National Oceanography Centre

An Information Management Framework for Environmental Digital Twins (IMFe)

Led by NOC and co-funded by NERC and the Met Office

Natural Environment Research Council

Met Office

April 2022

J. Siddorn[1], G. Blair[2], D. Boot[1], J. Buck[1], A. Kingdon[3], A. Kloker[1], A. Kokkinaki[1], G. Moncoiffe[1], E. Blyth[2], M. Fry[2], R. Heaven[3], E. Lewis[4], B. Marchant[3], B. Napier[3], J. Passmore[3], S. Pepler[4], P. Townsend[4], J. Watkins[2] and K. Winfield[4]

British Geological Survey

UK Centre for Ecology & Hydrology

Science and Technology Facilities Council

[1] NOC, [2] UKCEH, [3] BGS, [4] STFC

Acknowledgements

This work has been jointly commissioned by the Natural Environment Research Council (UKRI) and the Met Office under the project: *Towards An Information Management Framework for Environmental Digital Twins* (Grant Ref: 2021DTIMF1Siddorn).

We would like to thank the many contributors to this work. The case studies would not have been possible without the engagement of Aaron Hopkinson and Kirstine Dale from the Met Office, Scott Hosking, James Byrne and Maria Fox from the British Antarctic Survey and Ian Bailey and Steve Kochli from Telicent.

We would also like to thank Alex Luck of the Department for Business, Energy & Industrial Strategy (BEIS), Matthew West and Peter el Hajj from the Centre for Digital Built Britain (CDBB) and representatives of the European Commission's Destination Earth programme (Peter Bauer, Nils Wedi and Irina Sandu of ECMWF) who helped set the wider context.

Additional thanks go to Aaron Hopkinson for his particularly active and helpful engagement and his contribution reviewing the report.

A roadmap for developing the Information Management Framework for Environmental Digital Twins

Executive summary

Environmental science is primarily concerned with assessing the impacts of changing environmental conditions on the state of the natural world, whether affected by natural variability or by the impact of human activity. The Natural Environment Research Council (NERC) has recently published its digital strategy[1] , the first of its kind for NERC, which sets out a vision for digitally enabled environmental science for the next decade. This is echoed in the Met Office's Research and Innovation Strategy that includes the vision of transforming the weather and climate research and services through deploying transformative technologies such as Digital Twins[2]. This strategy places data and digital technologies at the heart of UK environmental science. One such set of technologies are digital twins.

A digital twin is a virtual representation of an object or system (for example the natural environment) updated as the system changes using observations. Observations may come from a range of sources, some traditionally used in the environmental science community such as satellite remote sensing or sensors on ships or weather stations, or through the emergence of sensors on everything from fridges to cars to large-scale built infrastructure. A digital twin then uses simulations or data-based methods such as machine learning to generate a replica ('twin') of the system that can be used to understand the system itself. Environmental digital twins therefore have the potential to significantly improve our understanding of the natural environment.

The emergence of increasingly large, diverse, observed data sources and the development of digital twin technologies combined provides an opportunity for the environmental science community to make a step-change in our understanding of the environment. But to realise the value of environmental digital twins they need to be developed following agreed standards to make sure the information can be trusted by the user, and so that data from twins can be shared, both between environmental digital twins and with other types of digital infrastructure.

To enable this, an information management framework (IMF) is needed that establishes the components for effective information management within and across the digital twin ecosystem. It must enable secure, resilient interoperability of data, and is a reference point to facilitate data use in line with security, legal, commercial, privacy and other relevant concerns. Previous work has highlighted the importance of developing an IMF, including the Centre for Digital Built Britain (CDBB) roadmap to an IMF (CDBB, 2020).

This roadmap follows the CDBB approaches and develops it further to outline the steps needed to develop an IMF that meets the demanding requirements of the environmental domain (an IMFe) whilst also ensuring interoperability with other digital twins.

1 www.ukri.org/publications/natural-environment-research-council-nerc-digital-strategy-2021-2030/

2 www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/research/approach/r-i_strategy_full_version_v2.pdf

Timeline

The conceptual sequencing of the IMFe development is shown below. The timeline runs from left to right, but no actual timescales are prescribed given the need to put together a coherent and funded programme of work to implement this Roadmap, without which any timescales are arbitrary. The agile approach allows parallel activities, although the development of the vision, a development framework and a management and governance framework naturally come early in the process.

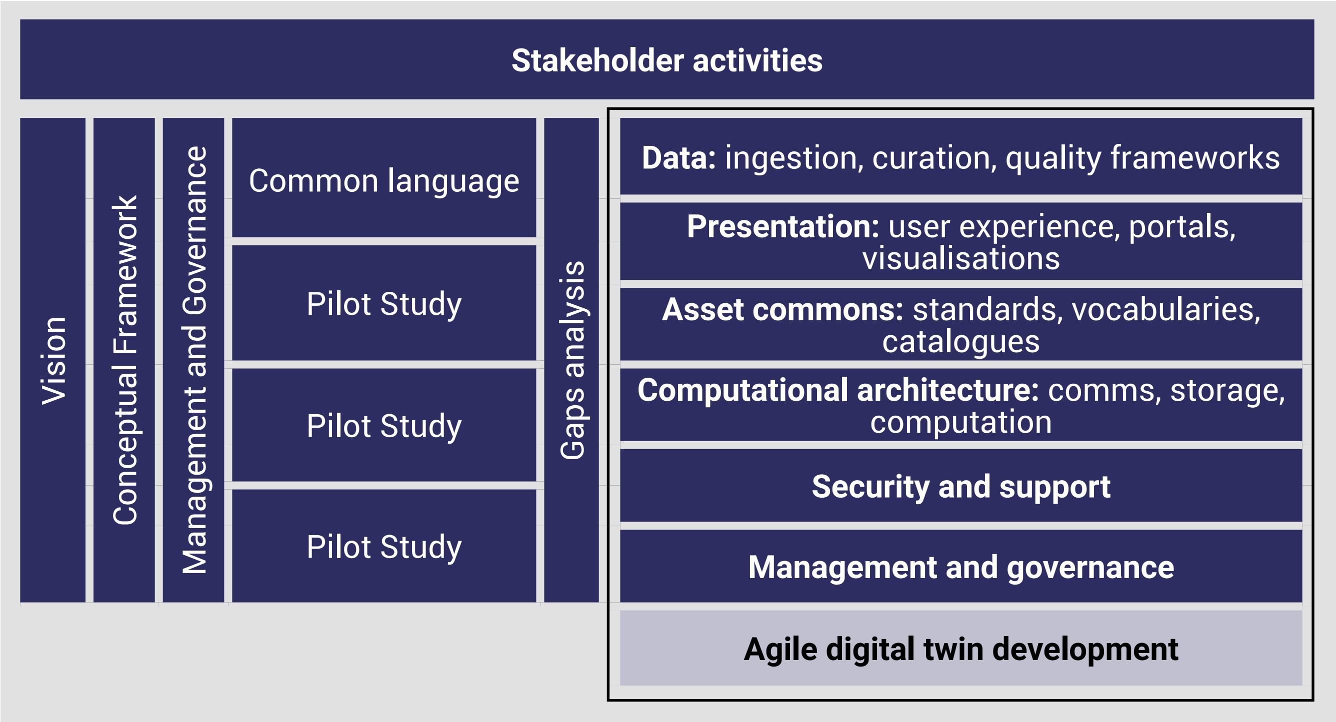

Figure 1: Schematic of the sequential activities required to develop the IMFe, running from left to right in time. Boxes do not represent relative sizes of the activity.

