ACCESS MY INFO

Software Design Document

Electronic version first published at openeffect.ca in 2017 by Open Effect. Open Effect is a Canadian not-for-profit applied research organization focusing on digital privacy and security.

The Citizen Lab at the Munk School of Global Affairs, University of Toronto contributed expertise and equipment in support of this project. Open Effect and the Citizen Lab are collaborative research partners. Together, the two groups engage in research that investigates the intersection of digital technologies and human rights.

This project was supported by a grant from the Canadian Internet Registration Authority (CIRA)'s Community Investment Program.

Document Version: 1.0

ABOUT THIS DOCUMENT

Document Version: 1.0

Access My Info (AMI) is a web application that helps people learn what data companies have collected and stored about them, and how that data is used. *AMI* guides requesters through a simple step-by-step process that creates a letter addressed to a particular company that requests access to a variety of personal information.

This document is primarily written for software professionals and researchers interested in learning how *AMI* was designed both conceptually and technically. It describes *AMI*'s design goals, its model of what an access request looks like, its technical design, its implementation as a web application, and a short guide on how to deploy *AMI* in new jurisdictions. It concludes with a discussion of the applications' impact to-date, as well as some limitations and future work.

ABOUT THE ORGANIZATIONS

OPEN EFFECT

Open Effect is a Canadian not-for-profit that conducts research and advocacy focused on ensuring that people's personal data is treated securely and accountably. It builds interactive advocacy tools to empower individuals to learn about and exercise their rights online. Open Effect's research on the adoption of has been published in peer-reviewed studies.

`https://openeffect.ca`

CITIZEN LAB

The Citizen Lab is an interdisciplinary laboratory based at the Munk School of Global Affairs, University of Toronto, Canada. It focuses on advanced research and development at the intersection of Information and Communication Technologies (ICTs), human rights, and global security.

`https://citizenlab.org`

ABOUT THE AUTHOR

Andrew Hilts
`andrew@openeffect.ca`

Andrew Hilts is a researcher working at the intersection of Internet-connected digital services and human rights, with a focus on privacy. He runs the Canadian not-for-profit Open Effect, through which he collaborates with the Citizen Lab at the Munk School of Global Affairs, at the University of Toronto. He has spoken at events organized by organizations such as the Council on Foreign Relations, Access Now, and USENIX.

Mr. Hilts' work has appeared in numerous academic publications, a wide variety of news media stories, and privacy-enhancing software authored by him has been used by thousands of people. His team's research into the privacy and security of fitness tracking devices led to changes in several companies' applications, and "Access My Info", a tool to help people create legal requests for their personal data from companies, is being localized and released in several different countries around the world.

CONTENTS

LIST OF FIGURES

1 WHAT IS ACCESS MY INFO?

Access My Info (AMI) is a web application that helps people learn what data companies have collected and stored about them, and how that data is used. Originally released in 2014 as a prototype, *AMI* guides requesters through a simple step-by-step process that creates a letter addressed to a particular company that requests access to a variety of personal information. Letters are generated from pre-written templates that cite relevant law and are carefully worded to remind the company of its obligations to respond. Letters can be saved as PDF files to be printed and mailed through the post, or can be generated as prewritten emails for the requester to send directly to a company's privacy contact.

This document describes *Access My Info*'s design goals, its model of what an access request looks like, its technical design, its implementation as a web application, and a short guide on how to deploy *AMI*. It concludes with a discussion of the applications' impact to-date, as well as some limitations and future work.

2 CORE DESIGN PRINCIPLES

The core problem that *Access My Info* addresses is a knowledge imbalance. Individuals know relatively little about what, why, and how their personal information is collected, processed, and shared by companies. Companies, on the other hand, use this personal information while only typically communicating broad and vague statements about the use of that data in privacy policies. *AMI* is intended to help level the playing field by **empowering its users to actively obtain new facts about how a company use their personal information**. What follows is an overview of the major considerations that went into the design of *AMI* so that the application could most effectivelyI w tackle this problem and help people better understand their own personal information.

2.1 ZERO USER DATA COLLECTION BY DEFAULT

We originally designed *AMI* to collect zero information about the requests that it helps to facilitate. We felt that a tool designed to help people understand what personal information companies have about them should not itself collect personal information. *AMI* asks that users input their name and often physical or email addresses in order to generate a personalized request. We designed the system to not transmit this inputted data outside of the user's web browser; everything happens locally on the user's computer. By collecting no information besides basic server logs, *AMI* can furthermore demonstrate that digital services can deliver value to people without having to collect troves of data about them. Another benefit of not collecting user data

is improved security; *AMI* delivers value to individuals without putting the personal data of previous requesters at risk in the event of a data breach. We believe this enhances the credibility and trust that people have in our system.

2.2 INDIVIDUAL AUTONOMY

When companies receive a request for access to personal information generated by *AMI*, our goal is for that request to be viewed as an authentic and legitimate exercise of a legal right by an individual citizen. We want to avoid the situation where companies ignore requests because they are all seemed to have been sent by an automated tool, and not by a different people. Therefore, the concept of individual autonomy is essential to embed in the design of *AMI*. In essence, we want to help citizens themselves send requests, not issue requests on their behalf.

2.3 FLEXIBILITY

AMI was designed to be a general-purpose system for creating legal requests for access to personal information. Our goal is to build flexibility into the system so it can be easily adaptable to new contexts. Some of these adaptability features include ease of translation, easy addition of new companies and industries, and easy inclusion of different sorts of requests, depending on the jurisdiction and other contextual factors.

3 REDESIGN REQUIREMENTS

After the initial launch of *Access My Info* in 2014, we determined that the tool was successful in its basic task of enabling Canadians to generate legal requests to telecommunications providers. However, we identified several areas for improvement that led to new requirements for the next version of the tool. Specifically, due to the zero-knowledge nature of the tool, we did not know how many people used the tool, or which companies people generated requests for. We also determined that *AMI* could provide more support to users after they created their requests, such as helping them deal with frequently encountered issues. Finally, our initial implementation left a great deal of room for improvement with regards to making it easy to launch *AMI* for new industries, in other languages, or focused on other jurisdictions.

3.1 BASIC STATISTICAL TRACKING

AMI should be able to collect the following data in order to provide basic statistical information:

- A record of a request being created at a particular date and time

- The organization to which the request was addressed
- The language the request was created in
- The jurisdiction that the request cites for its legal basis

These data will let us determine which companies receive the most requests, how the volume of requests changes over time, what types of industries that people are most interested in requesting data from, in what languages requests are most frequently issued, and in what jurisdictions most requests are created.

By limiting our default data collection to the above items, we cannot easily understand much about an individual user's requests, such as how many unique individuals create requests, and for how many different organizations on average individuals create requests.

A tradeoff of fulfilling this requirement is that the system can no longer fulfill our initial design goal of zero user data collection. If statistical data is collected about requests, even if no user-inputted personal data is collected, the user's IP address will still be associated with the data. The IP can be used for some identification purposes. Therefore, *AMI* should provide users the option to opt-out of statistical data collection, and not associate user IP addresses with statistical records in the application database. To prevent data from being intercepted in transit, *AMI* should employ HTTPS to secure every transmission.

3.2 USER ENGAGEMENT

Access My Info should support users throughout the request process, not just at the time of request creation. In Canada, companies are obligated to respond within 30 days of receiving an access request. Access My Info should remind users after 30 days that they should have received a request, and provide some information on what to do if they have not. After 60 days[1], *AMI* should follow up to check to ensure they have received a response, and let them provide feedback on their request. This timeline of events would vary in other jurisdictions depending on the legal requirements for response timeframes.

While *AMI* by default should not collect any personal information about the people using the tool, such collection is a prerequisite for contacting users. Therefore, *AMI* should provide an **opt-in** mechanism so that consenting users can provide their email address to the system. *AMI* should take steps to verify the ownership of the email address submitted and provide an easy method to unsubscribe.

