MorPhiC Consortium - Data Release and Use Policy

Version 1.1 – December 1, 2023

The MorPhiC Consortium Data Release and Use Policy will be reviewed and, if necessary, updated on a semi-annual basis.

Rationale: To catalyze research within and outside the MorPhiC Consortium and to provide a rich community resource.

This document describes the MorPhiC Consortium policy for releasing data, both internally and to the public. All data generated as part of a MorPhiC Consortium-funded project are subject to this policy.

This policy encourages the open sharing of data generated by the MorPhiC Consortium program with the use of a specific permissible license (CC BY 4.0 license). All data needs to be shared with the MorPhiC Consortium Data Resource and Analysis Coordinating Center (DRACC). All data (raw and processed) will be disseminated centrally through the MorPhiC DRACC. Dissemination through DRACC includes open access data as well as the distribution to external data resources for controlled access data (CAD) (e.g., databases like dbGaP). The controlled access data is defined by the consent permissions and data usage and modification restrictions. The data derived from CAD is considered open access.

Data Development Levels

Many groups in the MorPhiC Consortium are engaged in developing new technologies for investigating the functional roles of human genes in a multicellular system by generating functional null alleles. For data release, the policy differentiates between data generated with more established assays with agreed-upon standards and processing workflows, and data generated with emerging technologies, which require more internal vetting prior to public release.

Production Data is defined as an established data type which has an approved metadata model, a common protocol that has been approved by the MorPhiC Steering Committee, and a data processing workflow approved by the MorPhiC Data Working Group.

Technology Development Data is defined as an emerging data type for which a working protocol is available in PDF format, but no approved metadata model exists. As novel technologies and analytical workflows developed in the MorPhiC become more refined and standardized, these data types may be reclassified as production data.

Once a protocol has officially been approved for a technology development data type, a metadata model has been implemented, and a standard processing pipeline established, the data will become a production data type. Any previously submitted data for such data types will be made publicly available using the quality flags provided by the submitter.

Data Release Stages

Data is expected to be released in three stages: (1) *archival data* release consisting of data contributed by the Data Production Centers (DPCs), with base level of QC checking; (2) *secondary-processed data* generated by the standardized processing pipelines, where the data processing pipelines and data levels

will be designed jointly by the DRACC and the DWG (Data Working Group); and (3) *integrated data* release including analysis results such that the full uniformly processed MorPhiC dataset is queryable by e.g., gene (MorPhiC Data Portal).

Internal release will consist of archival and secondary-processed data. *Public release* will consist of archival, secondary-processed, and integrated data.

The data may be updated in subsequent data releases, according to guidance and feedback from the DWG. Release of some data may be restricted as a result of QC measures, consent restrictions, and other relevant considerations.

Data Release Schedule

All MorPhiC generated data needs to be shared with the DRACC immediately, or as soon as possible, after it has been generated and the QC performed.

Production Data

Production data and associated metadata will be shared *internally on a monthly basis* following completion of the submission to the DRACC, as the archival and secondary-process data.
Production data and associated metadata will be shared *with the public on a 6-months cycle basis*, as the archival, secondary-processed, and integrated data. Before the public release, the data needs to be signed-off by the data producer.

Technology Development Data

Technology development data will be shared *internally* on a monthly basis following completion of the submission to the DRACC, as archival release.

As a general rule, any MorPhiC Consortium data used in a manuscript submitted to a preprint server must be publicly available at that time.

Considerations for Use of MorPhiC Consortium Data

Use of public data by external users: External data users may download, analyze, and publish results based on the public open access MorPhiC Consortium data, under the terms of CC BY 4.0 license. Controlled access data is subject to consent permissions and data usage and modification restrictions.
Researchers using public, but yet to be published MorPhiC Consortium data must contact the specific data producer to discuss possible coordinated publication. Unpublished data are those that have never been described and referenced by a peer-reviewed publication. Use of MorPhiC Consortium data must be cited as described below.

Use of internally shared data by MorPhiC Consortium members: MorPhiC Consortium members have access to production and technology development data prior to public release. Use of these data in a publication prior to public release must be accomplished as a collaboration with the data producer. Compliance with this policy will be monitored and failure to abide by this policy will be reported to the MorPhiC Consortium Steering Committee and Program Directors.
Internal use and sharing of the identifiable data is defined by the MorPhiC Data Use Agreement.

Citation of MorPhiC Consortium data: Any data and data products released by the MorPhiC Consortium must be referenced using the official accession numbers and any relevant Data Release numbers (so that

downstream analyses are reproducible), provided by DRACC. Manuscripts using data released by the Consortium must also **cite (1) the** *latest Consortium publication* <mark>*(doi: MorPhiC white paper, to be added when published)*</mark> and **(2)** *all relevant publications and preprints describing MorPhiC Consortium data referenced in the manuscript*. The corresponding MorPhiC Consortium Data Production Centers or labs must be acknowledged.

Contributors

Policy Working Group
Co-Chairs: Mazhar Adli, Paul Robson

MorPhiC Grantees

Revisions

- Version 1.0 – February 1, 2023
  - draft based on '4D Nucleome Consortium' policy
- Version 1.1 – December 1, 2023
  - (Approved by the Steering Committee??)