

Barth syndrome is an ultra-rare disease that affects 130 affected individuals in the US and 300+ individuals worldwide, thus each piece of data generated through research participation is of outsized importance. Capitalizing on the infrastructure of the NINDS BRICS GUID system, GUIDs can efficiently connect retrospective and prospective Barth syndrome clinical research datasets that employ the same GUID platform, thereby maximizing the current and future uses of each dataset. Importantly, researchers will only need to do this process once for each research participant. Alongside BSF’s available data-sharing agreement template, protocol and consent language, our goal is to make data-sharing faster and more efficient across our entire research field.

Table of Contents

| | |
|--|-----------|
| GUID FAQ..... | 1 |
| What is a GUID? | 1 |
| How does GUID work? | 2 |
| How would a researcher implement the GUID?..... | 2 |
| What pieces of information are required to create a GUID? | 3 |
| But can’t we just manually de-identify the data? | 3 |
| Is the GUID compliant with data regulations outside of the United States?..... | 3 |
| How does somebody get a GUID? | 4 |
| Have there been other research groups that adopted the use of GUIDs?..... | 4 |
| Recommended language for the protocol and Informed Consent for collecting GUIDs in protocols..... | 5 |
| Protocol language – Example 1 (summarized) | 5 |
| Protocol language – Example 2 (detailed)..... | 5 |
| Informed Consent language: | 6 |
| Methodology language | 6 |
| Sample NYU/NIH Database Consent and Assent..... | 8 |
| Sample NYU/NIH Database Parental Permission 12-17 years old..... | 9 |
| Sample NYU/NIH Database Parental Permission up to 11 years old..... | 10 |

GUID FAQ

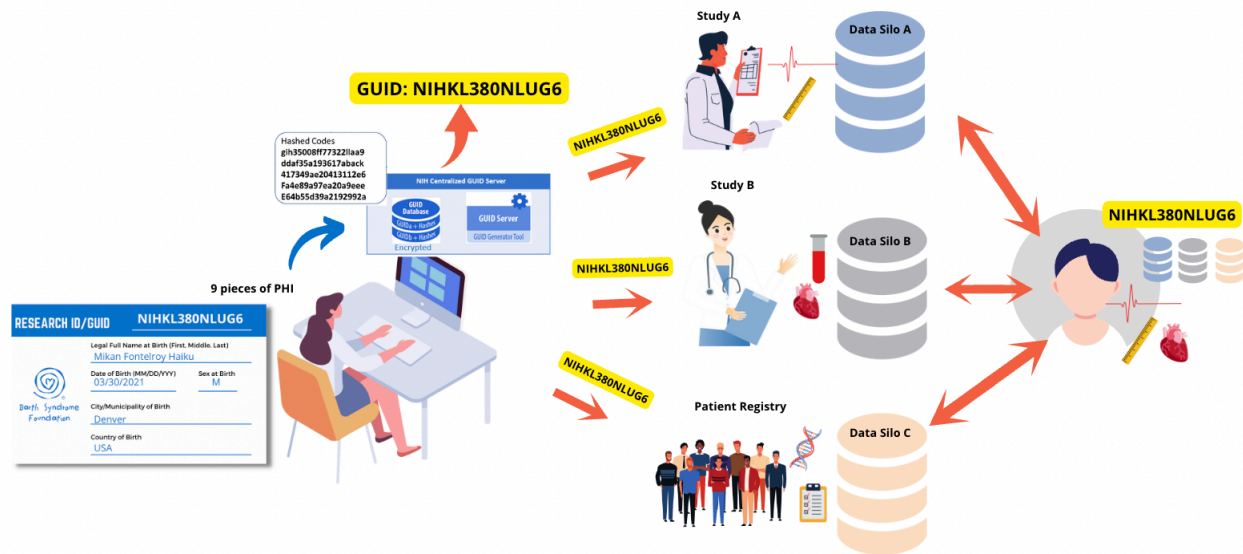
(adapted from <https://bricsguid.ninds.nih.gov/>)

What is a GUID?

