DyCons: Enabling Dynamic Consent on Distributed Bioinformatics Infrastructures

Final report

by Yuriy Pavlov

for Dr. Blanchette, COMP 402

for McGill University

Submitted April 25, 2020

Updated April 28, 2022

Abstract

Distributed bioinformatics data-sharing infrastructures enable the reuse of data between institutions, improving the statistical power of studies while permitting data to live under the oversight of its home institution. Respect for the right to informed consent has been a particularly challenging standard for large-scale bioinformatics data-sharing. Fortunately, the technologies and governance practices of distributed computing yield opportunities for digitizing informed consent. This report considers *dynamic consent* on distributed systems from both theoretical and technical perspectives. We will first review the dominant consent theories and practices that pertain to bioinformatics research and related fields such as biobanking. We will discuss some of the challenges presented to frameworks of informed consent by big data technologies. We will then review some of the authorization tools and metadata ontologies developed to address these challenges. Finally, we will propose the DyCons architecture for the enabling of dynamic consent on Canadian distributed bioinformatics infrastructures.

Introduction

Reuse of biological data improves the statistical power of bioinformatics studies, and usually reduces their cost. Distributed data-sharing infrastructures ("federations"), such as the Canadian Distributed Infrastructure for Genomics (CanDIG)[i], enable the reuse of data between institutions, while permitting the data to live under the oversight of its home institution. Data federation eliminates the need for the centralization of data in off-site banks. Thus, data can be overseen closer to the point of data collection, closer to the research participant and their trusted clinical intermediaries. Distributed infrastructures also permit projects with strict security protocols to participate in data sharing that may be impermissible on centralized infrastructures, which tend to require greater assumptions of trust.

Ethical use of distributed data-sharing infrastructures requires that the rights of the research participant be accounted for in the architecture, implementation, governance, and use of these technologies. Respect for the right to informed consent has been a particularly challenging standard in the development of bioinformatics data-sharing networks. Since there is little direct participant involvement after the completion of data collection, and because of the administrative costs of maintaining contact with participants, data is often shared between research projects without the informed consent of the participant.

The theory of *dynamic consent* looks to leverage networked computing technologies to enable continuous participant oversight of their data. The technologies and governance practices of distributed computing contain unique opportunities for dynamic consent, such as the ability to collect data use authorization claims from remote sources.

Dynamic consent is essentially the application of contemporary technologies to consent theory, especially informed-consent ideals. Therefore we will begin with a description of consent theory and the key tenets of informed consent. Much of this background is attributed to Faden and Beauchamp's theoretical work on consent.

Informed consent theory and privacy

Moral justification for consent solicitation predominantly draws from two principles: consent as a practice of self-determination, corresponding to the principle of *autonomy*, or consent as a mechanism for the protection of well-being, corresponding to the principle of *beneficence*. The application of consent doctrines varies between disciplines, for example between practices of biomedical and sociological research.[ii] We will focus on consent theory as it applies to bioinformatics research involving human data.

Information is crucial to meaningful consent, both in its ethical and legal definitions. The consent-giver must have an adequate *understanding* of the proposed task, its possible consequences, and the available alternatives in order to make a decision that is meaningfully theirs (self-determination) and is in their subjective best interest (well-being). One challenge for the process of disclosure is that the set of information that is *material* in decision-making usually differs between individuals. The process of disclosure should begin by conveying a generic set of information about the protocol and purpose, but there should also be feedback mechanisms in place so that the solicitee can follow up on this initial disclosure with questions pertaining to their subjective interests. Such a disclosure may satisfy the subjectively material needs of the solicitee, ethical standards, and institutional standards, without lapsing into information overload.[iii]

One may have more or less understanding of a task when making a decision. One may also be more or less controlled by an influence when deciding. As requirements for informed consent, understanding and autonomy are continuums. They are required *substantially* rather than in full.[iv]

Throughout this work, we will refer to consents as being quanta of *authorization* information. This is meant in both the general and technical sense. Documented consents are records of social relationships; they record the levels of trust and bilateral expectations present in the act of data disclosure. Thus, consent information is crucial metadata for the privacy and security of data.

To consider privacy in more rigorous terms, we will benefit from Nissenbaum's notion of *contextual integrity* as a benchmark for privacy. Nissenbaum argues that information revealed in a context "is always tagged with that context and never 'up for grabs'".[v] In addition to the explicitly named terms of the data-disclosure event, *informational norms* determine the expectations attached to an instance of data disclosure. Since data-disclosure decisions are made according to these norms, violation of informational norms constitutes violation of contextual integrity, and therefore a violation of privacy.[vi] In other words, the data-collector and -users are bound both by the explicit terms and the normal expectations of the research participant's data-sharing decision.

Extending on the contextual integrity benchmark for privacy, Woolley posits that the solicitation of informed consent demonstrates the researcher's intention to understand, record, and respect the contextual integrity of the data collected. Thus, the solicitation of informed consent is the very basis for the trust relationship between the participant and the researcher.[vii] Failure to solicit informed consent, or to adhere to the explicit and implicit terms of consent, may result in mistrust of research initiatives.

We will now discuss the consent practices employed in bioinformatics research. The research participant is connected to a bioinformatics project by a network of clinical and administrative

documents and practitioners. In order to digitize consent practices, it is necessary to understand the purposes of these connections. It is also helpful to examine the aspects of consent practices that fall short of informed consent ideals. These may be loci for improvement when developing new practices.

