

# 1. What is Databrary?

Databrary<sup>1</sup> is an open data library designed to facilitate data sharing and greater transparency within labs, among collaborators, with members of the larger behavioral/developmental science community, and with the general public. Investigators are encouraged to develop and share their data through Databrary, and to make use of the tools we offer for analyzing and visualizing their contributions.

Databrary contains *De-Identified Data*, *Non-Identifiable Data*, and *Identifiable Data* – video/audio *Recordings* and photographic images and other sensitive materials, such as *Self-Reported Health-Related Information*. The tools stored in Databrary consist of software that aids in the extraction, visualization, and analysis of time-series-based research data. These tools may be open source, freeware, or commercially licensed, and may be supplemented by sample data sets, web links to data sets shared elsewhere within Databrary, or documentation. Every data set and tool will have its own unique, permanent, uniform resource locator (URL) – web address – so that researchers may both take credit for their own Databrary contributions and properly and consistently cite the Databrary resources they use from others.

The Databrary ecosystem also includes a database of researchers, called *Principal Investigators* or *Affiliate Investigators*, or collectively *Investigators*. *Investigators* have authorization to use and contribute data. The database keeps track of the identities of each *Investigator* who contributes to, browses data within, or withdraws data or tools from Databrary.

This document sets out Databrary’s Standard Operating Procedures (SOPs). Section 2 describes the challenges that must be met in creating a data sharing system for sensitive, identifiable data. Section 3 describes Databrary operating procedures. Section 4 describes other documents involved in Databrary operations. Definitions of key terms, italicized throughout the text, may be found in a separate document<sup>2</sup>. For the latest versions of Databrary policies, see the website<sup>3</sup>.

## 2. The Challenges of Databrary

Implementing the Databrary ecosystem poses several challenges. The most crucial challenge concerns research ethics. Databrary must keep the promise of confidentiality given to research participants who provide identifiable data while upholding the principle of informed consent.

Any identification of a research participant or unauthorized disclosure of confidential information violates the *Promise of Confidentiality* given to the research participant. This requirement poses a challenge. Many of the sources targeted for sharing in Databrary contain *Personally Identifying Information* about research participants and other individuals involved in a study. Images and/or voices may be visible or heard. Names may be spoken out loud. Images and voices of visitors or other people who are not research participants may be recorded. If a study takes place in a home or other setting, aspects of that home or setting may be recorded. Thus, individuals or locations could be identified from *Recordings* directly or identities could be determined from *Deductive Disclosure*. The *Depicted Individuals* in *Recordings* may include research participants and other private persons who are not the targets of a research study.

It is technically possible to try to turn *Recordings* into *De-identified Data* through blurring, distortion and other means of modification, but doing so diminishes scientific value. Therefore, Databrary attempts to do everything possible to ensure that *Data*, especially *Recordings*, are used solely for research purposes and not for the investigation or identification of specific individuals. Databrary and its users must do everything practicable to minimize the chances that individuals can be identified from information shared within the library. Databrary staff and its *Investigators* must understand, certify competence in, and abide by the highest standards of ethics concerning research with human participants. This means that no *Data* can be deposited, stored, or shared within the Databrary ecosystem that contain *Personally Identifying Information* with two exceptions: 1. *Data* may contain a research participant’s date of birth since this is critical for answering scientific questions; 2. *Recordings* may contain faces, voices, names, and the physical location of the recording.

At the same time, the principle of informed consent requires that *Depicted Individuals* must give permission for their *Recordings* to be shared with others in Databrary. Similarly, individuals who provide sensitive information, such as *Self-Reported Health-Related Information* should be given the opportunity to give or refuse permission for this information to be shared. Depicted individuals and research participants must decide what level of sharing permission

---

<sup>1</sup><http://databrary.org>

<sup>2</sup><http://databrary.org/user-guide/policies/definitions.html>

<sup>3</sup><http://databrary.org/user-guide/policies.html>

they wish to give: no permission at all because they do not wish to share *Recordings* or *Self-Reported Health-Related Information*, or sharing permission with the assurance that shared data will be treated with care by other researchers. Only *Recordings* for which *Permission to Share* has been given by all depicted individuals may be shared openly in the Databrary. Similarly, only *Self-Reported Health-Related Information* for which *Permission to Share* has been given may be shared in Databrary. In turn, Databrary must ensure the *Level of Access* that depicted individuals have granted. This means that information about sharing permissions must be permanently linked with individual *Data* files in Databrary. It also means that researchers who access files in Databrary must agree to follow the wishes of individuals who grant permission for their *Recordings* or other sensitive data to be shared.

