

# MorPhiC Consortium - Software Sharing Policy

---

MorPhiC Consortium Publication, Data Release, and Software Sharing Policies distinguish two release tiers:

- Tier 1 concerns release to consortium members
- Tier 2 concerns release to the public

This policy will be reviewed and, if necessary, updated on a semi-annual basis.

---

This policy is based on the ENCODE software release policy (<https://www.encodeproject.org/about/data-use-policy/>).

*Rationale: To ensure reproducibility of analyses and to encourage reuse of software.*

Developers of significant new MorPhiC Consortium-related software will make their programs, including source code, freely available to consortium members and to the public. Examples include data processing pipelines and implementations of statistical, visualization, and modeling tools developed by MorPhiC Consortium funded groups to process or analyze data produced by the Data Production Centers (DPCs). Developers are strongly encouraged to integrate support for data format standards agreed upon by the MorPhiC Consortium Data Working Group into their software.

## What to Release

The MorPhiC Consortium requires the release of analysis pipelines used for major MorPhiC Consortium products. The consortium requires release of software tools and pipelines used to process any released data and used for major analyses and modeling in planned MorPhiC Consortium publications. When released together with specific datasets or modeling applications, the software should be released together with the parameter files that would allow fully reproducible runs as well as a minimum sample data set and expected outputs. Furthermore, a complete description of all external dependencies (such as libraries and their specific versions) of the software needs to be provided. To enhance reproducibility, software containers that include all dependencies are strongly recommended. Where possible, the files required to produce the container images, (Dockerfiles, executables) should be included. They should include exact versions of the dependencies and executables.

We also strongly encourage release of software likely to be useful to multiple groups either within the consortium or in the broader community.

## When to Release

The decision of when software should be released should balance the benefit to the consortium and the broader community, against the labor involved in software release and maintenance.

Software tools for data analysis and processing should be released as soon as they are sufficiently stable and no later than the time of the data release (according to the Tier system described at the beginning of this document).

Other software should be released with the release of the manuscript according to the publication policy and the Tier system.

**Software that has been released only within the consortium (Tier 1) cannot be used in any publications of consortium members, whether for projects funded by MorpHic Consortium or otherwise.** By requiring that any software used to generate hypotheses or results presented in a publication is publicly available (Tier 2), we will enable the broader community to repeat studies and reproduce findings.

## How to Release

Tier 2 (i.e. released to the public) software should be released by version controlled public repositories (e.g. Github, GitLab, BitBucket). These repositories should be linked via the MorpHic Consortium DRACC. Software should be well-documented and there should be a contact person for questions. Software development should continue through version-controlled deposition, and the software version used to generate each dataset must be documented using a consistent versioning scheme, e.g. Semantic Versioning (<http://www.semver.org>).

Whenever possible, MorPHiC Consortium funded software should be released under an open source license. It is, however, the sole responsibility of consortium members to follow any relevant rules of their institution with respect to software licensing.

In order to avoid multiple versions of the same software being maintained by different consortium members, consortium members who wish to contribute modifications of shared software tools (both Tier 1 and Tier 2) should do so using standard collaborative software development procedures such as Pull Requests on GitHub repositories. They should also clearly document what changes they made and the rationale for the modifications.

## Who Releases

Developers who release software publicly (Tier 2) should ensure that their software is included in the MorPHiC Consortium resource catalog.

## Other Considerations

**Accompanying publications:** In addition to the release of well-documented code, we strongly encourage developers to publish citable descriptions of their software. We recommend that authors describe their software in methodological papers so that they can receive credit for their work. These can be published in conventional journals, and/or disseminated pre-publication through pre-print servers (e.g. bioRxiv).

**Dissemination of more complex pipelines:** For most complex analyses, multiple software components are routinely combined to generate intermediate datasets. For reproducibility of these results, analysts should document all software components used, and the specific software versions utilized. We encourage (1) documenting these components; (2) providing scripts that reproduce key figures in MorPHiC Consortium publications; (3) establishing reusable, publicly accessible analysis pipelines (e.g. Biodepot, virtual machines, Docker, workflows compatible with systems provided by (commercial) cloud-based analysis providers); and (4) linking these through the MorPHiC website.

**Current and future support for released software:** Tier 2 Software that will be released for publication and to repositories (e.g., Github) should state the types and degree of support users can expect for them to download, run, and troubleshoot the available software as well as whether or not updates and “fixes” should be expected. In many cases, the expectation is that there is minimal support for users, ie, the

software is provided for reproducibility purposes, and not for outside use. The state of the software should be documented in a README file that is part of the software.

## **Contributors**

### **Outreach Working Group – Grantees**

*Chair:* Mazhar Adli, Paul Robson

Attendees: Ka Yee Yeung

### **Outreach Working Group - Program Staff**

Additional comments provided by ...

## **Revisions**

- Version 1.0 - ----Date----
  - Approved by Steering Committee during ..... meeting!