

DARE UK

TRE-FX

**Core federation services for a federated
network of TREs to enable Five Safes analytics**

Carole Goble (ELIXIR-UK, HDR-UK, University of Manchester)

Phil Quinlan (ELIXIR-UK, HDR-UK, University of Nottingham)

TRE-FX

<https://trefx.uk/>

Federated Analytics toolkits



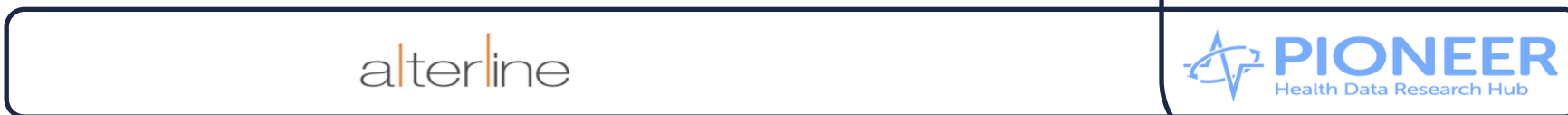
Infrastructure providers



TREs

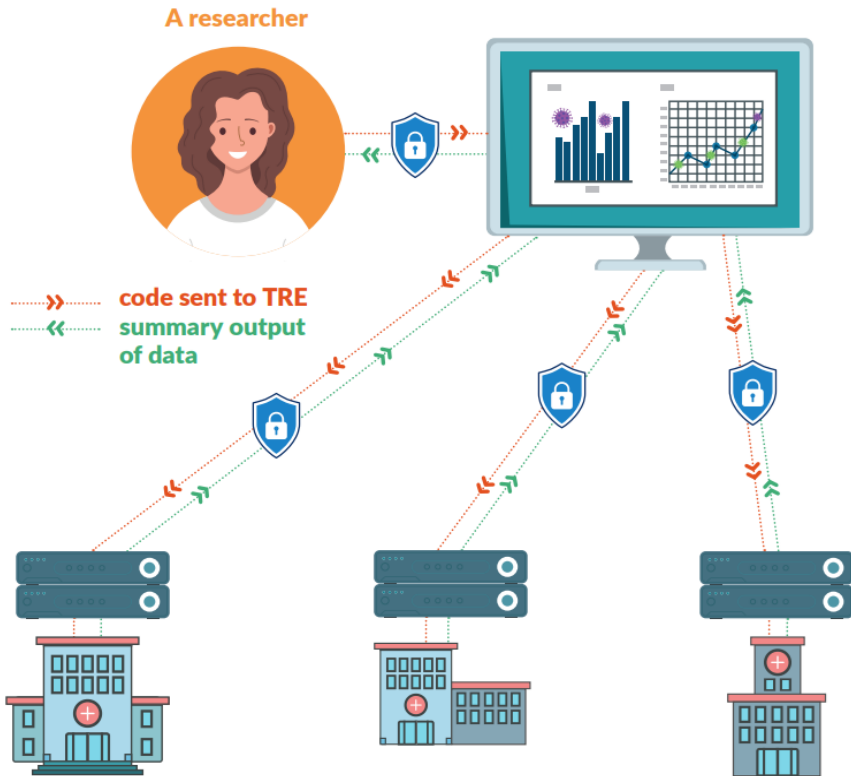


PIE



TRE-FX

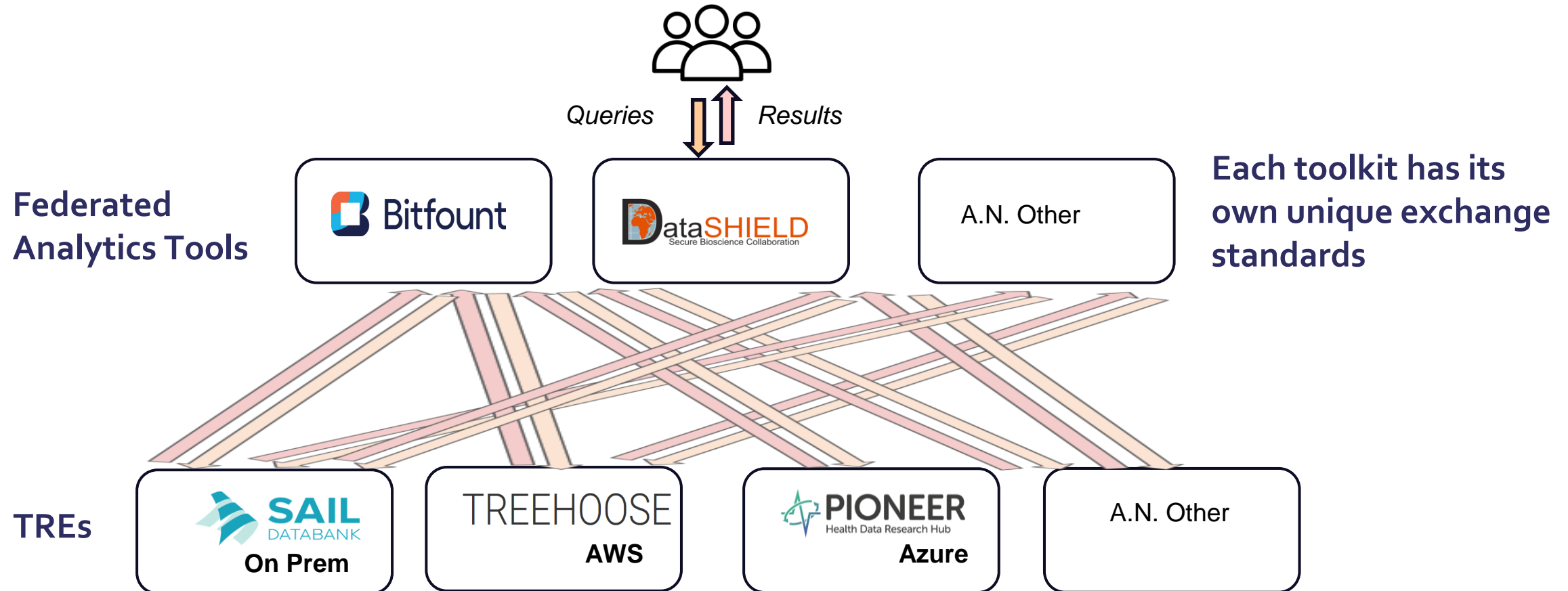
