



INPDR NP Global Unique Identifier (GUID) Standard

File name: INPDR NP GUID Standard v1.2 20250113

	Name	Title	Signature	Date
Author	David Storey	DPO	DS	13/01/2025
Reviewer	Shaun Bolton	COO	SB	13/01/2025
Authoriser	Conan Donnelly	CEO	CD	13/01/2025

Effective Date	13 th January 2025
Review Date	N/A

Version	Date	Author	Changes
V1.0	27 th Sep 2024	David Storey	To be approved NP GUID Standard for circulation and feedback
V1.1	2 nd Jan 2025	David Storey	Incorporated formal review feedback. Updated example to reflect "ToLowerCase" function
V1.2	13 th Jan 2025	David Storey	Clarified that NP GUID users are responsible for compliance.

Table of Contents

<i>Introduction</i>	3
<i>Scope</i>	3
<i>Terms and Definitions</i>	4
<i>NP GUID Specification</i>	4
<i>Use and Deployment Guidelines</i>	6
<i>About INPDR</i>	7

Introduction

Niemann-Pick diseases are a group of ultra rare conditions that are passed down in families. The conditions affect the body's ability to break down and use fats inside cells (such as cholesterol and lipids). Because of the buildup of fats, these cells don't work as they should and over time, the cells die.

There are two types of Niemann-Pick disease: Niemann-Pick Disease Type C and Acid Sphingomyelinase Deficiency (ASMD), which is also known as Niemann-Pick Disease Type A & B. ASMD is caused by changes to the SMPD1 gene. Niemann-Pick disease type C is caused by changes in the NPC1 and NPC2 genes.

A number of disparate data repositories exist to capture the natural history, genetics and trial outcomes of Niemann-Pick (NP) diseases patients. Whilst generally straight forward to establish, standalone repositories can quickly become data silos severely limiting their potential value. Researchers welcome the opportunity to aggregate and share data from multiple sources to develop new treatments and reduce timescales.

Given that Niemann-Pick diseases are categorised as rare diseases with quality data scarce, the ability to connect data from multiple sources can improve research. By implementing a standard NP community wide identifier, NP data sets can be linked, merged and aggregated by approved researchers.

In addition to providing researchers with a wider pool of research participants, the use of a standard identifier can improve data quality by removing duplicates and enabling accurate matching. Finally, the use of a standard identifier enables multiple facets of data to be taken into consideration when designing and performing studies (e.g. linking trial data with natural history data).

The use of a Global Unique Identifier (GUID) facilitates the linking of Niemann-Pick patient data across disparate data repositories. Whilst the concept of a GUID is not new, its application in medical research requires stringent compliance with global Data Protection Legislation and Ethical Obligations.

The adoption of privacy preserving techniques enables a consistent GUID to be generated per patient based on the input of a standard set of personal data. The resulting GUID's are unique to the patient but importantly, cannot be reverse engineered to reveal their true identity.

Scope

This standard specifies the requirements for a Niemann-Pick Global Unique Identifier (NP GUID). It is intended for use by organisations seeking to facilitate research into Niemann-Pick Diseases. It is applicable to organisations who wish to code the identity of registry, study or trial participants with a unique, pseudonymised standard identifier.

This standard establishes:

- the attributes of the participants personally identifiable data that shall be used as input to generate the NP GUID;
- the processing of the input required to ensure that the output is unique, irreversible and preserves the privacy of the participant;

- The format of the pseudonymised output to maintain readability and portability.

The standard provides guidelines for the use and deployment of NP GUID creation tools to ensure compliance with privacy legislation and ethical considerations.

Terms and Definitions

For the purposes of this document, the following terms and definitions apply:

CDC-32Hash

CRC-32 (Cyclic Redundancy Check 32) is a checksum algorithm that hashes input strings to 32 bit (8 character) values.

Global Unique Identifier (GUID)

A GUID is a Globally Unique Identifier. A GUID can also be known as a 'Universally Unique Identifier (UUID).

Pseudonymised

Pseudonymisation is the processing of identifiable personal data in such a manner that the personal data can no longer be attributed to a specific person without the use of additional information.

Salt

In cryptography, a “salt” is a piece of additional data added to the input of a hash algorithm. The use of a salt, adds an additional layer of complexity to further guarantee a unique output.

SHA-256 Hash

SHA-256 (Secure Hash Algorithm) is a cryptographic hash function that outputs a value that is 256 bits (64 character) long.

NP GUID Specification

Personal Identifiable Data

The following Personal Identifiable Data will be entered in the prescribed order to initiate the creation of the NP GUID.