## 3. How Databrary Works

This section describes in detail how Databrary works, from accessing data, *Metadata*, and tools stored in the library to the contribution of new information.

### 3.1 Accessing data sets or tools from Databrary

Some information may be easily shared with minimal restriction so long as *Personally Identifying Information* has been removed and there are no restrictions on the how long the information may be stored. Sharing *Recordings* or other data with *Personally Identifying Information* or sensitive information such as *Self-Reported Health-Related Information* must be shared only among researchers who understand, embrace, and formally agree to abide by principles of research ethics that include maintaining participant confidentiality. Thus, Databrary must monitor and control who has access to data sets and what level of access they are granted.

#### 3.1.1 Levels of access permitted to data or software

There are two primary levels of access: (1) public/unrestricted, and (2) restricted, but open to authorized Databrary *Investigators*. *Data Contributors* may choose whether to allow all authorized *Investigators* on Databrary to have access (Open Sharing) or only a selected subset of individuals (Collaborative Sharing).

Databrary will also allow the contributor to grant public unrestricted access to archival materials, materials that do not contain sensitive information (abstracts, researcher information, protocols, visual displays, coding manuals, etc.), excerpts, and links to data analysis/software tools.

Identifiable or sensitive data — *Recordings*, coded spreadsheets linked to *Recordings*, *Self-Reported Health-Related Information* — and *Metadata* will be assigned to the restricted, but open access level in almost all cases. These materials may be shared if the *Depicted Individuals* or research participants have given permission. See section 3.3.2 below for details about the different levels of sharing permission permitted.

#### 3.1.2 Means of access, user accounts and privileges

There are three means of access to Databrary's assets: browsing, downloading, and contributing. Individuals who wish only to browse Databrary's public/unrestricted assets may do so without authorization from Databrary, however for security reasons the system will log access by internet protocol (IP) addresses (see 3.1.7. Access Logging/Usage Metrics). Unauthorized/unregistered individuals may not comment in Databrary's comment sections, contribute *Data*, or carry out other functions reserved for *Investigators*. Those who wish to browse Databrary's restricted, but open assets must apply for Databrary Principal Investigator or Affiliate Investigator status.

#### 3.1.3 Who May Apply for Databrary Investigator or Affiliate Investigator Status

*Databrary Principal Investigators* must be eligible for *Principal Investigator (PI)* status at their home *Institutions*. Graduate and undergraduate students, postdoctoral researchers, research scientists, and research staff may be granted Databrary *Affiliate Investigator* status if they are carrying out research under the supervision of a *Databrary Principal Investigator*. *Databrary Principal Investigators* who supervise *Affiliate Investigators* must (1) grant permission to researchers under their supervision to view data on their behalf, and (2) agree to take responsibility for the data use of the sponsored members under their supervision, and (3) take responsibility for the security of the data (where data are stored, etc.) by following best practices in data security<sup>4</sup>.

---

<sup>4</sup><http://databrary.org/user-guide/policies/best-practices.html>

### 3.1.4 Timing of *Databrary Principal Investigator* or *Affiliate Investigator* Application for Researchers Who Want to Contribute Data

A researcher may apply for *Databrary Principal Investigator* status before formal permission has been granted by an *IRB* for collecting or contributing a given data set, but no contributions should be shared with Databrary until *Investigators* have received whatever approvals or waivers are required by the *IRB* for their *Institution*.

### 3.1.5 Application Process

Application for *Investigator* status may be made via a web form. The following information will be sought:

*Required:*

- Name, email, institutional affiliation.
- An electronically signed copy of the Databrary Investigator Agreement<sup>5</sup>.
- Certification of PI status eligibility from *IRB*, or for Databrary Affiliate Investigator applicants, their research supervisor

The agreement obligates applicants to follow principles of ethical research with human participants, to seek *Permission to Share Recordings* or *Self-Reported Health-Related Information* from participants, to convey those permissions to Databrary, to keep their research supervision offices informed about their activities, to follow Databrary's Best Practices in Data Security<sup>6</sup>, to follow the Databrary Bill of Rights<sup>7</sup>, and other principles.

