CReDo

CATAPULT Connected Places

DATA SHARING PRINCIPLES, FRAMEWORK, AND ARCHITECTURE

December 2024

Delivered in partnership with: UKRI Innovate UK

CONTENTS

Executive Summary (4)

Introduction (6)

Purpose and scope of this document (7)

Oerview of Data Sharing in CReDo (8)

Objectives of sharing data (8)
Sources and users (9)
Permissions (12)
Terms of use (12)

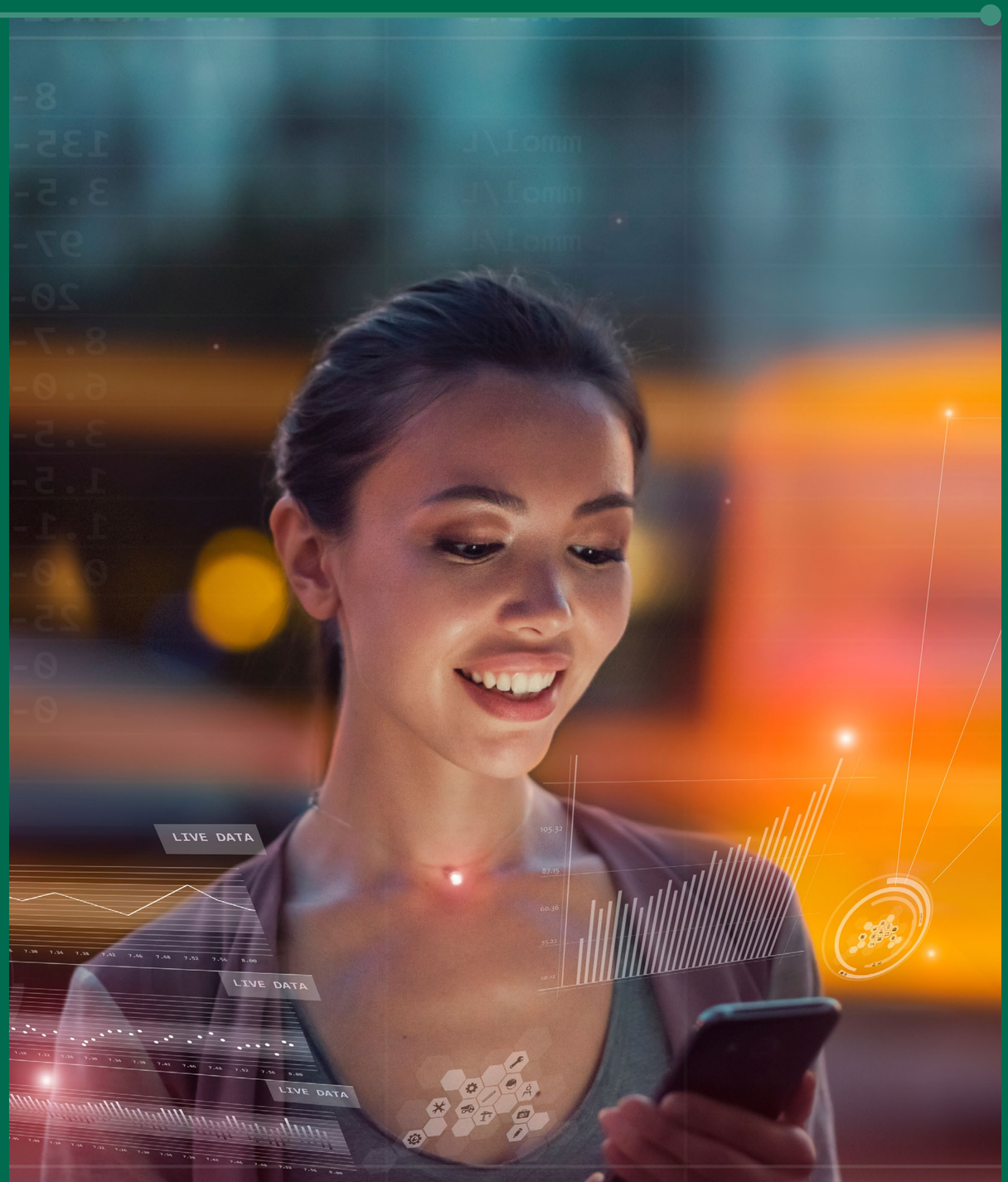
Credo Legal Framework: Insights and Research (13)

Key areas of interest (13)
Contract architecture (14)
Accession of new data providers (15)
Adding new shared data (16)
Refreshing shared data (17)
Accession of data users (18)
Licence conditions for new insights (19)
Review of existing data sharing licence structures (20)
Trust frameworks (20)
Open Energy (22)
iSHARE Trust Framework (22)
Automated smart contracts – Next Generation Initiative (24)
a. Comparison of assessed trust frameworks and considerations in selecting a potential format (25)
Future areas of discussion (27)

Development of The CReDo Distributed Architecture (28)

Distributed architecture on Azure – visualisation access (29)
Visualisation access process before the latest development (CReDo deployed on DAFNI) (29)
Visualisation access process after the latest development (CReDo deployed on DAFNI and MS Azure) (33)
Distributed architecture on Azure – evaluating scenarios (37)
The process of evaluating a scenario before the latest development (CReDo deployed on DAFNI) (37)
The process of evaluating a scenario after the latest development (CReDo deployed on DAFNI and MS Azure) (41)
CReDo distributed architecture – conclusions and future work (45)

Appendix 1. Interview Questions For Credo Partners And External Experts - Updating The Legal Framework (46)



EXECUTIVE SUMMARY

Background

The Climate Resilience Demonstrator (CReDo) is a climate change adaptation Digital Twin project that connects data across organisations and sectors (e.g. power, water, telecoms and others in the future).

CReDo's vision is based on the concept that cross-sector connected data can support decision-making in climate adaptation and improve the resilience of the "system of systems". This holistic view will help infrastructure owners make better-coordinated decisions at the lowest cost.

While CReDo has potential to realise benefits across the infrastructure network, the project's use of infrastructure asset data from multiple networks poses complex legal and technological requirements. This report outlines how the legal framework and technical architecture respond to these requirements.

Legal framework

The CReDo team has identified these areas of the current legal framework as issues for scaling, as they increase administrative load on CReDo partners:

- Contract architecture (two agreements vs. one agreement)
- The process for accession of new data providers
- The process for adding new shared data to the licence
- The process for refreshing data shared under the licence
- The process for accession of new data users (who are not also data providers)
- Licence conditions for new insights

Engagement with project partners and legal experts highlighted that some administrative load will be necessary in a CReDo-type context, as the process ensures data and asset security. The challenge is identifying the right level of administrative load for each partner.

The CReDo team recommends trust frameworks as a solution to these governance challenges. However, trust frameworks are still in development in the UK. Future discussions should focus on how CReDo can develop in tandem with emerging trust frameworks.

Technical architecture

The technical solution to data sharing challenges is provided by CReDo's distributed architecture, which aims to enable data sharing "at source" by connecting to asset owner systems. This will avoid long-term storage of all CReDo data on a central location, thus inherently reducing cybersecurity risks.

Currently, CReDo has demonstrated the concept of a distributed architecture by separately storing data from various asset owners on different servers (one per asset owner) and having a central node that coordinates the use of this data from the various locations. These servers are, however, still deployed on DAFNI (Data & Analytics Facility for National Infrastructure); this is designed to be representative of a distributed architecture, whilst ensuring security of the data during testing and development.

This year, the CReDo team has deployed servers containing synthetic asset owner data outside of DAFNI. This data setup more realistically represents the setup required to connect to asset owners' data at source. The new distributed architecture can demonstrate bilateral transfer of data to/from CReDo nodes in Azure into a central CReDo node in DAFNI, with a secure design that will help future development and deployment work.

Authorship

Main authors: Cara Navarro (Data Strategist – Connected Places Catapult), George Brownbridge (Senior Technical Architect – CMCL) and Stephane Fernandez Garcia (Data Strategist – Connected Places Catapult).

Reviewers: Chris Jones (Ecosystem Director), Elliot Christou (CReDo Lead)

Contributors: Alanna Gluck (Delivery and Engagement Manager – Connected Places Catapult), Jacob Langlands (General Counsel / Legal – Connected Places Catapult)

INTRODUCTION

The Climate Resilience Demonstrator (CReDo) is a climate change adaptation Digital Twin project that connects data across organisations and sectors (e.g. power, water, telecoms. and others in the future). CReDo's vision is based on the concept that cross-sector connected data can support decision-making in climate adaptation and improve the resilience of the "system of systems". This holistic view will help infrastructure owners make better-coordinated decisions at the lowest cost.

Connected Places Catapult (CPC) and the project partners – BT Group (BT), Anglian Water (AW), and UK Power Networks (UKPN) – have been developing CReDo as a demonstrator of the technology, with the aim of progressing CReDo to a minimum viable product (MVP) and, ultimately, a fully usable product. In its next phase, CReDo will enter a critical stage of product development, and as such, the development team has begun to address challenges around enabling the required data sharing at scale. These can be captured in two main categories:

- **Operational:** these include areas where CReDo will need to develop processes that allow it to operate as a product and integrate into "business as usual" processes in any organisations that hold a stake in it. A major challenge in this area is the legal framework to enable the sharing of data and insights, daily operation of CReDo, onboarding of new organisations and commercialisation.
- **Technological:** those that relate to the development of CReDo as a piece of technology that needs to enable different functions (e.g. asset and cascade modelling, insight generation, data/insight hosting and/or sharing etc.)

These challenges stem from CReDo's security needs: CReDo's use of infrastructure asset data from multiple networks poses complex legal and technological requirements.

CReDo is currently operating within the legal context set by two documents, a Data Exploration License (DEL) and a Participation Agreement (PA). The former governs how key project partners share data with CReDo and how it is hosted; the latter allows CReDo to share this data with third parties for the sole purposes of CReDo's development. These two documents have historically enabled CReDo to develop as a demonstrator and take the first steps to reaching an MVP. However, the current licence is not fit for scaling the product to more users. A new licence that addresses key issues regarding data sharing, usage, commercialisation and other challenges is needed to ensure CReDo can scale in subsequent development cycles and operate as a product once it is deployed.

The technical solution to data sharing challenges is provided by CReDo's distributed architecture, which aims to enable data sharing "at source" by connecting to asset owner systems. This will avoid long-term storage of all CReDo data on a central location, thus inherently reducing cybersecurity risks. Currently, CReDo has demonstrated the concept of a distributed architecture by separately storing data from various asset owners on different servers (one per asset owner) and having a central node that coordinates the use of this data from the various locations. These servers are, however, still deployed on DAFNI (Data & Analytics Facility for National Infrastructure); this is designed to be representative of a distributed architecture, whilst ensuring security of the data during testing and development.