In addition to directly contacting users, *AMI* should publish information, accessible online, that can aid requesters in engaging with companies to ensure they get satisfactory responses to their requests. Such information should be available online.

[1] Canadian organizations can obtain a 30-day extension for their response to a personal information access request, bringing the total maximum time to 60 days.

3.3 SUPPORT MULTIPLE INDUSTRIES

Access My Info should make it easy for a user to create access requests for various industries. Specifically, each industry should have its own request letter template, its own set of default data types added to the request, a set of personal identifiers that the requester needs to provide to verify account ownership, and a list of companies categorized into that industry.

3.4 SUPPORT MULTIPLE LANGUAGES

Access My Info should be easy to internationalize. It should be easy for users to switch languages. It should intelligently set a default language for a first-time user. For each new language, translations must be provided for the user interface, and for each:

- Request letter template
- Data type that can be requested
- Types of personal identifiers that users must provide
- Email communication templates
- Educational support material about access requests

The *AMI* system should be implemented to make it very easy for an administrator to input the above translations into the system.

3.5 SUPPORT MULTIPLE JURISDICTIONS

Access My Info should be able to support people from different legal jurisdictions to use the tool to create requests citing relevant law in their jurisdiction. *AMI* should furthermore make it simple to detect and set the jurisdiction in which the user is situated, and enable them to easily switch jurisdictions within the tool. This means that the following types of material should be customized and/or developed for each new jurisdiction:

- Specific request letter templates
- Educational support material
- Email communication templates

For some instances, it may be preferable to simply implement a fully customized version of *AMI* in a new jurisdiction, without providing the capability for the user to change jurisdictions. This may be preferable because user interfaces may need to be customized and custom branded for a new context, and people may want to host their own version of the tool in their own legal context.

4 CONCEPTUAL DESIGN

Conceptual models help software designers understand the different components of the problem or issue they attempt to address through their application. Determining and documenting the different entities that interact within a problem space helps to ensure that a proposed solution will address the right problem, from an informed position.

This section presents the primary conceptual model that informed the design of *Access My Info*: A model of an access request and response. The model is intentionally broad in scope, aiming to capture many possible ways someone could request access to their personal information from an organization. It is not meant to be exhaustive, but instead to provide a "good enough" foundation on which to develop and continue to improve *AMI* and related access request technologies.

We developed the model by reviewing the request letter template used in the first version of *AMI*, fact sheets about Canada's consumer privacy legislation PIPEDA, and analyses of Canadians' experiences with their telecommunications service providers when filing PIPEDA requests. We searched for the different actors involved, various types of data that should be included in a request letter, different aspects of the request process, and legal requirements that might affect the substance of a request or process of sending a request and receiving a response.

> The model below is an Entity-relationship diagram. Boxes represent entities – actors or information pertinent to a consumer personal data access request. Connectors represent relationships between entities. Connector start and end points indicate whether the relationship is one-to-one, one-to-many, or many-to-many. A crow's foot means that many instances of the entity adjoining to it can be related to the entity on the other end of the connector.

A detailed exposition of the various entities in the *AMI* model and their interrelationships follows.

4.1 COMPONENTS OF A REQUEST

4.1.1 ACTORS

REQUESTER The individual data subject who creates a `request`, an identifiable person. We assume they reside at least one `jurisdiction`. For simplicity, this assumes the `requester` is the data subject (the identifiable individual about whom personal data is collected). The `requester` decides which `data operator` they'd like to request information from, what `questions` they'd like to ask, what specific `data` they'd like to obtain, and is subjected to with `re-`

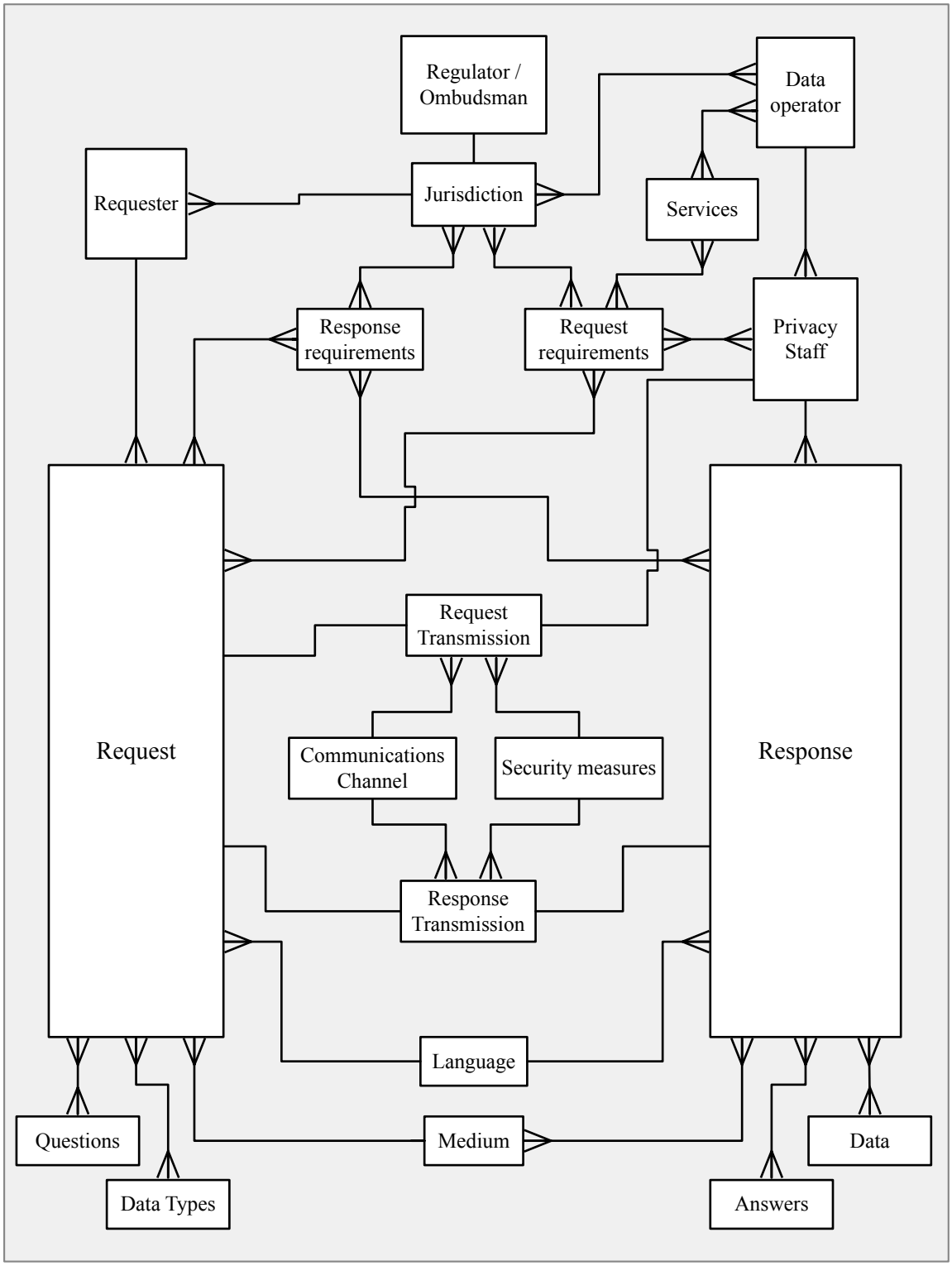

Figure 1: Model of an personal information access request

quest requirements as stipulated by the data operator and the regulator / ombudsman. The requester communicates the request to the data operator by sending a request transmission to the operator's privacy staff.

DATA OPERATOR The organization to which a request is sent. Data operators may provide services that a requester utilizes, and are thought to hold personal information about the requester. Data operators employ privacy staff who receive and handle requests. Data operators operate in one or more jurisdictions with specific response requirements for how they respond to requests and may be subject to oversight by regulators/ombudsmen.

