

# Project Donate

Brendon<sup>1</sup>

<sup>1</sup>McGill University, Montreal, Canada

April 29, 2018

## Abstract

The purpose of this project was to develop a module on the Opal app to allow patients to donate their de-identified medical data to a secure repository to be used for medical research. De-identified medical data can be used to help develop new treatment methods, to improve care, support public health initiatives, etc. In-person data donation can be very tedious as the patient is required to visit in-person, fill out consent forms and fax them. That is why having an easy-to-use data donation module that allows patients to choose which data they would like to donate to what study is desired. Among the many de-identification techniques, variable suppression and reduction in detail were utilized in this project. This report describes the user interface platform that was developed as well as the two back-end methods that were outlined to collect data requested from the hospital's database and send them to the secure repository for research. A prototype study for testing was made to test with, however, there is much future work that can be done on this project by another student.

## Introduction

Many people donate their organs when they die. Organ donation is very important as each deceased donor can save several lives. Among the public, there is an inordinate amount of support for organ donation. In some countries, up to 80% of the public is in support of organ donation ([Shaw, 2015](#)). However, organs are not the sole thing that can be donated to save lives, medical data can be donated as well. Medical data is very valuable, as medical research cannot take place without data. This project focuses on developing an easy-to-use module on the Opal app for patients to donate their medical data.

There are many projects that are pushing for people to be able to donate their medical data posthumously just as they would as an organ donor ([Shaw et al., 2015](#)). Most people assume their health data is just deleted after they die. Depending on the jurisdiction, researchers can in fact access the medical data of deceased patients under certain circumstances as part of ethically approved research projects—for example, if the data are anonymized ([Shaw et al., 2015](#)). However, there are obvious difficulties in using “deceased data” compared to using medical data from living individuals, for example, people who are dead cannot be asked to provide informed consent. These difficulties can be overcome by having patients donate their data while they are alive. Unfortunately, most people are not aware that they can donate their health data and would not even know-how to initiate that process. Additionally, it is difficult for researchers to find and get access to data. That is why developing a platform where patients can easily donate their de-identified medical data on the Opal app is significant.

For patients to donate their medical data, there are many regulations in place so that the data cannot be traced back to the patient (Taylor and Mandl, 2015). Data that cannot be connected to the identity of a patient is called “de-identified data”. It is patient information that has been wiped of all direct identifiers. From a privacy perspective, it is a basic principle of medical research to use health data in the least intrusive way in accordance with the specific research objectives. Common de-identification strategies involve removing identifying variables as well as generalizing “quasi-identifiers”. Quasi-identifiers are not unique identifiers but are sufficiently correlated to an entity that, when combined with other quasi-identifiers, can be used to form a direct identifier. The table below is a list of direct and quasi- patient identifiers:

Identifiers	Type
Name	Direct
Social Insurance Number/Social Security Number	Direct
Patient ID	Direct
Date of Birth	Direct
Postal Code/Zip Code	Quasi
Date of Admission	Quasi

Figure 1: Examples of Identifiers

The increased adoption of health information systems accelerates their potential to facilitate beneficial studies that combine large complex data sets from multiple sources. These systems represent an enormous, underused data resource for medical research (Jensen et al., 2012). For example, researchers used patient data to discover previously unknown adverse effects associated with diabetes medications with myocardial infarction, and to identify groups of individuals that were at risk for morbid events such as heart attacks (Brownstein et al., 2009). Additionally, with new computational techniques capable of analyzing large complex sets of data, the return on medical data donation is immense. Machine learning was from the very beginning designed to analyze medical data (Kononenko, 2001). There is particular interest in using machine learning for medical diagnostics. With a known correct diagnosis and patient records, these data can be used as input for a machine learning algorithm. This is of course an oversimplification, but the derived classifier can then be used to assist the physician when diagnosing new patients in order to improve the diagnostic speed, accuracy and reliability (Kononenko, 2001). The classifier can also be used to train students or physician non-specialists to diagnose patients in a special diagnostic problem. Due to the immense potential that medical data have whether used in clinical research studies, or in conjunction with new technologies, it is vital to develop a platform on the Opal app where patients can easily donate their de-identified medical data to a secure repository for research.

