Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Data access management system for relational accountability orchestration

Meeri Hakala
CSC - IT Center for Science (Finland)
Jaakko Leinonen
CSC - IT Center for Science (Finland)
Miska Husgafvel
CSC - IT Center for Science (Finland)
Miikka Kallberg
CSC - IT Center for Science (Finland)
Jarno Laitinen
CSC - IT Center for Science (Finland)
Ilkka Lappalainen
CSC - IT Center for Science (Finland)
Janne Lauros
CSC - IT Center for Science (Finland)
Heikki Lehväslaiho
CSC - IT Center for Science (Finland)
Mikael Linden
Gofore Finland Oy
Martin Matthiesen
CSC - IT Center for Science (Finland)
Dylan Spalding
CSC - IT Center for Science (Finland)
Juha Törnroos
CSC - IT Center for Science (Finland)
Tommi Jalkanen
Bruvo AI
Markku Rontu
Nitor Group Oy
Minttu Sauramo
Helsinki University Hospital
Tiina Wahlfors
Finnish Institute for Health and Welfare
Melissa Konopko

ELIXIR
Martin Kuba
Masaryk University
Bruno Pacheco
LNDS Luxembourg National Data Service
Oliver Hofmann
The University of Melbourne
Conrad Leonard
The University of Melbourne
Andrew Patterson
The University of Melbourne
Bernard Pope
The University of Melbourne
Tracey Godfery
University of Otago
Matt Pestle
University of Auckland
Claire E. Rye
University of Auckland
Tommi H. Nyrönen
nyronen@csc.fi

CSC - IT Center for Science (Finland)

Article

Keywords:

Posted Date: October 22nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10915605/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

While the scientific community has improved the findability and interoperability of research data, access to sensitive datasets remains a major bottleneck. Accessing data from human subjects often depends on interconnected, manual, and non-standardised administrative workflows that are incompatible with large-scale data federation. The Resource Entitlement Management System (REMS) defines data access request, review, and authorisation processes through configurable workflows; standardising complex procedures while ensuring auditability. REMS and its application workflow engine translate ethical, legal, and cultural requirements into machine-readable permissions. The system separates identity from access rights and coordinates accountability among data applicants, authorisation bodies, and data holders. REMS uses international standards to automate generation of application forms. It reduces administrative overhead and supports local and national data sovereignty. REMS does not harmonise national law, assess scientific merit or disclosure risk, or enforce secure processing environments. However, the use cases across diverse regulatory and federated settings in Europe, Australia, and New Zealand demonstrate versatility of the software to deploy trustworthy access control on sensitive data.

Introduction

In the decade since the formalisation of the FAIR (Findable, Accessible, Interoperable, and Reusable) principles [1], the scientific community has made significant strides in making high-value research data findable and interoperable. However, sensitivity of data like human genomic sequences, clinical health records and longitudinal linguistics the “A” or Accessible in FAIR imposes data protection requirements, and it is a significant bottleneck [2]. Access to high-quality data constrains the development of artificial intelligence models based on these regulated data, which is particularly important in developing predictive systems for improving human health [3]. While data may be findable via metadata catalogues, the actual process of gaining access is often via manual, non-standardised administrative workflows that fail to scale [2, 3] with the increasing volume of global data and the growing demand for its use.

Privacy laws worldwide, especially the General Data Protection Regulation (GDPR) [4] within the European Research Area and beyond, have been the central enforcement towards these principles. Specifically, the principle of data privacy by design (GDPR Article 25) mandates that organisations integrate data protection into the very heart of their processing activities [4, 5]. In the context of data access, this requires infrastructure that doesn't merely log data access permissions, but actively enforces data sharing through secure, indisputable, auditable, and transparent workflows. For research infrastructures, the goal is to transition from a reliance on manual oversight to a system of trusted tokens, where the researcher's identity and the scope of their permission are machine-readable and verifiable across borders and legislative domains.

A complementary set of core principles CARE (Collective benefit, Authority to Control, Responsibility, and Ethics) Principles for indigenous data governance [6] emerged and gained traction since they were first published in 2020. Developed out of the Research Data Alliance’s indigenous data sovereignty interest

group, the CARE principles shift data governance from a data-centric approach to one that is people and purpose focused. These principles have helped move data and research practices away from potentially exploitative and harmful approaches toward those that support Indigenous well-being and self-determination through control, sovereignty, and the building of relationships of trust. Taken together, the FAIR and CARE principles enable the responsible advancement of data-intensive research in a culturally respectful context with significant effort dedicated to operationalising their alignment in practice.

Building societal and ethical trust in managing sensitive data requires connection between digital solutions and legal framework, as well as established relationships. Trust is built between researchers and participants, infrastructure providers and community leadership. Examples of this kind of collaboration are managing genomic data of indigenous peoples such as Māori in Aotearoa, New Zealand, as part of studies such as Rakeiora [7, 8] and minority populations in the European 1 + million genomes initiative [9].

Due to these legal, technical and societal factors, the process of sensitive data access governance is complex and interconnected. The data access applicant, the data access authorisation body, and the data holder exist in a state of interconnected mutual dependency, in an “administrative three-body problem” (Fig. 1). Unless interactions between the three bodies are coordinated well, the access process can enter an administrative gridlock. The process typically begins with the applicant, who must provide verified identity before submitting a purpose-bound data access application. This application defines the specific justification and intended data processing scope. The data access authorisation body (e.g. ethical review boards, data controllers, or data access committees assigned by the data controllers) then evaluates the request. Their role is to establish the legal and ethical, and in many cases, cultural basis for data use, ensuring alignment with regulatory frameworks such as GDPR Articles 6 and 9 [4] or CARE principles [6]. This body issues the formal access decision, often accompanied by specific usage restrictions or obligations that the other actors must satisfy. Finally, the data holder (or digital service provider who makes the data computationally accessible) serves as the technical enforcement point. Rather than simply providing the datasets for download, data access typically needs to happen within an infrastructure service the user has access to. The data holder is responsible for the enforcement of data access according to the agreed purpose, and for the maintenance of usage logs, including sensitive data movement in and out of the environment, as per the agreed purpose. If a discrepancy is found the access decision should be withdrawn, and a new application made and authorised by the data access authorisation body.

The Resource Entitlement Management System (REMS) has emerged as a critical piece of open-source cyberinfrastructure that coordinates between these dependencies and their practical, production-level implementation [10]. REMS stabilises interactions by giving the data applicant, authorisation body and data holder a shared, traceable state for reaching and communicating an access decision. The interconnected system of data access governance requires supporting technology, and such technology is a prerequisite to implement FAIR and CARE principles on sensitive data sharing. To help address the administrative problems, the Global Alliance for Genomics and Health (GA4GH) has developed a suite of

standards designed to facilitate secure, interoperable data sharing [11] ("I"—Interoperable in FAIR) that scales globally. Established data access management standards among these are the GA4GH Data Use Ontology (DUO) [12] and the GA4GH Passport standard [13]. While the DUO provides a machine-readable language for data use restrictions, the GA4GH Passport (and its constituent "Visas") provide a standardised framework for communicating a researcher's specific data access authorisations with their digital identity to various platforms. These standards have been developed in sync with the regulatory and ethics experts to bridge between policy and technical implementation [11, 14, 15]. Implementation of GA4GH standards in REMS, and its orchestration capabilities of research purposes, applicant information, and governance requirements stabilises interactions needed for controlled data access. REMS provides configurable levels of automation while keeping human oversight (e.g. for a DAC) intact. It uses DUO codes to automate matching of an applicant's intended research against dataset consent restrictions. Once a DAC approves a request, REMS issues a GA4GH Visa and logs the entire process for auditing.