Informed consent in bioinformatics

It is useful to distinguish between three types of bioinformatics research projects in terms of proximity to the site of data collection and degree of direct participant trust:

Project type (data user)	**Data collected from**	**Data use authorized by**	**Degree of data us**
Consented primary projects (C1º)	Participant	Participant (directly)	Primary data use
Consented secondary projects (C2º)	Primary project	Participant (directly)	Secondary data use
Waived-in secondary projects (W2º)	Primary project	Data Access Committee (DAC) of primary project	Secondary data use
Tertiary projects (3º)	Secondary project	DAC of secondary project, if at all	Tertiary data use

Fig.1: Classification of data users by level of data use authorization

Direct consent and participation solicitation

Bioinformatics research projects that solicit direct participant consent (C1º/C2º) generally recruit participants through one of two pathways. For lower-risk population-level datasets, they may solicit through public channels such as those used for marketing. For studies seeking participants with specific characteristics, calls for participants are often made to representatives of the characteristic-bearing persons, and participation is mediated by those representatives. One benefit of this mediation is that it reduces the burden knowledge upon participants, and often improves the materiality of the information that reaches them. When participants are sought on the basis of clinical characteristics, the clinician is the representative that is initially solicited. The clinician determines the set of patients with whom it would be appropriate to initiate direct consent solicitation.

The process of C1º/ C2º consent solicitation generally involves an additional intermediary, who is both the *informant* and the representative of the participant within the project. This intermediary, who is usually affiliated with the primary project itself, should be skilled at helping the participant arrive at substantial understanding of the research protocol and purpose. This intermediary engages in the dialogue of informed consent with the participant, participates in decision-making over which secondary or tertiary projects their data may be shared with, executes the processes triggered by a consent withdrawal, and remains the main point of contact for the participant. In bioinformatics research, this intermediary is often a genetic counsellor, who can play a *trusted* role because of their joint training in genomics and its psychosocial outcomes.

Supplemental 1 summarizes the actors and flows involved in consent solicitation through clinical channels. One shortcoming of clinician-mediated consent solicitation is that patients that are in

clinic due to disruptive or distracting illness may not be able to give adequately intentional, informed, or uncontrolled consent to research. Encouraging the patient to take time deciding, and to request further information, can make for a more consensual process.[viii] Dynamic consent, as an asynchronous and participant-configurable process, solicits consent at the time and place of the participant's choosing.

DACs and waiver exceptions to informed consent

In *Fig.1* we distinguish between projects that receive authorization via direct, specific consent and those that receive authorization in the spirit of a broad consent given at an earlier point in time. The latter form of authorization has some legal foundation in the *waiver exception* to the duty to obtain informed consent. In the waiver exception, the participant absolves their clinical trustee of their duty to obtain informed consent by waiving their right to active and informed participation in decision-making.[ix] As described in *Supplemental 2*, exercise of the waiver exception is ethically ambiguous.

In research, a participant that waives their right to autonomously authorize the secondary use of their biological materials entrusts a Data Access Committee (DAC) with the power to authorize on their behalf. This informed-consent-waiving process is referred to as *broad consent*. Often, the participant understands the general conditions under which the DAC will and will not give this authorization, and may have some say in setting these conditions. For example, they may check-mark "lung cancer research" but not "cancer research" as an acceptable purpose for secondary use.

In non-therapeutic research, requesting a consent waiver without offering feasible alternatives is ethically dubious. Ideally, there should be an option to be directly solicited for informed consent; otherwise, the participant must choose between participation in the project and their own autonomy. When primary projects only share the data of consent-waiving participants with secondary projects, or secondary projects only request the data of consent-waiving participants from primary projects, a duty towards informed consent is not fulfilled, as there is no effort to solicit informed consent for participation in the secondary project. There may be further injustice if there are demographic differences in the willingness to waive the right to informed consent. However, the research projects may simply not have the means to solicit meaningful and continuous informed consent, due to logistical concerns such as a limited number of staff. Thus it is important to develop accessible and Research Ethics Board (REB)-approveable tools that lower the cost on informed consent. The integration of dynamic consent systems into large-scale data-sharing infrastructures can accomplish this goal.

Tertiary projects may be waived-in, for example if they have therapeutic goals that align with the purpose of primary data collection. The participant's primary project may permit secondary projects to share data with tertiary projects so long as the protocol and purpose align with intention behind the participant's consent. Participants are frequently unaware of tertiary data uses. When consent is revoked, it is difficult to propagate the revocation to all tertiary data-users. By leveraging many of the technologies presently used in data-sharing among bioinformatics projects, dynamic consent systems enable research projects to automatically adhere to the consents of their participants.

Bioinformatics today

-Omics research is a ‘Big Data’ discipline. The large, raw data output by sequencing is not particularly informative. Very many samples are needed to distinguish between the unique and common bytes of this data. Common motifs are indicative of a shared evolutionary history, and possibly of codes for features important enough to have been conserved in the genome over many generations. If a region with a known function contains a unique motif, the motif may illuminate the biology underlying a disease.

Sharing and reuse of data provides more samples per study. This both improves the statistical power of the study and enables the observation of rare alleles that are unlikely to be observed in smaller cohorts.[x] Data sharing also permits more research to be performed per instance of data collection, reducing overhead on exploratory/pilot studies. These benefits, along with the general encouragement of open science by funders, have yielded interest in the development of data banks such as biobanks and, more recently, networked digital platforms for data sharing. These infrastructures enable data reuse at great magnitudes and speeds. Since it was infeasible for informed consent processes to be applied to data use at this scale, *broad consent* became the new bar for respect of research participant rights.