GUID stands for Global Unique Identifier and it is a subject ID that allows researchers to share data on a study participant without revealing any personal identifiable information (PII) or protected health information (PHI). This allows researchers to match participants over time and across research sites, labs, and data repositories. GUIDs are generated through the secure infrastructure of the National Institutes of Health (NIH) National Institute of Neurological Disorders and Stroke (NINDS) Biomedical Research Informatics Computing System

Last Updated: September 2024

(BRICS). The core design and innovation of this collaborative research effort is that **PII/PHI never leaves the research institution or organization that is inputting the data.**



Adapted from (Nesbitt & Murphy, 2022)
NIH NDA GUID Tool

Figure 1: Schematic of the GUID creation process and its application for data sharing

How does GUID work?

GUIDS are created using the BRICS NINDS GUID Tool, which is a piece of software that accepts personal information of subjects, and uses that to create a series of hash codes sent to the NIH system and checked against the GUID database.

How would a researcher implement the GUID?

1. An authorized user of the research team [signs up for an account](#) and is vetted by the NIH.
 - Approval takes about 1 week, if the approval process is taking longer than that reach out to NINDSBrics-ops@mail.nih.gov to inquire.
2. Once the account is approved, the user navigates to the [NIH NINDS BRICS GUID tool](#) and selects 'Create a GUID' under the Subject Management menu.
2. The user enters the participant PII into the tool/GUID portal and submits the information.
3. **Without the PII ever leaving the user's computer**, the tool will generate a series of one-way hash codes based on the PII entered.
4. The hash codes (which do NOT contain PII) are encrypted and securely sent to the GUID system.
5. If the hash codes match an existing hash code, the associated GUID is sent back to the user.
6. If the hash codes do not match an existing hash code in the system, a newly created GUID is generated and sent back to the user.
7. Without revealing the PII of the individual, the GUID is created/retrieved and can be used to check for matches across other research studies that also employ the GUID.

What pieces of information are required to create a GUID?

| Information required to create a GUID | Comments/Considerations |
|---------------------------------------|---|
| Sex | Biological sex as appears on birth certificate. |
| Full First name | Only legal first name as it appears on birth certificate. No initials, nicknames, or blanks. <ul style="list-style-type: none">• If an individual legally changed their name, do not use the new name, only the one on the birth certificate. |
| Full Last name | Use the last name as it appears on the birth certificate. No initials or blanks. Omit any suffixes (e.g., Jr. III, etc.). <ul style="list-style-type: none">• If an individual legally changed their name, do not use the new name, only the one on the birth certificate. |
| Full Middle name | Only legal middle name as it appears on birth certificate. No initials, nicknames, or blanks If there is no legal middle name, you must use the tool to specify there is no middle name. <ul style="list-style-type: none">• If you do not have this information or the middle name is unknown, a GUID cannot be created. |
| Date of birth | The date of birth makes up 3 required fields (month, day, and year). |
| City/municipality of birth | Do not include state/country. Only the city/municipality name as it appears on the birth certificate should be entered. <ul style="list-style-type: none">• If the birth certificate contains both, check with the participant which one they used in the past and inform them to use that field consistently moving forward. |
| Country of birth | The country of birth as it appears on the birth certificate. |

But can't we just manually de-identify the data?

The advantage of the GUID is that once a researcher enters the required pieces of information, it will generate either a new GUID or match to the individual if they already exist in the system. This GUID will stay with that research participant forever and will only need to be generated once.

Manually de-identifying the data is a time and labor-intensive process, which can be expensive and inefficient. This is especially true for research that encompasses several datasets, exists across different institutions, or if protocol revision and reconsent of participants is required.

Is the GUID compliant with data regulations outside of the United States?

Data regulations do not apply in the creation of a GUID. However, the research site inputting the PII must handle and store the PII in accordance with their respective country's regulations. The NINDS BRICS GUID administrators have indicated that the GUID tool only receives hash codes during the data transmission

Last Updated: September 2024

process. When a user enters the PII onto the GUID tool the transmitted information are NOT PII, but rather de-identified hash codes. The only output that the inputting user receives back is the de-identified GUID number.

How does somebody get a GUID?

