DATA REPOSITORY IMPROVEMENTS & HONEST BROKER BREAKDOWN

National Institute of Child Health and Human Development

Regina Bures, PhD – Program Lead, DASH
Jennifer Weck, PhD – Program Lead, NICHD Biorepository

CIVIC DIGITAL FELLOWSHIP

NIH

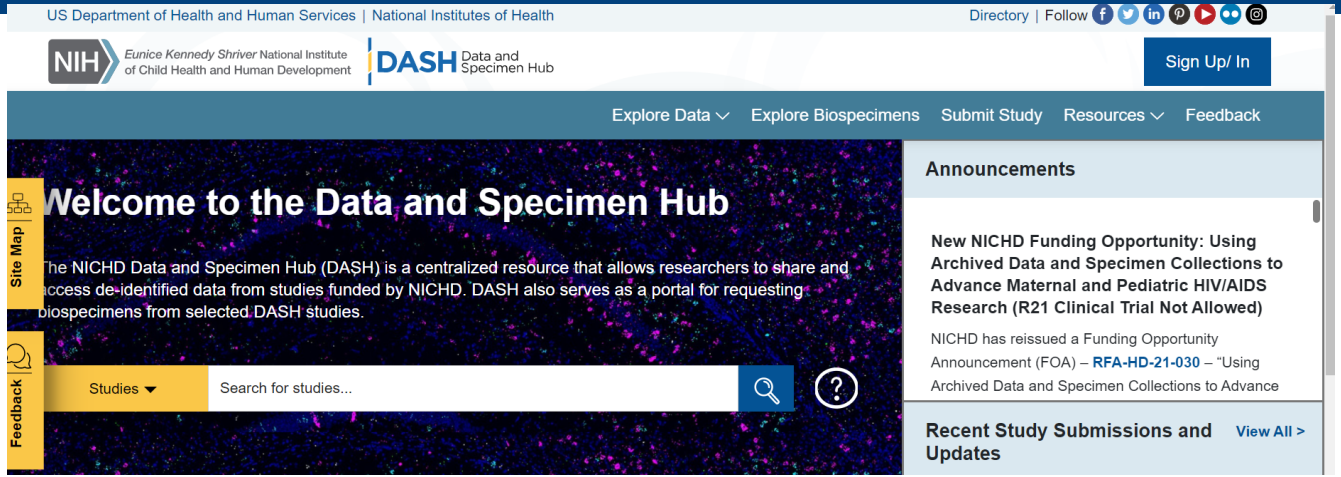
EMELIN FLORES
Rutgers University ‘21
Information Technology and Political Science

PROJECT OBJECTIVES

1. Develop or discover meaningful metrics to demonstrate or give benchmarks about where to move forward with the repository
2. Develop or discover privacy metrics to improve interoperability and accessibility to data available in DASH
3. Create a high-level guideline to curating and maintaining metadata organization in data repositories
4. Contribute in creating an ‘Honest Broker Hash Fact Sheet’ to explain how the Honest Broker works, implementation, and any questions that can be answered by research

CIVIC DIGITAL FELLOWSHIP

WHAT IS DASH?



DASH is a secondary research data repository. Researchers use DASH to conduct secondary research, test out research questions, build credibility, and train students. DASH also serves as a portal for requesting biospecimens from selected DASH studies.

CIVIC DIGITAL FELLOWSHIP

PROBLEM STATEMENT

DASH is looking to improve its usability by incorporating emerging technologies. The project looked to metrics for improvements. The main questions to answer to improve this data repository were:

- What metrics can help inform opportunities?
- Identify areas of improvement

CIVIC DIGITAL FELLOWSHIP

DASH REPO IMPROVEMENTS 1 – 2 – 3

What kind of data should we looking at to improve DASH?

In my findings, I found that we can find more insights about the user's journey through DASH by:

- looking into what Google Analytics tracks/the data they collect.
- Conducting usability studies and market research initiatives to understand a user's needs and pain points

CIVIC DIGITAL FELLOWSHIP

DASH REPO IMPROVEMENTS 1 – 2 – 3

What privacy metrics should we be looking at? What tools can we use?