Position	Field Title	Type	Size
1	Family Name	Character	40
2	Given Name	Character	40
3	Sex at Birth	Character (M or F)	1
4	Date of Birth	Date (DDMMYYYY)	8

The Personal Identifiable Data will be parsed and transformed into a concatenated string of lower-case characters for input to the NP GUID Creation module:

Field Title	Transformation
Family Name	<ul style="list-style-type: none"> • Convert to lower case (e.g. toLowerCase) • Prepend Family Name with “-” [ASCII – 2D Hex, 45 Decimal]
Given Name	<ul style="list-style-type: none"> • Convert to lower case (e.g. toLowerCase) • Prepend Given Name with “-” [ASCII – 2D Hex, 45 Decimal]
Sex at Birth	<ul style="list-style-type: none"> • Convert M or F to lower case (e.g. toLowerCase)

	<ul style="list-style-type: none"> Prepend Sex at Birth with "-" [ASCII – 2D Hex, 45 Decimal]
Date of Birth	<ul style="list-style-type: none"> Reorder DDMMYYYY to YYYYMMDD Prepend Date of Birth with "-" [ASCII – 2D Hex, 45 Decimal]
PID	<ul style="list-style-type: none"> Concatenate Family Name, Given Name, Sex at Birth, Date of Birth [e.g. String1.concat(String2, String3, String4)]

The resultant Parsed and Concatenated String shall be:

Field Title	Type	Size
PID	Character	96

Example

Personal Identifiable Data of Participant:

Family Name	Given Name	Sex at Birth	Date of Birth
Brown	Charlie	Male	30 th October 1950
Transformation			
brown	charlie	m	19501030
PID			
-brown-charlie-m-19501030			

NP GUID Creation

The parsed and transformed string of Personal Identifiable Data (PID) will be processed by the following privacy preserving methods leading to the creation of the NP GUID.

Field Title	Type	Size
Salt	Character	2
Sha256Hash	Character (Hexadecimal)	64
Cdc32Hash	Character (Hexadecimal)	8

The privacy of the participant shall be further protected by "Salting" the PID string with a constant value, in this case "np".

Field Title	Value
Salt	<ul style="list-style-type: none"> np

The Salt will be prepended to the front of the PID string.

Field Title	Transformation
PID	<ul style="list-style-type: none"> Concatenate Salt, PID [e.g. String1.concat(String2)]

The revised string consisting of the Salt and PID will be "Hashed" using a SHA-256 algorithm to create a 64 Character (32 Byte) Cryptographic Hash.

Field Title	Transformation
Sha256Hash	<ul style="list-style-type: none"> Generate SHA-256 Hash from Field Title PID [e.g. Node.js crypto.createHash (sha256)]

In order to aid usability and portability, the SHA-256 Hash will be further processed by a CRC-32 Hash to reduce the length of the hash from 64 to 8 characters.

Field Title	Transformation
Cdc32Hash	<ul style="list-style-type: none"> Generate CDC-32 Hash from Field Title Sha256Hash [e.g. Node.js crc32.js]

Example

Creation of a NP GUID:

PID
np-brown-charlie-m-19501030

Sha256Hash
d875475b351271291353ab034797086869dc53e281bd63a09d4b041eb52af944

Cdc32Hash / NP GUID
ac5517a3

Use and Deployment Guidelines

The generation of a NP GUID to facilitate the linking, merging or aggregating of Neiman-Pick data, should only be undertaken in compliance with local and global privacy legislation and ethical considerations. Compliance is the sole responsibility of the organisation making use of NP GUID's. The INPDR will not be involved in monitoring or enforcing compliance measures.

Organisations wishing to make use of GUID's should be transparent about their use, providing written information to participants, including the organisation's identity and any registrations with Ethical Bodies or Data Protection Authorities.

The explicit, informed consent of participants should be captured prior to the generation of a NP GUID. Records of consent should be held as documentary evidence as a minimum, for the duration of the trial, study or registry.

It should be as easy for a participant to withdraw consent should they wish to do so, as it was to initially grant it.

Depending on local geographic laws and obligations, each individual data source to be linked, may require individual consent to be maintained.

The deployment of a NP GUID Creation tool should be in a manner where processing takes place locally on the end user devices (e.g. client side - within the users browser). Any transfer or processing by a remote server should be avoided to prevent data export.

The ability to provide end users with a Copy / Paste capability should be provided to assist with data integrity, across data sets.

About INPDR

The International Niemann-Pick Disease Registry (INPDR) is a disease-specific patient registry. The INPDR mission is to document the Niemann-Pick patient experience in order to advance research and improve health outcomes.

As the INPDR is community owned and is independent of advocacy, policy and commercial interests, we are an appropriate organisation to define the standard for NP GUIDs, along with developing, supporting and promoting a generation tool.

The development of the Niemann Pick Global Unique Identifier (NP GUID) has been made possible by funding from the Ara Parseghian Medical Research Fund.

The INPDR are registered in the United Kingdom and can be contacted at:

- W: <https://inpdr.org>
- E: info@inpdr.org
- T: +44 (0)191 415 0693

Our postal address is: INPDR Registry Management Team, Suite 2, Vermont House, Washington, Tyne & Weir NE37 2SQ, United Kingdom