CReDo

CReDo phase 2

Technical Report – Distributed Architecture

Report written by CMCL

Table of contents

Authors

Author: Jethro Akroyd (**CMCL**)

Executive summary

CReDo, the Climate Resilience Demonstrator, is a climate change adaptation digital twin demonstrator project to improve system-wide resilience across infrastructure networks.

CReDo is a pioneering project to develop a digital twin across infrastructure networks to provide a practical example of how connected data and greater access to the right information can improve climate adaptation and resilience. CReDo demonstrates how it is possible to connect data across organisations to deliver efficiencies and wider societal benefits.

Enabled by funding from Innovate UK via Connected Places Catapult, CReDo looks specifically at the impact of extreme weather, in particular flooding, on energy, water and telecoms networks. CReDo brings together asset data, flood hazard data and asset failure models to provide insights into infrastructure interdependencies and how they would be impacted under future climate change flooding scenarios. The vision for CReDo is to enable asset owners, regulators and policymakers to collaborate using the CReDo digital twin to make decisions which maximise resilience across the infrastructure system rather than from a single sector point of view.

This second phase of CReDo running from April 2022 to March 2023 has sought to:

- Move existing CReDo functionality to a distributed architecture where asset owners can host their data assets in their own secure environments, and CReDo can access the relevant data and return insights securely.
- Enable asset owners to see insights from the combined infrastructure network and, by signing up to an additional part of the data licence, a system-level view of the network.
- Build in extensibility to accommodate the future addition of new asset owners and data types.
- Continue to explore and develop methods of promoting interoperability.
- Unlock the strategic resilience planning use-case by enabling asset owners to access real data for both their own and interconnected assets via CReDo.

This report documents the implementation of the distributed architecture, and how it is deployed to enable the extensible cross-sector sharing of data describing infrastructure, and the generation of insights to facilitate decision support to improve the climate resilience of the combined infrastructure network.

Introduction

This report documents the implementation of the second phase of the CReDo connected digital twin. CReDo integrates a description of assets from the energy, water and telecoms networks with data describing flood hazards for different climate change scenarios. It resolves the effect of floods on individual assets and the corresponding cascade of effects across the combined network.

CMCL[1] were engaged by Connected Places Catapult to develop a distributed architecture for the CReDo digital twin. The internal data structure of CReDo is based on the use of ontologies to represent information as a knowledge graph. The nodes of the knowledge graph represent data as instances of concepts that describe the type, operational state and location of each asset, and the edges of the graph represent relationships between nodes.

The first phase of CReDo concentrated on how to use the knowledge graph to enable interoperability between asset data from different sectors, how to interface with data describing flood hazards, and how to resolve the cascade of effects across the combined network. Full details are provided in the CReDo Technical Paper 1: Building a cross-sector digital twin report published at the end of the first phase of CReDo.[2]

Although the CReDo digital twin developed in Phase 1 represented the data as a knowledge graph, it did so by taking a copy of the data and combining it on a single secure host. The need to host a copy of data in a central facility raised several questions:

- How to scale CReDo to include other asset owners and other types of assets?
- How to enable data sharing to develop and share insights based on the combined network, while respecting the confidentiality and security of the data?

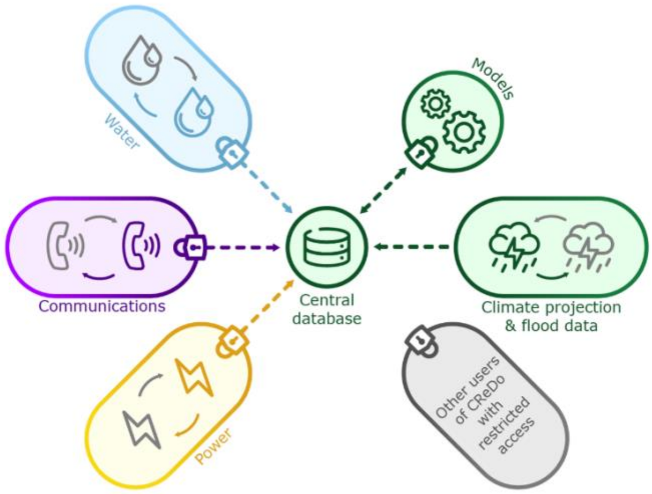
The ability to scale and extend CReDo is critical to understanding the interaction between networks and the impact of cascading risk. These questions motivated the development of an extensible distributed architecture in CReDo Phase 2 to enable the possibility of allowing asset owners to share data and benefit from insights based on the combined network, while retaining control of their data within their own IT systems.

Figure 1(see next page) illustrates the centralised data sharing architecture used during CReDo Phase 1 and the distributed architecture demonstrated in CReDo Phase 2. This report documents the implementation of the distributed architecture.

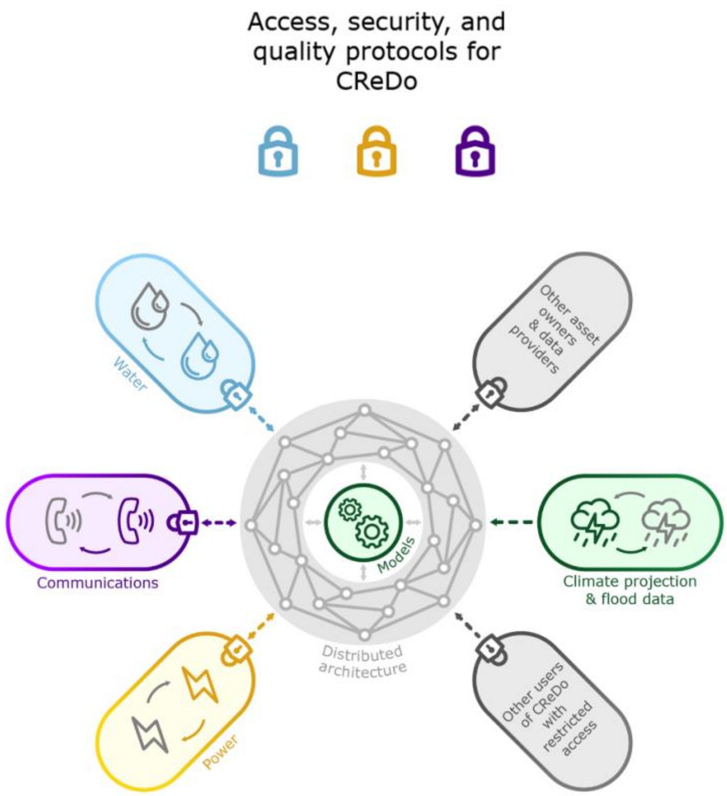
---

[1] https://cmcl.io

[2] https://doi.org/10.17863/CAM.81779



(a) Centralised data sharing architecture.



(b) Distributed data sharing architecture.

Figure 1. CReDo data sharing architectures.

CReDo architecture

The CReDo digital twin uses a distributed architecture. It allows asset owners to share data and benefit from insights based on the combined infrastructure network, with the intention to enable them to retain full control of their data assets.

The objectives of CReDo Phase 2 were to:

1. Move existing CReDo functionality to a distributed architecture where asset owners can host their data assets in their own secure environments, and CReDo can access the relevant data and return insights securely.
2. Enable asset owners to see insights from the combined infrastructure network and, by signing up to an additional part of the data licence, a system-level view of the network.
3. Build in extensibility to accommodate the future addition of new asset owners and data types.
4. Continue to explore and develop methods of promoting interoperability.
5. Unlock the strategic resilience planning use-case by enabling asset owners to access real data for both their own and interconnected assets via CReDo.

An additional requirement was for the distributed architecture to be able to host scenarios defined by a combination of asset and hazard data. By comparing different scenarios, we can understand the vulnerability of the combined asset network to different climate scenarios and can understand the difference in implications for different time horizons, 2030, 2050 and 2070 for example, and for different hazards, for example coastal, fluvial and pluvial flooding. The latter is important because the type of protection required from saltwater flooding may be quite different from that required for protection for freshwater flooding. The ability to host scenarios additionally provides the possibility of hosting data describing alternative asset bases, for example the planned configuration of the water network in 2030. The ability to assess the implications of future plans on the combined asset network is a vital component of strategic resilience planning.