## Related Work

Since there is a fair amount of motivation for the sharing of medical records with researchers, there are other projects working in parallel on simpler ways to donate medical data with researchers in accordance with the HIPAA act regulations (Taylor and Mandl, 2015). For example, “Sync for Science” or “S4S” is a collaboration

between researchers at the Harvard Medical School Department of Biomedical Informatics, electronic health record vendors, and the United States federal government. Their goal is to facilitate patients sharing their medical data with researchers in a manner that benefits research participants, researchers, EHR vendors, and healthcare providers. S4S intends to benefit research participants by providing a straightforward way for them to contribute to scientific progress. They plan to benefit researchers by presenting a simple path for to access donated medical data. This is advantageous to healthcare providers as it imparts a reduction in staff time to support data requests, as they flow automatically through their vendor-supplied patient portal. Additionally, S4S provides a method to empower their health care provider customers, to facilitate research, to participate in the development of stronger health care systems, and to meet electronic health record Incentive Program requirements for API-based patient access ([sci](#)). Sync for Science uses the SMART OAuth2-based SMART on FHIR authorization specification for the sharing of clinical data and they annotate each data type with its [MU Common Clinical Data set](#) label.

In a similar manner to S4S, Bitmark has created a data donation app to help facilitate research as much as possible. In providing optional push notifications, the app automatically informs donors of missing data, new studies, and pending tasks and donations for the researcher. Moreover, the app makes it easy for donors to give authorized consent for use of their data, simply by the touch of a button. New studies are presented in the form of study cards. These study cards efficiently outline basic information about each study for users to view. If the study interests the user, they can participate by simply tapping the join button. Bitmark uses a blockchain-based system, and end-to-end encryption to ensure security through the entire process.

## Methods

### De-identification techniques

There are many techniques used for the de-identification of data. This section provides an overview of some of the most common ways of de-identifying patient data. Refer to the table below for an overview four of the main de-identification techniques:

De-identification techniques	Description	Example
<b>Reduction in Detail</b>	Involves rounding or collapsing values into broader categories	Round date of admission to month of admission
<b>Variable Suppression</b>	Supress known direct identifier	Supress name
<b>Random addition of Noise</b>	Add or subtract random amount from variable while keeping within certain range	Add small random value to patient's height
<b>Pseudonymization</b>	Replace identifying variable with an irreversible pseudonym	Replace name with another random name

Figure 2: Table of de-identification techniques

The de-identification methods that were used in this project were reduction in detail and variable suppression. Reduction in detail is by far the most common method of de-identification ([can](#)). It involves reduction in the detail of a variable by rounding or collapsing values into larger categories of data. This method can

be used on numeric variables as well as on postal or zip codes and dates. Reduction in detail was used by rounding the “date of first appointment” field in the mockup study to the month of the first appointment if it was selected to be donated. Another very common de-identification method is variable suppression. This technique involves the removal or withholding of identifying data. Variable suppression was utilized by withholding direct identifiers, such as name, when the patients selected which data they wanted to donate.

## Software Development

The AngularJS view, controller, service framework was used when developing the user-interface of this project. Papa John’s style guide was followed when coding the AngularJS aspects of this project. The files that were created for the user-interface were *donate.html*, *donateController.js*, *donateService.js*, *individualDonate.html*, and *individualDonateController.js*. To retrieve data from the Opal database, *requestToServerService.js* was utilized through the listener and Firebase. A new request was made to get the data to be donated, that is lab results and month of first appointment. In the backend, *donateAPI.js* was used to send the data to the repository for research. Firebase was used as the new secure repository where the data are sent to in order to be used for research.

## Results and Discussion

### User-Interface

Simplicity and ease of use were kept in mind when designing the user-interface platform for patients to donate their medical data. The more intuitive the platform is, the more likely it will be that patients donate their medical data. The Donate tab was added on to the General view (Figure 3). When clicked it brings the user to the Donate view. On the Donate view studies are listed that the patient is able to donate to (Figure 4). Once the patient clicks on the study they selected, they are brought to the Study view (Figure 5).

For our purposes, we created one mock biomarker study to which patients could donate their blood work data and the date of their first appointment (this value would actually be rounded to the month of their first appointment). The patient would then be able to select what data fields they would like to donate and tap the green button near the bottom of the screen to donate those data (Figure 6). A thank you note would then pop up with the option for the patient to go back to the Donate view (Figure 7).