PRIVACY STAFF The individual(s) who receive public communications on privacy issues at the data operator, they receive and process access requests. They communicate with requesters, providing one or more responses via a response transmission. Privacy staff may communicate request requirements to requesters, for instance, if they have not been met in the initial request.

REGULATOR / OMBUDSMAN A government or arms-length public body that oversees compliance with privacy regulation in a given jurisdiction. Often responsible for publishing legally-mandated request requirements and response requirements. Often responsible for receiving and handling complaints created by requesters in response to their interactions with data operators when attempting to obtain access to their information.

4.1.2 INFORMATION

REQUEST A request is the requester's demand for access from a data operator. The request may ask questions about how their personal data is being used and specify the data types they are particularly interested in obtaining. The request may cite response requirements relevant to the requester's jurisdiction that the data operator must abide by. Similarly, the request may contain information that fulfills the request requirements determined by the data operator and the jurisdiction. The request is expressed in a specific medium and language. The requester communicates their request via a request transmission sent to the data operator's privacy staff.

RESPONSE A communication issued by the privacy staff at a data operator as a reply to a request they received. The response may contain data that may or may not correspond to the data types asked for by the requester. Similarly it may or may not contain answers that correspond to the questions asked by the requester. The response may ask the requester

to provide additional information to the privacy staff to comply with the data operator's or the jurisdiction's request requirements. The response is expressed in a specific medium and language. The privacy staff communicate their response via a response transmission sent to the requester.

DATA TYPES A request can include a list that specifies types of personal data that the requester thinks may be held by the data operator and would like access to. For instance, a requested data type could be "geolocation information", where the requester is interested in knowing what sort of record the data operator has regarding their whereabouts. Data provided by the data operator in its response may or may not be strongly associated with the data types requested.

QUESTIONS A request can include a list of questions that the requester is interested in obtaining answers to from the data operator. These typically will focus on the collection, use, disclosure, retention, or deletion of the requester's personal data. Questions can be directly associated with requested data types. For example, a question might ask "for how long do you retain my geolocation information?", where the data type in this case is "geolocation information".

ANSWERS A response can include a set of answers that may or may not correspond to the questions asked in the request. For example, data operators may combine multiple questions into a single answer, may not address questions at all, or may provide new information not asked for in the questions. Answers may be at a different level of specificity than asked for in the questions. For example, a question may ask "have you disclosed my information to third parties including law enforcement or state agencies?" An answer could only address one part of the question, such as "We have not disclosed your information to law enforcement or state agencies." In this example, the answer does not provide specific information about other third parties besides law enforcement, such as commercial partners.

DATA A response can include data that may or may not correspond to the specific data types asked for in the request. For example, if a data operator does not retain a data type, they will not be able to provide it. In other cases, the data operator may not include all the data they collect in their response. This can occur if the data operator instructs the requester to view their data in an online portal, or simply does not acknowledge they retain such data. Data may be provided in a variety of formats (spreadsheet, screenshots, etc), fixed in a specific response medium.

4.1.3 TRANSMISSION

REQUEST TRANSMISSION A request transmission gets a request to a data operator. The transmission needs an addressable sender (the requester), and an addressable receiver (the privacy staff at the data operator). It also requires a communications channel.

RESPONSE TRANSMISSION A response transmission gets a response to a requester. The transmission needs an addressable sender (the privacy staff at the data operator), and an addressable receiver (the requester). It also requires a communications channel, and should employ security mechanisms to protect the confidentiality and integrity of the information being transmitted.

COMMUNICATIONS CHANNEL A communications channel is the method by which a request transmission or response transmission is communicated. This is typically either postal mail or email. In some cases, telephone calls can be included, as can internet URLs directing requesters to websites to download their data.

SECURITY MECHANISMS A security mechanism should be in place for response transmissions to ensure the confidentiality and integrity of the contained information. Sometimes this includes requiring a signature on delivery for postal mail, or setting a password on transmitted documents (eg a PDF password), or serving the response data from a website protected by HTTPS. Additional security mechanisms may be in place to verify the requester's identity, per the request requirements, but such mechanisms are not relevant to the transmission of the data itself.

4.1.4 EXPRESSION

MEDIUM The medium in which the request and response is expressed. Typically, this will be in writing. In some cases, requests and responses may be communicated through voice communications. The medium is transmitted through a communications channel. A specific medium may in some cases be a request requirement or response requirement.

LANGUAGE A language represents the ideas contained in a request or a response and is realized in a medium. A specific language may in some cases be a request requirement or response requirement.

4.1.5 PARAMETERS

JURISDICTION The legal context in which a `requester` or a `data operator` exists. `Jurisdictions` will often have privacy laws that set out certain `request requirements` and `response requirements`. For example, a law might require that `data operators` respond to a `request` within thirty days. A law might require that `requesters` provide adequate information to `data operators` so that they can identify the `requester` in their records.

REQUEST REQUIREMENTS Requirements that a `request` must comply with in order for the `requester` to obtain access to their `data`. These can be stipulated by a `jurisdiction` and by the `data operator` itself. A jurisdiction, through law, might require that a request be made in writing. A `data operator` might require specific identifiers from the `requester` to verify their identity. For example, an online dating service might require that a requester provide their username and use the email address associated with their account in order to verify their identity.

RESPONSE REQUIREMENTS Requirements that a `response` must comply with in order for the `data operator` to meet the legal obligations set out in a `jurisdiction`. These can include a timeframe within which the `operator` must issue a `response`, requirements about the `medium` and `language` of the response, the format the `data` must be expressed in, and accessibility requirements, and whether or not the `operator` can charge a fee for access.

4.2 APPLICATIONS

We implement the majority of this model in *Access My Info*. We use a CMS to define and help manage various records of the different entities in the model. See Section 6 for a more detail about *AMI*'s technical implementation..

Implementing this model allows for *AMI* to be more readily adopted in new jurisdictions, to new industries, to new requirements for identity verification, among other advantages. The primary disadvantage is the complexity of the implementation makes it challenging to create, define, and manage the various data entities needed for a functioning *AMI* instantiation. Dedicated user experience design focused on the administrator experience for managing this data could help a great deal.

Defining the model also defines a vocabulary with which one can discuss *AMI*-related concepts. Given the complexity of the components of an access request, having a codified terminology can help when talking about any aspect of the request process, which we discuss in detail below.

5 THE ACCESS MY INFO PROCESS

In this section, we present the data request processes that *Access My Info* has been designed to support. The process is divided into three parts: 1) Awareness; 2) A typical request/response; 3) Exceptions to the typical process.

> We describe the various parts of the processes and include graphical process models to illustrate the flow of information between different actors during the processes. Each column in the models contains the tasks that a particular actor performs in the request process. Four different actors are depicted: 1) the requester; 2) *Access My Info*; 3) the data operator; 4) the jurisdiction.

5.1 AWARENESS

The first stage in the *AMI* process focuses on the conditions that enable an individual to send a request. A jurisdiction publishes information about its right of access, which helps citizens become aware of their rights. This also lets *AMI* provide educational material about access rights to potential requesters. Data operators make public what data they collect about users and how they are identified, which informs *AMI* administrators in developing a database of potential data types, questions, and identifiers relevant to the data operator's services. Once aware of their access rights, and perhaps having consulted public materials about their data operator's privacy practices, potential requesters might develop questions about how their data is used.

5.2 A TYPICAL REQUEST

From a requesters perspective, a typical request beings with a prospective requester when they are aware of their rights and may have some questions about their data. In order to successfully support a requester, Access My info must cultivate a database of data operators, data types, and identifiers, the requester can start the typical request process, supported by Access My Info. A typical process means that the user can create a request and get access to their data without taking any extra steps. Any exceptions to this typical process are discussed in the next section.

