
# Your Memories. Private & Secure.

Stacey Baradit, Sindhuja Jeyabal, Alexander Jones,
Aditya Mishra & Stephanie Snipes

Professor John Chuang & Professor Kurt Beyer

May 2016
Capstone Project Final Report
Master of Information Management & Systems

Berkeley SCHOOL OF INFORMATION

# Table of Contents

# Project Summary

### What We Built & Why We Built It

*Datensparsamkeit* is a German term that succinctly encompasses the belief and attitude that we should handle only the minimum amount of information needed to accomplish our purpose.[1] It is taken from current German data protection legislation -- strict rules that govern how data is captured, stored, handled, and shared. Several German companies strive to uphold *datensparsamkeit* by developing according to privacy-by-design principles. For example, at Deutsche Telecom, users are not required, by default, to give up their exact locations through their smartphones. If this data is not necessary, their locations are blurred automatically.[2]

We wanted to build something that embodies *datensparsamkeit* for our MIMS capstone project. Data privacy and ownership are key interests for the entire team, but as we conducted our user research, we quickly found that these interests are not regularly shared by those outside of academia. People often hyperbolically discount data privacy and ownership; they cannot picture the potential consequences of giving up control of their data now for a more immediate benefit. The question for us soon became: How might we design a product that promotes and educates the importance of data privacy and ownership, in a way that is relatable to a mass audience?

Our answer to this is Portal, a mobile application that aggregates data streams from a user's digital life on popular web services, such as Facebook, Instagram, Foursquare, and Spotify. On the outside, Portal appears to be like any other app, serving up "memories" or collections of users' digital footprints, allowing users to retrace what they did, saw, and heard on a particular day through a friendly interface. However, unlike many other apps, Portal makes it a point to not store any of its users' data. A user's aggregated data is shredded, encrypted, and distributed across the blockchain network, making it impossible for anyone but the user to retrieve and recompile without a personal key.

---

[1] Bliki: Datensparsamkeit. (n.d.). Retrieved May 06, 2016, from http://martinfowler.com/bliki/Datensparsamkeit.html

[2] P., & Preuschat, A. (n.d.). German Companies Push for Tough New Data-Protection Rules in Europe. Retrieved May 06, 2016, from http://www.wsj.com/articles/german-companies-push-for-tough-new-data-protection-rules-in-europe-1424774254#:yKBsUjLm_AaEKA

Through Portal, we aim to prove to other developers that it is possible to build applications that prioritize users' privacy and data control without compromising on user experience. Hoarding users' data would not improve how Portal works; it will not enhance the user experience. Instead, we should uphold *datensparsamkeit*, taking and handling only as much data as we need, and utilize the latest in data protection technology, such as blockchain, to protect our users. At this point, Portal upholds privacy at the "Information Aggregation" stage and as blockchain gains more traction as a storage platform, privacy would be an integral part of other activities as defined by Daniel J. Solove.[3]

**Objectives and the Journey**

The goal for the project has been to give people power, awareness, and control over the privacy and security of their data. Initially, we intended to meet this goal by working with connected devices in the Internet of Things (IoT) space. However, after significant research and hands-on experimentation, it became clear that the industry was too nascent for us to narrow the solution down to a specific problem space. There were too many devices, too many platforms, too many uses, as well as a large gap in user needs between expert/early adopters and more mainstream users.

The crucial pivot was changing the set of aggregated data from futuristic, connected IoT devices to focusing on the most common connected device that people own: smartphones. From the smartphone, we further focused on digital data sources with which users are most familiar, such as Facebook, Spotify, and Instagram.

Over the course of four months, 20 in-depth interviews were conducted to understand user experiences with IoT, personal health fitness, self-documentation, memories, and privacy. Users ranged from 21 years to 76 years and the topic of conversations varied, but carried the goal of understanding the user's behavior and possible privacy concerns. The majority of interviews were exploratory and ultimately guided our evolving project.

In addition, another dozen interviews were conducted with domain experts in the IoT, entrepreneurship, privacy, and security spaces. Their initial recommendations

[3] P., & Preuschat, A. (n.d.). German Companies Push for Tough New Data-Protection Rules in Europe. Retrieved May 06, 2016, from http://www.wsj.com/articles/german-companies-push-for-tough-new-data-protection-rules-in-europe-1424774254

suggested the team pursue business-to-business (B2B) opportunities because businesses are more willing to pay for heightened security and privacy whereas consumers are still accustomed to free services. However, it seems that consumers are becoming more educated and aware of how their data is being utilized and are starting to show willingness to pay $1.00-$5.00 for apps that do not unnecessarily access their contacts or track their browser history or location.[4] Considering the B2B space is already being targeted by Microsoft and others, the challenging consumer space seemed a ripe place to begin.

A team member had personally used a similar app, HeyDay, that aggregated data into collections or digital ”memories.” The app then proceeded to shut down their services without notice, leaving the team member concerned and frustrated. There was no communication as to where her collected data went, and when further research was conducted, it was evident that the parent company, Hey, Inc., had a record of unethical data collection and reuse. For example, the company was using users’ uploaded Twitter images to provide the main content for another of their apps called Stolen. The app allowed users to “steal” profiles, alter them, and trade them like cards with virtual currency. The ethical issues were substantial enough to garner action from U.S. Representative Katherine Clark, who wrote a letter to both CEOs of Twitter and Apple expressing concern over the ability for people’s Twitter profiles to be traded, altered, and purchased without any consent from the owner of the profile. She argued it would ”enable online abusers” and provide them another tool “to harass, bully, and intimidate.[5]”

Based on our user interviews and the data privacy trends, we realized data aggregation in a private and secure manner should be the key value proposition of the product. Portal takes on the features and functionality that many users seek out in these ‘walk down memory lane’ apps, and has decided to incorporate a novel backend that is significantly more secure and private. Though privacy is a priority, the user experience will not suffer, but instead will utilize opportunities to inform and educate the user about privacy-related decisions. The goal is to design the application’s interface and

[4](n.d.). Retrieved May 06, 2016, from http://www.theatlantic.com/technology/archive/2013/12/study-consumers-will-pay-5-for-an-app-that-respects-their-privacy/282663/

[5]Game Shut Down After Democrat Katherine Clark Complains of 'Harassment' - Breitbart. (2016). Retrieved May 2016, from http://www.breitbart.com/tech/2016/01/15/game-shut-down-after-democrat-katherine-clark-complains-of-harassment/

architecture in a way that empowers the user to have full control over their content while also maintaining a satisfying, delightful, and pleasant experience.

## The Economics of Privacy

### Economics of Privacy for Individuals

To understand the larger context of the issue of data privacy and ownership that we would be tackling in our project, it was necessary to figure out how much individuals and companies value privacy. How important is privacy to users? How much would they be willing to pay for privacy? How much do companies value privacy? Why don't companies place more of an emphasis on users' privacy? By answering these questions, we gained a clearer understanding of how to design, position, and price our product for maximum effect with a mass audience.

Alessandro Acquisti, a MIMS alumni and expert on the behavioral economics of privacy, served as the basis for our understanding of how much individuals value privacy. A key takeaway from Acquisti was that consumers' valuation of privacy largely depends on how their options for privacy are presented to them in the first place.

In "What is privacy worth?" Acquisti and several other researchers conducted a study with gift cards: the first card was worth $10 and anonymous, the second card was worth $12 but its transactions could be tracked. The researchers found that when people were presented with the $10 anonymous card first, 52% refused to switch to the $12 gift card. They chose instead to keep the lesser amount in lieu of giving up their privacy. However, when the researchers presented people with the $12 card first, fewer than 10% switched to the $10 option, choosing not to "trade cash for greater privacy."[6] The study simply and clearly demonstrated that when people have privacy to begin with, they value it. They are less likely to give privacy up. However, when "the starting point is that we feel we don't have privacy, we value privacy far less."[7]

Consumers take part in transactions for their privacy everyday -- transactions that are often beyond their realization and control. With every Google search, a user "[sells] his personal information and [buys] search results. But people do not think about, or are unaware of, the notion that typed search requests help determine the ads that Google displays and what its ad network knows about them."[8] Consumers have slowly become

[6] Game Shut Down After Democrat Katherine Clark Complains of 'Harassment' - Breitbart. (2016). Retrieved May 2016, from http://www.breitbart.com/tech/2016/01/15/game-shut-down-after-democrat-katherine-clark-complains-of-harassment/

[7] Ibid.