The distributed architecture used in CReDo Phase 2 is shown in Figure 2(in the next page) The asset owner data is hosted on separate servers, which could be hosted by asset owners in their own IT systems. The distributed architecture enables CReDo to connect securely to a copy of the data held on these servers, and could also enable the possibility to connect directly to asset owner data at source. By separating the data, the architecture seeks to avoid the risks associated with hosting all the data in the same place. Each asset owner server also hosts a visualisation, providing a means to facilitate different views of the insights and data.

In the current implementation, a central node is responsible for resolving the connectivity between assets, for running the models that simulate asset failures and the impact across the combined asset network, and for managing scenarios. The architecture enables other data and service providers to contribute to CReDo. For example:

- Climate projection and flood data. In Phase 2, data describing coastal, fluvial and pluvial flood hazards were contributed by Fathom.[3]

[3] https://www.fathom.global

- Models that contribute to the analysis of the combined network. In Phase 2, models describing (i) the failure of assets and the cascade of effects across the combined network, and (ii) the economic assessment of potential resilience measures were contributed by Connected Places Catapult and Frontier Economics[4] working in conjunction with Anglian Water, BT and UK Power Networks.
- Other sources of data. In Phase 2, the ability to accommodate new asset owners and data types was demonstrated by including open data describing NHS infrastructure.[5]

The communication between all elements of the architecture is encrypted using standard protocols.

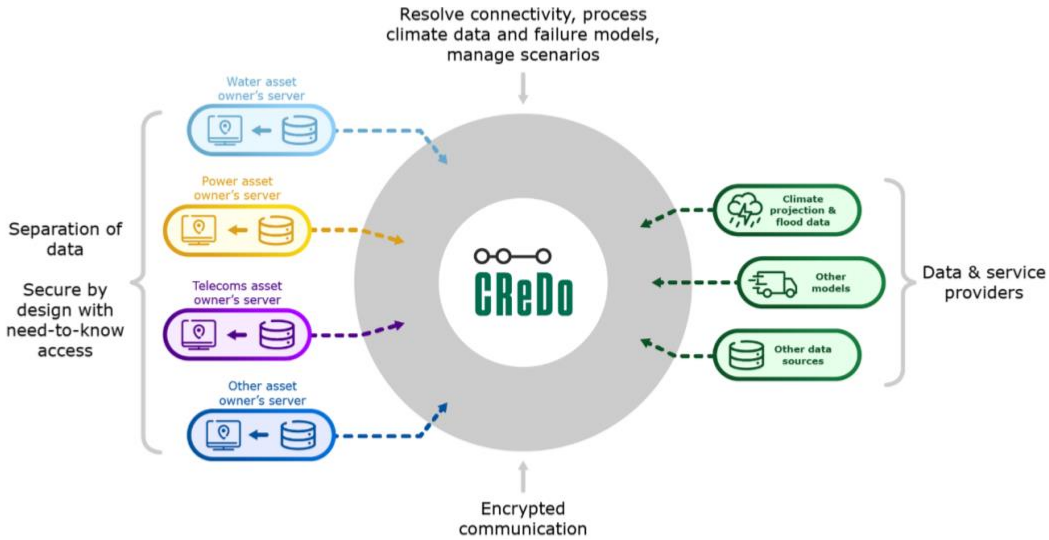


Figure 2. CReDo Phase 2 distributed architecture.

The functionality of CReDo continues to be based on the use of a knowledge graph as described in the CReDo Phase 1 report.[2] The difference in CReDo Phase 2 is that the knowledge graph is now distributed across multiple servers. The asset owner data are mapped to the CReDo ontologies using Ontop[6] (see the Phase 1 report). This permits the possibility of mapping data from whatever format is used by asset owners into the CReDo data structure. This maximizes outward compatibility, for example the ability to map from other data models. It means that CReDo (and other tools) can use whatever internal data structures suiting their purposes, while using mappings to access data in its native format. This includes the possibility of different asset owners having different data formats. This is important because digitalisation is a journey, with different parties being in different places on that journey.

CReDo is implemented using containers deployed and managed via Podman[7], so it is system and cloud agnostic. The asset owner data was hosted by STFC in a secure environment within DAFNI[8] during Phase 2. This included using separate virtual machines to play the role of each asset owner server. This mimicked the envisaged distributed set up in a safe environment while Phase 2 was in progress. A solution that uses multi-factor authentication (MFA) to provide security and access controls was developed by STFC to provide asset owners with a secure means to access the visualisations. The MFA solution is responsible for both authentication (confirming the identity of a user) and authorisation

[4] https://www.frontier-economics.com
[5] https://www.nhs.uk/about-us/nhs-website-datasets
[6] https://ontop-vkg.org
[7] https://podman.io
[8] https://www.dafni.ac.uk

(controlling which nodes are accessible to a user). The specifications of the secure hosts and the MFA solution are reported separately and are available via the DT Hub.[9]

The vision for CReDo is of an extensible system, where multiple providers can offer models and data. This extensibility is facilitated by the idea of computational agents that provide discrete elements of functionality, as opposed to a monolithic tool. The implementation is based on ideas[10] developed by CMCL in collaboration with the Computational Modelling Group[11] at the University of Cambridge and the Cambridge Centre for Advanced Research and Education in Singapore (CARES)[12] as part of the World Avatar research project.

The following sections describe the elements of the distributed architecture developed in CReDo Phase 2. Details reported previously are not repeated. Rather, the reader is referred to the CReDo Phase 1 report.[2]

CReDo data structure

The starting point to build the CReDo knowledge graph is to define ontologies. These form the internal data structure of CReDo. By first intent, CReDo Phase 2 sought to leave the ontologies from Phase 1 unchanged while developing the distributed architecture. The rationale for this was to avoid changing multiple things at once. However, minor changes to the ontologies were necessary and are described below. As per CReDo Phase 1, it is recommended that a future phase of CReDo review the ontologies to ensure that they describe the assets in the level of detail required to support future failure models.

The ontologies that CReDo uses are hierarchical. At the top of the hierarchy, we define concepts and relationships about asset and flood data that apply throughout CReDo. We then define specialisations of these concepts for each type of the water, telecoms, energy networks and NHS infrastructure networks. The ability to define such specialisations confers extensibility because it makes it easy to add new specialisations for new assets.

The main components of the ontology are shown schematically in Figure 3 (see next page) The filled boxes represent concepts, the hollow boxes represent data objects, and the arrows represent relationships. CReDo Phase 2 introduces a *Connection* concept to represent explicitly the links between assets in the networks. This is different from CReDo Phase 1, which implicitly encoded the connections using specialisations of the *supplies* relationship. This change facilitates asset owners having access to information regarding external connections, and their states, without having to expose the details of the assets in other networks. It is necessary to support the idea of connections having a state, and means that connections can now also have properties associated with them, for example, the number of pylons supporting an overhead line. This may be useful for future failure models.