In some cases Databrary may request:

- Research ethics with human participants training certification (e.g., Collaborative Institutional Training Initiative, CITI<sup>8</sup>), and expiration date of current authorization.
- Information about *IRB* that governs the applicant's scientific activity, if applicable, including name(s) of representative(s) of the institution, surface address, telephone number, email.

*Optional, but recommended:*

- Title, surface address, phone
- Applicant web site, ORCID<sup>9</sup>, Research Gate<sup>10</sup> profile, social media accounts, ERA Commons ID.

### 3.1.6 Application review

Once an application has been completed and submitted, applicants will receive an automated email confirmation that an application has been received by Databrary.

The welcome/notification email will include reminders about Databrary policies regarding data use and citation.

For *Databrary Principal Investigators*, Databrary will determine the appropriate contact representing the institution, e.g., at a sponsored programs office. This representative will be responsible for authorizing the applicant, subject to:

- Applicant identity information that is valid and correct.
- Applicant certification in human subjects ethics training from an *IRB*.
- Applicant has *PI eligibility* or its equivalent at the *Institution*.
- Applicant has not violated Databrary use standards in the past.

For individuals seeking *Affiliate Investigator* status, the supervising/sponsoring *Databrary Principal Investigator* will be notified via email that an application has been filed. The supervising/sponsoring *Databrary Principal Investigator* will log on to Databrary, approve (or not) the *Affiliate Investigator* application, and provide information about the level of access to be provided to this individual.

If an application requires revisions, project staff will communicate the nature of those revisions via email. All email correspondence will be logged. Once an application is approved, researchers and the institutional representative will be notified by email.

---

<sup>5</sup><http://databrary.org/user-guide/policies/investigator-agreement.html>

<sup>6</sup><http://databrary.org/user-guide/policies/best-practices.html>

<sup>7</sup><http://databrary.org/user-guide/policies/bill-of-rights.html>

<sup>8</sup><https://www.citiprogram.org>

<sup>9</sup><http://orcid.org>

<sup>10</sup><http://www.researchgate.net>

The Databrary Ethics Committee will audit the procedures annually. The Databrary Ethics Committee will consist of staff, Advisory Board members, and the research community at-large.

### 3.1.7 Access logging/usage metrics

Databrary will log access to the system by Investigator ID and by IP address. Most access logging information will be used by Databrary for internal security, user metrics, and management functions. Some summary information about files browsed, downloaded, shared, or modified may be announced via Databrary Twitter, Facebook, RSS, and other web/social media feeds. For example, Databrary may identify and publicize which data sets are most often browsed, downloaded, and cited. Databrary *Investigators* will have access to information about Databrary activity of *Affiliate Investigator's* through an administrative control panel.

### 3.1.8 Browsing and downloading

Any member of the public may browse data or download tools with unrestricted/public level of access. Individuals with *Investigator* status may also browse or download assets with restricted, but open access. When *Investigators* request data sets to download, the Databrary system will remind them to cite data sets and associated papers, to obtain appropriate approvals for conducting research, to follow best practices for data security<sup>11</sup>, and to remove and destroy data that are no longer needed. Databrary will also send periodic email reminders to researchers who download data reminding them to follow these practices.

Many Databrary uses will involve browsing or downloading data sets without conducting research on those data (e.g., finding excerpts for teaching, searching for exemplars to facilitate study design, pre-research activities to determine whether formal analyses are feasible or promising). If *Investigators* intend to conduct research with data sets, they may be required to obtain permission from an *IRB*. Decisions about whether *Investigators* may conduct research on *Data* or the type of research proposed are beyond Databrary's control. Thus, whenever permitted by the *IRB* that supervises their research, *Investigators* may conduct research on data sets derived from Databrary. It is the explicit responsibility *Investigators* to obtain whatever approvals or waivers are required by their *Institutions* and to follow local, state, and national laws. This is clearly stated in the Databrary Investigator Agreement<sup>12</sup>. While not required by Databrary, *Investigators* are encouraged to report *IRB* protocol names, identification numbers, and expiration dates.

