Seer Data Platform Architecture & Security

Overview of Seer Data & Analytics' Approach

PREPARED FOR:	**General distribution**
PREPARED BY:	Seer Data & Analytics
DATE:	Last updated 19 March 2024

Table of Contents

Introduction

This document provides an overview of the Seer Data Platform architecture and outlines Seer Data & Analytics' security practices as an organisation. These practices contribute to the strong foundations of trust that we have with our partners and customers.

Since its inception, Seer Data has evolved by working closely with our customers. This has afforded Seer Data the unique position of developing a solution that directly meets our customers' needs and addresses the very real problems of using data to answer community questions.

The Seer Data Platform is an intermediatory system where data custodians have control over who they share their data with. As an intermediatory, Seer Data enables individuals and organisations to realise the value of Data Mesh design principles, wherein Domain Data Products can be shared within an organisation or between organisations.

Seer Data & Analytics

Seer Data & Analytics was founded to solve the problem of equitable access to data for decision-making for people of all skillsets.

Seer Data is a data sharing platform co-designed with and for our community of users. Our users are organisations that build and deliver strategies for people and places including community organisations, not-for-profits, collaboratives, businesses and all levels of government.

Trust underpins the work Seer Data is doing across the data ecosystem - from policymakers, down to a local community level. Our work at the community level means we have deep understanding and strong trusted relationships with our community of users, and Seer Data is recognised by Federal and State government bodies as a trusted data sharing Intermediary.

As the Seer Data Platform has evolved, customers' needs have remained central to functionality, scalability and security.

Associated Document

Seer Data Governance & Privacy - provides an overview of Seer Data & Analytics' governance and privacy philosophies.

Seer Data Platform Overview

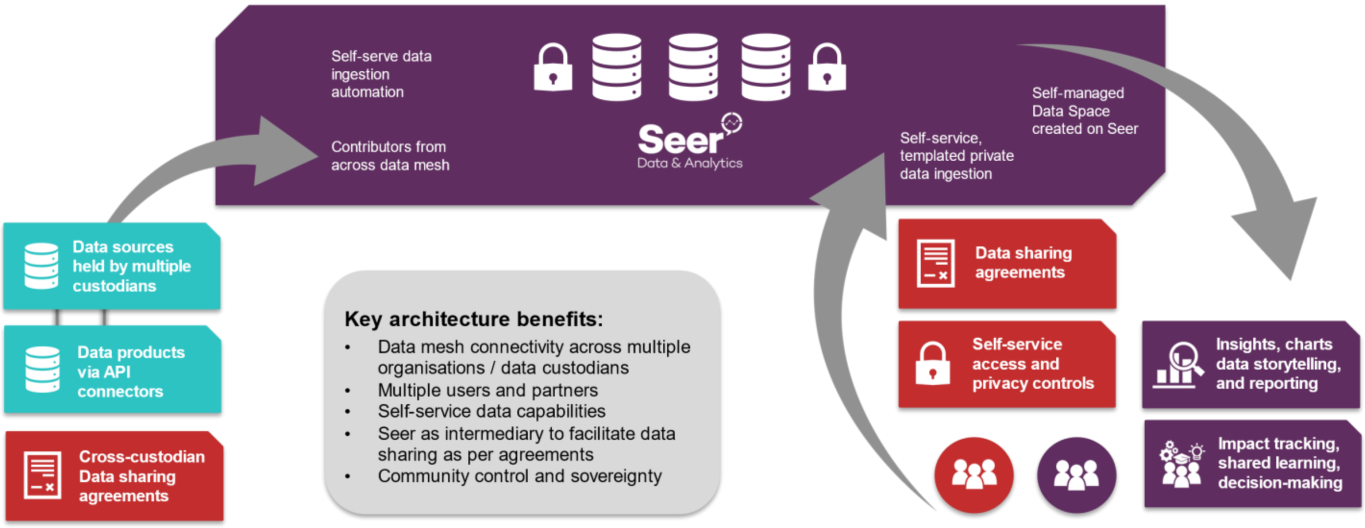
The Seer Data Platform enables individuals and organisations to realise the value of Data Mesh design principles by providing an easy-to-use Self Service Data Platform. Data custodians host their data (Domain Data Products) with Seer Data, allowing that data to be shared, discovered, queried, and analysed by their designated data consumers within their organisation or partner organisations. Seer Data aligns with the principles of Data Mesh design as follows:

Domain Ownership: Teams most knowledgeable about the content and context of their dataset are responsible for their design and ongoing validation, reducing reliance on third parties and centralised data teams.