Purpose and scope of this document

This report documents how CReDo's legal framework and distributed architecture have developed in the most recent phase of the project. Key data sharing flows for CReDo are summarised in the first section of this report, which inform subsequent sections on CReDo's legal framework and technical architecture.

The second section of this report outlines CReDo's legal framework and key considerations for its evolution. The contents of this section are based on engagement with key contacts within asset-owning organisations, including legal, cybersecurity, and operational teams. External legal and cybersecurity experts were also consulted. In addition, the CReDo team completed a short literature review of existing data sharing initiatives to identify potential learning opportunities.

The third and final section of this report documents development of CReDo's distributed architecture. This project aimed to use synthetic data from various infrastructure owners to advance the maturity, readiness level and deployment status of CReDo's distributed architecture. To do so, the CReDo team has deployed servers containing **synthetic** asset owner data outside of DAFNI. The purpose of this is to realistically represent the setup required to connect to asset owners' data at source. Since many asset owners' systems are supported by Microsoft-owned architecture and platforms, the external CReDo nodes (containing the synthetic data) were deployed on Microsoft Azure (a cloud computing platform).

Successful completion of this work has allowed CReDo to make progress in the technical development of the distributed architecture technology by creating a synthetic data setup that is closer to the one that would be used when connecting to asset owner systems at source. This newer version of the distributed architecture can demonstrate bilateral transfer of data to/from CReDo nodes in Azure into a central CReDo node in DAFNI, with a secure design that will help future development and deployment work.

In sharing CReDo's legal and technical data sharing frameworks, the project team aims to contribute to emerging common guidance in the data sharing space. Data sharing initiatives are developing across the ecosystem, such as the Virtual Energy System, the National Underground Asset Register, Stream, and the National Digital Twin Programme. Sharing learnings across as many initiatives as possible is essential to our interoperability.

OVERVIEW OF DATA SHARING IN CREDO

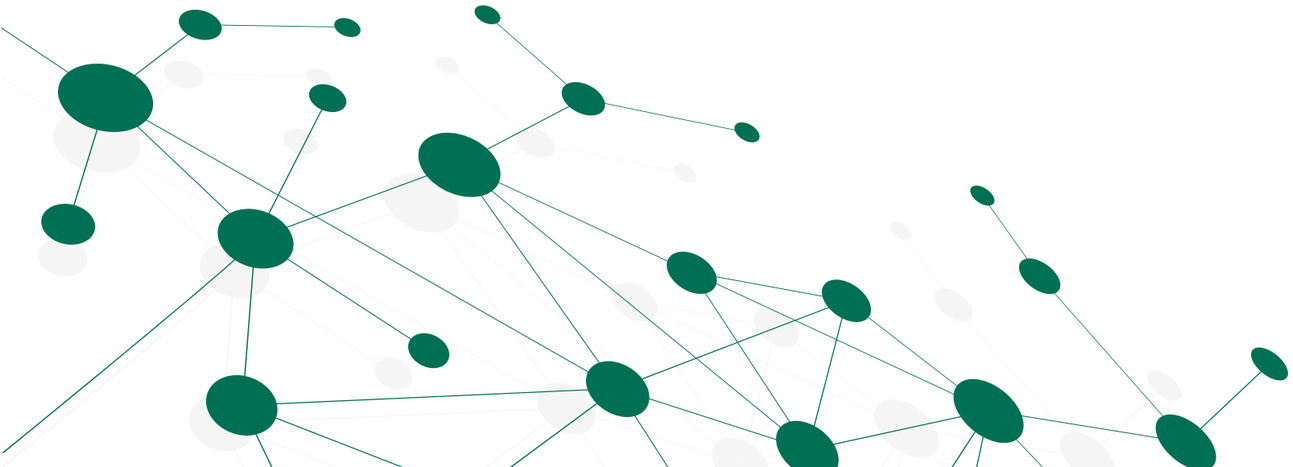
To understand and identify CReDo's future legal and technical enablers and how those might apply to other data initiatives, it is important to understand the current practical aspects of data sharing in CReDo.

The specific data sharing context of CReDo is important for selecting a suitable data sharing framework and architecture. The objectives and data flows are outlined in the following subsections.

Objectives of sharing data

The overall objective of sharing data in CReDo is to understand climate-related interdependency risks across infrastructure networks, with the aim of enabling asset owners to coordinate investment in asset resilience interventions. Secondary use cases under this umbrella are yet to be fully explored, but include 1) supporting emergency response planning at the systems level, and 2) cross-sector climate resilience reporting to network operators, infrastructure regulators, government and governmental agencies, and other decision makers.

The primary benefit is realised across the infrastructure system as a whole, with overall savings in climate resilience investment. However, infrastructure networks also see benefits independently. Each network sees fewer climate event-related outages, higher customer satisfaction, and fines avoided from regulators.



Sources and users

Figure 1 and Table 1 (below and in the next page respectively) outline data flows between sources and users within CReDo.

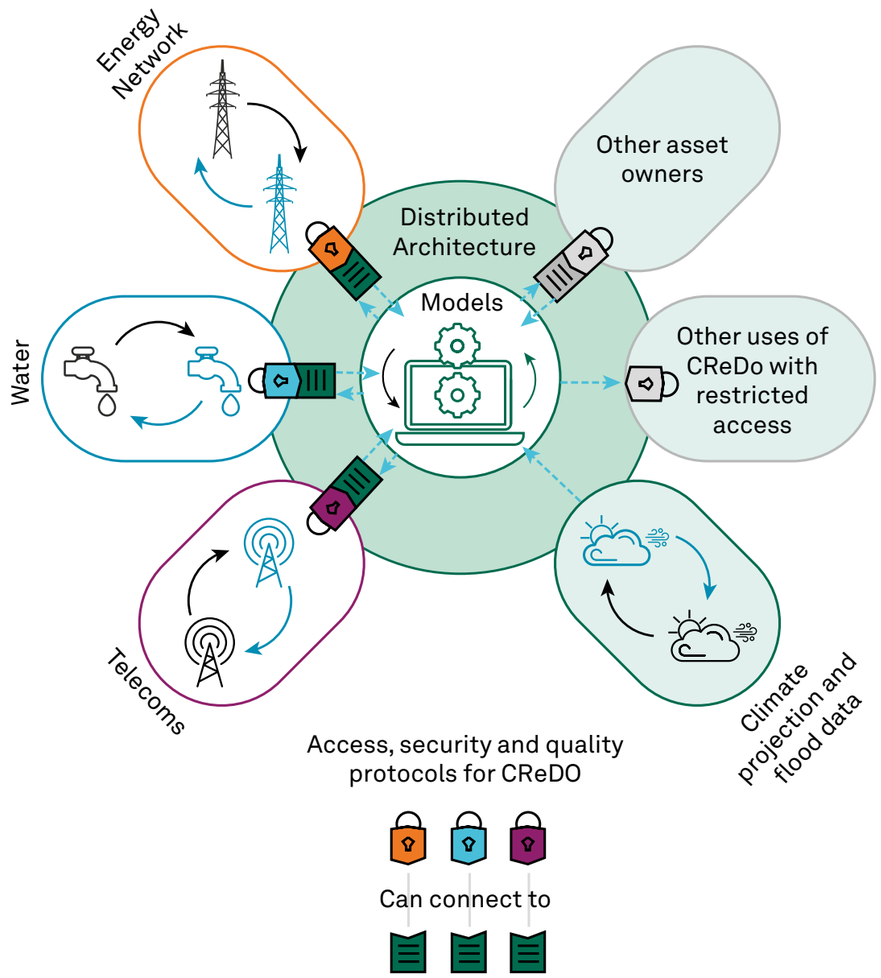


Figure 1. Data flows between sources and users represented in the conceptual architecture diagram of CReDo.

Value chain segment	Future export potential	Investment priority	Summary		
1	Asset information.	Asset owners.	CReDo models (with data currently visible to CReDo development team: CPC , Computational Modelling Cambridge Ltd. (CMCL), Science and Technology Facilities Council (STFC))	Modelling asset and cascading failure probabilities in climate scenarios, maintaining model.	Established with current CReDo partners.
2	Insights regarding asset and cascading failure risks to individual asset owner's asset base, derived from modelling all network assets.	CReDo development team.	Asset owners.	Understanding of risks and planning for interdependency-related climate risks to their own assets. Strategic planning to improve climate resilience.	
3	Insights regarding asset and cascading failure risks to selected asset bases, derived from modelling on all network assets.	Asset owner (underlying asset information) and CReDo development team (insights).	Regulators, insurers, government and potential other users.	Regulators and government to understand resilience of different networks, insurers to inform asset underwriting. Other commercial use cases to be determined.	Needs cross-relationship development and established legal framework.
4	Asset information, insights regarding failure risks to specific assets in one or more networks.	Asset owner 1.	Asset owner 2.	Coordinating asset resilience investment.	CReDo current partners are keen to collaborate, but processes and legal framework governing flow of information needs development.

Table 1. Details of data flows between source and users in CReDo.

It is important to consider than the data flows do not currently contain personal data, and they are not expected to. However, given that the insights expose critical infrastructure vulnerabilities, there is a high associated security risk. It is also important to note that future developments of CReDo will investigate whether it is possible to reduce the risk by extending the distributed architecture to distribute the intra-network modelling as well as the hosting of data (i.e. to reduce the data that is aggregated within the models via step (1) in Figure 1).

Permissions

Permissions should vary between CReDo partners depending on what is the expected use of data by each. Key considerations for the different users are captured below:

- The development team requires full access to source data for the development and testing of the failure model and CReDo technology (e.g. ontology development)
- By default, asset owners should only have access to failure model outputs for their own assets, with basic information about the status of connections to other assets. This should provide asset owners with an enriched understanding of climate related cross-sector risk (that would not be possible without CReDo), but without exposing proprietary data.
- Where two asset owners have agreed to share data about specific assets, they should have permission to see both basic information and failure model outputs about those assets. This would provide asset owners with an enhanced understanding of details around cross-sector dependencies which could prove useful in resilience planning or other collaborative initiatives these asset owners may be working on.
- Users that do not own assets, such as regulators and insurers, could see failure probabilities for specific asset bases under different climate scenarios. This level of access does not include the source data that informs these probabilities. This is based on our preliminary understanding of these types of users' wishes. Further engagement with asset owners and non-asset owning users is needed to further develop use cases and design a fit-for-purpose solution that addresses both user needs and data sensitivities.

