Secure safehaven for the ASPREE clinical trial – The need

8 March 2021
Categories: MeRC
Tags: Research Stories, Using the Cloud

The year 2017 began with the ASPREE Data Management Team seeking advice on an emerging need for a collaborative sensitive analysis environment. Back then HPC and clouds were very much the realm of categorically non-sensitive data, and secure ("red zoned") systems were very much the realm of categorically non-collaborative data. This bifurcation was rife in health and social sciences. This Research Cloud engagement with ASPREE was seminal work to transition the Monash research environment towards a continuum (rather than bifurcation) between collaboration and sensitive expectations.

The ASPREE team had a single commodity physical PC located at the ASPREE office in the School of Public Health and Preventive Medicine. Despite the ASPREE team streamlining processes to appropriately allow project collaborations, an innovation in its own right, collaborators could only perform analysis by being physically in the office. The protocol required data custodians to copy ASPREE phenotypic datasets (via USB sticks) onto the PC, whilst also physically disconnecting the ethernet cable to ensure no unintended access. Collaborators would fly into Melbourne just to run their analysis. This logistically-taxing workflow made collaboration hard and significantly delayed research outcomes. As new project requests emerged from ASPREE sub studies, it became apparent that data management, data governance and the analysis ecosystem would need to be revamped to support the growing demand. A scalable and secure "safe haven" was required.

The Research Cloud team approached the situation from a pragmatic point of view. The team first critiqued the scalability of the analysis environment. We discovered the environment would require security-hardening to protect against intentional and unintentional data leakage. Furthermore the interfacet needed improvement to become intuitive for non-academic, external and international collaborators.

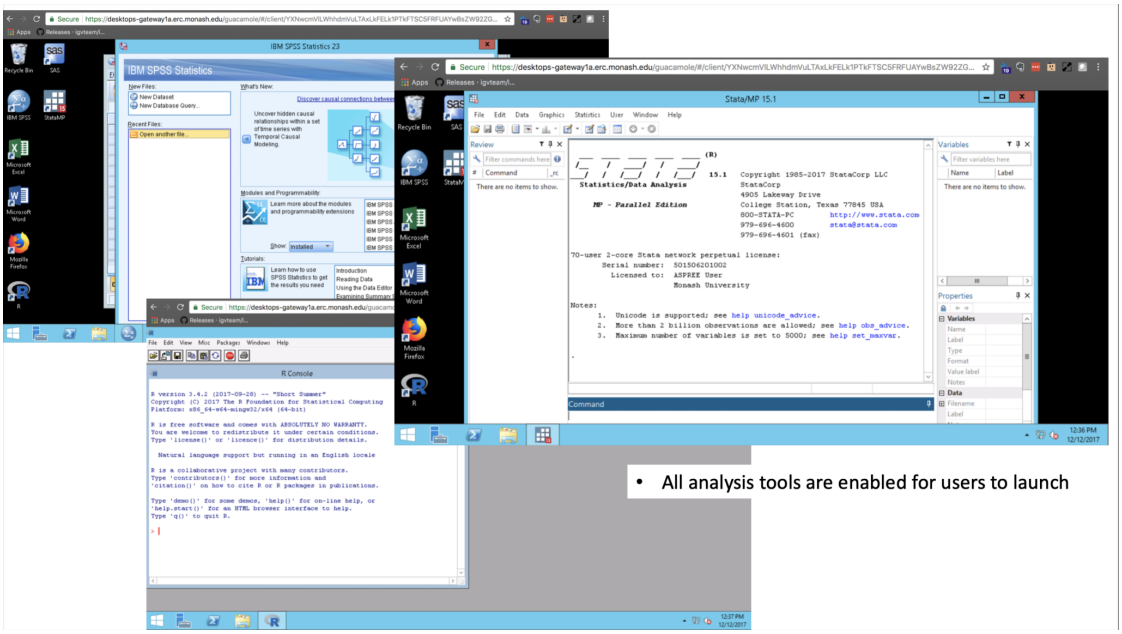

Figure 1. A typical ASPREE Analysis Environment

Fortunately Leostream, a remote desktop (VDI) scheduling platform, was already being used by virtual laboratories on the Monash zone of the Research Cloud. Leostream provides a high-level interface for allocating remote desktops to users. It also allows access to these remote desktops through a web-based (HTML5) viewer. To scale out the analysis environment, the team deployed a number of Windows-based instances on the Monash zone of the Research Cloud. These instances have been pre-configured with analytical tools chosen by the ASPREE community (e.g R/RStudio, SAS, SPSS) and connected to the Monash license servers. A typical analysis environment is shown in Figure 1. above. Access is managed through the Monash Active Directory (Domain) and each analysis instances have been configured with Group Policy Objects (GPOs). These GPOs enforced a number of rules or security controls inside the instances, e.g preventing users from changing desktop settings, access to registry tools and much more. A reserved set of hypervisors have been used to host these secure instances, which also reside on a segregated private network. Hyper-threading has been turned-off on the hypervisors to minimise the risk of Spectre/Meltdown-type vulnerabilities.