---

[9] https://digitaltwinhub.co.uk/credo
[10] https://doi.org/10.1017/dce.2021.10
[11] https://como.ceb.cam.ac.uk
[12] https://www.cares.cam.ac.uk

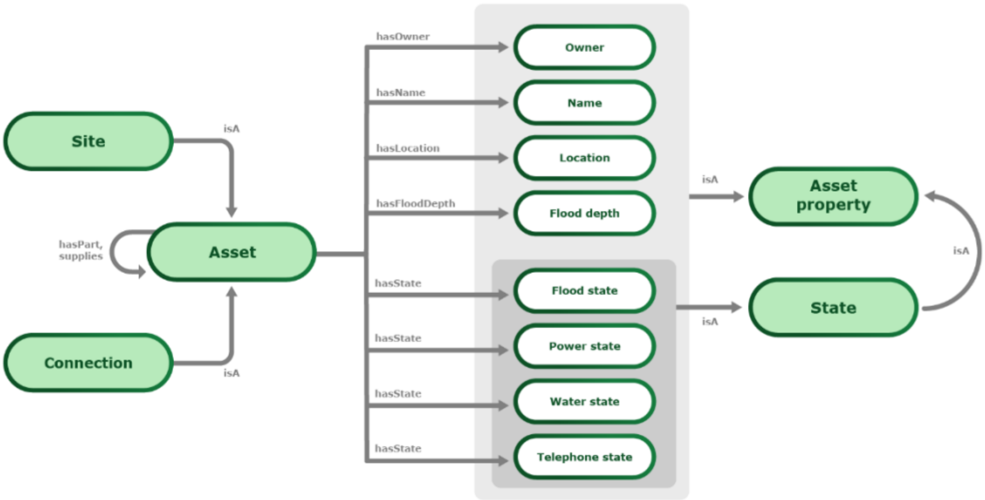


Figure 3. CReDo Phase 2 core asset ontology (simplified).

The main specialisations for each infrastructure network are shown schematically in Figure 4 – Figure 7. The green boxes represent concepts from the core asset ontology, showing the inheritance relations used to extend the ontology. The other filled boxes represent concepts used to describe the specialisations for each network. Hollow boxes represent data objects. The use of the *Asset*, *Site*, *AssetProperty*, *FloodState*, *FloodDepthState* and *State*, and their corresponding specialisations remains as described in the CReDo Phase 1 report.

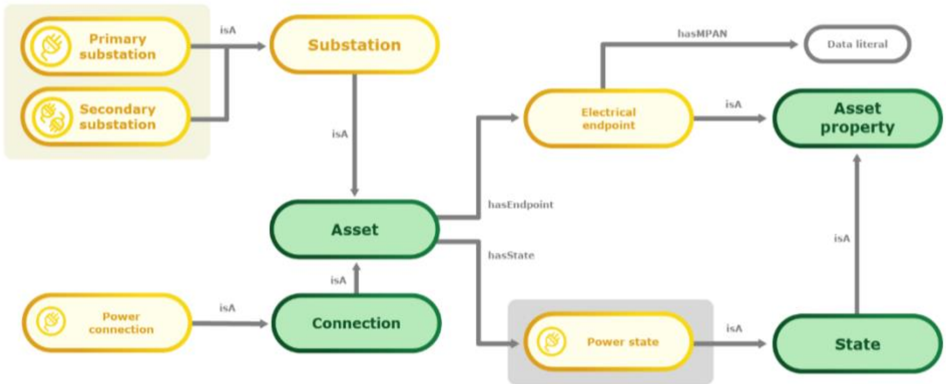


Figure 4. CReDo Phase 2 sub-domain ontology (simplified): energy network.

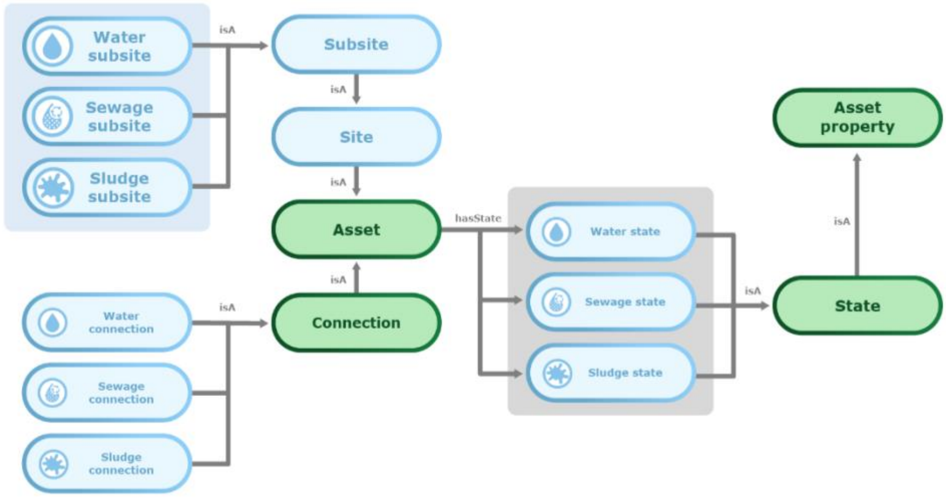


Figure 5. CReDo Phase 2 sub-domain ontology (simplified): water network.

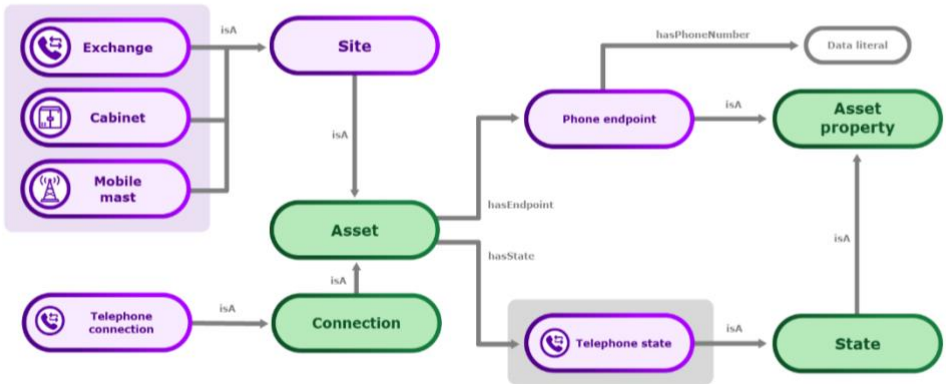


Figure 6. CReDo Phase 2 sub-domain ontology (simplified): telecoms network.

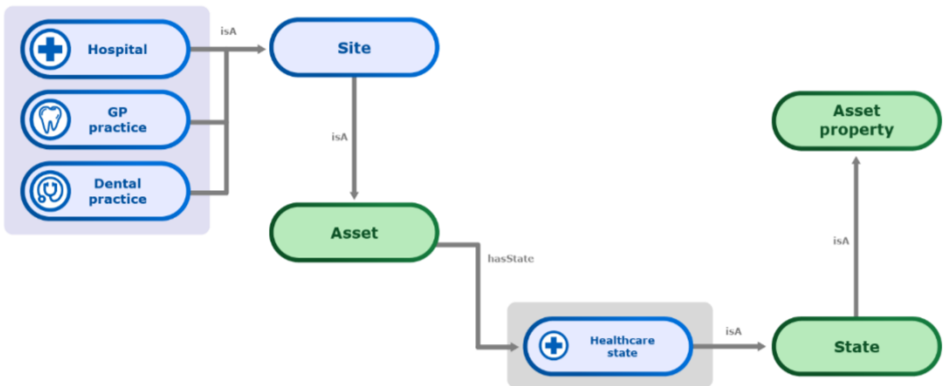


Figure 7. CReDo Phase 2 sub-domain ontology (simplified): NHS network.

When the ontologies are used to represent data, the data forms a directed graph. Figure 8 shows an example. The instances of concepts – in this case assets (large circles with symbols) and connections (small hollow circles) – form the nodes of the graph, while the instances of relationships between concepts, in this case instances of the *supplies* relationship, form the arrows between nodes.

Figure 8. Example CReDo knowledge graph.

Scenarios and datasets

CreDo Phase 2 formalises the idea of scenarios introduced in Phase 1. In CreDo Phase 2, a scenario consists of ***data sets*** describing:

- A hazard (*e.g.*, coastal, fluvial and/or pluvial flooding calculated under a given set of climate change assumptions).
- An asset base from each asset owner.

A data set can belong to more than one scenario. A future phase of CreDo may choose to extend the failure models so that it is possible for a scenario to consider combinations of hazards. For example, extreme heat followed by flooding or vice versa. Further work may also be required to refine the restrictions placed on the asset bases that can belong to a scenario, for example to allow for more complex situations where an asset owner has multiple asset bases covering different areas.