Introduction

The bringing together of observations of the real world with digital capacities is increasingly recognised as a powerful tool for understanding and monitoring our environment and the effects of the environment on people and our infrastructure. Cyber-physical systems, as they are often referred to, are already well-established in the engineering domain. They are also well established, although in a different guise, in the forecasting and prediction systems that we have, for many years, been developing in the environmental community for weather and hazard prediction and climate understanding.

The concept of cyber-physical systems is therefore not new, but what is perhaps new is the cross-over of concepts and ideas from engineering into the environmental sciences. It is therefore important to understand what is meant by cyber-physical systems (and the subset which are termed digital twins), what they can realistically achieve, and how they may drive developments in the environmental domain.

The vision for Cyber-Physical Infrastructure (CPI) is developed in the Robotics Growth Partnership (RGP) vision for cyber-physical infrastructure (CPI)[3] and reflected in the BEIS consultation on "Enabling a National Cyber-Physical Infrastructure to Catalyse Innovation"[4] which will provide the basis for the definition of digital twins and cyber-physical infrastructure used here.

There are two fundamental, perhaps existential, questions around digital twins that often create barriers to discussing their form and benefits. The first is how, and if, they differ from **models**. And the second is whether they have been hyped to such an extent that the reality of what they can deliver will inevitably lead to disappointment. We will look to address both questions here.

Unlike a traditional model, cyber-physical infrastructure is **federated**; in much the same way as the internet creates a world-wide web, a cyber-physical infrastructure is designed with interoperability and federation at its heart. In the vision of the CPI there is no one monolithic infrastructure, but myriads of component parts sharing and communicating to allow a powerful interconnectedness of information. The components of a cyber-physical infrastructure are **modular, reusable** and **networked**. This puts increasing emphasis on **common standards, semantics (languages)** and **interfaces** to allow communities to co-develop and link their systems. It is the roadmap for the framework for these standards, semantics and interfaces that is developed in this report.

Digital twins are one form of cyber-physical infrastructure. They have been widely applied in the engineering realm for tasks such as engine optimization and port management. Some of the different systems sometimes referred to as 'digital twins' are shown in figure 2.

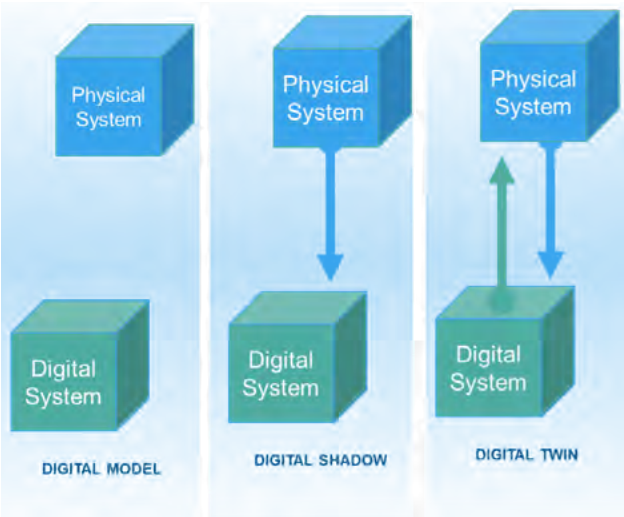

Figure 2: Schematics of different forms of Cyber-Physical Infrastructure, including Digital twins[5].

3 *The Cyber-Physical Infrastructure - Empowering innovation, people, robots and smart machines to enhance prosperity, resilience, sustainability and security (publishing.service.gov.uk)*

4 *Enabling a national cyber-physical infrastructure to catalyse innovation (publishing.service.gov.uk)*

5 *following Enabling a national cyber-physical infrastructure to catalyse innovation (publishing.service.gov.uk)*

Digital Model: an informative digital representation of a system which may use data techniques and / or process knowledge to generate the representation. Although it may be informed by observations, there is no automated continuous updating of the digital system because of changes in the real world. An example may be climate simulation or observation derived gridded climatologies.

Digital Shadow: a digital model that integrates information from its physical counterpart. This may be in real- or right-time. These are often used in systems without autonomous or direct feedback from digital to the physical twin. The most obvious example in the environmental domain is weather forecasting, where the physical twin information is assimilated into the model in real-time, but there is no corresponding feedback.

Digital Twin: a digital model with (real- or right-time) two-way information flows. This can enable autonomous optimisation, and remote and autonomous operation. These are not yet widespread, but are typically found in autonomously operated environments such as robotics and autonomous vehicles and maritime vessels.

Although only systems with two-way feedback as described above would by some definitions be considered a digital twin, the underpinning infrastructure and standards required to develop federated information sharing for all the above are likely to be common. The key distinction between the CPI instantiations of a digital model, or of a digital shadow, and traditional model representations of the world are in the federation between the systems and the interfaces that allow information to be passed into or out of the system seamlessly. Moreover, it is the federation and interoperability requirements of moving information between systems that drive the need for a focus on an **information management framework**.

Therefore, for the purposes of this report the distinctions above are not significant and it is perhaps more helpful to think about connected digital representations of the environment of varying levels of complexity and feedbacks that constitute a **federated cyber-physical infrastructure ecosystem**. Although the developments of the environmental Information Management Framework (IMFe) will allow for the most demanding of the categories of digital twin with federation and two-way, real-time exchange with the real twin, they will be relevant and useable in all types and will provide value across the spectrum of digital representations of the world we live in. If, for example, a global coupled earth system model run in real-time assimilating observations from around the world (a digital shadow) is developed in such a way as to make its outputs fully accessible and with appropriate data standards and governance to allow seamless connection for another (approved) group to generate a climatology of urban wind speed extremes (a digital model) then we will have seen significant value in having a federated system of digital capabilities. The shadow may also be connected to provide real-time gridded probabilistic, forecast information to aquaculture that adapts based on that (and other) data (a digital twin), with appropriate, openly available and quality assured tools to allow downscaling (also used in the climatology just described). And so on, adding value and increasing efficiency and effectiveness.

In summary, the paradigm shift of having an ecosystem of federated representations of the environment has massive potential to add value to what presently already exists.

Components of a digital twin

To realise the potential of digital twins there is the need to understand what underpinning infrastructures and capacities are needed for a digital twin.

