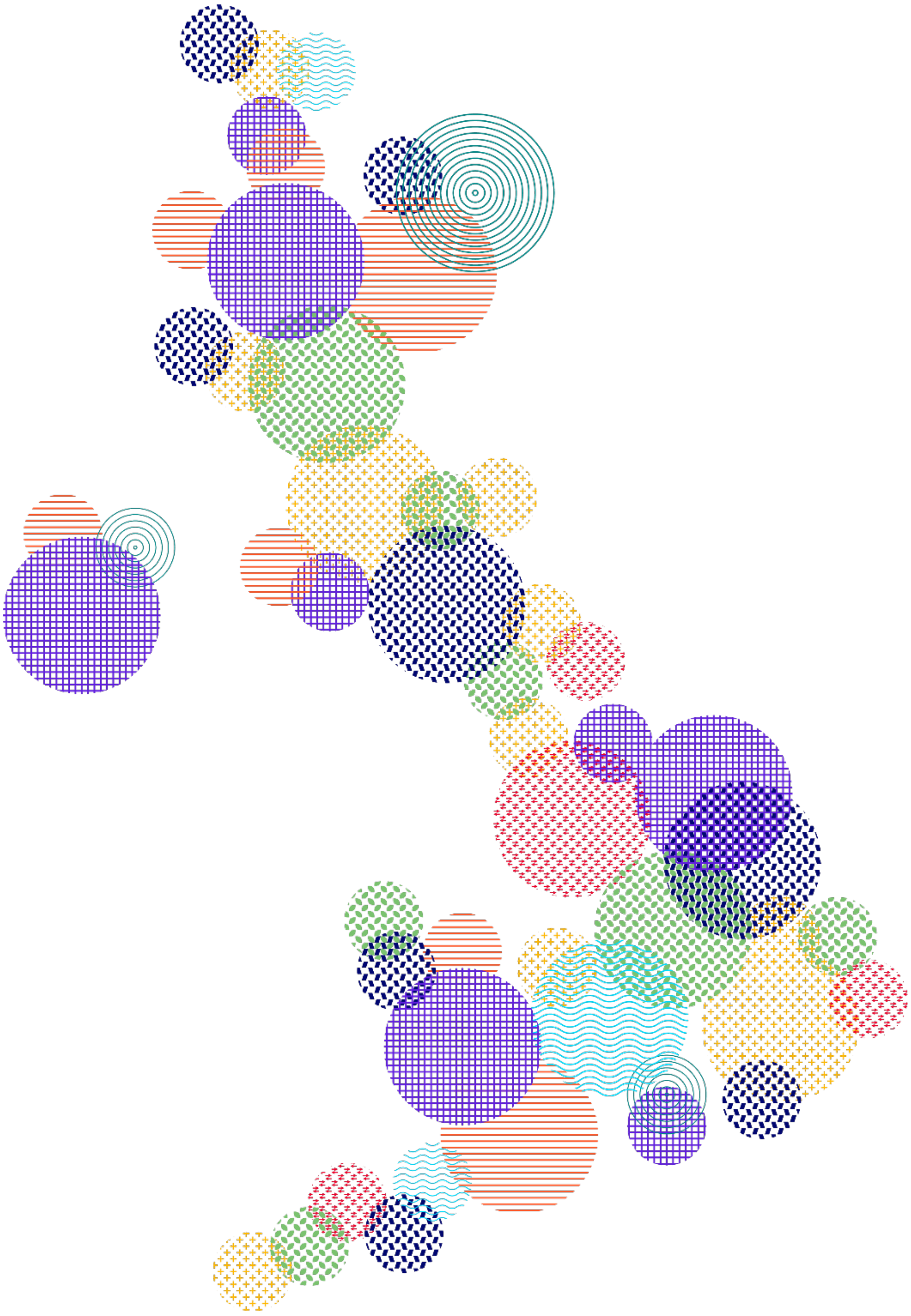
CReDo: an overview

March 2022



Contents

Foreword

Professor Jim W Hall FREng
Professor of Climate and Environmental Risks
University of Oxford

Adapting the infrastructure systems to the increasing impacts of climate change is one of the greatest engineering challenges that we face. Complex systems, like power networks, water supplies and telecommunications systems, need to be made more resilient to foreseeable and unexpected shocks. The use of computer simulation models will be absolutely essential for stress-testing our critical infrastructure systems and prioritizing adaptation interventions.

Computer simulation modelling has been around for a long time, but today it is becoming possible to link computer simulations with data from sensors and satellites, so that computer models are always up to date. This is a first step towards a digital twin, which uses real-time observations, simultaneous computer simulations, and feeds back to the physical world to operate systems more safely and efficiently.

The future of digital twins of infrastructure systems is coming into sight. The Centre for Digital Built Britain has been working from 2017 to advance the digital representation and management of the built environment. The CReDo project is a demonstrator of essential aspects of digital twinning. CReDo has made significant progress with respect to connecting datasets across organisations with models to predict the impact of future flood scenarios on the infrastructure system.

The risks to infrastructure from climate change and from failing to adapt to climate change are very real. Storms Arwen, Barra, Dudley, Eunice and Franklin wreaked havoc across the UK causing power cuts and disruption during 2021-2022, bringing further urgency to the CReDo project as the team embarked upon bringing together datasets across energy, water and communication networks in the East of England.

The CReDo film, *Tomorrow Today,* brings home the personal impact that climate change will have on us all. Climate change isn't something that happens to other people in other countries, it will affect us all and we need to adapt. CReDo is a step along the way in the journey of adaptation.

CReDo also a step along the way in the journey towards better information management and improved ways of sharing data, towards a National Digital Twin as an ecosystem of connected digital twins.

It can be a challenge to encourage and incentivise infrastructure asset owners to share data for public benefit and I am personally delighted as Chair of the DAFNI Governance Board, that

DAFNI has been a major part of this process in creating a secure area to work with asset owner data. Building trust by looking after data appropriately is essential to making progress in using data for the public good. As well as providing a secure space for data sharing, the DAFNI platform provides a convenient computational environment for coupling models, reliably and repeatably.

Reports from Climate Change Committee have highlighted the need to better understand infrastructure interdependencies urgently in order to adapt to climate change. It is just not possible to understand these interdependencies, if we continue a siloed way of thinking. Thinking only about sector resilience rather than system resilience will not provide all of the outcomes that we need. CReDo breaks the mould by encouraging and demonstrating collaboration across sectors. It isn’t just technical progress which is achieved through CReDo but also a change of approach to thinking about system resilience and to industry, academia and government collaborating in working towards greater system resilience to the benefit of us all.

JW Hall

Executive summary

Project Vision and Objectives

The National Digital Twin programme's Climate Resilience Demonstrator (CReDo) is the UK's first climate adaptation digital twin that adopts a novel approach to resilience planning at a systems level, building on the principles of connected digital twins across organisations and sectors. CReDo uses asset information from different infrastructure sectors, as well as climate data, to show how connected digital twins at the systems level can support better long term resilience planning. The use case involves three infrastructure asset owner organisations from the water, power and telecommunication sectors sharing data in a common data environment with the aim to understand the cascading failures across the system under a range of flooding scenarios. This helps to understand asset criticality beyond the boundary of individual sector-level networks and shows the most vulnerable points in the system, which require priority interventions to maintain existing levels of service in the face of climate change. The purpose has been to demonstrate how an Information Management Framework approach to connecting datasets and digital twins can enable connections between digital twins to scale and how connected digital twins can help to tackle climate change. Starting with a specific use case demonstrates how the National Digital Twin can be developed into an ecosystem of connected digital twins, on a case-by-case basis.

Approach and outputs from the first phase of CReDo

Building on the principles of the National Digital Twin programme’s Information Management Framework, asset data from multiple organisations has been integrated within one system model and equipped with a visual interface. This has allowed for a clear representation of the connectivity between assets and an analysis of the resulting interdependencies between sectors as shown in Figure 1

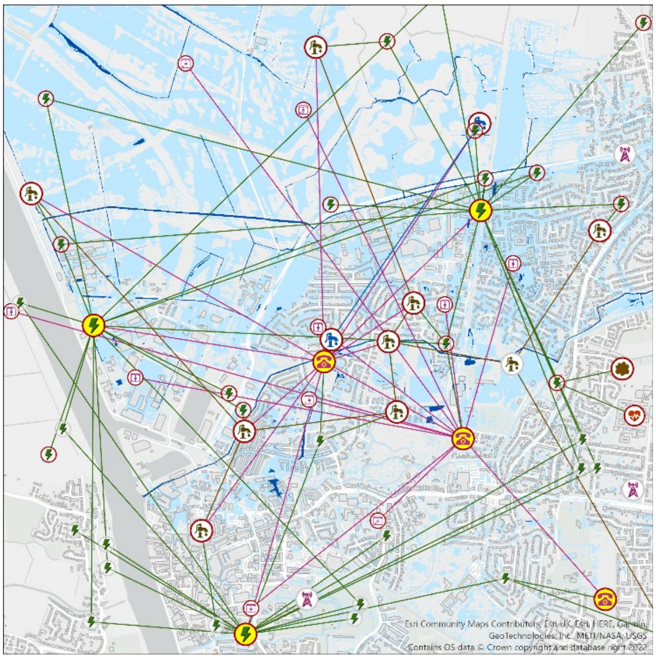


Figure 1 Data about assets is brought together with a flooding scenario (Note this diagram shows synthetic asset data)

Additional flood hazard information was processed, integrated and interrogated within the system model to represent the extent and depth of flooding across the studied area under various future flooding scenarios. This climatic data was used alongside asset information to investigate direct failures resulting from local flooding conditions, and to test the ability of the system to cope under a range of future flood scenarios. In order to build a picture of the wider system impacts, two processes were run in parallel;

- First, information about assets was gathered along with information concerning the likelihood of failure in various flooding scenarios. This was conducted using expert elicitation techniques and Bayesian modelling, assessing the probability of failure of a given asset as a result of local flooding conditions.
- Second, operational research techniques have been employed to better understand the infrastructure interdependencies and to propagate failures resulting from flood conditions across single networks, and further across the entire infrastructure system.

Learnings and Recommendations for future work

While the work in this first phase of CReDo allows for the visualisation of system impacts resulting from direct and cascading failures across the networks, it is based on a contained part of a wider system. Further system complexities and increased granularity at the asset level should be further explored to obtain a richer picture of single asset failures and how these propagate across the wider system considering actual operating conditions. This involves the consideration of existing asset redundancies (e.g. dual power supply in a pumping station) and other resilience measures (e.g. water storage, back-up generation capacity) that would prevent the failure of an asset. Future iterations of CReDo would also benefit from the development of dynamic models as opposed to the current static modelling. This would allow for a representation of asset and system failure and recovery over time, as the flood event propagates and while mitigation measures are put in place over time. More comprehensive modelling would provide a more accurate picture of vulnerabilities and impacts at the system level and would help to prioritise more targeted investment decisions for building resilience.

