

The Data Journey

Managing information and data across federated digital twins of the ocean

Pier Luigi Buttigieg
(HMC // GEOMAR // AWI)

A foundation: digitised data, information, and knowledge in 4D

“Data is a precious thing and will last longer than the systems themselves.”

— Tim Berners-Lee, A Framework for Web Science

“Metadata is a love note to the future”

— Jason Scott



The journey of our digital heritage is non-linear and will never end

As we build data management solutions for twins, avatars, and dashboards, we must engineer for our immediate needs as well as for an unknown and uncertain future

**Rock paintings from the Cave of Beasts (Gulf Kebir, Libyan Desert) Estimated 7000 BP
Clemens Schmillen - Own work**

1

Data: representational signs, symbols, about some entity

2

Information: Data which reduces uncertainty about some entity (i.e. is informative)

3

Knowledge: Familiarity and understanding of an entity, built from trustworthy information, enabling reasoning

Robust twinning starts with accurate, high-density digitisation across this spectrum

(and a clear accounting of provenance and bias)

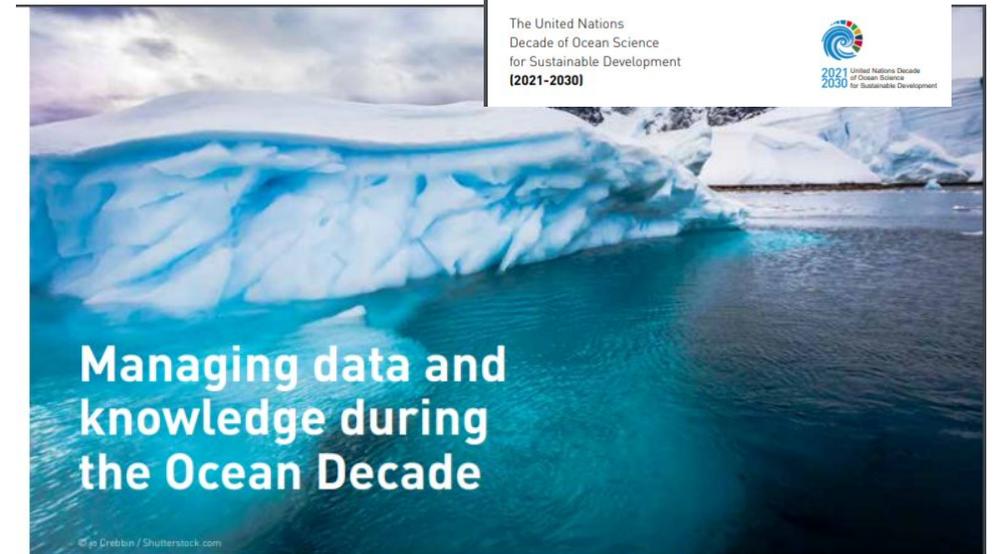
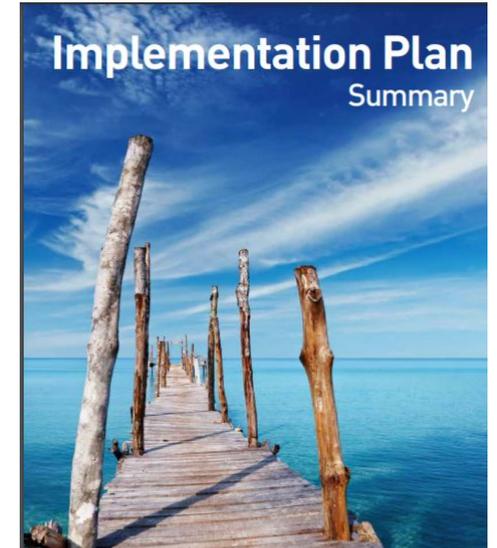
Robust twinning starts with accurate, high-density digitisation across this spectrum

(and a clear accounting of provenance and bias)

- Twin-enabling digitisation and data management will accurately represent (and therefore respect) the real-world processes and entities they are about
- We thus need to know – as comprehensively as possible through stakeholder engagement – what we’re trying to twin and for what purpose / community
 - With some exceptions, most twins will start out as avatars or simply structured digital shadows or digital exhaust – and that’s fine
 - The progression from a rough or distorted twin to a refined one is a

