Developing an information management framework for environmental digital twins (IMFe)

Justin Buck[1], John Siddorn[1], Gordon Blair[1], David Boot[1], James Byrne[1], Andrew Kingdon[1], Alice Kloker[1], Alexandra Kokkinaki[1], Gwen Moncoiffe[1], Eleanor Blyth[1], Matt Fry[1], Rachel Heaven[1], Edd Lewis[1], James Passmore[1], John Watkins[1], Ben Marchant[1], Bruce Napier[1], Sam Pepler[1], Poppy Townsend[1], and Kate Winfield[1]

[1]Affiliation not available

March 28, 2023

Developing an Information Management Framework for Environmental Digital Twins (IMFe)

Justin J.H. Buck, J. Siddorn, G. Blair, D. Boot, J. Byrne, A. Kingdon, A. Kloker, A. Kokkinaki, G. Moncoiffe, E. Blyth, M. Fry, R. Heaven, E. Lewis, B. Marchant, B. Napier, J. Passmore, S. Pepler, P. Townsend, J. Watkins and K. Winfield

NOC, UKCEH, BGS, STFC, BAS

PRESENTED AT:

AGU FALL MEETING
Chicago, IL & Online Everywhere
12–16 December 2022
SCIENCE LEADS THE FUTURE

INTRODUCTION

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. The roadmap was published in May 2022 and we recently started a project to pilot the IMFe using a biodiversity use case which will be challenged by using other environmental twins in the environmental doming to ensure it is extensible.

This digital poster describes the work undertaken to generate the roadmap and its recommendations before moving on to describe the work we are planning to undertake in the pilot project. We are keen to engage with and align to related efforts globally to maximise the impact and breadth of the IMFe.

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

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

IMFE SCOPING APPROACH

To ensure we capture a broad range of community requirements we have combined top-down, conceptual consideration of the domain with a bottom-up approach using case studies to understand the key aspects of information exchange. We identified and documented existing components of an IMF for environmental digital twins. The use of case studies and defining them by their scope and scale proved helpful in mapping the components that already. The approach is shown visually in Figure 2 The development of Personas to highlight key interactions with the information management framework within these use cases allowed us to analyse the use cases from a use perspective and brought helpful insights to the study.

Six generic user types were identified:

1. A data producer with relevant data who wants to feed data into a DT;

2. A data user who needs to access output from different DTs;

3. Someone who wants to develop a DT; i.e. a DT product owner in charge of scoping or developing a twin

4. A machine or robot accessing and feeding data into DTs;

5. n environmental research scientist who needs to use DT technology as an alternative or as an extension to physical fieldwork;

6. Someone developing, running, or using the output of an alternative instance of a DT.

The six personas represent users with a range of technical knowledge and with the following type of interactions with DT: feeding data in, getting data and information out, creating twins from scratch, coding a robot to interact with twins, non-expert plug-and-play interactions, and expert customisation of existing twins.