Broad consent (opt-out)

Initially developed in the biobanking context, broad consent involves the solicitation of consent to research towards a general purpose. Its goal is to reduce the administrative overhead involved in per-project ‘narrow consent’ solicitation. In the biobank, broad consent is given for a mission statement, and typically applies to both current and future projects covered by this statement. The DAC of the biobank commits to adhering to this mission statement in its data access authorization decisions, and judges applicant projects accordingly.[xi] Broad consent can be described as an *opt-out* model: the participant consents to DAC-approved research projects by default, and agrees to be responsible for opting out of projects with subjectively unsatisfactory protocols or purposes.[xii]

In terms of consent theory, participants cannot give meaningfully *informed* consent to research protocols that do not exist at the time of consent solicitation.[xiii] Although the participant is given the option to withdraw consent for specific projects, in order to make use of it they must be informed when a new project is approved for access to their data. If a participant’s contact information changes, they may not receive information about new authorizations. Further, in contrast to the centralized institution that governs the biobank, data sharing infrastructures may connect multiple institutions or projects with diverse mission statements. Thus, broad consent to a meaningfully specific mission statement is becoming less common, and participants are increasingly less aware of the practices and policies parametrizing access to their data.

Some proponents of broad consent posit that the purpose to which broad consent is given is not a specific mission statement, but more general *solidarity* with fellow humans.[xiv] Framed this way, a project protocol is covered by the broad consent so long as it is intended to benefit humanity and is of low enough risk to be approved by a decision-maker such as a DAC. From a similar perspective, Hansson argues that policies that require more narrow consent are more paternalistic: these policies

interfere with ability of research participants to make autonomous, broad decisions in the name of solidarity.[xv] Hansson's model does not account for the definition of autonomous acts as being *intentional* and done with substantial understanding. Whether it is possible to act intentionally through the passivity of the opt-out model, especially if one is not adequately informed of opportunities for action, should at least be a matter of contention. Moreover, whether a potential outcome is adequately solidaritious to justify potential risks is subjective, and this decision requires information of subjective materiality. For example, a study may intend to benefit one community yet may produce risks for another, and the involvement of the participant in the latter community may be a factor unknown to the DAC.

Rather than being justified in terms of the principle of respect for autonomy, broad consent could be considered in terms of an acceptable level of absence of consent. Multiple exceptions to the duty to obtain informed consent exist in both theories and policies of consent. Broad consent could perhaps be justified in terms of *waiver*[xvi] and *low-risk, high-benefit*[xvii] exceptions. Note however that these exceptions are functions of the level of risk, which may change over time, and that neither of these exceptions are guaranteed to protect the researcher or institution from liability.

Some people do prefer to give broad consent, while others prefer to give narrow consent.[xviii] Thiel et al. found that generally, people respond neutrally or positively towards better visibility and closer control of transactions made on their data.[xix] Similarly, in Flory and Emanuel's study, more information during the participation solicitation process did not reduce willingness to participate, nor overall participant satisfaction.[xx] Therefore, having options for narrow consent be available alongside broad consent should at least improve participant trust in data sharing infrastructures. It could also encourage selective participation in persons that would refuse when faced with an all-or-nothing broad consent request.

Risk

The risks involved in the collection and especially publication of -omics data have changed over time, and the governance of bioinformatics research is in the process of adjusting to these changes.

In Canada, exemptions from the obligation to obtain informed consent for research are usually granted on the basis of low risk.[xxi] Broad consent is often ethically and legally justified on this basis; if enrollment in additional research projects does not expose the participant to risk that exceeds the level that was broadly consented to, then these projects may be exempted from obtaining specific informed consent.[xxii] Participant anonymity is the mechanism for risk minimization that is most frequently referred to in these justifications.

Deidentified information differs from identifiable information by the removal of directly-identifying data, such as names. Anonymized information differs from deidentified information by the additional removal of indirectly-identifying data. However, increasing linkage between large databases, which contain data of varying levels of identifiability, makes it difficult for a single data collector to guarantee the full and future-proof anonymity of a data resource. Moreover, many formats of -omics data are difficult, if not impossible, to deidentify, given the innate uniqueness of an individual's genome. Therefore there is reason to doubt that anonymity will remain a substantially attainable

standard of privacy in bioinformatics research.[xxiii] At the same time, the increasing commercialization of data produces stronger incentives for violation of data privacy.[xxiv] These developments increase the level of risk involved in bioinformatics research participation over time.

When the risks associated with a research protocol change, REBs may request the solicitation of reconsent from the project's participants. Reconsent solicitation requires recontact. Up-to-date contact records are more likely to be kept by research projects that sustain continuous relationships with their participants, and may lack in projects that rely on broad consent. Difficulties with recontact result in high participant drop-out rates for projects faced with changing levels of risk. By enabling accessible active research participation, digital participant portals (such as dynamic consent portals) can reduce both the logistical burden of the reconsent process and the rate of recontact-related drop-out.[xxv]

State of the Art

As a result of the growing interest in digital infrastructures for the sharing of biological and clinical data, there is a need to digitize consent information. In clinical research, paper-based consent forms and data access applications are retained as records of interpersonal relationships that outline bilateral duties and rights. They may be referred to as a part of the process of judging new data access applications, or for informing an existing participant of past data transactions. Consents are therefore key metadata to the materials accessed by a research project.[xxvi] As autonomous authorizations, they inform whether or not a specific instance of data access is permissible. Respect for contextual integrity requires that the digitization of data be accompanied, as much as possible, by the digitization of the context of its disclosure. Therefore the digitization of -omics and related data calls for the digitization of consent records.