By serving as a centralised orchestrator, REMS serves as the common framework to provide standardised access authorisation metadata necessary to prevent administrative gridlocks. Originally developed to manage access to valuable biomedical and soon after linguistic datasets in Finland [16], REMS provides a generic workflow engine that translates ethical, legal and cultural requirements into actionable permissions. The benefit of a standards-based approach supporting different data types, legal compliance, user trust and sustainability is demonstrated by the implementations of REMS reported in this article. REMS is part of the ELIXIR Finland development programme at the CSC - IT Center for Science in Finland [10, 17].

A key goal of FAIR is the reduction of administrative friction. Several of the controlled-access repositories already have software for managing the access process [18]. These are usually fully integrated and proprietary, such as the case with EGA and dbGaP [19, 20]. While other open software exists, such as systems developed by the Broad Institute [15], these tools typically have dependencies to the processes of the hosting institute [18]. REMS allows repositories and institutions to deploy, host and flexibly control their own instances independently on their chosen technical platform [10, 17, 18]. As the standards-based REMS software scales within and across systems, it supports greater ecosystem interoperability by providing RESTful interfaces and a unified user experience.

In this article, we detail the technical architecture of REMS as implemented in 2026. REMS provides implementing organisations a tool for metadata-driven data access orchestration. To demonstrate the versatility of this technology across diverse regulatory and cultural landscapes, we present implementation in four distinct use cases: the European 1 + million genomes Genomic Data Infrastructure (GDI) project [21], the Aotearoa Genomic Data Repository in Aotearoa [23], Finnish biobank biological data at The Finnish Institute for Health and Welfare, and Language Bank of Finland for global use [16]. Through these examples, we show how REMS enables a FAIR and CARE data ecosystem by making "Access" as standardised and interoperable as "Findable", as well as supporting authority to control access, embed responsibilities and maintain ethics.

Methods

The Resource Entitlement Management System is an open-source product with permissive MIT licensing that provides a metadata-driven architecture to technically enable implementation of relational accountability for data access management. By functioning as a centralised orchestrator, the system manages the complex interdependencies between data applicants, access authorisation bodies, and data holders. The system separates identity management and access rights management, using international standards to automate data access form generation and issue machine-readable data permissions. The process of how REMS gathers access authorisation information is immutable and secure, making it a good tool for authorisation bodies like data access committees (DACs) to grant or withdraw access permissions. The system's adaptability is demonstrated across diverse regulatory landscapes and resources, proving its capability to enforce access programmatically, reduce administrative overhead, and support both global science and local data sovereignty.

System Architecture Overview. REMS is implemented as a web-based application to couple digital identities of data applicants to digital resource identifiers. It is designed to securely manage, automate, and provide an indisputable audit trail of access rights management to controlled resources, such as sensitive biomedical or academic datasets (Fig. 2., Fig. 3.). A central module of the REMS architecture is its dynamic workflow engine, which handles multi-stage, branching approval processes. Administrators can configure custom state-machine workflows to mirror the specific governance requirements of different DACs. The workflow engine evaluates the immutable event log to calculate the current state of any given application dynamically. It supports complex logic, including multi-party reviews, third-party consultations, and automating parts of application processing, such as reviewer invitations or approvals. The application adheres to a decoupled client-server architecture, strictly separating the user interface from the core data processing and access control logic. To ensure high reliability, concurrency, and maintainability, the system is fundamentally grounded in functional programming paradigms, utilising the Clojure ecosystem for both server-side and client-side operations. An extensive application programming interface (API) provides means to further extend or integrate REMS into existing infrastructure, and extensions can be implemented with any programming language.

Server-Side Core Backend Infrastructure and Data Persistence. The REMS backend acts as the authoritative component for business logic to encode the access rights of an applicant as machine-readable resource access permissions called entitlements. These entitlements are always tied to identified individuals. In this paper, entitlement means a right to do or have, i.e., permission to access a resource with a controlled access process. Implementations of REMS manage access to e.g. sensitive datasets, but a resource could also be any other digital asset, like a computing service. REMS is developed in Clojure, executing on the Java Virtual Machine. State persistence is governed by a PostgreSQL relational database (Table 1.). Rather than relying solely on traditional Create, Read, Update, Delete operations, the system's core design heavily incorporates event sourcing principles. Every state change within an application, for example a Data Access Applicant submitting a form, a Data Access Reviewer requesting modifications, or a DAC member granting approval, is recorded as an immutable

event in the REMS database. Each event depends on the previous event. This append-only event log guarantees an exhaustive and reliable audit trail, which is mandatory for compliance with strict global privacy regulations (e.g., GDPR).

Table 1
REMS data record types.

Data Record	Role in REMS	Storage Mechanism
Events	Immutable log of state changes for an application (e.g. submitted, approved).	Appended sequentially in PostgreSQL event tables with tampering prevention.
Applications	The data access request itself. State is reconstructed by replaying the event log.	In-memory cache (Clojure state), rebuilt from PostgreSQL events.
Resources	The underlying datasets or assets the user requests access to.	Standard relational PostgreSQL tables.
Catalogue Items	The UI-facing representation of a resource that users browse and apply for.	Relational PostgreSQL tables, linked to Workflows and Forms.
Workflows	The review process definitions, including handlers (reviewers/approvers).	Relational PostgreSQL tables.
Forms & Fields	The dynamic questionnaires applicants must complete to request access.	Relational PostgreSQL tables defining form structure and rules.
Licences	Terms of use attached to a resource that an applicant must accept.	Relational PostgreSQL tables.
Entitlements	Granted access rights given to a user once an application is approved.	Relational PostgreSQL tables (often synced to external registries).
Users & Orgs	Identities of applicants and handlers, usually federated via OIDC.	Relational PostgreSQL tables, augmented by OIDC ID tokens.

The primary output of the backend is the generation of machine-readable entitlements (using e.g. GA4GH standards). Once a workflow reaches an approved state, REMS binds the identity of the applicant to the requested resource. These entitlements are structured to be queried by external enforcement points via APIs. Examples of endpoints provided by REMS implementations are presented in Table 2.

Table 2
Projects, repositories and commons that utilise REMS along with the data types and links to the public instances for technical validation.