Terms of use

In the current setup, asset owners and the development team set the terms of use for source data. These terms are defined by what is needed for developing failure models and CReDo as a technology platform; this considers both the models themselves, the security requirements of the models' architecture and other general aspects of the CReDo development.

There are currently no terms of use for the insights derived from modelling on the source data, apart from a non-commercialisation clause. It is expected that these terms will be defined in the short term through existing channels of engagement with current CReDo partners and that the terms will apply to any new partners going forward. However, a governance process needs to be developed should any partners object to the terms of use.

CREDO LEGAL FRAMEWORK: INSIGHTS AND RESEARCH

Key areas of interest

The CReDo team has identified these areas of the current legal framework as barriers to scaling from demonstrator to MVP, as they increase administrative load on CReDo partners:

- Contract architecture
- The process for accession of new data providers
- The process for adding new shared data to the licence
- The process for refreshing data shared under the licence
- The process for accession of new data users (who are not also data providers)
- Licence conditions for new insights

In this section, the following points are explored for each area of interest:

- The current process that is present in the licence (the Data Exploration Licence, i.e. DEL)
- The CReDo team's proposed aim for a new process
- Consolidated insights drawn from interviews with CReDo partners, internal CPC stakeholders, and external legal and cybersecurity experts. (Please see Appendix 1 for the questions covered in these interviews.)

Contract architecture

Process in the current licence structure	Data sharing in CReDo is contractually governed by a Data Exploration Licence (DEL) and a Participation Agreement (PA). The DEL is a multilateral agreement between Connected Places Catapult (CPC), Science and Technology Facilities Council (STFC), CMCL, Anglian Water Services (AW), BT Group Plc, and UK Power Networks (UKPN). Each PA is a bilateral agreement between CPC and the relevant organisation. The DEL sets out the specific datasets AW, BT, and UKPN (referred to as licensors) are sharing for the development of CReDo and outlines the terms on which the CReDo development team (referred to as licensees) can access this data. The data shared under the DEL is referred to as Exploration Data. The PA establishes a legal obligation for AW, BT, and UKPN (referred to as data service providers) to contribute to CReDo's development and enable them to view data in CReDo for specific defined purposes (mainly for testing in the context of the development of the CReDo technology). For third parties involved in development, the PA acts as a DEL sub-licence, giving them access to data shared by AW, BT, and UKPN under the DEL.
Aim for the new licence structure	The CReDo team's aim is for asset owners to sign only the DEL, while third parties involved in development would sign only the PA. The aim is for new asset owners to sign only one document when joining CReDo, streamlining the onboarding process.
Engagement insights	• Internal stakeholders noted that the DEL and PA are complementary: the DEL sets up general terms of access, while the PA sets out liabilities and IP provisions more specifically for individual asset owners. • In the past, allocated work to support CReDo has been altered for certain asset owners, which would not be possible if the DEL and PA were merged. The DEL is multilateral, meaning that signatories would all have to agree to the same terms.

Accession of new data providers

Process in the current licence structure	Any party to the DEL can request the addition of a new asset owner as a DEL licensor. They must notify other parties to the DEL of this request in writing. Each party needs to respond to this request in writing within 21 days, with either an acceptance or a rejection. If a party rejects the addition of the new licensor, they must provide a rationale. If a party does not respond within 21 days, it is assumed that they accept the request. An acceptance of the request must be unanimous to go through. If one party to the DEL rejects the request, all parties need to discuss how to address their concerns. Once all parties accept the request, the licensors, licensees, and the new asset owner must sign an accession agreement, which names the new asset owner as a DEL licensor. The list of shared datasets is also updated to reflect data from the new licensor.
Aim for the new licence structure	The CReDo team's aim is for new participants to sign a data sharing licence without all other participants having to sign the licence again. That said, there should be some means by which participants can be accountable to each other if anything goes wrong.

Engagement insights	• Signing the same document creates a "contractual link" between existing and new participants, which provides more protection. In principle, this is the preferred way forward for CReDo partners even if it potentially creates additional administrative burden.
	  ○ If not everybody signs the same contract, then the creation of a mechanism where parties can still be liable to each other is needed.
	• An independent legal expert suggested that data providers and users form a community; then a licence condition mandating that community members publish their details in an online register can be created; this allows for constant transparency on/to data providers and users.
	  ○ Being a community, setting a code of conduct or good behaviour would be useful. This is about how the CReDo community operates – when anticipating getting new people in, setting behaviour standards is critical.
	• Potential data users are the main group of concern (see "Accession of new users"), other data providers are seen as inherently less problematic.

Adding new shared data

Process in the current licence structure	A member of the project team updates the list of shared datasets and circulates it by email to representatives from all current asset owners. Asset owners accept or reject the change via an emailed response.
Aim for the new licence structure	If a participant needs to share more data for the purpose of the data sharing project, this can be added to their own licence, without other participants needing to approve the additional shared data. New shared data includes new asset bases, new asset details, new geographies etc.
Engagement insights	• Interviewees strongly feel that additional data sharing must be managed contractually. • In a CReDo-type context, it is advisable that a security advisor reviews any risks from sharing additional data. • The general stance of some interviewees' organisations is open; there are no general red lines around sharing any dataset "in isolation". The red lines are highly dependent on the use case; it's about justification and value. • The CReDo use case of modelling cascading risk is well-developed and clearly linked to the shared data in the existing data licence. However, the uses of insights related to cascading risk – the follow-on use cases – are less well-understood, especially for non-asset-owning organisation types, and need to be defined more clearly before their inclusion in any sharing agreements. • If access to data and insights varies per use case, then it might be necessary to get asset owners' approval every time a new user type joins. This creates additional administrative burden. • Interviewees felt that there should be a difference between input data that is shared and only going to be used in the background of CReDo (e.g. only for modelling/creating insights and is not visible to users) and the data that can actually be visualised by users. ○ The CReDo team believes that a classification, where data is tagged as "viewable" (by users) vs "functional" (only used in the background) could be captured in the legal documents (i.e. a data schedule); this might prove useful in managing concerns around data sharing. The companies owning the data would be best placed to make those decisions. This would also allow for separate legal classifications of data for each organisation if needed. Different partners might have different views on the classification of same or very similar pieces of data.

Refreshing shared data

Process in the current licence structure	The current licence is designed for one-off data sharing. There are no provisions for refreshing the data when it becomes out of date.
Aim for the new licence structure	The CReDo team's aim is for a clause that encourages data providers to refresh their data at an appropriate frequency.
Engagement insights	• Interviewees were not concerned about contract clauses that request a refresh of data. Refreshing of data sets already shared could be covered in the initial contract, but data providers will need to be able to reject or approve new data. • It will be key to define what is considered an appropriate frequency. A solution that balances effort from asset owners and the CReDo team with adequate accuracy of outputs in CReDo is needed. • To avoid negative impacts on decision-making, it is important to keep CReDo's outputs as updated as possible. However, it may be difficult for asset owners to ensure enough resource is available for refreshing data. In the short term, therefore, a data licence should encourage rather than mandate a refresh of data.

Accession of data users

Process in the current licence structure	Currently, there is no process allowing parties to register for access to insights without also contributing data.
Aim for the new licence structure	The CReDo team aims for non-asset owning organisations to sign a separate contract outlining terms and conditions for accessing insights.

Engagement insights	• Interviewees have expressed concerns around verifying the identities of those who have access to insights. They would like to have a process in place that prevents nefarious individuals from claiming affiliation with legitimate organisations.
	  ○ Part of the process could require ensuring organisations are registered on the Companies House, but this by itself does not solve the problem of verifying individual access as, for example, it is very easy to set up a business on the Companies House.
	• In-house efforts to verify individuals' identities should be kept at a minimum. If a substantial number of users regularly join CReDo, e.g. daily effort, it is not reasonable to assume that verification can be undertaken by data licensors/partners.
	• A fit-for-purpose risk assessment about who is seeing the data will be needed.
	• The ISHARE Trust Framework has addressed this (see below bullet points for an outline) and, as such, might prove to be a useful resource when addressing this discussion area in the future.
	  ○ A central authority supervises identity provision.
	  ○ Identity providers are certified to levels of assurance based on how much information they collect on users. This sort of setup takes the burden of verifying user identities away from the partners / data licensors.
	  ○ With regards to verifying the identities of users, entities are assessed on three dimensions: how the real-world identities of users are verified upon application (do they provide official identity documents?), how users' electronic identities are connected to their real-world identities and how the electronic identity is authenticated upon accessing data.
	• It is important that the security barrier for user entry is adequate, but not unreasonable enough to discourage potential users from joining. There are potential learnings to consider in this area from other sectors, such as the financial sector. For example, bank account opening processes might prove to be a source of useful information on how to solve this challenge.
	• Identity verification is seen as a potential root cause that might have a negative impact in other challenges that are essentially based on building trust. Having an adequate process for this might streamline solutions for other challenges.

Licence conditions for new insights

Process in the current licence structure	As the current licence is focused on bringing asset owners' data into CReDo, there are no licence conditions around the insights that CReDo generates.
Aim for the new licence structure	The CReDo team would like a framework that addresses the following questions in relation to the insights generated from asset owners' data: • Who can see insights. • Using insights. • Adapting insights. • Combining insights with other data sources. • Redistributing insights. • Security provisions for new insights.
Engagement insights	Solving these challenges will most likely require: • Secure mechanisms to take insights/data off the platform (e.g. download) both from a technological and legal perspective. • Different licence terms that enable/regulate the activities for data sharing organisations and data/insight viewing or user organisations. Note that there are cases where a single organisation can both share and use/view data or insights. The roles of these different types of organisations are significantly different; there could also be organisation subtypes, in particular for data viewers/users. These should be appropriately mapped and defined in future versions of the legal documents. • This should cover CReDo sharing insights with the assets' respective owners. Also, asset owners sharing data and insights about specific assets with each other.

Review of existing data sharing licence structures

The challenges that need to be addressed in CReDo's data sharing licence are not unique to CReDo – these are shared by data spaces, connected digital twins, and industries looking to streamline data sharing. As such, a variety of solutions have emerged to address these requirements. This section reviews learnings from these solutions.