[8] Lohr, S. (n.d.). You Want My Personal Data? Reward Me for It. Retrieved from http://www.nytimes.com/2010/07/18/business/18unboxed.html

accustomed to, and perhaps unintentionally accepting of, the attitude that this type of transaction and demand for personal information is commonplace.

In an analogy to Acquisti's research study, consumers accept and stay with less private apps like the $12 gift cards because they make up the majority of apps on the mass market. Consumers are not first offered privacy-centered apps because privacy has not been the focus for many of these apps. With Portal, we aim to provide consumers with an alternative. We want to give consumers the chance to start with the $10 anonymous gift card -- to start with privacy first.

**Economics of Privacy for Companies**

Why aren't there more privacy-minded alternatives available to consumers? What are the economics of data and privacy to companies? For context, many companies believe that collecting and storing large amounts of data will afford the potential for uncovering trends, predicting results, and achieving the ultimate goal of consumer personalization. More data than ever is being collected -- and hoarded. By 2020, annual data production is expected to increase by 4,300%, though most companies will only use a fraction of the data that is collected and stored.[9]

Foursquare, a popular social networking app that tracks a user's "check-ins" at local businesses and phone-reported locations, is one such company that has touted the benefits of collecting all of this data. Through its trove of background location reporting by its user base, the company was recently able to predict the downturn in Chipotle's Q1 sales in 2016 based on foot traffic recorded around franchise branches. This type of data, collected daily in huge amounts by Foursquare's user base, is sold to a multitude of buyers, including "retailers, real estate developers, Wall Street traders, and consumer package-goods companies."[10]

Foursquare's newest technology, known as Pilgrim, runs quietly in the background of users' phones, recording "every time a phone with Foursquare installed stops moving. When it stops moving, Pilgrim tries to figure out where exactly you are, if you've been there before, or if there's anything going on in the area you might be interested in. [...]

---

[9] http://www.forbes.com/sites/bernardmarr/2016/04/28/big-data-overload-most-companies-cant-deal-with-the-data-explosion

[10] https://www.washingtonpost.com/news/innovations/wp/2016/04/28/how-foursquare-knew-before-almost-anyone-how-bad-things-were-for-chipotle/

Pilgrim makes these decisions millions of times per day."[11] According to the company's CEO, "the biggest misconception that still exists about Foursquare is that it's reliant on manual check-ins. Pilgrim has made it possible to check in without taking your phone out of your pocket."[12]

Acquisti, in "The Economics of Privacy," describes this data collection and usage as a "market for personal data" which "involves free products or services provided to consumers in exchange for their data. [...] In these exchanges, consumers are directly involved in the transaction, although the exchange of their personal information is not always a *visible*, explicit component of the transaction."[13] This is information asymmetry, where companies hold the upper hand in knowing how and what kind of data is being collected, and for how long and where this data will be stored.

In these asymmetric transactions, data collection is often bundled into the idea of offering the product or service for "free." Consumers act under the assumption they are receiving a product or service, such as Foursquare, for free, without full disclosure that they will actually be giving up -- constantly, in Foursquare's case -- their personal information in exchange. Unknowingly, consumers therefore give up their personal information to companies free of charge.

Companies believe they are gaining wildly from this asymmetric relationship. They amass huge troves of data, freely, from consumers. Spurred by success stories of utilizing all this data, such as Harrah's testimonial in "Diamonds in the Data Mine," companies have placed a high value on gaining all of this personal information. However, data security expert Bruce Schneier disagrees with this valuation.

As Schneier puts it, this notion of the importance of "big data" is really only based on the idea that these "large databases of seemingly random data about people" will someday be valuable.[14] It is cheap to acquire and collect data from consumers, and it is cheap to save this data -- "save as much as possible, and save it all forever. [...] Until recently, there was absolutely no downside to saving everything."[15] What companies have failed to realize, until the data breaches of late, is that hoarding all of this data can come at a negative cost. Denoting "data as a toxic asset," Schneier shows how all of this data is

[11]Lohr, S. (n.d.). You Want My Personal Data? Reward Me for It. Retrieved from http://www.nytimes.com/2010/07/18/business/18unboxed.html
[12] Ibid.
[13]Acquisti, Alessandro.The Economics of Privacy: Theoretical and Empirical Aspects, 2013
[14]Schneier on Security. (n.d.). Retrieved May 2016, from https://www.schneier.com/blog/archives/2016/03/data_is_a_toxic.html
[15]Ibid.

“dangerous” because it’s “highly personal” and “intimate.” This makes it an attractive good for attackers to exploit and therefore also makes it vulnerable and very difficult to protect.

**A Proposal for the Role of Data & Privacy in a Company**

Both Schneier and Acquisti question the need for such data to be collected. Schneier implores us to treat data as toxic, to collect, handle, and store it with care and for precise needs, rather than gorging at the buffet. Acquisti posits that the benefits of applications could be “enjoyed by consumers without their having to disclose personally identified data: the adoption of privacy enhancing technologies can make it possible to satisfy both the need for privacy and the need for sharing data, by selectively protecting and disclosing pieces of personal information.”

It is this question that supports the framework of Portal. With Portal, we have specifically chosen not to store any aggregated user data on our servers. Using blockchain storage technology, users are able to upload and store their data themselves, but without giving us access to any of it. We have realized, along the lines of Acquisti’s words, that the explicit user experience for this app -- to aggregate and display a user’s data in an enjoyable, pleasing way -- does not need us to see or collect any of that personal information. Saving this type of intimate data would be a mistake for us; it would be a toxic asset and vulnerable.

In the next few sections of this paper, we will continue to look at the state of storage in the web services industry and explore our business strategy of creating a successful, viable consumer application without storing any user data.

**Economics of the Big Data Industry**

The rise of Big Data in the recent past is a result of increased digitization of businesses and lives, the decreasing cost of storing data, and the increasing availability of third party data. It does not mean there was no data or data analytics before that. Companies have long been using reports, spreadsheets, and computer-aided tools to perform analytics to drive business intelligence. With Big Data, companies can do the same analytics faster and at larger scale. When competitors can deliver value faster and move toward cross-channel, real time customer engagement, companies worry they will be left behind in the race to acquire, develop, and retain customers. This has led to Data

as a Service (DaaS) and Big Data as a Service (BDaaS) as separate industries that help boost revenues and guard against competition.[16]

DaaS is a process that leverages the modern data ecosystem and real-time data analytics to create a customized, "always-on" dataset. DaaS combines a company's first-party CRM (customer relationship management) data with real-time triggers and Hard-to-Find-Data (HTFD) sources to deliver better targeting and a stream of in-market consumers. This is the era where Everything is offered as a Service (or XaaS) and IT spending that is cloud-based is estimated to increase from about 15% today to 35% by 2021. Given that the Big Data market will be worth $88 billion by that point, we can forecast BDaaS market to be about $30 billion.[17]

**What does the rise of BDaaS mean to everyone in the ecosystem?**

- *For companies*: They need not set up infrastructure or hire data experts or own exclusive rights to data to derive actionable insights. BDaas services also help in Decision Management and reduce the insight-to-action gap for companies.
- *For data brokers*: Sell the "Derived Data" or insights from "Service Data" or "Behavioral Data" to interested parties.
- *For the customers*: Discover 'Services' of better quality and relevance.

The data marketplace, as we see, is a great business model for *companies*. But the most important question to ask here is: "Is this ecosystem fair to all *actors*?" It helps to understand how consumers perceive the 'services' they receive and if they understand the exchange they are making in order to receive them. Also, the value of these 'services' varies across different consumers. Are these 'services' worth sharing all of the data and the personal information for an individual? Is the individual user entirely aware of the transactions with which she has engaged? Who decides what happens with the information or the data she creates? Should a user have control over whether or not her data can be collected? And if the user agrees to her data being collected by companies to "analyze," should she receive something she values in return?

[16] Datafloq - Connecting Data and People. (n.d.). Retrieved May 2016, from https://datafloq.com/read/real-life-examples-of-companies-using-DaaS/961
[17] Big Data-as-a-Service is Next Big Thing. (2015). Retrieved May 2016, from http://www.forbes.com/sites/bernardmarr/2015/04/27/big-data-as-a-service-is-next-big-thing/

This is an issue that is already discussed in circles concerned about user tracking and privacy. Fitbit is an example of a company that provides “premium services” for a subscription fee based on the data it collects from users.[18] These services primarily involve analytics on user data. What if this scenario was reversed? What if the customer had control over what she shares to the ecosystem and also gets paid if she allows for her data to be used?