Project / Repository / Commons	Data Types	REMS Instance Endpoint	Technical Validation Focus
THL Biobank (Finland)	Human Health, Clinical and phenotype	https://thl-biobank.elixir-finland.org	National Scale: Handling thousands of requests for clinical and biobank samples.
The European Genomic Data Infrastructure (GDI)	Human Genomic and phenotype	https://portal.gdi.lu	Interoperability: Federated access management across Europe.
CSC Sensitive Data Services	Health and generic sensitive	https://sd-apply.csc.fi/	Federated Access: European Genome-phenome Archive in Finland
FIN-CLARIN Language Bank of Finland	Language resources, corpora	https://www.kielipankki.fi/access	Production Longevity: Managing restricted linguistic data since 2015.
Rakeiora Genomics Platform (Aotearoa, New Zealand)	Human Genomic and Clinical	https://apply.rakeiora.ac.nz/ and https://browse.rakeiora.ac.nz	Indigenous Data Sovereignty: Implementing Māori Kaitiaki governance workflows.
Aotearoa Genomic Data Repository (AGDR)	Non-human Genomic	https://apply.agdr.org.nz/	Protected biological and environmental genomic samples.
Language Data Commons of Australia	Language	https://rems.ldaca.edu.au/	High-value collection: Linguistically diverse regions across Australia.
Coordinated Access for Data,	Social Sciences and	https://cadre.ada.edu.au/	Five Safes Framework: Distributed

Project / Repository / Commons	Data Types	REMS Instance Endpoint	Technical Validation Focus
Researchers and Environments (CADRE)	related fields		access management platform.
Childhood Cancer Insights portal (ZERO)	Cancer	https://insightsdev.zerochildhoodcancer.cloud/	Data selection and minimisation: Restful API integration
Garvan Institute of Medical Research	Human OMICS and Clinical	https://rems.public.garvan.org.au/	Sensitive human datasets: patient cohorts.
Australian Cardiovascular Disease Data Commons (ACDC)	Human OMICS and Clinical	https://acdc.rems.biocommons.org.au/	Service for data producers: Diabetes and heart diseases dataset cases.
Bigpicture	Digital pathology	https://bp-rems.sd.csc.fi/	Public-private data sharing: large image archive
Statistics Finland's Licensing Service	Statistics, register data	https://lupa.stat.fi/	Government institute process integration: Research data permits

Table 3. REMS software information card.

Software Name	Resource Entitlement Management System (REMS)
Source Code	https://github.com/CSCfi/rems
Licence	MIT Licence
Primary Language	Clojure / ClojureScript
Standards Support	GA4GH DUO, GA4GH Passport/Visas, OIDC

User Interfaces. The client-side interface is deployed as an asynchronous Single Page Application. It is written in ClojureScript, compiled to optimised JavaScript using Shadow-CLJS. The user interface is built upon the React.js framework via the Reagent wrapper, utilising the re-frame library for state management. The re-frame pattern dictates a unidirectional, event-driven data flow. This ensures that the user interface deterministically renders the current state of the system, eliminating race conditions when

displaying complex application forms or dynamic administrative dashboards. The REMS frontend includes a sophisticated form rendering engine that dynamically generates user interfaces based on administrative configurations. This allows resource owners to customise mandatory input fields, file attachments, and localised terms of use. Furthermore, the user interface natively integrates the GA4GH DUO. DUO codes are rendered as human-readable terms within the interface and stamped onto generated PDF summaries, ensuring both researchers and reviewers clearly understand the machine-readable constraints applied to the data. This client-side interface is optional, and implementations of REMS are free to create and use their own client-side interfaces, if they choose to do so.

User Identity and Authentication Management. Given its role as a gatekeeper for sensitive scientific data, REMS delegates data access applicant, data access reviewer and data access committee user authentication to external identity providers, ensuring that the software does not store sensitive credential payloads. The system integrates with academic and clinical federated identity networks, such as the ELIXIR Authentication and Authorisation Infrastructure, the Life Science Login service, the Australian Access Federation and the Finnish Haka identity federation [17] leveraging industry-standard OpenID Connect (OIDC) protocol. Upon successful authentication, the system ingests a configurable set of user attributes provided by the external identity providers (e.g., unique identifiers, organisational affiliations, and verified “bona fide researcher” statuses [22]). Furthermore, REMS can display the quality of user authentication (level of assurance of identity vetting), if the attribute is supported and provided by the external identity provider.

Systems Integration, Interoperability, and Security. To function seamlessly within a broader scientific ecosystem, REMS provides a comprehensive suite of interoperability tools leveraging GA4GH DUO, GA4GH Passport/Visas and OIDC standards. All operations available in the graphical user interface are supported by a fully documented RESTful API. This interface enables headless integration, allowing external data portals to embed REMS application processes or automated scripts to query entitlement statuses programmatically (see Zero Childhood Cancer Insight implementation in Table 2.). REMS event notifications also allow sending the application information to other systems such as customer relation management or archiving systems. An integrated security module intercepts file attachments uploaded during the application process. Utilising an external malware scanning path configured during deployment, the software asynchronously scans applicant attachments, piping output logs to the application logger to prevent the propagation of malicious payloads to reviewers or data archives.

Metadata-Driven Orchestration of Data Access. REMS can transform metadata standards into dynamic workflows by using GA4GH DUO [12] as the foundational logic for its application engine. When a dataset is registered, it can be tagged with specific DUO codes (e.g., DUO:0000006 for “Health/Medical/Biomedical research”). REMS uses these machine-readable terms to dynamically assemble a tailored data access application form, automatically injecting fields based on dataset restrictions. Upon data access approval, REMS functions as a GA4GH Visa Issuer [13], generating a digitally signed JSON Web Token (JWT) carrying the granted data access right. This tamper-evident token contains digitally signed metadata identifying the applicant, dataset, access type, and expiration

date. Downstream systems can verify the signature using the public key provided by the REMS OIDC endpoint.

API-First Architecture for Portable Workflows. REMS provides a comprehensive set of Application Programming Interfaces (APIs) to support modern portable workflows. By exposing permissions via APIs, automated analysis platforms can verify a researcher's Passport/Visa in real-time without manual intervention. This decoupling of the policy decision point from the policy enforcement point is essential for the scalability of large-scale genomic projects.

Data Formats. REMS leverages data-oriented formats rather than heavy object-oriented classes. Table 1. gives a high-level representation of the core data model. The primary file types defining the data model and configuration include: EDN (.edn): Extensible Data Notation. This is the native data format of Clojure used for system configuration (e.g., config-defaults.edn, dev-config.edn), test data definitions, and environment profiles. SQL (.sql): PostgreSQL schemas, database migrations, and queries. REMS maps SQL files directly to backend Clojure functions. Clojure/ClojureScript (.clj, .cljs): Source code files that define the in-memory data model. On the backend, data records and validations are often enforced using schema libraries. On the frontend, data flows through a single state using the ClojureScript re-frame framework. JSON (.json) is used exclusively at the interfaces of the system. JSON is the format for REST API requests/responses, OpenAPI/Swagger documentation generation, and OIDC user identity payloads.