Trust frameworks

The issues of the original CReDo licence – a lack of scalability to new users, lengthy negotiations – are in part reflective of a potential lack of trust between participants. Trust frameworks provide tools that establish this trust. They define common standards that members must agree to, certify that members adhere to these standards, and build authentication APIs to enforce these standards in data sharing transactions. This saves effort for data sources, who do not need to validate the trustworthiness of every organisation looking to use their data. Instead, this work is outsourced to the trust framework organisation.

Trust frameworks also enable scalability of a data sharing initiative. Rather than signing data sharing agreements with each other, all members sign an accession agreement with the trust framework organisation, which governs all data transactions between members. Crucially, this accession agreement covers the terms of data sharing, not access to specific datasets. Requirements for accessing datasets are set in separate licence conditions that members are legally bound to respect.

These licence conditions are machine-readable and enforced in APIs for data transfer. The data transfer flow shown below is for a specific trust framework, but others follow broadly the same sequence: data users' attributes are encoded in an access token from a central intermediary, which data owners, then verify through the same intermediary.

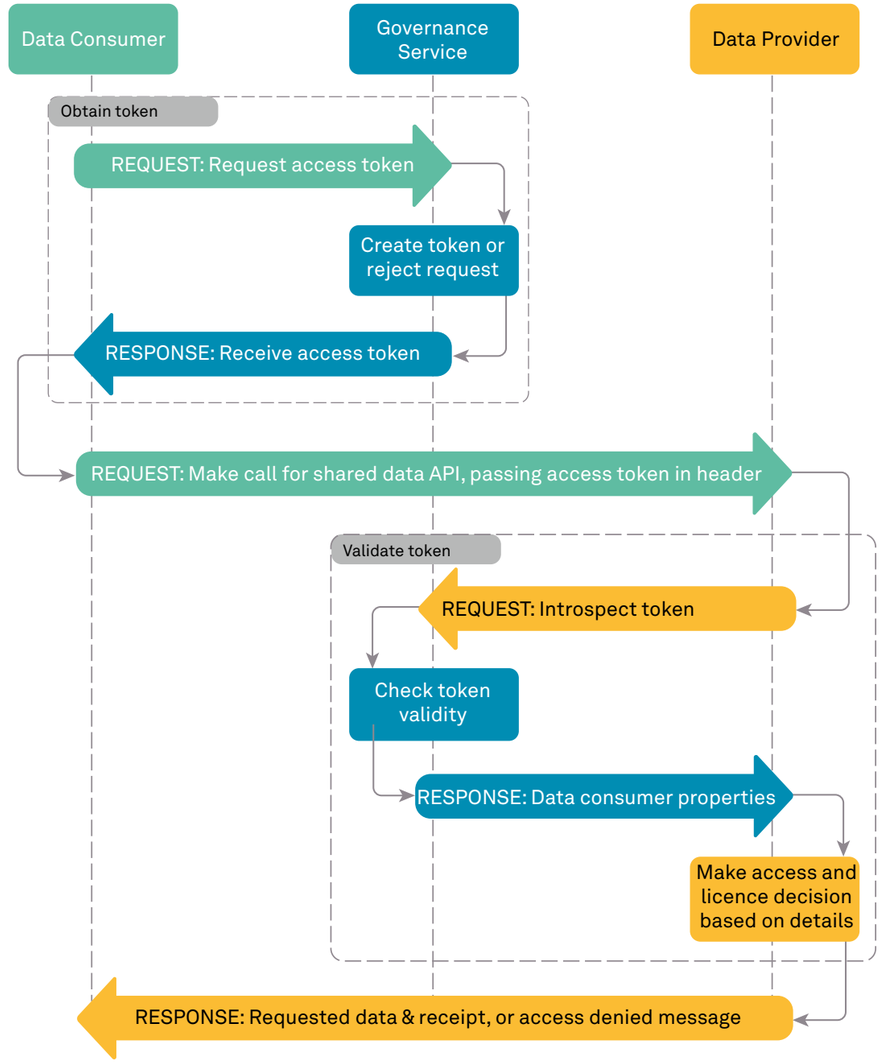


Figure 2. Example of a data transfer flow for a specific trust framework (Source: Open Energy)

Trust frameworks have been developed for sector specific (e.g. Open Energy, Open Banking) and for sector-agnostic data sharing initiatives (e.g. iSHARE). Open Energy and iSHARE are reviewed here. While they share key similarities in structure, as outlined above, they focus on different barriers to scaling data sharing, and pose different onboarding requirements for stakeholders.

Open Energy

Open Energy is a trust framework for the UK energy sector developed by Icebreaker One. Open Energy enables scalable data sharing by standardising data licensing and data transfer security.

Under this framework, the central intermediary is the Open Energy Governance Service (OEGS). Key steps in applying this framework to data sharing initiatives are outlined below.

- **Development team requirements**
  - Legal teams review and sign the membership contract.
  - Organisation details are sent to Open Energy.
  - Private keys and cryptographic transport certificate are generated via OEGS, which will be used in all data transactions.
  - Integration of Open Energy financial-grade API (FAPI) specifications into distributed (or corresponding) architecture.
- **Asset owner requirements**
  - Legal teams review and sign membership contract.
  - Organisation details are sent to Open Energy.
  - Creation of metadata files for datasets and publication of these on public web server for automatic OEGS indexing.
  - Development of Open Energy-compliant FAPIs for accessing datasets.
  - Assignation of datasets to standardised sensitivity class and specification of data access rules (by specified group, organisation type, or whether organisation is a paying customer of the data provider) and obligations from a standardised list[1].
  - Obtaining internal legal sign-off for access rules.
- **Non-asset owning organisations**
  - Legal teams review and sign membership contract.
  - Organisation details sent to Open Energy, who then create entry in the participant registry and set up the organisation's OEGS account
  - Private keys and cryptographic transport certificate are generated via OEGS, which will be used in all data transactions.

1 Access Control and Capability Grant Language (Icebreaker One, 2021)

iSHARE Trust Framework

iSHARE is a European trust framework developed for data spaces more broadly. As opposed to Open Energy's focus on data licensing and security, iSHARE standardises identity, authorisation, and authentication procedures.

The central intermediary is the iSHARE Foundation. Some responsibilities, such as admitting new participants to a data sharing scheme, are delegated to Satellites. CReDo (and similar initiatives) would be a Satellite under this framework. Key steps in applying this framework to data sharing initiatives are outlined below.

- **Development teams requirements**
  - Development and testing APIs required for access to iSHARE network:
    - Generating OAuth access token for accessing secured services.
    - Enabling other users to search distributed ledger-based participant registry.
    - Receiving iSHARE's list of trusted eIDAS certificate authorities.
    - Validating parties' authorisation levels for accessing data and services.
    - Publishing service capabilities.
  - Assessing identity, authentication, and authorisation services against the iSHARE framework.[2]
  - Providing an eIDAS certificate, nationally recognised company identification number, and signed Accession Agreement for Certified Parties to iSHARE Foundation.
  - Onboarding new CReDo/programme users onto the iSHARE network as they join.
- **User requirements**
  - Developing and testing APIs required for access to iSHARE network:
    - Generating OAuth access token for accessing secured services.
    - Providing data and services with the appropriate authentication and authorisation checks.
  - Providing an eIDAS certificate, nationally recognised company identification number, and signed Accession Agreement for Adhering Parties to iSHARE Foundation.
  - Assigning licence code to data and services from standard list[3], or otherwise negotiate if none suit the requirements.

2 Admission (iSHARE Trust Framework, 2024)

3 Licenses (iSHARE Trust Framework, 2024)

Automated smart contracts – Next Generation Initiative

As part of the European Commission's Next Generation Initiative (NGI), a team of researchers has developed a contract service that automates the generation of data sharing agreements for data sharing networks.[4] These agreements are implemented as smart contracts on the Ethereum blockchain: contracts that automatically execute code when parties digitally sign the agreement. Unlike trust frameworks, which digitise data licensing but maintain a traditional contract agreement, the automated data contract service introduces an entirely digital contract format.

The service consists of three components:

- A registry of participants, services, and data policies.
- An agreement management service for generating and verifying data sharing agreements.
- An agreement signing service.

Source code for the service is available on Github.[5] Signing an agreement using smart contracts would include the following steps:

1. Data owners set human- and machine-readable data policies regarding datasets, which cover conditions, obligations, restrictions, prices, certifications, data security, rights, data protection, and liability. These data policies are made publicly available in a registry.
2. Data users publish human- and machine-readable metadata about their security and compliance provisions for data access. These provisions are published in the same registry as data policies.
3. If data users agree to the policies set by data owners, a data sharing agreement is automatically generated, converting the relevant policies into human- and machine-readable clauses based on a standard template.
4. Data sharing agreements are signed cryptographically on the Ethereum blockchain.
5. Information on signed agreements is stored as a non-fungible token (NFT) on the blockchain, the presence of which is verified by a data intermediary before data transfers.
6. Both data owners and data users need to publish instructions on how to validate a data sharing agreement. These instructions are used by data intermediaries.

4 Rulebook Architecture Design Document (The Rulebook Consortium, 2021)

5 Human-Centric Rulebook: Data Contract Service (2022)

a. Comparison of assessed trust frameworks and considerations in selecting a potential format

Table 2 summarises key points of comparison for contract structures.

Table 2. Comparison of key points between Open Energy, iSHARE and Next Generation Initiative Trust frameworks.

Solution	Agreement format	Participant registry	Legal negotiation	Security standards	Technical integration
Icebreaker One Open Energy Trust Framework	Paper agreement for overall scheme terms with digital, automatically enforced licences.	Outsourced to OEGS.	Negotiate from standard capabilities and obligations.	FAPI for authentication, authorisation, and data transfer.	Integrate FAPI specifications and link to OEGS into distributed architecture APIs.
iSHARE Trust Framework	Paper agreement for overall scheme terms and liabilities with digital, automatically enforced licences.	Outsourced to iSHARE.	Negotiate from a list of standard licence codes.	OAuth 2.0 specifications, in addition to iSHARE-specific standards.	Integrate OAuth specifications and link to iSHARE into distributed architecture APIs.
NGI Data contract service	Smart contract for data sharing agreement between two parties.	Built by members of the data sharing initiative.	Negotiate from scratch, including acceptance of smart contracts.	Not defined by framework.	Integrate provided source code into the architecture of CReDo / data sharing initiative.

The project team believes a trust framework would provide the ideal solution for scaling CReDo's data sharing agreement or for other similarly complex data sharing programmes. Upon joining a trust framework, individual users would sign onto a pre-defined agreement. Additionally, given the multiple types of data flows and access permissions in CReDo, there will likely be multiple licences needed. The standard licence terms provided within trust frameworks would streamline the negotiation of those licences. Security of a participant registry would be maintained by the framework organisation rather than by CReDo.