The process begins with a requester visiting *Access My Info.* AMI provides the requester with a list of industries and data operators to which they can create their request. The requester selects a data operator, then chooses from a list of data types and questions in which they are interested. The requester can add their own questions if they like. Next, the requester views a list of required identifiers and adds those identifiers to their request so that the data operator can effectively identify the requester and retrieve their records. AMI then combines the requester-

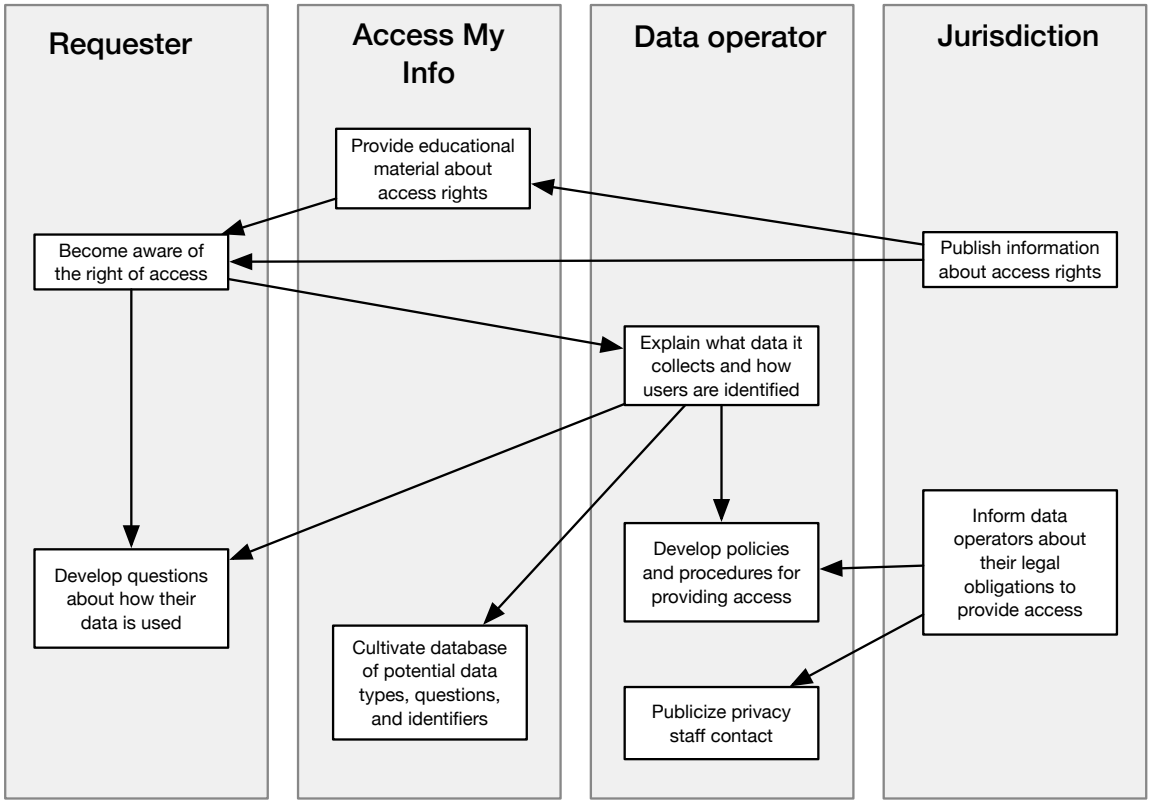

Figure 2: Awareness phase of the *AMI* process

submitted data with a request letter template and presents the generated request letter to the requester. The requester finally selects the request transmission channel, using AMI-supplied privacy staff contact information to help direct the transmission. The requester then sends the request to the data operator's privacy contact.

Next, the data operator's privacy contact receives the request. They verify the requester's identity, estimate time required to fulfill the request, decide whether or not to charge a fee for access, retrieve customer records based on the provided identifiers, redact other people's personal information from the records, answer any questions in the request letter, and finally send the response to the requester. There are many exceptions that could arise in this process, which are discussed in detail in the next section.

Note that in the above process, we do not know what exactly the data operator does once they receive the request. We have inferred the process described above based primarily on our findings from how Canadian telecommunications service providers respond to requests. Experiences could vary depending on the type of industry or the jurisdiction, among other variables.

Additionally, the jurisdictional authority does not play an explicit role in the typical process. This is because in an ideal process, when a requester can successfully access their data, the jurisdictional authority would not be actively involved. In some cases, jurisdictional authorities

could be alerted – such as if a data operator is compelled by law to notify authorities before letting a requester know if their data was provided to law enforcement or other state agencies.

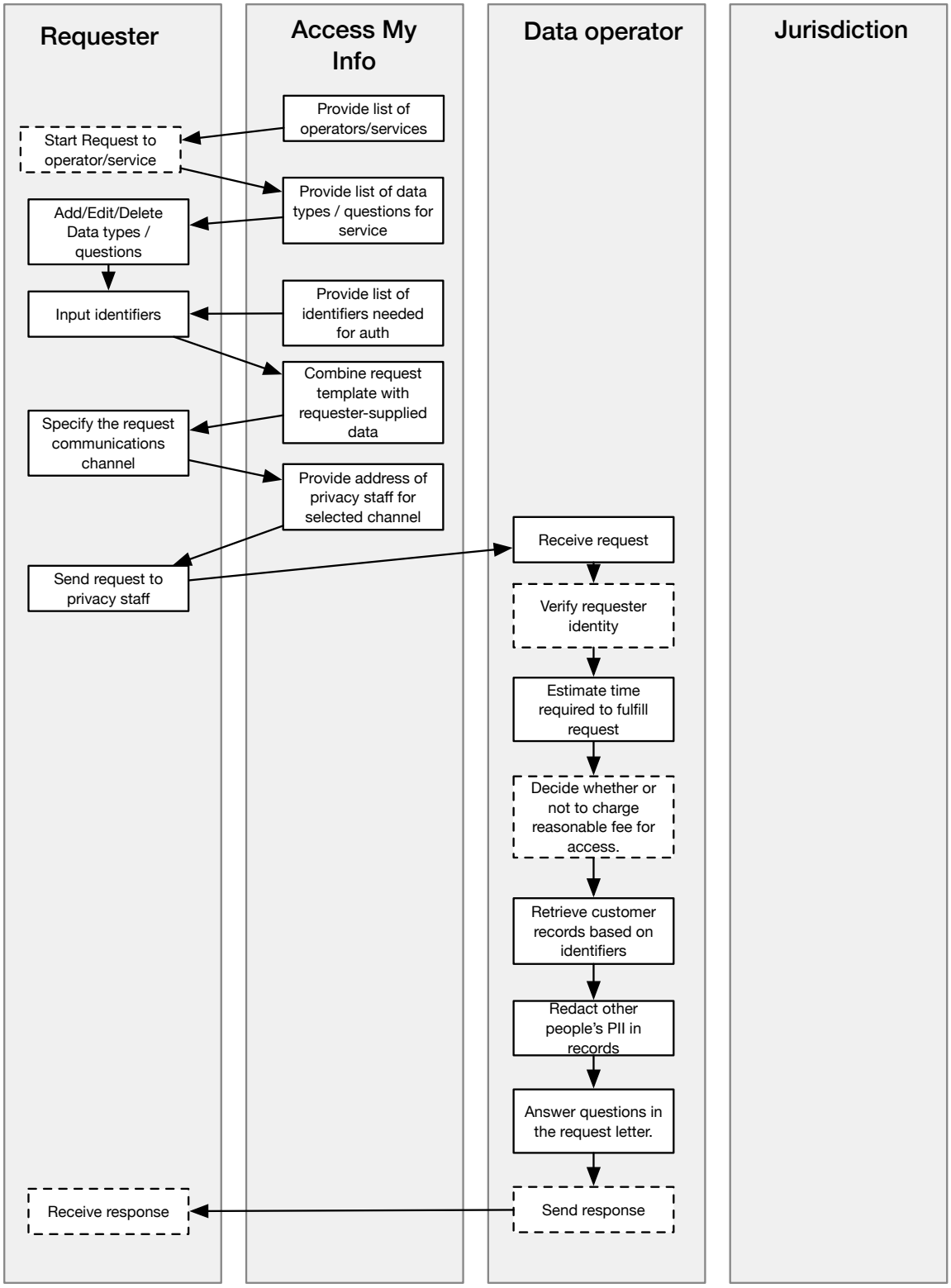

Figure 3: Typical *AMI* process

5.3 EXCEPTIONS TO A TYPICAL REQUEST

Having discussed how the typical process can unfold when someone requests their personal data using *AMI*, we now turn to how the process could get more complicated. In many tasks performed in the process, exceptions can occur. This section documents some of the most prominent exceptions that may occur during a request.