In most of the current products that collect or rely on data, the user's data is collected by default and the companies centrally own all the information and exercise rights over it. What if there was a system where data storage is decentralized and is not owned by a single entity which could monopolize it? This is not a new idea; P2P storage systems have been tried before and the idea has been around for a while.

The term Peer-to-Peer refers to a class of systems and applications that employ distributed resources to perform a function in a decentralized manner. Each node potentially has the same responsibility. Shared items can include CPU cycles (SETI@Home) or Storage Space (Napster, Gnutella, OceanStore).

Symmetry of nodes, reliable decentralization that is resilient to faults and has higher system availability is a major design issue in P2P systems. Getting a global view of the system state is difficult. There are several open problems in P2P file systems pertaining to untrusted peers, distributed DDOS, ever-changing network topology, malicious users creating artificial churns, and P2P systems not being able to provide performance guarantees.[19] There has not been a central authority maintaining the P2P system or an appropriate protocol to either ensure reliability, redundancy, and provide incentives for being a 'responsible peer' or ‘punish’ a 'malicious peer'. The following are few reasons why P2P networks have not been used as an actual system of deployment for production software.

[18]How Fitbit Makes Money? - Revenues & Profits. (2015). Retrieved May 2016, from http://revenuesandprofits.com/how-fitbit-makes-money/

[19] Hasan, R., Anwar, Z., Yurcik, W., Brumbaugh, L., & Campbell, R. (2005). A survey of peer-to-peer storage techniques for distributed file systems. *International Conference on Information Technology: Coding and Computing (ITCC'05) - Volume II*. doi:10.1109/itcc.2005.42

## Current Trends Supporting Privacy-First Applications

The current ecosystem has changed a lot in support of more decentralized data storage. The trends in technology, policy, and social awareness call for better data control and ownership.

### Evolution of Blockchain Technology

A blockchain, is a distributed database that maintains a continuously-growing list of data records hardened against tampering and revision. Blockchain, the technology underlying Bitcoin, is garnering more attention from financial institutions, governments and other businesses  who view it as a suitable technology to automate accounting and transactions through its open ledger system in a safe and secure manner. DARPA relies on blockchain as the platform separates the message creation, from the transfer of the message within a secure courier to the reception and decryption of the message. [20]

### Changes in law and policy

A recent Forrester study says with consumers prioritizing privacy over convenience, concerns about online privacy will prompt regulators to crack down on companies that are a greater risk of attacks by hackers. In early 2016, EU is set to approve the new General Data Protection Regulation (GDPR) which would make companies in violation of the law liable for up to 5 percent of global revenues. In the U.S., companies can expect severe financial penalties as regulators also flex their muscles, a trend already underway. There have been several privacy violation related cases where the companies were charged huge penalties which indicates the regulators are taking the issue more seriously.[21] In terms of laws in the U.S., California already has passed two regulations earlier this year where authorities require a warrant or court order for access to an individual's electronic communications, and will toughen and expand data-breach notification regulations.

[20] Not Just Bitcoin: Why The Blockchain Is A Seductive Technology To Many Industries. (n.d.). Retrieved May 2016, from http://www.npr.org/sections/alltechconsidered/2016/05/04/476597296/not-just-bitcoin-why-blockchain-is-a-seductive-technology-to-many-industries

[21] Privacy will hit tipping point next year: Study. (2015). Retrieved May 2016, from http://www.cnbc.com/2015/11/09/privacy-will-hit-tipping-point-in-2016.html

**Increase in public awareness**

With more hacks and data breaches reaching the mainstream media, the public is increasingly aware of the potential data risks. As traditional banner and display ads continue to decline in performance, advertisers are turning to new methods to increase ad relevancy, such as behavioral targeted ads. These ads are based much more on specific user actions, such as site history, browsing history, and buying behavior. Coupling together these known behaviors, advertisers can deliver more relevant ads. Many people are seeing a tremendous increase in the appearance of behavioral ads -- some with ads and recommendations crossing the border from relevant to intrusive. This, in turn, is giving rise to the concern that everything they do online is being tracked. Consumers are aware of their data's value and have shown a willingness to trade it for better service and better prices. Subscription content and user-based sites are growing dramatically as advertisers trade content for context. But it's even more critical that online data be protected. Unfortunately, many groups can't seem to differentiate between privacy and security -- which leads us to the debate today over regulation and rights. But when it comes to behaviors and advertising, let people decide how much they want to share and what they are willing to trade for their information.[22]

[22]Privacy vs. Security: What's Behind Ad-Targeting Concern? (n.d.). Retrieved May 2016, from http://www.cmo.com/articles/2012/2/22/privacy-vs-security-whats-behind-ad-targeting-concern.html

## Market, Technology & Political Trends

### Macro Economic Trends

Major macro economic trends support the evolution of products and services that will place a higher value on users' data privacy and control, actively reduce the amount and type of data collected, and highlight greater transparency and user choice on data collection. These trends include:

*Increased Occurrence of Data Breaches Will Lead to the Need for Better Data Security*

In 2014-2015, there were over 780 high-level data breaches that affected over 212 million households, leaking intimate and financial personal information.[23] [24] Experian, the credit reporting agency, stated in its 2015 Data Breach Industry Forecast report that the "risk of experiencing a data breach is higher than ever with almost half of organizations suffering at least one security incident in the last 12 months."[25]

Consumers have responded to these breaches. According to the Pew Research Center in 2013, 86% of internet users have taken steps to increase their privacy and anonymity online and a number of respondents reported experiencing some level of a data breach. 21% said their email or social media accounts were compromised, while 11% had crucial personal information, such as their social security number and banking details, hacked. It is unsurprising, with the rise in reports of data insecurity, that consumers are trusting companies less and less with their personal information, particularly companies working in the online space. A study conducted by the Harvard Business Review in May 2015 showed that for a survey base of 900 respondents, the majority ranked Internet giants like Google and Yahoo, technology firms, and social media companies last in terms of reliability with protecting user data. The study found that consumers "clearly worry about their personal data -- even if they don't know exactly what they're revealing."

This has led consumers to slowly demand a higher standard in data privacy and control, a trend that we believe will continue to increase as industry leaders respond. Several high-profile companies such as Facebook, Google, Whatsapp, and Apple have

[23] Data protection in United States: Overview. (n.d.). Retrieved May 2016, from http://us.practicallaw.com/6-502-0467
[24] 2015 Data Breaches | ITRC Surveys & Studies | ID Theft Blog. (2015). Retrieved May 2016, from http://www.idtheftcenter.org/ITRC-Surveys-Studies/2015databreaches.html
[25] Data Breach Industry Forecast. (2015). Retrieved May 2016, from https://www.experian.com/assets/data-breach/white-papers/2015-industry-forecast-experian.pdf

recently announced their intention to meet consumers' demand for greater privacy and security. Facebook has "grasped that [consumer] trust is no longer just a 'nice to have'", it is essential.[26] The company rolled out Privacy Basics in January 2015, to help consumers better and more easily understand how their online profiles appear to the public and how their content can be customized. Google has done similarly in July 2015 with their Privacy Checkup, a "user friendly means of taking control of account privacy."[27] In April 2016, Whatsapp, the messaging service used by millions around the world, enabled end-to-end encryption, by default, on all messages sent through the platform.[28] And in the most high-profile standoff of these four companies, Apple released a letter to all of its customers, promising to uphold data security and privacy to the highest level, even in the face of demands from the U.S. government.

