What non-policy approaches could NASA take to encourage open source licenses for NASA-funded codes (for example, bounties for opening closed codes or for creating new open codes…

We do not have value-add input on this question to contribute.

8. Over the long run, what would be the impact on the quality and reproducibility of research if NASA required all NASA-funded, peer-reviewed science papers to include an electronic compendium of (or pointers to) the source codes, inputs, and outputs that produced each scientific claim in the paper?

An open code policy inclusive of COTS technologies will help in the long run to ensure quality and reproducibility of research at NASA. It will encourage scientists and mission teams to make their data available in standardized formats, so data can be more easily used by others. In turn, this will encourage accessibility and peer review.

Final Author's Note: We are happy to provide additional information & answer any further questions the Committee may have in response to our submission. Thank you.