Federated Analytics: Researchers run their analysis at many TREs and combine the results.



How do we reduce the burden on a Researcher to query many TREs to answer their question

How do we reduce the burden on existing TREs and analysis toolkits to implement that within the Five Safes framework

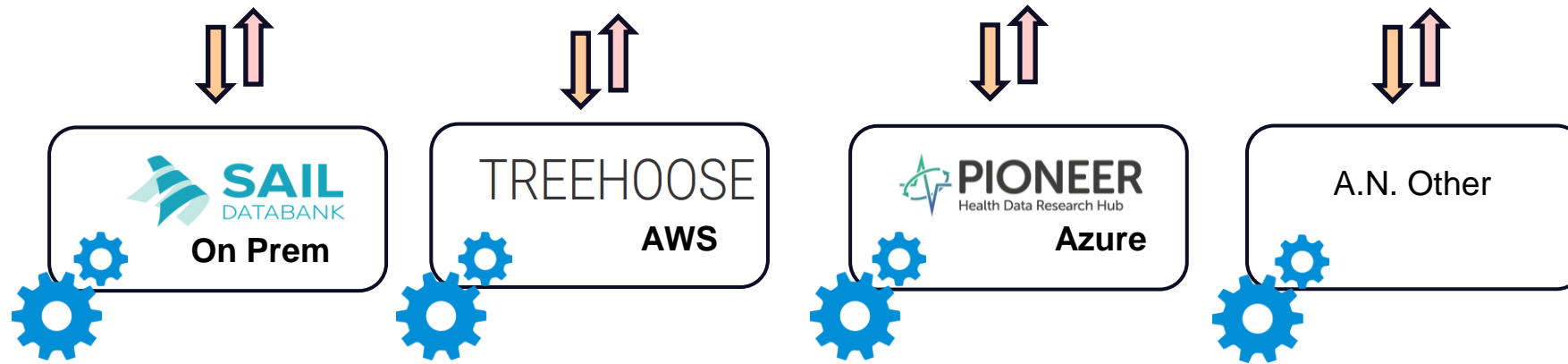
Reduce the burden of analysis across multiple TREs **TRE-FX** by streamlining the plumbing