### 3.1.9 Account renewal

*Investigator* accounts must be renewed regularly. Databrary will notify users on the website 1 month prior to when their accounts are due to expire and ask *Investigators* to update their account information. To renew their accounts, *Investigators* will also electronically sign the Databrary Investigator Agreement and again have this agreement co-signed by their institutional representative or sponsoring investigator. If a Databrary *Investigator* fails to renew his or her account by the renewal deadline, they will no longer be able to access Databrary shared data. Data contributed by *Investigators* who do not renew their accounts will remain on Databrary.

## 3.2. Contributions

Researchers who wish become *Data Contributors* to Databrary must do some planning in advance. They must:

- Determine the source(s) of the data to be deposited;
- For *Recordings* or *Self-Reported Health-Related Information*, determine whether *Depicted Individuals* or research participants must give permission for their identifiable or sensitive information to be shared; in most cases, the answer is yes.
- Seek and be granted, as appropriate, permission to collect and share *Data* from the *IRB* that governs research at the *Investigator* or clinician's home *Institution*;
- Secure *Permission to Share* from *Depicted Individuals* or research participants using the Databrary *Sharing Release Templates* or equivalent language that has been approved by the *Investigator's IRB* having equivalent protections as the Databrary language.
- Apply for and be granted Databrary *Investigator* status;

---

<sup>11</sup><http://databrary.org/user-guide/policies/best-practices.html>

<sup>12</sup><http://databrary.org/user-guide/policies/investigator-agreement.html>

- Prepare *Recordings* for sharing with Databrary, including tagging files with the *Level of Permission* granted by *Depicted Individuals*. Remove *Personally Identifying Information*, except date of birth, from other *Data* and *Metadata*.

### 3.2.1 Sources

There are two main sources of *Recordings*: (1) *Recordings* collected by researchers or clinicians in laboratory or clinical contexts, and (2) *Recordings* collected from incidental sources not intended for research or clinical purposes. The latter may include home movies or videos, *Recordings* from security cameras or microphones, cameras or other *Recordings* made in public places, and *Recordings* shared with the public via web portals such as YouTube<sup>13</sup> or Vimeo<sup>14</sup>. Laws governing whether *Recordings* made in public places may be used for research purposes vary. Researchers who wish to deposit *Recordings* from these or other non-laboratory/non-clinical sources should consult with their *IRB* for specific *Permission to Share Recordings* with Databrary.

There are two primary types of studies involving *Recordings* and other data streams: (1) new *Recordings* with associated *Metadata* (coding spreadsheets, analyses, manuscripts, etc.); and (2) reused *Recordings* (complete *Recordings*, subsets, or files culled from different original sources) with new associated metadata (e.g., new codes). The same *Investigator* or research team may reuse original data deposited in Databrary. Databrary will implement a version control/date system to record modifications to, reanalyses of, and extensions to *Data*. Versioning does not change the status of study types. Versions and reused data sets will retain links with the original and subsequent studies.

### 3.2.2 Seeking permission To share data

Researchers who wish to share data with Databrary should seek permission to do so from their *IRB*. Individuals whose scientific activities are not governed by an *IRB* may choose to have their protocol reviewed by another entity with similar authority. In preparing applications to the entity, researchers are strongly encouraged to refer to and use the document(s) provided by Databrary (see Section 4).

Links to all of the relevant Databrary documents concerning permissions and access may be found on the Databrary website<sup>15</sup>.

## 3.3 Securing permission from participants to share data

Under most circumstances, the collection of *Recordings* requires that individual participants, and for minors their parents or legal guardians, give written informed consent to participate in some specific research activity. Databrary will require that *Investigators* demonstrate that they have secured explicit written permission (or for minors assent and parent/guardian permission) to share *Recordings* with Databrary from each individual depicted in a recording — not just research participants — or that the *IRB* has waived the requirement for explicit written permission. Similarly, *Investigators* who wish to share *Self-Reported Health-Related Information* must also gain permission or assent to share these data with Databrary.