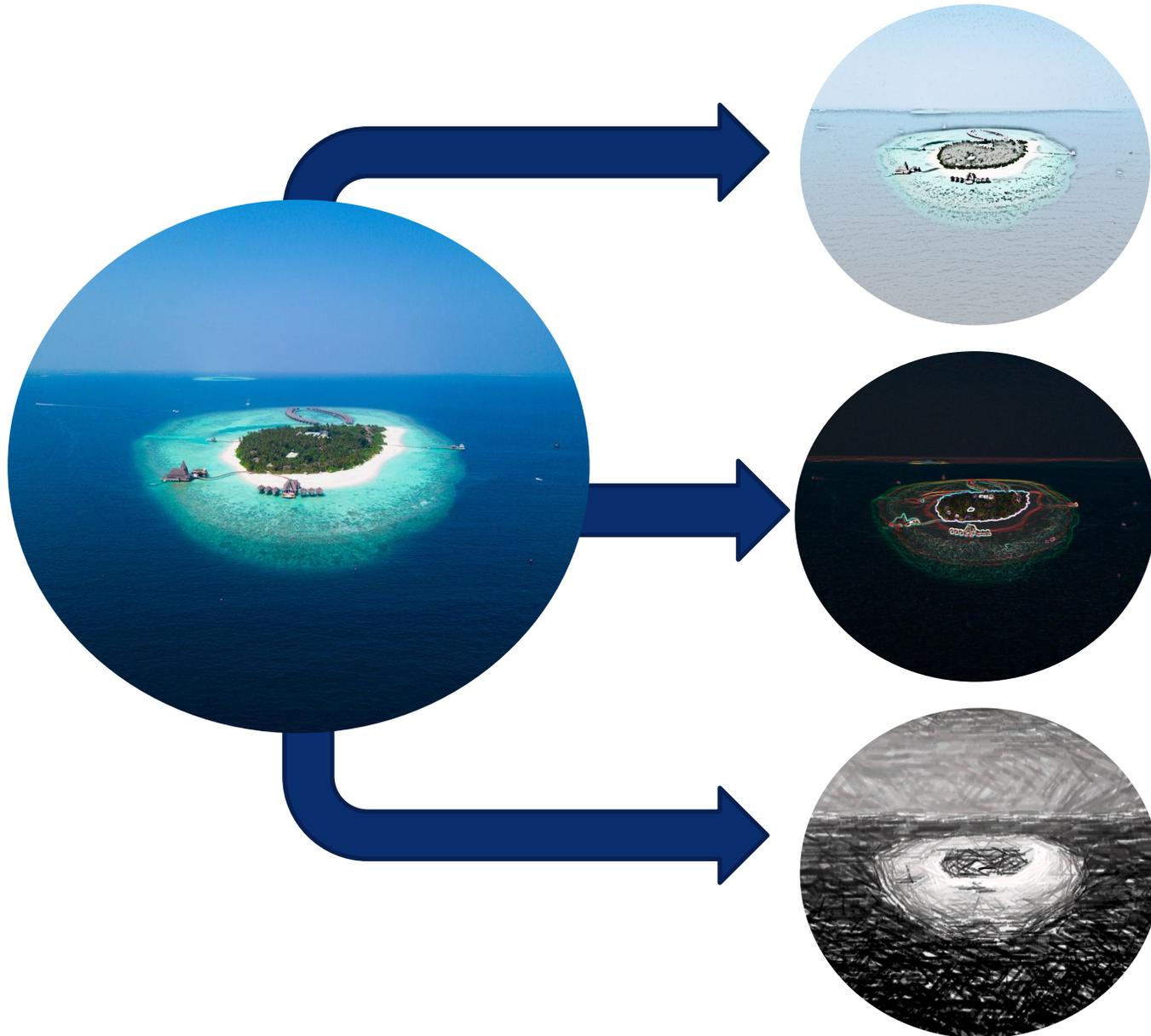
The Decade Implementation Plan

- Since 2018, over 2000 stakeholders have been consulted to develop the Ocean Decade Implementation Plan
- Two peer review processes generated nearly 300 submissions on the draft Plan from IOC Member States, UN agencies, and other stakeholders.
- **Evolving this plan with outputs and consensus from summits (such as this one) is essential to stabilise *global* federation**