Each TRE develops a bespoke solution, each researcher needs to make their code compatible
Must be **independently tested** against the Five Safes, and validated by PPIE & governance

1

Streamline how the processing is done by using adaptable and reproducible analysis pipelines



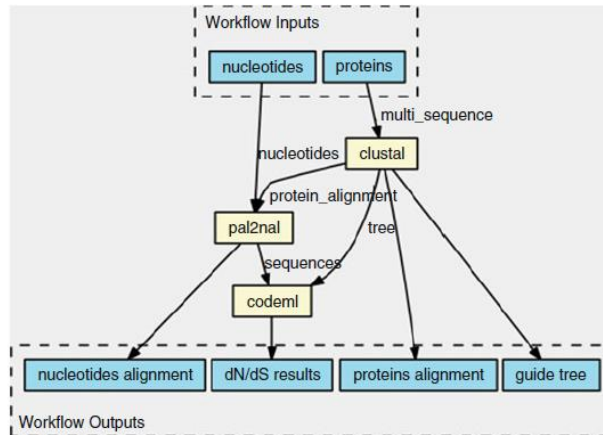
streamline how we run an analysis inside the TREs

Pre-approved workflows that are sharable and work with any technology
support transparency and reproducibility
could be peer reviewed for quality and methodology
handle all the hassle of making sure the code works in the TRE

1

Analysis using Computational Workflows

“Recipes” for linking processing steps on data.



Reusing technologies developed by European Research Infrastructure using International Standards



