
Open Science Champion

# ASAP Open Science Policy Handbook

# ASAP Open Science Policy Handbook

## Executive Summary

The Aligning Science Across Parkinson's (ASAP) Open Science Policy Handbook is a comprehensive reference guide. It provides fine-grained detail and unambiguous explanations of what ASAP expects from its grantees in the Collaborative Research Network (CRN) in terms of open science for their original research. This document constitutes the official ASAP Open Science Policy, itemizing all requirement specifications around policy compliance within our five (5) overarching requirements. Each item within an overarching requirement outlines a specific action that a grantee can take to comply with the Policy and lays the foundation for ASAP to have a consistent compliance monitoring and enforcement workflow. External-facing documentation summarizing the Policy can be found on the ASAP website and this checklist for authors.

The overarching goal of the ASAP Open Science Policy is for scientific findings that stem from ASAP funding to be verifiable, and for the associated research outputs to be reusable. Practically speaking, this means that someone looking at a figure panel from an ASAP-funded study should be able to identify and reuse the data that underlie that figure, the code used to analyze that data, and the lab protocol and key lab materials used to collect those data (Requirements 1 and 2). The other sections aim to ensure that this information is shared in a timely and open manner (Requirement 3), that ASAP is acknowledged (Requirement 4), and that a comprehensive record of ASAP-funded outputs exists and is shared within the ASAP Collaborative Research Network at an early stage (Requirement 5).

This document comes with an accompanying glossary. Everytime a keyword first appears in this document, we provide a link to the section of the glossary that outlines how this document is using that term. To improve clarity, this document italicizes words that refer to the level of our policy (e.g., *require*, *recommend*; see glossary).

The Policy aims to cover a broad range of research disciplines and we acknowledge that special circumstances may arise. Grantees may request exemptions by emailing openscience@parkinsonsroadmap.org with a clear justification.

# Table of Contents

# Overview of the Policy Requirements

The ASAP Open Science Policy is divided into five (5) main requirements:

1. **Share research outputs.** Data, code, and protocols generated as part of an ASAP-funded study must be deposited in a discipline-specific, community-recognized repository by the time of publication, with accompanying information to facilitate reuse of those outputs and a license that allows for reuse. Key lab materials generated as part of an ASAP-funded study must be registered by the time of publication.

2. **Identify research inputs.** Data, software, protocols, and key lab materials used in a study–but which were not generated as part of an ASAP-funded study–must be unambiguously identified in the study's publication.

3. **Ensure immediate open access.** Preprints must be posted no later than the date a manuscript is submitted to a journal for review. Preprints and publications must be immediately publicly available with a CC BY 4.0 or CC0 license and include an Availability Statement outlining where all research outputs (Requirement 1) and research inputs (Requirement 2) can be accessed.

4. **Acknowledge ASAP.** Manuscripts and other research outputs that were partially or fully funded by ASAP must acknowledge ASAP. Manuscripts must include an ORCID and a CRN author affiliation for CRN investigators.

5. **Share outputs with the ASAP network.** All ASAP-funded research outputs, including manuscripts, must be shared on the ASAP grantee virtual platform, no later than time of publication. Manuscript drafts must be sent to the ASAP Open Science Team no later than the time of posting a preprint.

# 1. Share Research Outputs

**Summary.** Data, code, and protocols generated as part of an ASAP-funded study must be deposited in a discipline-specific, community-recognized repository by the time of publication, with accompanying information to facilitate reuse of those outputs and a license that allows for reuse. Key lab materials generated as part of an ASAP-funded study must be registered by the time of publication.

## 1.1. Data

*1.1.1.* *Required.* **Raw data** (see glossary for definition). Data must be deposited in the format accepted by a discipline-specific, community-recognized repository, in their rawest reasonable form (e.g., the data have not been cleaned or preprocessed). We acknowledge that, depending on the type of data and the repository used, some level of preprocessing or curation may be required before sharing. Discipline-specific repositories are designed to hold specific data types (e.g., CRN Cloud, GEO, SRA); and are in contrast to generalist repositories that can hold all data types (e.g., Zenodo, Dryad). See here and here for recommended discipline-specific and generalist repositories.

Deposited data must come with a persistent identifier (e.g., a Digital Object Identifier–DOI) that is cited in the manuscript.

We *recommend* using a discipline-specific repository to deposit raw data, but acknowledge that discipline-specific repositories do not exist for all data types and that some repositories require substantial training to use effectively (e.g., DANDI for neurophysiology).

*Exception:* Sensitive data (see item 1.1.3).

*1.1.2.* *Required.* **Cleaned data** (see glossary). Cleaned data are the data entered into an analysis script and may represent the individual data points displayed in figures. Cleaned data used to produce all results, including figure panels, tables, graphs, and numbers presented in the results section of a manuscript must be deposited in a publicly accessible repository. This includes image quantification data. Cleaned data are often in tabular format and have often been preprocessed. It is possible, but unlikely, that the raw data and cleaned data will be the exact same.

Deposited data must come with a persistent identifier (e.g., a DOI) that is cited in the manuscript. Datasets that include only summary data (e.g., means, standard deviations–see glossary) can be shared, but are not

sufficient to meet this policy item. This item is in addition to item 1.1.1 (i.e., both raw *and* cleaned data must be shared).

*Exception:* Sensitive data (see item 1.1.3).