Understanding the impact of climate change hazards that cause flooding requires a predictive estimation of extreme weather changes as well as modelling the impact of those changes on the total magnitude of flooding. In this first phase of CReDo, this was achieved by adding a climate change allowance to uplift the selected pluvial and fluvial flooding extremes. The UKCP18 local projections for the high carbon emission scenario were used, and this enabled an understanding of changes in the intensity of convective storms. The probabilistic projections of extremes, based on UK Climate Projections (UKCP), also examined the full range of uncertainty in emissions and climate models. Although, in this first phase of development, the impact of climate change on a single extreme flood event has been used to demonstrate the asset failure risk, the long-term objective of CReDo is to expand the hazard information to include a greater number of events and additional hazards (e.g. heat waves) and test the ability of the system to cope under a range of future climate conditions. In addition, future modelling could consider how multiple hazards, such as high winds and lightning alongside flooding and heat, could lead to further cascading failures and could examine different types of failure mechanisms.

It is important for future phases of CReDo to consider how the lessons learnt from this initial phase can be scaled up to bigger and more complex infrastructure systems. The methods used in the first phase should evolve to allow for greater scaling, repeatability and consistency across wider networks of assets. Whilst a level of reproducibility and automation is necessary, future modelling at the systems level should consider how the necessary detail at the asset level (e.g. asset performance characteristics) affect asset failure, and in turn cascading failures. Furthermore, criticality, vulnerability and resilience metrics need to be defined and quantified at the asset and system level. Future phases of CReDo could consider integrating asset information from the transportation sector since site accessibility for all water, power and telecommunication parts of the system was found to be critical to how quickly recovery procedures could begin during a flooding incident.

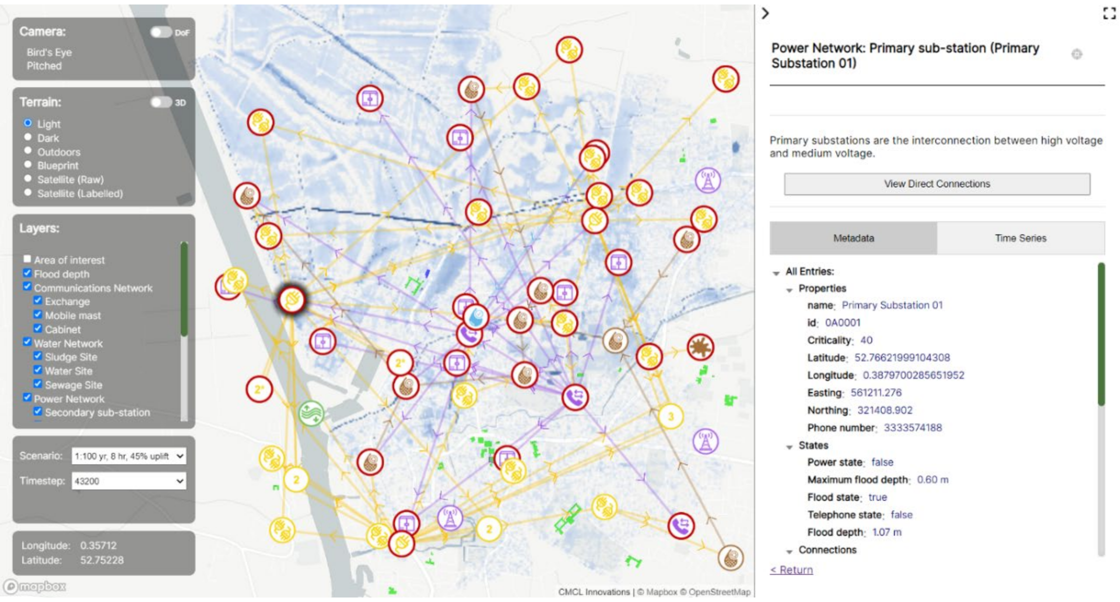
Data quality is another area of improvement for future work phases. Improving quality will require systematic and robust data cleansing processes to accelerate data formatting, structuring and suitability for modelling. The Information Management Framework principles allow for single digital twins and datasets to be connected in a scalable way. The design of the data model in the current CReDo digital twin should evolve towards full compatibility with the Information Management Framework to enhance its interoperability, repeatability and scalability. Furthermore, integrating the connected digital twin with asset owner data IT systems would allow for futureproofing technology applicability, ensuring that the most up to date data is fed into the digital twin, whilst retaining the ability for asset owners to control the data they share.

Due to time constraints, the IMF team were able to work only with Anglian Water to identify patterns in the data to improve the quality of select datasets by cleaning data, enriching and refactoring the semantics in order to raise the level of interoperability in the future. Continuation and expansion of this approach will facilitate data sharing at scale.

The Benefits report [1] includes an illustrative quantification of the potential benefits of CReDo based on a simulation using synthetic data. The Frontier Economics team found that the benefits from CReDo looking at the impact of future surface water flooding scenarios could range from £6m to £13m across East Anglia and from £81m to £186m across the UK over the period to 2050 (in constant prices from 2022-2050). The analysis used synthetic data and does not cover the benefits of other use cases such as extreme heat. Limitations from using synthetic data rather than real data mean that at this stage, it has not been possible to provide robust estimates of the potential benefits that CReDo can generate. The methodology underpinning the results can be adapted and refined to form the basis of a more detailed evaluation of a future version of CReDo. A refined version of this cost-benefit analysis methodology could be integrated in CReDo and would help the users of CReDo to understand the costs and benefits of different strategies to increase resilience of the infrastructure networks. An indicative public return on investment of 23:1 implies that connected digital twins like CReDo help to address a coordination problem. While the benefits accrue across multiple parties, individual actors may lack the incentive to invest alone in systems-based solutions such as CReDo. Therefore, work is required to demonstrate the benefits and to kickstart the coordination of effort to achieve those benefits.

Conclusion

The ambition of CReDo is to allow for better and more cost-effective climate resilience planning for infrastructure assets at the systems level using a connected digital twin approach. Providing a systems level approach to understanding asset vulnerabilities, criticality and service failure when exposed to climatic hazards is a powerful tool that allows different asset owner organisations to plan for resilience in a more dynamic way. Collaboration across sectors is essential to a systems level approach. Translating this data into a visual picture of the system as illustrated in Figure 2, below, also aids understanding of the system level view. This enables asset owners to better communicate their respective investment plans to regulators, funders and other stakeholders involved in resilience planning transparently. Future work on CReDo should build on the current digital twin demonstrator outputs to provide more dynamic modelling of the system vulnerabilities, facilitate important data sharing and expand to include more climate hazards and other cascading failures, as well as explore how different climate adaptation interventions could be tested and supported by built-in cost-benefit analyses.



○ Light
● Dark
● Outdoors
● Blueprint
● Satellite (Raw)
● Satellite (Labelled)

☐ Area of interest
☑ Flood depth
☑ Communications Network
  ☑ Exchange
  ☑ Mobile mast
  ☑ Cabinet
☑ Water Network
  ☑ Sludge Site
  ☑ Water Site
  ☑ Sewage Site
☑ Power Network
  ☑ Secondary sub-station

Scenario: 1:100 yr, 8 hr, 45% uplift
Timestep: 43200

Longitude: 0.35712
Latitude: 52.75228

Power Network: Primary sub-station (Primary Substation 01)

Primary substations are the interconnection between high voltage and medium voltage.

- All Entries:
  - Properties
    - name: Primary Substation 01
    - id: 0A0001
    - Criticality: 40
    - Latitude: 52.76621999104308
    - Longitude: 0.3879700285651952
    - Easting: 561211.276
    - Northing: 321408.902
    - Phone number: 3333574188
  - States
    - Power state: false
    - Maximum flood depth: 0.60 m
    - Flood state: true
    - Telephone state: false
    - Flood depth: 1.07 m
  - Connections

Figure 2 CReDo visual dashboard

1 Introduction to CReDo, Climate REsilience DemOnstrator

This report sets out an overview of the first phase of CReDo, running from April 2021 to March 2022, with a focus on the technical architecture developed to integrate different datasets and models into the connected digital twin.

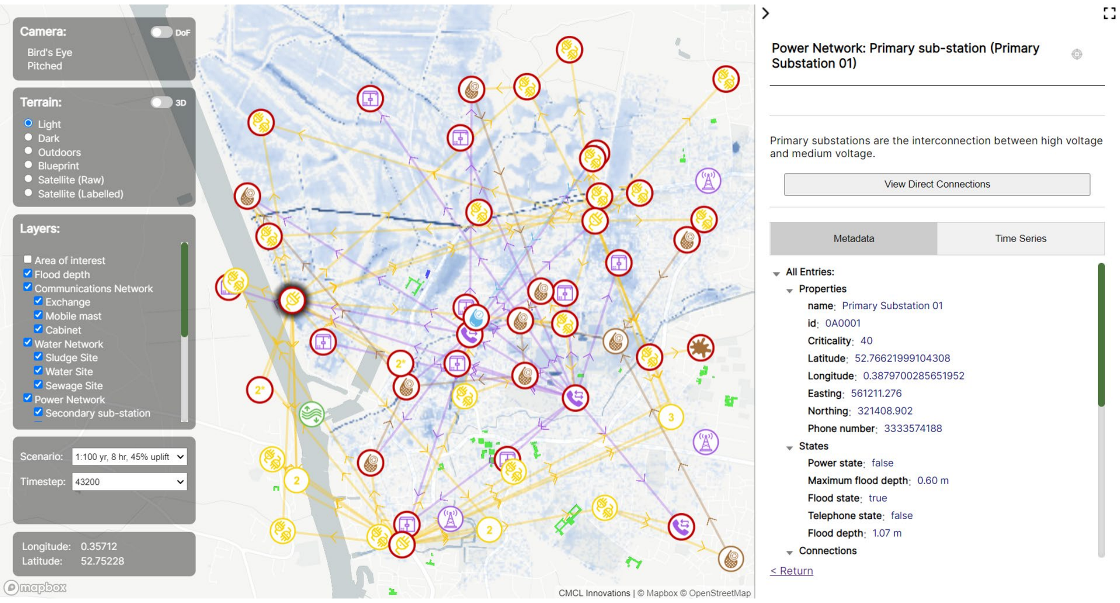
CReDo is a climate adaptation digital twin sponsored by UK Research and Innovation and Connected Places Catapult and is the pilot project for the National Digital Twin programme. CReDo's purpose is two-fold:

1. To demonstrate the benefits of using connected digital twins to increase resilience and enable climate change adaptation and mitigation
2. To demonstrate how principled information management enables digital twins and datasets to be connected in a scalable way as part of the development of the information management framework

Data about assets is brought together across three infrastructure asset owners — Anglian Water, BT and UK Power Networks — into a connected digital twin of the infrastructure system network. Combining data sets from three separate organisations into one system model is not straightforward. Principled information management techniques, such as using the appropriate ontologies and striving for semantic precision, are essential to bringing the data together to present the clearest picture of the infrastructure system without inaccuracies.