A new ontology was needed to represent the information required to describe data sets and scenarios as part of the knowledge graph. The ontology is shown schematically in Figure 9.

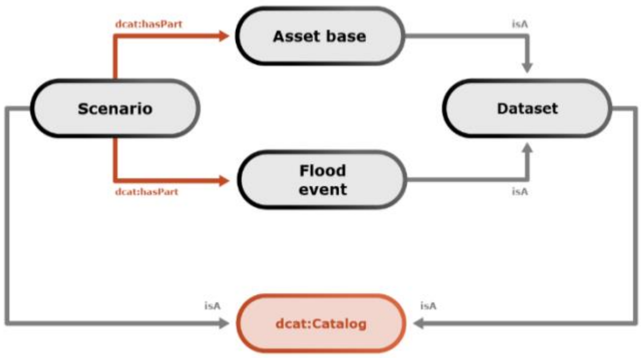


Figure 9. CReDo Phase 2 data set and scenario ontology (simplified).

In addition to allowing the specification of human-readable metadata, the ontology allows for the specification of an identifier and the information necessary to access each data set and scenario. The components that comprise each scenario are show schematically in Figure 10.

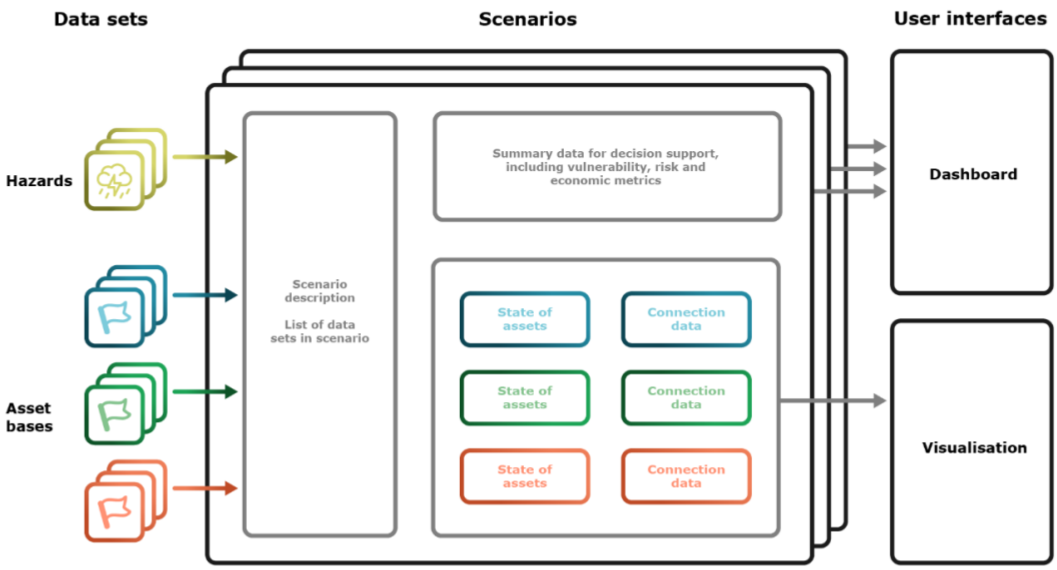


Figure 10. Components of CReDo Phase 2 scenarios.

Distributed data hosting

The components that comprise each scenario are stored in different locations within the distributed architecture depending on what they are. A description of each component and where it is hosted are given in Table 1.

Table 1. Description and location of CReDo Phase 2 data components.

Component	Description	Location
Ontologies	Definition of the concepts and relationships used by CReDo.	All nodes that use the ontologies
Hazard data	Raster data describing flood depth for a given (coastal, fluvial or pluvial) hazard under a given set of climate assumptions.	Central node
Asset data	Data describing assets that form an infrastructure network.	Asset owner node
Route data	Data defining the endpoint of each CReDo node.	Central node
Scenario description	Data defining scenarios, including which data sets belong to each scenario.	Central node
Asset states	Data defining the flood depth and state of assets.	Asset owner node
Connection info	Data defining the existence and state of connections from an asset base to other assets.	Asset owner node
Connection info	Data defining the type and geographic coordinates of each connection.	Central node[13]
Summary data	Per-asset data describing vulnerability and risk to support the dashboard.	Asset owner node
Summary data	Per-scenario data, for example cost-benefit metrics, to support the dashboard.	Central node

[13] This information is required to support the user interface and is a temporary solution. It should be designed out in the future.

Data processing

The CReDo Phase 2 digital twin calls a sequence of codes to evaluate a scenario and provide the functionality required by the CReDo user interface. A description of the codes is given in Table 2. The order of presentation follows the calling sequence of the codes, except for items marked as "asynchronous", which respond to events (*e.g.*, mouse clicks via the CReDo user interface).

Table 2. CReDo Phase 2 codes used to evaluate a scenario.

Code	Description	Location
Stack manager	Run once on each node to initialise services.	Each node
Stack data uploader	Run on each node to upload node-specific asset and/or hazard data.	Each node
Central stack agent	Call to create and run each scenario.	Central node
Access agent (asynchronous)	Runs as a server to route and federate queries and requests to the correct endpoints.	Each node
Dashboard agent	Collates data for the dashboard on each asset owner node. The agent queries all nodes; the authorisation and authentication solution controls which nodes are accessible, and therefore what data are retrieved and made available via the dashboard.[14]	Asset owner node
Feature info agent (asynchronous)	Retrieves metadata for an asset when it is selected via the CReDo user interface. Data describing the state of incoming connections from other asset owners are retrieved via a query to the central node.[14]	Asset owner node

It is recommended that a future phase of CReDo provide a graphical user interface to allow the management of data sets and scenarios (*e.g.*, addition and editing of data sets, creation, deletion, editing and evaluation of scenarios *etc.*).

[14] This aspect of the design will be subject to revision in the future.

Visualisation

The CReDo user interface provides a map-based visualisation and an analytics dashboard to allow users to interrogate the data. The map-based visualisation is based on the Mapbox GL JS[15] library, with GeoServer[16] being used to support the tiling of data to avoid overloading network connections. The dashboard is based on Apache Superset[17].

CReDo Phase 2 has demonstrated the possibility of using a distributed architecture to share insights to provide decision support based on the combined cross-sector infrastructure network, while respecting the confidentiality and security of the underlying asset data. The visualisation can be configured to permit either a restricted view of assets or a full system-level view of the combined network, depending on the terms of the prevailing data licence.

It is recommended that future phases of CReDo refine the models used to generate insights, and that the types of insight and user experience offered by the visualisation are optimised to facilitate decision support to improve the climate resilience of the combined infrastructure network. It was also observed that GeoServer occasionally crashed without providing an error message to explain the problem. Various issues were also encountered when working with Superset to develop the dashboard (see the Appendix for details). It is recommended that a future Phase of CReDo investigate the behaviour of GeoServer in more detail and consider alternatives to Superset.

---

15 https://www.mapbox.com

16 https://geoserver.org

17 https://superset.apache.org

CReDo login process

The CReDo Phase 2 digital twin was deployed at STFC in a secure environment within DAFNI. Separate virtual machines were used to play the role of servers belonging to each asset owner, who were each provided with secure access to the CReDo user interface hosted on "their" server at STFC via a multifactor authentication solution (reported elsewhere). Figure 11 shows screenshots from the login process.

Figure 11. CReDo login process using multifactor authentication.

CReDo user interface

The CReDo user interface allows the selection of different scenarios. The scenarios can support different combinations of (currently flood) hazard event and asset bases. Figure 12 shows the scenario selection interface.