1.1.3. *Required.* **Sensitive data** (see glossary) must be deposited to the extent allowed by the associated research ethics approval (e.g., after anonymization). If the data can be openly shared, then we *require* these data be shared as described in items 1.1.1 and 1.1.2. If the data cannot be openly shared, but can be shared with restricted or controlled access (see glossary), then it must be shared as restricted or controlled data, and come with instructions for how to request access to the data.

Note, we *require* that grantees collecting data from human participants provide evidence that there has been satisfactory review and approval of the plan to collect and/or share such data from the appropriate ethics committee(s) (or evidence that no such approval is required).

1.1.4. *Required.* **iPSC quality control table.** If iPSCs are used, an iPSC quality control table must be included in the associated manuscript or deposited to a publicly accessible repository and cited in the manuscript. If the grantee did not run a particular quality control test that appears in the table, they must explicitly state that the test was not run. For more information, see the ASAP iPSC QC Reporting Template.

*1.1.5.* *Required.* **No new primary data statement.** If a study did not collect any new *primary data* (see glossary), text must be included in the Availability Statement (see glossary) stating that no new primary data were collected in the study. See item 3.3 for more information on Availability Statements and boilerplate text.

1.1.6. *Recommended.* **Data access for compliance review** (see glossary)**.** A member of the ASAP Open Science Team should be able to access the cleaned data associated with a manuscript when the manuscript is submitted for ASAP compliance review. A grantee can accomplish this by sharing the link to an open data deposit, sharing data via web links or email, or depositing the data to the ASAP-Zenodo workspace.

We also *recommend* that the ASAP Open Science Team are able to access the deposited raw data; although we acknowledge that it may not always be straightforward to provide access to these data at the time of manuscript submission to a journal.

*1.1.7.* *Nice-to-have.* **Optional data types.** We encourage the sharing of quality control (QC) data, confirmatory data, and intermediate data (see glossary).

*Exception:* iPSC quality control tables are *required* (see item 1.1.4)

*1.1.8.* *Required.* **Reuse license (data).** Deposited data must come with a CC0 or CC BY license. “ND”, “NC”, and “SA” modifiers on the CC BY license are *not permitted*.

*1.1.9.* *Required*. **Readme file (data).** Each data deposit must include a readme file entitled “README”. Readme files must include enough information so that someone who was not part of the study team can access the data, identify which data produce which results (e.g., figure panels), and understand the data well enough to reuse them. Guidance is available here. If a repository does not allow readme files, the equivalent information must be shared in the appropriate metadata fields.

*1.1.10.* *Recommended.* **Formatting of tabular data.** For data types that are stored in tabular format, the data should be deposited as a .csv file or another plaintext, non-proprietary equivalent (e.g., .txt, .tsv). All the data should be contained within text. Color coding and text formatting are not retained in .csv files and are not best practice for data management. We *recommend* each column represents a single variable and each row represents a single unit of observation. We also *recommend* including a data dictionary that explains each variable. Additional guidance is available here).

*1.1.11.* *Recommended.* **Folder and file structure.** For types of data where a standard community-accepted file structure, folder structure, and/or specification exist, we *recommend* grantees use that structure or specification. For example, the Brain Imaging Data Structure (BIDS) for neuroimaging data and the Neurodata Without Borders (NWB) structure for neurophysiology data.

*Exception*: ASAP Labs working with CatalystNeuro are *required* to use NWB format.

## 1.2. Code

*1.2.1.* *Required.* **Code** (see glossary). Code generated by grantees must be deposited in a publicly accessible repository and issued a persistent identifier. This includes all scripts, software, packages, libraries, macros, pipelines, algorithms, executables, batch files, and any other code the grantees wrote to manipulate data in any way, including but not limited to cleaning data, preprocessing data, analyzing data, and producing figures, tables, and results. This item applies for any grantee-generated code regardless of whether the code was written to be used primarily for a single study or for reuse by other researchers.

We *recommend* depositing code in a GitHub repository and then issuing a persistent identifier (e.g., a DOI) to the repository using Zenodo (instructions here) and citing that DOI in the manuscript. We *require* that manuscripts cite a persistent identifier to the code as it was run for that study. We also *recommend* citing the GitHub URL, so that a reader can access any updates that were made to the code. Citing a GitHub URL alone is *not sufficient* because the content is not preserved and can be removed.

*1.2.2.* *Recommended.* **Analyses without code.** When data cleaning, analysis, and/or visualization was conducted without coding scripts (e.g., they were performed using GraphPad Prism or another software with a graphical user interface that does not output a script), we *recommend* that grantees include a detailed and unambiguous explanation of how the cleaning, analysis, and/or visualization was performed. We *recommend* sharing files (e.g., Prism files), screenshots of images or videos, step-by-step written instructions of the actions performed, and/or any other relevant material. This could be done in a separate document with an itemized explanation for how each panel figure was produced as well as the steps taken to clean and preprocess the data.

*1.2.3.* *Required.* **No code statement.** If grantees did not generate any code for their study, they must mention so in the Availability Statement (see item 3.3). We *recommend* using the following boilerplate text: *"No code was generated for this study; all data cleaning, preprocessing, analysis, and visualization was performed using [insert program name(s)]"*. If even a few lines of code were entered into the command line or a program, we consider this to be code; it should be shared, and in that instance, this no code statement does not apply and therefore should not be used.