In my research I found that there is potential to collect more privacy metrics than assumed.

- Some example privacy metrics are measuring how long does the de-identification process happen, measuring disclosure risk, linkages, modified risks, and much more
- There are ways to make de-identifying a document easier with certain tools like the NLM Scrubber or other third-party tools.

CIVIC DIGITAL FELLOWSHIP

DASH REPO IMPROVEMENTS 1 – 2 – 3

How do we organize metadata in order to use DASH better?

There are so many different standards and formats that it's hard to choose from. To better understand how you can organize your data to your best potential I created:

- A list of questions to consider when organizing data
- Suggested metadata standard and formats

CIVIC DIGITAL FELLOWSHIP

WHAT IS THE HONEST BROKER?

The Honest Broker is an automation tool for data sharing among de-centralized datasets. The Honest Broker utilizes a master patient index (MPI) and deidentified data to direct requested data to preauthorized destination systems while addressing concerns surrounding patient privacy, information security, informed consent, lexical equivalence, and identity matching.

CIVIC DIGITAL FELLOWSHIP

HONEST BROKER BREAKDOWN

Step 1 is completed by investigators

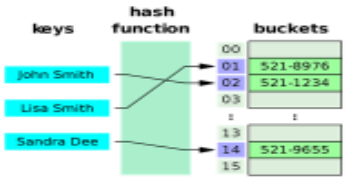
Steps 2, 3, 4 will be completed by either the site or honest broker and happen as rapid sequential steps

Step 1: Data Collection

Jon/John Smith
03/27/1945
Male

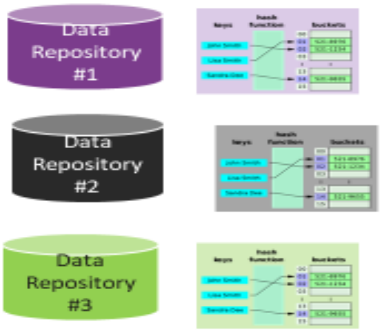
John Smith was admitted to IU Health (N3C Site) from his senior living facility due to shortness of breath.

Step 2: The data collected will be hashed and joined in a hash table.



This process is used to ensure the integrity of PHI

Step 3: Hashed data in previous step will be stored in appropriate data repository.



Step 4: Tokenization is used for data security and as an identifier to map back to sensitive data through system. The information hashed in each data repository is now tokenized.

Tokenization Algorithm → Output de-id tokens Patient 123

Tokenization Algorithm → Output de-id tokens Patient 456

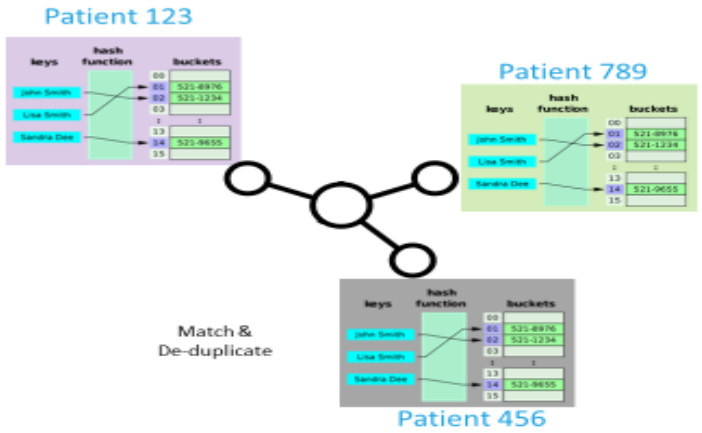
Tokenization Algorithm → Output de-id tokens Patient 789

CIVIC DIGITAL FELLOWSHIP

HONEST BROKER BREAKDOWN

Steps 5 & 6 will be completed by the honest broker

Step 5: The de-id tokens generated in step 4 will be used to find linkages between data repositories.



Step 6: In our example, we now know patient 123, 456, and 789 are the same patient and the patient data we are requesting. As a last step, we will collate the data and tokenize it.

CIVIC DIGITAL FELLOWSHIP

THANK YOU

Thank you to everyone who made this fellowship possible!

CIVIC
DIGITAL
FELLOWSHIP