Development and Deployment. To ensure consistent deployment across localised on-premise environments and cloud infrastructure, the software relies on modern continuous integration and containerisation practices. The REMS application, including its backend Java Virtual Machine processes and compiled frontend assets, is packaged into lightweight Docker containers. Linting tools (such as Hadolint) are integrated into the continuous integration pipeline to enforce image security. The repository encompasses a robust testing suite comprising unit tests and headless browser testing (via Karma and headless Chrome). Additionally, a specialised user simulator module is provided within the codebase for synthetic load testing, allowing deployers to model concurrent user access and ensure high-throughput stability before production deployment.

Results

REMS is currently deployed in operation across a wide spectrum of research environments internationally, demonstrating its capacity to adapt to diverse regulatory, cultural, and technical requirements. Table 2 illustrates public instances of REMS known to authors, and the data types they support access to. The implementations of REMS manage and generate critical access metadata across a wide spectrum of international research environments. The metadata and entitlement records processed by the system are distributed across multiple public instances, varying by data type and organisational scope. For example, the European Genomic Data Infrastructure (GDI) project uses REMS as the access gateway for national nodes across 20 European countries. REMS orchestrates the application process, ensuring member state legal sovereignty while using the GA4GH Passport standard

to eliminate redundant authentication for multi-country genomic analysis. Aotearoa Genomic Data Repository (AGDR) & Rakeiora operating in New Zealand use REMS to implement Māori Kaitiaki governance workflows. For AGDR the system links geography, species, datasets, and guardians to manage complex access control relationships, facilitating transparent, collective decision-making in alignment with Indigenous Data Sovereignty principles. As the national statutory operator for public health, The Finnish Institute for Health (THL) and Welfare Biobank has managed access to samples and data originating from the Finnish population since 2013. The THL Biobank and the Language Bank of Finland represent a high-volume, mature REMS deployment. REMS handles the entire request lifecycle, scalable for processing thousands of requests and demonstrating that data privacy by design approach is sustainable at a national scale.

Technical Validation

Four production-level data sharing ecosystems are highlighted in this article to technically validate REMS: European Genomic Data Infrastructure (GDI), Aotearoa Genomic Data Repository in New Zealand, the Finnish Institute for Health and Welfare Biobank in Finland, and FIN-CLARIN's Language Bank of Finland which provides language data for world-wide use.

Secure Cross-Border data infrastructure for human genomic datasets — the European Genomic Data Infrastructure (GDI) The GDI project is the technical backbone of the 1 + Million Genomes (1 + MG) initiative [9, 21, 25] aiming to establish a federated network of national genome collections across 20 European countries. Within this infrastructure, REMS serves as a core technology for managing cross-border data access requests. The fundamental challenge in 1 + million genomes is providing a unified access point while respecting the legal sovereignty of each member state. REMS functions as the access gateway for GDI nodes operated in countries like Finland, Sweden, and Belgium. When a researcher requests data from multiple national nodes, REMS orchestrates the application process, ensuring that each node's Data Access Committee (DAC) can review the request according to their specific national laws. By using the GA4GH Passport standard [13], REMS allows a researcher to obtain a "Visa" from a national authority that is then technically recognised by a central European GDI clearing house. This process eliminates the need for redundant authentication, enabling secure, multi-country genomic analysis at scale.

The 1 + million genomes initiative builds societal trust with public digital infrastructure to make international data sharing a success. This is particularly important in preparing for the implementation of large-scale health data regulations such as European Health Data Space (EHDS) [26]. Data donors and custodians must have confidence that their data is being used strictly within the bounds of consent and legal framework. Open-source tools like REMS, built on open GA4GH standards, provide the transparency required for this trust. By making the code and the logic of access visible and auditable, we move towards a transparent, decentralised digital infrastructure that supports both global science and local sovereignty.

Indigenous Data Sovereignty — Aotearoa Genomic Data Repository (Aotearoa, New Zealand) The AGDR is a secure within-nation repository for the storage, management, and sharing of non-human genomic data generated from biological and environmental samples originating in Aotearoa New Zealand. It was developed in alignment with the principles of Māori Data Sovereignty, the partnership ethos underpinning Te Tiriti o Waitangi (the Treaty agreement considered as New Zealand's founding document), and the cultural practice of kaitiakitanga (guardianship). The AGDR supports data management practices that enable iwi, hapū and whānau (tribes, subtribes and extended family) to effectively exercise their responsibilities as kaitiaki (guardians) over biological entities they regard as taonga (treasured) [23, 27]. The repository has a metadata standard which supports FAIR data standards, with additional fields to also support researchers in fulfilling their responsibility to the CARE principles.

AGDR was built in response to the strong push from largely Western science systems making scientific data fully open. Requirements of funders and journals are well known, including the joint data archiving policy of 2011 [23, 28] requiring data supporting publications be made publicly available. This is inconsistent with the views, values, beliefs, and practices of many Indigenous peoples, including Māori of Aotearoa New Zealand, especially in relation to genomic data originating from species or places of cultural significance [8, 27].

The challenge in Aotearoa was enabling both Māori data sovereignty, and appropriate research by approved researchers using genomic data from New Zealand species and environments. The implementation of REMS offers a transparent mechanism to allow collective decision-making on what data can be used, for what purpose and by whom. The use of different forms, and terms and conditions allows for respective kaitiaki to have different agreements in place with researchers as fits the context and values of the relevant iwi or hapū. REMS provides a basis for benefit sharing and potential pathways to commercial use through facilitating transparent communication, and relationship building between the researchers and kaitiaki, alongside the ability to develop bespoke licences for the agreed terms for use. This could include contribution to the Cali Fund for the Fair and Equitable Sharing of Benefits from the Use of Digital Sequence Information on Genetic Resources [29].

First used in the Rakeiora Genomics Platform [7, 30], a pathfinder in precision health research with Māori — this work drew on local experience in implementing and using the REMS platform for enabling Indigenous Data Sovereignty. REMS has been integrated into the AGDR production system using the same authentication and authorisation as the rest of AGDR, it is used both to approve data submissions and data access requests and has mediated 24 requests over the last 5 months.

Longevity and maturity — The Finnish Institute for Health and Welfare Biobank The Finnish Institute for Health and Welfare (THL) Biobank [31, 32] represents one of the longest-standing and most high-volume deployments of REMS, managing access to samples and data collected from the Finnish population since 2013. THL was a pioneer in federating sensitive data access authorisations. REMS handles the entire lifecycle of a biobank access request from the initial "shopping cart" selection of available cohorts (e.g., FINRISK, Health 2000) to the final approval under the Finnish Biobank Act [31]. The REMS solution

longevity demonstrates its stability and adaptability. REMS has successfully navigated the transition from local legislation to the strict requirements of the GDPR, proving that a Data Protection by Design (GDPR Art. 26 [4]) approach provides a sustainable foundation for national health infrastructures that can be implemented in digital tools that support respective data access authority processes. THL implementation verifies that maintaining a stable, legally compliant support based on REMS over a decade is feasible. With REMS integration THL is also prepared to scale to meet the needs of thousands of researchers and multiple national registries, if interest in the biobank data collections increases [33].