*1.2.4.* *Required.* **Reuse license (code).** Code must be deposited with a license that allows reuse. We *recommend* the MIT License (or another permissive license). Copyleft licenses are also *permitted* (e.g., GNU General Public License). Note, code uses a different licensing system than manuscripts and data (and that is why we recommend an MIT License rather than a CC BY license).

*1.2.5.* *Required*. **Readme file (code).** Each code deposit must include a readme file entitled "README". Readme files must include enough information that someone who was not part of the study team can access the code, identify which scripts execute which functions (e.g., clean data, produce figures), and understand the scripts well enough to reuse them. Guidance is available here.

*1.2.6.* *Required.* **Computational environment.** Code deposits must include a file that lists the session information, packages and their version numbers,

and any other information needed to recreate the computational environment in which the code was run. This item can be met in several ways. For example, by using packages like 'sessioninfo' and 'renv' in R, a jupyter notebook, or a dockerfile.

1.2.7. *Recommended.* **Code commenting.** We *recommend* that all code contains comments throughout so that someone reading the code can easily understand what the code is doing. This can also be achieved by including explanatory text in a jupyter notebook or markdown file.

1.2.8. *Nice-to-have.* **Reproducible container.** This allows the code to be run without needing to download materials or install software or packages. This can be achieved using a website like codeocean.com or a jupyter notebook.

## 1.3. Protocols

1.3.1. *Required.* **Unpublished protocols** (see glossary)**.** For all protocols that are mentioned in a manuscript but are not cited with a persistent identifier that links to recipe-style instructions (see glossary), grantees must create a recipe-style protocol, deposit the protocol to receive a persistent identifier (e.g., DOI) (we *recommend* using protocols.io), and cite the persistent identifier in the manuscript. See the glossary for a detailed explanation of published versus unpublished protocols. Additional guidance is available here.

*Exception:* We consider *in silico* methods to be code (these requirements are outlined in Section 1.2).

1.3.2. *Required.* **Reuse license (protocols).** Protocols must be deposited with a CC BY or CC0 license. "ND", "NC", and "SA" modifiers on the CC BY license are *not permitted*.

## 1.4. Lab Materials

1.4.1. *Required.* **Organism Research Resource Identifiers (RRIDs).** Grantees must register an RRID for the following newly-generated organisms: mouse, rat, zebrafish, and Drosophila using this form. Guidance for registering these novel organisms is located in the SciCrunch Resource Citation Guidelines.

1.4.2. *Required.* **Cell line RRIDs.** Grantees must register an RRID for newly-generated stable cell lines. This includes cell lines that have been engineered with CRISPR/Cas9, Zinc Finger Nucleases, and Transcription Activator-like Effector Nucleases (TALEN) gene-editing processes. We

*recommend* registering cell lines at *Cellosaurus* by emailing cellosaurus@sib.swiss

We do *not require* that transient cell lines are registered, as they are not stable cell lines. For example, cell lines engineered to (over)express a transgene through an expression vector (e.g., alpha-synuclein-GFP overexpression via a lentiviral expression procedure).

We also do *not require* that primary cell lines are registered due to their limited lifespan, and because they are also not considered stable cell lines. For example, primary neuronal cultures do not need to be registered with an RRID.

1.4.3. *Required.* **Plasmid RRIDs.** Grantees must register and deposit newly-generated plasmids at Addgene. Addgene plasmids are registered RRID items in the format of RRID:Addgene_####. See https://www.addgene.org/deposit for more information.

*Exception:* Viral tools developed at commercial entities using ASAP funds (e.g., Vector Builder) are often not able to be deposited at Addgene. In these instances we *recommend* sharing the complete vector sequence and catalog number.

*Exception:* When reporting the use of plasmids that were generated by a collaborator, we *recommend* grantees encourage the collaborator to deposit the plasmid at Addgene. If the plasmid is not deposited, we *recommend* grantees report the origin of replication, antibiotic resistance gene, and promoter region.

1.4.4. *Required.* **Antibody RRIDs.** Grantees must register an RRID for newly-generated antibodies at the Antibody Registry.

*Exception:* Antibodies developed at commercial entities using ASAP funds (e.g., AbCam through the ASAP/MJFF tools program) may be registered by the commercial entity directly rather than by the grantee.

*Exception* (for items 1.4.1 – 1.4.4). Grantees are not required to register key lab materials they did not generate.

1.4.5. *Recommended.* **Open Material Transfer Agreement (MTA)**. We *recommend* that the MTA associated with newly-generated key lab materials is open. A template Open MTA is available here.

# 2. Identify Research Inputs

**Summary.** Data, software, protocols, and key lab materials used in a study–but which were not generated as part of an ASAP-funded study–must be unambiguously identified in the study's publication.

## 2.1. Published Data

2.1.1. *Required.* **Persistent identifiers.** All data used in a study–but which were not generated as part of the study–must be cited using a persistent identifier (e.g. DOI or accession number).

*Exception.* If a persistent identifier does not exist for the dataset being used, please include a link to the repository webpage, release number (if applicable), and the date accessed.

2.1.2. *Required.* **Access instructions.** If the data are not openly available, the manuscript must include information on how a reader can access the data and who they would need to contact to request access. This information could go in the Additional Information column of a manuscript's Key Resource Table.

*2.1.3.* *Recommended.* **Portion of data used.** We *recommend* you state the portion of the data used in your study (e.g., which participants, which variables).