Data as a product: Considering data as a product necessitates it being maintained in an up-to-date and ready-to-query state and prompts custodians to consider the needs of consumers within their domain team and other domain teams with respect to their data.

Self-service infrastructure: A dedicated self-service layer (the Seer Data Platform) where custodians host and administrate access to their data, and where custodian-designated consumers discover, query, analyse, and communicate their findings and actions.

Federated governance: Unification of the diverse world of data products in many different native formats within a common simple interface enables interoperability between data from different organisations. A common syntax for describing access controls and privacy protection measures enables each custodian to apply and communicate their own governance requirements.



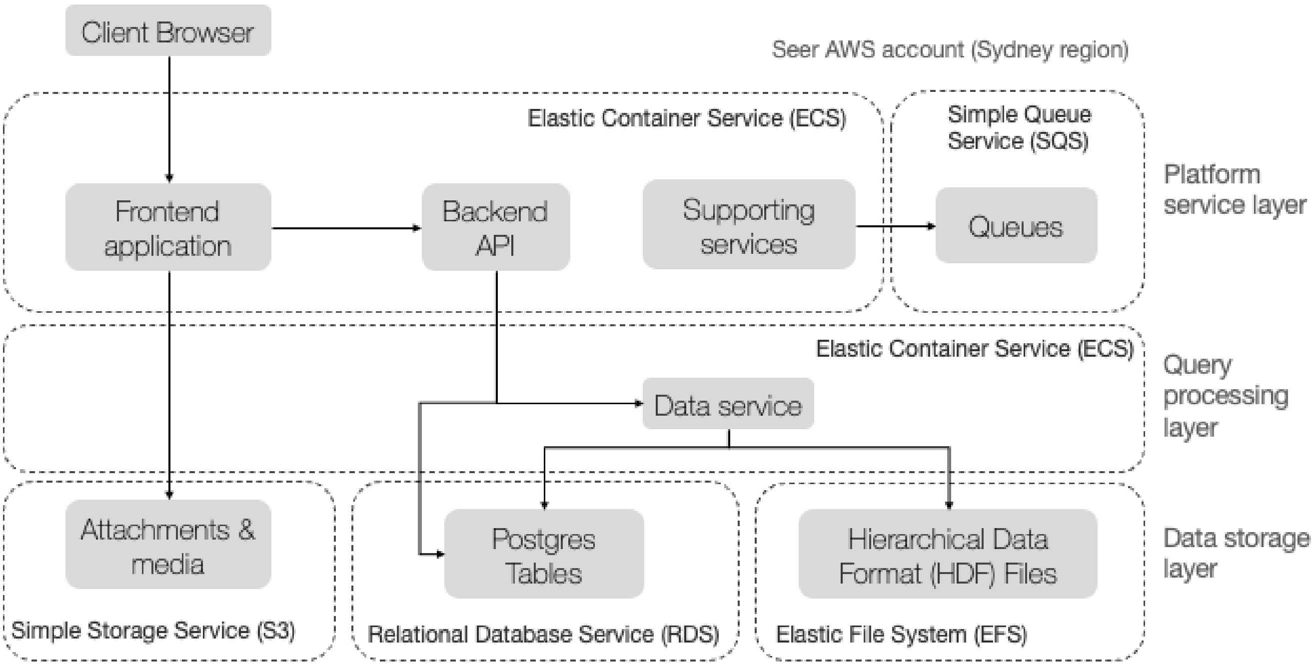
Seer Data Platform Architecture

Overview

The Seer Data platform is hosted on the Amazon Web Services (AWS) cloud provider, with all region-based services located in Australian datacentres based in Sydney. Seer Data uses a combination of AWS services to achieve outcomes for our clients in a way that improves our development cycle speed, and by extension, time-to-market of new features; centralises service management; reduces tooling overhead; and allows for scalability and service sustainability as our organisation, client base, and data ecosystem grows over time.

