BloodPAC Data Management and Data Sharing Plan

BloodPAC is committed to the open and timely dissemination of research outcomes and research data and followings the policies for data management and data sharing developed by the Open Commons Consortium (OCC).

What data will be shared? The BloodPAC Data Commons is designed to store all data types, including both primary data and derived data. In particular, the following data types are some of the data types supported: clinical/phenotype data, genomic and other omics data, imaging data, and biospecimen data.

The BloodPAC Data Commons is operated by a not-for-profit consortium. The BloodPAC Data Commons host data from consortium projects and contributed by consortium members as well as third party datasets selected to be of interest to the consortium and the research community that it serves.

FAIR Access. All data submitted to the BloodPAC Data Commons will be assigned globally unique IDs (GUIDs), with metadata associated to each GUID, and the data and metadata available through open APIs so that all data submitted to the BloodPAC Data Commons is findable, accessible, interoperable, and reusable (FAIR).

Who will have access to the data? The BloodPAC Data Commons supports data with different levels of sensitivity and corresponding levels of access controls.

- **Open access data** is available to anyone.

- **Registered access data** is available to anyone, but the user must register with a name, email, and optional organization. In addition, some consortia impose some additional requirements, such as the source of data must be acknowledged in any use or publications resulting from the data.

- **Controlled access data** is more sensitive and requires that the researcher or the researcher's organization sign data use agreements that impose requirements to protect the privacy and security of the data. More sensitive controlled access data may also require that the user fill out a Data Access Request (DAR) describing the intended use of the data. If data requests DAR, then each consortium operating an OCC commons also sets up a Data Access Committee (DAC) to review and approve DARs. DARs must be renewed regularly, usually once a year. The goal of the DAC is to provide data access to all qualified researchers.

- **Restricted access data** is so sensitive that the data is made available through an interactive portal that supports analyses that protect the sensitivity of the data, through data aggregates, or indirectly by sending an analysis workflow to the commons for approval and execution, with the results returned to the researcher (remote execution). In general, the BloodPAC Data Commons supports open access, registered access, and controlled access data, but in limited cases also supports restricted access data.

How will researchers locate and access the data? Researchers who use data the BloodPAC Data Commons will be required as a condition of access to acknowledge that the data was provided by the BloodPAC Data Commons in any publications and presentations that they make. The BloodPAC Data Com

- exposes APIs so that metadata about all hosted datasets are available through an open API;
- exposes APIs so that open access datasets may accessed by third party software services;
- exposes APIs to that controlled access datasets may be accessed by third party software services with an appropriate security token showing the user is authorized to access the data.

Submitting your data to BloodPAC. If you have a dataset that you would like to host and to distribute in the BloodPAC Data Commons, please submit a request to the commons of interest or to info@bloodpac.org. In the request, please specify whether the data will be open access, registered access, or controlled access. Datasets are reviewed by a data selection committee. In general, datasets are expected to be associated with a publication that describes the dataset or project that created the dataset. All datasets distributed by the BloodPAC Data Commons require that a Data Contributor's Agreement (DCA) be executed.

Public disclosures. Names and institutions of persons either given access or denied access to controlled access data, and the basis for such decisions, are generally available to the public.

Authorized workspaces. The BloodPAC Data Commons is designed to interoperate with trusted third party authorized workspaces.

Long term sustainability. BloodPAC managed data will be available in cloud storage for at least five years after the BloodPAC Data Commons project is over. If continued funding is not available to support the continued storage and access to BloodPAC Data Commons data after five years, data will be transferred to another data repository for longer term storage.

Appendix A – OCC Operating Principles

These operating principles are from the OCC website: https://www.occ-data.org/commons-principles. OCC Consortium activities, including the operations and management of OCC commons, are undertaken by Working Groups.

Open Data, Open Source Software, And Open Access Publications

In general, the Open Commons Consortium (OCC) and OCC Working Groups support open science by:

- Whenever possible, making the data associated with their projects, including both open access and controlled access data, publicly available using a data commons, data repository or similar resource accepted by the research community that the Working

Groups are supporting. In general, data should be released to the public as soon as possible, but with a data embargo lasting no later than six months after the project is concluded.

- Some data is sensitive, and, even though publicly available, is only available to the public after appropriate data use agreements are executed, with the understanding that the data will only be used in IT environments with the required security and privacy safeguard to protect controlled access data.
- When data is exceptionally sensitive, then instead of releasing the data to the public, the data may only be available via specialized environments, gateways that support federated queries, analysis or workflows, or via other mechanisms designed to protect especially sensitive data in accordance to applicable federal and state laws including contractual obligations.
- In general, OCC Working Groups will place preprints of their research paper in archives or repositories for research articles and publish their research in open access journals and publications.
- In general, OCC Working Groups will license their software using open source software licenses and distribute their open source software using applications designed for this purpose.

Open Access Data, Controlled Access Data And Fair Data

It is important to distinguish between closed data that is proprietary and not generally available, has fees associated with the use of the data, or has other restrictions of this type.

In contrast, data may be open and generally available to the public, or in some cases, generally available to the research community.

Open data may be *open access* and available to anyone or *controlled access* and only available to those who have signed the necessary data use agreements and have systems that have the necessary security and compliance policies, procedures and controls.

Data in the Commons is open and findable, accessible, interoperability and reusable (FAIR).

Operating Principles

OCC Commons are structured with: i) a core ("Core") and ii) an ecosystem of other systems, data sources and resources, computational resources, software services, applications, and notebooks.

- The "Core" is open source, standards based, uses open APIs, and the data that it contains is available without restriction (open data), except for those restrictions required to protect data derived from human subjects, or the privacy restrictions required by consumer apps that collect the data.
- The applications, systems, data and other resources outside the Core can be open or proprietary, free or fee-based, but are all required to satisfy not only all required legal and regulatory requirements, but also the ethical guidelines developed by the ELSI Working Group.

- It is important to note that term *open data* here means that the data is available via open APIs and without fees. As mentioned above, open data may be *open access* and available to anyone or *controlled access* and only available to those who have signed the necessary data use agreements and have systems that have the necessary security and compliance policies, procedures and controls. **It is important to emphasize that all controlled access data will be made available only to those that have signed the required data use agreements and are authorized to work with it.**
- Any philanthropic funds provided to support the Commons are expected to impose the restriction that the work they support is developed as open source and supports open data, but the components built in this way by Consortium Members may be part of a larger proprietary system or fee-based system or service.
- The core platform will use the security and compliance policies, procedures, and controls from NIST SP 800-53r4 at the Moderate level.