We estimate that shifting from physical sample requests to REMS-powered digital data access could increase request volumes to THL by one to two orders of magnitude [2, 33]. Unlike physical samples, digital information is not depleted by use, allowing for infinite reuse. If the present access is streamlined via automation, the Finnish biobank infrastructure could transition from handling hundreds of complex manual requests annually to thousands of automated, high-velocity data “Visas”. This digitalisation shift makes machine-actionable orchestration not just a convenience, but a prerequisite for the survival of national health infrastructures to manage their dataset assets successfully.

Providing restricted language data world-wide — The Language Bank of Finland In 2014 the Language Bank of Finland [16], also hosted at CSC, needed to replace their own ageing access right management system. REMS could be directly applied to manage data sharing at the Language Bank. Since 2015 the system operated at CSC has processed thousands of applications for access to restricted linguistic and social science datasets.

The Language Bank has a three-way licensing model for language data. Data can be public (“PUB”), for “academic use” (“ACA”) and restricted (“RES”)]. “RES”-labelled data requires a personal application managed by REMS. Furthermore, depending on the type of data, assurances of GDPR-compliance or, in case of undergraduates, additional approval from their supervisors is needed. Data owners can process the applications themselves or delegate approval to Language Bank staff.

For certain datasets a usage permissions span a collection of datasets, like variants and even newer versions. This streamlines administrative burden: If a user is granted permission to download the dataset, access through a web browsing interface can be granted simultaneously, and REMS provides an entitlement to authorise the data use in both services.

REMS also offers a solution for handling grey areas of “ACA” licensed data. Users from academic institutions world-wide that provide their active organisational affiliation status to the eduGAIN federation can access “ACA” licensed data. If this information is not transmitted, users have the option to apply for “academic access” via REMS, effectively granting them the same role-based access rights. The Language Bank also uses REMS to provide data against a fee [35]. In this case the application is not granted on academic merits but on proof of payment by the payment processor at the University of Helsinki.

The use of REMS in a long-term production environment and supporting the choice of the standards-based approach to access management has been a success story. It enables researchers to share datasets which would be otherwise difficult to share and also can handle role-based and commercial use cases.

Discussion

Integration of REMS into production environments of European GDI, Aotearoa Genomic Data Repository, THL Biobank, and the Language Bank of Finland and other implementations in Table 2. technically validates the usefulness of REMS. The evolution of the FAIR ecosystem over the last decade has reached a critical point where the sheer volume of global sensitive (e.g. genomic and clinical) data has outpaced the capacity of manual, human-mediated oversight. As research infrastructures transition from siloed repositories to federated networks, the role of systems like REMS is shifting from a simple administrative portal to a high-performance orchestration engine. REMS enables the transition from paper-based permissions to digitally signed tokens that travel with the researcher.

As we enter the next decade of FAIR with availability of advanced artificial intelligence to analyse and interpret valuable data, the focus must remain on interoperable authorisation to ensure proper data management practice. This is one of the key aspects of development and use of AI for high-value and sensitive data [3], also recognised by the European Union regulations in the Digital Omnibus package [34]. By leveraging e.g. OIDC, GA4GH standards, and robust open-source implementations, we can build a global data ecosystem where the “A — Accessible” in FAIR is as seamless and reliable as the data it protects. REMS tackles this bottleneck by introducing configurable levels of automation while keeping human oversight intact. It uses DUO codes to automate matching of an applicant's intended research against dataset consent restrictions. Once a DAC approves a request, REMS issues a GA4GH Visa and logs the entire process for auditing. The technical pursuit of automation must also not come at the expense of cultural safety or data sovereignty. The Rakeiora platform in Aotearoa New Zealand serves as a global model for how technical standards can support, rather than override, indigenous governance. By integrating Māori values into the technical REMS workflow, Rakeiora demonstrates that building trust through the inclusion of community-led oversight is both feasible and as critical as the technical digital infrastructure.

Declarations

Usage Notes

REMS is designed around open standards to function as a component within the global research ecosystem. By implementing standards such as OIDC and the GA4GH Passport specifications, the system ensures vendor neutrality and regulatory compliance. Organisations deploying REMS can tailor the software to meet specific national or institutional regulatory requirements without seeking permission from a proprietary vendor. All operations available in the graphical user interface are

supported by a fully documented RESTful API, enabling headless integration for external data portals or automated scripts.

Code Availability

All the data is available within the article. The source code for the Resource Entitlement Management System (REMS) is publicly available under the MIT Licence on GitHub at https://github.com/CSCfi/rems. The software is primarily developed in Clojure and ClojureScript and natively supports GA4GH DUO, GA4GH Passports/Visas, and OIDC standards. To ensure long-term archival stability and reproducibility, the specific version of the code referenced in this manuscript is deposited in Zenodo [36]. GitHub acts as an interaction hub for code contributions to develop REMS forward. Development is coordinated by ELIXIR Finland operated at the CSC - IT Center for Science (Table 3).

Competing interests

The authors declare no competing interests.

Author Contribution

All authors read and approved the manuscript. M. Hak. is the REMS software product lead, writing; J. Lei. development manager, writing; M. Hus. review, writing; M.K. implementation; J. Lau developer; J. Lai. implementation; I. L. original author, review; H. L. implementation, writing, visualisation; M. L. original author, writing; M. M. implementation, writing; D. S. implementation, writing; J. T. implementation; T. J. developer; M. R. developer; M. S. implementation; T. W. implementation; B. P. implementation; M. Kub. implementation; M. Kon. implementation, writing, visualisation; O. H. implementation; C. L. implementation; A. P. implementation; B. P. implementation, writing; T. G. implementation; M. P. implementation; C. E. R. implementation, writing; T. H. N. original author, correspondent, writing.

Acknowledgement

Research council of Finland grants No. 137370, 141512, 263164, 271642, 273655, 292265, 319968, EU Horizon 2020 grants 676559, 824087, EU Digital Europe Programme grant number 101081813, Horizon 2020 / Innovative Medicines Initiative 2 (IMI2) grant number 945358, the Nordic e-Infrastructure Collaboration (NeIC) and CSC - IT Center for Science.

Data Availability

All the data is available within the article. The source code for the Resource Entitlement Management System (REMS) is publicly available under the MIT Licence on GitHub at https://github.com/CSCfi/rems.

The software is primarily developed in Clojure and ClojureScript and natively supports GA4GH DUO, GA4GH Passports/Visas, and OIDC standards. To ensure long-term archival stability and reproducibility, the specific version of the code referenced in this manuscript is deposited in Zenodo https://doi.org/10.5281/zenodo.21701368. GitHub acts as an interaction hub for code contributions to develop REMS forward. Development is coordinated by ELIXIR Finland operated at the CSC - IT Center for Science (Table 3).