*Increased Demand for More Secure Cloud Storage*

While Whatsapp has released end-to-end encryption for its messages, this practice has not yet extended fully to other communication and cloud storage services. Most cloud services do not offer this level of encryption. Many users' files continue to remain on these companies' servers, not fully protected. One leader in the secure cloud storage space is Box, which is largely acknowledged to be the storage provider of choice for government entities and many private companies due to its emphasis on client- and server-side encryption. In early 2015, it finally released Box Enterprise Key Management, a feature that allows clients to be the only controllers with the keys to their files stored with Box. The feature was developed in part to meet the strict needs of clients in highly-regulated industries, such as banking, government, and law.[29] In contrast with Box, Dropbox, a close competitor in the industry, does not yet provide encryption on the client side, though it allows integration with third-party services that can do so.[30]

[26]Customer Data: Designing for Transparency and Trust. (2015). Retrieved May 06, 2016, from https://hbr.org/2015/05/customer-data-designing-for-transparency-and-trust
[27] Take the time to walk through the new Google Privacy Checkup, 2015 http://www.techrepublic.com/article/take-the-time-to-walk-through-the-new-google-privacy-checkup/
[28]Forget Apple vs. the FBI: WhatsApp Just Switched on Encryption for a Billion People. (n.d.). Retrieved May 06, 2016, from http://www.wired.com/2016/04/forget-apple-vs-fbi-whatsapp-just-switched-encryption-billion-people/
[29](n.d.). Retrieved May 06, 2016, from http://www.forbes.com/sites/alexkonrad/2015/02/10/box-unveils-key-management-to-woo-big-banks/#772e8af86929
[30]Can I specify my own private key for my Dropbox? (Dropbox Help Center). (n.d.). Retrieved May 06, 2016, from https://www.dropbox.com/en/help/28

We expect that as higher standards for encrypted messaging trickled down to consumers through Whatsapp, the standards for greater data security and privacy will also trickle down from highly-regulated clients to the mass consumer.

*Increased B2B Adoption of Blockchain Services*

The cloud computing industry is beginning to adopt blockchain technology, providing developers with easy-to-access and easy-to-deploy blockchain-based alternatives to less-secure services. Microsoft Azure and Amazon Web Services (AWS), both heavyweight cloud computing platforms, recently announced their adoption and integration of BaaS or Blockchain-as-a-Service. Both companies hold their own portfolio of blockchain storage startup partners, and the Azure ecosystem notably includes Storj, the platform that Portal uses to decentralize and distribute user data, as of April 2016. By backing these startups and forming such partnerships, Microsoft Azure and AWS will ultimately help to drive the momentum of the blockchain-based development.

As recent as May 4, 2016, Microsoft Azure announced that it is joining the Chamber of Digital Commerce. This is significant as the Digital Chamber is a politically active trade association representing a mission to "promote the acceptance and use of digital assets and blockchain-based technologies."[31] As large technology firms, such as Microsoft, begin to work more closely with policymakers and regulators to integrate distributed, decentralized storage technology like blockchain, we expect to see acceptance and adoption of blockchain rise.

**Technology Trends**

In addition to the macro economic trends supporting the adoption and development of blockchain and other privacy-centered services, there are several technological trends that will promote the growth of apps like Portal. These trends include:

*Increased acceptance and usage of cryptocurrency*

Cryptocurrency, most popularly known in the form of Bitcoin, is on the rise. According to a report released by PricewaterhouseCooper, fewer than 10%

[31] ABOUT THE CHAMBER. (n.d.). Retrieved May 06, 2016, from http://www.digitalchamber.org/about.html

of respondents surveyed by the authors labeled themselves as extremely familiar with cryptocurrency.[32] However, despite the low recognition rate and while "cryptocurrency growth over the next year is expected to be solid but not spectacular," the authors state that the majority of their respondents are "bullish" over the impact of cryptocurrencies on banking and retail. Many of these respondents believe that cryptocurrencies will "redefine banking as we know it."[33]

It is also worth noting that cryptocurrencies, or "virtual currencies," are recognized as taxable by the U.S. Internal Revenue Service as of March 2014, indicating that the federal government recognizes the growing impact and potential. As cryptocurrencies continue to gain mainstream acceptance, we expect that attention from regulatory and financial bodies will increase as well.

*Kryder's Law & Nielsen's Law Applied to B2B*

In 2005, Mark Kryder of Carnegie Mellon University predicted that the "capacity of computer semiconductors will double every 18 months" as the physical size of a hard drive shrinks and its memory innards become more dense. Known as Kryder's Law, the accompanying thought is that as storage becomes more abundant, the cost of storage will go down (disk prices expected to drop 40% per year over a 30-year history). However, as recent research shows, Kryder's Law has not fully come to fruition.

David Rosenthal of Stanford University Libraries and several other authors assert that cloud storage, in particular, has not followed this expected pattern of price dropping. They posit that cloud storage providers, such as Amazon's Simple Storage Service (S3), recognize the high lock-in price consumers encounter when storing their files on Amazon's servers. If a user felt that the price paid for 1 TB of storage was too high, he would need to move his data to a cheaper location. However, doing so would not only cost a high amount of time, but it would also incur its own cost in bandwidth charges. The authors state that even if a competitor were to enter the market with lower prices than Amazon S3, Amazon would simply undercut these prices, which are already kept artificially high.

Nielsen's Law is similar to Kryder's Law, except it deals with the growth of bandwidth over time. According to Nielsen's Law, "users' bandwidth grows by 50% per

[32]Money is no Object: Understanding the Evolving Cryptocurrency, 2015, http://www.pwc.com/us/en/financial-services/publications/assets/pwc-cryptocurrency-evolution.pdf
[33] Ibid.

year (10% less than Moore's Law for computer speed.")[34] The law is affected by three factors: the conservative nature of telecoms companies which must invest in physical equipment and installation to improve bandwidth speeds; the reluctance of users to spend on bandwidth; and the broadening of the user base as more people get online.

Considering both laws, we believe this means that cloud storage prices will slow over time despite increased density in physical storage technology; but they will not reduce as quickly as originally predicted. In contrast, users' bandwidth speeds will continue to increase, though slowly due to the limitations of installations by telecoms companies. In total, this outcome supports the evolution of blockchain or peer-to-peer based storage systems and apps, such as Portal. If cloud storage prices do not decrease, consumers will look for alternatives, such as blockchain. The bandwidth required to participate in blockchain storage, especially the continuous uploading and downloading of blocks of data within the ecosystem, will increase, meeting the supply as according to Nielsen's Law.

**Political Trends**

In late May 2016, the European Union will implement the General Data Protection Regulation (GDPR), a strict set of guidelines directly applicable to all members of the EU. The GDPR strongly protects consumers' "fundamental right" to data protection and aims to "give citizens back control over their personal data."[35] Notably, the GDPR sets boundaries for data collection, specifying that personal data must be collected for "explicit and legitimate purposes and not further processed in a way incompatible with those purposes." The legislation also defines the standards for storage, processing, accuracy, and accountability of such data.

The explicitness of the GDPR is of interest as it demonstrates a shift in the worldview of data collection and storage. It recognizes the need to protect consumers, and by this, also acknowledges the current and common wrongdoings in handling users' data. The graphic following this section depicts how data protection legislation has increased as the pervasiveness of information technology has risen.

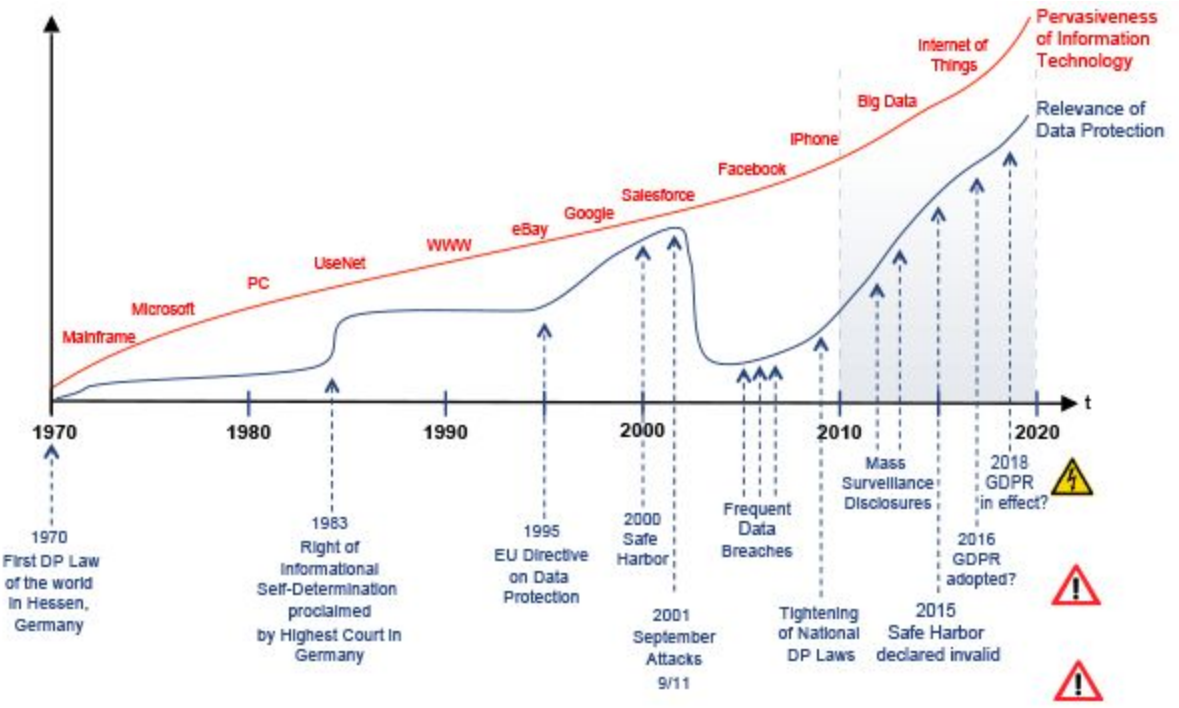
---

[34] "Nielsen's Law of Internet Bandwidth." *Nielsen's Law of Internet Bandwidth*. N.p., n.d. Web. 06 May 2016.