Coding consents

Consent and revocation records can be digitized in formats of varying fidelity. Images of paper-based consent documents can be made by C1º and C2º projects. These may be useful for the population of electronic Personal Health Records (ePHRs) but are mostly impractical for data sharing, as they are highly identifying. They are also difficult to aggregate and enforce as authorizations on larger infrastructures. Therefore an important problem in the digitization of consent is consent *coding:* the translation of consent information obtained during a human encounter into a standardized, and usually machine-readable, format. A small subset of the various technologies developed to address this problem are described below. There exists substantial interoperability between these and other standards.

Data Use Ontology (DUO)

DUO provides a hierarchical vocabulary for coding consents, data use requirements, and research purposes. Data resources can be tagged with these codes, enabling the implementation of data-discovery filters for matching the data-use intentions of the W2º/3º project.[xxvii]

Automatable Discovery and Access Matrix (ADA-M)

ADA-M Profiles constitute a robust generic data model for the governance of data resources. This metadata standard encodes data access authorizations as machine-readable fields and permits free-text elaborations upon these codes. A single data resource can have multiple authorization Profiles defined, each corresponding to a different data use. For example, different data-use authorizations may apply depending on whether the user is browsing the data for discovery purposes or has been approved to access the data for use in research.[xxviii]

Data-use application review

Consent codes can be used to reduce the triage burden on DACs by automatically disqualifying clearly ineligible data access requests.[xxix] This is one way to improve the efficiency of DAC review of W2º or 3º data-use applications with automation. Two other solutions for streamlining data-use application review are described below.

Resource Entitlement Management System (REMS)

REMS automates the workflows involved in the data access application process. Datasets can be catalogued with their associated licenses and data access applications. Prospective users can browse the catalogue and apply for access to datasets. REMS then circulates this application according to an owner-defined workflow and accrues reviews and judgements from the appropriate DAC members. Once a final judgement is made, it is communicated to the applicant. An approval is recorded as an E*ntitlement* to the requested dataset. REMS can be integrating into the authorization and authentication infrastructure of a data-sharing architecture to serve Entitlements as authorization metadata (user 'claims'). REMS can even be tasked with authorization policy decision-making and enforcement.[xxx]

Data Use Oversight System (DUOS)

DUOS semi-automatically mimics the data governance processes involved in matching W2º/3º projects with pertinent datasets, as well as the processes involved in DAC review of data access applications. A pilot trial comparing DUOS' automatic authorization decisions with those made by a DAC will investigate the outcomes of deeper automation of DAC review.[xxxi]

Dynamic consent

Dynamic consent combines the technologies and policies developed for informed consent, broad consent, and patient portals. Widespread digital literacy and networked technologies are leveraged to enable consent solicitation outside of the clinic, thereby reducing the administrative cost of narrow research consent. The movement of per-project consent solicitation from the clinic to the personal computer enables the participant to engage in making consent and refusal decisions in the setting of their choice, improving their capacity to act autonomously. Dynamic consent portals are also promising as opportunities for deepening participant engagement in research and reducing participant drop-out.[xxxii]

Enabling Consent and Revocation (EnCoRe)

The EnCoRe architecture is a comprehensive and highly extensible solution for enabling dynamic consent provision and revocation digitally-mediated data transactions, including biobanking. A browser plug-in informs the data subject of the data consumer's practices, including dynamic validation of the consumer's respect for the subject's past consent and revocation choices. Consent and revocation choices may be communicated to secondary data-users automatically using homomorphically encrypted 'sticky policies'. EnCoRe is especially thorough in its operational control, featuring automated compliance checking in its policy decision-making and enforcement components.[xxxiii]

Problem Motivation

The increasing commodification of data and the obsolescence of anonymity increases the risk inherent to participation in research projects involving the collection, analysis, and especially sharing of human biological data. Thus, bioinformatics research projects are no longer able to avoid the duty to obtain informed consent by appeal to low risk. Dynamic consent, a digital approximation of the clinical process of continuous informed consent, is a logistically feasible mechanism for the fulfillment of this duty.

In their theory of informed consent, Faden and Beauchamp posit that "in one important sense of the term, an 'informed consent' is an autonomous authorization by a patient or subject".[xxxiv] They also acknowledge a second sense of 'informed consent': "the rules governing informed consent in public policy and institutional contexts".[xxxv] One must integrate both senses of the term when designing tools for informed consent and its revocation. The DyCons model of dynamic consent on distributed infrastructures will predominantly consider the governance rules applicable to inter-provincially distributed bioinformatics infrastructures in Canada. Such considerations may substantially alter the specifications, and therefore the design, of similar models in other contexts, even if the theoretical principles are the same.

High-level principles

- Substantial autonomy (intentionality, understanding, absence of external control) [xxxvi]
  - Consents may be given in the formats that render them most meaningful [xxxvii]
    - The participant is able to consent broadly or narrowly
    - The participant is able to control their level of engagement
  - Research protocol information is made available to the participant according to their preferences
  - Consenting participants may be representatives of the data subjects rather than actual data subjects
- Respect for contextual integrity and transparency (as bases for trust)
  - The participant has access to information on all transactions on their data [xxxviii]
  - The participant is able to alter consent choices in reasonably *real time* [xxxix]
  - The participant's consent choices reliably and predictably factor into access control policy decision-making [xl]
  - The participant's consent choices and engagement preferences are enforced as system-wide obligations [xli]
  - The participant's consent choices are linked to their data and can be communicated to secondary/tertiary data users [xlii]
- Accessibility
  - Maintenance of human/paper-based alternatives to digital participant portals
  - Respect for existing research norms and REB standards, both at broad and local scales

- Strengthening of participant engagement in research [xliii]

Solution: DyCons architecture

We now present the DyCons architecture for enabling consent and revocation on distributed bioinformatics data-sharing infrastructures (DSIs). See *Supplemental 3* for information about the development of the DyCons reference implementation.