References

1. Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data*. **3**, 160018 (2016).
2. Lähteenmäki, J. et al. Integrating data from multiple Finnish biobanks and national health-care registers for retrospective studies: practical experiences. *Scand. J. Public. Health*. **50**, 482–489 (2022).
3. Resendez, V., Yıldırım, F., Gaeta, E., Fico, G. & Borsci, S. Toward a common set of interface requirements for genomic data management: scoping review. *J. Med. Internet Res.* **28**, e78405 (2026).
4. European Parliament & Council of the European Union. Regulation (EU) 2016/679 (General Data Protection Regulation). *Off J. Eur. Union*. **L119**, 1–88 (2016).
5. European Parliament & Council of the European Union. Regulation (EU) 2024/1183 amending Regulation (EU) No 910/2014 as regards establishing the European Digital Identity Framework. Off. J. Eur. Union L. /1183 (2024). (2024).
6. Carroll, S. R. et al. The CARE principles for Indigenous data governance. *Data Sci. J.* **19**, 43 (2020).
7. Rye, C. E. et al. Rakeiora Genomics Platform: a pathfinder for genomic medicine research in Aotearoa New Zealand. *J. R Soc. N Z.* **55**, 2481–2505 (2025).
8. Kukutai, T. & Black, A. CARE-ing for Indigenous nonhuman genomic data—rethinking our approach. *Science* **385**, eadr2493 (2024).
9. Saunders, G. et al. Leveraging European infrastructures to access 1 million human genomes by 2022. *Nat. Rev. Genet.* **20**, 693–701 (2019).
10. Linden, M., Nyrönen, T. & Lappalainen, I. Resource entitlement management system. In Selected Papers of the TNC2013 Conference (TERENA, 2013); Zenodo (2026). https://doi.org/10.5281/zenodo.18402461
11. Rehm, H. L. et al. GA4GH: international policies and standards for data sharing across genomic research and healthcare. *Cell. Genom*. **1**, 100029 (2021).
12. Lawson, J. et al. The Data Use Ontology to streamline responsible access to human biomedical datasets. *Cell. Genom*. **1**, 100028 (2021).
13. Voisin, C. et al. GA4GH Passport standard for digital identity and access permissions. *Cell. Genom*. **1**, 100030 (2021).

14. Thorogood, A. et al. International federation of genomic medicine databases using GA4GH standards. *Cell. Genom*. **1**, 100032 (2021).
15. Rahimzadeh, V., Lawson, J., Rushton, G. & Dove, E. S. Leveraging algorithms to improve decision-making workflows for genomic data access and management. *Biopreserv Biobank*. **20**, 429–435 (2022).
16. FIN-CLARIN. The Language Bank of Finland. re3data.org (2026). https://doi.org/10.17616/R3ND16
17. Linden, M. et al. Common ELIXIR service for researcher authentication and authorisation. *F1000Res* **7**, 1199 (2018).
18. Carnuccio, P. et al. Human Genomes Platform Project: Data Access Committee Management Systems Candidate Solutions Evaluation Report. Zenodo. https://doi.org/10.5281/zenodo.10720100 (2024).
19. Lappalainen, I. et al. The European Genome-phenome Archive of human data consented for biomedical research. *Nat. Genet.* **47**, 692–695 (2015).
20. Mailman, M. D. et al. The NCBI dbGaP database of genotypes and phenotypes. *Nat. Genet.* **39**, 1181–1186 (2007).
21. European Commission. European '1 + Million Genomes' initiative. (2026). https://digital-strategy.ec.europa.eu/en/policies/1-million-genomes
22. Dyke, S. O. M. et al. Registered access: authorizing data access. *Eur. J. Hum. Genet.* **26**, 1721–1731 (2018).
23. Te Aika, B. et al. Aotearoa genomic data repository: an āhuru mōwai for taonga species sequencing data. *Mol. Ecol. Resour.* **25**, e13866 (2025).
24. 24 et al. Leveraging synteny to generate reference genomes for conservation: Assembling the genomes of Hector's and Māui dolphins. *Mol. Ecol. Resour.* **25**, e14109. https://doi.org/10.1111/1755-0998.14109 (2025).
25. Pacheco, B., Sarvari, A., Lebioda, P., n der Werf, H. C. & Zenodo J.va GDI D4.3 – User Portal Live Cataloguing Data Within the GDI. (2025). https://doi.org/10.5281/zenodo.15309427
26. European Parliament & Council of the European Union. Regulation (EU) 2025/327 on the European Health Data Space and amending Directive 2011/24/EU and Regulation (EU) 2024/2847. Off. J. Eur. Union L. /327 (2025). (2025).
27. Hudson, M. et al. Te Mata Ira: Guidelines for Genomic Research with Māori. (Te Mata Hautū Taketake—Māori & Indigenous Governance Centre, University of Waikato, ISBN 978-0-473-36937-8 (2016).
28. Liggins, L. et al. The future of molecular ecology in Aotearoa New Zealand: an early career perspective. *J. R Soc. N Z.* **52**, 92–115 (2022).
29. Conference of the Parties to the Convention on Biological Diversity, Sixteenth meeting, Cali, Colombia, 21 October–1 November (2024). https://www.cbd.int/dsi-gr (2024).

30. Robertson, S. P. et al. A primary care pharmacogenetic precision medicine pilot based on specific Māori tribal ethical frameworks and principles. *J. Community Genet.* **17**, 79. https://doi.org/10.1007/s12687-026-00914-7 (2026).
31. Parliament of Finland. Biobank Act (688/2012). Finlex (2012). http://data.finlex.fi/eli/sd/2012/688/ajantasa/2023-04-14/fin
32. Finnish Institute for Health and Welfare. THL Biobank. (2026). https://thl.fi/en/our-services/thl-biobank1
33. Tupasela, A. et al. Estimating the use of biological samples in Finnish biobanks and hospital collections. *Eur. J. Hum. Genet.* **33**, 1493–1498 (2025).
34. European Parliament & Council of the European Union. Regulation (EU) 2026/1744 amending Regulations (EU) 2024/1689, (EU) 2018/1139 and (EU) 2023/1230 as regards the simplification of the implementation of harmonised rules on artificial intelligence (Digital Omnibus on AI). Off. J. Eur. Union L. (2026). (2026)/1744.
35. FIN-CLARIN. Accessing the Language Bank of Finland. (2026). https://www.kielipankki.fi/access/
36. CSC – IT Center for Science Ltd. Resource Entitlement Management System (REMS) (Version 2.39.1), Computer software. Zenodo. (2026). https://doi.org/10.5281/zenodo.21701368