In the *AMI* interface, requesters are required to select a data operator they wish to create a request for. If the organization isn't listed, *AMI* provides an alternative step where a user fills in the name of the company and its privacy contact details themselves, and carries on with the request. *AMI* instructs the requester that privacy contact information can usually be found on a company's privacy policy. This exception handling is only useful if the requester wants to send a request to a company that can be categorized into one of the industry types supported by *AMI*.

Among the most common exceptions is that the requester sends in their request, but does not receive a response after the mandated response period has expired. In these cases, *AMI* can help support requesters by automatically sending a reminder email after the period has ending. This email typically recommends that the requester send a reminder to the data operator, and includes suggested phrasing to help elicit a response. This information is duplicated on the *AMI* website itself, so that requesters who did not opt-in to receiving email notifications can still benefit from this information.

If the requester never receives a response even after sending a reminder, the requester can often be justified to complain to the jurisdictional authority. *AMI*, in its guide on nonresponsiveness, informs requesters of this right.

Another common exception is when the data operator determines they require additional proof of identity from the requester prior to responding to the request. Typically, the data operator will contact the requester, demanding documentation to prove the requester is who they say they are. The data operator can demand that the requester use the email address associated with the requester's account with the data operator's service (which *AMI* recommends requesters do from the outset). In some cases, requests for identity verification may be overbroad, such as a request for a notarized copy of a driver's license or passport. In those cases, *AMI* provides guidance on how requesters can often negotiate with data operators to prove their identity through less onerous methods.

Another common exception is when a data operator demands payment in exchange for retrieving data and providing access. Whether or not a fee can be demanded depends on the parameters of the law granting the right of access. In some cases in Canada, we have found that telecommunications companies often provide price quotes for different types of records, often at high rates, such as $100 per month of records. The fee demands may inhibit requesters from obtaining access, as they often may require requesters to provide a more detailed version of their request in order to reduce fees, which is a significant amount of extra labour for the re-

quester. *AMI* can help by providing guidance on how requesters can pare down their request to get a sense of the personal data held about them, without having to pay high fees for complete access. In other cases, where fees may be deemed unreasonable, *AMI* can provide information about where requesters can complain to the relevant jurisdictional authority.

The final exception we discuss here occurs when the requester receives a response that includes some data and answers to their questions, but is unsatisfied with the response. Reasons for not being satisfied could include:

- Data known to be collected is not provided;
- Answers to questions were vague or combined together;
- Data is not in a useable format or is corrupted

In these situations, the requester could send another request to the data operator, asking for additional data, clarifications on some points, or for the data to be provided in another format. If the data operator does not respond satisfactorily, the requester could complain to their jurisdictional authority. AMI can provide guidance on how a requester might seek clarification or additional data from their data operator, and on how they can go about filing a complaint to their jurisdictional authority.

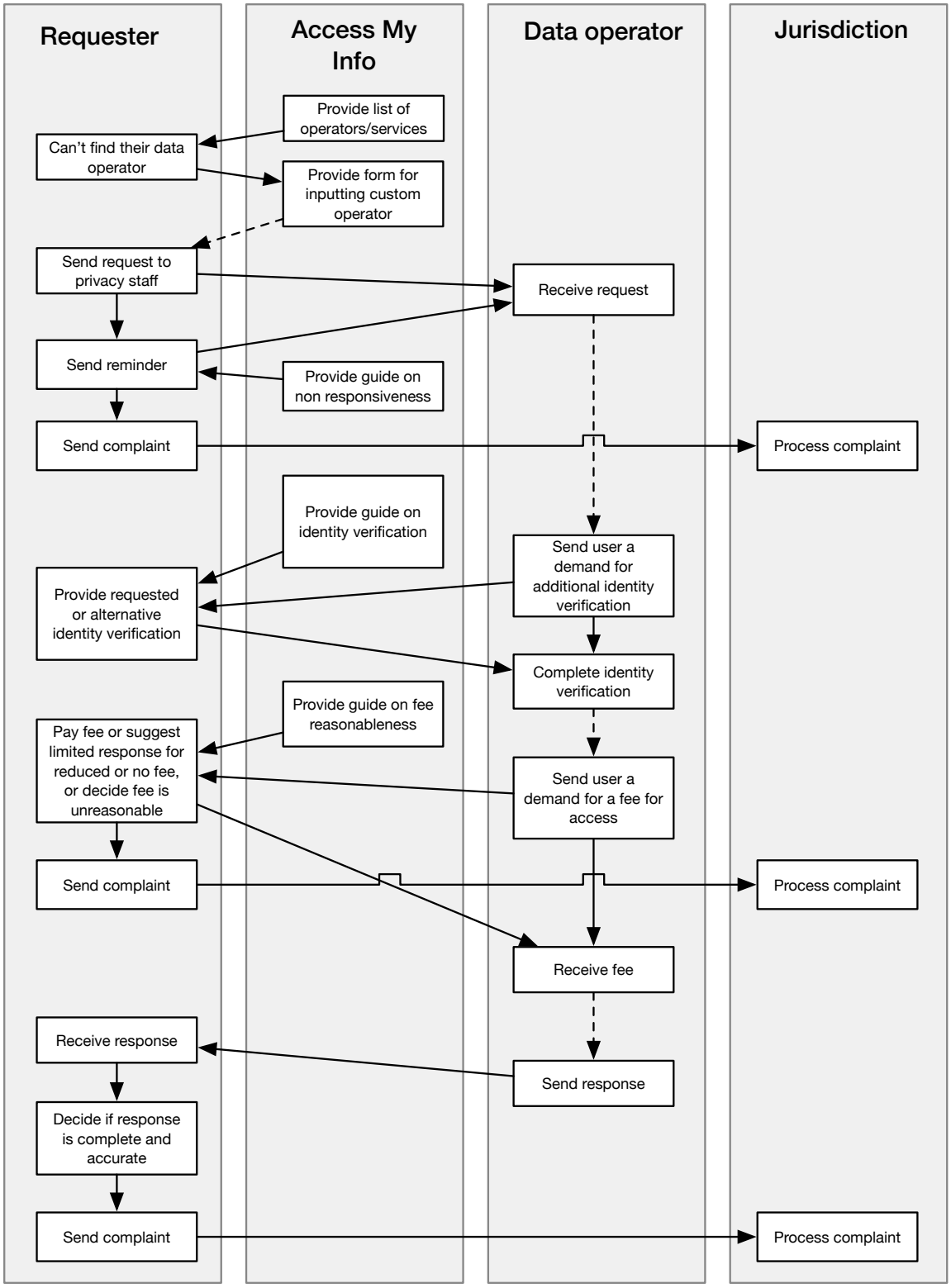

Figure 4: Common exceptions to the typical *AMI* process

6 TECHNICAL IMPLEMENTATION

We implemented *Access My Info* as a web application comprised of three subsystems: The frontend, CMS, and community tools.