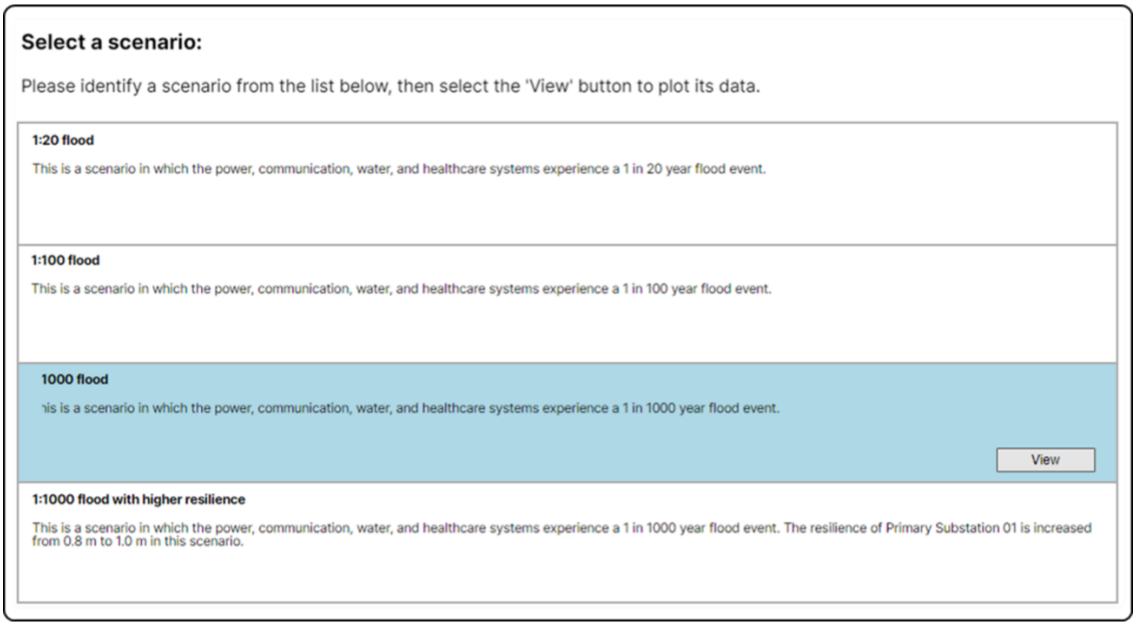


Figure 12. CReDo supports scenarios with different hazard events and asset bases.

The flood data, asset location, connectivity and operational states of assets and connections can be visualised via a map-based interface. Figure 13 (next page) shows an example for a 1-in-1000 year event under a high emissions climate scenario for a visualisation that has permissions to see the full-system view. The depth of colour in the flood layers on the visualisation is proportional to the water depth. The borders around the assets allow users to see how the cascade of problems caused by the flood traverses the combined network.

- A red circular border around an icon indicates an asset that has failed because it is flooded.
- An orange octagonal border indicates an asset that has failed because of a problem upstream. For example, a secondary substation that is no longer receiving power.
- A grey square border indicates an asset that is still working, but that has some other problem. For example, a substation that is providing power but that has perhaps lost telecoms or water.

Although CReDo was developed using real asset data, **the figures in this report use synthetic data to respect the confidentiality of the real asset data.** This is reflected in the names of the asset bases listed in the controls on the left of the visualisation.

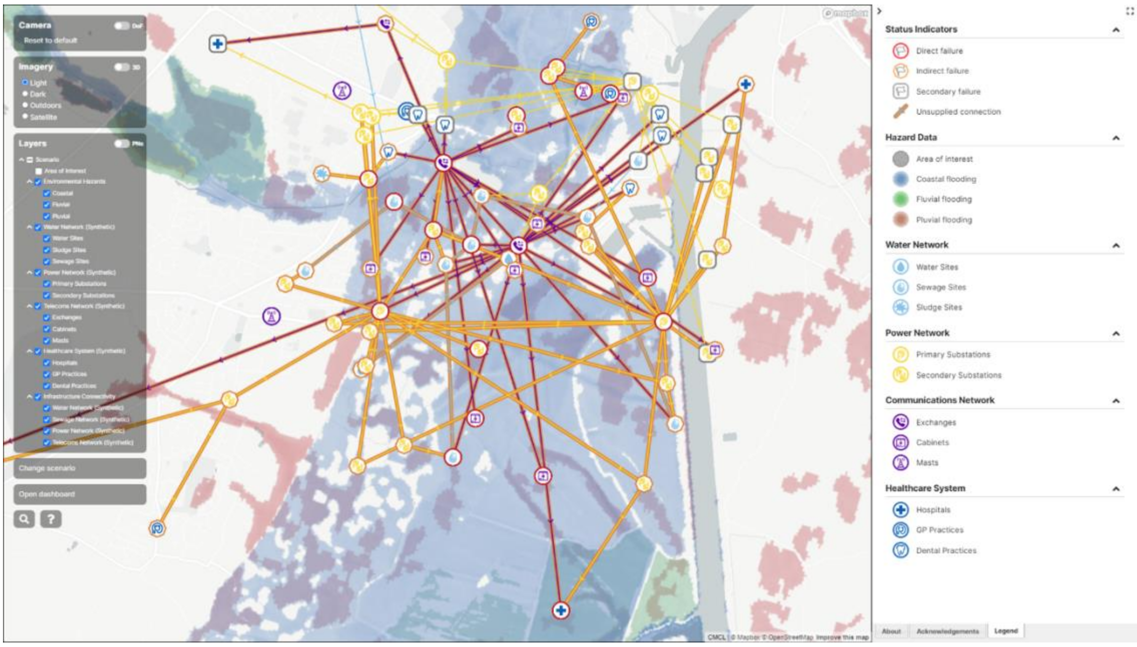


Figure 13. Flood, asset, connectivity and operational state data are represented on a map-based visualisation.

Assets can be selected by clicking on them on the map. If an asset is selected, a sidebar on the right of the visualisation will display detailed information about the asset, including its name and metadata about the asset, the operational state of the asset and the incoming and outgoing connections. This helps understand the upstream cause of failure. Figure 14 shows an example. The selected sewage site has suffered a direct failure due to flooding, and has additionally lost incoming power and telephone connections.

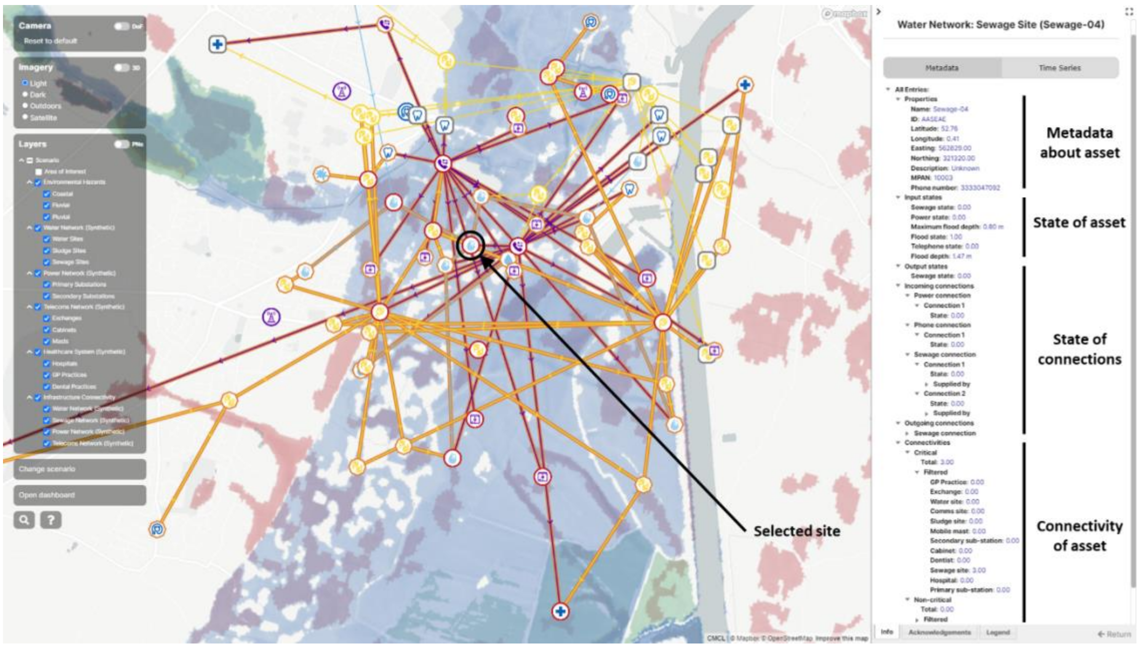


Figure 14. Asset metadata is shown on the sidebar of the map-based visualisation.