### 3.3.1 How will *Permission to Share* data be secured?

Databrary has published a Sharing Release Template<sup>16</sup>. This template should be submitted with *IRB* applications for research, and should be used by *Investigators* when they seek *Permission to Share* from research participants or depicted individuals.

### 3.3.2 Levels of permissions to share recordings

*Depicted Individuals* (or their parent or legal guardians) or research participants (or their parent or legal guardians) may grant different levels of *Permission to Share*. Permission/assent to share *Recordings* should be sought separately from *Permission to Share Self-Reported Health-Related Information*. The sharing levels for each are as follows:

*Sharing only within a research group (Level 0)*

Databrary may store *Data*, *Recordings*, or *Self-Reported Health-Related Information* but will not permit access by *Investigators* unless the *Data Contributor* gives permission. *Data Contributors* may choose to permit access to their

<sup>13</sup><http://www.youtube.com>

<sup>14</sup><http://vimeo.com>

<sup>15</sup><http://databrary.org/user-guide/policies.html>

<sup>16</sup><http://databrary.org/user-guide/policies/release-template.html>

*Data* to other *Investigators* as permitted by their *IRB*. *Data Contributors* assume full responsibility for identifying who is allowed access to the *Data*, for controlling that access using Databrary’s tools, and for any actions others with whom the *Data Contributor* has shared take with the *Data*.

#### *Sharing with all Databrary Investigators, Public Excerpts Prohibited (Level 1)*

If this level of permission is granted, Databrary will share *Recordings* or *Self-Reported Health-Related Information* with all Databrary *Investigators*. *Data* that retrieved from Databrary for viewing, downloading, and analysis must be handled according to best practices in data security. However, no images or parts of *Recordings* may be shown outside of a research setting. This means that no *Excerpts* from recordings may be shown in public settings.

#### *Sharing with all Databrary Investigators, Public Excerpts Allowed (Level 2)*

Many researchers now ask research participants for permission to show *Excerpts* in various public settings, such as classrooms and scientific talks. Databrary staff believe that from the point of view of a *Depicted Individual*, these distinctions between public venues are virtually meaningless and impossible to monitor or enforce. Researchers don’t consistently control who is in their classrooms or presentations, who records a lecture or talk, or how materials are distributed. Thus, under current practice, well-meaning researchers may be inadvertently implying that they have more control over an *Excerpt* than is actually the case.

Accordingly, Databrary’s Template Sharing Release requires that *Depicted Individuals* give separate *Permission to Share Excerpts* with the public. The template permission document asks whether depicted individuals give permission for *Investigators* to select *Excerpts* and to share them with the public for scientific or educational purposes.

In addition to sharing *Recordings* and *Self-Reported Health-Related Information*, this level of sharing permission enables Databrary *Investigators* to select and show *Excerpts* in public settings — in classrooms and at scientific or technical meetings -- as long as no *Personally Identifying Information* about a *Depicted Individual*, with the exception of faces, voices, and location information typically contained in *Recordings*, is included with the *Excerpt*.

*Investigators* must exercise professional judgment and uphold ethical principles in determining which *Excerpts* to show and to what audiences. *Investigators* must select *Excerpts* and edit them to minimize the possibility that a *Depicted Individual* can be identified, embarrassed, or harmed by the public showing of an *Excerpt*. Generally speaking, *Investigators* should refrain from creating excerpts for all individuals in a data set.

### **3.3.3 When should *Permission to Share* data be sought?**

Permission to share *Recordings* or excerpts with Databrary should be sought after a testing session is complete. This creates a clear separation between the consent an individual (or parent or legal guardian) gives to participate in research and the *Permission to Share* sensitive data. It also ensures that individuals can give informed consent about what sorts of behaviors were recorded during a testing session, and thus, what will be shared. Separating the consent to participate from the *Permission to Share* may also reduce the likelihood that participant in-take procedures are unduly prolonged by the additional procedures involved in sharing data. No *Investigator* should coerce depicted individuals or research participants to share.

