

REGISTRY IDENTIFIERS

A Case Study

Joy Payton

Kathryn Lowe



WHO ARE WE?

- Katie Lowe – Human subjects protection expert, genetic counselor, keeps studies compliant and safe.
- Joy Payton – IT generalist (mostly web development, project management, data science). Keeps study data nimble.

WHO ARE YOU?

Quick Poll:

I have worked with research data or researchers, I know this area.



I don't know much about healthcare research.

I really want implementation / algorithm details



A quick technical overview is plenty, thanks!

ITINERARY

- Research data vs clinical data
- Privacy and consent issues to keep in mind
- National Database of Autism Research (NDAR) as an example of a research registry
- Solutions for privacy

WHY DOES THIS MATTER TO YOU?

- You may have patient or research subject data that is shared with disease repositories
- You may need to understand why and how researchers share data
- You may be interested in developing new id's for your patients or research subjects

VOCAB

- Autism / ASD
- Subject / Participant
- NDAR is a registry, NDAR GUID is an identifier

WHAT IS THE CENTER FOR AUTISM RESEARCH (CAR)?

- Outreach
- Training
- Research
 - Gold standard diagnostic measures
 - Behavioral questionnaires and assessments
 - Brain imaging
 - Genetic testing
 - Clinical trials

WHO OVERSEES OUR RESEARCH?

- Institutional Review Board
- Protects the rights of research subjects
- Approved research must
 - Meet well-established ethical principles
 - Comply with CHOP policies
 - Comply with federal policies

BELMONT REPORT

Summarizes basic ethical principles for research involving human subjects

- Respect for Persons
- Beneficence
- Justice

INFORMED CONSENT

- A process, not just a document
- Serves as a basis of discussion
- Informs potential participants
 - Purpose of the study
 - Study procedures
 - Risks and benefits
 - How information will be used and protected

INFORMED CONSENT

To put in context for this presentation:

- Child appearing in this slide deck has a signed model release from parents
- All data shown here is fabricated
 - (But we could use real data if presenting to autism researchers, with consent)
- NDAR GUIDs are also fabricated

DATA SHARING

- Why share data?
 - Large sample sizes increase statistical power
 - Encourage collaborations
- Research Registries and Repositories
- IRB oversight of data sharing
 - Deidentified data
 - Coded data
 - Limited data set
 - Identifiable data

TYPES OF DATA

- Deidentified data: Participant 123 has BP of 140/80
- Coded data: Participant 123 (whose identity we know) has BP of 140/80
- Limited data set: Participant 123, in Philadelphia, PA, had a BP of 140/80 in January 2017
- Identifiable data: Jane Doe, MRN 12341234, had a BP of 140/80 measured on January 17, 2017

DATA SHARING LANGUAGE IN CONSENT FORMS

- HIPAA authorization
 - Sponsors; other CHOP employees
- Use of data for future research
 - May be optional or mandatory
- Sharing with national databases

EXAMPLE LANGUAGE

Use of Audio-Video Recordings

As part of this research study, we will be recording your behavioral testing and motor activities (which may include images of your face which can identify you, as well as your voice). If you allow CHOP to release these recordings to other individuals or organizations, please understand the recipients could use, distribute, broadcast, and/or publish them in ways that do not protect your privacy and that CHOP cannot control. Please indicate whether you will allow us to use these recordings for the following reasons by choosing the following:

For future research analysis at CHOP. Yes No

For future research analysis outside CHOP. Yes No

For presentations to professional and/or medical groups. Yes No

For publication in medical journals. Yes No

IMPORTANT TAKEAWAYS:

- Research subjects (or their families) make an agreement with researchers about data
- Research centers like ours have to abide by these agreements!

- (Sometimes, knowing if we're complying with the agreement is tricky)

NDAR

- National Database for Autism Research
- NIH: We give you federal dollars to research autism, so share your data back with the research community.
- Registry where data goes both ways

WHAT IS THE NDAR GUID?

- Global Unique Identifier (\neq the term in the IT sense – not 128 bits)
- **Uniquely identifies a subject for life**

NDARCJ743PV3

WHAT IS THE NDAR GUID?

- Form: NDAR AA NNN AA C
 - Alpha excludes {I, O, Q, S}
 - Final character (AN) is checksum
- Published in 2010 in the *Journal of the American Medical Informatics Association*, Available at <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3000750/>.

NDARCJ743PV3

GUID GENERATION

- PHI is hashed client side and the hash (and *only* the hash) is sent to the NDAR server
- A nondeterministic string is generated (for a new hash value) or retrieved (in the case of an old hash value) and supplied as the NDAR GUID

NDAR GUID Client version 5.1

Functions Settings Help

Please enter the information below TWICE for each field.