The metadata about an asset also includes information about its connectivity. The connectivity is calculated by counting the number of direct and indirect connections to things that depend on the asset. It is broadly proportional to the number of things that would have a problem if the asset failed. Fractional values arise from the consideration of redundancies in supply routes.[18]

The distributed architecture enables the visualisation of insights derived from the combined infrastructure network with different views of the asset data. Figure 15 shows a version of Figure 14 from a visualisation that is not authorised to view data from the power or telecoms networks. While the assets and the detailed connectivity of the other networks can no longer be visualised, the visualisation continues to show the same summary information about the incoming power and telecoms connections as per Figure 14.

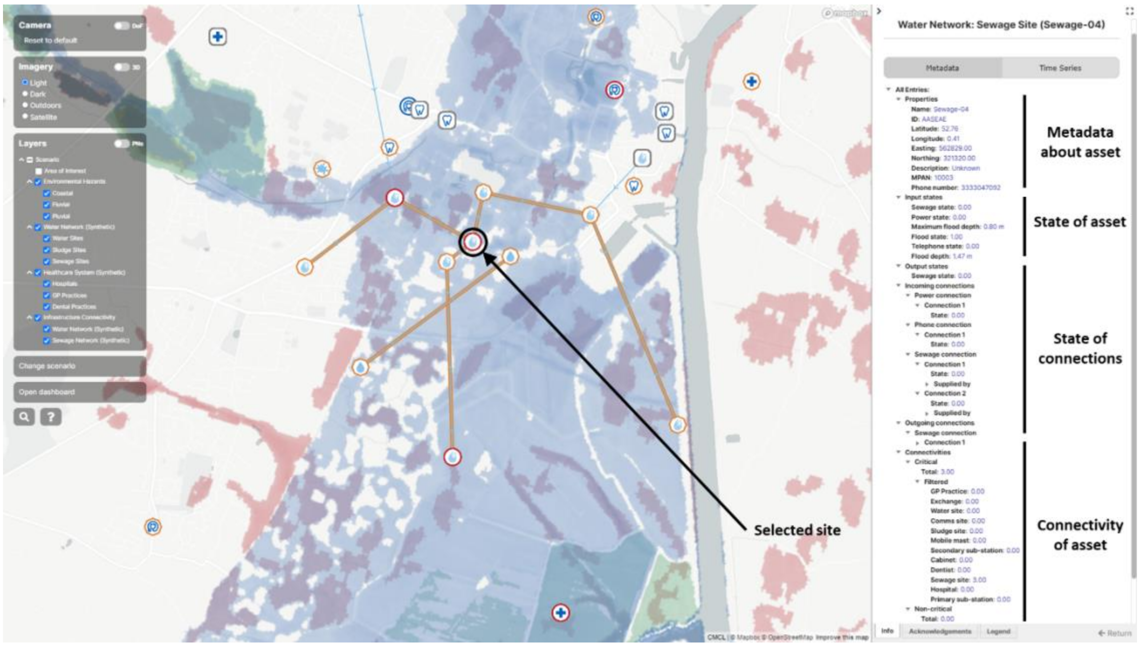


Figure 15. Different view of shared data based on a different authorisation level.

In addition to the map-based visualisation, CReDo now also provides a prototype of an analytics dashboard to summarise information across multiple scenarios. The vision is to provide a succinct summary of the information needed to support decision making about interventions. For example, an overview of the most vulnerable assets in each network, measures of the extent of the cascading risk, and aggregated economic metrics about the cost and benefit of different options to increase the resilience, with all quantities calculated based on the full set of shared data and the cascade of failures across the combined infrastructure network.

Figure 16 shows the current iteration of the dashboard. It provides a filter to control which scenarios are included in the summary, and reports the number of failed assets in each network, the assets with the highest connectivity and risk (currently assessed as the product of the failure probability and connectivity – a measure of consequence), and most vulnerable assets (currently assessed in term of flood depth). A basic feature is provided to enable data to be exported from each element of the dashboard. The immediate future iteration of CReDo will add an economic model so that aggregated cost-benefit metrics can be reported via the dashboard. As the complexity of the failure models evolves,

[18] The connectivity was previously described as "criticality" in CReDo Phase 1.
See the Phase 1 report (https://doi.org/10.17863/CAM.81779) for details of the calculation.

the dashboard should be updated to refine the metrics used to summarise the vulnerability of assets, the number of connections at risk and the level of cascading risk.

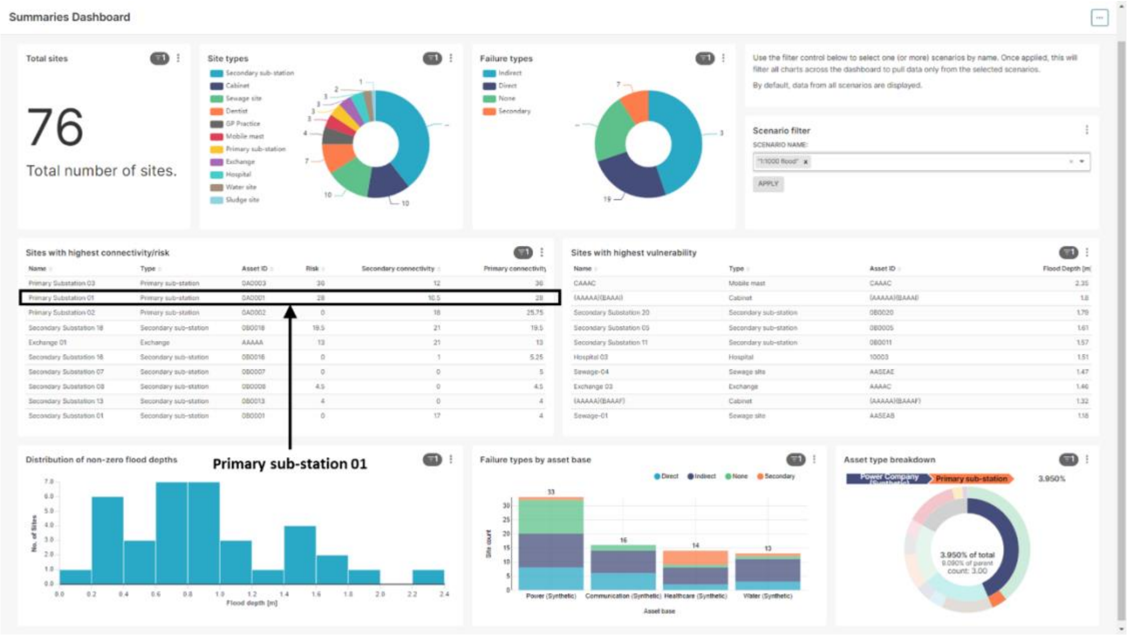


Figure 16. Dashboard to summarise information across multiple scenarios.

The visualisation also includes a simple search feature that can be used to locate items on the map. Figure 17 shows the result of a search for Primary sub-station 1, which appears with high connectivity and risk in the scenario on the dashboard shown in Figure 16.