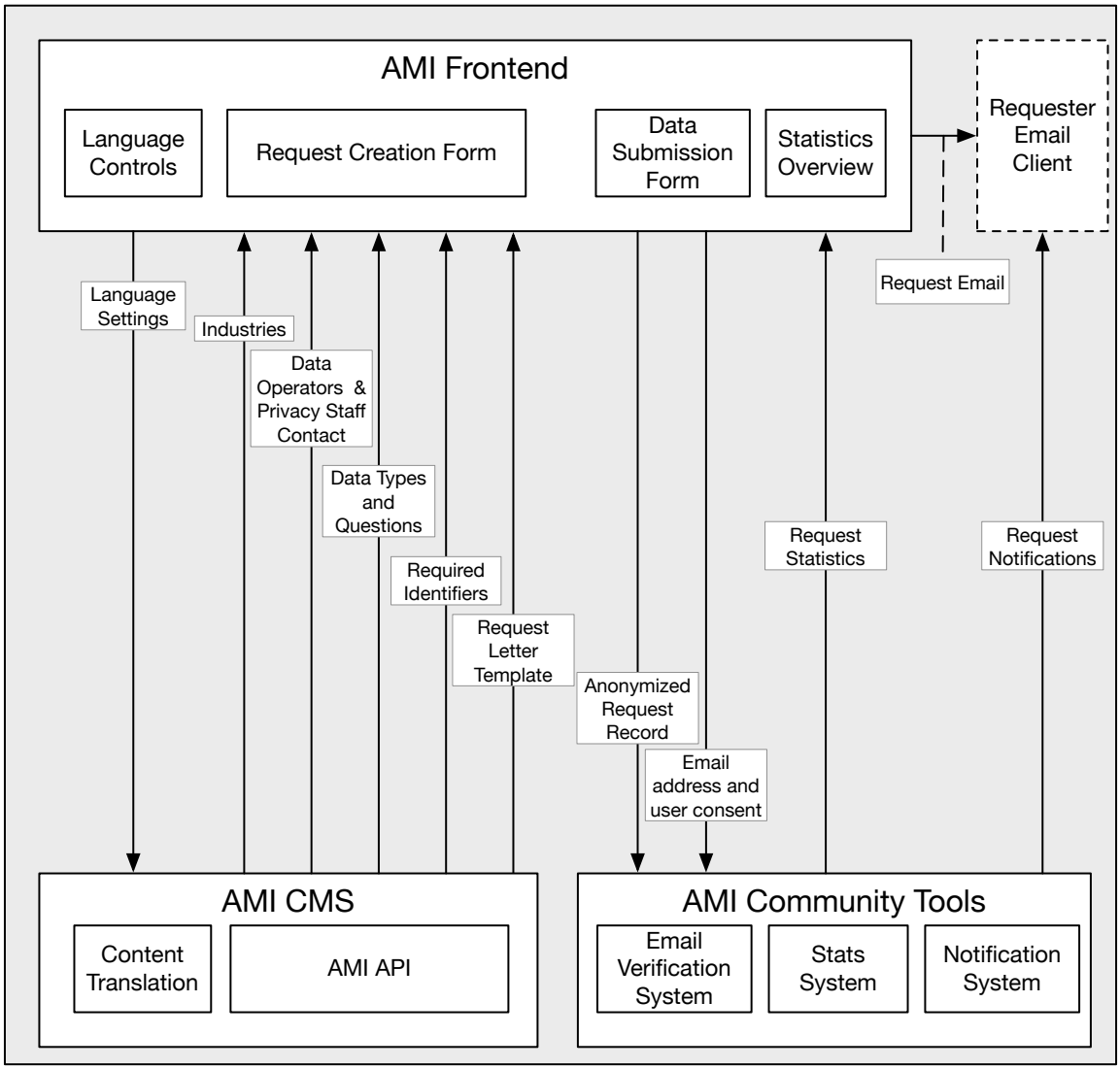

Figure 5: *AMI* system architecture

6.1 AMI FRONTEND

The frontend is what the requester sees. It provides educational content about access rights and exceptions that can occur in the request process. It guides the requester through a step-by-step process of selecting a data operator, refining questions to ask, inputting personal identifiers, and obtaining a request letter based on a pre-written template. Figure 6 shows different stages in the frontend's step-by-step process.

End users can opt-out of sharing anonymized statistics about their requests, and opt-in to receiving email notifications at key points in the request process.

The frontend is a single page Javascript application using the AngularJS framework. It supports multiple languages.

The frontend generates PDFs clientside using a custom AngularJS directive, the PDFMake Javascript library, and a custom script we authored called CanvasDoc. CanvasDoc converts HTML to a canvas image so that unicode characters can be rendered in a PDF. We go through this technical process so that the frontend can generate the PDF clientside without having to send personal information about the requester over the internet to the AMI server. This helps preserve the privacy of the requester.

> For more information on the frontend, please consult its code and documentation repository on Github.
> `https://github.com/andrewhilts/ami`

6.2 CONTENT MANAGEMENT SYSTEM

The *Access My Info* content management system (CMS) provides a mechanism for administrators to create and manage content corresponding to many of the entities described in the conceptual model described in Section 4.

The CMS is a Wordpress website that uses some custom plugins to act as an API that provides content to the frontend. When a user loads the frontend, the frontend queries the AMI CMS to obtain a list of different industries. When the user selects an industry, the frontend queries the CMS for a list of data operators in that industry. The process continues throughout the whole request creation cycle, and can be seen in the architecture diagram in Figure 5

CMS administrators define industries, data operators, data types, personal identifiers, jurisdiction, and perhaps most importantly, can author request letter templates for particular industries directly in the CMS.

> For more information on the CMS, please consult its documentation on Github.
> `https://github.com/andrewhilts/ami-system/blob/master/docs/amicms.md`

6.3 AMI COMMUNITY TOOLS

The AMI Community Tools system manages the statistical tracking of requests, and provides email notification functionality to requesters who have opted-in to the feature.