As demonstrated, this data donation platform is very simple to use. Data donation on the Opal app is advantageous over other data donating services because all the patient’s data are already accessible through the app. This means that there are no forms that require to be filled out, or the data to be manually typed in by the user.

### Data Flow

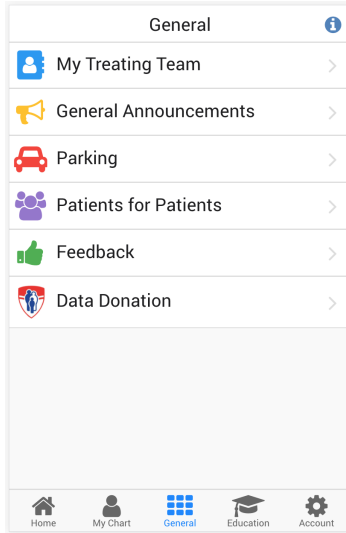


Figure 3: General view with added data donation tab

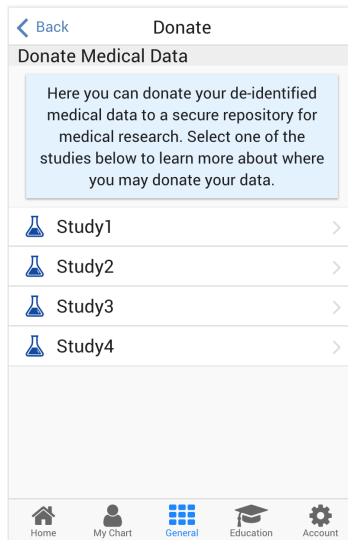


Figure 4: Donate view with list of studies

## Request Method

There are two main methods that were utilized on the back-end of the data donation platform. The request method (Figure 8) involved using the *requestToServerService* to request the data from the hospital database and bringing it to the app in the usual fashion. From here, the data from the app would then be sent to a separate Firebase, which would act as the central repository for research.

This method is convenient because most of the frame work was already built in; only the new requests had to be added. Then the data could be pushed to Firebase where they would be used for research. However, there are still a few bugs that need fixing before this method is fully functioning, particularly with the MySQL query. It would be helpful going over *AddingRequests.pdf* in the Github wiki to begin troubleshooting this.

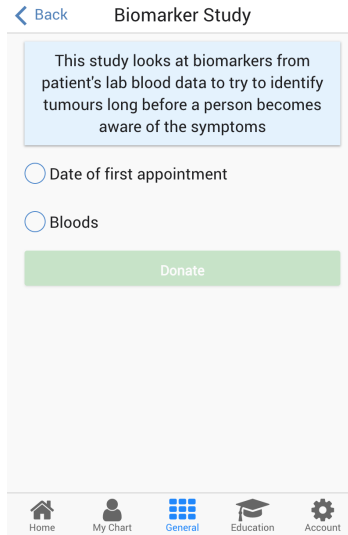


Figure 5: Study view

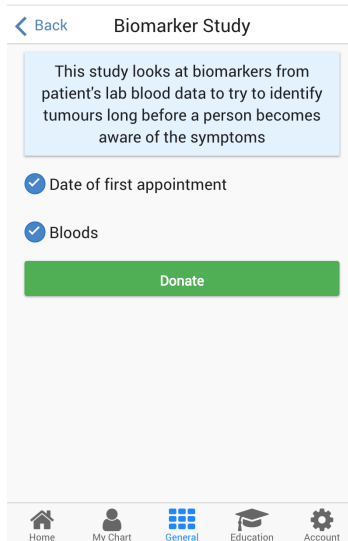


Figure 6: Study view with data for donation selected

The data The limitations with this method would be that it will not be able to scale well due to the Opal app not being designed to bring that all the data to the app of a potential study that needed many data fields.

### Harvesting Method

This method scales much better but unfortunately has not yet been implemented. The second method, or “harvester” method, involves harvesting data that is flagged to be donated. The logic is that a harvester is implemented to copy flagged data straight to the secure repository for research. These data would be flagged with a code which refers to the study that the data are being donated to.