Coastal and fluvial flood data has been sourced from the Environment Agency and the HiPIMS (High-Performance Integrated hydrodynamic Modelling System) [2] model has been used to generate surface water flooding data that could be expected under a range of future climate change scenarios. Expert elicitation techniques have been employed to understand the impact of the flood scenarios on asset failure within the infrastructure networks. Operational research techniques have been employed to better understand the infrastructure interdependencies and to identify the propagation of asset failure, both across single networks and across the infrastructure system as a whole, resulting from the flood scenarios. This builds a picture of system impact from flooding scenarios that would not otherwise be available to the individual networks or regulators who would only see the impact of flooding on single networks.

This cross-sector picture in Figure 3 below, demonstrates the impact of extreme weather events on the infrastructure system and can enable asset owners and regulators to better understand infrastructure interdependencies and identify the most effective, least cost and lowest carbon impact interventions to increase resilience. In addition, the incorporation of live data feeds would demonstrate the potential to inform shorter term operational response leading up to and during extreme weather events.



○ Light
● Dark
● Outdoors
● Blueprint
● Satellite (Raw)
● Satellite (Labelled)

☐ Area of interest
☑ Flood depth
☑ Communications Network
  ☑ Exchange
  ☑ Mobile mast
  ☑ Cabinet
☑ Water Network
  ☑ Sludge Site
  ☑ Water Site
  ☑ Sewage Site
☑ Power Network
  ☑ Secondary sub-station

Scenario: 1:100 yr, 8 hr, 45% uplift
Timestep: 43200

Longitude: 0.35712
Latitude: 52.75228

Power Network: Primary sub-station (Primary Substation 01)

Primary substations are the interconnection between high voltage and medium voltage.

All Entries:
  Properties
    name: Primary Substation 01
    id: 0A0001
    Criticality: 40
    Latitude: 52.76621999104308
    Longitude: 0.3879700285651952
    Easting: 561211.276
    Northing: 321408.902
    Phone number: 3333574188
  States
    Power state: false
    Maximum flood depth: 0.60 m
    Flood state: true
    Telephone state: false
    Flood depth: 1.07 m
  Connections

Figure 3: Cross-sector picture of impact of extreme weather events on the infrastructure system

A project like CReDo brings benefits in terms of improved information management and increased system resilience to extreme weather events, and it is necessary to demonstrate and evidence these benefits to encourage investment in information management and climate resilience digital twins, both the people and the technology. The benefits report [2] sets out the benefits of using digital technologies to share data across organisations to improve resilience.

The CReDo interactive app presents the concept of a resilience score to measure the resilience of an infrastructure system. The CReDo film, *Tomorrow Today*, presents a fictional scenario to demonstrate why infrastructure asset owners need to collaborate and share data to increase resilience. Both the film and the app can be found on the DT Hub.

The first phase of CReDo, running from April 2021 to March 2022, has been focused on proof of concept and showing how data can be connected in one secure digital twin. This first phase is going to be taken forward by Connected Places Catapult to explore further the opportunities to use digital twins for anticipating and managing the impact of climate change on infrastructure. For example, future work could explore how possible interventions could be assessed within the digital twin, with a view to creating a deployable tool for supporting decisions to increase the resilience of the overall infrastructure system.

The many elements of CReDo

CReDo has been developed as a socio-technical digital twin project embracing both the human and the technical elements of digital twin development. This start-up phase has been used to explore the following elements of a connected digital twin programme:

- Vision setting
- Communications
- Technical planning
- Data sharing agreements
- Technical development
- Domain-specific factors
- Operational research
- Information management approaches
- Visualisation

Vision setting

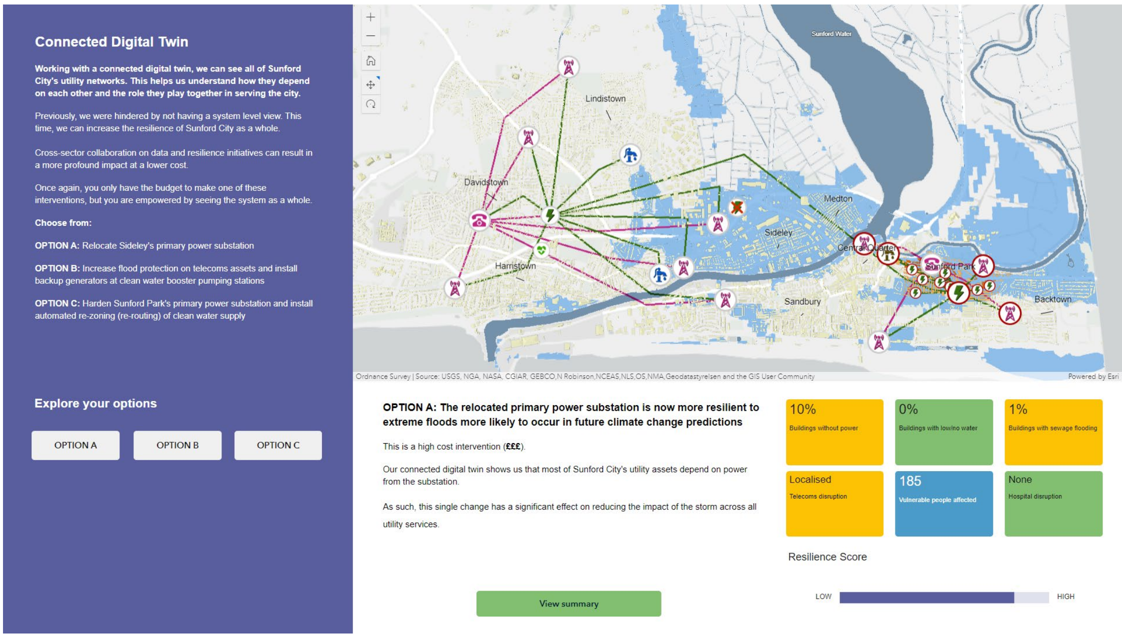
The vision for CReDo is to explore the art of the possible with respect to connected digital twins and to show tangible benefit in the climate resilience space and progress in maturing approaches to information management and in reducing the complexities of scaling connected digital twins. This is a long-term vision of the National Digital Twin programme where many CReDo-type digital twins are joined together as part of an ecosystem of connected digital twins. The intention was to present the concept of connected digital twins as a tool to aid collaboration in tackling climate change at the COP26 Climate Change conference.

The set-up phase

In order to put together a demonstrator project, the National Digital Twin programme approached Anglian Water, BT and UK Power Networks to participate in a regional pilot in the East of England. Through workshops, the parties agreed to focus on a climate resilience use case looking at the impact of future flooding scenarios on infrastructure interdependencies across energy, water and telecoms networks, and at the cascade of failure across the system. An initial proposal was produced to secure funding from UK Research and Innovation. At the same time a Memorandum of Understanding was signed between Anglian Water, BT and UK Power Networks to agree to participate in the demonstrator project and to contribute data should funding be obtained. Following the award of the grant, the initial phase was focused on getting the team in place, setting out a technical plan and getting the legal agreements drafted and signed.

Communications

Whilst the technical plan and the legal agreements were under development, communications were developed explaining what CReDo was with a view to increasing communications about the demonstrator project during COP26. It was decided that in order to explain the complex topic of digital twins to a public audience, the team would commission a film and an interactive app. A film was developed to explain how connected digital twins can have real impact on people’s lives. While the film tells the story of why resilience is so important to public safety and wellbeing in the climate emergency, the CReDo interactive app uses fictionalised data and a fictional series of storms to show how the forthcoming technical demonstrator could work in a decision-making context. Users of the app can make decisions and model the outcomes, viewing a visual dashboard shown in Figure 4, below, powered by a digital twin to help them choose between the available options.



Connected Digital Twin

Working with a connected digital twin, we can see all of Sunford City's utility networks. This helps us understand how they depend on each other and the role they play together in serving the city.

Previously, we were hindered by not having a system level view. This time, we can increase the resilience of Sunford City as a whole.

Cross-sector collaboration on data and resilience initiatives can result in a more profound impact at a lower cost.

Once again, you only have the budget to make one of these interventions, but you are empowered by seeing the system as a whole.

Choose from:

OPTION A: Relocate Sideley's primary power substation

OPTION B: Increase flood protection on telecoms assets and install backup generators at clean water booster pumping stations

OPTION C: Harden Sunford Park's primary power substation and install automated re-zoning (re-routing) of clean water supply

Explore your options

Ordnance Survey | Source: USGS, NGA, NASA, CGIAR, GEBCO,N Robinson,NCEAS,NLS,OS,NMA,Geodatastyrelsen and the GIS User Community

Powered by Esri

OPTION A: The relocated primary power substation is now more resilient to extreme floods more likely to occur in future climate change predictions

This is a high cost intervention (**£££**).

Our connected digital twin shows us that most of Sunford City's utility assets depend on power from the substation.

As such, this single change has a significant effect on reducing the impact of the storm across all utility services.

10% Buildings without power	0% Buildings with low/no water	1% Buildings with sewage flooding
Localised Telecoms disruption	185 Vulnerable people affected	None Hospital disruption

Resilience Score

Figure 4 The CReDo app visual dashboard

The film (shown in Figure 5 below) and the app are available on the DT Hub. Both the film and the app proved to be extremely useful in conveying the concept of CReDo to a wide audience and to gaining support for the programme.