A GUID can be generated for research participants by contacting the Barth Syndrome Foundation and/or in a research study that employs the GUID. Through the registry or participating studies, Barth Syndrome Foundation (BSF) or study staff collects the 9 pieces of PII required. To employ quality assurance and the fidelity of the overall effort, BSF equips research participants with a GUID/ResearchID card (see Fig. 2) such that individuals can provide the same information across the studies they participate in. BSF also provides training to external research teams on the GUID creation process.


| | | |
|---|--|------------------------------|
| RESEARCH ID/GUID | | NIHKL380NLUG6 |
|  | Legal Full Name at Birth (First, Middle, Last) | <u>Mikan Fontelroy Haiku</u> |
| | Date of Birth (MM/DD/YYYY) | <u>03/30/2021</u> |
| | Sex at Birth | <u>M</u> |
| | City/Municipality of Birth | <u>Denver</u> |
| | Country of Birth | <u>USA</u> |

Figure 2. Example GUID card

Have there been other research groups that adopted the use of GUIDs?

- Simons Foundation SFARI & SPARK Autism research programs
 - (see: <https://academic.oup.com/jamia/article/17/6/689/843868>)
- Parkinson's Foundation
- The Association for Frontotemporal Degeneration
- NYU (see: <https://www.nyu.edu/research/resources-and-support-offices/getting-started-withyourresearch/human-subjects-research/forms-guidance/NIHGUID.html>)

Recommended language for the protocol and Informed Consent for collecting GUIDs in protocols

(courtesy of Association for Frontotemporal Degeneration, extracted from the Rare Diseases Clinical Research Network)

Protocol language – Example 1 (summarized)

De-identification will be accomplished using a Global Unique Identifier (GUID) generated via the Biomedical Research Informatics Computing System (BRICS) at the National Institute of Health. GUIDS are generated using the following personal identifiable information (PII): complete legal name given at birth, date of birth, city of birth, country of birth, physical sex of subject at birth. The PII is entered into the BRICS GUID tool and the software generates a one-way encryption (i.e., one-way hash), which is sent to the GUID Client to determine if the research subject hash codes have been seen by the system before. The encrypted hash codes do not have information to recreate the PII. However, they do have enough information to determine if a research participant GUID already exists in the system. The PII never leaves the institution entering the data.

Protocol language – Example 2 (detailed)

(courtesy of NIH BRICS GUID manual)

Every individual who provides consent will be assigned a Global Unique Identifier (GUID). The GUID enables data to be associated with a research participant without exposing or transferring the research participant's personally identifiable information (PII). It is randomly generated, alphanumeric code that is not generated directly from PII. This capability allows data about a research participant to be accumulated across projects over time, regardless of where and when that data were collected.

Researchers collect PII from their participants and store that data in a local database that is not made available outside the research institution. These data will only be available to a limited number of individuals. It is the availability of these data at the investigator's site that is used to generate GUIDs. This is made possible by issuing special software that runs at the research site on an investigator's computer or allowing users to access it through an application. This software performs a one-way encryption, often called a one-way hash, which is sent to the NIH GUID Client to determine if the research subject hash codes have been seen by the system before. The encrypted hash codes do not have information to recreate the PII. However, they do have enough information to determine if a research participant and associated GUID already exists in the system.

The PII fields that are used to generate a GUID are listed below. The PII fields are pieces of data that will not change over the lifetime of the participant and are uniquely specific to the participant. Each PII field has an associated probability of a match in the general population. By combining full legal name, date of birth, and municipality of birth, the probability that two individuals share the same information and thus the same hash codes (i.e. a false positive) becomes negligible and is the minimum required information to generate a valid GUID. Additional data that is provided beyond the required minimum further decreases the probability of a false positive.

Last Updated: September 2024

In order to generate a GUID, the following information and PII is required:

- Complete legal given (first) name of the subject at birth
- If the subject has a middle name
- Complete legal family (last) name of subject at birth
- Day of birth
- Month of birth
- Year of birth
- Name of city/municipality in which subject was born
- Country of birth
- Sex at birth

Informed Consent language:

You will be assigned a Global Unique Identifier (GUID). The GUID is a unique code made up of letters and numbers that allows researchers to share data from other studies in which you have participated without letting others know who you are.