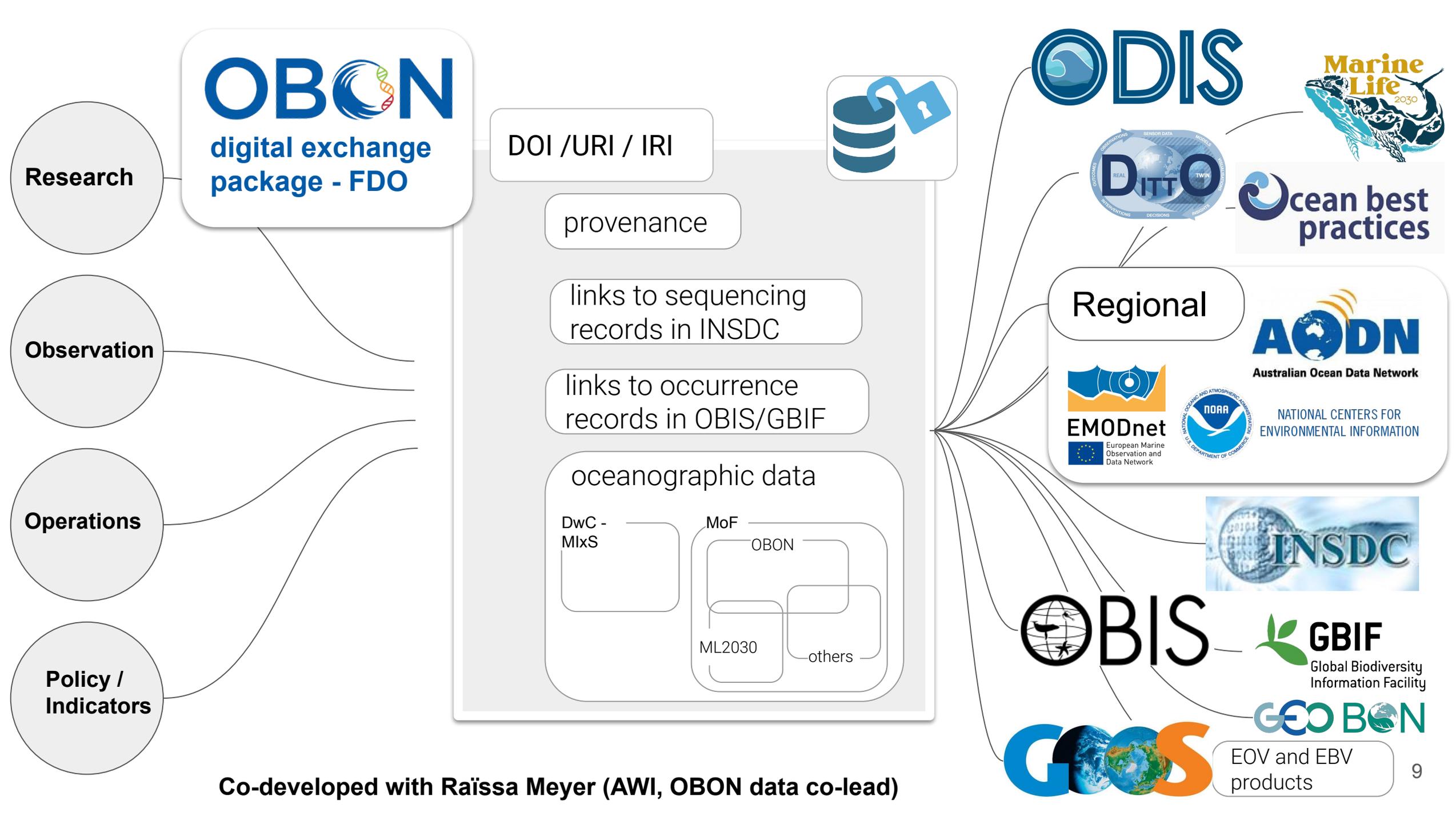


<https://ioc.unesco.org/ocean-decade>

Prepare for diversity



- Different stakeholders will find different data and model outputs (mis)informative or usable
- Knowledge and societal value systems vary, and twin-bound (meta)data will be influenced by this
- Data management for federated twins relies on **true** interoperability of all source (meta)data and careful modularisation