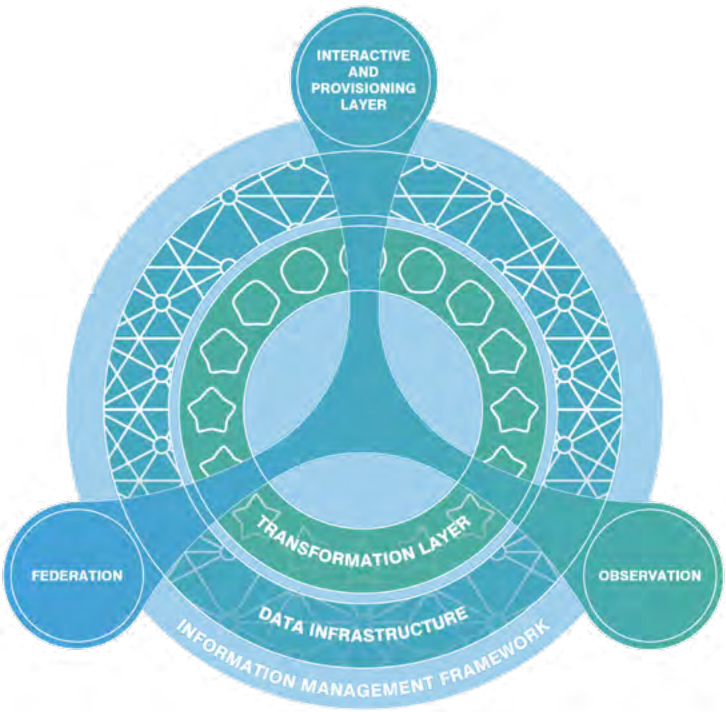

Figure 3: Schematic of the components of a digital twin and the role of the IMFe in supporting the flow of data and information from and to the user.

An observing system

Not normally considered a part of the digital twin itself, but it is an underpinning requirement for digital twins to have access to information about the asset it is representing. Often the concept of internet of things, with ubiquitous sensing of the asset, is invoked here although it is challenging in the environmental domain to have the density of observations at the range of scales required to fulfil the needs of many digital twins. The Information Management Framework must provide the specifications that allow the digital twin to interface with the required observing networks.

Federation

Digital twins are normally expected to obtain data and information from other twins or cyber-physical infrastructures. As with the link to observations, the Information Management Framework must provide the specifications that allow the digital twin to interface with other data and information sources.

Data infrastructure

Data infrastructure provides access to inputs (often, but not always, observations) through data communication channels as well as the computing capacity to perform simulation, emulation or other data transformations. It also provides the compute infrastructures to support user access to the information within the twin and to store data and information as required.

A transformation layer

Within a digital twin there will be capability to transform data to maximise the understanding and value from environmental observations. This transformation may be through predictive modelling, emulation, or other application of artificial intelligence / machine learning. Digital twins often incorporate the additional capacity for the user to modify the prediction engine to explore options, scenarios, and consequences.

An interactive and provisioning layer

The connection between the digital twin and the real environment requires a well-formulated interface between the digital twin, environmental data, and the user. User interaction is therefore an essential function that is embedded in the design of digital twins, including visualisation, user-driven data transformation and data-science tools.

This layer provides a powerful interface to the information and tools in the data engine that is easy to adapt and use, and represents one of the characteristic features of digital twinning. These provisioning layers are tailored towards human users and are often visually pleasing front-end interfaces with easy, intuitive access. Machine-to-machine provisioning is also common, in which case the provisioning layer is tailored to enable federating with other twins or systems.

An information management framework

The CDBB define an IMF as "a built environment information management landscape and data environment which will adequately define parties, processes, information, technology to support the National Digital Twin"[6]. Fundamentally this description is also a good starting point for environmental digital twins, although due to the fundamentally trans-boundary and collaborative nature of many environmental applications the geographic and geopolitical scope needs to be broader.

In the context of the environment, therefore, and noting that the scope of the IMFe encompasses the range of cyber-physical infrastructures, including digital twins, discussed above, this definition of an IMFe can be reframed as:

> ***"an environmental information management landscape** and data environment which will **adequately define parties, processes, information and technology** to support the **development and use** of federated environmental cyber-physical infrastructure, including digital twins, for national digital twin programmes and recognising international engagement"*

An IMFe produces the building blocks for effective information management throughout its lifecycle. It enables secure, resilient interoperability of data. It is a reference point to facilitate data use in line with security, legal, commercial, privacy and other relevant concerns.

As a digital twin requires information and data to be compatible across the built and natural environment, information from a digital twin must be presented in consistent formats to allow for sharing and integration between different digital twins. It also requires curation and mapping of existing and future models and data.

The IMF will define how to produce a **catalogue of environmental digital twins** to ensure that there is discoverability between twins of diverse backgrounds developed by a range of actors. This will ensure digital twins can discover what information is available from other twins and how it may be accessed.

The IMF will also define the **standards that allow twins to interact**. Interoperability is fundamentally about interfaces and ensuring they are accessible and described using appropriate standards and with mappings between ontologies and vocabularies that allow digital twins both within the environmental domain and more broadly to interact with each other. Having a common set of standards, ontologies or vocabularies is unrealistic across even the environmental domain, so the key is defining how to map across at the interface between twins to ensure information provenance is maintained.

And finally, the IMFe will define **the components required to develop and operate** a digital twin in a way that allows data and information to flow into, through (with multiple diverse transformations) and then out again with clarity around the data provenance, data quality and security and intellectual property maintained.

6 https://digitaltwinhub.co.uk/glossary/

Scope

The above defines the components of a digital twin, one of which is the Information Management Framework which governs its development and use. The scope of this roadmap can be described by using the conceptual description of an IMF (shown in figure 4) which is a combination of a technical structure (the orange core) and a structure based on management theories (the blue and green outer circles).

A complete set of definitions or common standards that describe all components within all digital twins of the environment is not possible. The ambition of this framework is therefore to define where commonality is needed and identify best practice. An IMFe is particularly important in defining interfaces to observation infrastructures, to other digital twins and to users. It also plays a critical role in ensuring that, as data are passed through the system, important information about that data is retained and appropriately adapted.