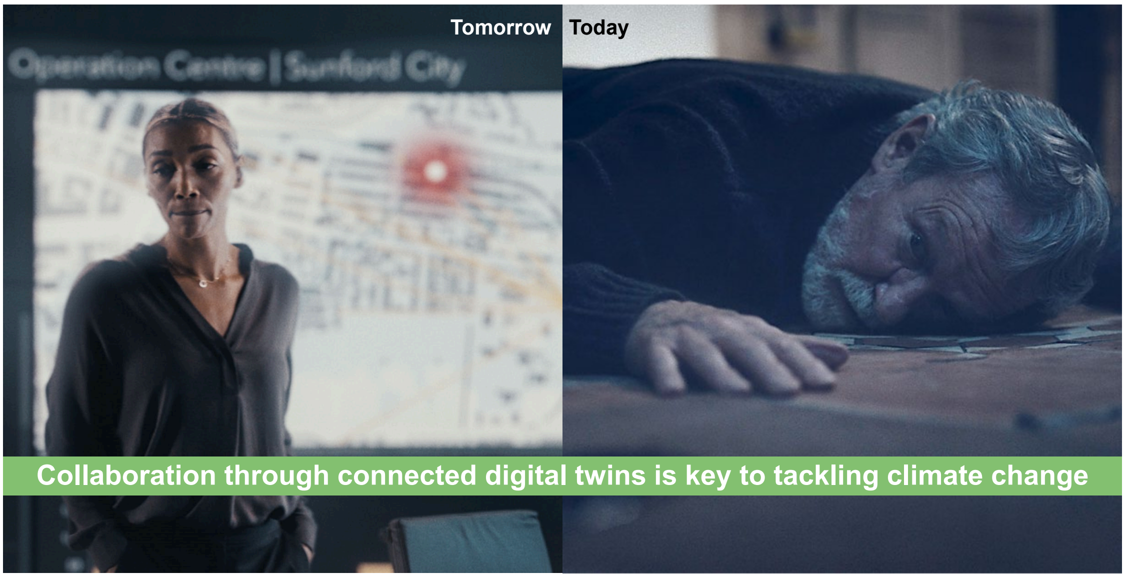


Figure 5 Stills from the CReDo Film, Tomorrow Today

Data sharing agreements

Discussions amongst the parties led to the identification of a data exploration licence as the appropriate form of legal document that could be tailored to the requirements of the CReDo project. A challenge that is often encountered in data sharing projects is how to obtain multi-party agreement to a data licence. Sorting out these matters takes time. In this phase of the project, Cambridge University contracted a range of technical experts to undertake the technical development work as part of a consortium. Cambridge University also contracted with contributors under participation agreements and the full range of contributors is listed on the DT Hub. All participants who were required to work directly with the asset owner data had to sign the data licence, which was agreed between the asset owners, DAFNI[1] as the data hosting platform, and the signatories. Information about the Data Exploration Licence and template, for future projects to use as a guide, are available on the DT Hub.

This overview report focuses on the technical development of the CReDo digital twin and the technical architecture required to piece together the different parts of CReDo. It shares an insight into the nuts and bolts of piecing together both the teams and the data required to build a connected digital twin. It should be read as an overview document in conjunction with the other reports published on the DT Hub that are designed for a technical audience and explore the detail of different technical elements of CReDo:

- Technical Report 1: Building a cross sector digital twin
- Technical Report 2: Generating flood data
- Technical Report 3: Assessing asset failure
- Technical Report 4: Modelling system impact
- The Benefits of CReDo

---

[1] Data and Analytics Facility for National Infrastructure developed by Scientific Computing Department at STFC

2 Technical planning and development

2.1 Approach

This section details the overall technical approach taken by the project team. With complex technical challenges faced by a remote team distributed across multiple organisations, flexible project and risk management was required to meet project deadlines and deliver robust outcomes.

2.1.1 Project partners

Our technical outputs are the culmination of a concerted effort across multiple organisations:

- Anglian Water, BT, and UK Power Networks (UKPN), who provided vital asset data and domain expertise, and are collectively referred to in this report as the Asset Owners;
- University of Cambridge / Centre for Digital Built Britain (CDBB), who provided overall project leadership and project management;
- University of Edinburgh, who provided the project Technical Lead and the team who led the implementation of system impact modelling within networks;
- Newcastle University and University of Warwick, who led expert elicitation processes to model individual asset failure due to flooding;
- Science and Technology Facilities Council (STFC):
    - The Hartree Centre, who provided: a technical architect; a senior data scientist to support asset failure modelling; and research software engineers who supported flood modelling, system-wide impact modelling and general integration of the technology stack;
    - Data and Analytics Facility for National Infrastructure (DAFNI), who provided: secure platform, tools required for multi-factor authentication, storage and compute; capabilities via their public portal; a secure private development environment to enable collaboration on sensitive data across the distributed team; and staff effort for implementing CReDo software on their systems applying systems of systems approaches and integration of workflows on the DAFNI platform;
- CMCL Innovations, who carried out data engineering work, provided the underlying Knowledge Graph technology for our Digital Twin, developed the front-end visualisation tools for end users and supported the integration with DAFNI;
- Connected Places Catapult, who contracted CMCL Innovations through competitive tender and supplied a data scientist to lead data synthesis efforts along with project management resources;
- Mott Macdonald, who provided domain expertise to support flood and asset failure modelling and support throughout all areas of the CReDo project;
- The Joint Centre of Excellence for Environment Intelligence (JCEEI), a collaboration between the Met Office and University of Exeter, who provided climate expertise and designed experimental runs to develop surface water flooding data as a vital input to our connected digital twin.

- The Information Management Framework (IMF) team, who undertook an extreme collaboration with Anglian Water to apply a “clean once, reuse many approach to improving data interoperability, looking at how to reduce the costs and barriers to sharing data in the future.

2.1.2 Project management

With extensive legal discussions required before it was possible to put data sharing agreements in place, technical work had to be accelerated to successfully deliver within a shorter period of project time. Facing multiple unknown quantities, the approach taken was to prioritise shorter term details by:

- quantifying upcoming requirements and plan accordingly where sufficient detail about each task was known to make this approach possible; and
- tracking interdependencies between each required outcome where detailed planning was impossible, maintaining a watching brief on these areas until emergent details provided enough clarity to refine and plan these as above.

With this approach, progressive elaboration of project plans became feasible; by focussing on interim outcomes over the process of getting there, a quasi-agile project methodology was enacted, with time taken to develop and agree outcome specifications (such as the specification of a Minimum Viable Product) and to clarify the explicit purpose the project outcomes must support in key enabling areas.

Visualisation
This was undertaken primarily for evaluation of the data and outcomes of the modelling process rather than as a standalone product for final deployment. We agreed with the Asset Owners that final visualisation integration would be handled by each of their teams in turn, rather than requiring the deployment of our developed prototype. This freed up the visualisation team to work on what was needed for the specific use case at hand.

Synthetic data
This was used primarily to support the dissemination of project outputs without sharing real infrastructure data. Creating this dataset enabled us to ensure that we would be able to both widely share our outputs and maintain Asset Owner confidentiality. Providing a clear approach that solved both of these requirements was a key enabler for the data sharing agreements.

The CReDo technical architect set out task lists for regular review with a focus on highlighting the need for, and supporting, any further scope reviews if the various trajectories looked to be heading off course. This enabled teams to work in parallel on their own areas of interest, and the technical architect played a connecting role to ensure blocking outcomes were prioritised to help other teams continue to progress at pace.

The ambitious nature of the project and the particular requirements for delivery meant that technical upgrades to the DAFNI platform were required to support the final end-to-end model workflow, which meant integration of models and visualisations were moved to the final stages of the project. During the project, the presence of the development environment meant work could continue on real data, and regular demonstrations of CReDo on synthetic data ensured that the outputs would meet the needs of the Asset Owners. Encouraging this feedback throughout the development cycle, and ensuring compatibility with the data that was available, meant that the final results would need minimal changes at the integration stage. Future work building on the DAFNI infrastructure will have the opportunity of iterative development cycles with direct feedback from the end users, and an environment fully prepared for developing these approaches. The time taken to accurately explore the requirements of a platform for hosting CReDo provided key lessons that will be vital to future projects of this nature.

2.1.3 Risk management

With so many moving pieces, both proactive and reactive risk management approaches were used.

Proactive measures included:

- Allowing time for legal discussions to reach conclusion prior to commencement of technical work;
- Creating a film and demonstrator application to promote interest and encourage engagement in shaping the project outcomes, meaning we could build these comments into technical work as it was being developed;
- Adopting a modular approach to the technology stack and using placeholder models to ensure the end-to-end modelling workflow could be completed in parallel with the development of more sophisticated asset failure and system-wide impact models;
- Encouraging the release of initial iterations of each model to ensure deliverables were met (partially or wholly) while providing the opportunity for deeper technical exploration if time permitted[2];
- Encouraging transparent and asynchronous communication across the distributed team through use of a multi-organisational Microsoft Teams environment to share files, discuss modelling approaches/requirements and share updates.

Reactive measures included:

- Floating Research Software Engineer (RSE) effort from the Hartree Centre and DAFNI to support project partners as and where needed, which proved vital when tasks emerged as larger in scope than originally envisioned, such as wrapping an existing open-source flood model into a generic script to run using multiple GPUs on DAFNI;
- Maintaining stretch targets and adapting these where required as the project progressed, such as codifying desired technical work into detailed implementation plans for future phases, ensuring a higher quality plan is available and working towards the achievement of longer-term CReDo objectives;
- Initial weekly stand-up calls scheduled between the technical teams and asset owner staff. These had no fixed agenda and were provided purely to facilitate open discussion during the exploratory phases of technical development. As the project progressed and each technical team's focus narrowed to their specific deliverables the technical teams concentrated their time on delivery.

2.2 Use case

This section details the process of defining a specific use case as the underpinning structure on which the project is based.

To focus development efforts within the project timeline, a single use case was developed and agreed upon with the asset owners, taking data availability and common interest into account.

The project focussed on developing a digital twin, modelling system-wide impacts due to asset failure caused by future flooding scenarios. The Asset Owners were eager to explore the cascading effects of failure across the networks in order to inform a potential future deployable tool that could both aid resilience planning across the infrastructure system and aid emergency response to future extreme weather events caused by climate change. By enabling evaluation of individual network and whole system responses to a range of flood scenarios quickly and easily, the intention is to supplement the information available to resilience and operational planners, enabling them to make decisions both within and across their organisations to maximise resilience of the infrastructure system as a whole. As such, this puts CReDo in the space of a pioneering climate change adaptation digital twin.

Despite working to a single use case, the CReDo team have held future expansion in mind throughout development, and our technical stack has been developed with extensibility as a requisite feature for future iterations and follow-on developments.

2.3 Metadata exploration

This section details the investigative work undertaken to understand the available data to schedule the project accordingly in order to maximise the utility of the demonstrator.

The project sought to include data from a wide range of sources, comprising differing formats and digital availability. In descending order of ease of use for a software application, these sources were:

- Digital, structured data (with varying availability across systems, Asset Owners and geographies);
- paper-based (whether hand-written, printed, image-based or other);
- human knowledge (not captured anywhere but living memory).

The CReDo team held exploratory sessions with technical asset owner staff to identify areas of mutual interest to the communications, power, and water network staff. This included iteratively refining the area of interest to ensure it contained sufficient variety of assets and geographic features for each party. This process required a round-robin approach, whereby the area of interest proposed by a single asset owner was shared with the others in turn, achieving consensus through individual analyses before data sharing agreements were in place.