The Seer Data platform is deployed as a set of multi-tenant applications in AWS Elastic Container Service (ECS) to enable quick, continuous delivery of value with minimal management and configuration overhead and high availability with automatic service scaling rules.

Platform Architecture Diagram



Platform Architecture Description

The architecture diagram in the previous section depicts Seer Data's clients interacting with the Seer Data Platform frontend web application through their computer's web browser. Seer Data makes use of AWS Elastic Container Service (ECS) to run applications in scalable Docker containers, including the frontend web application.

The Seer Data Platform frontend web application has been designed as a single-page application (SPA) to allow for smooth and frictionless transitions between different functionality of the Seer Data Platform. The frontend application is developed in JavaScript (Typescript), using the popular ReactJS library.

Throughout a User session, the Seer Data frontend JavaScript application makes requests on behalf of the User to authenticate them and perform various actions by making Application Programming Interface (API) request calls to the primary Seer Data Backend API service.

When these requests include requests for Open Data and/or Private Dataset querying or retrieval, the Seer Data Platform Backend service makes another API call to the Seer Data service, which is responsible for querying and returning data from the raw data.

Based on the type of data, all datasets in the Seer Data Platform are stored in one of two formats:

- Hierarchical Data Format (HDF), which is a file format that is stored on a regular file system on disk, and q
- Postgres tables, which is a relational database storage object.

Both storage formats ultimately enable returning aggregated data, information about the data catalog, and/or a filtered representation of the catalog through the Data Service to the Backend API. This is used in combination with Access Control Limits that are specified at the Dataset level or at the data variable/category level to return to the User the information they requested, limited to the Datasets the User has access to, for display and further exploration in the frontend browser application.

Data related to the operations of the platform itself – such as User details, login credentials, User-generated tables and charts, and User-specified data access controls – are stored in a managed relational PostgreSQL database running on AWS Relational Database Service (RDS) in data centres in Sydney.

The PostgreSQL database is primarily accessed via the Seer Data Backend service discussed earlier, which functions as the core API to access the Seer Data Platform operational data and perform regular Create-Read-Update-Delete (CRUD) operations on database objects.

The Seer Data frontend application also accesses AWS Simple Storage Service (S3), as depicted in the section above, to retrieve media and files that relate to a specific User's platform usage, such as User profile pictures, video assets, and PDF attachments.

Seer Data makes use of AWS Application Load Balancers (ALBs) to direct network traffic to the right target ECS service and enforces a redirect from HTTP to HTTPS on certain request types to ensure we use a more secure network communication protocol wherever possible.

Approach to Architecture & Security

Introduction

Seer Data treats data and application security with the utmost importance. Seer Data employs a variety of measures based on industry best practices to minimise exposure and mitigate risk of security breaches. Our security controls are in place to protect the confidentiality, integrity and availability of data hosted through our platform, as well as our overall IT environment. For obvious security reasons, we do not publicise details of the security mechanisms of the Seer Data Platform.

The Seer Data Platform was initially designed to host publicly available Open Datasets, which contain population-level statistics that had typically undergone significant privacy and deidentification processes prior to being published as an Open Dataset. One example of this is the Census dataset published by the Australian Bureau of Statistics (ABS).

The Seer Data Platform has evolved and provide the ability for users ingest their own non-publicly available data to collaborate and analyse this data in conjunction with Open Data and with other Users of the Seer Data Platform. The data custodians of these private datasets are able to provide controlled access to the data to Users or partner organisations whom they trust or only share within their own organisation. Please refer to **Seer Data Governance & Privacy** for more details.

It is the policy of Seer Data that all Datasets (Aggregate and URL) hosted in the Seer Data Platform must **not** contain PII. This is reflected in our Terms of Service. Customers are responsible for adhering to this policy. Seer Data does not actively assess customers' Datasets to identify PII, nor does Seer Data warrant that any Dataset does not contain PII. Seer Data assists Organisations to process their Datasets to remove PII data if requested. Seer Data does not warrant that processed Datasets do not contain PII. For details of the Seer Data Terms of Service refer to https://seerdata.ai/terms-of-service/.

Other kinds of data currently hosted by Seer Data include basic user and organisation information for our client Organisations and Users, file attachments uploaded by users, and open-ended text commentary that can be supplemented with rich media (e.g. video, images).

