DyCons: Enabling Dynamic Consent on Distributed Computing Infrastructures

KAT PAVLOVA[1], JONATHAN DURSI[1], GUILLAUME BOURQUE[2], MICHAEL BRUDNO[1]

1: University Health Network; 2: Department of Human Genetics, McGill University

DyCons extends distributed data-sharing infrastructures to support dynamic consent

Efforts like GA4GH enable the sharing and reuse of human medical and genomic data, which can greatly accelerate health research. However, it can be prohibitively expensive to gather paper-based and clinician-mediated participant consent for each instance of data reuse.

Dynamic consent leverages developments in access to internet and digital literacy to involve, inform, and empower research participants in their data-use decision-making.

DyCons is an architecture for dynamic consent that is developed to address the specific needs of distributed computing infrastructures for bioinformatics research. It can be scaled to handle many projects that share data across institutions, and even between provinces, all within the unique constraints of their own local access control.

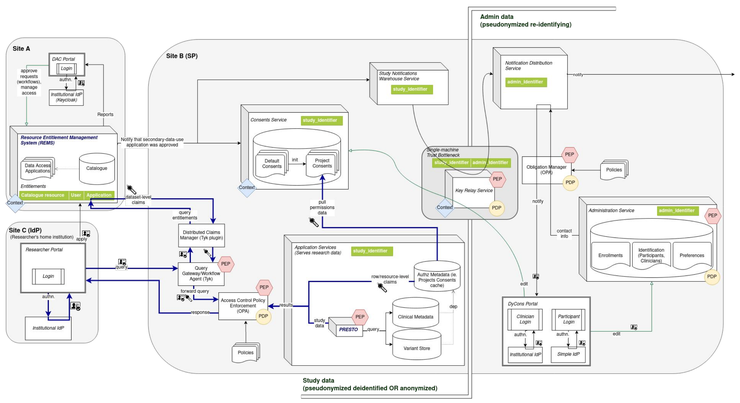

Fig. 1: The initial architecture for DyCons. Data Access Committees (DACs) and data users may be located far away from the research data storage and servers, in different institutions or even provinces. Queries on the data are authorized using data access claims collected from both local and remote services.

Each human role in the consent solicitation and data sharing process has a unique entrypoint into the system

Each class of user portal (DAC member, participant, clinician, or researcher) is restricted to communications with a specific subset of DyCons services. This minimizes data visibility and supports multi-site projects, which may have data storage and stewards residing in locations far from their governing DACs.

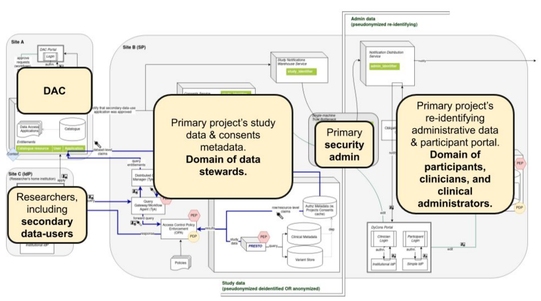

Fig. 2: The domains represented in the DyCons architecture. Each domain is expected to be physically separated, and most of them feature their own distinct user portals.

De-identified study data and consent metadata are physically separated from identifying administrative data

The physical separation of study data from administrative data ensures that a data breach in one domain does not compromise data in the other domain. The study and administration domains are connected by a carefully secured single-machine relay service, for the purposes of notification and participant consent-editing. The relay service reduces the burden of trust between services across domains.

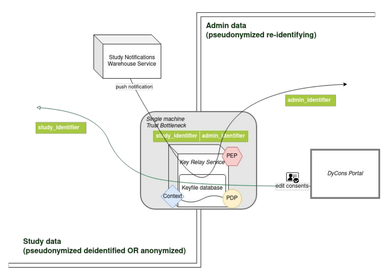

Fig. 3: The study and administration domains of the data-serving application are physically separate. A relay service enables and controls communication between them.

Authorizing queries: consent-as-authorization and distributed claims

Distributed claims from DAC entitlement lists inform dataset-level authorization decisions

Data Access Committee (DAC) decisions on data-access applications can be used to determine the entitlements of a prospective data-user to a dataset. These claims can be collected from services such as REMS (the Resource Entitlement Management System by CSC – IT Center for Science Ltd.)

Since multiple data-serving sites may be governed by a single DAC, DyCons treats the DAC claims server as remote to the core application services. These claims are therefore retrieved and managed as distributed claims.

Participant consents inform resource/row-level authorization decisions

Once the DAC approves a project for use of a dataset, the project-specific consents of each participant in the dataset are initialized from their default preferences. For example, opt-out preferences grant consent by default, while opt-in preferences refuse consent by default.

Project consents may remain static (in the case of anonymized data), or may be dynamically modified by the participant or their clinician (in the case of de-identified pseudonymized data). In all cases, the consents are used to make access authorization decisions on the participant's data.

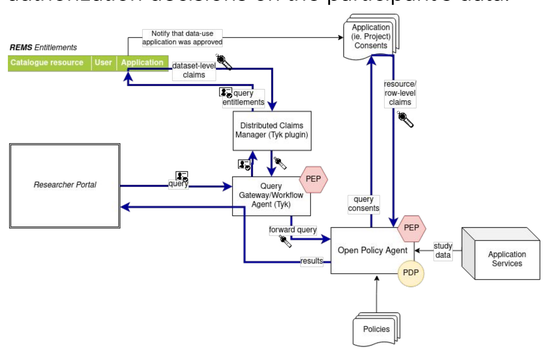

Fig. 4: Context for query authorization decision-making is collected from both local and remote sources, including: DAC entitlements lists, dynamic or static participant consents, and policies defined by data stewards and administrators.

Keeping research data in the context of its consent

In designing and implementing dynamic consent systems, it is important to keep research data close to its consent metadata. However, it is also important to keep some distance between them, such that consents may be accessed and edited by participants without risking the compromise of sensitive health information.

DyCons synthesizes these considerations by housing consent metadata and research data in seperate microservices, but enforcing strong dependencies between them during data registration and query authorization.

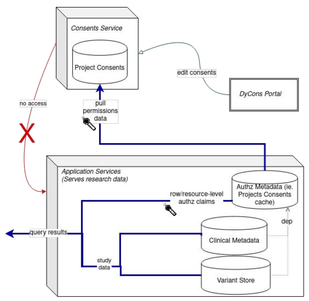

Fig. 5: A Consents Service enables the storing and modification of consents, away from the Application Services and the sensitive data within them. Consent information is pulled from the Consents Service into the Application Services, but there is no flow of data in the opposite direction.

Current and future work

A prototype implementation of the DyCons architecture is currently in development. This work will improve our understanding of the technical opportunities and challenges of distributed dynamic consent. At the same time, we continue to work closely with a distributed bioinformatics infrastructure on practices and technologies of data stewardship that respect participant autonomy.

DyCons is being developed as the model for dynamic consent on CanDIG (the Canadian Distributed Infrastructure for Genomics).

More information available in the "National Health Research for a Federal Canada - CanDIG in 2020" poster by Jonathan Dursi et al.

By enabling dynamic participant information and decision-making, DyCons can facilitate the broader sharing of health data, accelerating health research.