Whilst each individual asset owner has visibility of their own data and could seek authority to release this from their respective security teams, linking datasets across organisations was far less straight-forward. Given the need to preserve customer privacy, asset owners could not share full datasets on each meter/connection supported by their assets, and thus we took a bottom-up approach:

- Anglian water and BT shared their own site MPANs with UKPN, for UKPN staff to determine which of their assets supplied power to each site
- Anglian Water and UKPN shared their own site landline phone numbers with BT, for BT to determine which of their assets supplied communications to each site

This element of data gathering required co-ordinated effort across each organisation pairing, and highlighted initial errors in the datasets made visible by combining datasets. Whilst each Asset Owner has confidence in their individual data stores, the activity of linking across organisations is not trivial and required considerable manual effort, which was only possible due to the limited geographic scope of our project. Future iterations modelling connectivity across broader regions and/or sectors may be required to support this activity being performed programmatically, which will require innovative approaches and/or application of IMF-like approaches to data preparation. For more information on this element of the project, please see the Information Management Framework project on the DT Hub, and the Anglian Water presentation from the CReDo March 2022 webinar.

This area of the project culminated in the specification of a formal data request, which was added as a schedule to the legal agreements entered by project partners. Only once the full extent of the data request was known could Asset Owner security and legal teams prepare the relevant documentation and specify proportionate security measures.

2.4 Platform provision

This section details the specification and establishment of a common development platform which enabled multiple organisations to work together on private data in a secure and productive manner.

2.4.1 Specification

Discussions between STFC staff and Asset Owners allowed the Asset Owners to communicate any key data storage requirements, such as the presence of multi-factor authentication, and for DAFNI staff to ensure that these could be met. Following this consultation, the Asset Owner data store took the form of a virtual machine (VM) within a DAFNI-exclusive VMWare cluster, which has access restricted to a select number of STFC staff to ensure security. Eight cores providing 64GB of memory were used with an expandable SSD disk which could adapt to project needs, while data at rest on the VM was not encrypted. For accessing the VM, connections were made using SSH and the number of user accounts with access to the data were kept to a minimum, with these permissions reviewed regularly. Once on the VM, code development was enabled through the installation of Podman to encourage container-based approaches.

2.4.2 Data loading

For this initial phase of CReDo, asset data was loaded onto the secure development platform manually. Staff at Anglian Water, BT and UKPN extracted, cleaned, and encrypted data from their respective systems and the encrypted data were then shared with an STFC staff member to either download from private SharePoint sites or to upload from an encrypted physical storage device. Decryption keys were shared orally over calls, rather than sending them in written form.

Note due to the time constraints of technical delivery of this initial phase of the CReDo project, the asset data was cleaned between being extracted from source locations before being shared. The IMF approach recommends improving quality at source as explained further in the Anglian Water presentation.

A read-only copy of the asset data was loaded into a shared volume for project developers to access it from within the developer environment.

2.4.3 Developer environment

To support productive collaborative development, a secure environment was provisioned on DAFNI servers in line with security requirements agreed with the Asset Owners to ensure they were comfortable storing their data among the authorised project partners.

The development environment was established through a virtual machine running the CentOS Linux distribution, with system access protected by multi-factor authentication. De-mobilisation processes were implemented to ensure any staff leaving the project were denied access to the data in a timely and controlled fashion. DAFNI standard Acceptable Use Policy terms were supplemented with additional terms forbidding the download and storage of asset data by any project partners; in practice, there is only so far technical limitations can be managed before procedural controls such as these are required.

A canonical, read-only copy of the asset data was provided within a shared volume for authorised project collaborators to access the data in situ, develop code supporting their models directly on the system and then push copies of their code to a private code repository located on STFC servers. By working in this fashion, developers were able to work with the data without downloading or storing copies on their own machines, and many worked via Integrated Development Environments (IDEs) such as Visual Studio Code, to access the data and system environment in a familiar, low-friction fashion. Others deployed containerised software directly on the developer machine and connected to this remotely, such as using LibreOffice to interactively explore the data in graphical applications via Apache Guacamole.

2.4.4 Runtime environment

Once individual models had been developed and tested using the developer environment, they were exported for use on the runtime environment, the DAFNI portal. A CReDo workgroup was established on DAFNI to enable teams to share data, models and workflows as needed. Further information on the composition of the overall workflow is provided below.

2.5 Modular model architecture

This section details the approach taken to modularise the workflow to enable future upgrades and support different ownership models for ongoing work in future.

With multiple workstreams progressing in parallel, a modular approach was adopted to facilitate integration, decouple workflows and to support present and future ownership of individual components of the final demonstrator deliverables. This is well-suited to the DAFNI platform, as it provides a workflow orchestration service via versioned model containers, which are built and uploaded to the system by authorised users.

To support the parallelised workflow with an ephemeral backing service, the DAFNI portal and underlying architecture was updated to a newer Argo setup involving elements of parallel/backing-service workflows. This was required to meet a core requirement of the demonstrator stack for final deployment, but also serves as a valuable upgrade to DAFNI workflow capabilities for future collaborations.

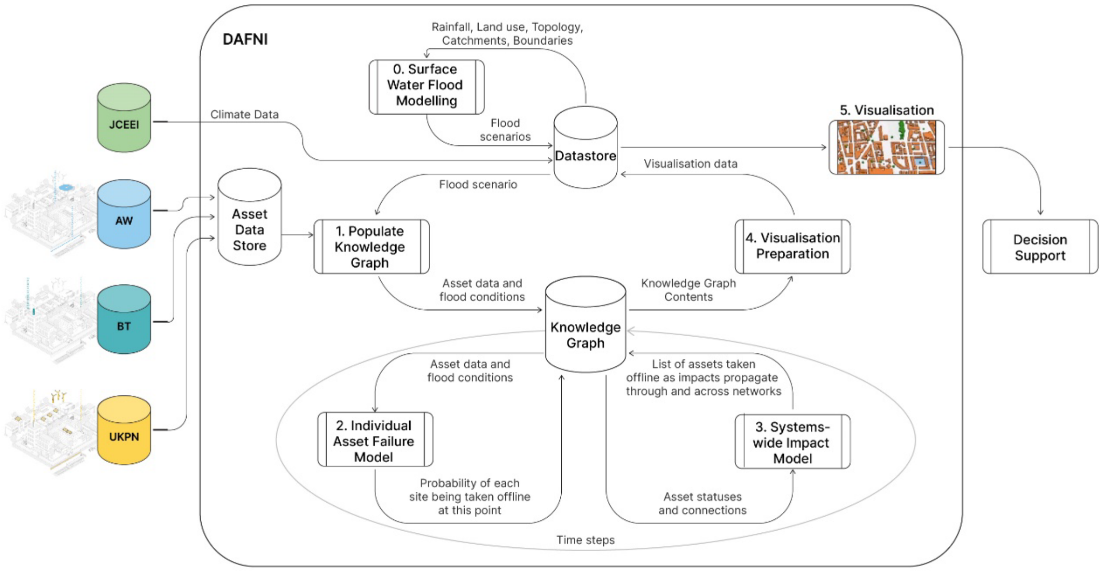


Figure 6 Information flow within the CReDo Connected Digital Twin

Figure 6 provides a high-level illustration of the information flow within the system; this does not include all models involved in the final workflow, which contains additional data processing stages between most of the steps included.

The underlying knowledge graph has the capability to specify which models should be used for different events and/or assets, which gives rise to the ability to implement multiple models (e.g., separate asset failure models for a specific asset type, class, or network). In addition to providing a means of updating each component within the stack iteratively, this capability also enables separate entities to own, upgrade and control their own models as they choose. In this first phase of CReDo, placeholder asset failure and system-wide impact models were initially used for testing, whilst parallel project teams worked on developing new models independently. We found that swapping the placeholder models out for their separately developed replacements to be a significant demonstration of how larger composite systems can be constructed from individually developed components, which is vital for future iterations of the connected digital twin.

2.6 Underlying knowledge graph

This section details the enabling technology used by the digital twin. The knowledge graph developed by CMCL Innovations, with funding from Connected Places Catapult, links the flood and asset data together with models describing the behaviour of the assets. For more information, please see the detailed technical report *Building a cross sector digital twin* and the high-level summary [3].

CMCL Innovations provided a knowledge graph as the underlying information integration mechanism. The knowledge graph used ontologies to represent the data in a way that enabled interoperability between the data and models within the connected digital twin. This brought many benefits, not least of which are the capacity to select specific models to run for specific assets and the ability to cascade changes efficiently through the network of connected assets, which will be true enablers of more diverse and sophisticated model catalogues in future iterations.

In the CReDo use case, the knowledge graph was regarded as ephemeral; a new one was created at the start of a modelling run, populated with asset data and a single flooding scenario, and used to calculate asset failures due to flooding, the failures were then propagated throughout the system connections (for as many time steps as required) and finally the results exported to the DAFNI datastore for later visualisation.

Upon ingestion, asset and flood data are mapped to ontologies developed specifically for this project, with an abstraction layer above that to enable extensibility. Ontologies help with scaling up and knowledge graphs naturally align with the use of ontologies so this approach is ripe for further development into a future “thin slice” of the National Digital Twin through further development and refinement of the ontology over time. Due to time and resource constraints, this was not an area we were able to explore in this initial phase of the project and thus is a recommended for future investigation, both as a key evolution of CReDo and as an exemplar of how existing frameworks in the digital twin community can be linked to the IMF.

2.7 Flood modelling

This section details the modelling work undertaken by the Joint Centre of Excellence in Environmental Intelligence, supported by the Mott Macdonald hydrology team, to provide a range of potential flood scenarios to use as inputs to the final demonstrator. For more information, please see the detailed technical report, *Generating flood data*, and the high-level summary [4].

The use case requires flood scenario data to subsequently model the impacts on individual assets, their immediate networks, and the system in its entirety. Three categories of flooding were considered within the use case at hand:

- Fluvial flooding
- Coastal flooding
- Pluvial/surface water flooding

Fluvial and coastal flood scenarios were prepared using data provided by the Environment Agency, providing maximum depths and maximum flood extents for a given event, but with no temporal aspect, which limited use of the data where the temporal aspect would be crucial (e.g., a 60cm deep flood developing gradually in a single location over the course of 24 hours would entail different mitigation efforts to a flash flood of the same depth arriving in minutes).

Surface water flooding scenarios were developed within the CReDo project team, using UKCP18 data as inputs to a freely licensed, open-source hydrology model known as the High-Performance Integrated hydrodynamic Modelling System (HiPIMS). Multiple flood scenarios were developed in individual model runs on DAFNI and a curated selection was stored in the DAFNI datastore for use within the connected digital twin. By modelling the flood scenarios separately from the asset failure/propagation, we can use and reuse these inputs for multiple network layouts without having to run repeats of the computationally expensive flood models each time. This saves both runtime and energy consumption, and this collection of flood scenarios can continue to be expanded for additional analyses on an ongoing basis.

The overall magnitude of surface water flooding is lower than that of the coastal data, in line with domain expert expectations, but the HiPIMS data are potentially of greater value in this specific use case given the ability to simulate flood depths at arbitrary time steps for a more granular view of how a flood may develop over time within the modelled area. There was particular interest in surface flooding arising from at-risk locations being less obvious than for coastal and fluvial flooding.