Distributed architecture

The DyCons architecture supports the distribution of actors, data, and queries. The architecture is thus extensible to research coalitions that span institutions subject to diverse regulatory standards and research norms. The architecture is split up into 3 *sites*, each of which contains the microservices and interface(s) pertaining to a particular set of actors in the bioinformatics research process. Research data and access control-related metadata may flow between these sites, within or across institutions.

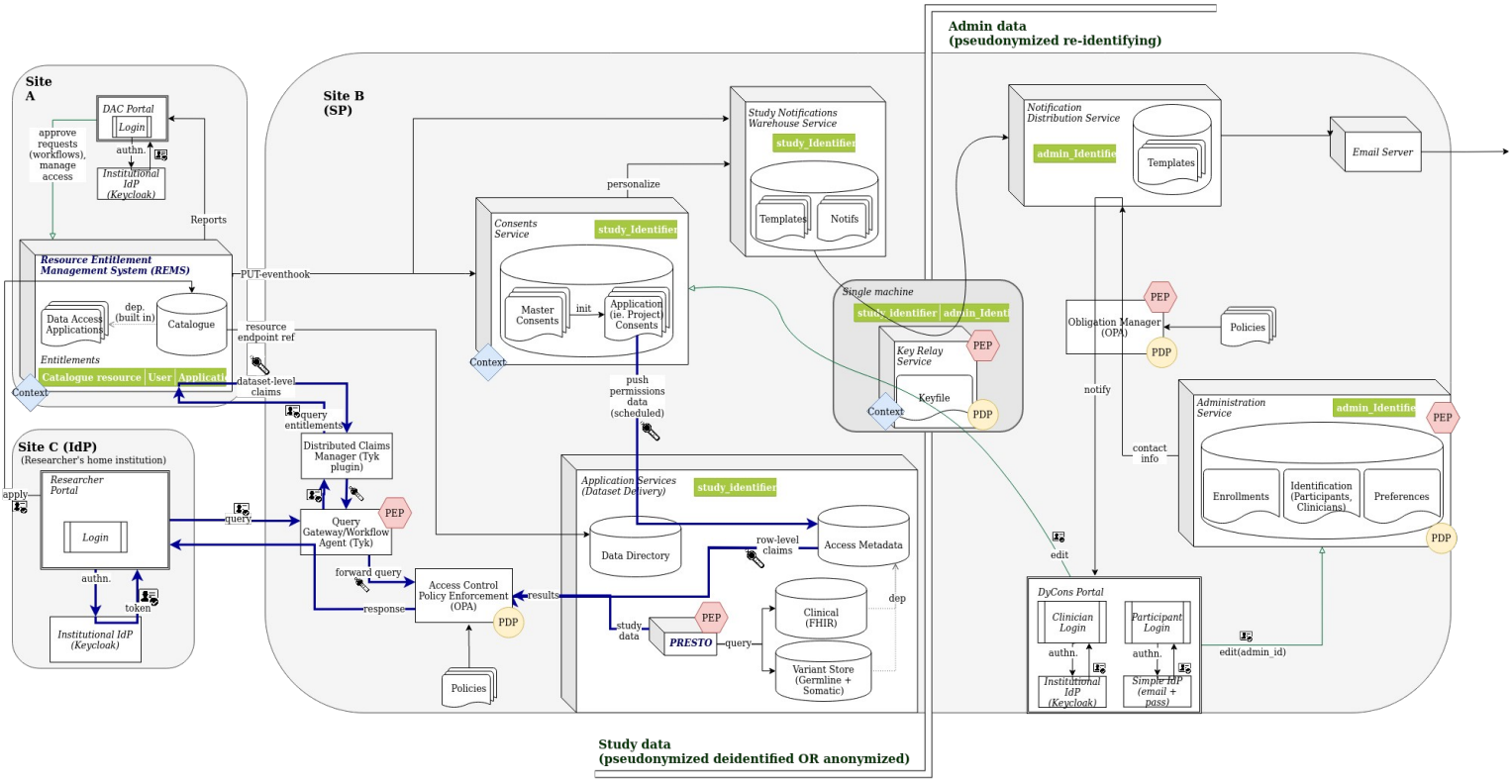

Fig.2: The DyCons architecture and major data flows

The primary project (C1º)

The DyCons solution specifically addresses the enabling of dynamic consent and revocation for secondary data use, so that more secondary projects can operate at a C2º standard. It is expected that the primary project that introduces a participant's data to the data-sharing infrastructure (DSI) has obtained consent to do so and has their own consent revocation mechanism in place. Therefore DyCons *never* intervenes when data is shared through the DSI between members of the C1º project. Such primary projects may be distributed, and desire to use the DSI internally rather than for sharing with secondary users. A consequence of this is that if a participant wishes to modify their consent as it pertains to their C1º project, they cannot do so through a DyCons interface.

Site A: DAC (C1º)

There can only be a single data access committee (DAC) per primary project, and the decisions of this committee must be communicable to all sites that host the project's data. *Site A* contains the collection of DAC decisions pertaining to secondary data access. The major component of this site is a Resource Entitlement Management System (REMS)[xliv] server, which handles the documents, licenses, and workflows involved in data access application and review. Applications are circulated among DAC members for review and judgement according to customizable workflows. The data access permissions of specific users known to the DSI's identity federation can be queried for by data servers on other sites. Upon application judgement, REMS can initiate workflows in other sites through event-hooks.

