XENON eResearch NZ 2026 – Building Small Scale Trusted Research Environments
SLIDE NOTES

Slide 1

Hi

Slide 2

We'll start with asking the question, what is a Trusted Research Environment (TRE)?

We'll take a look at the where Trusted Research Environments come from.

And then we'll review some approaches for establishing one in an organisation.

Slide 3

What exactly is a Trusted Research Environment?

Based on my experience, it's a carefully managed computing environment that enables researchers to use specific software to securely access and process data.

With the key elements being:

- A user accessible computing environment configured with appropriate software and tools.
- User access controls, to ensure only the right people have access to the computing environment.
- And data access controls, to ensure only the right systems and users have access to their relevant datasets.

In addition, there are two more elements that I feel are worth including because of how useful they are for establishing and maintaining a TRE:

Data lifecycle management, to ensure project data is properly curated through its lifecycle.

And a reproducible computing environment, which enables rollbacks in case something goes wrong and allows the user to return to earlier versions of the data processing software in case newer versions change the output in unexpected ways, or don't support older data formats.

Slide 4

The term ‘Trusted Research Environment’ started showing up on the internet over the last few years.

There are a few other ways people label the same idea, such as:

Secure Data Environment,

Secure Access Environment,

Secure Research Environment… You get the idea.

There’s also a Five Safes framework that’s referred to by several groups who have written about implementing these environments.

The Five Safes Framework focuses on how to manage the risks associated with processing sensitive data in a way that protects privacy while maximising research value.

The framework was initially published by the UK Office for National Statistics in 2003 as the Virtual Microdata Laboratory Security Model.

Slide 5

When it comes to what shape these Trusted Research Environments can take, we can look to the major cloud providers for examples.

Here’s a scalable architecture published by Amazon.

You can see some key elements, Researchers (very important)

Various controls and governance across different aspects of the environment.

Specific mapping of data sets to projects.

And even specialised software and code pipelines.

AWS even has an implementation diagram breakout out the various services…

Slide 6

Wow, OK so lots of parts here. Definitely geared towards the larger and more scalable side of things.

If you squint, you can see the user environment, access controls, data sources and data pipelines.

Slide 7

Here’s an architecture diagram produced by Azure.

While there's a notable lack of researcher icons present, we can still see data pipelines and workflows, controlled data ingest, and specific mapping of datasets to projects.

We can also see the computing environments configured with different tool sets.

Slide 8

In Azure's breakout diagram we can see user environment, access controls, and indications of how this system could scale.

The core elements are there, along with several optional components to help with scaling and monitoring.

Slide 9

Taking a closer look at the Five Safes framework, we can see it's taking a wider view of the system.
To expand on each of the elements:

Safe Projects: Is the data being used ethically and for public benefit?

Safe People: Are the users trustworthy, trained and authorised?

Safe Data: Has the data been anonymised?

Safe Settings: Is the computing environment secure, preventing unauthorised access and data exfiltration?

Safe Outputs: Have the results been reviewed to confirm no identifiable information is published?

All good points worth considering depending on your TRE's purpose.

Slide 10

For a small, on-premise Trusted Research Environment, the architecture could look something like this.

Here we have a data source, perhaps an instrument, writing to a specific project folder.

There's a user environment, configured with specific software. Perhaps the user environment can be virtualised to make snapshots and environment reproducibility easier to manage.

We'll need to implement user access controls across each system, so some kind of user and role management system will be required.

We could also implement some basic data security, perhaps the user environment can have read only access to the project's raw data directory and write access to the project's processed data directory.

Slide 11

So, where do trusted research environments come from?

Ideally, your organisation recognises the value in creating a TRE. If that's the case, then the TRE project can be driven by leadership coordinating between the research, governance, project, finance and IT groups.

Though if your organisation doesn't already have a TRE, it might not be obvious to everyone that one is required.