[35] A brief history of the General Data Protection Regulation. (n.d.). Retrieved May, 2016, from https://iapp.org/resources/article/a-brief-history-of-the-general-data-protection-regulation/

In the United States, our data protection laws comprise a "patchwork system of federal and state laws, and regulations that can sometimes overlap, dovetail and contradict one another." Centralized and cohesive data protection legislation such as the GDPR does not yet exist, though trends show that focus on such lawmaking is gradually becoming more commonplace, particularly in response to high-profile data breaches affecting millions of citizens.

We expect that this focus on data protection legislation will continue to increase in the U.S., helping to fuel support for data privacy and data control-centered apps such as Portal.




*Chronicling the development of the GDPR*[36]

**Micro Economic Trends**

*Focus on Storj, Provider of Blockchain-based Cloud Storage*

Portal is built on a blockchain-based cloud storage framework called Storj. Storj is one of several such frameworks in the blockchain space. Cryptocurrency, specifically Bitcoin, embraces the philosophy that such digital currency can economically "level the

[36] Protection of personal data. (n.d.). Retrieved May, 2016, from http://ec.europa.eu/justice/data-protection/

playing field." Bitcoin is mined and shared directly between peers. The system does not involve "middlemen" such as the Federal Reserve or other "big banks;" it is a currency that can be earned by anyone, rich or poor, or purchased outright with physical currency. It also affords its holder direct control over its movement. Holding large amounts of Bitcoin does not give a person greater status in the system.

Storj's system fundamentally involves free market economics. Under this system, users become "farmers" as they farm out extra storage space for other system users to keep and transfer their files. The currency earned under this arrangement is StorjCoin X. Under a pure free market scenario, farmers can negotiate directly with other users on how much they charge to store files, depending on features such as file size, duration and uptime required. Farmers can negotiate file-sharing contracts depending on the demand/supply equilibrium for a certain consumer's requirements.

A tussle with this scenario and an easier, status-based system exists, however. Newcomers to the Storj service must continuously navigate the difficulties of always needing to negotiate with farmers for their prices. Storj is currently navigating this tricky spectrum of solutions, where the underlying contract-level negotiations are abstracted to make file storage seamless for consumers. One of the solutions Storj attempted to explore at one point was to create a status-based model that could feasibly be gamed. Status could be bought by users in the form of exchanging physical currency (such as dollars) for StorjCoin X. Higher status would afford greater preferential treatment for Storj's services, such as uploading and downloading files. This fundamentally would go against Bitcoin's philosophy as the playing field would no longer be level; users would be able to outright purchase better services rather than earning them through farming and negotiating.

### *Projections of Storj Growth*

Storj does not use Bitcoin for their microtransactions because the cost of a proof/contract is only 0.0047 Bitcoin for a typical 2TB farm size. This means users have to establish a longer trusted connection (typically 24 hours compared to five minutes) if a higher valued currency like Bitcoin is to be used.[37]

[37] S. (n.d.). Why SJCX? Why not Bitcoin? Retrieved May, 2016, from http://blog.storj.io/post/111594471093/why-sjcx-why-not-bitcoin


*Graph Depicting Cryptocurrency Market Capitalization*[38]

The chart above depicts SJCX value trends in the past 2 years, The rise in the value seen around February 2016 was when the FBI demanded Apple to open up the back door to their iPhones. This is an indicator that the community views Storj or blockchain technology as a secure way of storing information where users have the control over their data.

With more applications developed on the platform, the value and the popularity of the network increases. This increases the value of SJCX and eventually reduces the storage cost on the platform for the end user. Storj has a cap of 500 million SJCX that can ever exist and currently there are 50 million coins in circulation. Because of the finite nature of the resource, it is inherently deflationary.

[38] Storjcoin X (SJCX). (n.d.). Retrieved May, 2016, from http://coinmarketcap.com/assets/storjcoin-x/

## Portal's Business Model

In order to uphold our mission of protecting user data and promoting data privacy and greater control, Portal will not collect or store any intimate, personal information. The explicit purpose of the Portal app is to aggregate external network data streams and to display back this data through an attractive, enjoyable-to-use interface. In order to accomplish this purpose, Portal should not need to store any copy of a users' data streams. To do this would be, as Schneier deems it, to take on the unnecessary risk of storing a toxic, vulnerable asset.

Despite having a lack of data assets, Portal will deliver value. Its primary value will be the higher level of privacy and security that users will be able to derive from the decentralized, distributed, blockchain-based model. The Storj network also provides users with higher likelihood for continuous uptime using industry standard data replication techniques such as erasure encoding. We believe that our decision to forgo storing user data will be a competitive advantage as it can better guarantee our users' privacy and data security. The lack of such toxic assets will also decrease the potential for costly data breaches and leakages. Through such a model, we believe we can sufficiently meet the goals of our intended user experience without exposing our business model to unnecessary breach costs.

The obvious debate that arises from our decision not to store user data is if we are forgoing the benefits of "big data." We, in turn, question the relevancy of this for Portal specifically. For Portal, we place a high value on data privacy and security. We believe the value of these attributes is higher than the addition of a new targeted feature created through machine learning.

We recognize, however, that some users may want to avail themselves of the benefits gained through machine learning and that we should not deter users from accomplishing this. If users want a more customized experience, they can opt-in through a transparent and clearly designed choice mechanism. By default, users are opted-out of data sharing and thus are not required to give up their personal information. This is an issue of user trust gained through transparency and choice, and it forms the backbone of our core design principles and data strategy.

Portal and Storj also both depend on the participation of a mass user group in order to keep costs low and efficiency of file transfer high. As with any peer-to-peer

system, Portal is therefore highly reliant on the users involved and must place high value on engaging and maintaining the blockchain user base.

**Pricing**

Portal's business model will integrate directly with Storj's business model. We see Portal as an application interface layer on top of the Storj framework, and it will be necessary to bundle the cost of Storj's storage services with the costs needed to support our app. Bundling will be crucial to deriving the ideal cost for Portal as we believe people will not want to deal with paying for both services separately.

Storj's current proposed cost structure to users is $0.015 per GB per month of storage and $0.05 per GB of bandwidth at time of launch. We estimate an average user on our platform will be onboarded with 30 GB of historical data, and will add 0.5 GB of new data per month. Factoring in Storj costs per user for the above data needs, and also operating costs per user for our infrastructure, we have come up with the following cost plan to break even. Our assumptions also consider the cost of "sharing files" since they add an overhead of bandwidth costs. On average, a user will consume three "memory portals" per day, which amounts to 1.35 GB of bandwidth consumed (also accounting for in the "Bandwidth Out" cost).