Global Alliance
for Genomics & Health



Workflow Registry to find and share



Run workflows made for different workflow systems



COMMON
WORKFLOW
LANGUAGE



Galaxy
PROJECT

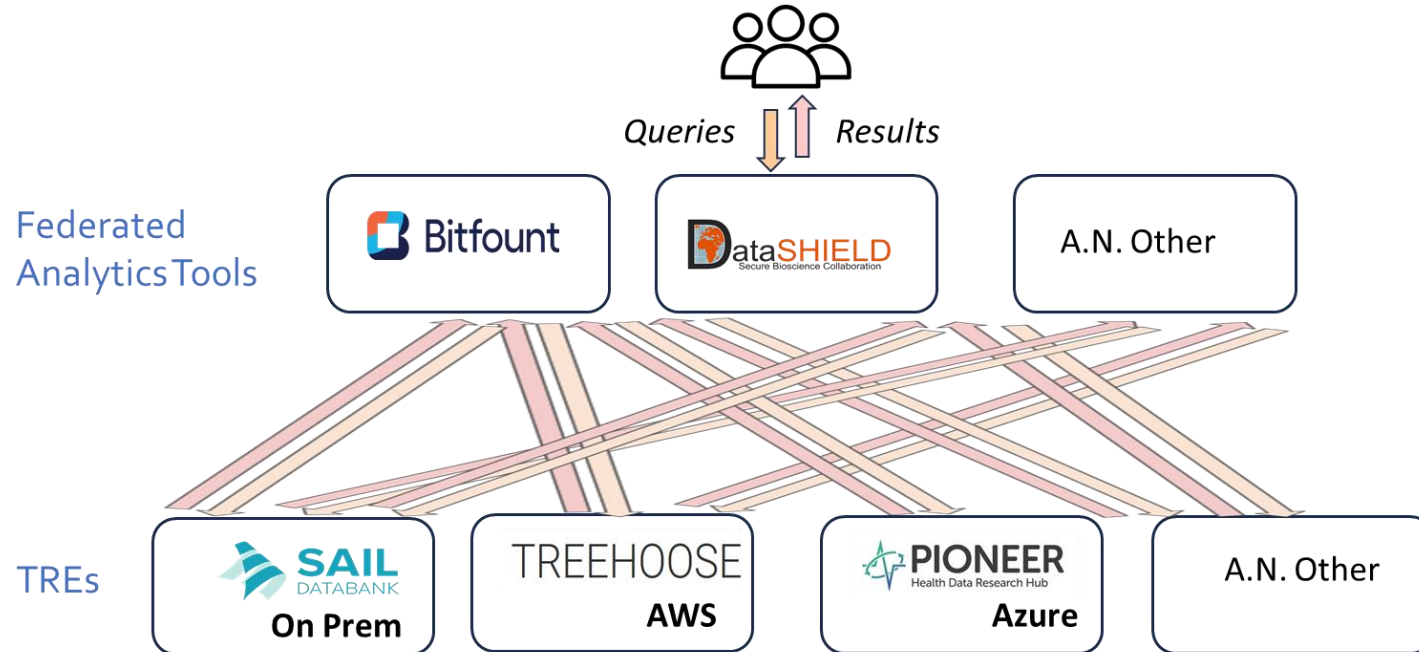


nextflow



2

Streamline the flow of information for Five Safes



standardise the flow of information (metadata)

works with all the platforms

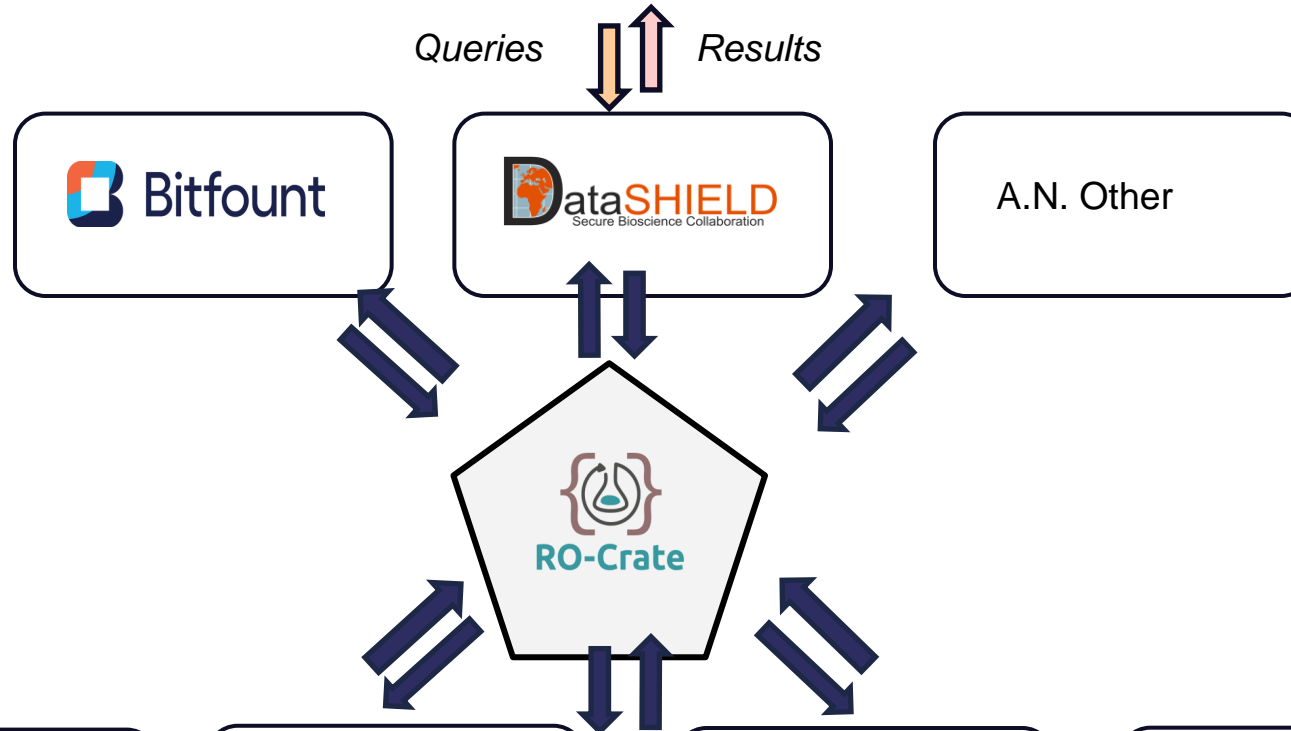
supports transparency and reproducibility