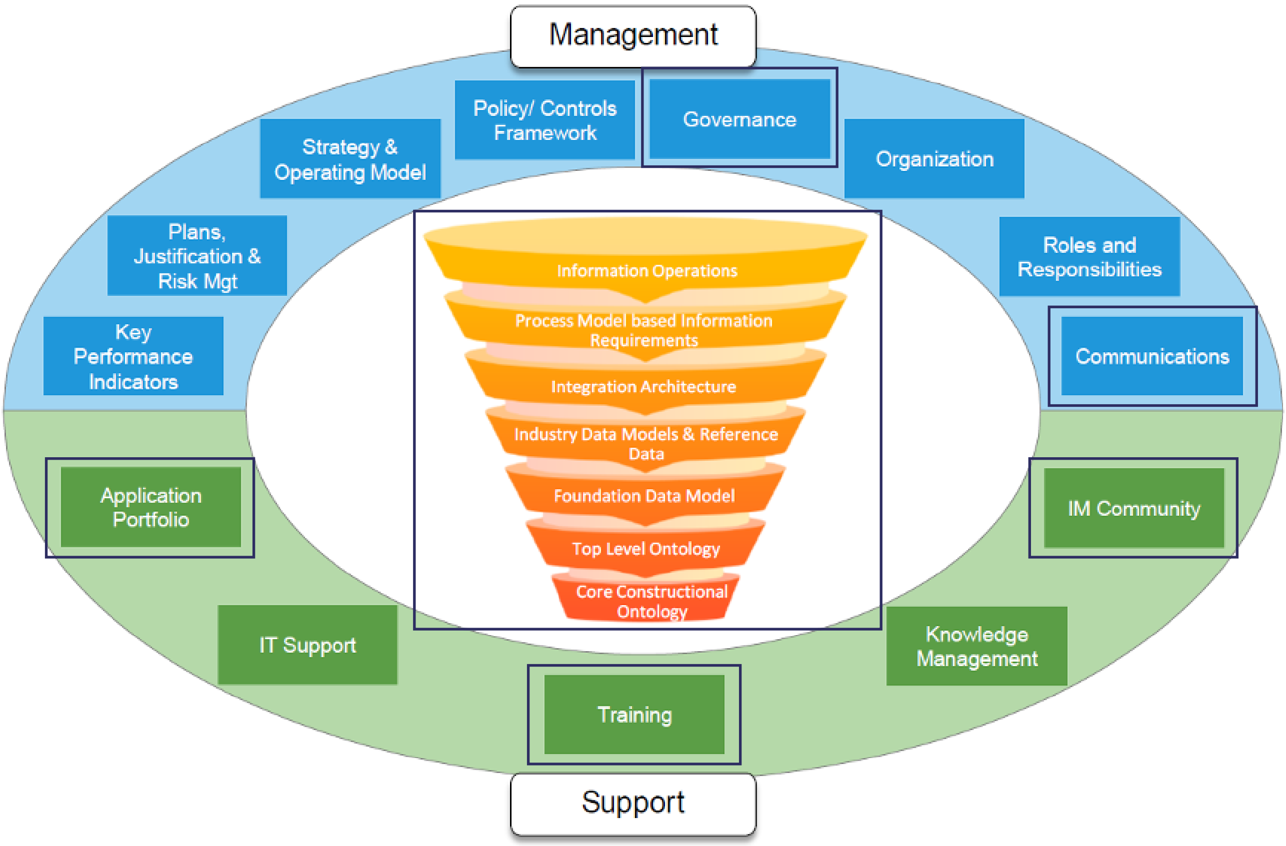

Figure 4: The management and technical components of an IMF, with those considered in this project highlighted[7].

7 *Courtesy of Dr Matthew West, CDBB Cambridge UK, paper in preparation, presentation describing the diagram available on the Digital Twin Hub*

Vision

Although this work has gone some way to setting an initial vision and roadmap for the IMFe, refining this to create the long-term vision for digital twins of the environment is still needed. This vision needs to clearly articulate the value proposition: why society should care, and why taxpayer should fund it. It should define a decadal strategy and vision for digital twins that ensures the value in developing digital twins is maximized for stakeholders.

Key recommendation 1

Clearly articulate a UK vision for the long-term ambition for digital twins of the environment.

Key recommendation 2

Building on the current momentum of stakeholder engagement, establish fora for engagement that ensures cohesive community engagement on digital twin governance, coordination and development approaches.

Engaging across communities of interested parties is critical to progressing the roadmap. Stakeholder communities must come together to stimulate, and coordinate, future activity and should include governance groups, ensuring that the IMFe architecture is fit for purpose for multi-disciplinary environmental digital twins. There is a recognised requirement to develop national (shared) cyber-physical infrastructure and a clear role for government and individual funders to come together to plan and coordinate future investment opportunities.

There was a clear desire by the community to maintain the stakeholder momentum, and a stakeholder roundtable in six months to chart progress, to push ahead with the creation of milestones to detail progress and build, momentum has been requested. The ambition is to build a community that engages with developing and delivering this vision and has clarity on the potential benefits.

Conceptual framework

Fundamental principles that underpin the IMFe have been identified which provide the basis upon which the IMFe should be further developed. In common with the CDBB roadmap for the IMF for the built environment (CDBB, 2020), which discussed incremental delivery as a fundamental principle, we also believe that developing the IMFe should be incremental. Digital twins are evolving, and their needs will also consequently evolve.

There will not be a single, monolithic project to develop the components of an IMFe. Both digital twins, and the IMFes needed to support them, will therefore be best developed using an agile approach. **A modular, incremental, development cycle** will allow component reuse and a broad community to contribute. This will allow the development of IMFe components in achievable, bite size chunks, aligned with project needs, and thus ensure the barrier to involvement in digital twin development is as low as possible, whilst ensuring long-term sustainability through reuse and sustainable, incremental community activity.

The priority of this roadmap is to define an approach that allows **incremental improvement** in the digital capacity of the UK's environmental science community in support of digital twinning, whilst also offering a step change in prediction and forecasting capability through an approach that achieves synergy between process and data understanding. It is neither desirable nor achievable to develop an IMFe in a single project. This roadmap outlines how one might develop digital assets and approaches that benefit a broad community, that are sustainable in the long-term and which are achievable through multiple activities reinforcing each other.

Key recommendation 3

A conceptual IMFe with guidance on principles for developing its components should be produced at an early stage. The approaches outlined in the IMFe project should be followed.

This will only work if the IMFe components are developed with interoperability and reusability at their heart and using **open-source approaches** where possible.

The development framework needs to be **documented and made easily available** and usable by a broad community of developers, with appropriate mechanisms to ensure the guidelines are widely adopted.

It should reflect **principles of backwards compatibility**, important to ensure resilience to change in the IMFe.

It should recognise the **differences and unique requirements** of digital twins operating in the environmental domain and develop (federated) approaches that meet these requirements **whilst being consistent and ensuring interoperability with other IMF instantiations**, including the CDBB architectural approach.

It should adopt a **'systems thinking'** and overall **'systems of systems'** approach that recognises the importance of interactions and feedbacks in complex environmental systems.

The approach to developing the IMFe must recognise the **diversity** of different communities both using and developing digital twins allowing them freedom to move at their own pace and seek their own solutions, while expediting the need to interoperate. These principles are all consistent with and strongly point towards a **'commons' approach** as embraced increasingly in other fields, for example bio- and medical informatics. A data commons "brings together (or co-locates) data with cloud computing infrastructure and commonly used software services, tools & applications for managing, analysing and sharing data to create an interoperable resource for a research community" (Grossman, 2018). A commons approach implies a community-driven approach where components are published and shared for the common good. We therefore advocate a community-led approach to service selection and standardisation, avoiding imposing standardisation as the core mechanism for interoperability.

More generally, it must recognise that **interoperability is an evolutionary journey**, not a destination and that the important role of the roadmap is to plot out this journey over time.

The principles outlined above form the basis for the IMFe future development. Further challenging this conceptual basis to ensure it is robust would ensure the future IMFe development is based on solid foundations.

Management and governance

Effective governance is critical and we see a commons approach as providing a pathway to embed governance into the heart of an IMFe architecture. Online, and widely available, resources that are collaboratively developed and managed by a community for common good are needed. The broader governance framework for digital twins and IMFe should include the development guidelines outlined above to ensure they become widely adopted.

