



DARE UK

Interoperable Digital Infrastructure for Sensitive Data

Rob Baxter, Technical Lead, DARE UK

About Us: A pan-UKRI infrastructure programme for secure, cross-domain sensitive data research

“Enable a step-change in the UKRI’s/national research capability for secure sharing, linkage and analysis of sensitive data for research and innovation, timely and at scale for public benefit.”

DARE UK



Phase 1, Current Work

- Phase 1, Design and Dialogue
 - Landscape review
 - Community engagement
 - Public dialogue
 - Guided research
 - National and international best practice
- Phase 1a: pathfinding, 2021-2022
 - 9 sprint exemplar projects
 - Recommendations: <https://bit.ly/41I5mTM>
- Phase 1b: design, 2022-2023
 - 5 driver projects
 - Architecture 1/3: <https://bit.ly/3oiodGJ>

DARE UK

Paving the way for a coordinated national infrastructure for sensitive data research

A summary of findings to date from Phase 1 of the UK Research and Innovation DARE UK programme

August 2022

Federated Architecture Blueprint

DARE UK Delivery Team

The Sensitive Data Landscape: Where Are We Now?

- Across the UK we have a vibrant service and data ecosystem
- Trusted Research Environments (TREs)
 - Large and small: national, regional, institutional, even departmental
 - Old and new: from decades old to coming soon
- Sensitive data providers
 - Big: NHS, GEL, UK BioBank, ONS, Scottish Government, UK Government, Welsh Government, ...
 - Small: individual surveys, studies and trials
- Others
 - Indexing services, for linking datasets on different spines
 - Discovery services, for finding what's out there
 - Platforms for building or hosting TREs
 - Open data sources to be linked

Vibrant Ecosystem = Smörgåsbord of Tech

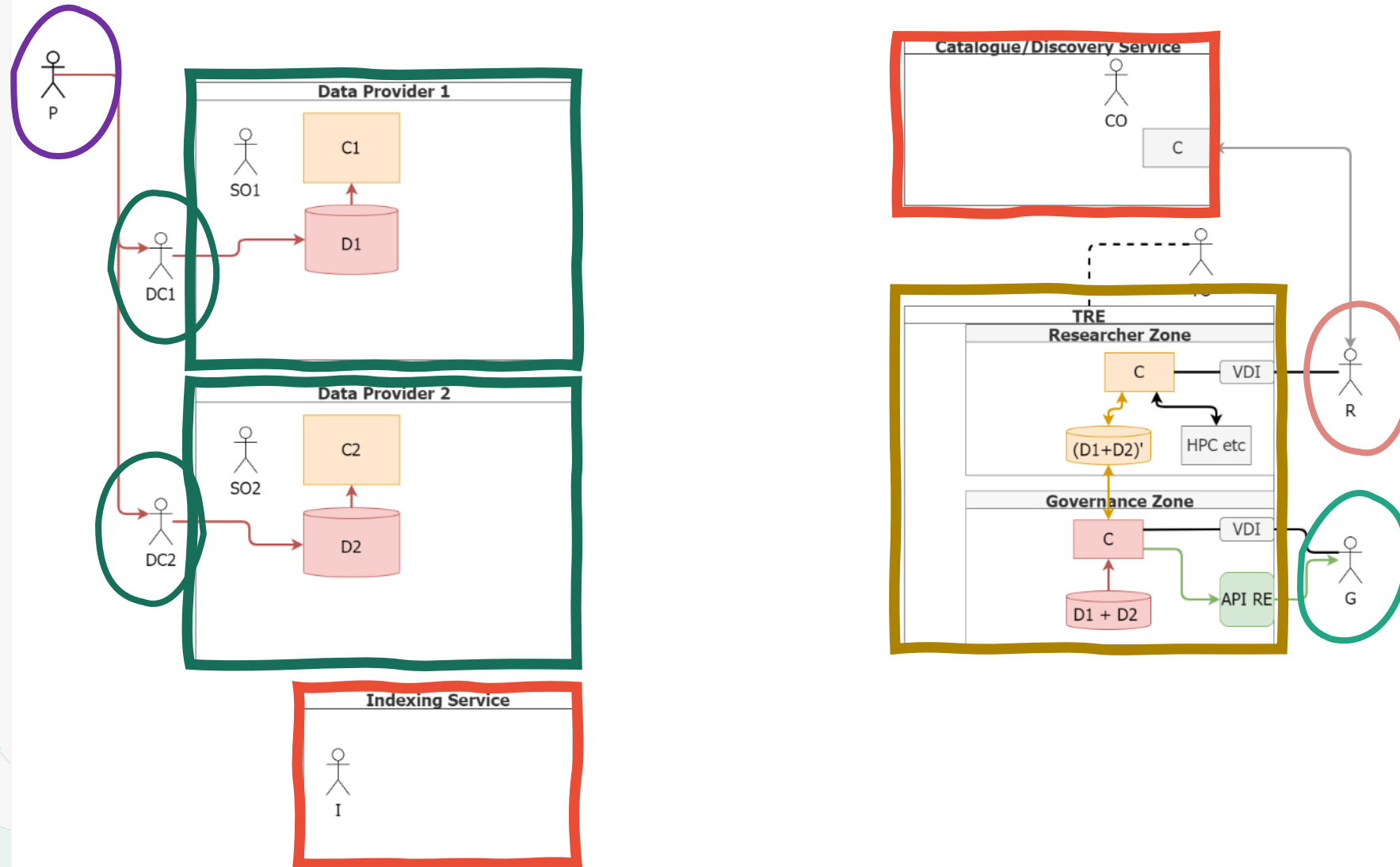
- TREs are generally cloud-like
 - Meaning virtualised infrastructure
 - But a mix of on-premises, Azure cloud, AWS cloud, GCP cloud, ...
- Data movement is always secure, but...
 - uses a variety of protocols and tooling (https, s/ftp(s), managed file transfer, ...)
 - sometimes (still!) follows the download-to-laptop model *
- Analytics services are a mix of traditional and modern
 - Stata, SPSS, SAS; R, Python, Julia, Scala; PyTorch, Tensorflow; CPU, GPU, TPU, HPC
- And we haven't even mentioned the data yet...