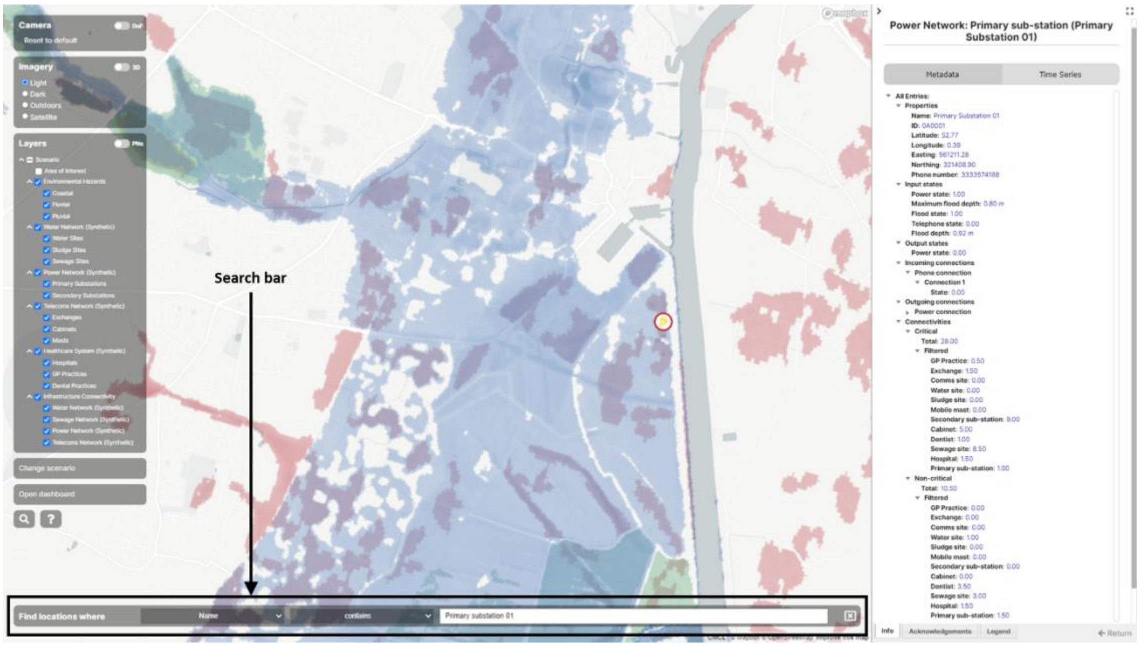


Figure 17. Search feature to find items on the map-based visualisation.

The hierarchical use of the ontologies described in Section **Error! Reference source not found.** makes is straightforward to extend CReDo to include new asset types from new asset owners. Figure 18 (next page) shows an example where this has been done for NHS infrastructure. The figure demonstrates the ability to resolve the effect of the failure cascade on new assets. Primary sub-station 1 indirectly supplies two hospitals. The first hospital is flooded. It has directly failed and has lost power and telephone service. The second hospital has indirectly failed. It is not flooded, but has still lost power and telephone service.

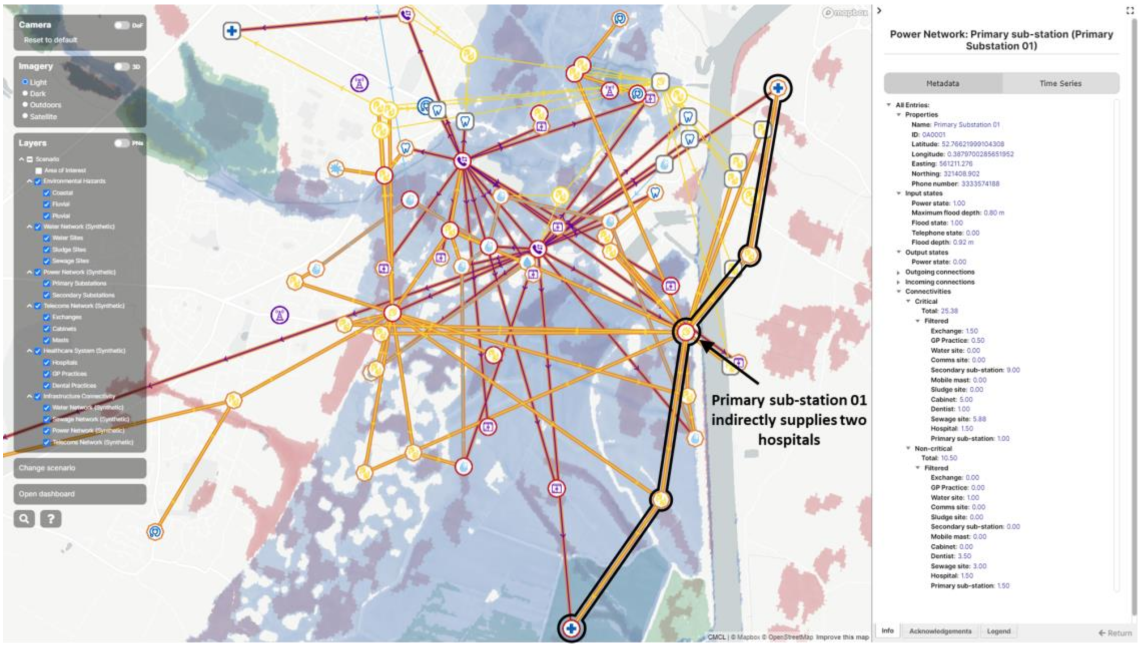


Figure 18. Loss of power to hospitals supplied via Primary sub-station 1.

The ability of CReDo to represent different scenarios allows visualisation of the impact of changes to improve the resilience of the asset bases. Likewise, the impact of different flood scenarios. Figure 19 shows the effect of the same 1-in-1000 year event considered in Figure 18, but with an alternative asset base for the power network in which the resilience of Primary sub-station 1 has been increased. In this new scenario, we avoid the loss of power to both hospitals**Figure 18**.

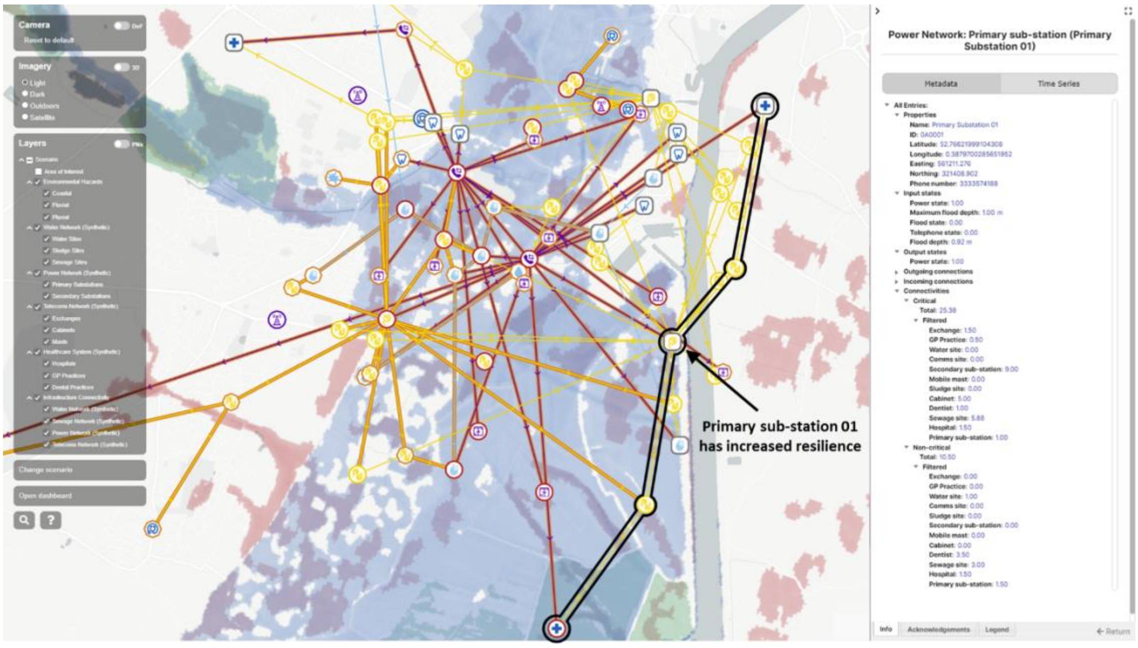


Figure 19. An alternative scenario with additional flood protection for Primary sub-station 1. The hospitals no longer lose power.