Key recommendation 4

An IMFe management and governance approach built upon the concept of the commons, consistent with the CDBB and future National Digital Twin programme approaches, should be produced and made widely and freely available.

We recommend adopting the management approaches as developed by CDBB (Figure 3). These cover aspects such as performance measures, plans, risk management, strategic and operating models, organisation design, roles and responsibilities, communications and governing guidelines and activities, although not all have been directly considered in this work. This will ensure consistent management approaches across all digital twins for the built and natural environment. However, we recommend management and governance approaches that are specialised for environmental data. This approach needs to be documented and made easily available alongside the more technical development framework.

Management and governance approaches will only be effective if they have broad support across digital twin developers and operators. Communities of stakeholders should be closely engaged in the development of this part of the IMFe to ensure the approaches agreed will be adopted and are sustainable. Mechanisms to ensure adherence to the governance and management principles of the IMFe can (and should) be included within funding calls where appropriate.

There is a need to support sustainably integrating and, if needed, collaborating with legacy developments to ensure that new developments are able take advantage of the work that has come before.

As part of the management framework, measures of performance of the digital twin will be required and need to be developed. As well as operational metrics required to monitor the functioning of the twin, these measures should also include performance against quality of the information and should be able to demonstrate the benefits of digital twins.

Governance must include appropriate consideration of ethical issues, including equitable access to data and information. Where there are gaps in ethics policies on digital twins these will need to be addressed by the IMFe and where appropriate adopted in wider ethical policies.

Common language

Future development of the IMFe will need to engage stakeholders to get buy-in, to maintain energy and funding and, most importantly, to ensure the evolving needs in this dynamic area continue to be taken into consideration.

The stakeholder community must encompass more than the science and research community. Digital twins are best developed from the user perspective, and they must be represented throughout. Taking a user led approach to digital twinning approach is most likely to achieve successful outcomes and mitigates against the risk of overselling digital twins. It is also important to engage the private sector, who are already developing digital twins, to support the work done within the environmental science community. The engagement of industry also has the potential for improving data access, both through unlocking data from private sector and through shared, and improved, data access infrastructure.

To build up and maintain expert communities, the environmental science community should develop a standing, open forum to engagement developers and users, including from the commercial sector.

Key recommendation 5

Engage the user community actively in developing the ideas for digital twins at an early stage. The community forum already started in this IMFe project should be developed further to stimulate this user led approach. An additional, open forum, to engagement developers and users, including from the commercial sector would also be recommended.

Key recommendation 6

Building on the framework developed in this project, develop a common language to describe digital twins, their components and in particular the components of an IMFe.

Key recommendation 7

Interoperability, by definition, needs buy-in across the spectrum of digital twin developers and this work needs engagement with national and international actors. The activities of large programmes such as Destination Earth should be aligned with UK activities.

One of the clear limitations in developing digital twins is in having a common language to share between the diverse range of users and developers. This lack of a common language hampers the ability to describe what digital twins are, what they do and how they might benefit users.

It also makes it difficult for those engaged in developing parts of the twin to work together. A common language to be used when discussing digital twins is fundamental to gathering the community and managing progress in a collaborative way. Even the term digital twin itself needs to be carefully considered. Communication strategies are needed to ensure the language used is clear, unambiguous, and inclusive across a broad spectrum of communities. Many of the components of digital twins, and concepts required to describe them, remain opaque due to the lack of a common and agreed language to describe them.

Following a broad consensus on defining what a digital twin is, it is important to develop briefs and visuals that reflect this and convey their role and importance to non-technical audiences.

Environmental information by its nature requires international as well as national cooperation. The geographic nature of the ocean, atmosphere and cryosphere domains mean information and data gathering often crosses national domains and covers regions that are beyond any single national jurisdiction. The development of a federated, interacting, body of data and information in the environmental domain therefore by necessity must be seen in the context of interacting national and international efforts. As a significant player in the development of standards for interoperability and with a well-established national programme of coordination of digital twins and their associated infrastructure and guidelines, the UK is leading in many areas relevant to this roadmap. However, there is the obvious danger if we define national activities outside the scope of international agreements and collaboration that the UK's developments will become superseded by work done elsewhere. It is incumbent upon the developers of the components defined in this roadmap to address this through engaging widely, and early, in the appropriate international fora that will support long-term uptake both in the UK and internationally of the approaches developed in UK projects. That means working in collaboration with non-UK partners to develop the IMFe with other suitably skilled groups, as well as playing an active part in the appropriate community fora that guide our communities.

Pilot studies

Following the approach used in this study, exemplars of digital twins for developing and testing the development of IMFe components are needed. Pilot studies will be used to iterate and validate models for the key IMF components which can then be duplicated in larger-scale studies.

The components that constitute an IMFe are largely understood and can be mapped in abstract but pilot studies are required to validate these components and assess their relevance. Delivery of the roadmap is dependent upon the capacity to develop multiple, multi-disciplinary digital twins that challenge the existing IMFe conceptual framework across a range of dimensions. Through iterative development of the IMFe through pilot studies, ideally with complementary foci, an increasingly complete IMFe definition can be developed.

Providing decision makers outside the science community with access to an example digital twin pilot they can engage with, in particularly in decision support, would add significant value, and support the understanding of their potential value.

Key recommendation 8

A number of diverse environmental digital twin pilots, with explicit reference to the IMFe building blocks as detailed in this project, need developing to challenge and test the IMFe and our readiness to implement it.

Key recommendation 9

The use of personas for describing the users of the IMFe have proved a useful approach to highlighting a user needs of an IMFe. Pilot studies should be developed in the context of key policy or service end user personas, ideally worked through with the user themselves.

The need for easy-to-use interfaces and tools to visualise outputs from the digital twins is important and should be delivered through these pilot studies.

Stakeholder engagement highlighted areas where there may be opportunities to develop digital twins suited both to delivering on the IMFe and to demonstrating to users their utility.

Sustainability, and the link to Net Zero, is a key driver for environmental science. The Net Zero Oceanographic Capability (NZOC) project highlighted the importance of the data ecosystem around ships and autonomous observing for reducing the carbon cost of ocean observing, a significant contributor to NERC's carbon footprint. Digital twins (together with the use of artificial intelligence, robotics and autonomous vehicles) have the capacity to underpin operational planning and reduce the carbon expenditure for data capture. A digital twin of the RRS Sir David Attenborough is already in development at BAS and could be extended (perhaps to other NERC ships and the NERC autonomous observing capability) to challenge the ability to include observations from diverse sources in a digital twin of our observing system. This has the added benefit of having strong Equality, Diversity and Inclusion benefits through opening up observing to those not able or willing to spend time in the field. It will support the ambition to attract a more diverse intake to environmental science. The Land Insight case study also has significant potential to answer key Net Zero related scientific questions.

A strong policy and / or impact link is also important to ensure engagement with stakeholders outside the science community. The Met Office 3DT (demonstrator dispersion digital twin) case study is at a relatively mature stage and would be well-placed as a pilot for developing the IMFe. Other examples were also discussed at the stakeholder workshops, including the idea of digital twins supporting understanding and monitoring biodiversity within Marine Protected Areas.