Site B: Data server (C1º)

Site B should be managed by the local research team that collects participant consents and data. It may federate calls to other *Site B*s hosted by other research teams. Various *Site B*s may even rely on the same DAC for management of secondary data use, so that the relationship between *Site A* and *Site B* is one-to-many.

This site is split into two domains: study data, which may be anonymized or deidentified, and consent and revocation (C&R) admin data, which is potentially re-identifying. Participants may access the DyCons dynamic consent interface through enrollment in the admin domain, but it is not mandatory for them to do so. They may participate in data sharing over the DSI anonymously with static broad or narrow consent. The two domains are connected by a secure relay service.

Study domain

Site B's study domain contains the core application services of the DSI, including the actual datasets under study. To respect contextual integrity [xlv], this data is linked to consent metadata through a mandatory dependency. Research subjects that do not initially consent to enrollment in the DyCons dynamic consent service may still give static consent that delimits secondary use of their data by a W2º standard of broad consent or broad refusal. They may later modify these initial consents following deidentified enrollment in DyCons, or continue to refuse enrollment and remain maximally anonymized.

The core application services contain a data directory that manages associations to access metadata. It packages the research data into discrete units (datasets) that can be governed by a DAC in REMS. The data directory also maintains the row-level dependencies upon consent metadata.

Queries to the application services are handled by the query workflow agent, which collects distributed claims from REMS and enforces access control policy. Queries carrying DAC-approved data access claims are then forwarded to an agent within the core application services, which fetches row-level participant authorizations (consents) and filters the response according to these fine-grain, dynamic permissions.

Consent information is stored in its own microservice and pushed to local storage in the core application services. In its latter form, it can be combined with other policies such as DAC overrides to fine-tune the row-level authorizations actually applied to a query. For the purposes of trust and respect for autonomy, use of such overrides is discouraged in most cases and DACs are expected to disclose any instances of override to affected participants.

Two levels of consents are supported in the consents service: master consents and project consents. Master consents define the default response to a new solicitation for data use and constitute mandatory metadata to the research data. Project consents constitute the row-level claims that are actually applied upon query by the pertinent secondary project. They are initially auto-populated from

the master consents and are regularly pushed to the core application services for use in row-level authorization. Both types of consents may be edited by C&R administrators (participants or their trusted clinical intermediaries.)

A team of notification services is split across the two domains. For the purpose of informing consent, they bundle and propagate information from secondary researchers or DACs to C&R administrators.

C&R administration domain

Site B's C&R administration domain contains the DyCons participant interfaces for dynamic consent. These interfaces are accessible from both a participant portal and a clinician portal. Although enrollment of clinical intermediaries in DyCons C&R administration may be costly, it is important to maintain alternative pathways for those with limited direct access to networked computerized technologies. It would be an issue of both justice and potential research bias to exclude such populations from being able to dynamically participate in bioinformatics research. The two portals differ in terms of user interface and identity provision (IdP) in authentication.

While clinicians may authenticate against an institutional IdP, participants do not have access to a trusted institution for claims provision. Therefore some re-identifying information must be stored within the C&R administration domain. In determining the level of pseudonymity to be employed here, tradeoffs may occur between security, accessibility, and trustworthiness. Email addresses are quite accessible but frequently insecure. Random tokens may be secure but inaccessible, and may be handled inappropriately as a result. In simulating a dynamic consent portal, Thiel et al. found participants to be uneasy about certain forms of public-domain multi-factor authentication,[xlvi] likely because it violated contextual integrity.[xlvii]

To ensure that re-identifying information is handled appropriately, all calls between the microservices of this domain are intercepted by an obligation manager.

Key relay service

The relay service shuttles calls between *Site B*'s two domains. It is a bottleneck that condenses the trust relationships between the microservices of the two domains for the purpose of reducing risk. Ideally, it should translate call *guid*s using a decryption key on a single machine. The bottleneck qualities of this service extend to its processing speed, especially if *guid*s are only rapidly searchable in one direction. Study -> C&R admin domain calls are higher-risk than those in the opposite direction, and are generally only made upon non-urgent notification (urgent notifications should move directly through clinical intermediaries in the C1º project, not through the data-sharing infrastructure.) Therefore it is acceptable to have these calls spend a long time in the decryption step at the relay service. Given that the arrival of notifying information following some study-domain event is itself informative, it may even be preferable to have incoming calls be additionally delayed at the relay for random amounts of time.

Another factor in relay design is whether it plays a call-forwarding or call-reformatting role. Given that the APIs and outgoing event-hooks of the communicating microservices may frequently change, it may be more secure to have the relay play a more involved role in intercepting and reformatting calls, so as to reduce the burden of trust between microservices.

Site C: Client (C1/ºC2º)

Site C contains the data user, who may be a primary or a secondary researcher, and their federation-trusted IdP. Secondary researchers may browse for research data through the REMS catalogue served by *Site A*s, and may send data access applications to these sites. Upon approval, data

retrieved from the *Site B*s of own or other projects may be temporarily stored in *Site C* for processing, depending on DSI design.

