# Data Management Plan

ASU Research Computing projects and systems development as well as potential related experiments can produce a number of differential data artifacts. Collection, curation, management, protection and dissemination of these artifacts is essential to the goal of broad impact that is central to the proof of concept model adopted by research computing and the University at large. Therefore Research Computing has carefully formulated distinct data management policies according to artifact types, regulatory compliance and information management strategies. Both publicly funded engagements as well as private research initiatives result broadly in two data classifications:

1. Open-source software.
2. Research data sets (products) produced through experiment and scientific discovery to include publications as well.

**Data Stewardship**

Arizona State University holds an appointed position known as ‘Director of Data Management’ (DRDM) who serves as steward for all code and data resulting from named research initiatives, assuming responsibility in the primary role as overseer of the development and management of the research data lifecycle. Additionally the DRDM will serve to determine the appropriate classiﬁcation listed in this document for particular data artifacts in collaboration with the research project lead, typically the principal investigator (PI) or the holder of an independent grant administered by the University and/or the lead researcher for the grant project, this could include artifacts produced in a laboratory study or a clinical trial.

**Data-Handling Policies**

Research Computing adheres to the policy of releasing all source code resulting from publically funded proposals as open-source software under suitable nonrestrictive licenses, and will make use of repositories, e.g., Github, to support this practice, unless specifically directed otherwise by the PI, Provost or University CIO. In situations where software is authored as a product of a publically funded award and said software product will remain restricted the PI will furnish a validation letter to support the decision.

Additional data artifacts will become available to the research computing community of researchers as well as the general public to the greatest possible extent, as consistent with the privacy considerations that surround domain-specific research. The methodology for broad classification of datasets will align with the following as public (P), conditionally releasable (C), or conﬁdential (R). The associated policies for each resulting data artifact will be classified as deemed appropriate by the DRDM as aligned with policy (P), (C), or (R) to category (1) data as deemed appropriate by the associated data steward. We will apply policy (P) to category (2) data by default. Research Computing data-handling policies are defined as follows:

* Policy (P): Public Data. Public datasets will be those suitable for posting online, e.g., data derived from research experiments, or the outputs of experiments (e.g., data, source code) that themselves do not involve any privacy-sensitive data. Public data will be made available via a project website or a public cloud via Internet2 and publically accessible Data Transfer Nodes (DTN). Larger data sets that cannot be disseminated by either such means will be stored locally and instructions will be published for interested researchers and others to obtain access to the data.
* Policy (C): Conditional Release Data. Data artifacts produced through research will carry either temporary sharing limitations or individually requested moratoria on the release of artifacts or on event permanent ones. Such data will be retained for an equivalent duration of time as speciﬁed in policy (P). These data will not be made public, but stored locally with appropriate access-control mechanisms to restrict both external and internal access or in the University public or private cloud with protections aligned with the compliance controls assigned to the data. Should internal or external researchers or others submit appropriate requests for data access, once confirmed appropriate, the DRDM will determine a practicable minimal-release strategy, speciﬁcally exploring time-limited and sanitized data-sharing practices, as well as whether data should be released directly or through a query interface. Data will be published/ released as expeditiously as possible, consistent with resource and policy constraints.
* Policy (R): Conﬁdential / Restricted Data. As data steward, the DRDM may deem selected data sets or artifacts as temporarily or permanently unsuitable for release outside of Arizona State University. Potentially sensitive research-related data collected in the course of a proposal will be deemed conﬁdential in all cases until otherwise classified by the DRDM. Derived data such as coarse-grained aggregate statistics, publically available clinical data or datasets may be categorized as (C) or (P) immediately. By default, the data will be preserved according to Policy (C), with no access granted outside the institution of the data stewardship of the DRDM.

**Data Storage and Lifecycle**

Whenever feasible volumes of data produced through publically funded proposals will reside primarily within the existing data storage facilities of the University. All data in suitable standard formats will securely reside in University facilities and will include access controls and encryption as suitable for the handling of speciﬁc data artifacts.

**Vulnerability Disclosures**

Research projects stored in the University public cloud offering will be classified as (P) and will not explicitly encompass vulnerability assessments. It is very well possible vulnerabilities in the public cloud could be discovered and deemed by the University Chief Information Security Officer (SSO) as unforeseen and therefore classified as acceptable risk. Security vulnerabilities or inappropriate data disclosures in the course of our work are possible in the public domain and the University assumes no responsibility or liability for such vulnerabilities.

In the event of an inappropriate data disclosure or breach including access level elevation the DRDM will adhere to community-standard responsible disclosure practices. Speciﬁcally, the following steps will be taken in disclosing a vulnerability:

1. Stakeholder Identification: Primary stakeholders, entities developing or managing the affected systems or data, as well as secondary stakeholders, those potentially harmed by the vulnerability, e.g., users of the impacted system or subjects of the relevant data will be identified. Research computing in coordination with the University Technology Office (UTO) SSO and DRDM will work as advocates for secondary stake-holders throughout the disclosure process.
2. Vulnerability Disclosure: Primary stakeholders will be notified of the vulnerability provided with tangible evidence for conﬁrmation and assessment of scope. The University SSO and DRDM will provide this disclosure as expeditiously as possible.
3. Vulnerability Remediation: Primary stakeholders will be advised on technical remediation strategies, as appropriate and engaged in assessment to further determine scope of impact.
4. Public Disclosure Plan: In consonance with research community practice, a default public disclosure speciﬁcally providing identification of the vulnerability and scope of impact preventing injury to secondary stakeholders and interested parties. The DRDM will engage the primary stakeholders to determine the appropriate response. The SSO
5. Primary Stakeholder Review: Drafts of the root cause analysis (RCA) and of the public disclosure will be delivered to primary stakeholders, soliciting their feedback and working with them to ensure that details are correct and amending the disclosure as appropriate, taking into account any harm that may affect primary and secondary stakeholders as a result of disclosure.
6. Public Vulnerability Disclosure: The office of the SSO and DRDM will publish the disclosure, including both technical detail and explanations accessible to secondary stakeholders, as warranted by the vulnerability and scope of impact as assessed by the University CIO, SSO and DRDM.