### **3.3.4 How will the levels of permission granted by participants be linked to files?**

*Investigators* whose recordings contain *Identifiable Data* will be responsible for ensuring that every *Depicted Individual* gives *Permission to Share* with Databrary. *Permission to Share* may be secured in writing on paper forms or by electronic signature, but *Depicted Individuals* must be given paper records if electronic signatures are given. Paper copies of the signed *Permission to Share* form(s) must be kept in secured files comparable to those used for storing informed consent forms. If no *Permission to Share* is secured or written documentation of a research participant or depicted individual’s preference is lost or damaged, then the data in question may not be shared.

*Investigators* may keep electronic records, in tab or comma-delimited text files, spreadsheets, or comparable formats that provide information about the participant and the level of permission granted.

When uploading a data set to Databrary, *Data Contributors* will enter the permissions information into a web form along with any *Data* that they upload, so that Databrary can associate permissions levels and any demographic information available with each data file. *Investigators* are strongly encouraged to provide participant’s age-at-testing and sex, but are not required to provide any information. A missing permissions level will be treated as the lowest level.

### 3.3.5 How will levels of *Permission to Share* be maintained within the Databrary system once files have been uploaded?

Databrary will maintain a database for each study, and each study will include individual-level records for each data set for which sharing permission has been granted. Any data with permission level 0 will never be shared with all Databrary *Investigators*, and in some cases these files may not be uploaded at all. However, Databrary will still collect demographic and de-identified data associated with these participants. This is necessary in order to allow meta-analyses and subsequent data set users to determine what proportion of the shared or published data was actually shared with Databrary. For files with *Permission to Excerpt*, *Investigators* may create *Excerpts* for use in public settings.

Files denoting the individual-level *Sharing Levels* will accompany each data set when it is downloaded. Before downloading data, *Investigators* will be reminded of their promise to upholding sharing permissions granted by the original participants and to their other obligations that are part of the *Databrary Investigator Agreement*.

A future version of the Datavyu<sup>17</sup> coding tool may use the *Permission to Share* file to provide reminder alerts/warnings to *Investigators* about the level of access permitted for a given data set. But, for the time being, *Investigators* will have to monitor the use manually.

## 3.4 Sharing with Databrary

Before a data set may be shared with Databrary, *Investigators* must evaluate *Recordings* and associated individual-level *Metadata* for volatile or sensitive content. Even if a participant (or parent or legal guardian) has given *Permission to Share*, the ultimate responsibility for determining whether data should be shared lies with the *Investigator*. If an investigator determines that an individual's data set should not be shared, the investigator will mark it as such in Databrary.

### 3.4.1 Investigators who contribute data provide study-level metadata

Databrary will include investigator-level, study-level, participant-level, and measure-level metadata. Researchers provide investigator-level metadata at the time they apply for Investigator status and preliminary study-level metadata at the time they apply to deposit a specific data set. Prior to or during the process of uploading participant-level data to Databrary, *Data Contributors* may provide enhanced study-level metadata. This may include manuscripts, posters or talks, figures, statistical analysis scripts, code books, or grant proposals. The default will be for study-level metadata to be open to the general public (open/unrestricted).

### 3.4.2 Data Contributors upload/deposit data to Databrary

*Investigators* may either upload data to Databrary manually, using the Databrary site's upload tool or use the automated file transfer system embedded within a future version of Datavyu. *Investigators* must log-on to the Databrary system in either case. With manual deposit, *Investigators* or *Affiliate Investigators* must select the study for which data are being uploaded, the participant permission file, and the individual files that will be uploaded.

Users of software compatible with the (to-be-developed) Databrary application program interface (API) will simply indicate that a data set is ready to share. The software will manage the upload/transfer process.

### 3.4.3 Databrary translates video and metadata into common format(s)

Once data sets have been uploaded to Databrary, they will be automatically converted into a standard set of formats. The specific formats will be determined based on an assessment of how best to preserve the data for the long term using open (non-proprietary) formats, but UTF-8 and H264 video formats are leading candidates for text and video files, respectively. Databrary will publish information about the data formats accepted for deposits and those used and supported within the data sharing system.

### 3.4.4 Databrary data quality checking

Once the file conversion process has completed, a set of automated and manual data quality checks will commence. The automated checks will assess whether sharing permission levels have been provided for every individual participant in the uploaded data set and whether *Personally Identifying Information* have been provided in the uploaded *Data*.