| Operating Costs | | | Storage Costs | | | | |
|---|---|---|---|---|---|---|---|
| Infrastructure | Bwidth | Total Data | New Data | Storage cost | Bandwidth In cost | Bandwidth Out cost | Total Cost |
| 0.166666667 | 1.35 | 30 | 0 | 0.45 | 1.5 | 0.0675 | 2.0175 |
| 0.0025 | 1.35 | 30 | 0.5 | 0.4575 | 1.525 | 0.0675 | 2.05 |
| 0.0025 | 1.35 | 30.5 | 0.5 | 0.465 | 0.025 | 0.0675 | 0.5575 |
| 0.0025 | 1.35 | 31 | 0.5 | 0.4725 | 0.025 | 0.0675 | 0.565 |
| 0.0025 | 1.35 | 31.5 | 0.5 | 0.48 | 0.025 | 0.0675 | 0.5725 |
| 0.0025 | 1.35 | 32 | 0.5 | 0.4875 | 0.025 | 0.0675 | 0.58 |
| 0.0025 | 1.35 | 32.5 | 0.5 | 0.495 | 0.025 | 0.0675 | 0.5875 |
| 0.0025 | 1.35 | 33 | 0.5 | 0.5025 | 0.025 | 0.0675 | 0.595 |
| 0.0025 | 1.35 | 33.5 | 0.5 | 0.51 | 0.025 | 0.0675 | 0.6025 |
| 0.0025 | 1.35 | 34 | 0.5 | 0.5175 | 0.025 | 0.0675 | 0.61 |
| 0.0025 | 1.35 | 34.5 | 0.5 | 0.525 | 0.025 | 0.0675 | 0.6175 |
| 0.0025 | 1.35 | 35 | 0.5 | 0.5325 | 0.025 | 0.0675 | 0.625 |
| 0.0275 | | | | | | | 7.9625 |

| Storj cost rate | |
|---|---|
| Storage | Bwidth |
| 0.015 | 0.05 |

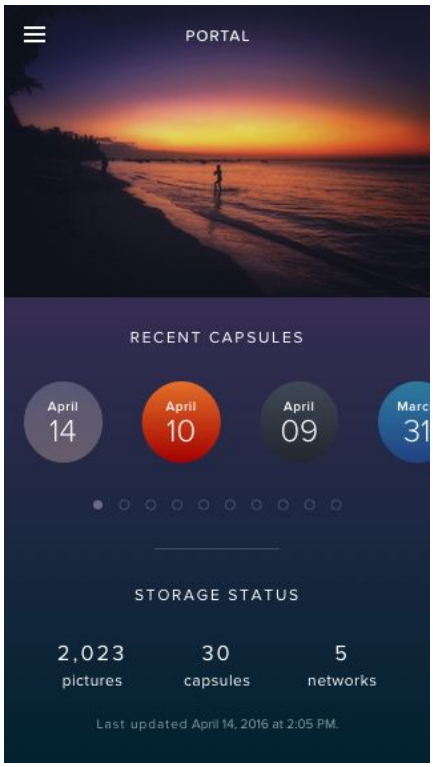
| Per User Annual Cost | $7.99 |
|---|---|

*Estimated User's Cost for Portal over a 12-Month Period*

*Assumptions: 30 GB monthly usage with 0.5 GB monthly addition of new storage*

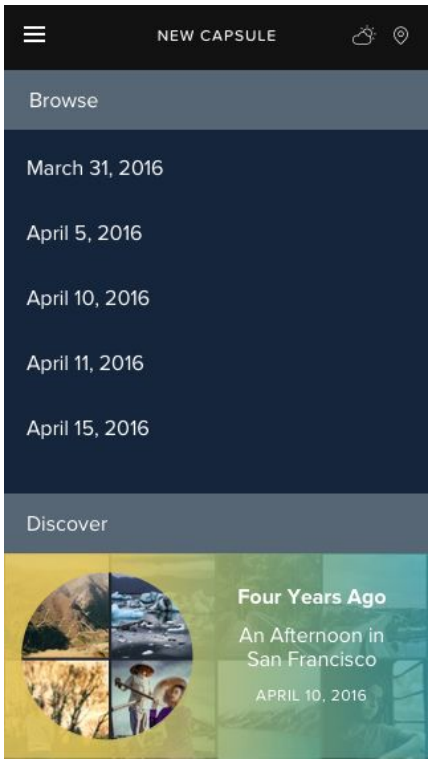
In the future, as Portal, Storj, and blockchain-based storage gains mainstream acceptance, we think it will be possible to introduce tiered pricing or premium features. We intend to carry pricing experiments and sensitivity analyses to help determine the right pricing tiers that can help sustain our operations and still remain economically competitive with competition.

## Overview of Portal's Design



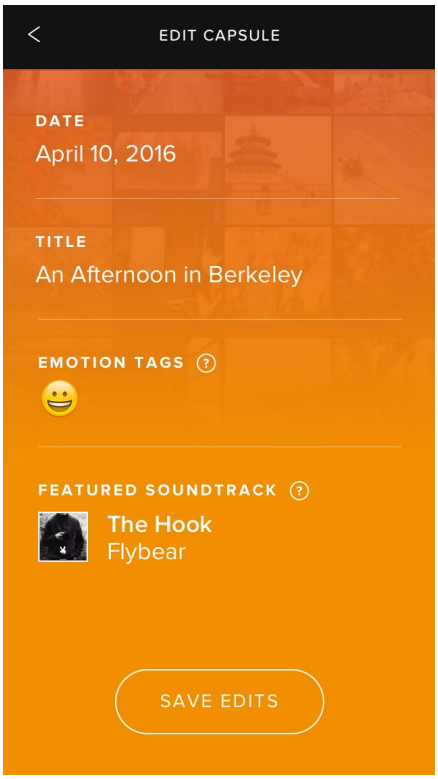

**Home Screen**
On this screen, a user will see a horizontal list of the "memory portals" that have been aggregated and combined based on the date entered. The portals are colored based on a combination of the weather (if available) and the timestamps of the photos taken (if available).

A status section at the bottom allows the user to gain a quick understanding of their usage of Portal.

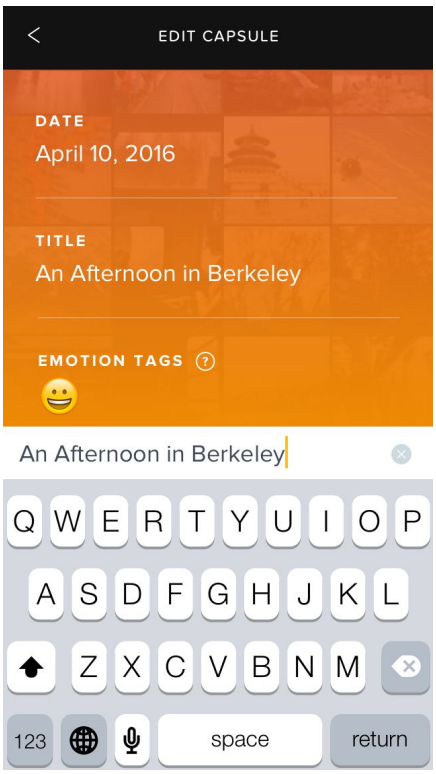



**Browse Portals**
Here, a user can browse through a list of portals available or swipe through a list of recommended portals to view. The portals listed in the Discovery section can be triggered by time, location, and weather.

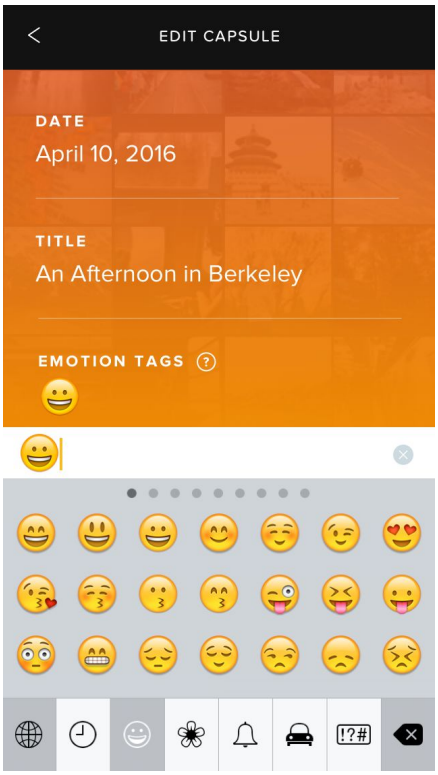

**Individual Portal**
This is the main view of an individual portal. The portal's background is a collage of the pictures available for that date.




**Portal Editing**
A user can edit a portal's metadata, fine-tuning the details.




**Portal Editing - Text**
An example of what the user view will look like when editing a text field on the edit screen.




**Portal Editing - Emotions**
We allow users to tag portals based on emotions so that they can later filter through the available list in this way.


**Portal Photos**
Once a portal is opened, a Masonry grid view of the pictures associated with the portal will appear. On this screen, a playlist bar is shown at the bottom (if music was recorded for that date), allowing the user to replay the songs she was listening to on that date. The user can also quickly access the weather and locations recorded for that date.