With this initial phase of CReDo complete, our DAFNI-compatible HiPIMS model wrapper has been added to DAFNI's public model library for use in future research. This wrapped model is containerised in such a way that it can also run on any platform capable of deploying containers, from a laptop to cloud computing resources.

2.8 Individual asset failure modelling

This section details the modelling work undertaken by Newcastle University and the University of Warwick, with support from the Mott Macdonald domain expert teams, to predict the impact of a given flood scenario on the operational status of individual assets. For more information, please see the detailed technical report *Assessing asset failure* and the high-level summary [5] .

With flood data under development, Newcastle University and the University of Warwick, supported by a Senior Data Scientist at the STFC Hartree Centre, focused on modelling how flood conditions might cause individual assets to fail. The objective of this model is to provide an estimate of the likely operational status of an asset for a given flood scenario, based on both information about the asset within our dataset and specialist domain expertise contained within the minds of asset operators.

To enable this approach, a Bayesian method was adopted, which takes a probabilistic view of how events might unfold using calibrated prior expectations where observable data are not available. Bayesian statistics uses probability distributions to represent all uncertainties, which means that the full apparatus of probability theory is then available as a coherent framework in which to work through the consequences of estimates, assumptions and judgments.

By way of illustrative example, information about an individual water network asset's physical location and dimensions were included within our digital twin, but the impact of a given flood on this asset depends on more than the relative dimensions and locations of the asset and a flood. Backup generators on site may provide a period of resilience in the event of a flood taking the site's mains power offline, yet this contingency plan relies heavily upon the availability of fuel for the generator. Based on historic events, this specific site had a slim but non-zero probability of that fuel being stolen. In the absence of live fuel level data integrated into this initial iteration of our digital twin, our Bayesian approach can include the potential likelihood of this occurring, and hence feed it (among other factors) into the flood impact model for this specific asset.

The benefits of this approach are two-fold; firstly, by eliciting key components of the model structure through discussions with the experts who operate the specific asset in question, our team could create a list of data sources to target for integration within the model to reduce future uncertainty. And secondly, in adopting a Bayesian approach we can provide sensible initial estimates for those likelihoods and subsequently test, refine, and eventually replace them with real observations if/when that data becomes available in future. This supports an iterative development cycle which will be crucial in a changing climate, and models of differing complexity can be developed depending upon the information known about specific asset classes and instances without restricting to solely modelling what is available within the digital twin at any given point in time.

Modelling any single asset in detail requires structured elicitation workshops with the staff who operate the asset in question, and with the limited time available in this initial phase of the project a universal model was developed first, which can be applied to each asset across the system to provide a sensible first estimate of asset failures under flooding conditions. The project team then arranged elicitation workshops with Anglian Water staff to refine the model for a single asset, as covered in the report *Assessing asset failure* [5] [6]The code to construct Bayesian Network models from a description of the individual variables and their respective probability distributions was written in an extensible manner, whereby construction of the model is carried out programmatically, using a single JSON file to define model nodes, states, and probability distributions. This enables future model iterations to be refined and could also be used in conjunction with the knowledge graph's existing capability to specify individual model definitions for each asset/class affected by flooding in a convenient and efficient manner.

2.9 System-wide impact propagation modelling

This section details the modelling work undertaken by the University of Edinburgh to predict the knock-on effects of flooding impacting individual assets. For more information, please see the associated technical report *Modelling system impact* and the high-level summary [6].

Once the direct consequences for individual assets of flooding have been modelled, it is necessary to analyse the system consequences of dependencies between assets (for instance the reliance of communications or water assets on a power supply, or the need for communications to coordinate operation within or between networks). This will generally require calling an external model from the underlying knowledge graph structure that is described in the next section.

CReDo implemented in code propagation of failures through dependence relationships based on that available within the knowledge graph, demonstrating linking to external code from the knowledge graph. Propagation across all three networks was achieved through constructing a single graph structure covering the three networks. There is also a conceptual design for how objectives such as protecting critical assets can be considered where there are decisions to be taken in an emergency situation.

This provides a platform for future work to consider more detailed system-of-systems modelling, such as network reconfiguration to mitigate effects of flooding, prioritising tasks during system restoration, or where greater detail of the engineering and physics of the systems must be accounted for in analysing consequences of failures. Ideally this analysis should be possible through linking together distributed models of the individual networks, rather than the present approach of linking through creating one central data repository and system model.

2.10 CReDo synthetic data

This section details the development and release of synthetic data to support wider dissemination of our outputs without compromising the security of private asset data.

Development of the CReDo demonstrator was made possible thanks to the sharing of national infrastructure data by Anglian Water, BT, and UKPN, but the intention of the programme was to develop an approach that can eventually be applied across all infrastructure networks. To support ongoing development of this and subsequent projects, a synthetic dataset has been developed and publicly released to enable us to show the system in use without exposing sensitive data to a broader audience not authorised for viewing.

To synthesise representative data, the intended purpose of the data had first to be clarified; understanding what the synthetic data is to be used for is key to its production, as it determines what techniques are relevant and the features of the synthetic dataset on which effort should be concentrated. In coordination with the Asset Owners, we defined the intended purpose as follows:

- Enable re-use of our flood data and asset failure models for the area of interest without identifying the location of any real assets, as re-running flood models for some other area of the UK would be time-consuming and provide little additional value.
- Support the narrative of the need for connected digital twins by providing an example of flooding impacts cascading across multiple networks – e.g., maintaining cross-network links such as water/comms sites requiring mains power feeds.
- Help third parties to instantiate a sample set of digital twins for each of the three asset networks and test out integration with their own datasets (e.g., National Highways or local councils), either in self-led research or in workshops/collaborative exploration with project partners.
- Generate screenshots and/or videos from our visualisation tools for promotional use in public dissemination materials.
- To ease ingestion of the synthetic data into our connected digital twin, the same base file structure was used as that of the data shared by the Asset Owners; filenames were preserved, but all real asset data within the file contents were replaced by new, artificial data to represent a similar, but smaller, network. The volume and density of assets included in the synthetic data are far lower than the true data; this supports the development of cleaner graphics for dissemination whilst preserving high-level relationships (primary substations provide power to secondary substations and so on).

The code used to generate the synthetic data will not be released, as this includes comparisons to the real asset data to satisfy concerns around giving away sensitive information. If the reader is interested in developing similar datasets to facilitate the addition of further industries into the connected digital twin, please contact Connected Places Catapult via the Digital Twin Hub.

2.11 Visualisation of results

This section details the user-facing front-end of our demonstrator. A public demonstration of the visualisation toolset is now available on the DT Hub, using the synthetic data described in the previous section.

Finally, the value of a digital twin is dependent upon users being able to extract useful information from it. Therefore, we placed high importance on creating an intuitive and accessible visualisation tool. Even before introducing the modules relating to asset failure and systems analysis, it was found there is a significant benefit to visualising the data originally provided by the Asset Owners. For example, displaying the networks on maps allows users to see the connections present in the dataset and to explore their assumptions around what this would mean for asset reliability, as well as serving as a valuable additional level of data verification.

The visualisation tool that was constructed by CMCL Innovations was designed to be compatible with a series of outputs from running the full digital twin when considering different flooding scenarios. It was predominantly written using JavaScript and utilised mapping libraries for rendering the geographical data. Development also ensured there was sufficient capability to extend the tool into using new datasets describing any of 1D, 2D, or 3D objects. For a full description of the development of this tool, see Section 3.7 of *Building a cross-sector digital twin* [3].

The visualisation software is available in three locations: one publicly accessible version with synthetic data on the DT Hub website; a second containerised version on the DAFNI platform that can connect to alternative datasets using their web browser interface; and a public release of the code on a CReDo Github repository. Each location hosting the visualisation serves a different purpose. The publicly available website explores synthetic data and displays the capabilities to a potential front-end user, without requiring any coding expertise and provides minimal overheads to interacting with the tool. The DAFNI-hosted version is designed to be used with data stored on the platform and provides users the opportunity to run the visualisations with their own data, or to integrate with an alternative analysis pipeline. Publicly hosting the visualisation codebase allows for continued improvement and collaboration of the software into the future. Ensuring that tools developed in this project are reusable and can grow with the digital twin sector is a key goal of the CReDo project.

3 Findings

This section details the findings from our initial phase of the CReDo demonstrator development. Additional findings and recommendations are available within the individual technical reports referenced throughout this document, and the intention of this section is to supplement them with context here, rather than replace or reduce their original form.

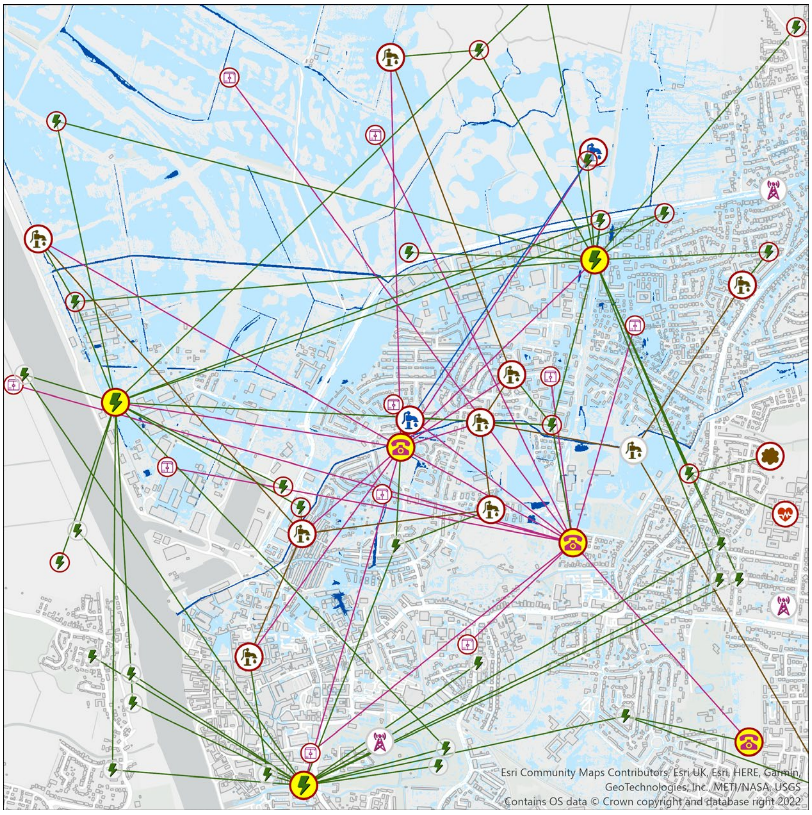
The CReDo project found that it is possible to piece together different datasets across different organisations into one coherent digital twin that conveys a realistic picture of the infrastructure system and the impact of future flooding scenarios caused by climate change. The work undertaken in the first phase of the CReDo project shows what will happen in the future with current networks if we do nothing to adapt. This has been shown with the real data, with synthetic data and in artistic form through the CReDo film. Digital twin projects do not need to be entirely technical in nature, in order to communicate the purpose of the digital twin, such projects need to embrace the human and personal elements of impact.