* This must stop! Download considered harmful! No TRE, no data!

Interoperation Between TREs and Data Providers

- A problem of three layers:
 1. Infrastructure layer: easy!
 - Services and functionality needed for inter-TRE federation
 2. Data layer: hard!
 - Discovery, linkage, encapsulation in projects
 - Syntactic (easy-ish) + terminological (harder) + semantic interoperability (ouch)
 3. Governance layer: harder!
 - Whether and how data may be used
 - Augment not replace existing (good) governance
- Our work to date focuses on infrastructure
 - Data and governance are TODO!

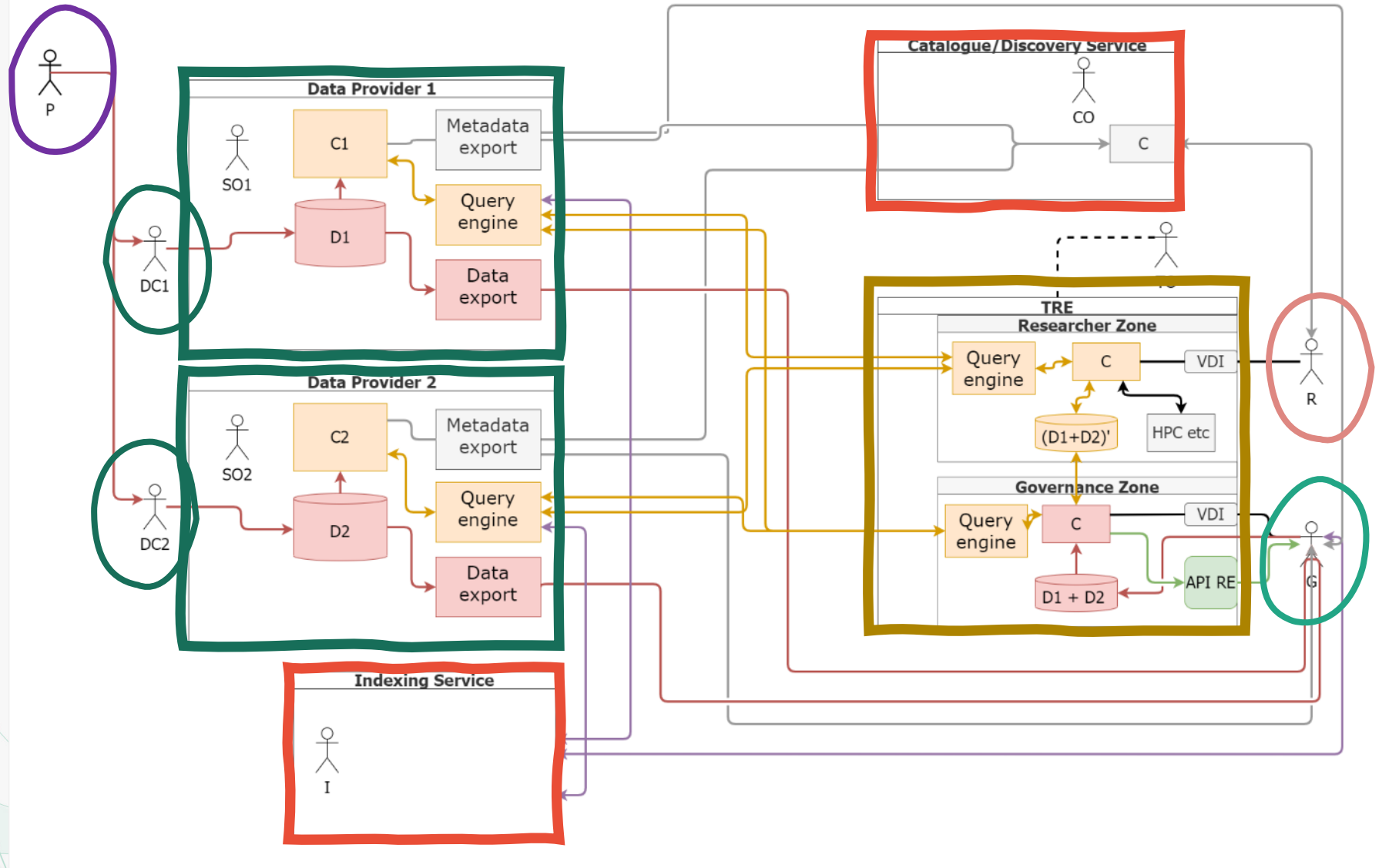
Infrastructure: from this...



As-Is

- Data providers, TREs and other services working together, kind of
- Ad hoc connections, varying tech and a lot of manual toing and froing
- It works, but it could be better!