Core Principles

Seer Data's philosophy is centred around several core principles in terms of our general approach to the Seer Data Platform architecture and associated security:

Accessibility

The purpose of the Seer Data Platform is to make data more accessible to community. To Seer Data, this means that access to the platform needs to be easily available to as wide of a community of Users as possible, with the necessary controls in place to ensure that this ease of access does not come at an unreasonable cost to security and privacy.

Trust & Privacy

While access is important, trust and privacy underpin the Seer Data Platform as well as all our work with people on the ground. Without features that allow the safe and secure access to data and insights, the Platform would not be aligned with Seer Data's core values and the values belonging to the organisations we work with. Please refer to **Seer Data Governance & Privacy** for more details.

Scalability

The Seer Data Platform is designed for scalability and considers growth in Seer Data's User base, volume of data in Seer Data's Open Data Library and in Seer Data's User's Private Data collection.

To deliver a scalable solution and in conjunction with the scalability features of AWS, Seer Data contemplates the performance and usability requirements of Seer Data Platform Users and Seer Data Platform administration as part of the scalability architecture.

It is Seer Data's engineering philosophy that scalability be architected and designed in a way that scaling can be achieved as and when needed through a just-in-time approach.

AWS Technical Validation

Seer Data & Analytics are a qualified Amazon Web Services (AWS) Software Partner. The Seer Data Platform has been validated by successfully completing the Foundational Technical Review (FTR).

The FTR's primary purpose is to ensure that the Seer Data Platform has implemented a set of architectural, security, and operational best practices. It defines a set of security controls and recommendations based on industry best practices that AWS have identified. This serves as a gate to ensure the products AWS endorses to customers have appropriate mitigations for the most common risks that impact end customers.

What does this mean for our customers? By continually optimising the security, scalability and agility of our services in line with best-practice industry-leading standards, Seer Data is able to deliver better solutions to communities, not-for-profits and Government.

This is a significant milestone on our data journey leading to exciting opportunities for collaboration, partnerships and growth with AWS and partners to drive innovation to benefit society.

Physical Security

Office Access

Seer Data employs flexible working practices. Seer Data staff work from Seer Data's shared office coworking space or remotely from home offices.

Access to the coworking space is via registration of all team members. Each team member is supplied with an access card for building access. Allocation and management of access cards is controlled by the coworking space and building management teams.

Seer Data has its own dedicated office within our Sydney-based coworking space. Staff who work from the office are allocated individual keys to the office. The office is locked when unattended and the facility has security cameras in all common areas of the building.

Employee Devices

Seer Data employees typically use a laptop computer for their work. Employee computers are password protected and are configured to log out after a period of inactivity. After a reboot, computers can only be accessed by a username and password combination.

Seer Data requires hard disk encryption for all our employees to reduce the risk of data breaches in any scenario involving manual unauthorised access to the hard disk directly.

Internal audits are conducted periodically to ensure device compliance.

Physical Documents

All physical documents are scanned and saved to a secure Microsoft SharePoint server. No hard copies of documents are retained by Seer Data. All documents are shredded in a secure bin when no longer required.

Platform Security

Cloud Provider

The Seer Data platform, including Open and Private data, is hosted on the Amazon Web Services (AWS) cloud provider and benefits from the security assurances that AWS makes to their customers.

AWS is a cloud storage and computing platform that is trusted by many public and private sector organisations world-wide.

Details of AWS security measures can be accessed at https://aws.amazon.com/security/.

Identification & Authentication

Seer Data Platform Users are authenticated through a code sent to their registered email address.

Users are granted a JSON Web Token (JWT) bearer token that contains the access and authorisation information for the authenticated user.