Alternatively, parts of each trust framework could be adopted without fully joining a trust framework. There could be an accession agreement with CReDo or data sharing initiatives themselves, specifying broad terms of use for data and insights. The standardised licence conditions developed by Open Energy could be used as a starting point for negotiation of data licences, for both input data and the insights developed within the specific programme. The mechanisms for automatically checking data licences could also be integrated into the corresponding technology architecture.

Another option would be for CReDo to leverage future sector-specific trust frameworks, such as those being explored by the Virtual Energy System and Stream.[6] CReDo could sign up to these trust frameworks as a user, and organisations would then join CReDo under their sector's trust framework. However, the trust frameworks themselves would need to be compatible. CReDo would also still need agreements for organisations in sectors without trust frameworks.

Before committing to an option, further questions need exploring:

- Are liabilities and dispute resolution processes independent of data licence conditions? Under a trust framework, they are, but this assumption needs to be tested with CReDo organisations' legal teams.
- How can machine-readable data licences be incorporated into the CReDo's or other projects' interfaces?

6 Cross-sector UK Data Sharing Infrastructure (2024)

Future areas of discussion

This section lists areas that will be important to discuss in detail in the future; it also captures additional comments shared by interviewees. It is important to take into account that this first round of engagement did not aim to discuss in detail the topics listed in this section. The project team considered that in order to do so, a proposed CReDo commercialisation plan would need to be shared with partners first; the plan is still being developed and was not available before this first round of engagement with partners took place.

Future discussions should include, but are not necessarily limited to:

- **Liabilities** (after the commercial plan dissemination stage)
  - A point to consider is that there is still a possibility that the data can be misused and, therefore, it is important to consider what is the liability to data users is (see more detail in the next collection of bullet points).
  - Also, interviewees feel that the wider the user base is, the less the per-user liability should be.
  - The National Underground Asset Register, NUAR; could provide valuable learnings for the project in this area.
  - o
    - is that all parties who give data have to have low liability to anybody else.
  - **Who could a data provider be liable to?**
    - Data users (including other data providers)
    - Asset owners whose assets are connected to CReDo users' assets (e.g. Gatwick Airport)
    - Liability of data providers to data users in the event of decisions taken on inaccurate data.
    - Liability of data sharing initiative team (e.g. CReDo) to data providers, in the event of a breach, data misuse, or some sort of reputational damage arising from sharing of insights.
    - Liability of data providers to each other, in the event of a breach or data misuse.
    - Liability / protection of "CReDo" if there are inaccuracies in the underlying data, which then leads to inaccurate model outputs.
  - Partner organisations believe **business as usual approaches should not change the fact there is no warranty on the data.**
  - **Commercialisation rights with data** and insights will need to be discussed in detail after a commercial plan is shared with partners. It will be important to distinguish and consider commercialisation of services and insights from modelled outputs from the data, compared to commercialisation of the data itself.

DEVELOPMENT OF THE CREDO DISTRIBUTED ARCHITECTURE

CReDo's distributed architecture is a key operational piece of technology that ultimately aims to enable data sharing "at source" by connecting CReDo to asset owner systems. This will avoid the future need to gather and store all CReDo data in a central location for the long-term, thus inherently reducing cybersecurity risks.

CReDo is designed to be distributed for both security and extensibility reasons. In previous CReDo projects this has meant storing and, where feasible, processing asset operator data on separate virtual machines (VMs) running on DAFNI servers. This separation limits the amount of data that could be accessed in the event of a problem with one of the servers.

In this latest piece of work, CReDo has been updated to support distributing the VMs across multiple enterprise systems. The main benefits of this are the increased security due to the greater decentralisation of the data and its processing; and the simplification of the asset operator onboarding process by removing the need for their data to be copied outside their organisational boundary.

As part of this development, CReDo now performs user authorisation checks at the boundary of each enterprise. An initial solution to this problem has been implemented.

The different components that are relevant to the current discussion can be described as follows:

- **Reverse proxy** – a gateway performing authorisation and routing tasks.
- **Authentication server** – a credentials store that implements OAuth2 functionality to perform authentication and enable authorisation. As part of the authentication process, it also handles a multi-factor authentication process.
- **CReDo web interface** – the user facing parts of CReDo, including the landing page that initially guides the user and initialises the authentication, the geospatial visualisation and data dashboards that display the results of CReDo.
- **Asset operator stack** – a CReDo node that ingests, stores, processes and, as appropriate, serves asset operator data.
- **Central stack** – a CReDo node that co-ordinates between the asset operator stacks and runs the models.

This section describes the latest progress in developing and demonstrating CReDo's distributed architecture. Two primary workflows, (i) visualising results and (ii) running scenarios, are described comparing how it worked before and after this latest development. The section ends with a summary of potential future work in this area.

For a more detailed understanding of the technical implementation of CReDo, please refer to:

1. **CReDo Technical Report 1:** Building a cross sector Digital Twin. This document provides an understanding of key technical elements of CReDo during the early development phases.[7]
2. **CReDo Phase 2:** Technical Report – Distributed architecture. This report describes aspects of the initial development efforts in relation to moving CReDo to a distributed architecture.[8]

Please be aware that these reports are intended to provide historical context and detail; CReDo is being continuously developed and it is possible that current aspects of the technology are different from those described in this report. If you are interested to find out more, please contact the CReDo team at credo@cp.catapult.org.uk.

Distributed architecture on Azure – visualisation access

One of the key aspects of CReDo is that users can visualise the impacts (failures, etc.) of different extreme weather scenarios on the overall connected network. What a user can see in the visualisation is determined by the permissions they have been assigned based on the licence agreements that are in place between the different participants of CReDo.

Visualisation access process before the latest development (CReDo deployed on DAFNI)

The main steps in the process are outlined below and are visually represented in architecture and flow diagrams in Figures 3 and 4 respectively.

1. A user goes to the CReDo landing page.
2. They login via the authorisation server and an access token is returned.
3. They go to one of the protected visualisation pages that they have permission to access, and data requests are sent to the stacks.
4. The reverse proxy requests information about the user from the authorisation server.
5. The reverse proxy authorises the data requests accordingly.
6. The data requests are forwarded to the relevant stacks and the data is returned and shown in the visualisation.

7 CReDo Technical Report 1: Building a Cross-Sector Digital Twin - Digital Twin Hub
8 CReDo phase 2: Technical Report - Distributed Architecture - Digital Twin Hub

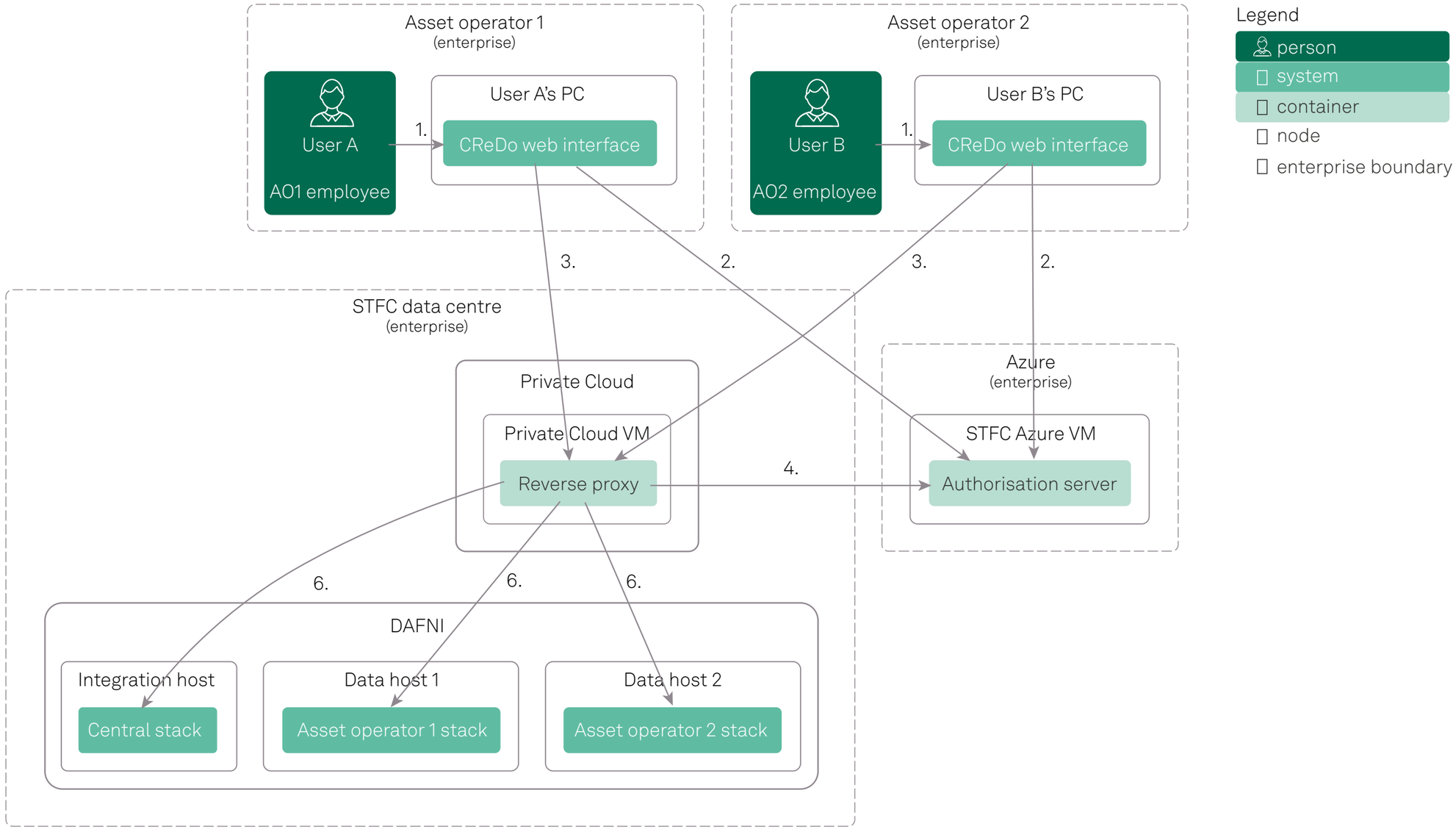


Figure 3. Architecture diagram showing the connectivity of components that are engaged when a user accessed the CReDo web interface before the changes made in the latest development cycle.