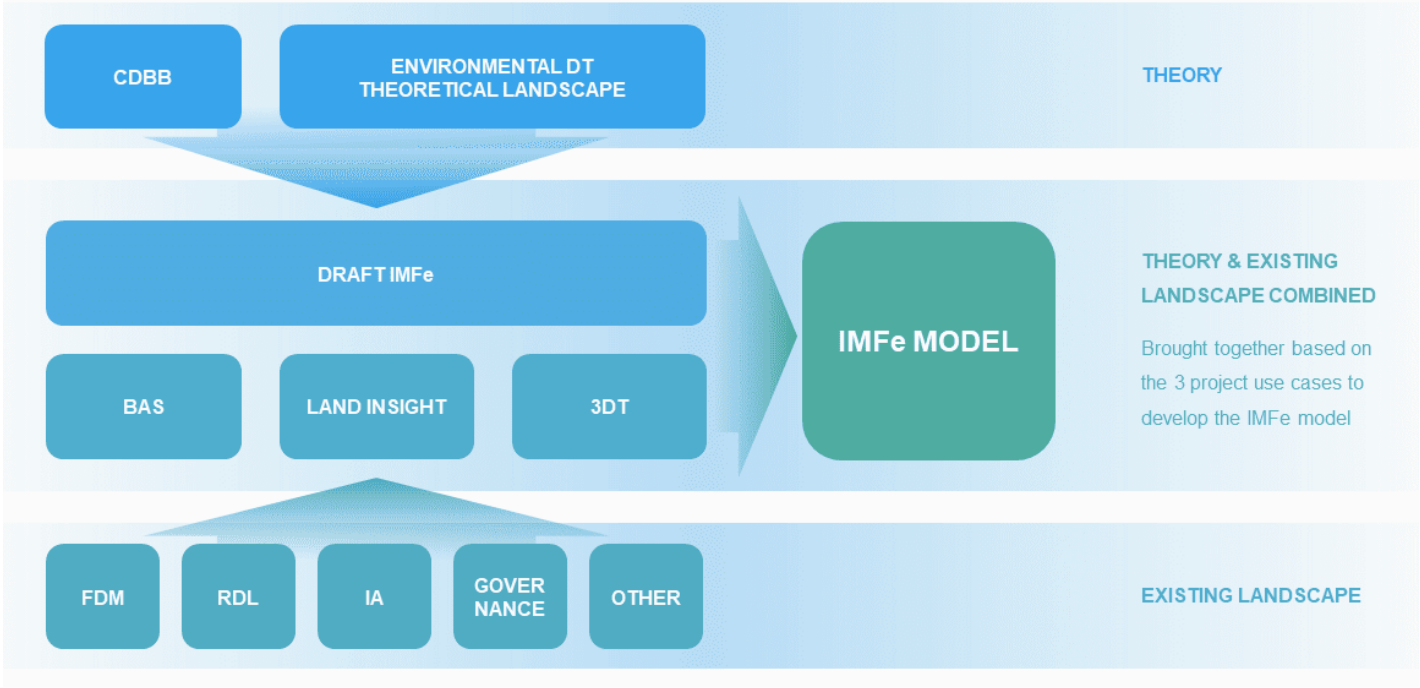


Figure 2: schematic showing how top down and bottom up requirements were combined to generate the IMFe.

Particularly helpful in defining the approach has been the use of workshops. This approach has been taken in the expert analysis, bringing together the project team and including experts from partner institutes (notably the Met Office, BAS and Telicent) to quickly brainstorm concepts and in-depth knowledge. It has also been taken to engage stakeholders including government to ensure this work has had a user-led focus.

To develop the conceptual framework the project started from the CDBB IMF [1] (Courtesy of Dr Matthew West, CDBB Cambridge UK, paper in preparation) and then mapping these digital twin components identified by the project to the conceptual framework of CDBB and the iteration developed for the IMFe that described in the recommendations section of this poster. The bottom-up assessment of the IMFe was then mapped to the conceptual framework to verify that the conceptual framework would stand the test of real digital twins and gave confidence in the approach.

Finally, recommendations arising from the project were then consolidated into a narrative framed around the recommendations to create guidance that can support the development of the IMFe.

[1] https://www.cdbb.cam.ac.uk/DFTG/DFTGRoadmap

IMFE RECOMMENDATIONS

The core recommendation from our project is that digital twins of the natural environment should be developed on top of an Information Management Framework tailored towards the needs of the environmental sciences domain (IMFe). This is important to ensure twins adhere to the FAIR Principles (Wilkinson, 2016) in that environmental assets used in the construction are readily findable and accessible and can also interoperate and be re-used across different contexts. This applies both to individual digital twins and also to richer, federated structures.

We advocate a commons-based approach to the development of an IMFe, as shown in Figure 3. We further advocate an asset commons covering a range of environmental assets including data, models, methods, and so on. This asset commons provides a common place supporting asset discovery, access, interoperability and asset re-use, tailored for a community and managed by that community for the common good. The asset commons is constructed on top of a range of standards - typically cloud standards - for data storage, communication, computation and service deployment. Services are also provided to input of data/assets into the commons and to enable results to be presented to a variety of audiences. Other important aspects of the architecture include security and governance.