The principles of diversity across the axes of spatio-temporal scale, theme and user requirements should be used to ensure the pilot studies provide a breadth of challenges to the IMFe. Personas were used within the IMFe project and proved helpful and should continue to be used in pilot study development. These personas should represent a diversity of user experience and interaction with the data, and a range of security and access control requirements, to represent a significant span of the likely uses of environmental digital twins.

Gap analysis

There will never be a complete information management framework, and the components will be built in an incremental way. The key is to make sure that developments are reusable and shareable and, wherever possible, to use or build on existing capabilities. However, the IMFe project highlighted gaps and missing capabilities required to develop the IMFe. Informed by continued development of the conceptual structure pilot studies, a better understanding of where the gaps lie will inform the long-term digital twin development needs.

Key recommendation 10

Bringing together the outcomes from digital twins Pilot Studies, develop an improved analysis in the gaps in the IMFe to inform future targeted developments.

Data: Ingestion, curation, quality frameworks

A fundamental aspect of digital twins is the complexity of the data journey, from observation to ingestion into a twin, transformation into intermediate information products and then passing through to the data and information forms that the user needs. Throughout this journey the data provenance must be maintained, and the quality assured throughout. Machine-readable trustworthy data and processes are needed. The IMFe therefore needs to enable the synthesis of multiple data sources and scales together with models to deliver understanding of interactions of complex processes in comprehensible ways in machine-readable forms.

Key recommendation 11

User-led data quality and data provenance capabilities are required and need development. Specific pilot studies designed to develop tools and methods for carrying quality and provenance information through digital twins are needed.

The IMFe also needs to promote the adoption of data quality frameworks to ensure data ingested into digital twins are fit for purpose. These quality measures need to be carried with the data to allow the user to monitor the data journey. The provenance of digital twin outputs should also be clearly documented for users, and the information required to provide that documentation should be tracked and made available throughout the data journey.

There will be elements of the information desired by users that will be difficult to provide, and a cascade of increasingly complex digital twins would be a sensible approach in dealing with this problem.

An asset commons: Standards, vocabularies and catalogues

Standards are essentially an agreement between providers, regulators and users. They provide rules, guidelines and characteristics that enable connection between systems, data, people, hardware, software and procedures. Standards reduce effort, time and cost of implementing technologies, improve return on investment, and help future-proof systems by enabling new capabilities to be added with minimal effort. Underpinning the concept of interoperability is the exchange of FAIR[8] (meta)data between digital twins to enable whole systems to be simulated and these results to be presented in a unified way to stakeholders. Semantic interoperability is facilitated by standards, be those metadata, data formats, or unified data models.

Use of standards across all aspects of the architecture will be critical to the long-term success of digital twins of the environment. Standards for metadata will drive discoverability. Standards for APIs or standards for hypermedia formats provide access to data through interfaces. Standards for the data models drive semantic understanding of the data, important for humans and for machine learning and artificial intelligence. Standards for dealing with authentication and authorization where it is required support digital twin security. And of course, not least, standards for the data formats themselves encapsulate the data. The NERC community is active in this area and progress is being made in the research community in the implementation of FAIR (and other) principles for environmental data.

Metadata standards have developed over time, some of the earliest being associated with environmental domains. ISO 19115 metadata (International Organisation for Standardisation,

Key recommendation 12

The NERC Environmental Data Services already have a role in developing and supporting standards in their thematic areas, and that role should be extended (and funded) to include the standards needed to support environmental digital twins.

Key recommendation 13

Explicit recognition of the overhead of designing to, or contributing to developing, common standards should be included in funding calls for environmental digital twins to ensure that appropriate resource is placed in this activity.

Key recommendation 14

Following the review of ontologies in the IMFe project, ensure that recommendations are made for ontologies as part of the IMFe governance.

8 *Findability, Accessibility, Interoperability, and Reuse* https://www.go-fair.org/fair-principles/

2014 and extensions to it), for example, describes geographic data, data services and data models. It can document persistent identifiers, provide citations, give detailed lineage and provenance, use and access constraints, quantitative and qualitative conformance statements, locate the data in time and space, provide detailed contact information to provide support. The challenge will, in the main, be to advocate for good metadata and ensure it is provided in projects relevant to digital twins of the environment.

It will largely be outside the scope of the digital twin development community, or indeed any one community, to develop data standards alone and where possible the community should use existing data standards and to develop them in line with FAIR principles. To ensure interoperability, community activities to develop data standards should be broader than digital twin activities, but digital twin developers should take an active role in influencing and supporting the appropriate standards as appropriate. Where possible standards should be developed through existing national and international organizations, building on progress already made with FAIR and respecting the differences that have emerged between communities and disciplines.

Digital twins of the environment should be designed to be flexible to an evolving landscape of data inputs, including where standards don't yet exist. Using standardised data formats through prescribed interface standards such as Application Programming Interface (APIs) allows the use of common functions regardless of the domain. Common visualisation capabilities and dashboards can be developed that meet multiple requirements if all accessed via a uniform interface.

A balanced approach to standards should be used to ensure the minimum required standards to facilitate the interoperability needed are implemented. Full interoperability across all perceivable twins or domains is not possible. Some twins will need to interoperate at the discovery level, others at the data in/out, and others at the process level. Whilst not all digital twins will need to be fully interoperable, interoperability will enhance the opportunities presented for their utilisation by other disciplines and should be encouraged.

Key recommendation 15

Future digital twin development projects should include appropriate effort in the development of vocabularies and ontologies, including cross-disciplinary top-level ontologies as well as the bridging mid-level ontologies and domain ontologies.

Key recommendation 16

The asset commons, and in particular asset registers and architecture descriptions, are underdeveloped and need focussed activity to support their development. The IMFe framework should include the recommendation for digital twin developers to develop systems using, and contributing to, an asset commons.

To conform to the "system of systems" principles inherent in the IMFe framework, digital twins will be incorporating translation engines that allow transfer from one standard to another. Mapping of concepts through ontologies will be an essential process for practically delivering interoperability of data. **Ontologies** describe the relationships between terms and can be used for matching and translating terms from different dictionaries that relate to common concepts. The concept of ontologies is therefore essential for enabling semantic interoperability for environmental data, particularly where related concepts are described differently across discipline boundaries. Top-level ontologies are general concepts that are common across all domains, and so cannot be developed in the environmental domain in isolation. These ontologies provide reliable and reusable definitions of abstract concepts and allow mapping of concepts across discipline boundaries. Engagement to decide how we might obtain national and international agreement around the use of a top-level ontologies would greatly progress environmental digital twins towards wider interoperability.

Mid-level ontologies bridge the abstract concepts of top-level ontologies and the rich details of domain ontologies by providing common terms (e.g. space and time). These ontologies are rarely well-developed and will need to be development using multi-disciplinary pilot studies. Domain-level ontologies define specific domains and how to extend concepts from mid-level ontologies. Domain-level ontologies are largely well-developed and are unlikely to need significant development to allow digital twin developments.