It is important that this information be the same as it appears on a birth certificate. To run many subjects at once, use the batch template.

[GUID Template](#)
[Pseudo-GUID Promotion Template](#)

* All fields are required

Complete legal given FIRST name of subject at birth

First Name * Retype First Name *

Complete legal given LAST name of subject at birth

Last Name * Retype Last Name *

Does subject have a MIDDLE name? * Yes No
 Re-enter answer * Yes No

If YES, please enter twice below. If NO, go to the next question.

Middle Name * Retype Middle Name *

Subject's BIRTHDATE (please enter twice)

Month * Day * Year *
 01-12 01-31 XXXX

Month * Day * Year
 01-12 01-31 XXXX

Physical SEX of subject at birth * Male Female
 Re-enter physical sex of subject * Male Female

Name of CITY/MUNICIPALITY in which subject was born

Birth Place * Retype Birth Place *

WHY IT MATTERS...



NDARCJ743PV3

Andrés Gallego-Smith
Jan. 1, 2004
Port St. Lucie, FL



HOW CAN WE USE THE NDAR GUID?

- Is it PHI? Is it identifying?
- Can we use it to share data outside of our submissions to NDAR? Say, if we want to collaborate with researchers by sending them a limited data set?
- What do we have to tell parents in our informed consent process?

HOW CAN WE USE THE NDAR GUID?

- If it's identifying, we need to say that to subjects (or their parents / consenters) before we get their permission to share it with NDAR.
- If it's PHI we should only share it if it's necessary, not as a matter of course or convenience.

JUST A REMINDER...

1. Names
2. All geographical subdivisions smaller than a state, except for the initial three digits of a zip code, if all zip codes with the same three initial digits has > 20,000 people
3. All elements of dates (except year) for dates directly related to an individual (including visit dates), and all ages related to the v. elderly
4. Phone numbers
5. Fax numbers
6. Electronic mail addresses
7. Social Security numbers
8. Medical record number
9. Health plan beneficiary numbers;
10. Account numbers
11. Certificate/license numbers
12. Vehicle identifiers and serial numbers, including license plate numbers
13. Device identifiers and serial numbers (URLs)
14. Biometric identifiers, including finger and voice prints
15. Full face photographic images and any comparable images
16. Any other unique identifying number, characteristic, or code, unless otherwise permitted by the Privacy Rule for re-identification.

“...used alone or in combination with other information to identify the individual.”

AN OBSERVATION

“Data can either be useful or perfectly anonymous but not both.”

Paul Ohm (see his excellent overview of re-identification in the UCLA law review <http://www.uclalawreview.org/pdf/57-6-3.pdf>)

USEFUL – BUT PHI

NDAR's success as a longitudinal identifier
make it PHI.

It's like a SSN – uniquely identifies subjects
for life, potentially in multiple institutions.

PROBLEM!