Recommendations

CReDo Phase 2 has demonstrated the use of a distributed knowledge-graph-based architecture to share data across sectors to investigate the impact of extreme weather, in particular flooding, on energy, water and telecoms networks. This section summarises what has been achieved and makes recommendations for how CReDo might be extended in the future.

CReDo demonstrates how sharing models and data across sectors can be used to provide decision support to increase climate resilience. CReDo uses a hierarchy of ontologies to create a knowledge graph that integrates and enables interoperability between data describing flood hazards with data describing:

- Anglian Water's water and sewerage infrastructure.
- BT and Openreach's communication infrastructure.
- UK Power Networks' power network infrastructure.
- NHS infrastructure.

CReDo uses models that operate on the knowledge graph to simulate the response of the combined infrastructure network to different hazard scenarios. Prototype models have been demonstrated to identify the vulnerable parts of the network and evaluate options to improve the climate resilience of the combined network.

CReDo Phase 2 modified the implementation of the knowledge graph to adopt a distributed architecture that integrates remotely hosted data. The distributed architecture enables the possibility of asset owners sharing data about their infrastructure and receiving insights based on the combined infrastructure network, while simultaneously retaining control of their data assets within their own IT systems.

The use of hierarchical ontologies together with the distributed architecture combine to make CReDo scalable and extensible. CReDo can accommodate:

- Additional infrastructure, for example new asset owners and new types of assets, by defining new specialisations of the core ontology and by connecting to a new server hosting the data. This has been demonstrated by the addition of the NHS data in CReDo Phase 2.
- Other climate projection and weather hazard data. This has been demonstrated by the addition of the flood hazard data supplied by Fathom in CReDo Phase 2.
- Multiple views of the infrastructure data. This has been demonstrated by the availability of both system-wide and restricted views of the infrastructure data in CReDo Phase 2. All views report insights based on the combined infrastructure network, while the restricted views do so without revealing confidential data.

The distributed architecture also permits partially restricted views, for example where a subset of asset owners has an agreement that permits data sharing to calculate insights *and* visibility of the of the underlying infrastructure data. It is anticipated that the system-wide view may be useful, for example to regulators such as Ofgem, Ofwat and Ofcom, while the flexibility of the (partially) restricted may remove obstacles to the participation of other asset owners to play a role like Anglian Water, BT and Openreach, and UK Power Networks.

The use of ontologies maximizes outward compatibility. It enables the possibility of mapping data from whatever format is used by asset owners into the CReDo data structure. This is important because it enables compatibility with other data models and standards for information exchange. This allows CReDo to accommodate different asset owners having different data formats. This is critical because digitalisation is journey, and different parties may be in different places on that journey.

The CReDo digital twin has been deployed at STFC in a secure environment within DAFNI. The distributed architecture makes provision for security, access and quality protocols. This has been demonstrated by the development and deployment of a multifactor authentication solution to provide assets owners with secure access to the CReDo on DAFNI.

Future considerations

It is recommended that future developments extend CReDo to include the coverage of other asset owners, other assets and other regions of the country, and to consider other hazards and examine the impact of other types of intervention. The types of intervention might include modelling the impact of increasing flood defences, relocating assets, different configurations of future versions of the networks *etc*. The choice of interventions and what hazards, or combination of hazards, to consider should drive the choice of what data to include. Extreme heat has been identified as an important hazard, while the road network has been identified as an important component in the response to many scenarios, for example is a road passable so that someone might reach an asset? Extending the coverage should include a process of understanding the failure modes of assets, understanding what data are necessary to describe the failure, and developing suitable failure models. This should be accompanied by the development of improved metrics to assess vulnerability, connectivity and risk, and models of interventions to facilitate decision support. Such an extension must include a review of the internal data structure of CReDo to ensure that the ontologies describe assets and the connections between them in the level of detail required to support the models.

The design of the distributed architecture and how data is split between nodes should be reviewed to determine what else might need to be done to address security issues. For example, future developments should seek to distribute the description of connections, and should design-out the need to cache information to support visualisation of connections and the dashboard. It is recommended that the distributed architecture should also be extended to facilitate the development and deployment of agents. This will facilitate the extensibility of CReDo. In particular, the use of object-graph mappers to support the development of agents would be beneficial. Future technical developments should also investigate the issues with Geoserver and should consider alternatives to Superset. The CReDo user interface should be extended to allow better exploration of the connections between assets and the cascade of effects through the combined network. Controls should be added to allow the management of data sets and scenarios.

This second phase of CReDo has developed a distributed architecture that enables the scalable and extensible sharing of models and data across sectors to provide decision support to increase climate resilience. The capabilities have been demonstrated using data shared by the energy, water and telecoms networks to explore the resilience of the combined network to floods. However, CReDo can do much more. Future work should consider other use cases, should consider how to link CReDo to other digital twin and data sharing initiatives, and should engage additional partners, for example, local, regional and national distributors, regulators and agencies.

Appendix

Source code

The code developed by CMCL on behalf of CReDo is held under version control at STFC (https://gitlab.stfc.ac.uk) and will be published under a permissive open-source licence.

The code depends on container images and library functions developed in collaboration with the Computational Modelling Group (https://como.ceb.cam.ac.uk) at the University of Cambridge and the Cambridge Centre for Advanced Research and Education in Singapore (CARES) (https://www.cares.cam.ac.uk). These dependencies are published under a permissive open-source licence and are available via The World Avatar package repositories on GitHub (https://github.com/cambridge-cares/TheWorldAvatar). All other dependencies can be resolved via Maven Central (https://search.maven.org), PyPI (https://pypi.org), and Docker Hub (https://hub.docker.com).

All software used in the final product of this project are licensed under terms that allow their use for commercial purposes with some basic conditions. The relevant terms and conditions can be accessed via the links in the relevant sections below. Simplified summaries of most of these licences can be found at TLDRLegal.

The following additional software were used in CReDo Phase 2, in addition to the software detailed in the report published at the end CReDo Phase 1. [2]

NGINX

Name: NGINX.

Licence: FreeBSD Licence.

Website: https://nginx.org.

Documentation: https://nginx.org/en/docs.

Source code: https://hg.nginx.org/nginx.

Docker Hub page: https://hub.docker.com/_/nginx.

Purpose: Open-source web server and reverse proxy for Windows and Linux platforms.

- **Pros**
    - Widely supported and documented.
    - Simple installation and setup process.
    - Regularly maintained and updated.
- **Cons**
    - No notable issues encountered so far.

Superset

Name: Apache Superset.

Licence: Apache Licence.

Website: https://superset.apache.org.

Documentation: https://superset.apache.org/docs/intro.

Source code: https://github.com/apache/superset.

Docker Hub page: https://hub.docker.com/r/apache/superset.

Purpose: Data exploration and visualization platform.

- **Pros**
  - Easy-to-use interface for exploring and visualizing data.
  - Create and share dashboards with rich data visualisation.
  - Enterprise-ready authentication integrations.
  - Integration with most SQL-speaking RDBMS through SQLAlchemy.
- **Cons**
  - Awkward to use with reverse proxy servers.
  - No analytics support without expert knowledge of SQL.
  - Inadequate permissions system, *e.g.*, unable to disable export of data.
  - Difficult to customise visualisation options, *e.g.*, to control precision of displayed numerical data.
  - Incomplete documentation.

Delivered in partnership with:

Innovate UK

CReDo

Project partners

Acknowledging Phase 1 contributors

Centre for Digital Built Britain

Joint Centre for Excellence In Environmental Intelligence

Department for Business, Energy & Industrial Strategy

National Digital Twin Programme

Cambridge University

Crocodile Media

Mott MacDonald

ESRI UK