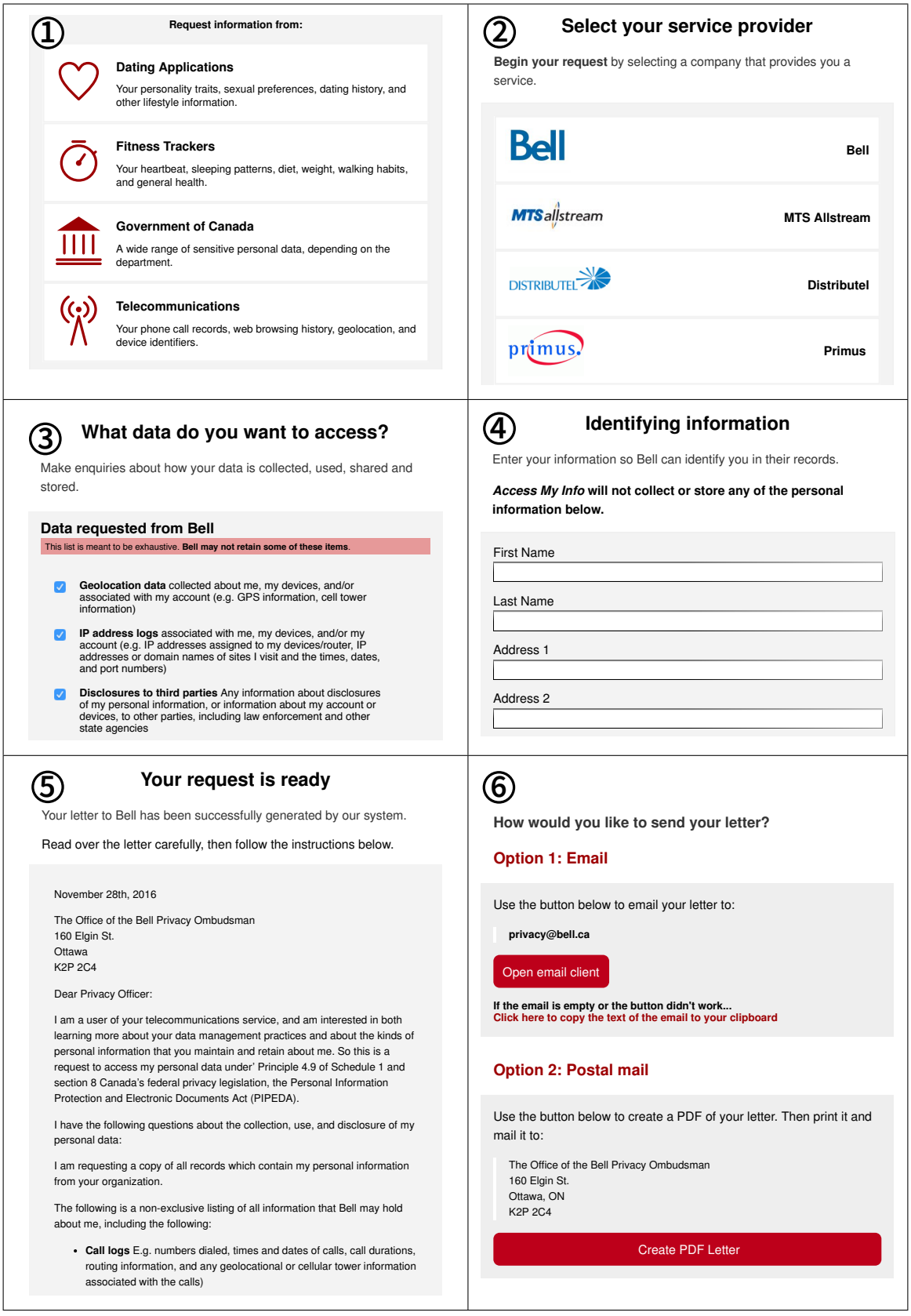

Figure 6: User interaction stages in the *AMI* frontend

The Community Tools is a Node.js system that runs an API that the Frontend communicates with. When an end user completes their request, the Frontend will send anonymized information about the request to the Community Tools, if the requester hasn't opted out. The frontend will also send the requester's email address to the Community Tools if the user opts into email notifications.

Request statistics include the date of the request, the request company, and the request jurisdiction. When a user opts into email notifications, the Community Tools sends a verification email with a unique token to the user's email address. Once the user clicks on that link, they are taken to the Frontend, which in turn sends the verification token to the Community Tools service, completing the process. Subscribers can also unsubscribe using a similar process. When requesters subscribe, they are scheduled to receive several emails from the system, based on key events such as the expiration of the mandatory response timeframe.

> For more information on the Community Tools, please consult its documentation on Github.
> `https://github.com/andrewhilts/ami-community`

6.4 SECURITY AUDIT

In April 2016, the security firm Cure53 conducted a security audit on *Access My Info*'s source code and a development implementation of the system. The audit was funded through the Open Technology Fund's Red Team Lab. Overall, the system was found to make many good security decisions, with the weakest point being the CMS' use of WordPress.

> Full security audit report:
> `https://cure53.de/pentest-report_accessmyinfo.pdf`

7 DEVELOPING AMI

As described in Section 6, there are 3 technical subsystems underlying the entire AMI system. To help developers starting working on AMI, configuring it for new instances, and compiling and deploying it on remote servers, we have developed automated scripts that streamline many of these processes.

For complete information on how this works, please consult the AMI System project documentation and code on Github.
`https://github.com/andrewhilts/ami-system`

8 RUNNING AMI IN A NEW JURISDICTION

Running *Access My Info* in a new jurisdiction is not as simple as copying the application's source code and changing the branding to a new country. Before even considering deploying the application, it is sensible to develop an understanding of the right of access in the new jurisdiction. This can be done by researching the law and how the industry sectors you want people to be able to request their data from respond to requests. We recommend conducting a pilot study, with some simple paper-based requests to leading companies in a jurisdiction, prior to any customization of *Access My Info* for the region. By understanding how companies respond to requests, you can better customize AMI to help requesters get the most out of the process.

9 IMPACT

This section briefly describes some of the results of releasing *Access My Info* in both Hong Kong and Canada in 2016. The numbers of requests created, basic patterns in when requests were created, and general information about media coverage are presented.

9.1 AMI HONG KONG

Access My Info was released in Hong Kong in April 2016. At the time of writing, over 1400 requests have been made using the system, to seven different telecommunications service providers. The majority of requests were created within the first month of the tool's release, followed by a rapid decline. Short-lived bursts of new requests emerged several times after the release, seemingly tied with media coverage about the tool.

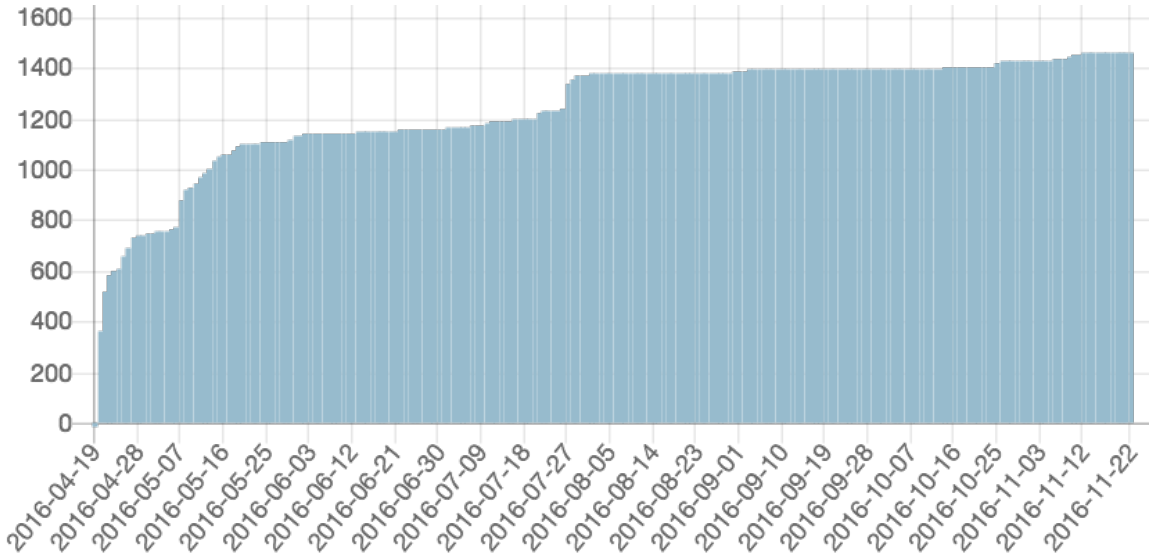

Figure 7: Cumulative requests created by *AMI* Hong Kong over time

Media coverage about *Access My Info* in Hong Kong primarily served to inform the public about the existence of the tool, with some focusing on how telecommunications companies in the region responded to receiving requests. Coverage included:

- Kris Cheng. "New app helps citizens find out what Hong Kong companies know about them." *Hong Kong Free Press*, April 19, 2016. `https://www.hongkongfp.com/2016/04/19/new-app-helps-citizens-find-out-what-hong-kong-companies-know-about-them/`
- Kris Cheng. "Telecom companies fail to provide sufficient responses to personal data requests, transparency advocates say". *Hong Kong Free Press*, May 6, 2016. `https://www.hongkongfp.com/2016/05/06/telecom-companies-fail-to-provide-sufficient-responses-to-personal-data-requests-transparency-advocates-say/`
- Josh Horwitz. "It's time to ask your telco how it's tracking your data, Hong Kong activists say." Quartz, May 9, 2016. `http://qz.com/678923/its-time-to-ask-your-telco-how-its-tracking-your-data-hong-kong-activists-say/`

9.2 AMI CANADA

Our redeveloped version of *AMI* was released in Canada in June 2016, with French language support added in September of that year. At the time of writing, over 4000 requests have been made to over 40 different organizations. The organizations operated in the online dating, fitness tracking, telecommunications, and government of Canada categories. Similarly to the case of Hong Kong, *AMI* Canada saw an initial flurry of requests being created, with another much smaller

uptick after media coverage. The release of the French version of the tool led to over 1000 new requests.