If that's the case, you may see a Trusted Research Environment organically form in the wild. Based on my experience, this usually begins with the researcher, triggered by a change in requirements, such as:

- Change in the data source such as planning for a new instrument or data source to come online.
- New data processing software.
- Perhaps data processing has outgrown easily accessible compute (which is usually just local workstations).
- There could be a requirement to develop and run unsigned code, not something most corporate IT Standard Operating Environments are geared for.
- Perhaps a change of team dynamic, or a requirement to collaborate with external users.

Slide 12

In its early stages, the TRE may face friction as requirements are presented to the various stakeholders involved.

I've found one approach that helps is to take people along for the journey. Help people to see how the success of this project aligns with the goals of the organisation and their team.

Work with stakeholders at a leadership level to find agreement on the project's vision and collaborate with them to build out a roadmap. What details do people need from you to help make this project a success?

Do storage and compute systems require additional capacity? How can these requirement be modelled?

Do policies need to change to support unusual software and enable external collaboration?

Is a change in workflow required to scale beyond local workstation processing to a larger computing environment?

Slide 13

Even when a basic Trusted Research Environment is established, without appropriate ongoing support it may not be able to adapt and survive.

Plan to onboard ongoing resources and establish funding to provide support into the future. Does your organisation like to plan things 5 years out? Perhaps 10 years? These costs should be modelled so there are no surprises down the road.

Now you may be asking yourself, how can we avoid the too hard basket?

Slide 14

It depends on your organisation's business process maturity.

Perhaps you have existing frameworks in place in your organisation that you can use.

Within the research team, there may be some experience with similar changes. How are projects approved and prioritised? How is funding allocated? Talk with people involved in previous projects, find out what worked well and what could be improved upon next time.

Do you have a governance and project team? Find out what they like to know, how they like to be involved and start to build out a roadmap with them.

For the first phase of the project, aim to establish a minimum viable product. Don't let perfect be the enemy of good. Aim to make incremental improvements in subsequent phases.

The finance team love expenditure forecasts, the more accurate the better. They will probably have feelings about the level of detail they require, and how far into the future they like to plan.

Close collaboration with the IT team will be critical to the success of the project. An IT team who is enthusiastic to support the research will need to understand the requirements from an IT perspective, things like data generation rates, data

storage over time, estimated compute and ram requirements for a typical and worst-case scenario.

On the other hand, some teams find embedding specialist scientific IT resources can help align IT support with scientific IT requirements.

While there will be technical implementation challenges in establishing a TRE, building the business case for IT resourcing to support both the initial implementation and ongoing maintenance may actually be the greater challenge.

Slide 15

There will be questions about the compute and storage required, not just initially but also projected into the future. This information can be defined using tools such as compute and data storage policies.

You can use data policies to define the lifecycle of the different categories of data.

Include information about data storage compliance obligations and when data may be deleted.

Specifics around the size of the data will be useful to include to help with data storage capacity planning. If you have any known increases expected in the future due to new instruments or data sources, be sure to add them too.

The policy can be owned by the research team and should be seen as a business rather than technical document.

Slide 16

Combined with historical and projected data storage utilisation, you can build a model that feeds into a data storage capacity plan.

The storage capacity plan can be owned by the IT team, where the planned capacity is a flow on effect of enabling the data policy.

This chart shows data storage requirements projected into the future, including a reserve to mitigate the risk of underestimation. You can also see a planned expansion every second year represented by the dotted line.

A similar kind of policy and capacity plan can be created for compute resources too.

Slide 17

In summary, I recommend starting with the business case, aiming for a minimum viable product and highlighting how the project aligns with the organisation and the team.

Reach out to your stakeholders, anyone who can recognise the value of the project and is interested in seeing this project succeed. Find out how they would like to be involved and what they need from you.

A phased implementation roadmap can help everyone pull in the same direction. Include an outline of future phases to highlight the fact that this is a living project requiring maintenance and support to keep up with the ever-changing requirements of research into the future.

Take ownership of, and create compute and data storage policies if they don't exist. Contribute to them if they do, they will help ensure the services provided by IT align with research requirements.

Slide 18

I hope this talk has been helpful, thank you for your time.

<<Talk Ends>>