could be peer reviewed for quality and methodology

2

Standardise using Five Safes Digital Objects as carriers

Federated
Analytics Tools



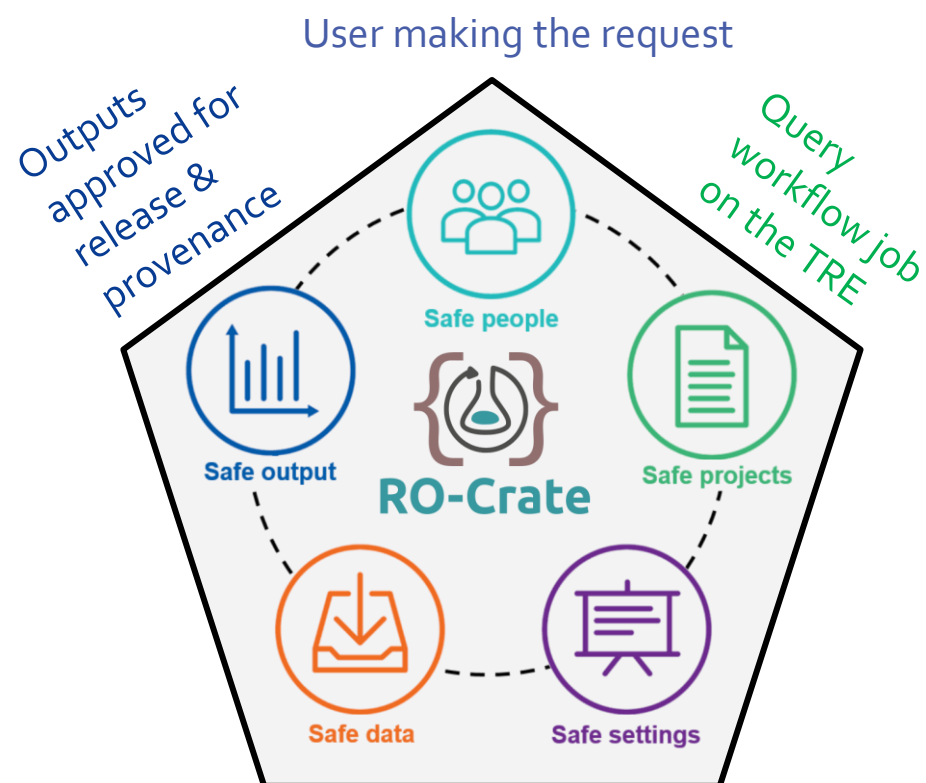
Each toolkit
handles RO-Crates



TREs

Each TRE handles RO-Crates

Five Safes RO-Crates carries, collects and documents the information needed: developed by our partners



Metadata of
dataset and ref to
dataset

Pre-approved
workflows and
containers on TRE

A specification the tools and TREs follow

5s-crate

Five Safes RO-Crate profile

Note: The organisation under `affiliation` is typically the employing organisation, e.g. a university or hospital. Virtual organisations such as research projects MAY additionally be listed using `memberOf` (see also [Responsible Project](#) below).

Responsible Project

The project that the request is sent on behalf of, typically related to permission to use a TRE, MUST be indicated from the root dataset using `sourceOrganization` to a [Project](#). The responsible project SHOULD be referenced from the requesting agent's `memberOf`.

Note: The *responsible project* is not necessarily a [ResearchProject](#) corresponding to a funded grant, but may be more specific studies within a funded project. Various TREs may have different granularity and identifiers for the responsible projects. A project `Grant` MAY be referenced using `funding` from the responsible project.