Authorisation

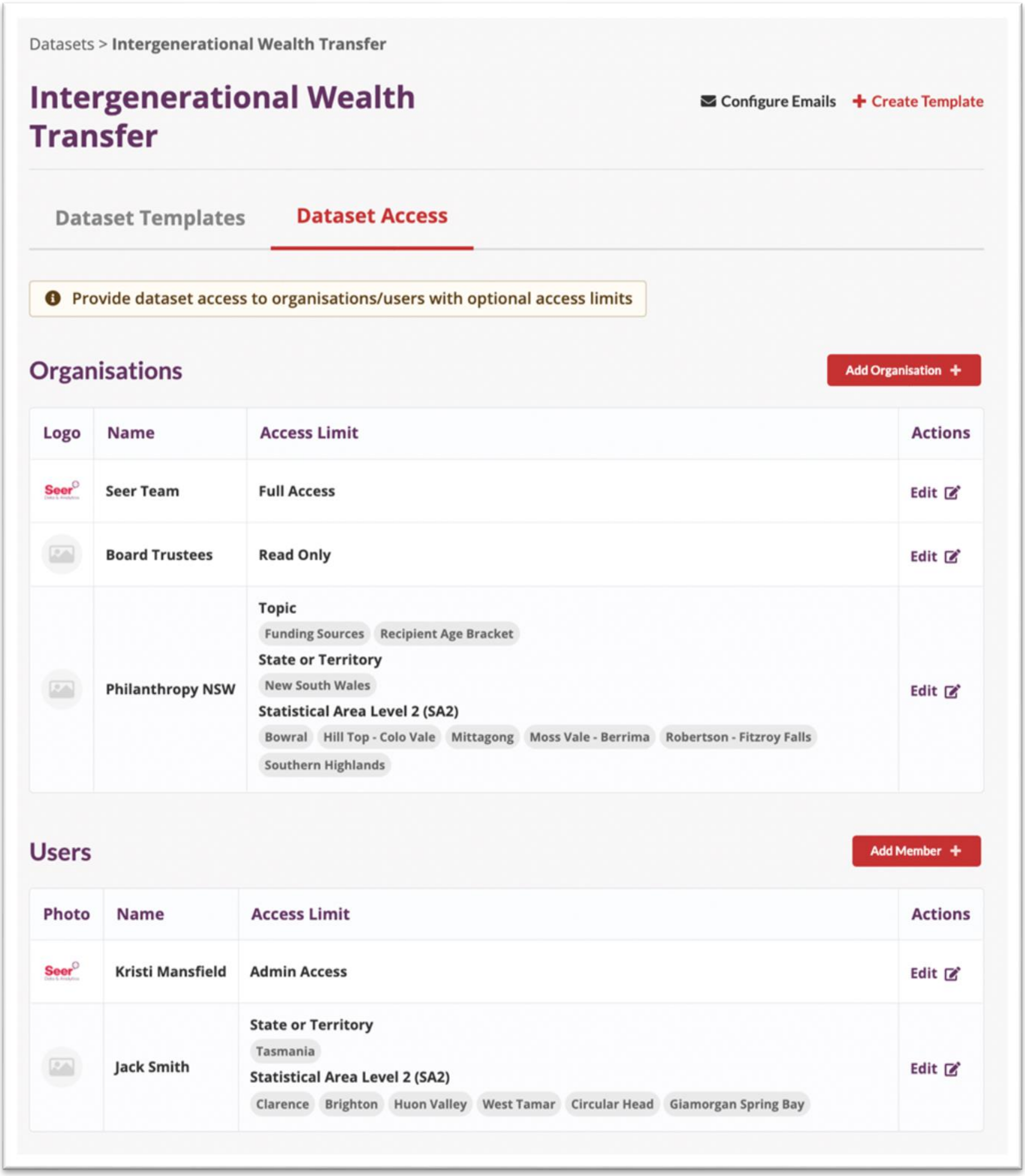
Authorisation is managed via the privacy and governance structures set out in the **Seer Data Governance & Privacy** document.

Authorisation includes access to datasets, parts of datasets, user-generated tables and charts, and other parts of the Seer Data Platform. Access is granted at the appropriate level set depending on the object being accessed – this might be Administrator rights, Full Access, Read Only or No Access.

Users who are organisation administrators are able to add users, remove users, modify organisation settings, invite partners and grant dataset moderator rights to team members.

All users are able to self-manage the editing and sharing permissions of assets they have created on the Seer Data Platform. This typically involves Insights consisting of tables, charts and storey telling content that are stored in a folder-like structure on the platform.

Users who have been granted dataset moderator access are able to manage the editing and sharing permissions of the given Dataset, to a granular degree of control that allows access to be specified on the variable and category-levels within a Dataset (e.g., it is possible to only provide a user access to data for a specific State or Territory within a dataset that may contain data for all States and Territories) and to do this at the User or Organisation level as depicted in the image below.