A key finding of this first phase is that a technical approach alone does not lead to successful delivery. A technical approach needs to be supported through effective communications across a multi-disciplinary team. Adopting an agile approach is beneficial, but still requires a clear technical plan to ensure delivery within set timeframes. In this way, allowing for diversity of approach and interpretation as part of an overall project team is important for harnessing a wide range of capabilities needed to build a connected digital twin where new thinking is required in approaching data sharing, integrating models and developing visualisations. Whilst technology is crucial, people matter most when putting together a connected digital twin.

Whilst this first phase of CReDo has demonstrated that it is possible to connect datasets across multiple organisations in a secure manner, many of the initial findings are in the area of data sharing and digital twin development as below.

Shared data offers insights.

- Shared data can offer immediate insights by visualising a number of datasets in one place. This gives a group of asset owners greater visibility over the whole system, rather than just their own network. When this shared data is presented in a visual format, it is easier to identify key interdependencies across the infrastructure system such as the assets highlighted in yellow in Figure 7, below.



Esri Community Maps Contributors, Esri UK, Esri, HERE, Garmin, GeoTechnologies, Inc., METI/NASA, USGS
Contains OS data © Crown copyright and database right 2022

Figure 7 key interdependencies across the infrastructure system

- This aided timely definition of a targeted area of interest for HiPIMS modelling. It was possible to identify asset interdependencies by plotting the connections between each asset, and infer crucial assets that were highly connected yet looked prone to flooding based on the EA flood data.
- This also aided data quality improvement during the above by highlighting anomalies. Visualisation of the interconnected assets highlighted anomalies in the data, where assets appeared to be connected incorrectly to assets in other systems. This allowed for cross-checking with the source data and for corrections.
- Using shared data to generate synthetic data is useful for working with the data outside of a secure environment and to present in public fora.

Use of climate data in digital twins is not yet straightforward.

- There are immense challenges in the use of climate data for decision support, arising from the complexity of the climate system, the corresponding complexity of any climate model, and the distinction between the real system and the model. Specifically, running current climate models require substantial supercomputer time. Because of this, the higher temporal and spatial resolution UKCP products required for this work provide runs only for a single climate model setup (model assumptions and parameters), and a single greenhouse gas concentration background. Similar restrictions apply to other analogous climate scenario studies.
- In general, for well-founded decision support, one would wish to have access to multiple scenarios or multiple model runs, in order to explore a representative range of possible futures. Because of the scale of climate models, with present computing technology general purpose climate projection products such as UKCP do not contain this exploration of uncertain space in their higher resolution versions, and bespoke model runs to support specific decision questions are generally not possible.

- It is possible to perform bespoke runs for particular storm scenarios of hydrology models such as HIPIMS, over a limited area such as that studied in CReDo, using a computing facility such as DAFNI that would be accessible to many studies. Extending these hydrology model runs to a substantially wider area or longer time series would require specialist high performance computing facilities.

Connected digital twins require multiple assumptions and a challenge is integrating non-digitised assumptions.

- As with any modelling approach, assumptions need to be relied upon in the absence of factual data. There are inevitable boundaries at the edge of simulations (until national and global-scale digital twins are available) that require a number of assumptions. Knowledge graphs can create placeholder entities to represent the assets outside an area that supply power outside an area (e.g. transmission network), but with no information about whether those sites are operational it was necessary to work with some underlying assumptions about their likely states.
- The asset failure work revealed that a substantial part of the key data for building a meaningful digital twin, for instance on failure modes of assets, is not available as numerical data, such as nameplate properties of assets, but may exist in other forms, such as the expertise of relevant organisations' employees.
- Techniques exist for quantifying such other forms of knowledge and integrating this into digital twins, however these rely on specialist skills that few possess even in relevant research disciplines. There is a similar issue of skills base in the scientific basis for coupling digital twin models of different systems together for decision support.

Flexibility in workflows can enable teams to add in further models and complexity as the project develops.

- The CReDo team sought to run software development in parallel across project subteams through using pre-agreed requirements for integration into the DAFNI architecture through designing a modular workflow model. This meant that software development could be led by members with expertise in different modelling approaches and ensure minimal work for integrating into the digital twin during the concluding phase of the project. CReDo has produced a workflow which can incorporate these models, or be adapted to include new models, in a simple and demonstrated fashion. This means that future work can focus on single aspects of the twin without requiring a complete overhaul of CReDo's architecture and provides a foundation for adding partners with new expertise to the project in a non-disruptive way.
- In parallel to this workflow, CMCL Innovations developed an alternative approach to modelling failures known as the information cascade model. This model chose a different approach, incorporating a deterministic flood failure model whilst also having the capability to propagate failures across the network using a depth first search. An alternative version of the twin was valuable as it provided an initial approach to building the CReDo workflow, for demonstrating integration with the knowledge graph, and for connecting to the visualisation tool. The modelling approach used by the information cascade was relevant to the likely outputs of early development of the asset failure and systems impact models, so could provide feedback to this parallel process. While the components of the information cascade model are not as interchangeable as with the other workflow, the current logic inherent in the model is an important initial prediction of network reliability and further work could build on the capability of the information cascade model.

Having both the information cascade and modular workflows meant that each pipeline could be tested against the results of the contrasting workflow to ensure reliability. Results and improvements from one approach could be adapted into the development of the companion pipeline and lessons were shared across teams.

Communicating the benefits is essential to getting a connected digital twin project underway and to ensure continuity.

- In order for asset owners to commit data to a connected digital twin and for their lawyers to agree to a data licence, the benefits of the proposed digital twin project need to be articulated. As the project develops, it is necessary to re-evaluate those benefits and to add to them as the project reveals new unexpected insights.

The Benefits report [1] sets out a logic model that shows how benefits from CReDo accrue across asset owners, customers and society, as set out in Figure 8.

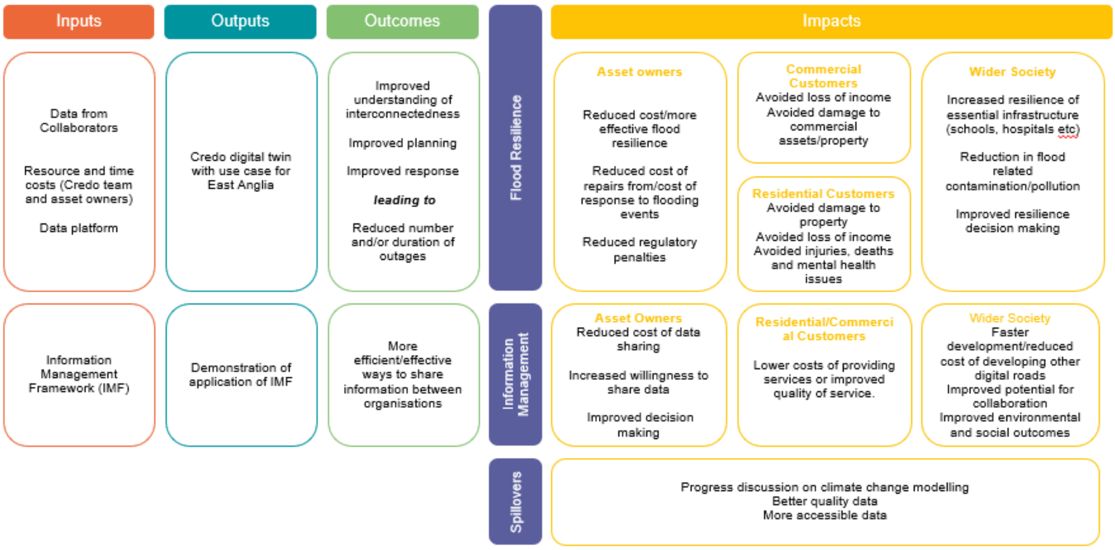


Figure 8 CReDo benefits logic model

The Benefits report includes an illustrative quantification of the potential benefits of CReDo based on a simulation using synthetic data. The Frontier Economics team found that the benefits from CReDo looking at the impact of future surface water flooding scenarios could range from £6m to £13m across East Anglia and from £81m to £186m across the UK over the period to 2050 (in constant prices from 2022-2050). The analysis used synthetic data and does not cover the benefits of other use cases such as extreme heat. Limitations from using synthetic data rather than real data mean that at this stage, it has not been possible to provide robust estimates of the potential benefits that CReDo can generate. An indicative public return on investment of 23:1 implies that connected digital twins like CReDo help to address a coordination problem. While the benefits accrue across multiple parties, individual actors may lack the incentive to invest alone in systems-based solutions such as CReDo. Therefore, work is required to demonstrate the benefits and to kickstart the coordination of effort to achieve those benefits.

4 Recommendations

4.1 Future CReDo development

Connected Places Catapult will be taking CReDo forward from April 2022 onwards. The ultimate ambition is that asset owners and regulators will recognise the opportunity CReDo presents to increase system resilience and save future costs, and will invest in development and deployment. Given the need to coordinate across sectors, it is necessary to continue proof of concept until the business case is clear. During phase 1, CReDo has made progress on technical development and the following recommendations set out further work to build upon the initial technical findings.

4.1.1 Data model and knowledge graph approach

CReDo has demonstrated how data from different asset owners can be joined in a common structure and how it is possible to work towards a greater level of interoperability across organisational datasets. In the project to date there has been a large manual effort in preparation of the data for sharing. This work could be extended with the appropriate security and authorisation processes in place towards direct access to company databases, and to live streaming of data for operational applications. There will be a range of different issues associated with the data requirements for offline analysis and operations: the additional requirements for model as opposed to data interoperability; the likelihood of poor system observability in emergency situations; and the challenge of sharing key information that exists in less structured forms such as expert knowledge.

As these aspects are developed in future phases, where data or computation are hosted on facilities such as DAFNI this should be done in a way that provides a step towards field application. For instance, for operational purposes a possible setup is a persistent knowledge graph that calls external model runs, using a server such as DAFNI if its computational power is needed.

Cleaning data at its source as recommended by the IMF approach is time consuming but reduces the cost and effort in combining data sets so that data can be reused rather than gathered for each specific purpose. A move towards higher quality data applying the IMF approach will minimise the costs of repeated data collection and consistency checking and improve the quality of the data and ultimately the quality of the outcomes from using the data.