## 2.2. Published Software

*2.2.1.* *Required.* **Programs** (e.g., GraphPad Prism, ImageJ; see glossary)**.** All programs used in a study must be unambiguously identified. The manuscript must include (1) the program version number, if one exists, (2) the program name, and (3) a URL, DOI, or persistent identifier where information about the program can be found.

2.2.2. *Required.* **Discipline-specific packages** (see glossary). For all discipline-specific packages, the manuscript must include (1) the version number, if one exists, (2) the package name, and (3) a URL where information about the package can be found.

2.2.3. *Required*. **Generalist packages** (see glossary). We do *not require* that generalist packages are listed in the manuscript or KRT. We *require* they are listed alongside the deposited code, as outlined in item 1.2.6 (computational environment). They must include at least (1) the version number, and (2) the package name.

2.2.4. *Recommended.* **RRIDs (software).** We *recommend* including RRIDs for all programs and discipline-specific packages. RRIDs can be found by searching SciCrunch.

## 2.3. Published Protocols

2.3.1. *Required.* **Published protocols** (see glossary)**.** For all protocols mentioned in a manuscript where recipe-style instructions already exist (i.e., published protocols), the manuscript must include (i) a DOI to the protocol, (ii) an explanation of what steps of the protocol were followed (this could be a statement that all steps were followed), and (iii) whether any modifications were made, and if so, what those modifications are (this could be a statement that no modifications were made). If a grantee made substantial modifications to a protocol, we *recommend* they *fork* the protocol (if it already exists on protocols.io), or publish a new protocol.

## 2.4. Pre-existing Key Lab Materials

2.4.1. *Required.* **RRIDs (key lab materials).** Key lab materials (see glossary) including organisms, cell lines, plasmids, viruses, and antibodies must be identified with a source/vendor, catalog number, and RRID. If an RRID does not exist, the manuscript or KRT must explicitly state that an RRID does not exist.

2.4.2. *Required.* **Shared materials.** Grantees are *required* to include source information for key lab materials that were shared (e.g., material name, donor lab, and donor institution). We *recommend* that grantees discuss registering these materials with the donor lab, but recognize that this may be outside a grantee's control to do if the donor lab is not part of our ASAP network.

# 3. Ensure Immediate Open Access

**Summary.** Preprints must be posted no later than the date a manuscript is submitted to a journal for review. Preprints and publications must be immediately publicly available with a CC BY 4.0 or CC0 license and include an Availability Statement outlining where all research outputs (Requirement 1) and research inputs (Requirement 2) can be accessed.

## 3.1. Preprints

3.1.1. *Required.* **Preprints.** A preprint must be uploaded to a community-recognized preprint repository no later than the time a

manuscript is submitted to a journal. We *recommend* bioRxiv or medRxiv. ArXiv is also appropriate for preprints that are primarily computational.

*3.1.2.* *Required.* **Reuse license (preprints).** Preprints must come with a CC BY or CC0 license. “ND”, “NC”, and “SA” modifiers on the CC BY license are *not permitted*.

*3.1.3.* *Recommended.* **Publication DOI in preprint.** Upon having a manuscript accepted for publication and receiving a DOI, grantees should add the publication DOI to the preprint. BioRxiv and medRxiv generally make this link automatically. However, if the link is not made within 2-3 weeks of publication, we ask that you email biorxiv@cshl.edu (see bioRxiv FAQ: *My preprint has now been published in a journal. What happens next?*)

*3.1.4.* *Nice-to-have.* **Author-Accepted Preprint.** Upon having a manuscript accepted for publication, the author-accepted manuscript can be uploaded as a new version of the preprint. The updated version of the preprint can mention that the manuscript has been accepted for publication, the date of acceptance, the journal name, and publication DOI. For example, by including the text “*This manuscript version was accepted for publication at [journal] on [date]: [DOI].”*

## 3.2. Publications

*3.2.1.* *Required.* **Immediate open access.** Publications must be made immediately publicly accessible with no embargo period. Immediate open access can be achieved by publishing in an open access journal or by posting the author-accepted manuscript version to a publicly-accessible, community-recognized repository (e.g., PubMed Central, institutional repository), with no embargo period. Posting to a non-permanent location is *not sufficient* (e.g., a researcher’s lab website).

*3.2.2.* *Required.* **Reuse license (publications).** Publications must come with a CC BY or CC0 license. “ND”, “NC”, and “SA” modifiers on the CC BY license are not permitted.

3.2.3. *Nice-to-have.* **Preprint DOI in the publication.** In the publication, (i) indicate that a preprint exists, (ii) include the preprint DOI, and (iii) include the date when Version 1 of the preprint was posted. This information could go in the Publication Info section of a publication, the Publication Info section of a publication, the Availability Statement, or elsewhere depending on the section headers used by the journal.

## 3.3. Availability Statement

3.3.1. *Required.* **Availability statement.** Preprints and Publications must include an Availability Statement (see glossary) outlining where all the data, code, protocols, and key lab materials from Requirement 1 (research outputs) and Requirement 2 (research inputs) can be accessed. The Availability Statement may point to a Key Resource Table that contains this information.

Note, it is *not required* that all research inputs and outputs are made publicly available at the time of posting a preprint. If a resource is not yet publicly available at the time of posting a preprint, you may indicate in the Availability Statement, or associated Key Resource Table, where and when that resource will be made available.