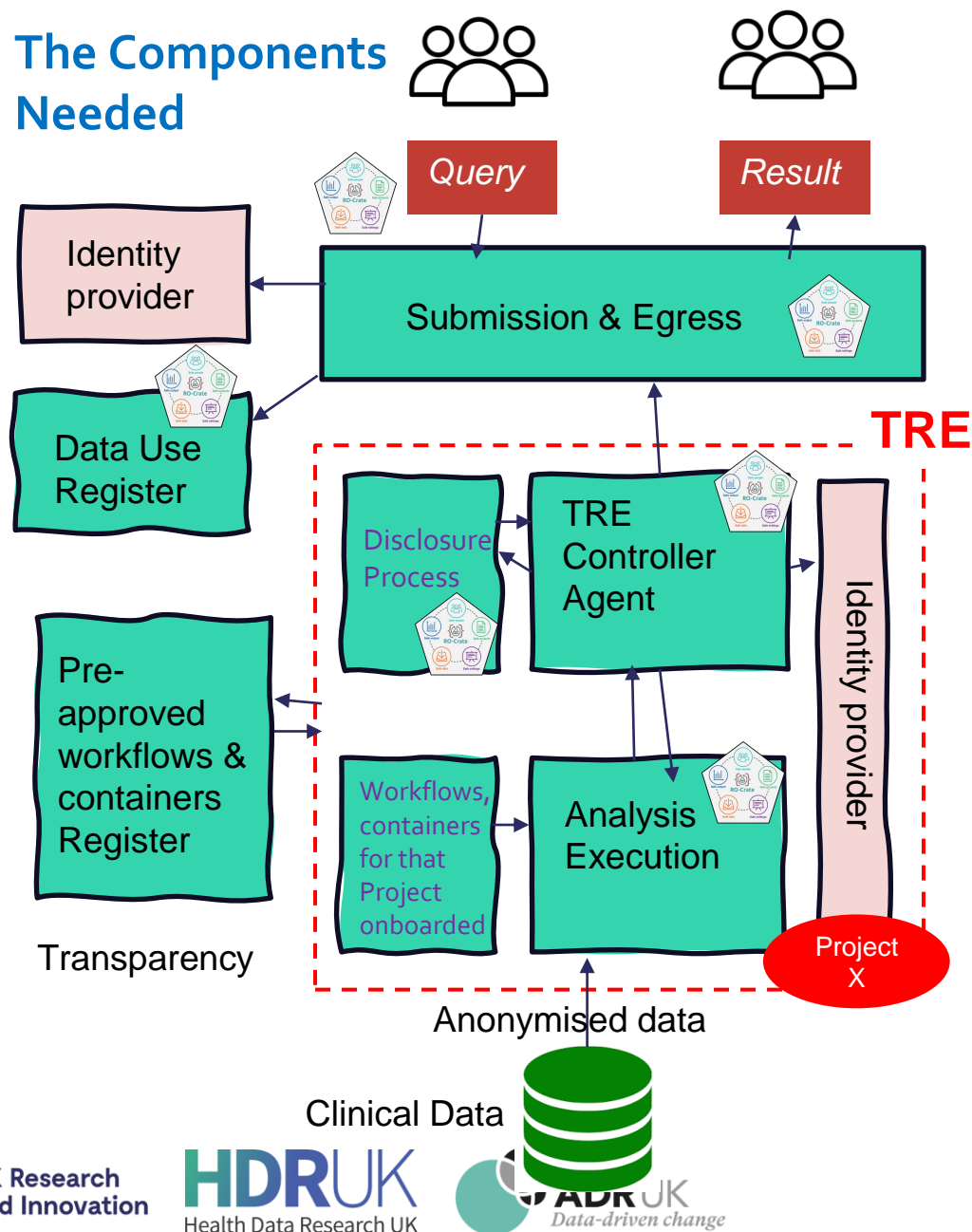
It is RECOMMENDED to include TRE-specific ids under `identifier` (which MAY be an array). If the identifier is not globally unique (e.g. an integer rather than an UUID or URI), it is RECOMMENDED to add a [repository-specific identifier](#) and provide the local identifier as `value` of a `PropertyValue` entity. Multiple repository-specific identifiers MAY be included for different TREs from a single Project entity.

The project MAY indicate the `member` organisations, in which case one of them SHOULD match the `affiliation` of the *Requesting Agent* with a `memberOf` to this project.