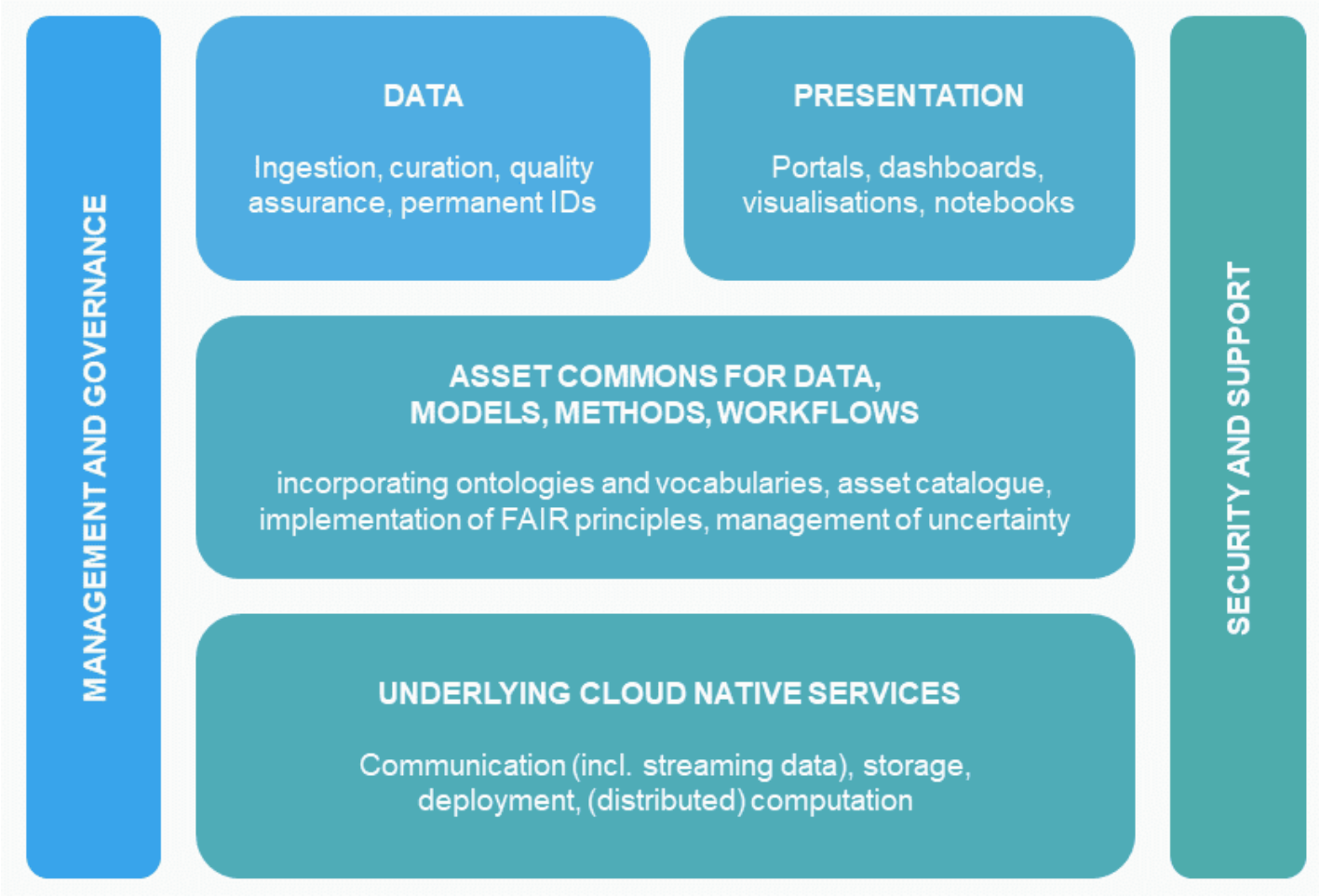


Figure 3: Proposed IMFe architecture.

In developing the commons, we further advocate the application of Grossman's end-to-end architecture for a data commons (see Figure 4. This proposes a narrow middle approach whereby the commons takes a minimalist approach to what is required at the heart of its architecture, in particular suggesting basic services to discover assets (through a cataloguing mechanism) accompanies by a means of offering permanent IDs, of being able to access the underlying resources and also offering authentication and authorisation. This minimal approach then allows a rich ecosystem of services and tools to develop around this narrow middle, supporting innovation at the edges of the architecture around this basic provision.