Availability Statements that state that data, code, and/or protocols, are available upon request are *not sufficient.*

If the ASAP Open Science Team is sent a manuscript without an Availability Statement, they will return it to the grantee and ask for an updated version that includes an Availability Statement before conducting compliance review.

3.3.2. *Recommended.* **Boilerplate Availability Statement.** We *recommend* the following text be included in an Availability Statement, depending on which outputs were generated in the study:

*"The data, code, protocols, and key lab materials used and generated in this study are listed in a Key Resource Table alongside their persistent identifiers at [enter the Table number or Zenodo DOI]."*

*"No code was generated for this study; all data cleaning, preprocessing, analysis, and visualization was performed using [insert program name(s)]"*

*"No new primary data were collected in this study"*

*"An earlier version of this manuscript was posted to [preprint server] on [date] at [DOI]."*

3.3.3. *Required*. **Key Resource Table (KRT)**. We *require* that manuscripts include, or cite, a KRT that outlines all research outputs and research inputs. We *recommend* using the ASAP KRT template. Guidance is available here.

If the ASAP Open Science Team is sent a manuscript without a KRT, they will return it to the grantee and ask for an updated version that includes a KRT before conducting compliance review.

### 3.4. Manuscript Submission

3.4.1. *Recommended.* **Reuse license in manuscript submission.** We *recommend* grantees include the following language in their manuscript when submitting to a journal: *"For the purpose of open access, the author has applied a CC BY [replace with CC0, if appropriate] public copyright license to all Author Accepted Manuscripts arising from this submission."* This language is in line with CoalitionS and serves the purpose of unambiguously informing a publisher that you plan to publish with an open access license.

# 4. Acknowledge ASAP

**Summary.** Manuscripts and other research outputs that were partially or fully funded by ASAP must acknowledge ASAP. Manuscripts must include an ORCID and a CRN author affiliation for CRN investigators.

### 4.1. Funding Acknowledgement

*4.1.1.* *Required.* **Funding (manuscripts).** ASAP must be acknowledged as a source of funding for all preprints and publications. This applies regardless of whether the preprints and publications are posted or published during the funding period or after the funding period has ended. The grant number(s) must be included. We *recommend* the following language:

> "*This research was funded [in whole or in part] by Aligning Science Across Parkinson's [Grant number(s)] through the Michael J. Fox Foundation for Parkinson's Research (MJFF).*"

*4.1.2.* *Required.* **Funding (research outputs).** ASAP must be acknowledged as a source of funding for all deposited data, code, protocols, and key lab materials. This applies regardless of whether data, code, protocols, and key lab materials are deposited or registered during the funding period or after the funding period has ended. We *recommend* the grant number to be included in the research output acknowledgement.

Zenodo and protocols.io have specific fields where the funder (Aligning Science Across Parkinson's) and the associated grant number(s) can be listed. For GitHub repositories, we *recommend* including the following statement in the README file:

> "*This research was funded [in whole or in part] by Aligning Science Across Parkinson's [Grant number(s)] through the Michael J. Fox Foundation for Parkinson's Research (MJFF).*"

## 4.2. ASAP Affiliation

*4.2.1.1.* *Required.* **ASAP CRN affiliation.** For preprints and publications, the ASAP CRN must be listed as an affiliation for all authors who are part of the Collaborative Research Network (defined as the people listed on the ASAP CRN Hub) with the following wording: "Aligning Science Across Parkinson's (ASAP) Collaborative Research Network, Chevy Chase, MD, 20815." For data, code, protocols, and key lab materials deposits, listing the ASAP CRN as an affiliation is *recommended*.

## 4.3. ORCIDs

4.3.1. *Required.* **ORCIDs.** Preprints and publications must include ORCIDs (Open Researcher and Contributor IDentifiers) for all authors who are part of the Collaborative Research Network (defined as the people on the ASAP CRN Hub). We *recommend* that all authors include their ORCID, regardless of whether they are affiliated with ASAP.

# 5. Share Outputs with the ASAP Network

**Summary.** All ASAP-funded research outputs, including manuscripts, must be shared on the ASAP grantee virtual platform, no later than time of publication. Manuscript drafts must be sent to the ASAP Open Science Team no later than the time of posting a preprint.

## 5.1. Compliance Review

*5.1.1.* *Required.* **Compliance review.** Manuscripts must be sent to openscience@parkinsonsroadmap.org for compliance review (see glossary) within 5 business days of posting a preprint (which must occur no later than the time of submission to a journal).

*5.1.2.* *Required.* **Response to compliance review.** When a grantee receives a compliance review outlining actions that are required to meet the ASAP Open Science Policy, they must send an updated version of the manuscript to openscience@parkinsonsroadmap.org that addresses the issues raised. The updated version must be shared *before* the date of publication. We *recommend* responding to the compliance review within 60 calendar days of receipt, when feasible, and including a brief description explaining how the items were addressed.

### 5.2. ASAP Hub

*5.2.1.* *Required.* **Upload to the CRN Hub.** Identifiers for publications and research outputs (data, code, protocols, and key lab materials) must be uploaded to the CRN Hub no later than the date of publication. Identifiers for preprints must be uploaded to the Hub within 5 business days of posting a preprint.

*5.2.2.* *Recommended.* **Timely upload to the CRN Hub.** We *recommend* that all manuscripts and research outputs are uploaded to the Hub as soon as it is possible to share them (e.g., at the time of posting a preprint, or earlier). The Hub is intended to facilitate collaboration among ASAP researchers throughout the research process, rather than only when a manuscript is published.