Figures

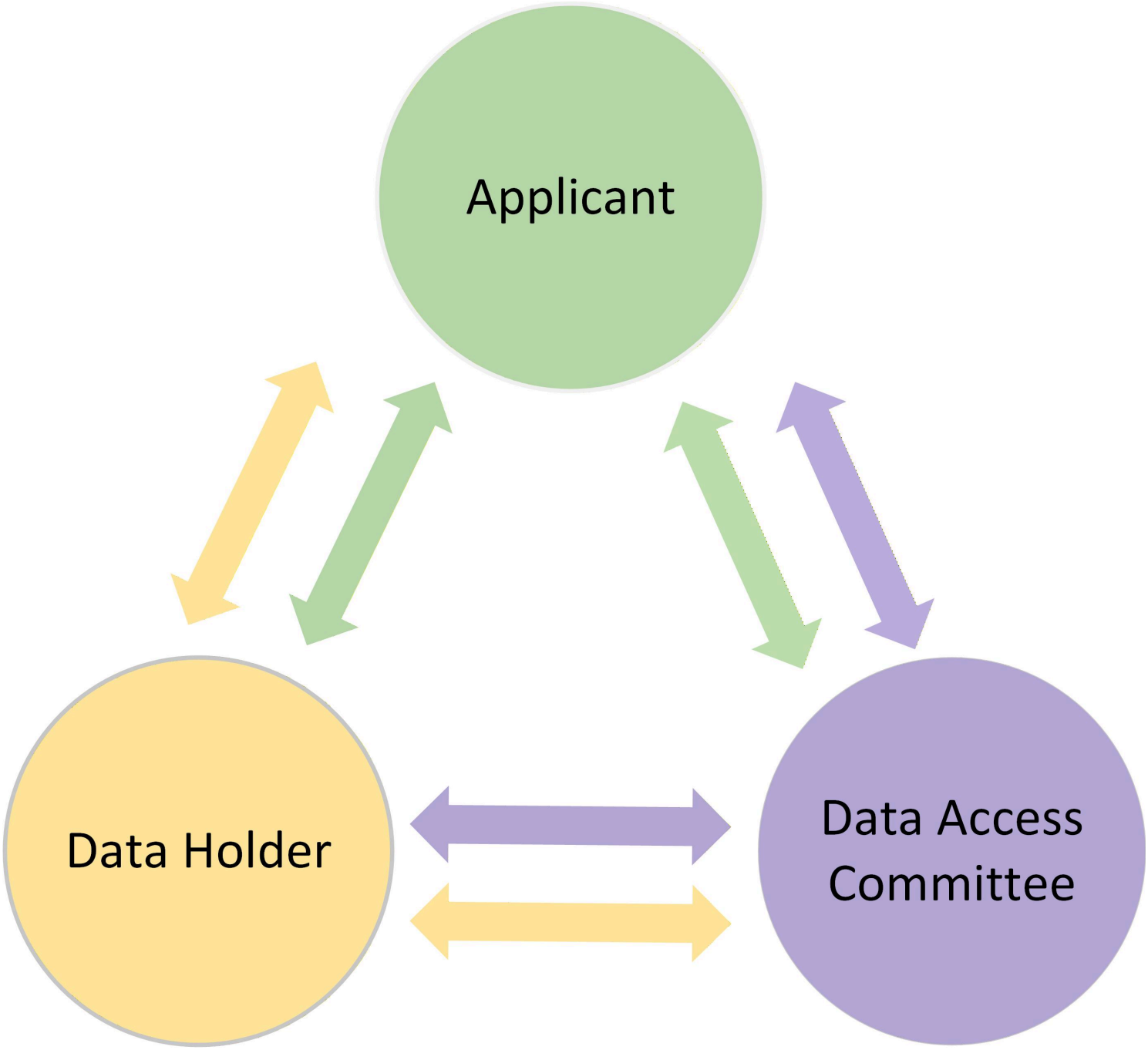


Figure 1

Administrative three-body problem. Interactions between data applicants, authorisation bodies, and data holders depend on how each of them understands the data access request (purpose, cohort, methods), their view of risk and compliance constraints, and technical expectations (secure processing environment, formats, etc.). Each actor can update their state (bidirectional arrows) independently or in response to the other two, forming a triadic feedback system.

Figure 2.

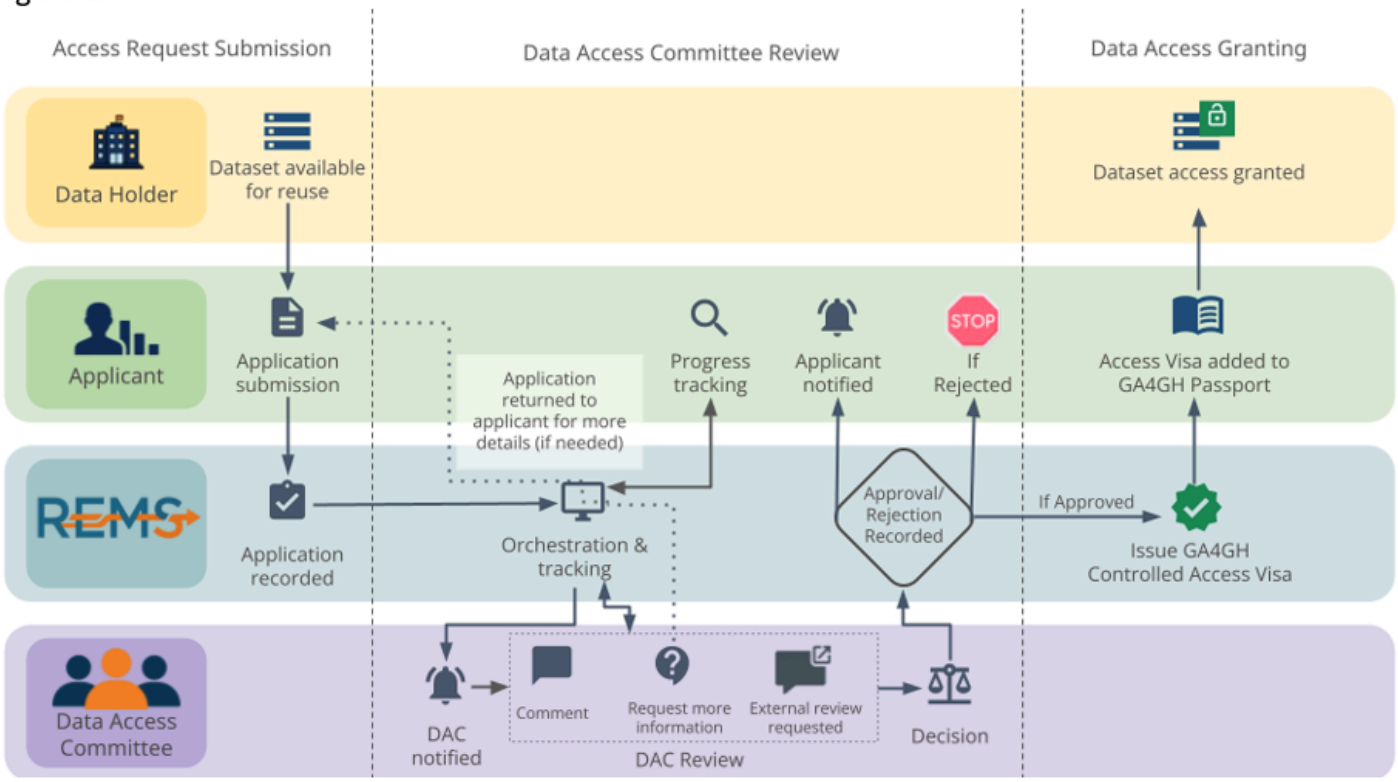


Figure 2

REMS application process. REMS orchestrates the process between data applicants, data access authorities and data holders by providing functionalities that streamline data access application management.