We will ask you for your full date of birth (day, month, year), first name at birth, last name at birth, middle name at birth (if applicable), city/municipality of birth, country of birth, and biological sex at birth. This personally identifiable information (PII) will be entered into the NIH Centralized GUID generator software program. Once the GUID is produced, there is no way to get back to your PII. The software will not keep your PII, but will have enough information to determine if you already have a GUID assigned in the system. If you participate in another project and provide the same PII, you will be assigned the same GUID. Your GUID will be part of our research records.

You do not have to agree to have a GUID number used to participate in this research.

Do you agree to having a GUID number used in this trial?

Methodology language

(courtesy of Simons Foundation <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3000750/>)

A GUID for research purposes is a random sequence of characters that is unique to each research participant, regardless of the study. The process for generating a GUID for a participant is implemented here as a web service (GUIDWS), involving a client application and a server application. The two applications communicate via hypertext transfer protocol and simple object access protocol. The following steps provide a broad outline of the process, with details described in the following sections.

Last Updated: September 2024

Using the client application, the researcher enters a specific set of PII obtained from the research participant, such as name, date, and city at birth. The client application processes the identifiers into several intermediary codes using a one-way hash function and transmits the codes in a secure manner to the server application.

The server application compares the transmitted hash codes against an internal database. If there is no match, the server application generates a random GUID according to a particular format, and stores the association between the hash codes and the GUID for future use. If the codes match those from a previous transmission, the associated GUID is obtained. In either case, the server returns the resulting GUID to the client application.

The researcher obtains the GUID from the client application, and stores it locally to establish an association between the GUID and the participant's research data.

In order to share the participant's research data with other investigators, the researcher removes all identifying information except the GUID. This anonymized dataset can now be linked with other datasets that follow this process (assuming appropriate consent agreements are in place).

Sample NYU/NIH Database Consent and Assent

NIH Data Sharing Consent Addendum for [IRB STUDY NUMBER]

If you agree, your data from this study will be submitted to the National Institutes of Health (NIH) database. This large database is where deidentified study data from many NIH studies is stored and managed. If you participate in more than one NIH study, then your data from each of these studies may be combined. To do so, the researchers need to collect certain identifiable information about you (see bottom of page). However, the researchers will submit only deidentified data to the database. Deidentified study data means that all personal information about you (such as name, address, birthdate and phone number) is removed and replaced with a code number.

During and after the study, the study researchers will send deidentified study data about your health and behavior to the database. Other researchers across the world can then request your deidentified study data for other research. Every researcher (and institutions to which they belong) who requests your deidentified study data must promise to keep your data safe and promise not to try to learn your identity. Experts at the NIH who know how to keep your data safe will review each request carefully to reduce risks to your privacy. Sharing your study data does have some risks, although these risks are rare. Your study data could be accidentally shared with an unauthorized person who may attempt to learn your identity. The study researchers will make every attempt to protect your identity.

You may not benefit directly from allowing your study data to be shared with database. The study data provided to the database may help researchers around the world learn more about health. You will not be contacted directly about the study data contributed to the database.

It is your choice whether your data is added to the database. You may decide now or later that you do not want your study data to be added to the database. *You can still participate in this research study even if you decide that you do not want your data to be added to the database.* If you decide any time after today that you do not want your data to be added to the database, contact the study staff, and they will tell the database to stop sharing your study data. Once your data is part of the database, the study researchers cannot take back the study data that was shared with other researchers before they were notified that you changed your mind.

Select one:

Yes, I agree to have my deidentified data added to the NIH database. (Sign and complete bottom).

No, I do not agree to have my deidentified data added to the NIH database. (Do not sign or complete bottom. You can still participate in the research study).

You have received a copy of this document to keep.

Subject's Signature & Date

Provide information as it appears on your birth certificate

First name: _____ Middle name: _____ Last name: _____

Date of birth (mm/dd/yyyy): _____ Sex: _____ City/municipality of birth: _____

Last Updated: September 2024

Sample NYU/NIH Database Parental Permission 12-17 years old

NIH Data Sharing Parental Permission Addendum for Children 12 – 17 Years [IRB STUDY NUMBER]

If you agree, your child's data from this study will be submitted to the National Institutes of Health (NIH) database. This large database is where deidentified study data from many NIH studies is stored and managed. If your child participates in more than one NIH study, then their data from each of these studies may be combined. To do so, the researchers need to collect certain identifiable information about your child as they appear on their birth certificate, including sex; first, middle, and last name; date of birth; and city/municipality of birth. However, the researchers will submit only deidentified data to the database. Deidentified study data means that all personal information about your child (such as name, address, birthdate and phone number) is removed and replaced with a code number.