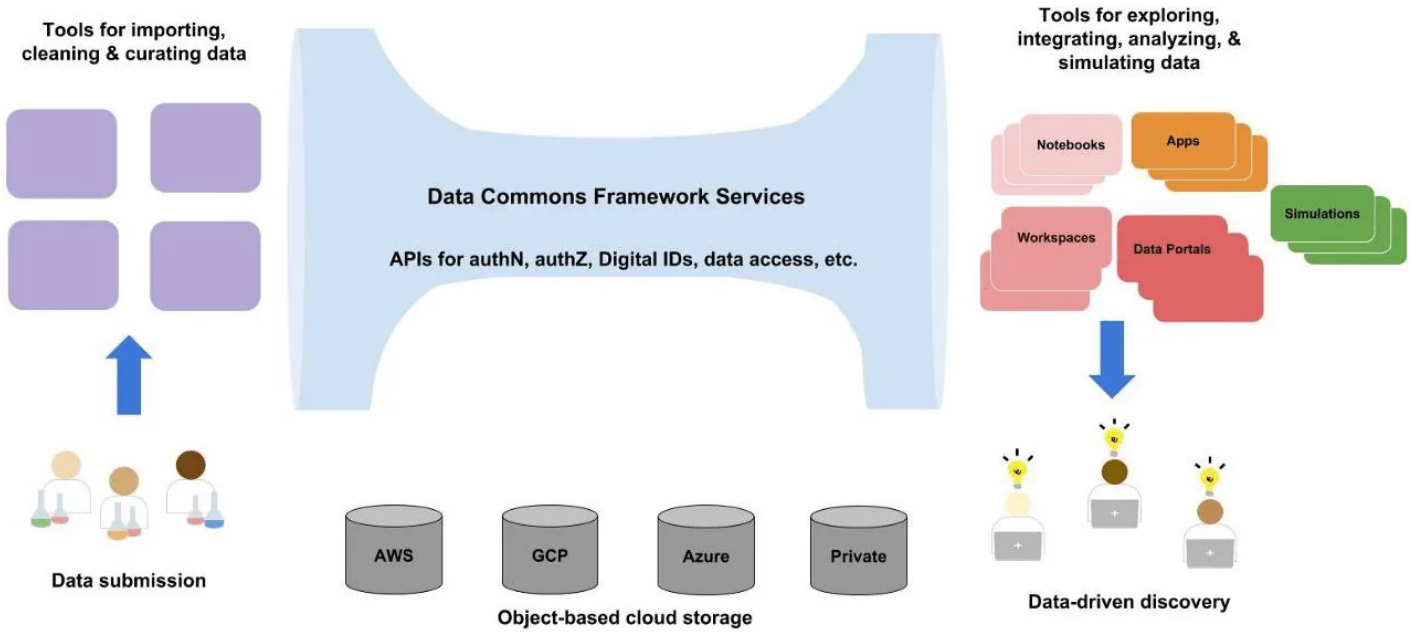


Figure 4: Grossman's end-to-end architecture for a data commons, from (Grossman, 2018).

Beyond this short summary of findings a complete list of the consolidated key Recommendation follows:

Key Recommendation 1: Clearly articulate a UK vision for the long-term ambition for digital twins of the environment.

Key Recommendation 2: Following on the successful stakeholder activities in the IMFe project there needs to be a forum to continue the senior stakeholders' engagement, and to initiate discussion on broader governance and coordination activities.

Key Recommendation 3: An IMFe development framework, with guidance on approaches to developing IMFe components, is needed and must be produced at an early stage of the IMFe development, following the approaches outlined in the IMFe project.

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.

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 the 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. A clear UK position on collaboration and international alignment is needed.

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.

Key Recommendation 10: 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 11: 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 12: Following the review of ontologies in the IMFe project, ensure that recommendations are made for ontologies as part of the IMFe governance.

Key Recommendation 13: 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 14: 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.

Key Recommendation 16: 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.

Key Recommendation 17: The asset commons, and in particular asset registers and architecture descriptions, are underdeveloped and need focussed activity to support their development. The IMFe framework must include the recommendation for digital twin developers to develop systems using, and contributing to, an asset commons.

Key Recommendation 18: The IMFe must 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.

Key Recommendation 21: 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

PILOTING THE IMFE

The IMFe is being piloted in a follow on project that commenced in October 2022 and will run until November 2023. The project includes goals to address high priority recommendations from the IMFe roadmap (Siddorn et al., 2022) including:

1. Developing the stakeholder community and exploring the potential role of communities of practice in the development of environmental digital twins

2. Iterating the and publishing the IMFe conceptual model.

3. Establishing governance principles around the development of environment digital twins

4. Building an example asset register with components to facilitate digital twin interoperability

5. Developing a marine protected areas biodiversity digital twin to evaluate the IMFe

6. Challenging the IMFe components though the use of related digital twins development to build a generalisable IMFe

The structure of the project is shown in Figure 5. Each of these goals will be briefly introduced. The project follows an agile approach to allow parallel activities, although the development of the vision, a development framework and a management and governance framework naturally come early in the process.