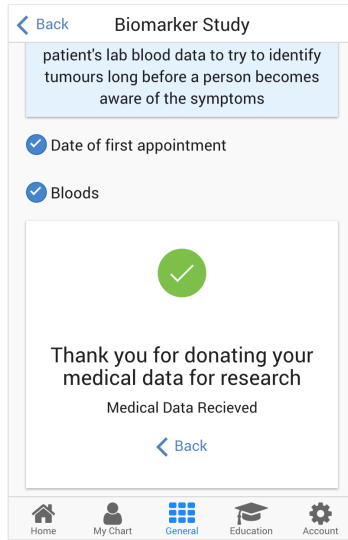


Figure 7: Thank you note with option to go back to the previous page

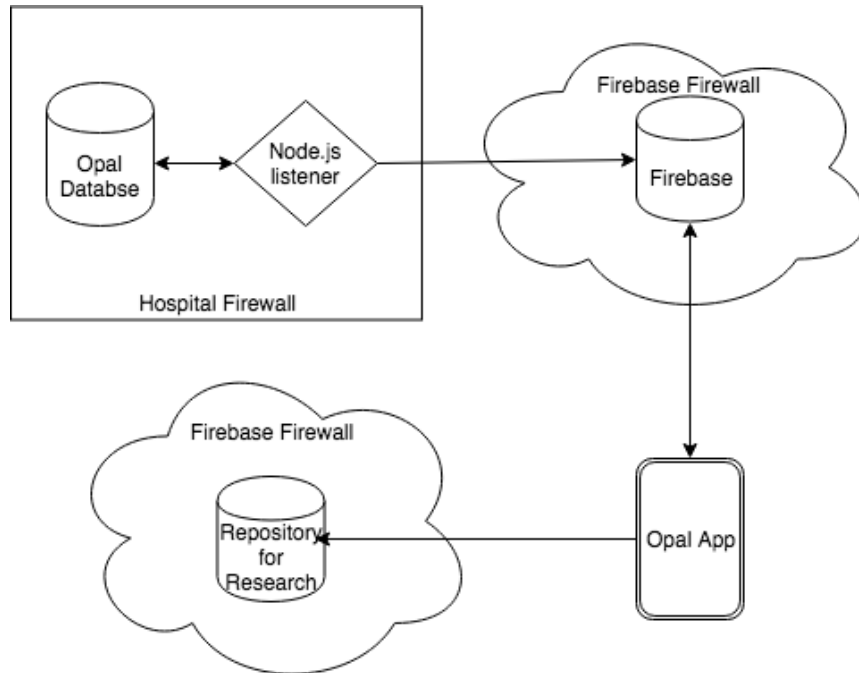


Figure 8: Data flow diagram for the “request” method

The “harvesting” method scales much better than the other, and is simpler as well as more secure. Making it overall the better method. However, the downside of the harvesting method is that the data are still within the “institution” and getting them out of the institution may not be simple, while the request method already gets the data out of the institution.

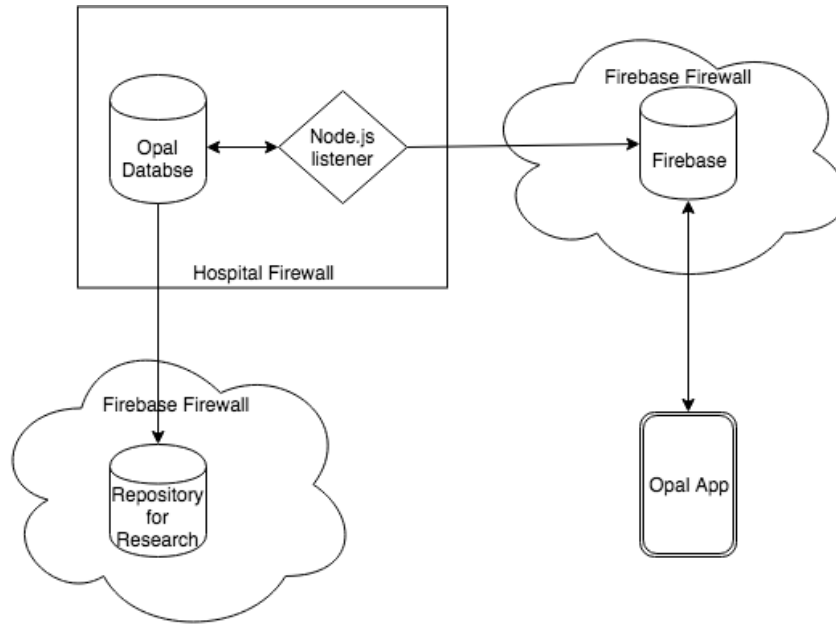


Figure 9: Data flow diagram for the “harvesting” method

## Future Work

The first next step would be to finish implementing the harvesting method. This method is preferred over bringing all the data for donation to the app. This is due to the app being designed to bring data to it and show the data, not to do anything with the data from there, such as sending it to a repository for research. This could be implemented by sending a request to the Opal database to add a flag which refers to the study that the data is being donated to. From there, implementing the harvester to copy the flagged data to the secure repository for research.