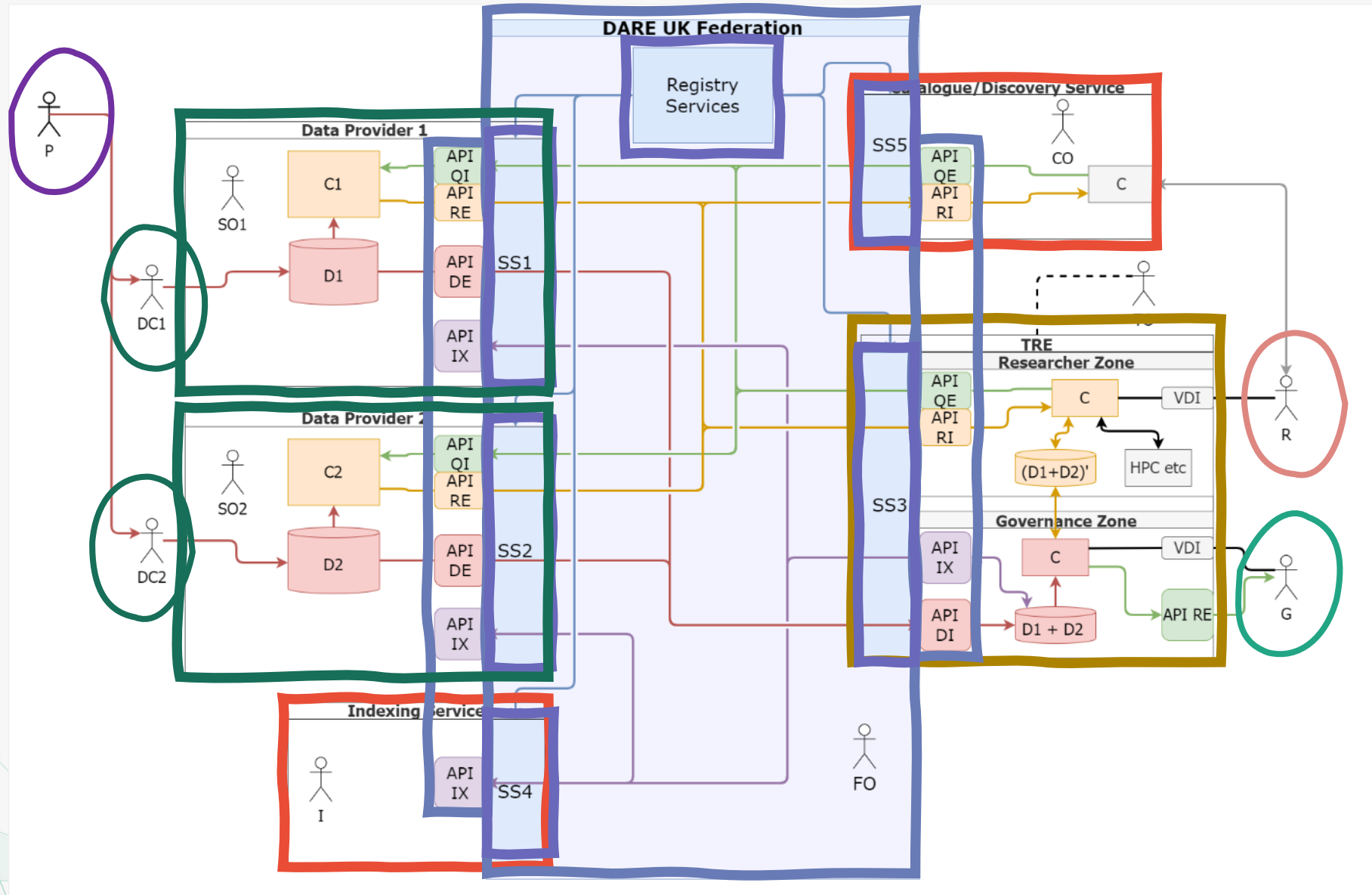
Infrastructure: from this...



As-Is

- Data providers, TREs and other services working together, kind of Ad hoc connections, varying tech and a lot of manual toing and froing
- It works, but it could be better!