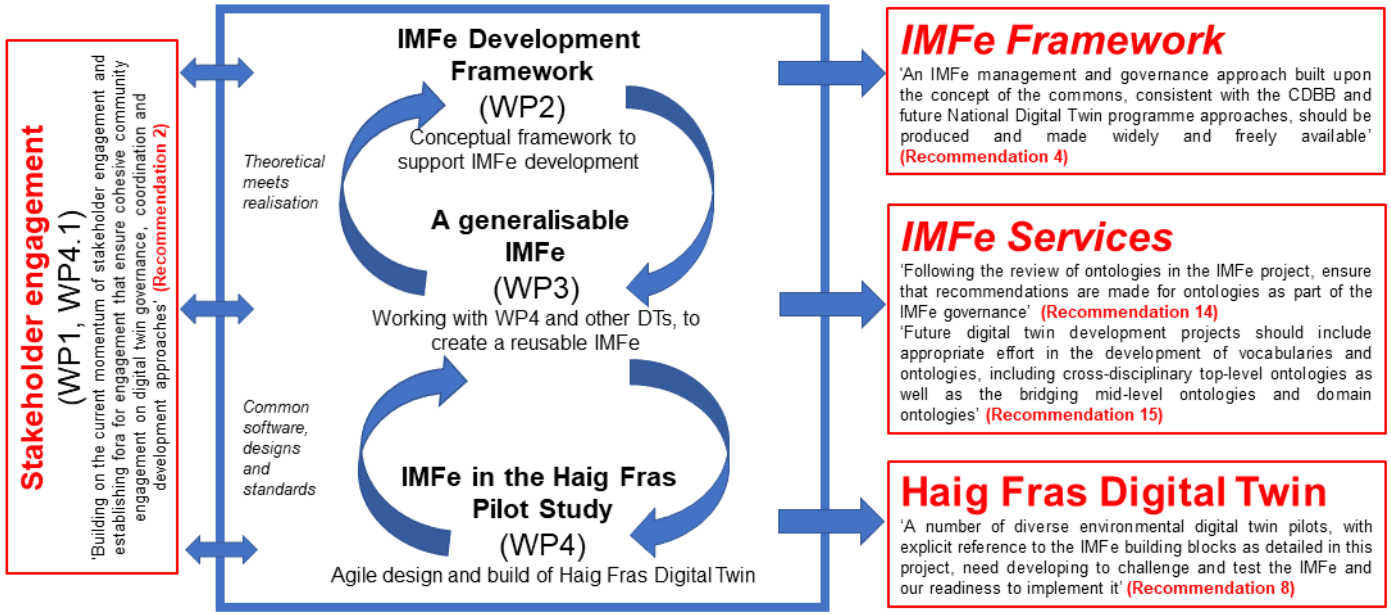


Figure 5: The project structure within the Piloting IMFe project.

Developing the stakeholder community and exploring the potential role of communities of practice in the development of environmental digital twins

Building on the stakeholder mapping and engagement activities that took place as part of the IMFe project, we will establish a Senior Stakeholder Forum for a UK Community of Digital Twin developers and users.

A community of practice will be developed to ensure digital twin practitioners have a forum for co-developing digital twins and the IMFe further. The IMFe will need to develop into a dynamic entity encompassing numerous documents, standards and techniques to cover the broad scope described in the IMFe report. These will need to be discoverable and iterable by the wider digital twins' community. Options will be explored based on open science techniques and liaison with the digital twins' stakeholder community and will integrate with existing initiatives such as the Digital Twins Hub. This will enable the initiation of a community beyond the end of the project to enable the creation of a sustainable IMFe that evolves with time with results feeding into IMFe governance development.

Iterating the and publishing the IMFe conceptual model.

In the IMFe project significant steps were taken towards delivering a conceptual framework, revolving around the concepts of an asset commons and associated asset register, coupled with the development of appropriate ontologies and vocabularies. This work will refine this concept towards a mature framework for an IMFe. The resultant output will be in the form of a framework that can be adopted and federated across domains.

Establishing governance principles around the development of environment digital twins

The previous project identified the importance of governance in the context of an IMFe and the opportunities presented by a commons approach to embed governance at the heart of the IMFe. We will implement this key recommendation and produce a governance approach tailored for the environmental domain, with an emphasis on practical approaches to ensure community management for the common good, as advocated by commons technology, whilst also delivering FAIR principles for environmental assets.