ID	IQ	SCQ	SP
NDARCJ743PV3	110	32	5
NDARXX123XX1	90	30	79

NDARCJ743PV3



Andrés Gallego-Smith
Jan. 1, 2004
Port St. Lucie, FL



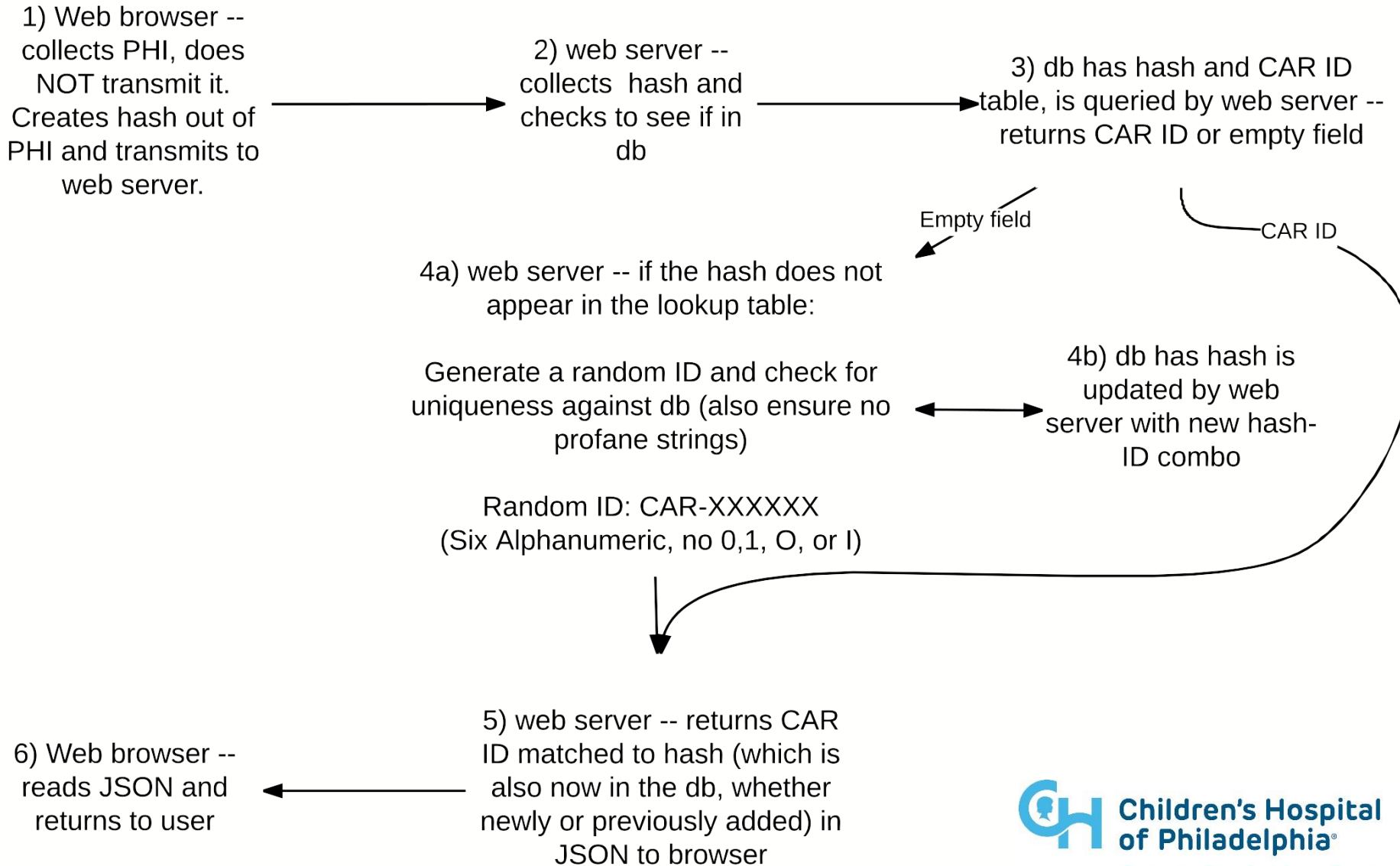
BUT WE LIKE NDAR GUIDS!

- We want to use something like the NDAR because we, as a center, often see the same subject as part of several different studies.
- We'd like to be careful not to “double count”, as well as conduct longitudinal studies.

BUT WE LIKE NDAR GUIDS!

- The solution? Take the “useful” part of NDAR (the adheres-for-life part) and leave out the “identifying” part (multiple centers know that Andrés is NDARCJ743PV3).

SOLUTION OVERVIEW



COMPARISON WITH NDAR GUID

- ONLY available to CAR researchers, so no re-identification possible by other institutions
- Uniquely identifies subjects for life

EXAMPLE

CAR Hash Generator and ID Lookup

Why?

Use a .csv File

Single Subject

First Name:

Please note that only the part of the first name before the first space will be kept, so Mary Jane will be reduced to mary.

Punctuation (apostrophes, hyphens) will be removed, so De'Angelo will be reduced to deangelo and John-Pierre will be johnpierre.

Last Name(s):

Punctuation (apostrophes, hyphens) and spaces will be removed, so D'Alessandro will be reduced to dalessandro and Smith Johnson will be smithjohnson.

DOB (must be yyyy-mm-dd: Include leading zeros as needed!):

CAR ID: **CAR** (This is the CAR unique ID for this subject).



WHAT ABOUT NDAR SHARING?

We have a solution to sharing with peers (don't use the NDAR GUID, but use something just as internally useful).

What about sharing with NDAR? **We've established that re-identification is a risk.**

NDAR GUID LANGUAGE

- Protocol
 - The **goal** of the GUID is to de-identify the research data such that the identities of participants cannot be **easily** ascertained.
- Consent form
 - Your name will not be shared with NDAR, **but the data may be identifiable** because the ID given to your data may be considered protected health information.

TAKEAWAYS

- Do you submit data to registries, or extract data from them? If so, consider:
 - What you tell families
 - How you classify those identifiers
 - The real usefulness of for-life identifiers
 - The hash-then-randomize algorithm

QUESTIONS / CLARIFICATIONS?