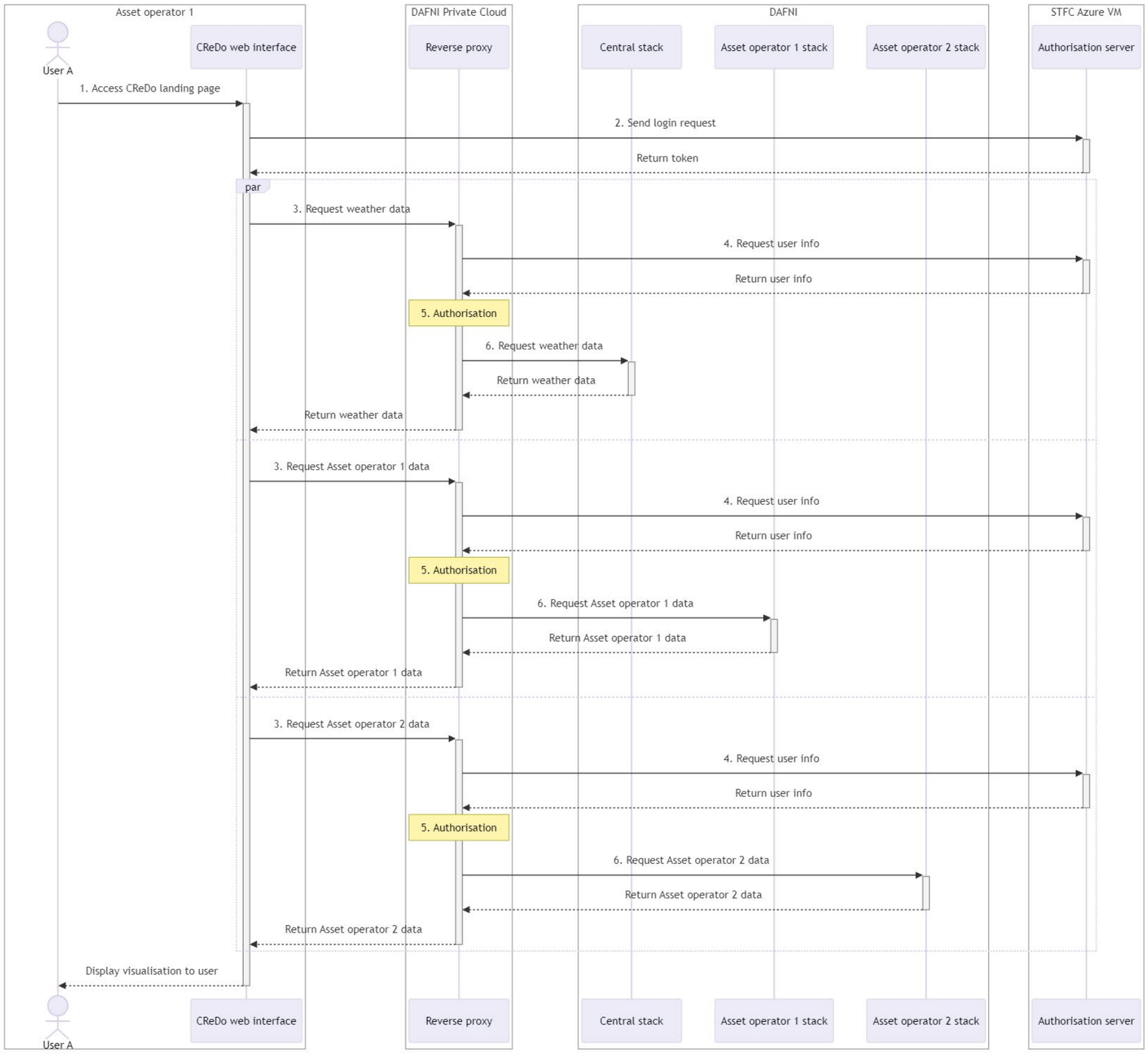


Figure 4. Flow diagram showing the communication between components that are engaged when a user accesses the CReDo web interface before the changes made in the latest development cycle.

Visualisation access process after the latest development (CReDo deployed on DAFNI and MS Azure)

The main change in this new setup is that the "Asset operator 2 stack" has been moved from the DAFNI infrastructure to a virtual machine (VM) running in a cloud platform (in this case Microsoft Azure). To ensure that this stack is still protected, authentication and authorisation also occurs in a reverse proxy running on the "Asset operator 2 Azure VM".

Similarly to the previous subsection, the main steps in the process are outlined below, and are visually represented in architecture and flow diagrams in Figures 5 and 6.

1. A user goes to the CReDo landing page.
2. They login via the authorisation server and an access token is returned.
3. They go to one of the protected visualisation pages that they have permission to access, and data requests are sent to the stacks.
4. The reverse proxy requests information about the user from the authorisation server.
5. The reverse proxy authorises the data requests accordingly.
6. The data requests are forwarded to the relevant stacks and the data is returned and shown in the visualisation.

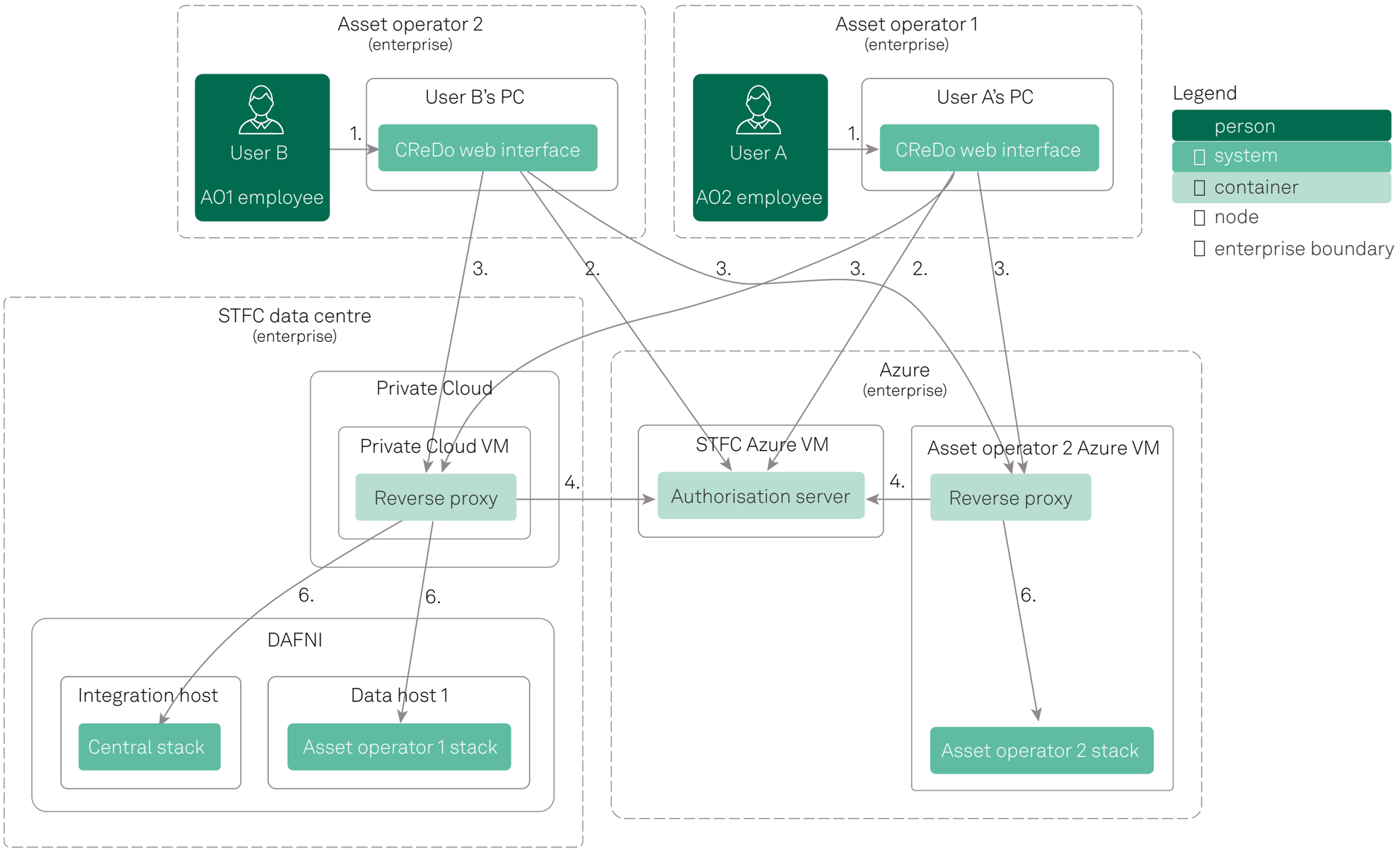
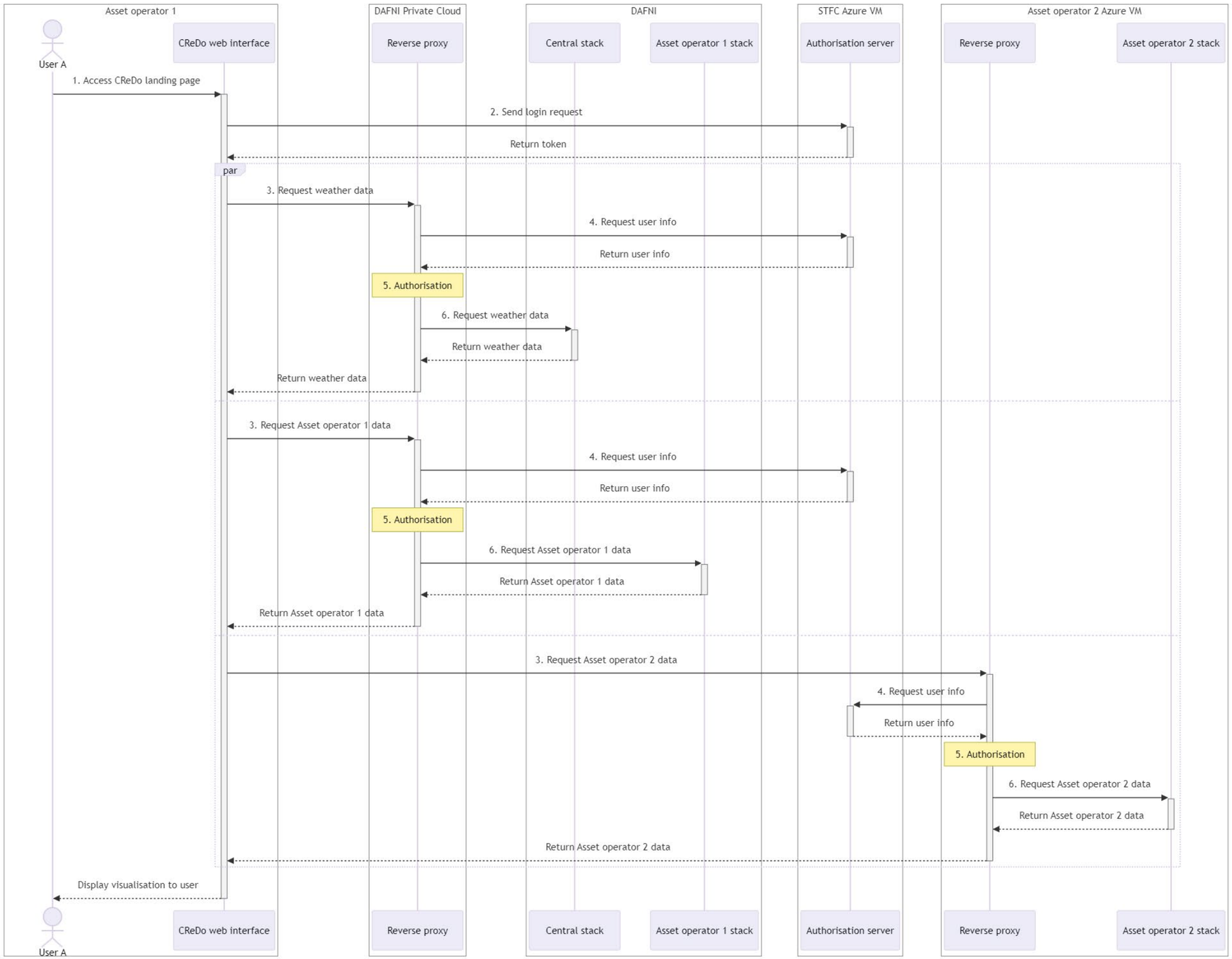