Figure 3.

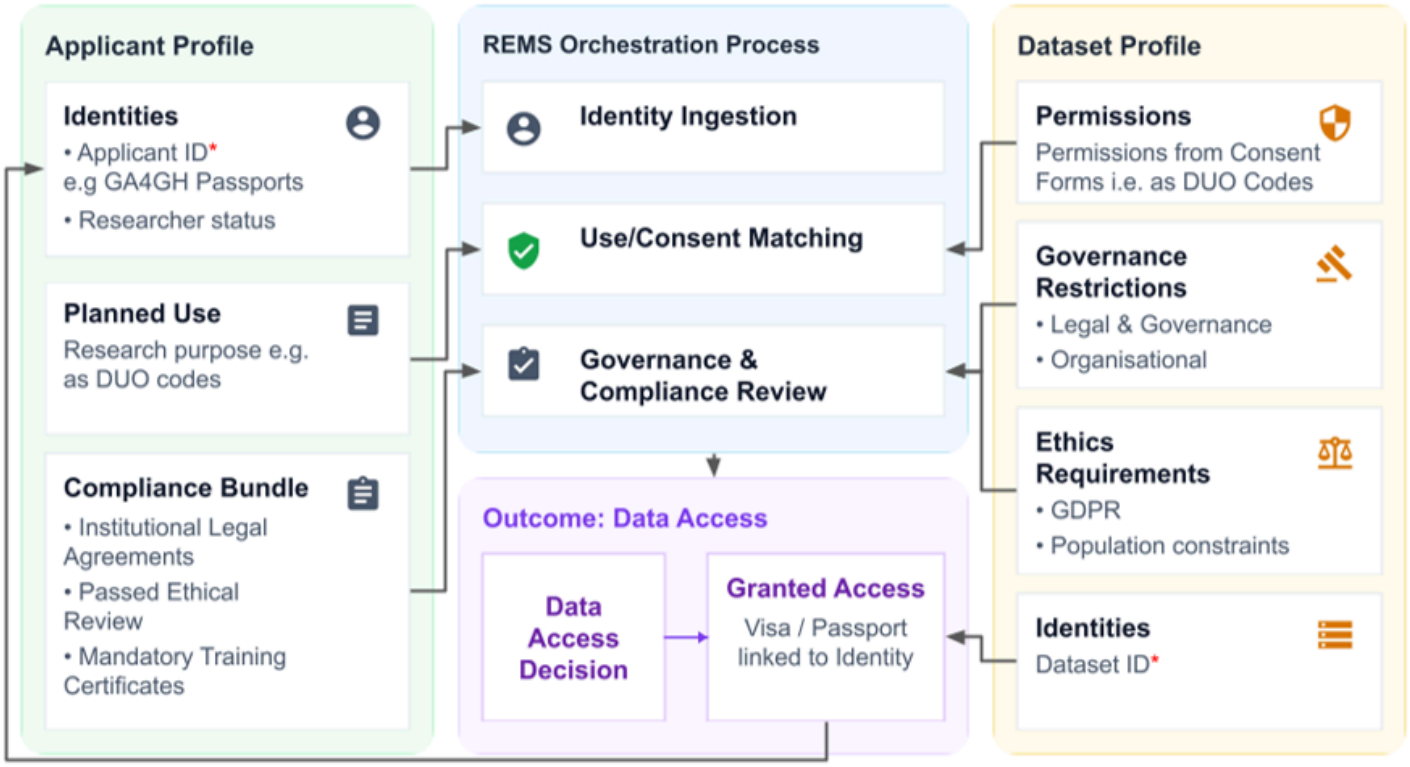


Figure 3

REMS orchestration. REMS collects the user information required for a resource access application in accordance with the process defined by the access authorisation body. It securely offers this information to service providers to implement access using the GA4GH Visa/Passport standard either through OpenID Connect protocol or using REST API calls. REMS only needs applicant and dataset IDs, but additional required fields can be added by those creating application forms.

Figure 4.

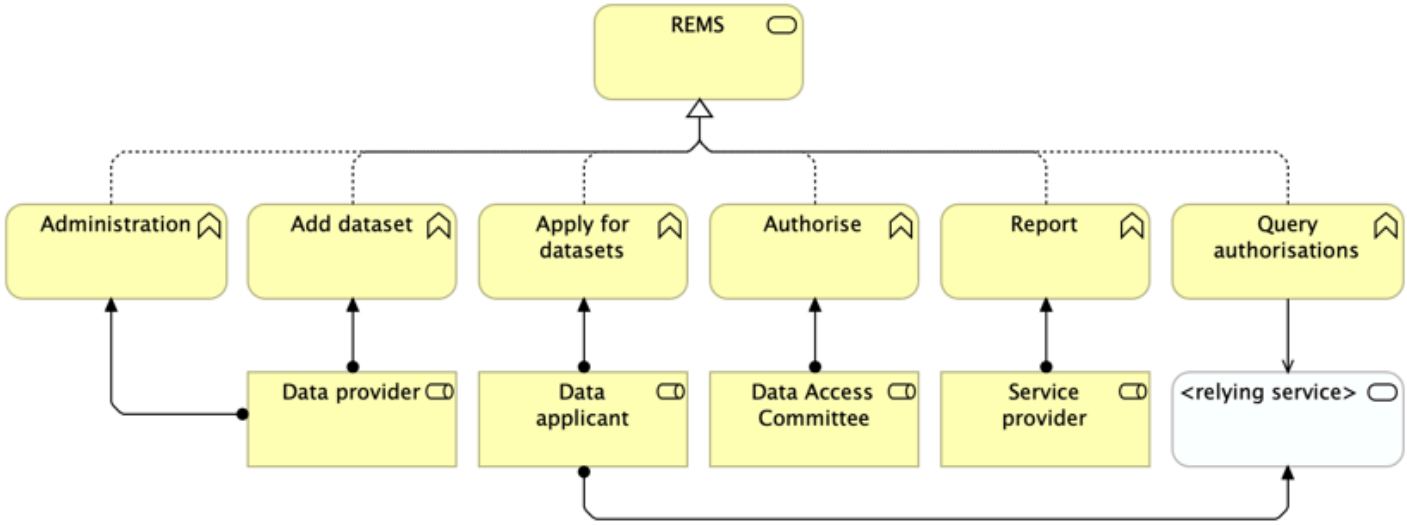


Figure 4

REMS service architecture. Software functions and roles using the Archimate notation.