Once this is complete, adding a database of studies and their required data fields would be the next part of the platform to work on. Adding a matching algorithm that matches patients to studies well suited towards them would make the platform easier to use so. This way patients would not see a whole bunch of studies on the app to which they are not applicable to make a donation. Additionally, figuring out what happens to the data once they reach the secure repository for research needs to be done. Specifically, how do researchers access these data and how does the app only let researchers see the data that has been donated to them? One method that is possible would be for there to be a user interface for researchers to post their studies and retrieve the data that is donated to them from the Firebase. Further work needs to be done on how the researchers would access the medical data from the repository. The limitations of this project were my unfamiliarity with technologies used and with web developing in general, however, I believe a good foundation was laid out for future students to build off of before integrating this data donation platform with the Opal app.

## Conclusion



The results in this report detail a data donation module that is easy-to-use. Additionally, this report presents an outline for a better data flow method that involves harvesting data from the database and directly copying it to the secure repository for research. With further work done on this platform, this data donation service can be integrated into the Opal app and could potentially lead to new medical research advances and saving lives.

## Acknowledgements

I would like to thank John Kildea and Laurie Hendren for their mentorship throughout this project. I would additionally like to thank James Brace and the rest of the Opal developing team for helping me get started with the software developing side of the project and aiding with the learning curve of working with multiple unfamiliar technologies.

## References

- EHealthInfo Canada. <http://www.ehealthinformation.ca/wp-content/uploads/2014/08/2011-Best-Practice-Guidelines-for-Managing-the-Disclosure-of-De-Identificatied-Health-Info.pdf>. URL <http://www.ehealthinformation.ca/wp-content/uploads/2014/08/2011-Best-Practice-Guidelines-for-Managing-the-Disclosure-of-De-Identificatied-Health-Info.pdf>. Accessed on Sun, April 22, 2018.
- Sync For Science. <http://syncfor.science/>. URL <http://syncfor.science/>. Accessed on Sat, April 21, 2018.
- J. S. Brownstein, S. N. Murphy, A. B. Goldfine, R. W. Grant, M. Sordo, V. Gainer, J. A. Colecchi, A. Dubey, D. M. Nathan, J. P. Glaser, and I. S. Kohane. Rapid Identification of Myocardial Infarction Risk Associated With Diabetes Medications Using Electronic Medical Records. *Diabetes Care*, 33(3):526–531, dec 2009. doi: 10.2337/dc09-1506. URL <https://doi.org/10.2337%2Fdc09-1506>.
- Peter B. Jensen, Lars J. Jensen, and Søren Brunak. Mining electronic health records: towards better research applications and clinical care. *Nature Reviews Genetics*, 13(6):395–405, may 2012. doi: 10.1038/nrg3208. URL <https://doi.org/10.1038%2Fnrg3208>.
- Igor Kononenko. Machine learning for medical diagnosis: history state of the art and perspective. *Artificial Intelligence in Medicine*, 23(1):89–109, aug 2001. doi: 10.1016/s0933-3657(01)00077-x. URL <https://doi.org/10.1016%2Fs0933-3657%2801%2900077-x>.
- David M Shaw. The ghost of donor passed. *BMJ*, page h6244, dec 2015. doi: 10.1136/bmj.h6244. URL <https://doi.org/10.1136%2Fbmj.h6244>.
- David M Shaw, Julianne V Gross, and Thomas C Erren. Data donation after death. *The Lancet*, 386(9991):340, jul 2015. doi: 10.1016/s0140-6736(15)61410-6. URL <https://doi.org/10.1016%2Fs0140-6736%2815%2961410-6>.
- Patrick L. Taylor and Kenneth D. Mandl. Leaping the Data Chasm: Structuring Donation of Clinical Data for Healthcare Innovation and Modeling. *Harvard health policy review: a student publication of the Harvard Interfaculty Initiative in Health Policy*, 14(2), 2015.