Figure 5. Architecture diagram showing the connectivity of components that are engaged when a user accesses the CReDo web interface after the changes made in the latest development cycle.



Distributed architecture on Azure – evaluating scenarios

To get insights from CReDo, scenarios need to be evaluated; each scenario requires one or more sets of asset operator data and appropriate weather data to be specified by the user. The datasets are then queried for the information required by the model, which is then evaluated. The results from the model are then passed back to the relevant asset operator stack, so that it can be accessed through the geospatial visualisation and data dashboard within the web interface.

The process of evaluating a scenario before the latest development (CReDo deployed on DAFNI)

The main steps in the process are outlined below and are visually represented in architecture and flow diagrams in Figures 7 and 8.

1. A user logs into to the CReDo integration host using ssh and sends a request to the Central stack to evaluate a specific scenario.
2. The Central stack requests the Asset data from the Asset operator stacks.
3. The Asset operator stacks request data about the scenario from the Central stack.
4. The weather data is accessed and queried at the site/asset locations.
5. Data is sent to the model and the model is evaluated.
6. The results of the model are returned to the Central stack.
7. The Central stack passes the results back to the relevant Asset operator's stack.
8. The user is informed that the scenario evaluation is complete.

Figure 6. Flow diagram showing the communication between components that are engaged when a user accesses the CReDo web interface after the changes made in the latest development cycle.

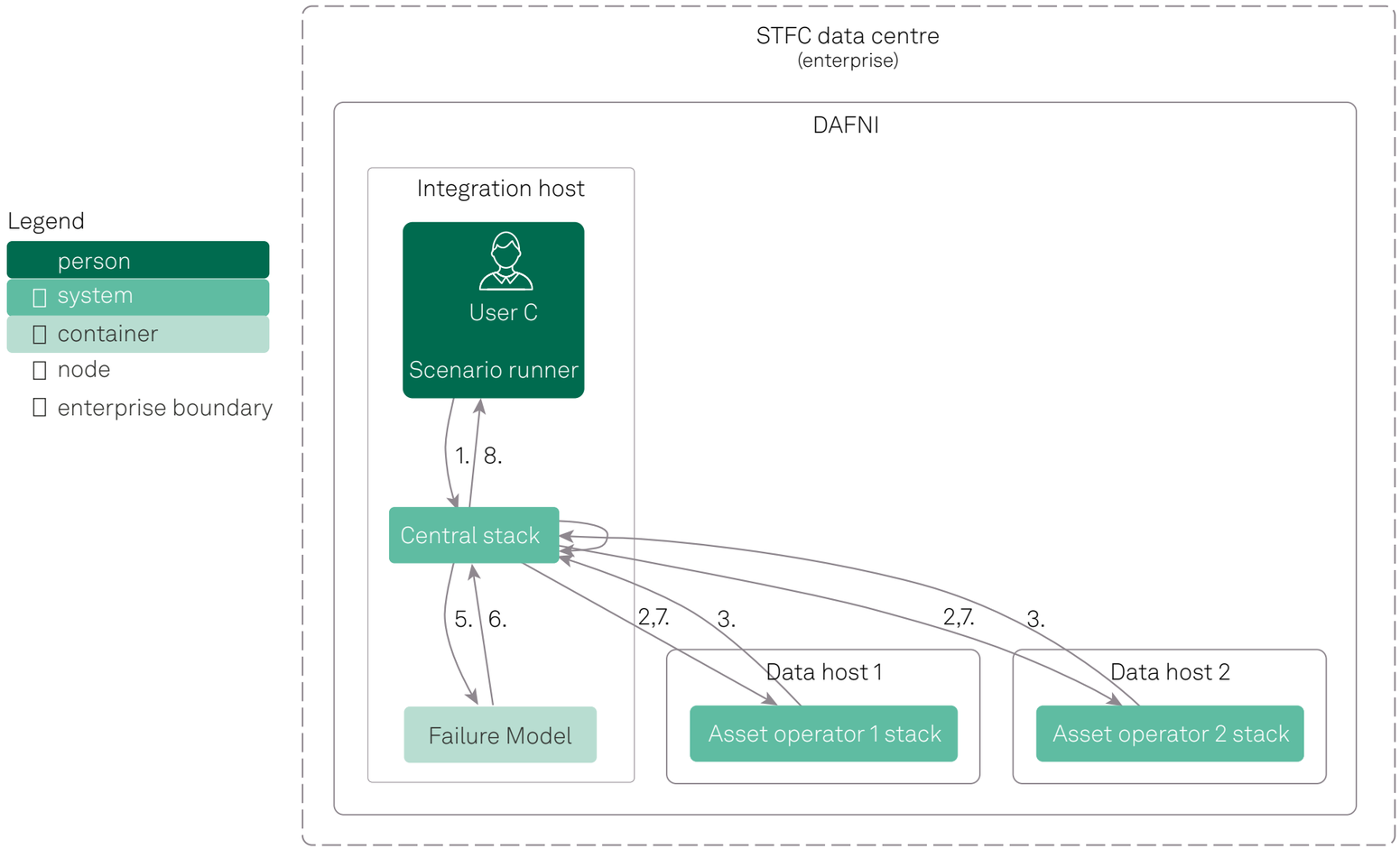


Figure 7. Architecture diagram showing the connectivity of components that are engaged when a scenario is evaluated in CReDo before the changes made in the latest development cycle.

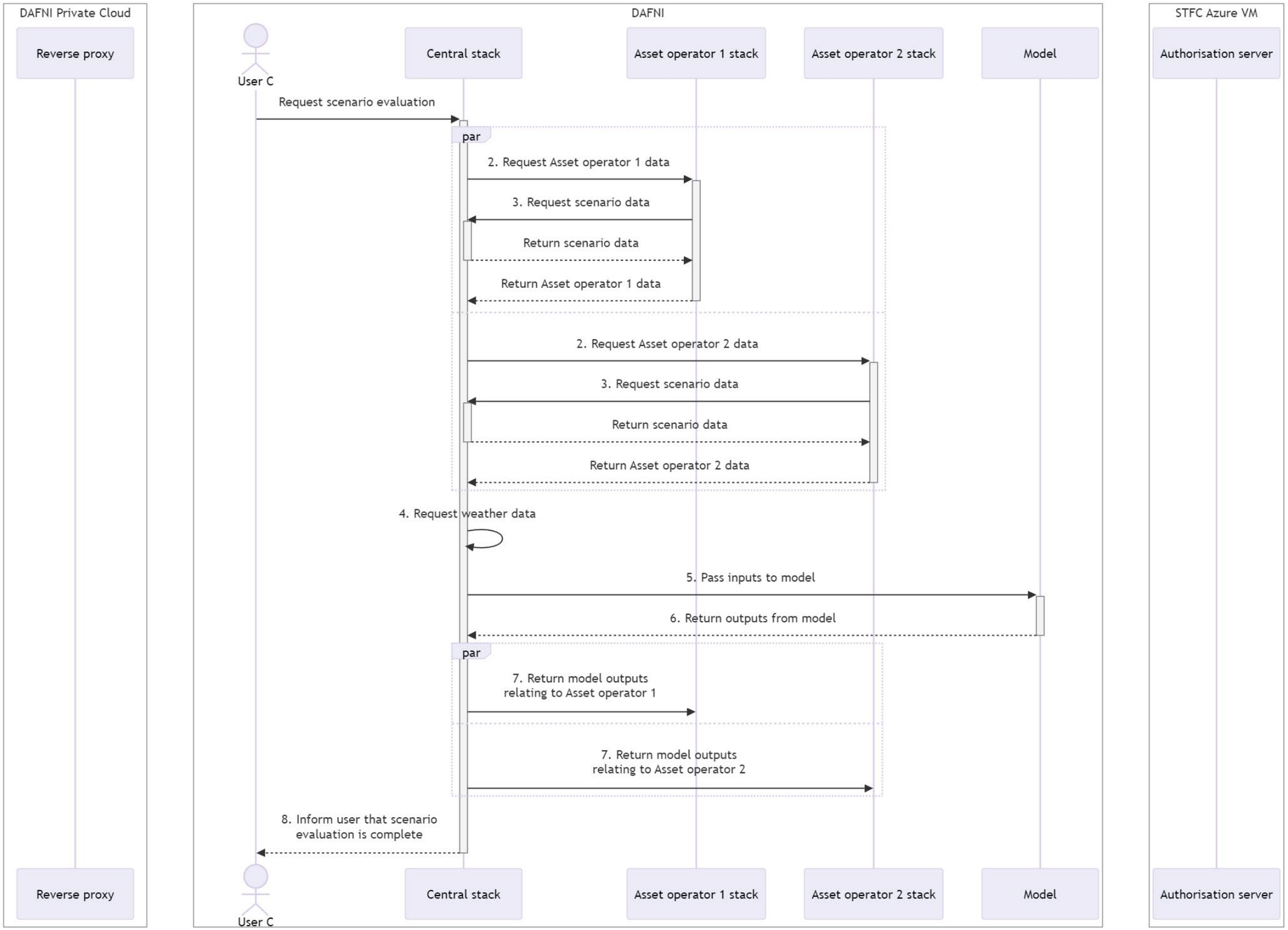


Figure 8. Flow diagram showing the communication between components that are engaged when a scenario is evaluated in CReDo before the changes made in the latest development cycle.