Seer Data Platform Dataset Access configuration screenshot.

For more information on Authorisation and access granting mechanisms or possible combinations to achieve specific outcomes, see the **Seer Data Governance & Privacy** document.

Data Security and Backups

All datasets are stored on AWS Elastic File System (EFS). These file systems are encrypted at rest using Amazon Web Services (AWS) encryption mechanisms. Amazon EFS uses industry-standard AES-256 encryption algorithm to encrypt EFS data and metadata at rest.

The AWS key management infrastructure uses Federal Information Processing Standards (FIPS) 140-2 approved cryptographic algorithms. The infrastructure is consistent with National Institute of Standards and Technology (NIST) 800-57 recommendations.

Automated daily backups are performed on Seer Data's AWS EFS dataset stores through AWS Backup. These backups are retained by AWS for a rolling period covering the most recent 5 weeks and are deleted thereafter.

All unit record level datasets are stored in a relational PostgreSQL database, running on AWS Relational Database Service (RDS).

Automated backups of the PostgreSQL database are captured daily and retained for a rolling period covering the most recent 7 days before they are deleted.

Application Security

Seer Data runs its applications as scalable Docker containers on AWS Elastic Container Service (ECS) providing the ability to easily scale up or down the number of running instances.

Service isolation is implemented by deploying the Seer Data Platform multi-tiered application as separate services within environment specific AWS Elastic Container Service (ECS) clusters. Communication between the services is enabled via AWS Elastic Load Balancers (ELBs) where requests are forwarded to the appropriate service target groups.

Administration of the Seer Data platform, including User details, records of content, and activities on Seer Data are stored in PostgreSQL databases running on AWS Relational Database Service (RDS). Access to Seer Data PostgreSQL databases is restricted to authorised Seer Data personnel, authenticated via username and password authentication methods.

Seer Data uses AWS Simple Storage Service (S3) to store media and files that relate to Users' specific platform usage, including items such as profile pictures, video assets, and PDF attachments. Access to Seer Data's AWS S3 environment is restricted to authorised Seer Data personnel, authenticated by AWS Identity Access Management (IAM)'s password-based authentication protocol.

Where authentication protocols rely on passwords, these are auto generated where possible and managed within a password storage application called 1Password.

Employee Access

Seer Data employees interact with our platform services via the AWS console. The AWS Console is a graphical user interface (GUI) designed to enable easy management of all AWS services without using code or manual commands. This approach reduces the likelihood of human error.

Members of the Seer Data Engineering team log into the AWS console with individual AWS Identity Access Management (IAM) accounts in order for AWS security policies to be applied according to the members role. Use of the root AWS account is heavily restricted.

When server access is required, authorised employees use the Secure Shell (SSH) network communication protocol, involving a public and private keypair, to log in to the Seer Data Platform virtual machines running on AWS.

Logging & Auditing

All systems utilise a central logging store in AWS CloudWatch.

Audit logs are protected with tamper proof controls and are reviewed on a periodic basis.

Seer Data Platform User activity messages and alerts that assist with Seer Data business goals, or to proactively manage the security and availability of the production environment, are logged to our instant messaging platform with notifications configured to alert the appropriate personnel.

Firewalls

Access to AWS EC2 servers is managed through the inbound and outbound rules specified in AWS Security Groups. Developer access is granted for a time-limited period when there are management, configuration or debugging needs that warrant access.

Glossary of terms

Concept	Definition
Aggregate and Unit Record Level (URL) Data	Aggregate Data is information that is comprised of statistics summarising an underlying set of Unit Records. The collection of Unit Records are Unit Record Level (URL) Data. For example, the values representing the number of people who live in a certain area or belong to certain demographics is Aggregate Data, while the individual records describing each person and their location and demographic information are Unit Records. Pivot Tables, Contingency Tables, and Cross Tabulations are all common formats for presenting Aggregate Data.
Datasets	Datasets are collections of data in the Seer Data Platform. Each Dataset belongs to one (and only one) Organisation. Access in full or in part to a Dataset for Discovery and Querying is administrated by the Owner of the Organisation to which the Dataset belongs. Datasets that Seer Data makes available to all Users are "Open Datasets", while Datasets that belong to Organisations who administrate access to them are "Private Datasets".
Insights	An Insight is a document in Seer Data which presents data, Queried from a Dataset, in the form of a Table and Chart(s), as well as space for observations and storytelling. Insights are saved in Suitcases.
Open Data & Open Data Library	Open Data is data that is in the public domain. Seer Data ingests a variety of Open Datasets into the Seer Data Platform and makes these available for the benefit of all Seer Data Users for free. The Seer Data Open Data Library contains more than 30 Open Datasets.