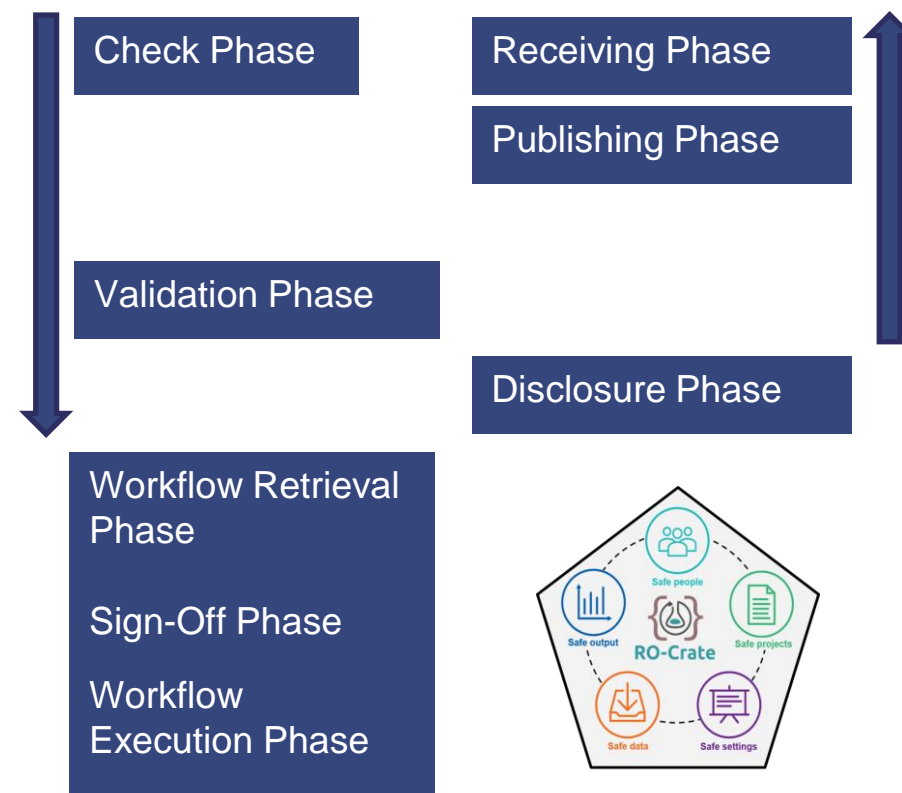
Example

```
{
  "@id": "#project-be6ffb55-4f5a-4c14-b60e-47e0951090c70",
  "@type": "Project",
  "name": "Investigation of cancer (TRE72 project 81)",
  ...
}
```