Dictionaries of controlled **vocabularies** are sets of terms related to a specific topic with standardised definitions. They are widespread across the environmental sciences and underpin many areas of scientific enquiry. These allow standardised documentable descriptions of parameters within the models that underpin delivery of digital twins. Work is required to catalogue these vocabularies and to make sure they underpin the development of the IMFe.

The pilot studies mentioned above will help to determine gaps in ontologies and vocabularies. Follow-up activities to define critical, missing ontologies and vocabularies is needed otherwise the lessons from the pilots will not be able to be implemented.

A data commons brings together data with computing infrastructure, tools and services to provide a resource for the community. We advocate extending the commons concept to incorporate a more general view of environmental digital assets, including data, models, workflows, notebooks, and indeed digital twins themselves, focussing on any re-usable asset. The assets to be included in the commons would not only include the list of digital assets above, but also the standards and policies required by the different stakeholders benefitting from the commons. The assets of the commons, including the policies associated with their governance, need to be identified within an asset catalogue.

The asset commons catalogue will need to use identifiers for each asset so that they can be located and reference either when they are identified individually or when they are linked together in semantic descriptions of the information flows through the digital twin. A call to this identifier in the catalogue will return the standardised metadata for that asset enabling access, its semantic relationships within the digital twin architecture and the polices governing its use (such as the authorisation required and licenses governing reuse). The IMFe needs to include a well-designed catalogue of digital twins with easy-to-understand and searchable high-level features, and easy access to more detailed technical information. For digital twins to interoperate there needs to be a standardised description of the digital twin, it's components and its interfaces. The architecture descriptions should incorporate a list of ingredients for the digital twins that describes how they may be reused across different applications.

Security and support

We recommend adopting the security and support elements as advocated by CDBB and future National Digital Twin programme guidance to have a consistent approach across UK digital twin communities. IT support, training and community building need a particular focus for the environmental community.

Tailored support for security processes and technologies for different use cases of environmental digital twins is needed, given the diversity of security needs expected. We should embrace this diversity and support a variety of solutions. In line with the commons philosophy, we advocate a minimal but sufficient security requirements in terms of authentication and authorisation.

Key recommendation 17

The IMFe needs to support access control and authorisation layers to facilitate different access/user roles in digital twins. The governance framework should be developed to include guidance on security and support.

Computational architecture

We advocate a cloud native approach recognising the desired feature of portability across different cloud and non-cloud environments. The location of data and the need to build an internet of things approach that connects the edges to the core is especially challenging in communications limited environments. There are bodies of expertise in this area already existing in the technology world we need to engage with. The challenge here is to understand how the design of the IMFe interacts with the design of the communications and compute architectures, and to develop the two together.

The IMFe needs to consider the best way to capture and manage the relationships between digital twins and digital twin components to enable users to plug-and-play with multiple digital twins or modular components of digital twins. Workflows and operational management tools need to be managed and made available, considering links to other workflows to define quick wins to delivery early impact from digital twins. These workflow or operational management tools need access control of services and elements of the twin.

Although the ambition is for modular and transferable components to be developed that can be agnostic to the compute environment, it became clear as part of the IMFe project that clear digital strategies across the developers of digital twins will support the application of an IMFe.

This is particularly important when considering the implication of the compute needs for digital twins in meeting net zero strategies. The basis for the forming of the digital twins in the use cases is the synthesis of multiple data sources and scales together with models to deliver understanding of interactions of complex processes in comprehensible ways.

Key recommendation 18

The IMFe should be developed in alignment with broader UK Digital Research Infrastructure strategies.

Key recommendation 19

Approaches to workflow and operational management tools that scale appropriately need to be developed.

Key recommendation 20

At institutional and research council level a clear policy and implementation strategy is needed for developing the mix of on-premise and cloud provision of compute infrastructure.

User experience

The user interface is an essential part of digital twins and is key to releasing the value in the twin. Stakeholders for digital twins may require both the capability to manipulate / customise data to be input into digital twins and then to interact with output data using them to drive access applications such as (but not limited to) portals, dashboards, visualisations, decision support systems and notebooks are all elements of digital twins, although not all will necessarily be in all twins. Such targeted outputs will increase the utility and impact of investment in digital twins and thereby drive the requirements for future investments.

Users of environmental information emphasise that a digital mechanism for providing actionable information must start with a clear understanding of the users' need. Whether this information is provided from a digital twin or other form of infrastructure is largely irrelevant to the user; what is key is the ability to refine, within the limits of the usability and quality of the underlying information, the underlying data to provide something that meets the user need. In some cases that may be maps or GIS layers that integrate with another expert system, in other cases it may be warnings or alerts. The provision of the ability to test scenarios and to integrate data from diverse sources to provide the outcomes required are also all expected to be driven through a user interface. It is also important to permit different viewpoints for different users.

Key recommendation 21

Given the fundamental importance of the user experience, use pilot studies to develop reusable user interfaces that allow visualisation, transformation and data combination and lower the barrier to access to the data in digital twins.

Developers of environmental information products are rarely able to pre-emptively design in the mode of delivery or the form in which the final information product will take. Designing in the ability for users to iteratively refine the outcome is therefore an important outcome of any interface design. To prevent the infinite numbers of possible interactions between the data and the user having to be independently, and therefore inefficiently, developed a modular approach to developing the user interface should be taken.

The IMFe needs to promote user focused interfaces enabling the full breadth of user communities to interact meaningfully with digital twins.

Roadmap summary

The development of this roadmap has been an incredibly positive and enriching process involving strong collaboration across the sector and folding in the views of a range of stakeholders as a fundamental part of the process. This has resulted in a common initial vision, approach and associated roadmap for the development of an IMFe. More than this, this project has built considerable momentum within the UK environmental sciences community and a shared enthusiasm about what digital twins can bring to this sector. Building on the IMFe there is also a real opportunity to develop a coordinated approach whereby the implementation of an IMFe would enable federation and interoperability by design. The IMFe consortium have also realised in their work that what is being proposed underpins not just digital twin development but a broader range of efforts in supporting integrative science, bringing together data and modelling assets into a common framework to allow them to work together.

Key recommendation 22

There is a need for significant investment in an IMFe to build on the current momentum and take advantage of the opportunities to support a coordinated approach to their development.

The moment is right to build on this momentum and thrust the UK to the forefront for digital twins of the natural environment, whilst also enabling exciting developments linking digital twins of the built and natural environment. This will require investment and a broader strategic vision to ensure that development of digital twins is aligned with other related developments around digital research infrastructure.