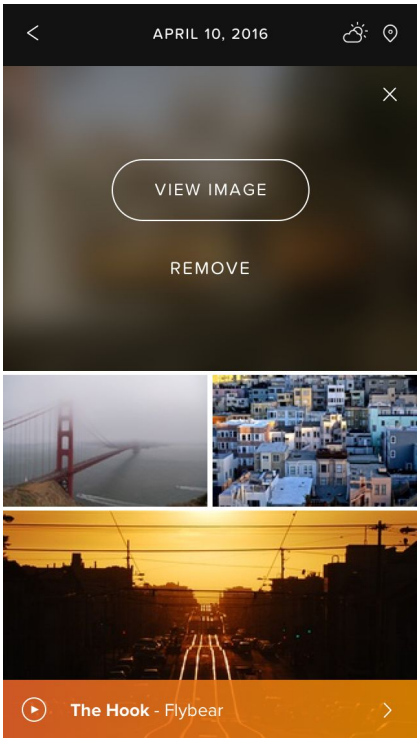



**Image Options**
Sometimes images may be shown that may be sensitive; a user may not want to see this image in the future. The user can quickly adjust these settings by tapping on the image and removing it from the portal.

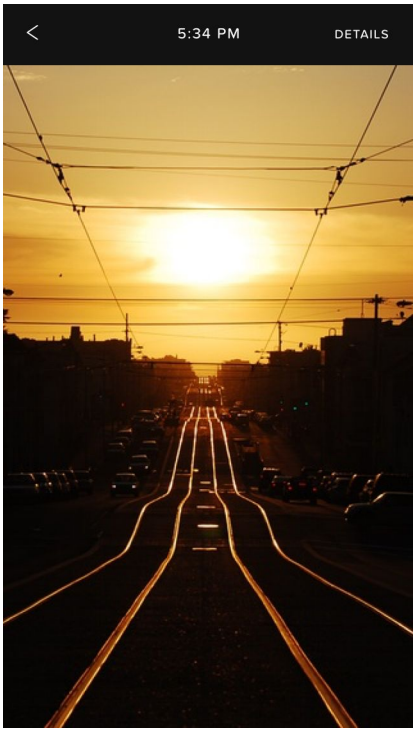



**Image View**
To view an image in a larger size, the user can open the image and see it full-screen.

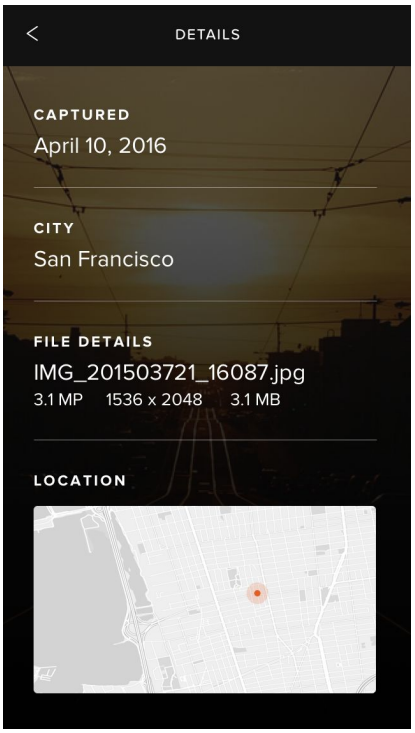
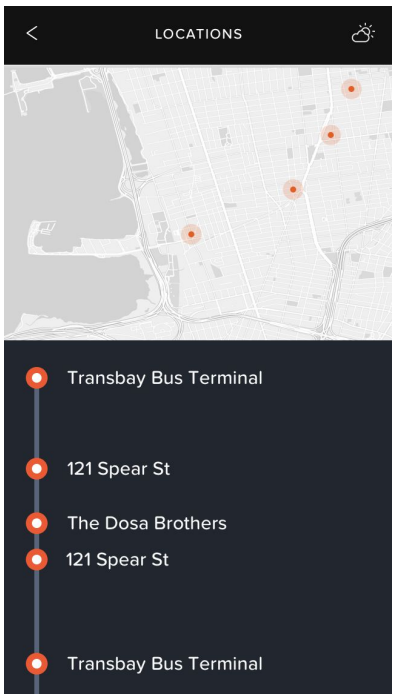


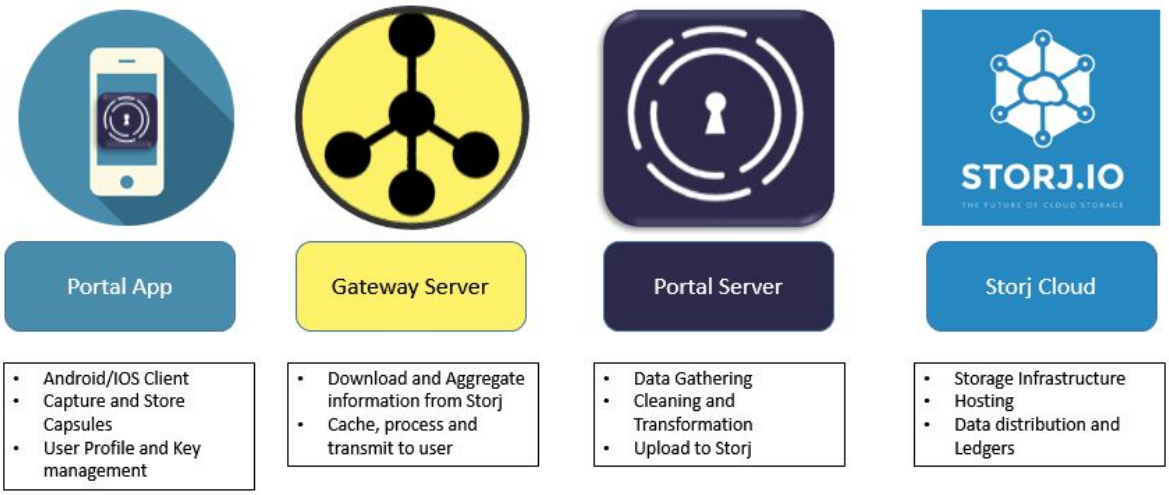

**Image Details**
If the user wants to learn more details about the image, such as when and where it was captured, the user can open and view the metadata.




**Location Details**
The user can access a list of the locations visited during a portal's date.


**Weather Details**
To add even more context to the portal, a user can tap the weather icon (shown if weather data was recorded for that date) to see what the environment was like on that day.