Discussion and conclusion

The DyCons architecture depends upon notification systems for the delivery of research protocol information to participants. However, the aspects of a research protocol that are subjectively material for a prospective participant's decision are difficult to predetermine. The disclosure of material information often occurs as the result of a *process* of information, where the participant is able to give feedback such as questions and the informant is able to integrate the feedback into their information.[xlviii] It is difficult to mimic this process with an architecture that linearizes the information loop into a unidirectional notification workflow. Enabling participant-clinician interfaces in the dynamic consent architecture is helpful in countering this effect, but it may be logistically difficult for clinicians to partake of this role. A cultural and financial shift towards the prioritization of dynamic consent practices would be necessary to fully enable clinician collaboration in the disclosure process. It may also be worthwhile to explore the fulfillment of a similar informant role by neural networks.

Kaye et al. point out the need for developments in the technical literature around operational control as it pertains to dynamic consent.[xlix] EnCoRe provides an exemplar architecture to look to in terms of operational control.[l] Homomorphic encryption and other types of sticky policies can help extend the life-cycle of a consent outside of the initial data server,[li] but more needs to be done in the development of strategies for the external enforcement of these autonomous authorizations, both in terms of technology and governance.

The DyCons architecture proposes a solution for the enabling of dynamic consent on distributed Canadian bioinformatics infrastructures. Distributed infrastructures build in data-stewardship options that centralized infrastructures lack, including local data storage and custom policy enforcement. At the same time, they introduce broader governance difficulties that cannot be remedied by purely technical solutions. Technical, ethical, and regulatory literature on dynamic consent and data federation must continue to evolve in conjunction.

Supplemental

S.1: Clinical consent flows for secondary data-use

Diagram of the processes of consent solicitation for C2º clinical bioinformatics research. The process for C1º projects is often similar, but with additional enrollment steps and a protocol for the sample collection process.

S.2: Ethical ambiguity of the waiver exception to informed consent

Exercise of the waiver exception is ethically ambiguous. A participant may waive their right to informed consent because they do not trust their own understanding of the research protocol and purpose, or due to controlling influence from the clinical trustee. Rather than to absolve the trustee of the duty to obtain this consent, the ethically appropriate action is to improve understanding, or to reduce the clinician's control, so as to enable the exercise of substantially informed consent/refusal.

On the other hand, a participant may have a strong interest in receiving care or participating in research, yet a strong disinterest in participating in the processes of decision-making. After being informed of the general risks of sharing their data, they may feel that the risks are broadly low enough to warrant the deferral of decision-making to a trustee. Alternately, in clinic-mediated research, they may be aware of their reduced capacity for autonomous action (for example due to significantly distracting pain) and prefer the role of research beneficiary over autonomous participant.

S.3: Reference implementation of the DyCons architecture

Development of the reference implementation of DyCons lives at github.com/dycons. See the design repository for sequence diagrams of major events, as well as OpenAPI/Swagger specifications of the component microservices.

Bibliography

Árnason, Vilhjálmur. “Coding and Consent: Moral Challenges of the Database Project in Iceland.” *Bioethics* 18, no. 1 (2004): 27–49. https://doi.org/10.1111/j.1467-8519.2004.00377.x.

“Broad Data Use Oversight System.” Accessed April 26, 2020. https://duos.broadinstitute.org/home_about.

Erlich, Yaniv, James B. Williams, David Glazer, Kenneth Yocum, Nita Farahany, Maynard Olson, Arvind Narayanan, Lincoln D. Stein, Jan A. Witkowski, and Robert C. Kain. “Redefining Genomic Privacy: Trust and Empowerment.” *PLOS Biology* 12, no. 11 (November 4, 2014): e1001983. https://doi.org/10.1371/journal.pbio.1001983.

Faden, Ruth R., and Tom L. Beauchamp. *A History and Theory of Informed Consent*. New York, NY: Oxford University Press, 1986.

Flory, James, and Ezekiel Emanuel. “Interventions to Improve Research Participants’ Understanding in Informed Consent for Research: A Systematic Review.” *JAMA* 292, no. 13 (October 6, 2004): 1593–1601. https://doi.org/10.1001/jama.292.13.1593.

Hansson, Mats G, Joakim Dillner, Claus R Bartram, Joyce A Carlson, and Gert Helgesson. “Should Donors Be Allowed to Give Broad Consent to Future Biobank Research?” *The Lancet Oncology* 7, no. 3 (March 1, 2006): 266–69. https://doi.org/10.1016/S1470-2045(06)70618-0.

Kaye, Jane, Edgar A. Whitley, David Lund, Michael Morrison, Harriet Teare, and Karen Melham. “Dynamic Consent: A Patient Interface for Twenty-First Century Research Networks.” *European Journal of Human Genetics* 23, no. 2 (February 2015): 141–46. https://doi.org/10.1038/ejhg.2014.71.

Linden, Mikael, Tommi Nyrönen, and Ilkka Lappalainen. “Resource Entitlement Management System.” Maastricht, Netherlands: TERENA, 2013. http://tnc2013.terena.org/getfile/870.

Mont, Marco Casassa, Vaibhav Sharma, and Siani Pearson. “EnCoRe: Dynamic Consent, Policy Enforcement and Accountable Information Sharing within and across Organisations,” n.d., 79.

Nissenbaum, Helen. “Privacy as Contextual Integrity Symposium - Technology, Values, and the Justice System.” *Washington Law Review*, no. 1 (2004): 119–58.

Prainsack, Barbara, and Alena Buyx. “A Solidarity-Based Approach to the Governance of Research Biobanks.” *Medical Law Review* 21, no. 1 (March 1, 2013): 71–91. https://doi.org/10.1093/medlaw/fws040.

“The Data Use Ontology.” Accessed December 21, 2019. https://www.ebi.ac.uk/ols/ontologies/duo.