Building an example asset register with components to facilitate digital twin interoperability

Delivering environmental digital twins, which simulate whole environmental systems, is inevitably a multi-disciplinary activity. They depend upon combining components across domains including baseline data, real-time monitoring data from sensor networks, time-variant process modelling and visualisation of outputs to enable understanding by stakeholders. This is fundamentally a problem of interoperability. The solutions require both semantic tools, to enable machine readability of parameters across domains, and tools for effectively parsing data between components in a standardised way for maximum computational efficiency.

The project will develop a prototype asset register, aligned with community standards identified by the project, and populated with the components needed for the pilot study This will serve as the basis for comparison with related efforts beyond the project and for the challenging of the components later in the project.

Developing a marine protected areas (MPA) biodiversity digital twin to evaluate the IMFe

This pilot case study will build a digital twin to monitor and predict changes to benthic communities in the Haig Fras MPA. The development of the pilot will be very closely linked to the activities in across the projecrt, to ensure that the pilot provides an effective demonstration and test of the IMFe. This case study has been chosen because the diversity and volume of data needed provide a good technical challenge to the IMFe. Marine imagery is also a data type that is not well handled in most existing digital infrastructures (Durden et al., 2017). The data needed to create the twin already exist, although are not currently made widely available to decision-makers. NOC’s existing relationships with key end users and stakeholders will be utilised to ensure that the digital twin addresses real end-user needs and generates impact.

Challenging the IMFe components though the use of related digital twins development to build a generalisable IMFe

To ensure adequate development, refinement and adoption of IMFe principles, the project establishes collaboration across NERC centers to apply it as an accreditation framework. In doing so, established infrastructure intended for use in environmental Digital Twins, as well as the newly created pilot study for the Haig Fras MPA, allows assessment of the IMFe's effectiveness in ensuring realised benefits are consistent across adopters and infrastructures. This work also contrasts the IMFe as applied to newly developed and established Digital Twin component infrastructures, which is important for understanding how to assert suitability of existing assets.

In order to achieve this, the conceptual framework will be challenged through development of the pilot study, assisting it's refinement. At select stages, existing infrastructures at BAS and CEH will be reviewed in relation to the same points that underpinned the development of the pilot study, to assert their suitability as existing developments and promote further potential development paths. In addition, other functional requirements that drove those original developments in these organisations will be used to drive coverage analysis within the IMFe, filling in requirement areas that might not be fully appreciated within a fresh pilot study.

With multiple studies at varying levels of maturity, adoption and coverage driving the development of the IMFe framework, it is likely to be more resilient and comprehensive. Further ongoing developments offer new scope to then assess conclusively, achieving a reasonable assertion that concepts have been proven.

IMFE STAKEHOLDERS

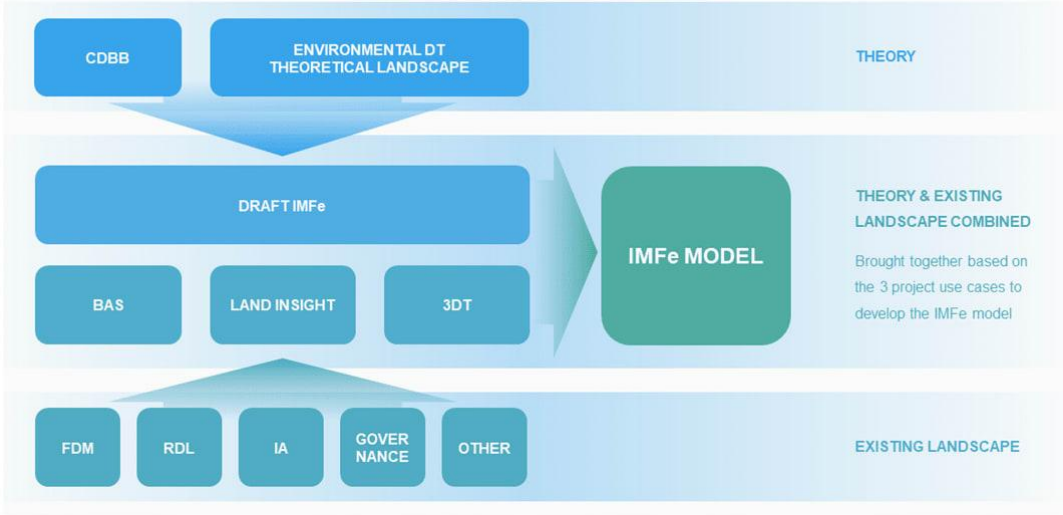
The aim of IMFe project stakeholder engagement was to deliver a stakeholder analysis of key organisations across a range of sectors with an interest in environmental digital twins, and from this analysis provide different engagement opportunities for key stakeholders to share their needs, priorities and commitments with the project team. Stakeholders from industry, government and the research sector came together for a two-hour hybrid workshop where priority use cases and recommendations for next steps were captured. In addition, the project delivered an in-person forum at senior leadership level where broad consensus was found around the need to deliver a pilot environmental digital twin to garner broader support from government and other investors.