# Additional Notes

- **Spirit of the Policy.** We will consider a requirement to be unmet if actions are taken that are not in line with the spirit of the ASAP Open Science Policy. For example, if a protocol uploaded to protocols.io consists of blocks of text that were copy-pasted directly from the manuscript text rather than written in recipe-style; if a data deposit contains only summary data; or if a readme file exists but contains almost no information, then the requirement is not considered to be met.

- **Grantee responsibility.** It is ultimately the grantee's responsibility to ensure that they meet the Policy. The ASAP Open Science Team performs compliance review to help grantees make their manuscripts compliant with the Policy. However, because compliance review draws from the manuscript text, this process can only identify content that is mentioned in the text. For example, if a grantee used code to analyze their data, but they do not mention code or the associated software in their manuscript text, then compliance review is unlikely to ask for code to be shared. In this situation, the ASAP Open Science Policy nonetheless requires that the code be shared.

- **Supplementary material.** We do not consider the Supplementary Material section of a publication or preprint to be registered with a persistent identifier. For example, data stored in supplementary material generally does not come with a DOI. Placing data, code, and protocols in the Supplementary Materials section does not constitute depositing to a repository and receiving a persistent identifier.

- **Research funded in part by ASAP.** The Policy applies to both research that is funded in whole by ASAP and research that is funded in part by ASAP. If a grantee feels that they will be unable to bring a partially ASAP-funded manuscript into compliance with the Policy, they must request an exemption (see next bullet point).

- **Exemptions.** Exemptions from these policy items may be granted on a case-by-case basis by emailing openscience@parkinsonsroadmap.org with a clear justification for why the Policy item cannot, or should not, be followed in a particular circumstance. We ask that requests for exemptions be emailed to us as early as reasonably possible. Exemptions requested at the time of publication or after publication are unlikely to be granted.

- **Policy Adherence.** We expect grantees to make reasonable efforts to comply with the above requirements and we will work cooperatively to ensure understanding and continually reinforce best practices. Compliance will be monitored and evaluated at regular intervals; failure to take corrective action(s) after repeated attempts by ASAP to rectify issues related to open science will be taken into account for future funding opportunities and could result in withholding of annual disbursement of funds and/or other support mechanisms.

# Appendix A. Policy for Non-Original Research Manuscripts.

The ASAP Open Science Policy, as outlined in the main section of this document, pertains to original research from grantees with the CRN. Appendix A outlines where the Policy differs for non-original research. We often refer to these articles as thought leadership manuscripts and they can include review articles, communication articles, letters, and other article types that do not include the collection of new data nor a formal analysis of existing data.

The text below outlines how the 5 overarching policy requirements from the ASAP Open Science Policy (as outlined in the main section of this document) apply to non-original research manuscripts. A checklist for authors is available here.

- **Requirement 1. Share research outputs.** *Not applicable***.** If data, code, protocols, or lab materials were generated, then we consider the manuscript to report on original research and deem it subject to the ASAP Open Science Policy as described in the main section of this document. This categorization holds even if the manuscript is being published in the review section of a journal.

- **Requirement 2. Identify research inputs.** *Recommended.* If any items from Requirement 2 apply to a non-original research manuscript, we *recommend* that the policy item be met.

- **Requirement 3. Ensure immediate open access.** Some items within Requirement 3 apply to non-original research manuscripts, as follows:

    - All items within 3.1, as well as item 3.2.3, are *nice-to-have*. Note that bioRxiv and medRxiv only accept original research. Thus, we suggest posting a preprint to an institutional repository or another service that accepts non-original research manuscripts.

    - Items 3.2.1 and 3.2.2 are *required* (i.e., immediate open access and reuse license for publications).

    - Items within 3.3 are *not required.*

    - Item 3.4.1 is *recommended.*

- **Requirement 4. Acknowledge ASAP.** All items are *required*, except for 4.1.2, which is *not applicable*.

- **Requirement 5. Share outputs with the ASAP network.** As for original research manuscripts, items 5.1.1, 5.1.2, and 5.2.1 are *required*, and item 5.2.2 is *recommended*.

# Appendix B. Glossary

## Policy Levels

The ASAP Open Science Policy uses six specific terms to describe the policy items: *Required, Recommended, Nice-to-have, Permitted, Not permitted,* and *Exception.* We use each of these terms with a specific definition.

### Required

**Definition.** If a *required* item is not met, we consider the related output (e.g., publication, data deposit) to be non-compliant with the Policy. Each *required* item comes with a workflow that the ASAP Open Science team employs to assess whether the item has been met.

**Related terms**. In relevant documents, we may also use the terms *must* or *ensure* to indicate that an item is required**.** For example, "to make this output compliant, a researcher *must*…" or "We ask that researchers *ensure* that…".

### Recommended

**Definition.** ASAP advises that grantees comply with *recommended* items. However, the ASAP Open Science Team will not systematically check that all recommended items are met. An output can be fully compliant, even if *recommended* items have not been met.

**Related terms**. We may also use the terms *should* or *encourage* to indicate that an item is recommended.

### Nice-to-have

**Definition.** Items that are *nice-to-have* can be thought of as items that ASAP encourages, but not as strongly as a *recommendation.* This term could be used in a few different contexts. For example, to encourage aspirational practices or for practices which ASAP has yet to establish a clearly operationalized recommendation or requirement.