During and after the study, the study researchers will send deidentified study data about your child's health and behavior to the database. Other researchers across the world can then request your child's deidentified study data for other research. Every researcher (and institutions to which they belong) who requests your child's deidentified study data must promise to keep their data safe and promise not to try to learn your child's identity. Experts at the NIH who know how to keep the data safe will review each request carefully to reduce risks to your child's privacy. Sharing your child's study data does have some risks, although these risks are rare. Study data could be accidentally shared with an unauthorized person who may attempt to learn your child's identity. The study researchers will make every attempt to protect your child's identity.

Your child may not benefit directly from allowing their study data to be shared with the database. The study data provided to the database may help researchers around the world learn more about health. You and your child will not be contacted directly about the study data contributed to the database.

It is your choice whether your child's data is added to the database. You may decide now or later that you do not want your child's study data to be added to the database. *Your child can still participate in this research study even if you decide that you do not want their data to be added to the database.* If you decide any time after today that you do not want your child's data to be added to the database, contact the study staff, and they will tell the database to stop sharing your child's study data. Once your child's data is part of the database, the study researchers cannot take back the study data that was shared with other researchers before they were notified that you changed your mind.

Select one:

Yes, I agree to have my child's deidentified data added to the NIH database.

No, I do not agree to have my child's deidentified data added to the NIH database. (Do not sign or complete bottom. Your child can still participate in the research study).

You have received a copy of this document to keep.

Parent's Signature & Date

Sample NYU/NIH Database Parental Permission up to 11 years old

**NIH Data Sharing Parental Permission Addendum for Children under 12 Years Old
[IRB STUDY NUMBER]**

If you agree, your child's data from this study will be submitted to the National Institutes of Health (NIH) database. This large database is where deidentified study data from many NIH studies is stored and managed. If your child participates in more than one NIH study, then their data from each of these studies may be combined. To do so, the researchers need to collect certain identifiable information about your child (see bottom of page). However, the researchers will submit only deidentified data to the database. Deidentified study data means that all personal information about your child (such as name, address, birthdate and phone number) is removed and replaced with a code number.

During and after the study, the study researchers will send deidentified study data about your child's health and behavior to the database. Other researchers across the world can then request your child's deidentified study data for other research. Every researcher (and institutions to which they belong) who requests your child's deidentified study data must promise to keep their data safe and promise not to try to learn your child's identity. Experts at the NIH who know how to keep the data safe will review each request carefully to reduce risks to your child's privacy. Sharing your child's study data does have some risks, although these risks are rare. Study data could be accidentally shared with an unauthorized person who may attempt to learn your child's identity. The study researchers will make every attempt to protect your child's identity.

Your child may not benefit directly from allowing their study data to be shared with the database. The study data provided to the database may help researchers around the world learn more about health. You and your child will not be contacted directly about the study data contributed to the database.

It is your choice whether your child's data is added to the database. You may decide now or later that you do not want your child's study data to be added to the database. *Your child can still participate in this research study even if you decide that you do not want their data to be added to the database.* If you decide any time after today that you do not want your child's data to be added to the database, contact the study staff, and they will tell the database to stop sharing your child's study data. Once your child's data is part of the database, the study researchers cannot take back the study data that was shared with other researchers before they were notified that you changed your mind.

Select one:

Yes, I agree to have my child's deidentified data added to the NIH database. (Sign and complete bottom).

No, I do not agree to have my child's deidentified data added to the NIH database. (Do not sign or complete bottom. Your child can still participate in the research study).

You have received a copy of this document to keep.

Parent's Signature & Date

Provide information as it appears on your child's birth certificate

First name: _____ Middle name: _____ Last name: _____

Date of birth (mm/dd/yyyy): _____ Sex: _____ City/municipality of birth: _____

Last Updated: September 2024