Infrastructure: ...to this



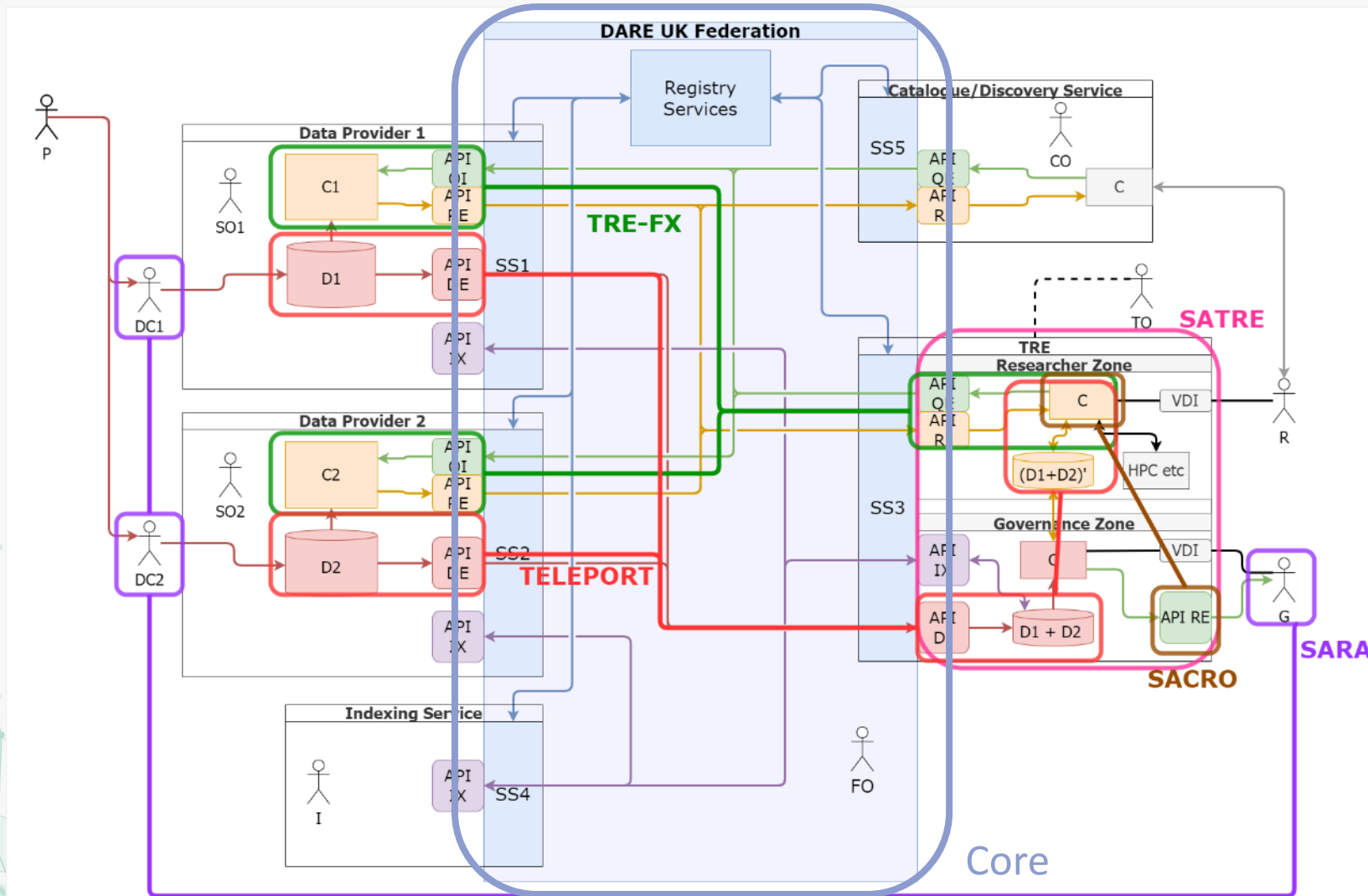
To-Be

- Standardise connections into “API types” (for role-based security control)
- Give each TRE, data provider and other service a standard security “gateway”
- Add a central set of registry services
- We get a high-trust network between all participants