### Permitted

**Definition.** Items that are *permitted* are considered in line with the ASAP Open Science Policy. However, we often use this term in a context where the policy recommends another option. For example, we *require* posting preprints, and *recommend* posting preprints to bioRxiv or medRxiv. However, we also *permit* posting preprints to other repositories, as long as they are freely accessible. All these options would be compliant with the Policy.

### Not Permitted

**Definition.** *Not permitted* can be thought of as the inverse of *required.* The absence of something that is *required* can also be considered *not permitted.* This term can be useful to provide additional specification about specific policy items. For example, we require that preprints come with a CC BY or CC0 license, and the “NC” modifier on the license is *not permitted*.

**Related terms.** In some text, the term *not sufficient* may be used.

### Exception

**Definition.** *Exceptions* indicate specific conditions where a *required* item does not need to be fully met for an output to be compliant with the Policy; or when an item that is *not required* must be addressed to be compliant with the Policy.

## Data Types

### Raw Data

Raw data, also known as primary data, are data in the initial form they were captured. They have not yet been preprocessed or cleaned. They are appropriate to deposit to a discipline-specific repository. For example, the ASAP CRN Cloud, GEO (Gene Expression Omnibus), SRA (Sequence Read Archive), DANDI (Distributed Archives for Neurophysiology Data Integration), or OpenNeuro (more information on data repositories is available here).

Some discipline-specific repositories may require raw data to undergo some degree of preprocessing or cleaning before being deposited. The ASAP Open Science Policy requires data to be deposited in the ‘rawest reasonable format’ in which it makes sense to share data. The rawest reasonable format will generally be defined as the format in which a discipline-specific repository requires data to be deposited.

### Cleaned Data

Cleaned data are the data entered into an analysis script and may represent the individual data points displayed in figures. Cleaned data originate from raw data. The amount of preprocessing and cleaning that occurs will vary greatly depending on the data type. Cleaned data are often in tabular format and can be uploaded to generalist data repositories like Zenodo. Some discipline-specific repositories may also allow cleaned data to be deposited alongside raw data.

Cleaned data are not that same as *summary data* (see below). For the purposes of the ASAP Open Science Policy, in the context of image quantification, we consider the image files to be the raw data, and the image quantification to be the cleaned data.

Some researchers in the life sciences also use the term *source data* or *processed data* to describe what we are calling cleaned data.

### Summary Data

We define *summary data* as the data that summarize more than one observation (e.g., means, standard deviation). These values may appear in figures and tables or as other numbers in the results section of a manuscript. Sharing these data can be useful, because the exact numbers are not always visible in figures. We consider *summary data* to be the output of a statistical analysis, and *cleaned data* to be the input of that statistical analysis.

### Sensitive Data

Sensitive data include data that contain personal information, such as protected health information, and any other data that is likely to negatively harm an individual or community if publicly released.

### Open Data

*Definition from data.bris:* Open data is the most permissive data access level and is used for data which has no particular sensitivities. Where research participants are involved, they have given consent to share anonymised data as 'Open' data; the risk of re-identification of participants is considered to be extremely low. Open data is released under a re-use license.

### Restricted Data

*Definition from data.bris:* Restricted data has some degree of sensitivity involved. For example, research participants have not given explicit consent to share data as Open data. However, the risk of re-identification of participants is considered low. Data is made available to approved bona fide researchers only, after their host institution has signed a Data Access Agreement.

### Controlled Data

*Definition from data.bris:* Controlled data has a large degree of sensitivity involved. For example, research participants have not given explicit consent to share as Open data and/or the risk of re-identification of participants is medium to high. Requests for Controlled data are referred to an appropriate Data Access Committee for approval before data can be shared with bona fide researchers, after their host institution has signed a Data Access Agreement.

Note, after depositing restricted or controlled data, it is not the researcher who collected the data who decides who gets access, it is the organization that manages the data that makes these decisions

### Quality Control (QC) Data / Confirmatory Data / Intermediate Data

The data collected to ensure the reliability, accuracy, and consistency of experimental results. This includes measurements and checks performed to validate sample integrity, instrument performance, and procedural accuracy. These data are used to identify and correct errors and to ensure that the data are trustworthy. Confirmatory data include many assays, such as chromatography, genotyping, and confirmatory sequencing. Intermediate data are produced part way through the process of converting raw data to cleaned data.

## Code and Software

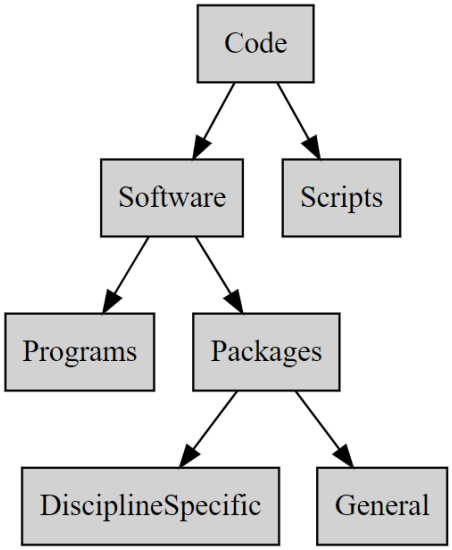


**Figure 1.** Schematic of the terminology the ASAP Open Science Policy uses regarding code. We use the term *code* to encompass all software, programs, packages, and scripts. We use *software* to indicate code that was written for reuse by others, and *scripts* to indicate code that was primarily written to clean and analyze the data from a particular study.