---

<sup>17</sup><http://datavyu.org>

Manual data quality checks will assess transcoding quality—whether video/audio is readable. If the quality of the transcoded *Recordings* is poor, they will be checked against the uploaded original files. If the original files are in good condition, they will be transcoded again. If the original files are in poor condition, the contributor will be asked to upload the files again.

Once Databrary has certified that an uploaded data set is ready to share, the data contributor will receive an email indicating that fact. Contributors may then review the data set. When contributors are ready to share the data set, they will log on to Databrary and share the data set.

### 3.4.5 Databrary publishes data with permanent URL

When a new data set is released for sharing, Databrary will announce the event on its website with the data set’s description, author(s), and URL. Databrary will also announce new data sets on Twitter, Google+, Facebook, and other social media, as appropriate. Investigators may choose to subscribe to a data set or another *Investigator’s* news feed to receive updates about new contributions or revisions to old ones.

*Investigators* may choose to share partial data sets and update or modify them as new data becomes available or new analyses are conducted. Databrary will maintain a versioning system similar to that used by Github<sup>18</sup> to maintain records about data provenance. Significant modifications or additions to existing data sets will also be published via the web, email, and social media.

## 3.5 Ensuring Ethical Practices, Handling Violations

### 3.5.1 Databrary Ethics Standards

Databrary intends to uphold the highest standards of integrity. These standards include the following:

- every Databrary staff member and *Investigator* will complete training in research ethics involving human participants from an IRB and keep current their human participants training certification with such an entity.
- every Databrary staff member and *Investigator* is expected to embrace and enact the principle of informed consent. This means that individuals who are recorded on video, audio or other means have the right to determine whether their data may be shared, and if shared, with whom.
- *Investigators* must accurately record the sharing permission preferences expressed by users. Investigators must transfer information about individuals’ sharing preferences accurately to Databrary.
- Databrary must label files accurately. *Investigators* who browse or download data must respect the sharing preferences expressed by participants. Databrary staff and *Investigators* are expected to embrace and enact principles of participant confidentiality. This means that all individuals who are recorded on video, audio or other means have the right to have their identities kept confidential.
- *Investigators* must not upload or share files that contain *Personally Identifying Information* beyond that contained in *Recordings*. *Investigators* must not create new or modify existing *Data* files in ways that increase the risk that a participant’s confidentiality may be violated.
- If an individual participant’s identity is determined inadvertently, that information must not be shared with anyone, including members of the *Investigator’s* research team. Information about the individual’s identity may not be recorded in any form.
- Databrary staff and *Investigators* are expected to monitor the use of Databrary for violations of the principle of informed consent by others and to report any concerns about possible violations to the Databrary Ethics Committee at (ethics@databrary.org<sup>19</sup>) or via the Databrary website<sup>20</sup>.
- In selecting excerpts of data for public viewing, *Investigators* must exercise professional judgment in order to minimize the chance that an individual participant could be harmed.
- In selecting data sets to share with Databrary, *Investigators* must exercise professional judgment to avoid sharing data of an offensive or prurient nature. Databrary reserves the right to remove data or data sets that are deemed to be of an inappropriate nature.

### 3.5.2 Databrary Ethics Committee

Databrary will form an Ethics Committee to monitor the project, review and modify ethics policies, and to adjudicate ethics complaints. The Ethics Committee will consist of five members of Databrary’s Advisory Board,

---

<sup>18</sup><http://github.com>

<sup>19</sup><mailto:ethics@databrary.org>

<sup>20</sup><http://databrary.org/contact.html>



staff representatives from the NYU and PSU compliance offices, three members of the community of *Investigators*, and the Databrary Staff. The committee will meet via conference call or webinar quarterly.

## 4. Related Policies and Procedures

- Databrary Investigator/Affiliate Investigator Agreement<sup>21</sup>
- Databrary Best Practices for Data Security<sup>22</sup>
- Databrary Bill of Rights<sup>23</sup>

---

<sup>21</sup><http://databrary.org/user-guide/policies/investigator-agreement.html>

<sup>22</sup><http://databrary.org/user-guide/policies/best-practices.html>

<sup>23</sup><http://databrary.org/user-guide/policies/bill-of-rights.html>