Glossary

Artificial intelligence	The ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings.
Autonomous	Machines or software designed to operate without human intervention.
CDBB	Centre for Digital Built Britain was a partnership between the Department for Business, Energy & Industrial Strategy (BEIS) and the University of Cambridge. It aimed to deliver a smart digital economy for infrastructure and construction.
Cloud native	The concept of building and running applications to take advantage of the distributed computing offered by the cloud delivery model.
Commons	Commons are resources that communities manage for individual and collective benefit. A data commons is both the technology for sharing data sets and the set of principles, governance strategies, and utilities that make use of those data sets possible. An asset commons extends that concept to the broader digital twin assets.
Cyber-physical systems	Cyber-physical systems integrate sensing, computation, control and networking into physical objects and infrastructure, connecting them to a compute environment and to each other.
Data	Factual information (such as measurements or statistics) used as a basis for reasoning, discussion, or calculation.
Digital Model	An informative digital representation of a system. Models may use data techniques and / or process knowledge to generate the representation.
Digital Shadow	A digital model that integrates information from its physical counterpart, such as weather forecast models, but there is no corresponding feedback to the physical counterpart.
Digital Twin	A digital model with (real- or right-time) two-way information flows. This can enable autonomous optimisation, and remote and autonomous operation. These are not yet widespread, but are typically found in autonomously operated environments such as robotics and autonomous vehicles and maritime vessels
Federation	The inter-operation of two distinct, formally disconnected, computing infrastructures (that may have different internal structures) through agreed standards of operation.
Information	Processed, organized and structured data, providing context to enable the communication or reception of knowledge or intelligence.
Information Management Framework for Digital Twins (IMF)	A built environment information management landscape and data environment which will adequately define parties, processes, information, technology to support the National Digital Twin
Information Management Framework for Digital Twins of the natural environment (IMFe)	An environmental information management landscape and data environment which will adequately define parties, processes, information and technology to support the development and use of federated environmental cyber-physical infrastructure, including digital twins, for national digital twin programmes and recognizing international engagement.
Internet of Things (IoT)	A network of electronic devices embedded with software and sensors that enable the interaction between machines
Interoperability	The ability of different systems, devices, applications or products to connect and communicate in a coordinated way.
Model	See digital model.
Observation	A measurement of some element of the real object.

Ontologies	An ontology is a formal version of a thesaurus where relations are described using a formal system such as Description Logic (DL) to mathematically classify individuals of classes and properties.
Personas	Characters that represent user types that might use or develop digital twins in a similar way. Creating personas helps the developer to understand users' needs, experiences, behaviours and goals.
Pilot Study	A small-scale study developing examples or prototypes of a capability to explore methods and approaches.
Real-time	Data, information or outcomes available without any (significant) delay from the actual time they are valid.
Right-time	Data, information or outcomes available when its impact is the greatest.
Semantics	Relating to meaning or understanding.
Simulation	The use of models to build a digital representation of a real object or system, and often includes predictions of future states based on a measured initial state.
Systems of systems	A collection of systems, each capable of independent operation, that interoperate together to achieve additional desired capabilities.
Systems thinking	Considering things holistically, as connected wholes rather than separate parts.
Vocabularies (controlled vocabularies)	A controlled vocabulary is a normalised, restricted list of terms for a specific use or context. Thesauri and taxonomies are types of controlled vocabularies, but not all controlled vocabularies are thesauri or taxonomies

Annex

Key recommendation 1
Clearly articulate a UK vision for the long-term ambition for digital twins of the environment.
Vision – Page 10

Key recommendation 2
Building on the current momentum of stakeholder engagement, establish fora for engagement that ensures cohesive community engagement on digital twin governance, coordination and development approaches.
Vision – Page 10

Key recommendation 3
A conceptual IMFe with guidance on principles for developing its components should be produced at an early stage. The approaches outlined in the IMFe project should be followed.
Conceptual framework – Page 10

Key recommendation 4
An IMFe management and governance approach built upon the concept of the commons, consistent with the CDBB and future National Digital Twin programme approaches, should be produced and made widely and freely available.
Management and governance – Page 11

Key recommendation 5
Engage the user community actively in developing the ideas for digital twins at an early stage. The community forum already started in this IMFe project should be developed further to stimulate this user led approach. An additional, open forum, to engagement developers and users, including from the commercial sector would also be recommended.
Common language – Page 12

Key recommendation 6
Building on the framework developed in this project, develop a common language to describe digital twins, their components and in particular the components of an IMFe.
Common language – Page 12

Key recommendation 7
Interoperability, by definition, needs buy-in across the spectrum of digital twin developers and this work needs engagement with national and international actors. The activities of large programmes such as Destination Earth should be aligned with UK activities.
Common language – Page 12

Key recommendation 8
A number of diverse environmental digital twin pilots, with explicit reference to the IMFe building blocks as detailed in this project, need developing to challenge and test the IMFe and our readiness to implement it.
Pilot studies – Page 13

Key recommendation 9
The use of personas for describing the users of the IMFe have proved a useful approach to highlighting a user needs of an IMFe. Pilot studies should be developed in the context of key policy or service end user personas, ideally worked through with the user themselves.
Pilot studies – Page 13

Key recommendation 10
Bringing together the outcomes from digital twins Pilot Studies, develop an improved analysis in the gaps in the IMFe to inform future targeted developments.
Gap analysis – Page 14

Key recommendation 11
User-led data quality and data provenance capabilities are required and need development. Specific pilot studies designed to develop tools and methods for carrying quality and provenance information through digital twins are needed.
Data: Ingestion, curation, quality frameworks – Page 15

Key recommendation 12
The NERC Environmental Data Services already have a role in developing and supporting standards in their thematic areas, and that role should be extended (and funded) to include the standards needed to support environmental digital twins.
An asset commons: Standards, vocabularies and catalogues – Page 15

Key recommendation 13
Explicit recognition of the overhead of designing to, or contributing to developing, common standards should be included in funding calls for environmental digital twins to ensure that appropriate resource is placed in this activity.

An asset commons: Standards, vocabularies and catalogues – Page 15

Key recommendation 14
Following the review of ontologies in the IMFe project, ensure that recommendations are made for ontologies as part of the IMFe governance.

An asset commons: Standards, vocabularies and catalogues – Page 15

Key recommendation 15
Future digital twin development projects should include appropriate effort in the development of vocabularies and ontologies, including cross-disciplinary top-level ontologies as well as the bridging mid-level ontologies and domain ontologies.

An asset commons: Standards, vocabularies and catalogues – Page 16

Key recommendation 16
The asset commons, and in particular asset registers and architecture descriptions, are underdeveloped and need focussed activity to support their development. The IMFe framework should include the recommendation for digital twin developers to develop systems using, and contributing to, an asset commons.

An asset commons: Standards, vocabularies and catalogues – Page 16

Key recommendation 17
The IMFe needs to support access control and authorisation layers to facilitate different access/user roles in digital twins. The governance framework should be developed to include guidance on security and support.

Security and support – Page 17

Key recommendation 18
The IMFe should be developed in alignment with broader UK Digital Research Infrastructure strategies.

Computational architecture – Page 18

Key recommendation 19
Approaches to workflow and operational management tools that scale appropriately need to be developed.

Computational architecture – Page 18

Key recommendation 20
At institutional and research council level a clear policy and implementation strategy is needed for developing the mix of on-premise and cloud provision of compute infrastructure.

Computational architecture – Page 18

Key recommendation 21
Given the fundamental importance of the user experience, use pilot studies to develop reusable user interfaces that allow visualisation, transformation and data combination and lower the barrier to access to the data in digital twins.

User experience – Page 19

Key recommendation 22
There is a need for significant investment in an IMFe to build on the current momentum and take advantage of the opportunities to support a coordinated approach to their development.

Roadmap summary – Page 19