### Scripts

These include all code that a researcher wrote to clean, curate, preprocess, analyze, visualize, or manipulate data in any way. We opt for the term script as opposed to 'code' to describe this because it may reduce ambiguity with the more broad use of code.

### Software

We use this term to describe code that was written with the intention to be used more than once. For example, software may be maintained and designed to be used by others who were not involved in writing the software. Software includes both programs / commercial / consumer software (complete and executable) and packages (which extend the functionalities of programs). We discuss software in terms of (i) programs, (ii) discipline-specific packages, and (iii) generalist packages (defined below).

### Program

Software that is made by a company, open source community, or individual that is executable on its own. The software is designed to be used by others and may include a Graphical User Interface–GUI. For example: GraphPad Prism, ImageJ, Fiji, R, Python, Matlab, STAR (Spliced Transcripts Alignment), FastQC, SAMtools.

### Discipline-specific Package

Packages are collections of code that developers have written to add specific functionalities to programs like R, Python, and Matlab. If the package serves a specific scientific purpose, we call that a *discipline-specific package.* For example: MetaPhlAn 3.0, AccuSleep, MERINGUE, Slingshot.

### Generalist Package

These are packages that were not designed for a specific scientific discipline. For example, the package *tidyverse* was made to make coding more streamlined and *ggplot* was made to plot various types of data. For example: dplyr, ggplot, tidyverse, kable, numpy, pandas.

## Protocols

### Recipe-style Instructions

We define recipe-style instructions as clearly written documentation on how to perform a particular protocol, procedure, methodology, or technique. Recipe-style instructions will often include a list of all the lab materials required and a step-by-step explanation of what to do. We only consider documents where the writing is specifically directed toward a reader who is trying to repeat that protocol to be recipe-style instructions. Guidance is available here.

### Published Protocol

We consider a protocol to be published if an identifier (e.g., DOI, URL) exists that brings a reader to *recipe-style instructions.* Published protocols can include:

- Protocols on repositories like protocols.io.
- Protocol articles or methods articles (e.g., articles in journals like *STAR Protocols* and *Nature Methods*).
- Manufacturers' instructions.
- Outsourced procedures.

Note, if you cannot provide an identifier that a reader can follow and arrive at receipe-style instructions for any of the four types of protocols listed above, then we consider them *unpublished protocols.*

### Unpublished Protocol

We consider a protocol to be unpublished if there is not an identifier that can bring a reader to *recipe-style instructions*. Unpublished protocols can include:

- Protocols that your team developed, but have not yet deposited to a repository like protocols.io.
- Methods outlined in regular research articles.
- Protocols that were developed based on a published protocol, but contain substantial modifications (i.e., where it would be easier to write a new or "forked" protocol, rather than list every modification).
- Protocols with several steps, even when some of those steps are to follow a specific manufacturer's instructions.
- Outsourced procedures.

### Forked Protocol

On protocols.io, there is an option to "fork" a protocol. This allows a user to copy a published protocol, edit the text, and then publish a modified version of the protocol that will be linked to the original version of protocol. We encourage forking because it provides a clear link to the original protocol.

## Lab Materials

### Key Lab Materials

Key lab materials include organisms, cell lines, plasmids, viruses, and antibodies. ASAP's Open Science Policy requires that manuscripts report RRIDs for key lab materials. *Other* lab materials include kits, reagents, hardware, and primers.

## Additional Terms

### Availability Statement

We use the term Availability Statement broadly to encompass any and all sections of a manuscript that appear under a header that is specifically related to the availability of data, code, protocols, and/or lab materials. These section headers could be: Data Availability, Data and Code Availability, Materials Availability, Resource Availability, Open Science Practices, or contain other relevant terms.

For example, the journal *Cell* requires both a Data and Code Availability Statement and a Materials Availability Statement. We use the term *Availability Statement* to encompass both of these manuscript sections.

## ASAP-funded Work

ASAP funds are project-based, meaning they are given to each team based on the work they proposed. ASAP-funded work should relate to the team's proposed project and consists of any of the following:

- Projects that are listed in the ASAP-funded proposal.
- Methods or resource papers that enable the ASAP-funded proposal.
- Pivots stemming from the findings of the ASAP-funded proposal.
- Thought leadership pieces (reviews, communications, letters) on the topic that the ASAP-funded proposal aims to address.

## Compliance Review

This is the process where a member of the ASAP Open Science Team performs a systematic check to see whether a manuscript is compliant with the ASAP Open Science Policy. This process consists of an automatic compliance check performed by dataseer.ai and a manual check from an ASAP Open Science team member. ASAP then sends the grantee the output from the automated dataseer report and a written list of actions that would be necessary to bring the manuscript into compliance with the ASAP Open Science Policy.

## Manuscript

We use the term manuscript to refer to a document reporting a research study. This definition includes all of the following: a draft that is not yet publicly shared, a preprint, a publication.

## Original Research

We define original research as any piece of research that includes the collection of new data or a formal analysis of existing data. Original research is subject to all items in the ASAP Open Science Policy. Non-original research includes reviews, communications, letters, and thought leadership pieces that do not include the collection of new data nor a formal analysis of existing data.