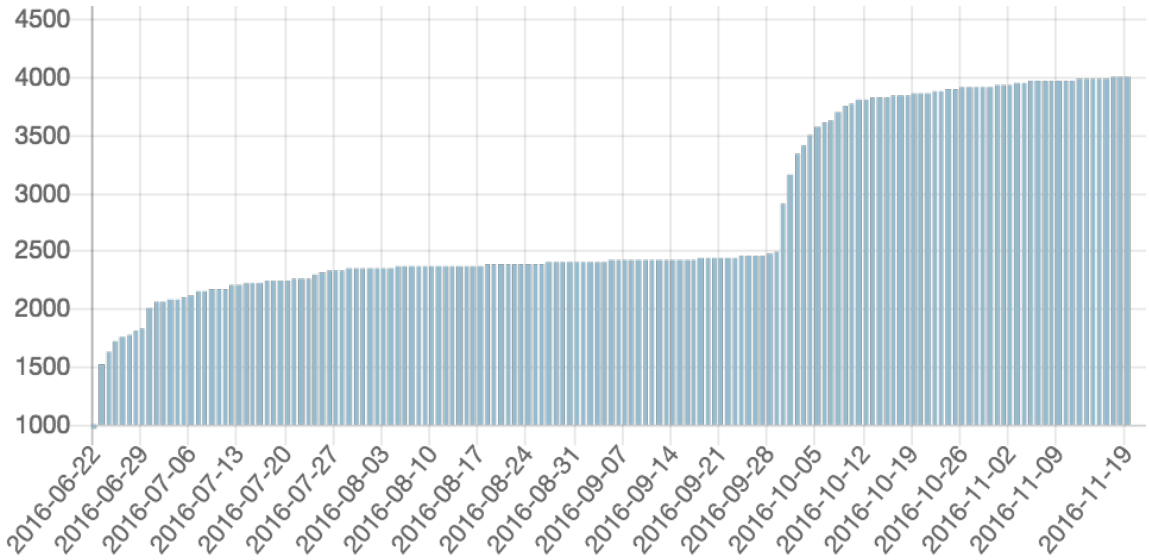

Figure 8: Cumulative requests created by *AMI* in Canada over time

Media coverage about *Access My Info* in Canada primarily served to inform the public about the existence of the tool. Particular attention was paid to the new industry categories that became available for Canadians following the new release of *Access My Info*. Coverage included:

- Emily Chung. "What are your dating and fitness apps sharing about you?" *CBC News*, June 21, 2016. `http://www.cbc.ca/news/technology/access-my-info-1.3644019`
- Karolyn Coorsh. "Is Tinder or Fitbit using your personal data? Online tool shows you how to ask." *CTV News*, June 22, 2016. `http://www.ctvnews.ca/sci-tech/is-tinder-or-fitbit-using-your-personal-data-online-tool-shows-you-how-to-ask-1.2956721`
- Matthew Braga. "How to Ask Dating Apps and Fitness Trackers For Your Personal Data (in Canada)." *Motherboard*, June 21, 2016. `http://motherboard.vice.com/read/how-to-ask-dating-apps-and-fitness-trackers-for-your-personal-data-in-canada`
- Danny Bradbury. "Online PIPEDA tool highlights Canadian trust issues'.' ITWorld Canada, July 11, 2016. `http://www.itworldcanada.com/article/online-pipeda-tool-highlights-canadian-trust-issues/384837`
- "Access My Info on Metro Morning." *CBC Radio Toronto*. `https://www.youtube.com/watch?v=rtH8soRIe40`

10 LIMITATIONS AND FUTURE WORK

While *Access My Info* has seen success in empowering thousands of people to create legal requests for access to their personal information, the system is not without limitations. Some of these limitations are about the request process in general and others are of a more technical nature. Additionally, some of these limitations can be relatively simply addressed through future work, while others may be general limitations of the *AMI* approach.

10.1 PROCESS LIMITATIONS

The core of *AMI* is arguably the request letter templates that the application assists users in filling out which they then transmit to data operators. Request letter templates are currently authored for entire categories of data operators. As a result, all data operators in a given industry will receive letters based on identical templates. This approach is useful from a comparative research perspective – it enables researchers to see how companies respond to the same letter. However, different data operators respond in different ways to requests, with some companies responding with clarifying questions. Requesters could be better served if request letter templates were written in a manner tailored to each individual company and kept up-to-date as companies change the procedures and substance of their responses. Customized letters for each data operator, however, introduce huge time and labour costs for *Access My Info* administrators, and therefore do not seem at present to be a scalable solution.

By tracking request creation statistics, we can clearly see that people use *AMI* largely in response to media coverage about the tool. *AMI* has not been advertised outside of issuing a press release, along with some simple social media posts about the tool when the tool was released. A corollary of this is that data operators receive requests in huge deluges, not at consistent rates. One way to encourage more consistent awareness levels and stable request rates would be to invest a small sum in advertising *AMI* online. Other methods could include periodized updates to the tool, such as adding new industry categories on a quarterly interval, or the routine writeup of what people have learned from their requests. These outputs could lead to more regular media coverage of the tool as well as organic social sharing.

AMI technically supports the ability for a user to change the jurisdiction in which they create their request within the tool itself. They could click a different country from a menu, and the entire application would change to present potentially different sets of industries, data operators, and request letter templates. In practice, however, each version of *AMI* has been released focused on a single jurisdiction. The ability to switch jurisdictions has been hidden from the user. By focusing on a single jurisdiction, the user interface, messaging, and branding of the application can be customized for that context. This reduces complexity for the end user. While the ability to change jurisidictions may introduce unneeded complexity in current releases, it is

possible that for new deployments of *AMI*, that target an entire region of linguistically, and culturally close countries (such as Spanish-speaking Latin American countries, for instance), the ability to change jurisdictions could still prove valuable because it would save developers from re-implementing several versions of the same tool.

10.2 TECHNICAL LIMITATIONS

The conceptual model upon which *AMI* is based is fairly closely adhered to in the CMS providing content to the *AMI* frontend. However, the frontend itself is request-oriented, and thus, groups all other data underneath a top-level `AMIRequest` object. As a consequence, it is more difficult for requesters to reuse some basic personal identifiers across different requests, and for identifier fields, data types, question descriptions and request letter templates to be easily replaced with their equivalents in different languages during the request creation process. In addition, if the data structures utilized in the frontend become more faithful to the conceptual model of an access request, it will become easier to implement a feature enabling *AMI* to output the request in other usable formats, such as XML or JSON.

Generating PDFs in the web browser presents its own set of technical limitations. Non-latin characters are not included in the core PDF standard, and thus aren't available for use in the web browser without downloading huge font files at runtime. Such downloads which would cripple application performance. Because of this limitation, we developed CanvasDoc[2] to convert HTML into a image files, and then save those images to PDF. However, our implementation is quite limited and we do not support much styling, tables, or images. As a result, the formatting of the letter is basic, and could be made to look more professional with dedicated effort to improve CanvasDoc.

A clear area for technical improvement is in the integration of the Community Tools and CMS subsystems. For example, certain values from the CMS are hardcoded into the Community Tools system. Unique IDs representing jurisdiction records in the CMS are stored as parameters in the Community Tools system, and entries in the Community Tools database represent CMS data operators. In an ideal system, the Community Tools and CMS could be combined into a more streamlined API without the need for separate configurations.

Another area that could benefit from integration of the CMS and Community Tools is a feedback mechanism. *AMI* has (under the hood) the capability to store feedback about specific requests. The Community Tools system could send an email, inviting feedback on a specific request. If the Community Tools were more tightly integrated with the CMS, it would be simpler to incorporate user feedback into the overall user experience, potentially opening the door for data displays such as visible statistics about various companies' responsiveness.

[2] Andrew Hilts. "CanvasDoc". *Github*. `https://github.com/andrewhilts/canvasdoc`

11 PROJECT TEAM AND CONTRIBUTORS

Craig Choy, Masashi Crete-Nishihata, Jakub Dalek, Ronald Deibert, Andrew Hilts, Kelly Kim, Jeffrey Knockel, Jason Li, Adam Molnar, KS Park, Christopher Parsons, Alexandre Plourde, Irene Poetranto, Sinta Dewi Rosad, Lokman Tsui, Glacier Wong, Yee Ting Yu, Sonny Zulhuda.