The concepts and structure of the ontologies used by the current CReDo digital twin could also be aligned with hierarchical ontologies arising from the Information Management Framework. Thus, the design of the data model for the current CReDo digital twin should evolve towards full compatibility with the Information Management Framework. This will aid interoperability and scalability of CReDo solutions and will also provide a case study of how an existing industry data model can be integrated into the IMF framework and the benefits of the IMF for scalability and reducing friction of data exchange. It will also demonstrate for the first time how common data items emerge from the combination of multiple IMF thin slices.

The knowledge graph approach used in CReDo has proved to be very effective for the current work of joining data from different organisations. Planning for next phases should have particular regard to issues such as how this approach can best be scaled, and how the particular structure of the knowledge graph can best align with calling external models that operate on the data.

CReDo connected the different asset owner data stores together by transferring a copied version of the asset owner databases to an agreed central repository as a one-off event. This method for sharing data is not necessarily desirable for future projects as it does not ensure consistency between asset owner data stores and CReDo, potentially allowing for deviation between the two over time. Approaches to integrating a digital twin with asset owner data systems will ensure the lasting applicability of the technology. This will require communication between the builders of digital twins and for asset owners to prepare their systems for this level of integration through an agreed data model. Exploring this area will improve the feasibility of decentralised data models and allow asset owners to have continuous control over the data that they share while still encouraging collaboration. Other approaches, such as homomorphically encrypted data, allow the potential for data sharing that reduces privacy concerns for the outputs of the digital twin, instead running on an encrypted version of the data and returning an output that only the asset owner would be able to decrypt.

4.1.2 Additional domain detail

The digital twin should be extended to include more granular detail such as components within assets (e.g., individual poles or pylons on power lines, or battery backup to sites), or signal strength around phone pasts; partial failures of assets and the failure mode that indicates why they are not working; and richer detail on how faults propagate and possible mitigation actions such as reconfiguration. This will support the additional applied capabilities discussed here.

Road transport data should be brought into the modelling, as road access is vital for system restoration after failures. This would be sourced from National Highways and local councils as relevant. It will be necessary to investigate whether data on detail of topography and on drainage systems are sufficient to understand the flood risk to roads, and the extent to which this must be supplemented by other sources of information.

4.1.3 Climate and weather

The approach should be expanded to consider other weather hazards, such as heat, wind and lightning. In particular, there is a concern that soil shrinkage due to hotter summers could become a dominant driver of pipe bursts, which currently are more common in winter during freeze/thaw cycles. This example would bring a need for geotechnical analysis of soil conditions and their effect on underground assets.

One specific research need is the calibration of climate model outputs for consideration of multiple weather hazards simultaneously, such as the coincidence of rain with wind or lightning. This is critical for infrastructure resilience and is not as well developed as calibration for single hazards.

Guidance and tools on use of climate data and modelling are required for the resilience community, with opportunities on multiple timelines. In the short term, this might include clear guidance on how present climate data such as UKCP can best be used in decision support within current frameworks, for instance through improved consideration of the uncertainty in present climate projections, or how present planning approaches based on standard design storms can be supplemented by the richer information in UKCP and a capability to run multiple scenarios of realistic weather events.

In the medium term (1-3 years) one area for innovation is how uncertainty in climate projections can be propagated through other modules of a decision support system such as hydrology or infrastructure models, which may themselves be computationally intensive. In the long term (5-10 years) there is need for research on how future generations of climate models using exascale supercomputing facilities can best be used to produce climate data for decision support, either providing richer information in general purpose products such as UKCP, or carrying out bespoke climate model runs for applications of very large scope. These will require methods such as fast-evaluation statistical emulators of the original models to provide the framework for uncertainty management and for computational tractability, and in addition to industry frameworks should sit within general approaches such as the IPCC risk management framework.

4.1.4 Asset failures, criticality and system consequences

The CReDo project has provided a significant original proof of concept in the use of expert elicitation methods for supplementing conventional structured data. It has also demonstrated how external system models can link to the core digital twin to analyse system consequences of asset failures. These need to be extended to consider the combination of failure modes, propagation of consequences across multiple systems, time evolution of the system, and reconfiguration/restoration, that are required in a deployable digital twin used to support decision making.

Development of both operational and planning digital twins used for decision making should consider assessment of criticality of assets, and use of this knowledge in making investment decisions and prioritising assets for restoration of service in an operational emergency.

In the present CReDo project we have had two expert teams working on this from backgrounds of Bayesian statistical analysis and Operational Research; the extension of this area in a subsequent phase should first identify the functionality required, and then the methodology that will be needed to provide this functionality.

As well as work on the underlying science, there is an urgent need for research and innovation on how specialist methods such as expert elicitation and Bayesian analysis can be productionised or automated, i.e. how protocols and software tools can be developed so that a wider range of analysts in application communities can use such methods in support of decision making within their organisations.

4.1.5 Benefits

The benefits analysis simulated the impact of some of the potential benefits of CReDo. The results of the simulation provide a conservative preliminary indication of the order of magnitude of benefits that CReDo could be expected to generate. The methodology underpinning the results can be adapted and refined to form the basis of a more detailed evaluation of a future version of CReDo. A refined version of this cost-benefit analysis methodology could be integrated in CReDo and would help the users of CReDo to understand the costs and benefits of different strategies to increase resilience of the infrastructure networks

The ultimate ambition is to turn CReDo into a deployable tool to support decision making in asset owners, regulators and, government departments to improve resilience across the infrastructure system.

4.2 Recommendations for projects taking a CReDo-type approach

It is hoped that many of the findings and recommendations emerging from the CReDo project will be useful to the wider resilience and digital twin communities for replications of CReDo looking to increase resilience across infrastructure networks and similar projects. These are recommendations for developing connected twins:

- Successful development of connected digital twins requires the involvement of multi-disciplinary teams. Connected digital twins cannot be developed by data scientists or domain specialists alone. The CReDo project team comprised the following skillset as set out in Figure 9. Collaboration across disciplines is essential to achieving a connected digital twin that will fulfil its purpose.

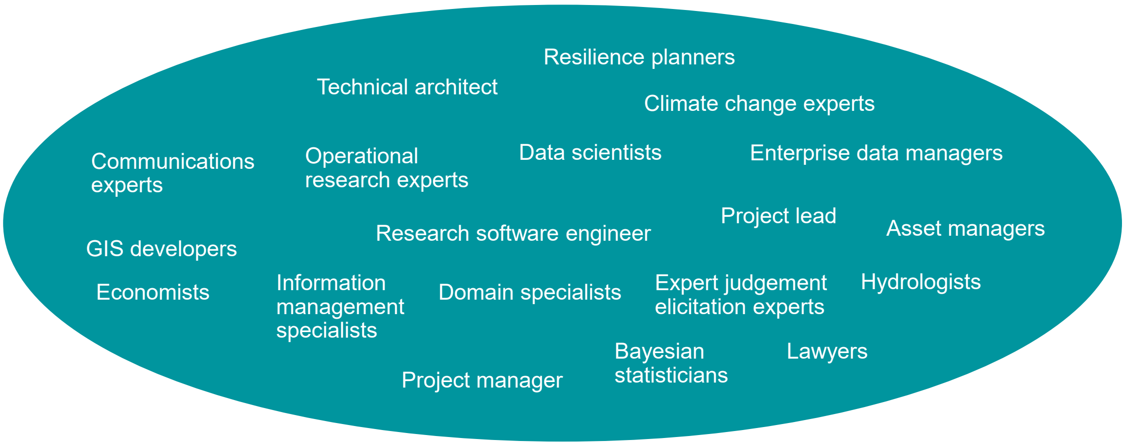


Figure 9 CReDo project team skillset

- A discovery phase is essential when the desired outcome and path of technical development is not clear. The vision needs to be set and a use case clearly defined during the discovery phase allowing for the involvement of representatives from all the relevant teams. As the Gemini Principles set, out a digital twin needs to have a clear purpose. A discovery phase needs to allow for technical freedom and exploration to identify what work is to be done and how it is to be done and what data is available before that project plan is translated into legal agreements. To support a discovery phase, it is recommended that parties agree to a non-disclosure agreement or an exploratory licence before putting the full legal agreements in place to share the data and enable full technical work to take place. These initial legal agreements will allow for the articulation of a clear project purpose, scope and plan to be communicated to the lawyers to inform the development of the necessary legal agreements.
- It is important to curate a data and model register from the outset to identify gaps and move to close them quickly or rescope the use case
- Ontologies help with scaling up and knowledge graphs can support ontologies. Ontologies and the corresponding knowledge graph played a key role in enabling the interoperability between data from the asset owners, and are anticipated to be important for the ability to scale up the approach.
- When working to tight technical deadlines, it is important to keep a multi-disciplinary team motivated on the purpose of the digital twin. Public engagement events like webinars and public reporting help to maintain momentum and unite the team in collaborating on a common purpose – using data for the public good.

5 Nomenclature

CDBB	Centre For Digital Built Britain
CPC	Connected Places Catapult
CReDo	Climate Resilience Demonstrator
CSV	Comma Separated Values
DAFNI	Data & Analytics Facility for National Infrastructure
EA	Environment Agency
HiPIMS	High-Performance Integrated Hydrodynamic Modelling System
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
IPCC	Intergovernmental Panel on Climate Change
IMF	Information Management Framework
NID	National Infrastructure Database
SSH	Secure Shell Protocol
UKCP	UK Climate Projection

6 References

[1] University of Loughborough, “HEMLab,” 2020. [Online]. Available: https://github.com/HEMLab.

[2] Frontier Economics, “Evaluating the impact of CReDo,” 2022. [Online]. Available: digitaltwinhub.co.uk/projects/credo/benefits.

[3] J. Akroyd, “CReDo Technical Report 1: Building a cross-sector digital twin.,” Centre for Digital built Britain, Cambridge, 2022.

[4] D. Ocio, G. Shaddick, J. M. Salter and E. Vanvyve, “CReDo Technical Report 2: Generating Flood,” Center for Digital Built Britain, Cambridge, 2022.

[5] C. Dent, B. Mawdsley, J. Smith and K. Wilson, “CReDo Technical Report 3: Assessing asset,” Centre for Digital Built Britain, Cambridge, 2022.

[6] L. Schewe, “CReDo Technical Report 4: Modelling system impact,” Centre for Digital Built Britain, Cambridge, 2022.

7 Authors and Contributors

Sarah Hayes

Chris Dent

Ben Mawdsley

Tom Collingwood

The Centre for Digital Built Britain at the University of Cambridge's National Digital Twin programme is funded by the Department for Business, Energy and Industrial Strategy via UK Research and Innovation.

Hayes, S; Dent, C; Mawdsley, B; Collingwood, T (2022). CReDo: an overview
https://doi.org/10.17863/CAM.82908

The Centre for Digital Built Britain at the University of Cambridge's National Digital Twin Programme is funded by the Department for Business, Energy and Industrial Strategy via UK Research and Innovation.