

# External Data Sharing Policy (FINAL 2019-07-24)

[DRAFT: Provided to SC - 2019-06-30 for comment, FINAL 2019-07-24 for approval by SC 2019-08-05 - Done]

*Policy Working Group Subgroup: Phil Blood, Ajay Pillai, Nils Gehlenborg*

This policy governs data access and data release by the HuBMAP Consortium to the public. We begin by articulating the guiding principles that the HuBMAP Program will strive to adhere to. We then outline the public data release, data access and software release policies and finish with appropriate use and requested acknowledgment statements.

## Guiding Principles

The HuBMAP Consortium will strive to support;

- FAIR principles: Findable, Accessible, Interoperable, and Reusable: <https://www.nature.com/articles/sdata201618#bx2>
- Rapid data release: Data will be released by HuBMAP after a verification step which is understood to mean that the HuBMAP lab considers the experiment to have produced usable data. It is expected that data may continue to be collected by improving SOPs and other QA/QC metrics and may be replaced in future releases.
- No embargos: There will be no publication embargo placed on the non-HuBMAP community once HuBMAP data has been publicly released.
- Data will be released under a permissive license, such as CC-BY 4.0 (for open data)
- Open source software release (this document covers only the software that is used in processing data for release) using a permissive open-source license like MIT or BSD. The HuBMAP project will endeavor to not use closed-source (e.g., vendor) software, but that may not always be possible.
- Release both raw and processed data. The definition of raw data will reflect community input and needs. The specific processing level will be agreed upon by the HuBMAP data release and QC teams, who will also be collecting input from the broader research community.

HuBMAP, like many other NIH programs, will follow “rapid data release” policies to the community. We will publish data release schedules (with a minimum frequency of every six months after the first release) on our website for the community’s benefit. Working papers and/or white papers generated by the HuBMAP consortium will be used to inform the research community, and we will accept feedback on all aspects of the data release process. It is expected that HuBMAP data will be released after “data verification”, which is defined as assessing the usability of the data. That is, the data has sufficient quality to support its use in follow up experiments. The working/white papers will provide additional and assay specific details at the time of data release.

## Public Data Release

All public releases of HuBMAP data will be associated with working/white papers or publications that describe the choices made in the experimental design, data collection and processing workflows and will assess the potential impact on the research community. As a general guideline, these papers will describe how we implemented our guiding principles described above and will include the following aspects: (1) how we implement FAIRness; (2) choice of data formats; (3) definition of raw data for each assay; (4) description of the data processing; (5) minimum QA/QC criteria; (6) analytical software including parameters; etc.

The HuBMAP program will endeavor to correct errors identified in released data in subsequent releases using an appropriate versioning processes.

## Data Access

The HuBMAP Data Portal will provide access to both open and restricted access data and will be guided by the rules set forth by existing NIH GDS Policy and other applicable laws.

There may be both controlled and uncontrolled access data available through the Data Portal. Permission to access controlled data will be reviewed and granted by a designated NIH Data Access Committee.

Controlled access data available on the HuBMAP web portal will be handled according to the [NIH Security Best Practices](#) document. As of this writing we only expect controlled access to be relevant to the following datatypes: DNA and RNA sequencing raw data. Controlled access is not applicable to other data types that are part of HuBMAP consortium assays. It is also not applicable to higher level processed data. HuBMAP consent groups (consent codes) under which data and metadata will be released will include two kinds (1) “no restrictions for research use” or (2) General Research Use (GRU).

## Appropriate Use Statement

Users of HuBMAP open-data or processed data agree not to use the requested datasets, either alone or in concert with any other information, to identify or contact individual participants (or family members) from whom data and/or samples were collected.

This provision does not apply to research investigators operating with specific IRB approval, pursuant to US 45 CFR 46, to contact individuals within datasets or to obtain and use identifying information under an IRB approved research protocol. All investigators conducting “human subjects research” within the scope of US 45 CFR 46 must comply with the requirements contained therein.

# Acknowledgement

Investigators using HuBMAP data in publications or presentations are requested to include an acknowledgement of the HuBMAP Program. Suggested language for such an acknowledgment is: "The results <published or shown> here are in whole or part based upon data generated by the HuBMAP Program: <https://hubmapconsortium.org>."