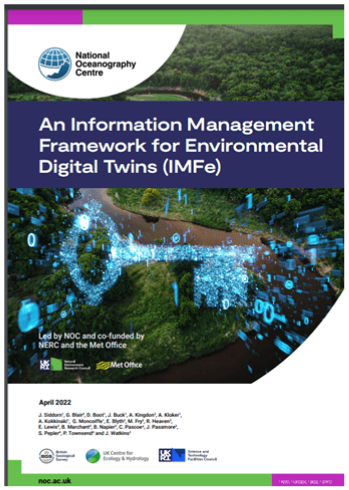
[VIDEO] https://res.cloudinary.com/amuze-interactive/image/upload/f_auto,q_auto/v1670514709/agu%20fm2022/46-aa-ec-aa-97-21-c0-56-46-5a-34-26-db-60-7f-09/image/belcher4web_udekxl.mp4

Figure 1: Prof Stephen Belcher, Chief Scientist, Met Office, chairing the March 2022 Senior Stakeholder Forum

Building IMFe digital twin stakeholder community

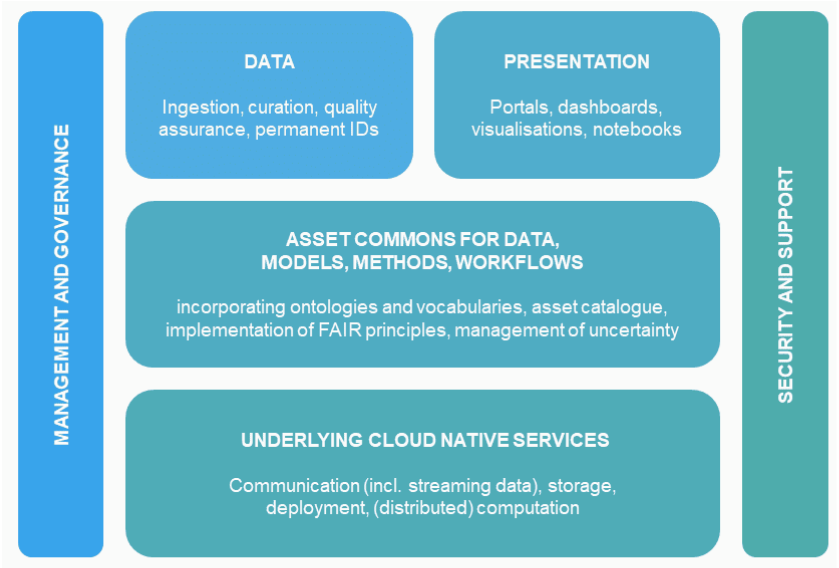
IMFe Roadmap developed by combining top-down and bottom up requirements