Concept	Definition
Organisation	An Organisation in the Seer Data Platform is one or more Seer Data Users organised as a group. Seer Data Users can be members of more than one Organisation. Each Dataset in the Seer Data Platform belongs to an Organisation, so the Organisation is the Custodian of the Dataset in Seer Data (the Custodian Organisation). Each Organisation has one (and only one) User who is the Owner of the Organisation. The Organisation Owner is responsible for administrating members and assets such as Datasets. Organisations can also have Administrator Users designated by the Owner who are able to assist the Owner in administrating the Organisation.
Partners	Organisations can agree to Partnership in the Seer Data Platform. This enables these Organisations to grant access to discover and query their Private Datasets, as well as share Suitcases.
Personally Identifiable Information (PII)	Under Australian law, Personally Identifying Information (PII) is defined under the Privacy Act 1988 as *"information or an opinion about an identified individual, or an individual who is reasonably identifiable"*. The Office of the Australian Information Commissioner (OAIC) provides commentary and expansion on this definition. It is the policy of Seer Data that all Data (Aggregate and Unit Record Level) hosted in the Seer Data Platform does not contain PII. Seer Data assists Organisations to process their data to remove PII data and protect privacy prior to uploading their data to Seer Data.
Private Data	Private Data is any data hosted with Seer Data that belongs to an Organisation (the Custodian Organisation) and is available in full or in part only to those Users or Organisations that the Custodian Organisation designates.

Concept	Definition
Querying Datasets	Datasets are available to be Queried in the Seer Data platform by selecting and arranging available Variables and Categories into a Query which describes a request for data from the Seer Data Platform.
Seer Data API	The Seer Data Application Programming Interface (API) provides an alternative means for interfacing with the Seer Data Platform. The Seer Data API can be used to import data and related content directly from Seer Data into third party software environments. Users are required to Authenticate through the Seer Data Platform before accessing the Seer Data API and can only access Datasets and Content through the Seer Data API that they would have access to through the Seer Data Platform web application.
Sharing Private Data	Private Data can be shared through the Seer Data Platform. Private Data can only be shared between Partner Organisations. Access to Private Data can be granted to individual Users or to whole Organisations. Access can be granted to the whole Dataset, or to parts of the Dataset based on Variables and Categories specified by the Owner of the Custodian Organisation.
Suitcase	A Suitcase is a space within the Seer Data Platform for saving Insights and collaborating with other Users. Suitcases can contain other Suitcases, creating a tree structure. Each Suitcase belongs to one (and only one) User, the Owner of the Suitcase. The Owner of a Suitcase is also the Owner of all content inside it, including Suitcases and Insights.
Suitcase Sharing	The Owner of a Suitcase can grant Full Access, Read Only, or No Access to the Suitcase to other Users in their Organisation or Partner Organisations. If a User is granted access to a Suitcase, that User will have the same level of access or greater to all Suitcases within it. Suitcase Owners can also publish the Suitcase with a Public Link, which allows non-Users to view the Suitcase content as Read Only.

Concept	Definition
User Identification & Authentication	To access the Seer Data Platform, Users are identified by email address and authenticated using a one-time password which is sent to the user's verified email address.
User / Member User	A User is a person with access to the Seer Data Platform at any subscription level. Individuals accessing Suitcases published with Public Links do not need to be Users.

The Seer Data Platform is continually enhanced and updated with new features to meet customer needs, requiring amendment of this and other referenced documents. Seer Data & Analytics (Seer Data) reserves the right to alter this document and other referenced documents without notice from time to time and at its sole discretion. Seer Data makes commercially reasonable efforts to ensure that the information in this document is accurate and up to date, but errors may occur.