The process of evaluating a scenario after the latest development (CReDo deployed on DAFNI and MS Azure)

The process is equivalent to the one illustrated in the previous subsection; the difference is that when requests cross enterprise boundaries then authentication occurs as a part of the initial request (1a.) and authorisation proceeds as follows:

a. Request sent to the relevant reverse proxy.

b. The reverse proxy requests information about the user from the authorisation server.

c. The reverse proxy authorises the data requests accordingly.

d. The data requests are forwarded to the relevant stack and the data is returned.

The corresponding visual representations (architecture and flow diagrams) can be found in Figures 9 and 10.

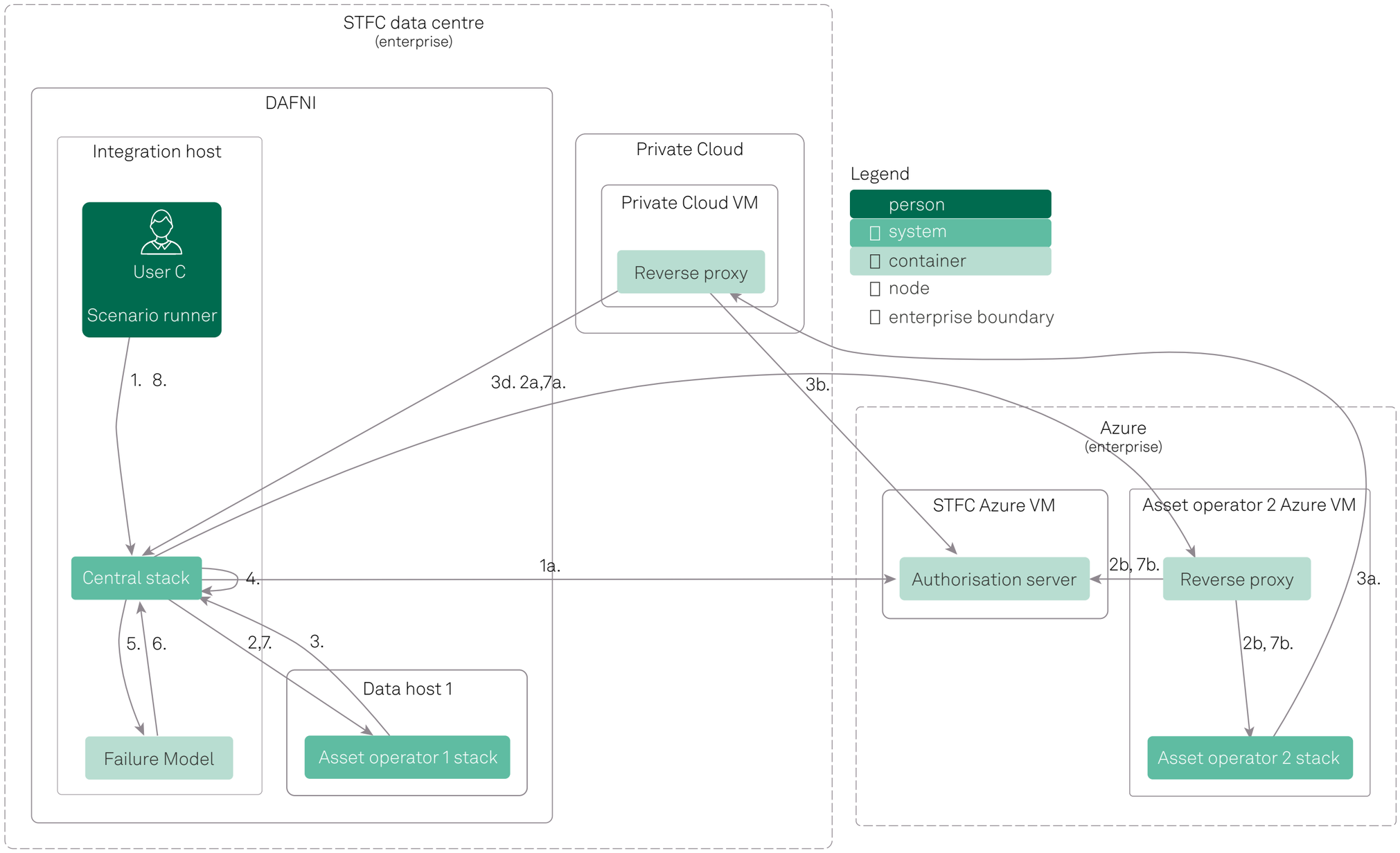
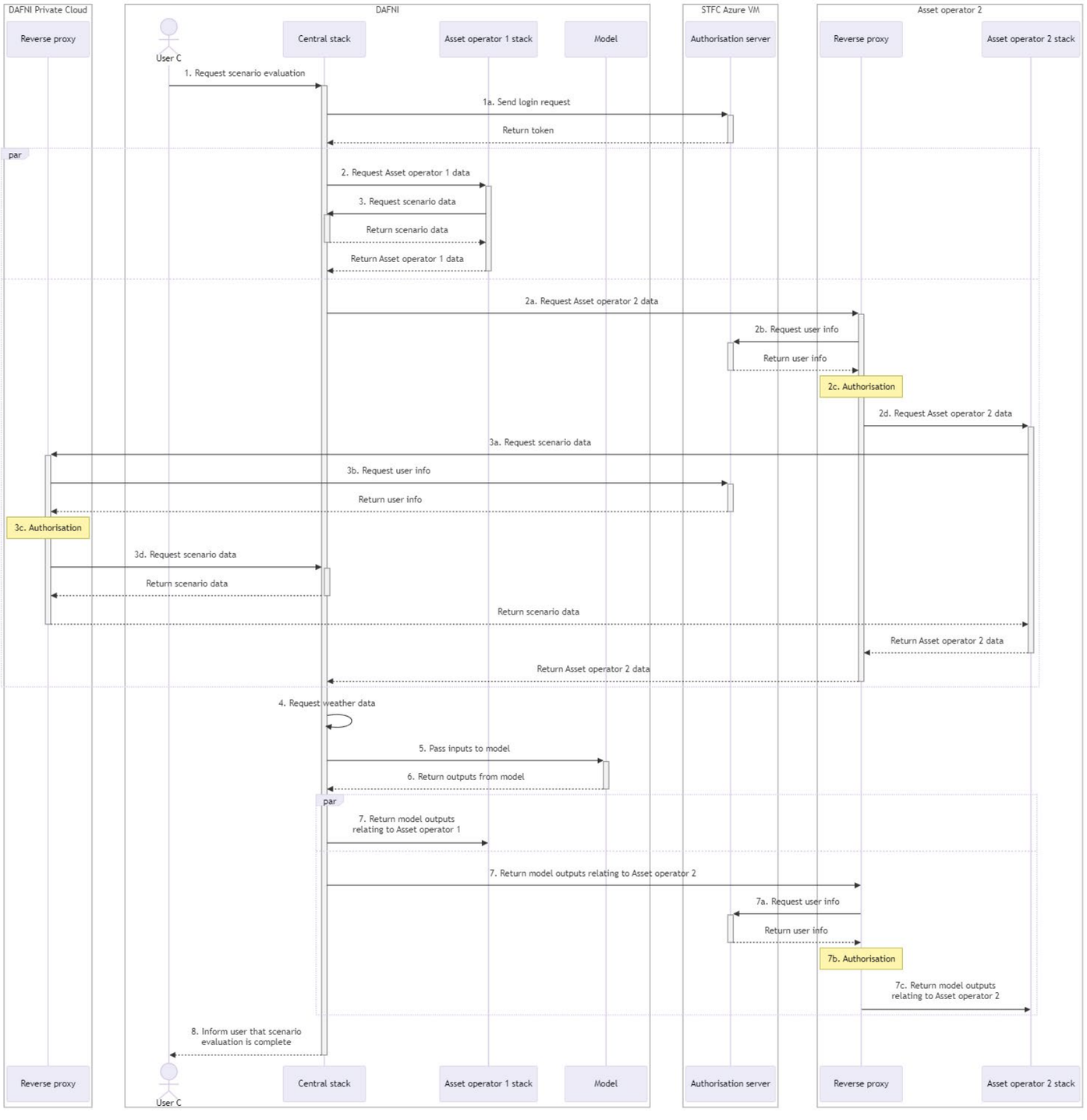


Figure 9. Architecture diagram showing the connectivity of components that are engaged when a scenario is evaluated in CReDo after the changes made in the latest development cycle.



CReDo distributed architecture – conclusions and future work

As part of this work, we have successfully deployed part of the CReDo system on a Microsoft Azure VM outside of DAFNI whilst maintaining comparable levels of security. This opens the door to allowing asset operators to deploy their CReDo stack on DAFNI or a cloud platform of their choosing.

Future work could include looking into adapting CReDo to work on other cloud platforms such as AWS (Amazon Web Services) or Google Cloud, and/or within their organisational boundaries.

Another area of potential future development could be to investigate more advanced authorisation and authentication methods that might include technologies such as token exchange.

Figure 10. Flow diagram showing the communication between components that are engaged when a scenario is evaluated in CReDo after the changes made in the latest development cycle.

APPENDIX 1

Interview questions for CReDo partners and external experts – updating the legal framework

- Are there any legal risks with our planned approach to data sharing?
- How feasible is it for asset owners' initial contract to cover future data updates?
- Is it necessary for you to know who else is signing the licence? Why or why not?
- Do you need to know who is seeing derived insights?
- What needs to change in the existing licence to make it a commercial data licence? What would the key issues be in negotiations?
- Where should liability sit in a distributed architecture?

Visit Connected Places Catapult website **here**

Follow us on Twitter **@CPCatapult**

Follow us on LinkedIn **Connected Places Catapult**

Email us **info@cp.catapult.org.uk**

CATAPULT Connected Places