IMFe Roadmap report publicly available for download

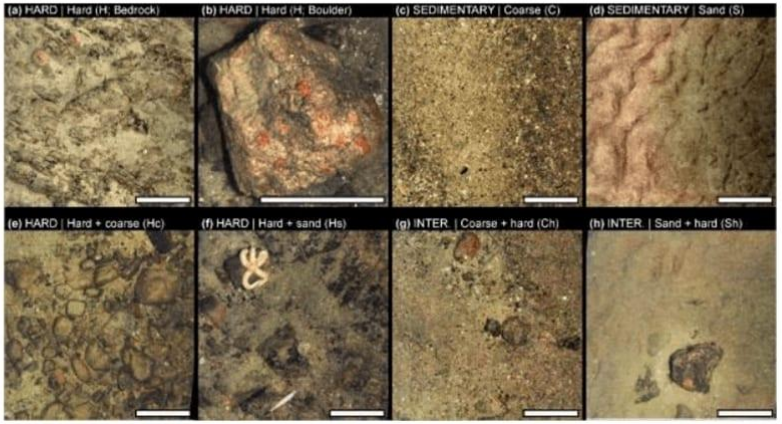
SCAN ME



Developing a conceptual model for digital twin development and interoperability

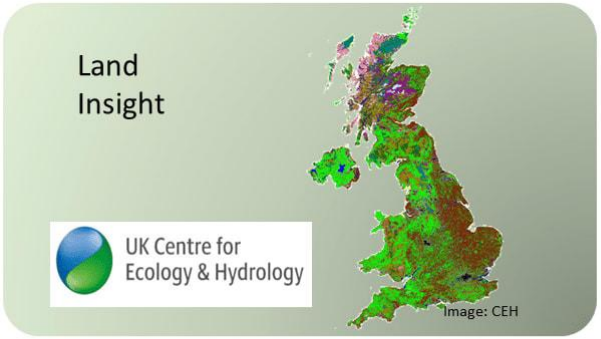
Piloting the IMFe with a marine biodiversity use case

Appendix S1 Seabed-type classification



From Benoist et al., (2019). Monitoring mosaic biotopes in a marine conservation zone by autonomous underwater vehicle. *Conservation Biology*, *33*(5), 1174–1186. https://doi.org/10.1111/COBI.13312

Challenging the IMFe with other environmental use cases



ENGAGING WITH IMFE

Broad engagement with the digital twin community is a crucial element of developing and IMFe that is fit for purpose, align with related efforts, and extensible beyond the use cases considered within the project. As such, we are keen to engage with other global efforts that will further the project goals. Our contact details are on our project website or you can engage with any of the project participants that span the environmental research and informatics domains.

Piloting an Information Management Framework for Environmental Digital...

Roadmap Developing the information...

noc.ac.uk

SCAN ME

ABSTRACT

Environmental science is primarily concerned with assessing the impacts of changing environmental conditions on the state of the natural world. The UK Natural Environment Research Council (NERC) has recently published its digital strategy, 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 UK 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. This strategy places data and digital technologies at the heart of UK environmental science. One such set of technologies are digital twins. Environmental digital twins 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 realize the value of environmental digital twins they need to be developed following agreed standards where interoperability is required to make sure the information can be trusted by the user, and so that information can be shared between digital twins.

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. We present recommendations for the development of an IMF for Environmental digital twins (IMFe) including the application of concepts such as an asset commons and balanced approach to standards to facilitate minimum interoperability requirements between twins while iteratively implementing an IMFe.

[1]https://www.ukri.org/publications/natural-environment-research-council-nerc-digital-strategy-2021-2030/ (https://www.ukri.org/publications/natural-environment-research-council-nerc-digital-strategy-2021-2030/)

[2] https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/research/approach/r-i_strategy_full_version_v2.pdf (https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/research/approach/r-i_strategy_full_version_v2.pdf)

REFERENCES

CDBB, 2020, The pathway towards an Information Management Framework: A ‘Commons’ for Digital Built Britain, available from: https://www.cdbb.cam.ac.uk/what-we-do/national-digital-twin-programme/pathway-towards-information-management-framework [Accessed 28th April 2022]

Durden et al. (2017). Integrating "Big Data" into Aquatic Ecology: Challenges and Opportunities. Limnology and Oceanography: Bulletin, 26, 101-108.

Grossman, R., 2018, A Proposed End-to-End Principle for Data Commons, available from: https://medium.com/@rgrossman1/a-proposed-end-to-end-principle-for-data-commons-5872f2fa8a47 [Accessed 29th April 2022]

Siddorn, John, Blair, Gordon Shaw, Boot, David, Buck, Justin James Henry, Kingdon, Andrew, Kloker, Alice, Kokkinaki, Alexandra, Moncoiffé, Gwenaëlle, Blyth, Eleanor, Fry, Matt, Heaven, Rachel, Lewis, Edward, Marchant, Benjamin, Napier, Bruce, Pascoe, Charlotte, Passmore, James, Pepler, Sam, Townsend, Poppy, & Watkins, John. (2022). An Information Management Framework for Environmental Digital Twins (IMFe). Zenodo. https://doi.org/10.5281/zenodo.7004351

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18