Work-in-Progress: Core Architecture & Driver Projects

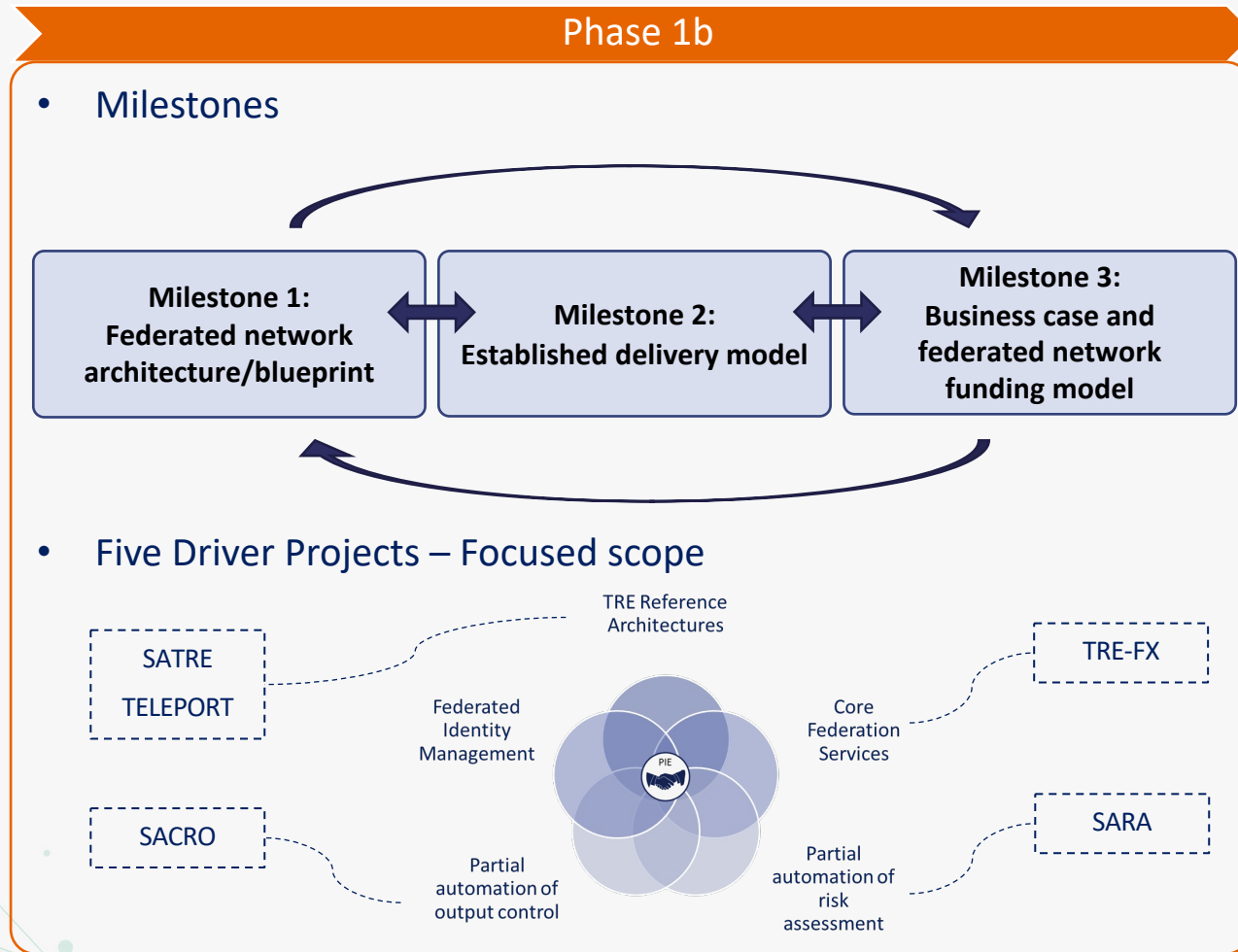
- **DARE UK Core** is developing the architecture for the federation and central services.
- **SARA** is developing and evaluating tools to support data controllers and information governance professionals in improving the risk assessment of unstructured free-text data.
- **SACRO** is extending the ACRO toolset for semi-automating statistical disclosure, providing researchers with a pre-disclosure validation check to reduce churn.
- **SATRE** is distilling existing TRE architectures into a common blueprint and reference implementation.
- **TRE-FX** is demonstrating the use of some ELIXIR technologies for information exchange in the federated query use-case (“horizontally partitioned” data).
- **TELEPORT** is demonstrating the use of “pop-up” TREs as a mechanism to encapsulate data in multiple policy wrappers when they have to move (“vertically partitioned” data).

Fitting At All Together



- We have a number of concurrent activities!
- Good coverage of overall needs
- Design, prototype, pilot, test
- Will fit together over course of 2023

The Programme: 2023 and Beyond





UK Research
and Innovation



DARE UK

Thank you for listening

dareuk.org.uk

© 2023 DARE UK, licensed under CC BY 4.0