The Components Needed

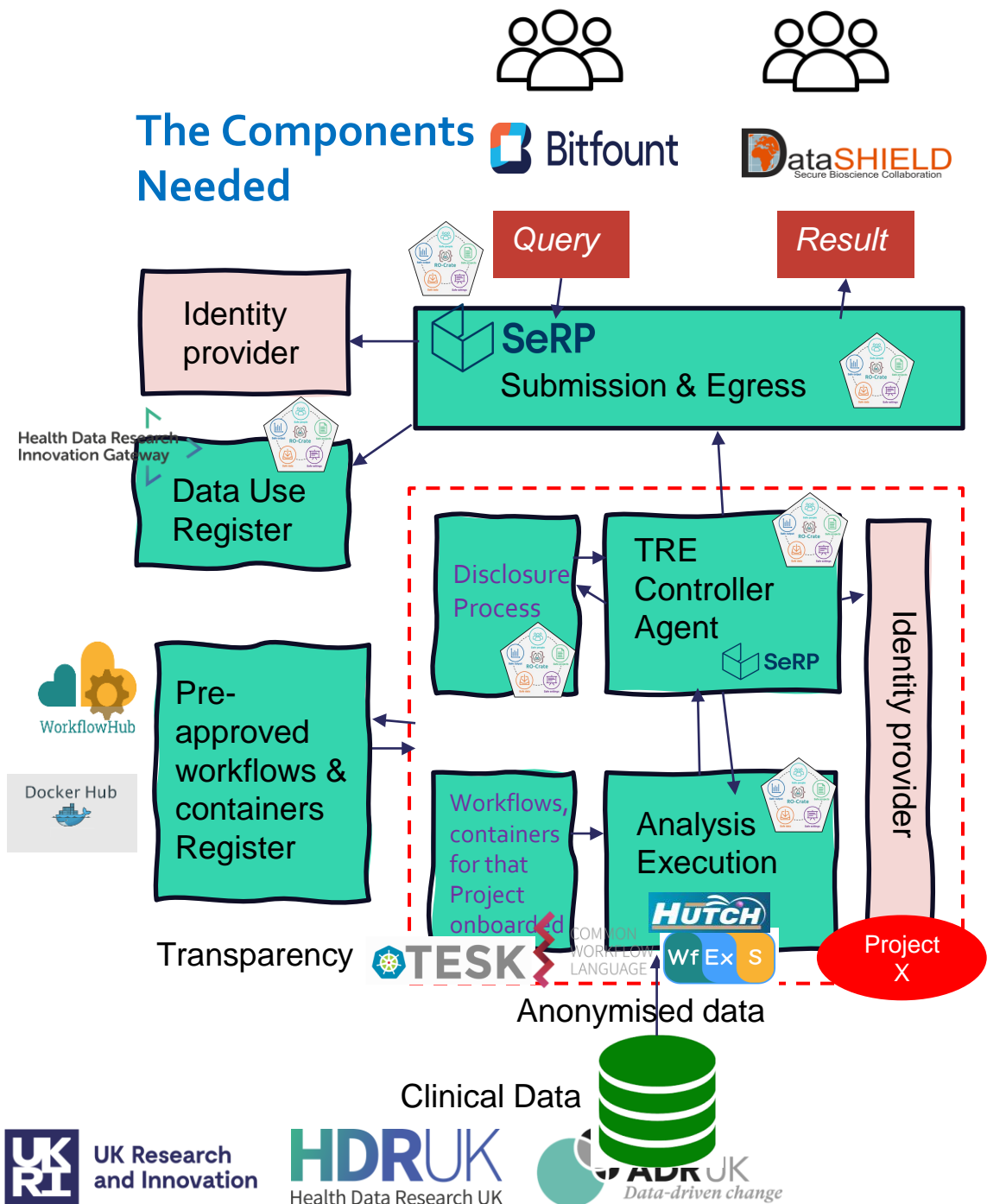


The RO-Crate travels around and collects & carries information at different phases. It doesn't touch or carry any data.



TRE-FX

Pilot implementations



Using RO-Crates, components in analysis execution



Bitfount

Using RO-Crates Execution manager with Bitfount Pods



Using RO-Crates

Submission layer

Execution manager

with WfexS and TESK

How the public were involved

PIE Leadership
Jillan Beggs



- Involved in Phase1 projects
- Supported the bid process
- Chaired bi-weekly meetings
- Input into the reporting

DARE UK

Glossary
of terms

alterline

8 Focus Groups

2 reports



2 Videos



2 Brochures

Highlights the need for clear and transparent communication, trust, accountability, and robust data governance to win over public trust.

Reflections



Accommodate existing and Five Safes TREs governance processes



Developer friendly and adoptable by existing systems and their software stacks



Flexible for different ways of federating, compatible with TELEPORT



Transparent and reproducible analysis



Reuse existing services and standards

Five Safes RO-Crate Objects standardize the information flow ✓

Technology agnostic mechanisms for running workflows –deployment detail into TRE software stacks ✓

Next steps - Operational deployments into TRE



- Onwards with technical deployments
- Development of process deployment
 - RO-Crate journey (what, when and where goes into RO-Crates)
 - How to approve the workflows and when to onboard them
 - How to organise the RO-Crates

TREs: TREEHOOSE and PIONEER

Moving Forward

HDR-UK Federated Analytics work programme

Inflammation and Immunity Driver

Combine with TELEPORT to cover different federated analytics patterns

Review other DARE Sprint results.

European work programme



Federated Analysis Platforms & TREs



DARE UK TRE Blueprint



DARE UK

Thank you for listening

<https://trefx.uk/>