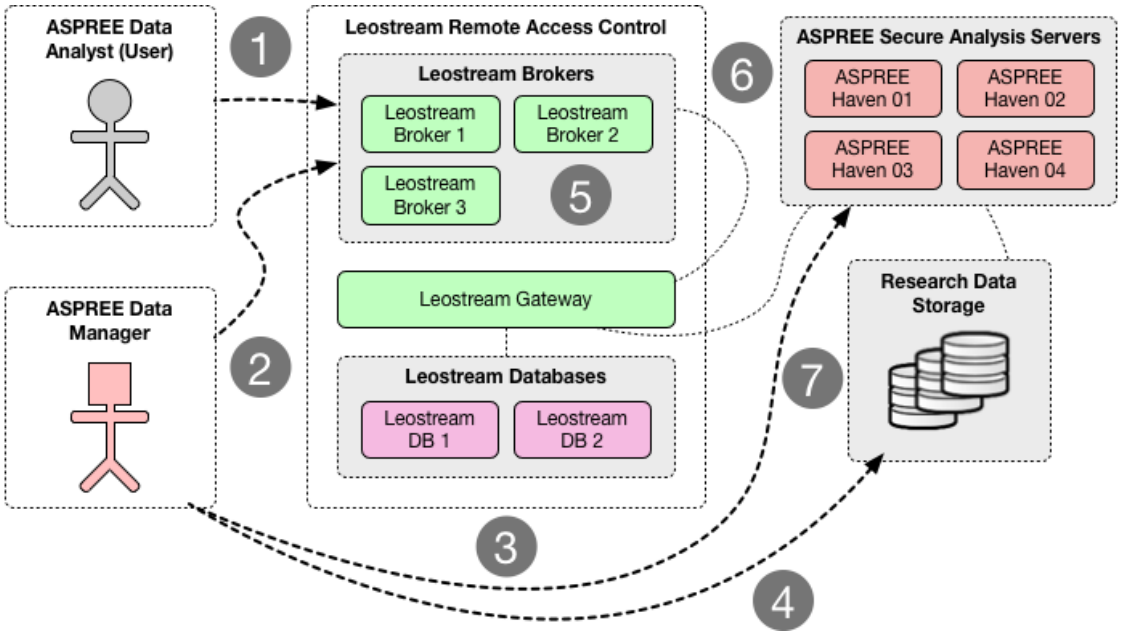

Figure 2. ASPREE safehaven architecture

A high-level architecture diagram for the ASPREE safehaven is shown in Figure 2. Monash eResearch Centre's Research Data Storage (RDS) provides a scalable storage backend to the safehaven. The team augments the storage pool with further controls to appropriately segregate the data. A dedicated user share is created for each approved ASPREE user. This user share is autonomously mounted into the analysis environment upon user login. ASPREE data custodians (managers) have elevated rights to the safe haven storage. They can review (approve or deny) what data goes in (ingress) and data going out (egress). Thus the technology / workflow automates the overall data governance of the ASPREE clinical trial by incorporating it to their own access management system (AMS).

Now operational for more than 3 years, the Research Cloud at Monash and Helix teams cooperate to provide user support for ASPREE safe havens. Several other registries and clinical trials have leveraged this ASPREE solution as their own safe haven. To date, over 100+ internal and international collaborators have used the ASPREE safe haven. This work has be foundational to Monash eResearch, the Research Cloud and Helix's initiatives towards the next-generation safe havens (e.g. SeRP, which further automates and audits generalised governance workflows).

> *"The ASPREE data is an NIH-supported clinical trial, and the NIH rightly demands full accountability for data handling. The team has been understanding, professional, flexible and fast. They gave extra consideration for ASPREE's urgent need (in 2016-17) to share our large and unique dataset to collaborators, whilst also supporting confidentiality in an active clinical trial. The co-design approach took into consideration our Data Manager's detailed requirements and produced an excellent environment for effective use and international collaboration centred on ASPREE data. The successfully funded extension study ASPREE-XT depended on getting this right."*
>
> *Dr Carlene Britt, ASPREE Senior Research Manager and ASPREE Data Custodian*

This article can also be found, published created commons here [1] (https://rcblog.erc.monash.edu.au/blog/2021/03/secure-safehaven-for-the-aspree-clinical-trial-the-need/#easy-footnote-bottom-1-2537).

1. Revote, Jerico; Aung, Swe Win; Quenette, Steve; Padmanabhan, Komathy (2021): Secure safehaven for the ASPREE clinical trial – The need. Monash University. Online resource. https://doi.org/10.26180/14176766.v4 (#easy-footnote-1-2537)