## Overview of Portal's Technology



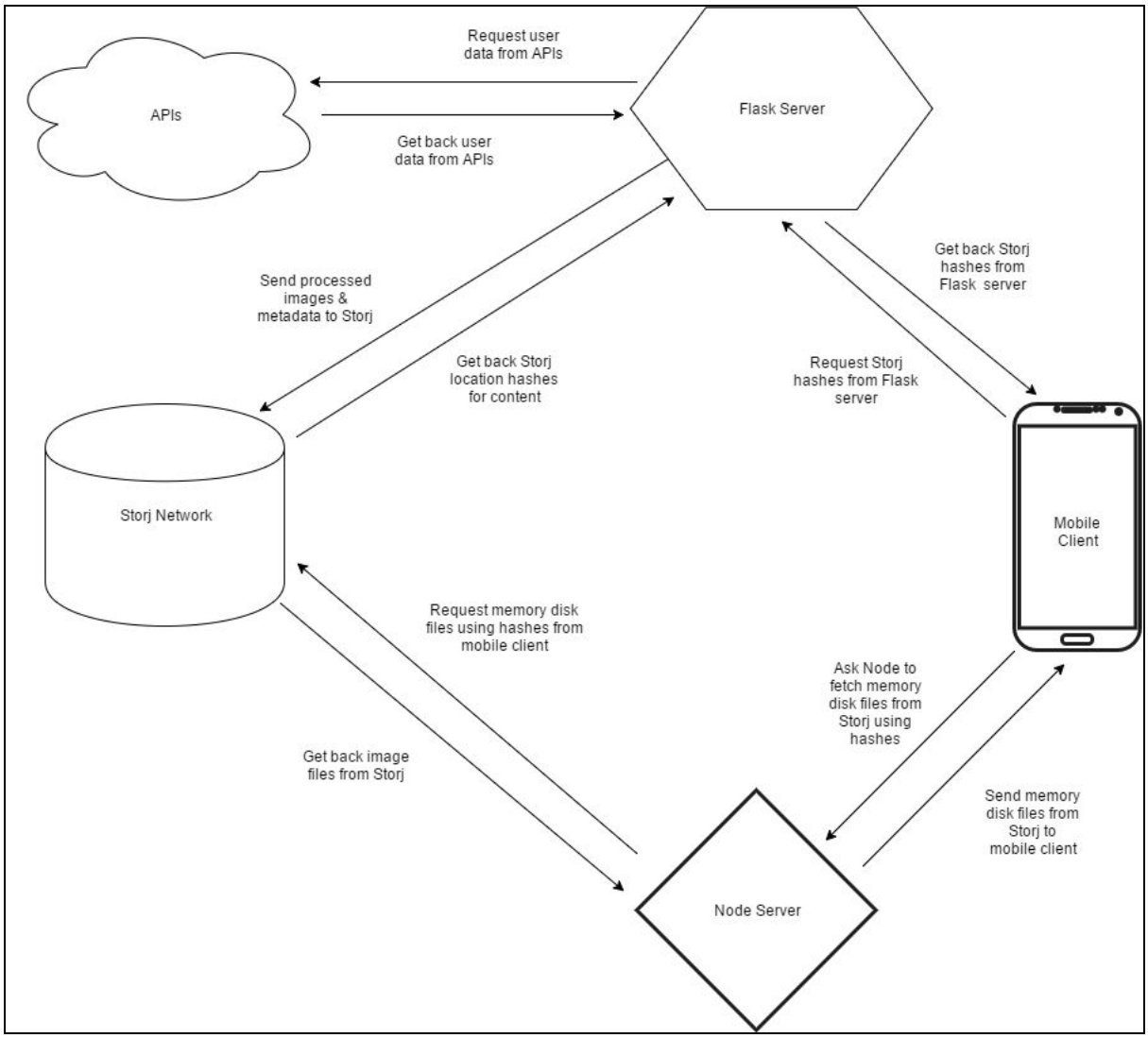

There are four major components in our technology stack, with a future vision of removing the gateway server. The front-end consists of the Portal mobile app and a gateway server. The Portal mobile app and mobile gateway server collect, download, and display memory capsules to the user. The Back end consists of the Portal server - a web application server, which is responsible for gathering a user's social media, creating memory capsules, and uploading them to the  Storj network.

Technologies Utilized:
Ionic, Cordova, JavaScript, AngularJS, Python, Flask, MongoDB, Node, Express

**Front-End**

Our customer-facing product is a mobile application, with builds for iOS and Android. When a user wants new content from their Storj account, the mobile application first requests hashed references to content on the Storj network by hitting an API endpoint on the Portal server . The Portal  server responds with the bucket ID hash and file location hash on the Storj network. These hashes are stored in a Mongo database on the Portalserver. The mobile client then requests content via a Node gateway server using these returned hashes. The Node gateway server then fetches the content from the Storj network using the hashes. Finally, the files and associated metadata are passed

back to the user's mobile client. This entire process of a user wanting new content to having it downloaded on their phone takes about 5 seconds, on average, including downloading all of the images files.




Our original plan was to include the Storj JavaScript library within the mobile application so a user could download his or her files directly to their mobile device from Storj. However, because the Storj JavaScript library was written for Node and wasn't working within the mobile client, we implemented the separate Node gateway server to interface between the mobile client and Storj. The files are not stored on the Node server. The Node gateway server is a temporary fix until Storj releases a mobile library.

We hope to achieve better performance and security once our mobile clients can download content directly from the Storj network.

We used Ionic as our front end framework. Ionic uses AngularJS, is built on top of the Apache Cordova hybrid mobile application engine, and comes bundled with dozens of pre-made components and plugins. Ionic allowed us to develop a mobile application using our existing Web development skills (HTML, CSS, and JavaScript). There was a slight learning curve with Angular, Cordova, and Ionic, but we managed to leverage many of the powerful features of each. We are using a few key, third-party libraries on the front end, including Masonry to create dynamic photo grids, Swiper for advanced touchscreen interactions, and Leaflet as a mapping engine. A user's downloaded images and metadata are stored in the device's local filesystem.

**Back-End**

The Portal server is the 'heavy-lifter' for data processing on the entire system. It serves three main purposes:

1. Storing user's Portal authentication and OAuth credentials.
2. Aggregating and curating data from multiple APIs and curating them.
3. Creating memory capsules and uploading to the Storj network.

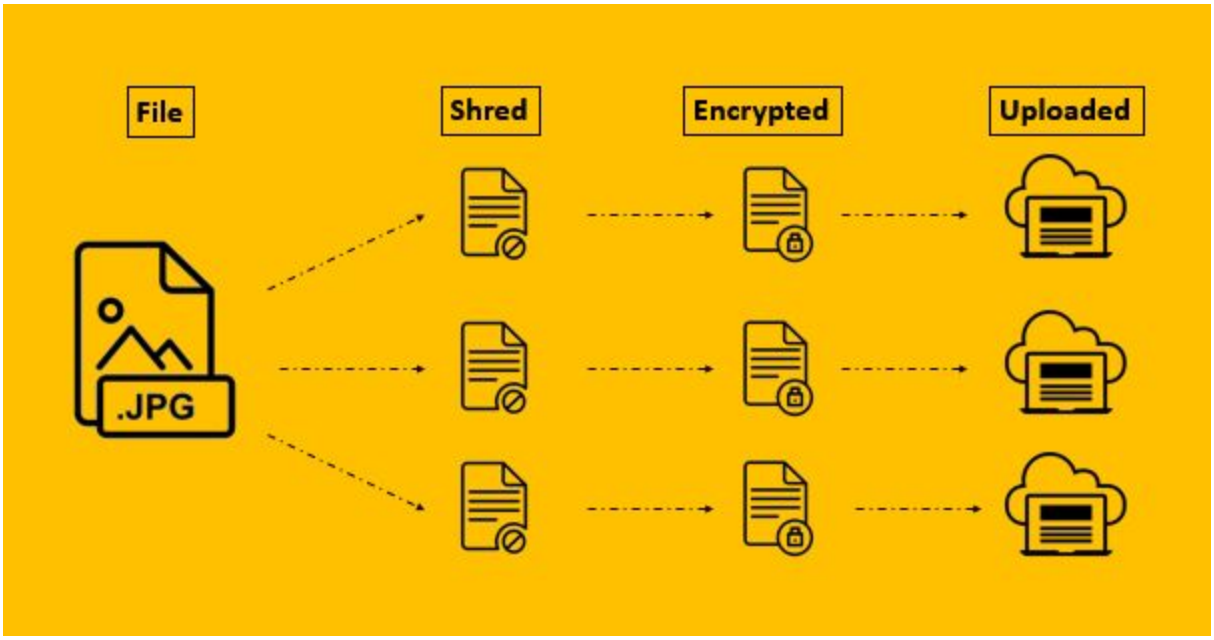
The Portal server is built on the Flask web framework and uses a hybrid of non-relational and relational database servers. The Portal server uses the relational database to store user logins and OAuth accesses to third party APIs such as Instagram, Dropbox, and Spotify. After user onboarding, the Portal server does a one time pull of information from all sources. This is the high level overview of what happens:

1. Pull in files from third party APIs one by one.
2. Assemble images, locations, soundtracks, and weather. Group them for each day.
3. Create a memory capsule by aggregating information for each day.

Currently the memories are grouped by day, but as next steps we want to cluster memories at an even higher granularity than on a day-level. Our algorithm runs through these aggregated data points and creates information for each memory capsule. A complete memory capsule consists of images, corresponding locations, places visited, songs listened, and weather from different APIs, along-with the metadata for each of the individual data points. This information -- raw files, capsule ids, metadata -- is staged on

the Portal server and waits for a process to pick them up and push them to the Storj network.

A daily daemon process runs during low-volume server load and uploads staged data individually to Storj. While uploading, we group each memory capsule and its associated files under a virtual data structure called ‘buckets’ and push them onto the network. During the upload, the server shreds each file, encrypts each shred using the end-user’s public key and stores each encrypted shred on a “farmer’s” machine. A farmer is just an actor on the blockchain network who has shared disk space on Storj.




Once the files are successfully uploaded to Storj, we clean up the staging cache and no files stay with us. We only retain a hash-key that serves as a file pointer to the uploaded “memory portal’s” bucket. This hash-key is later returned to the client where it can authenticate the user’s private key, then download and decrypt the files in each bucket. Only the client has access to user keys and therefore remains the only system where files can be assembled and decrypted.