Thiel, Daniel B., Jodyn Platt, Tevah Platt, Susan B. King, Nicole Fisher, Robert Shelton, and Sharon L. R. Kardia. “Testing an Online, Dynamic Consent Portal for Large Population Biobank Research.” *Public Health Genomics* 18, no. 1 (2015): 26–39. https://doi.org/10.1159/000366128.

Thorogood, Adrian. “Canada: Will Privacy Rules Continue to Favour Open Science?” *Human Genetics* 137, no. 8 (August 1, 2018): 595–602. https://doi.org/10.1007/s00439-018-1905-0.

“Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans – TCPS 2 (2018),” April 1, 2019. https://ethics.gc.ca/eng/policy-politique_tcps2-eptc2_2018.html.

Woolley, J. Patrick. “Trust and Justice in Big Data Analytics: Bringing the Philosophical Literature on Trust to Bear on the Ethics of Consent.” *Philosophy & Technology* 32, no. 1 (March 1, 2019): 111–34. https://doi.org/10.1007/s13347-017-0288-9.

Woolley, J. Patrick, Emily Kirby, Josh Leslie, Francis Jeanson, Moran N. Cabili, Gregory Rushton, James G. Hazard, et al. “Responsible Sharing of Biomedical Data and Biospecimens via the ‘Automatable Discovery and Access Matrix’ (ADA-M).” *Npj Genomic Medicine* 3, no. 1 (July 23, 2018): 1–6. https://doi.org/10.1038/s41525-018-0057-4.

i Thorogood, “Canada.”
ii Faden and Beauchamp, *A History and Theory of Informed Consent*.
iii Faden and Beauchamp.
iv Faden and Beauchamp.
v Nissenbaum, “Privacy as Contextual Integrity Symposium - Technology, Values, and the Justice System,” 143.
vi Nissenbaum, “Privacy as Contextual Integrity Symposium - Technology, Values, and the Justice System.”
vii Woolley, “Trust and Justice in Big Data Analytics.”
viii Faden and Beauchamp, *A History and Theory of Informed Consent*.
ix Faden and Beauchamp.
x Hansson et al., “Should Donors Be Allowed to Give Broad Consent to Future Biobank Research?”
xi Prainsack and Buyx, “A Solidarity-Based Approach to the Governance of Research Biobanks.”
xii Kaye et al., “Dynamic Consent.”
xiii Árnason, “Coding and Consent.”
xiv Prainsack and Buyx, “A Solidarity-Based Approach to the Governance of Research Biobanks.”
xv Hansson et al., “Should Donors Be Allowed to Give Broad Consent to Future Biobank Research?”
xvi Faden and Beauchamp.
xvii “Tri-Council Policy Statement.”
xviii Kaye et al., “Dynamic Consent.”
xix Thiel et al., “Testing an Online, Dynamic Consent Portal for Large Population Biobank Research.”
xx Flory and Emanuel, “Interventions to Improve Research Participants’ Understanding in Informed Consent for Research.”
xxi “Tri-Council Policy Statement.”
xxii Hansson et al., “Should Donors Be Allowed to Give Broad Consent to Future Biobank Research?”
xxiii Erlich et al., “Redefining Genomic Privacy.”
xxiv Woolley, “Trust and Justice in Big Data Analytics.”
xxv Kaye et al., “Dynamic Consent.”
xxvi Woolley, “Trust and Justice in Big Data Analytics.”
xxvii “The Data Use Ontology.”
xxviii Woolley et al., “Responsible Sharing of Biomedical Data and Biospecimens via the ‘Automatable Discovery and Access Matrix’ (ADA-M).”
xxix Woolley et al.
xxx Linden, Nyrönen, and Lappalainen, “Resource Entitlement Management System.”
xxxi “Broad Data Use Oversight System.”
xxxii Kaye et al., “Dynamic Consent.”
xxxiii Mont, Sharma, and Pearson, “EnCoRe: Dynamic Consent, Policy Enforcement and Accountable Information Sharing within and across Organisations.”
xxxiv *A History and Theory of Informed Consent*, 3.
xxxv Faden and Beauchamp, *A History and Theory of Informed Consent*.
xxxvi Faden and Beauchamp.
xxxvii Kaye et al., “Dynamic Consent.”
xxxviii Kaye et al.
xxxix Kaye et al.
xl Mont, Sharma, and Pearson, “EnCoRe: Dynamic Consent, Policy Enforcement and Accountable Information Sharing within and across Organisations.”
xli Mont, Sharma, and Pearson.
xlii Kaye et al., “Dynamic Consent.”
xliii Kaye et al.
xliv Linden, Nyrönen, and Lappalainen, “Resource Entitlement Management System.”
xlv Nissenbaum, “Privacy as Contextual Integrity Symposium - Technology, Values, and the Justice System.”
xlvi Thiel et al., “Testing an Online, Dynamic Consent Portal for Large Population Biobank Research.”
xlvii Nissenbaum, “Privacy as Contextual Integrity Symposium - Technology, Values, and the Justice System.”
xlviii Faden and Beauchamp, *A History and Theory of Informed Consent*.
xlix Kaye et al., “Dynamic Consent.”
l Mont, Sharma, and Pearson, “EnCoRe: Dynamic Consent, Policy Enforcement and Accountable Information Sharing within and across Organisations.”
li Kaye et al., “Dynamic Consent”; Mont, Sharma, and Pearson, “EnCoRe: Dynamic Consent, Policy Enforcement and Accountable Information Sharing within and across Organisations.”