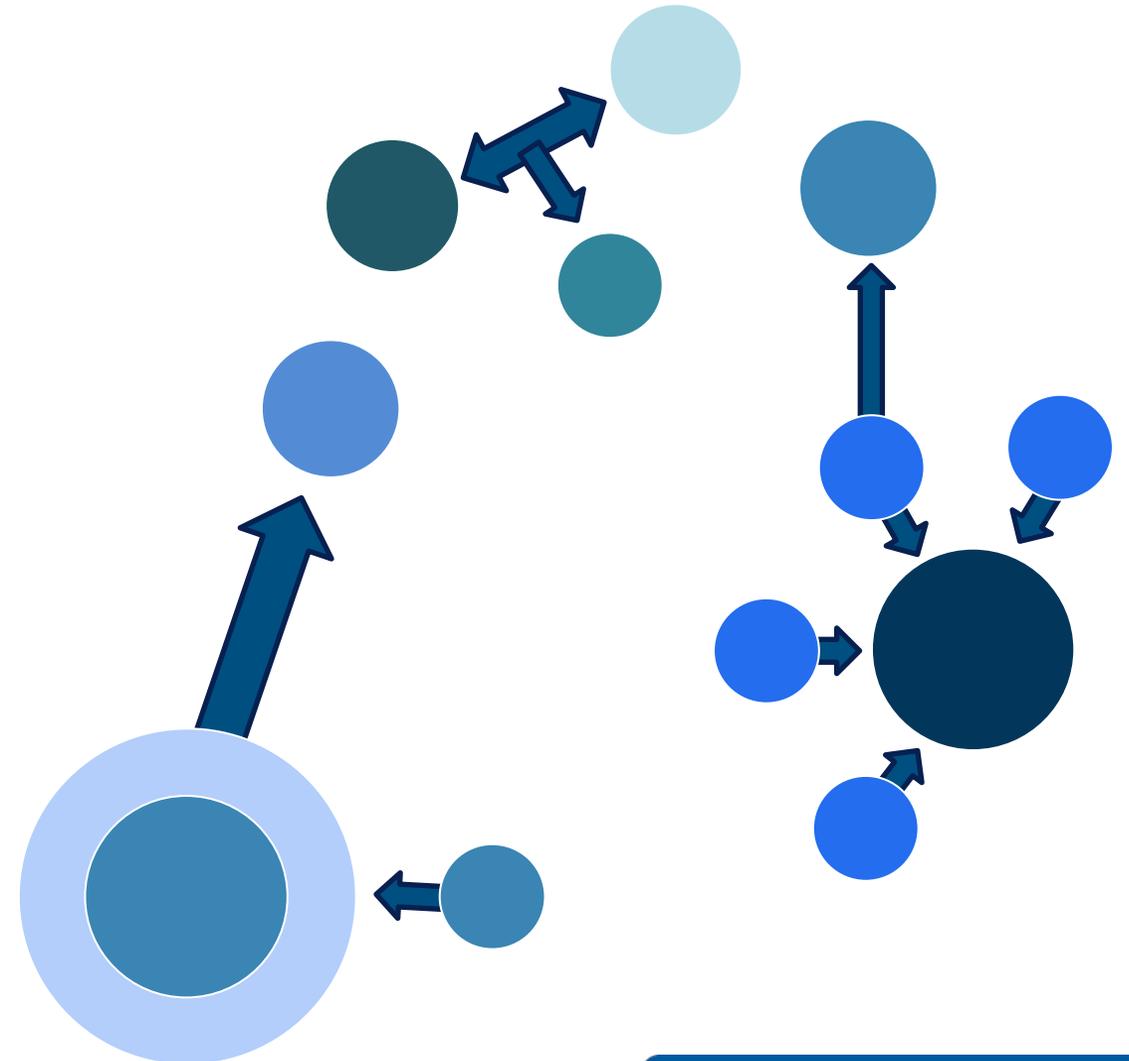


- High-level intent and implementation must match and scale (e.g. FAIR for whom?)
- Collaboratively identify what needs to be twinned and incentivise rich, dense (meta)data collection
- Atomise (meta)data with grim abandon – prepare for an unknown future
- Manage (meta)data and tooling/technology independently – avoid lock ins at all cost
- Be extremely clear about provenance and bias
- Layers of abstraction are needed
 - Engineer, develop, and manage digital assets in the most sensible way for your local context and create multiple routines to automatically *project* data/information/knowledge compliant

Embedding twins in digital ecosystems

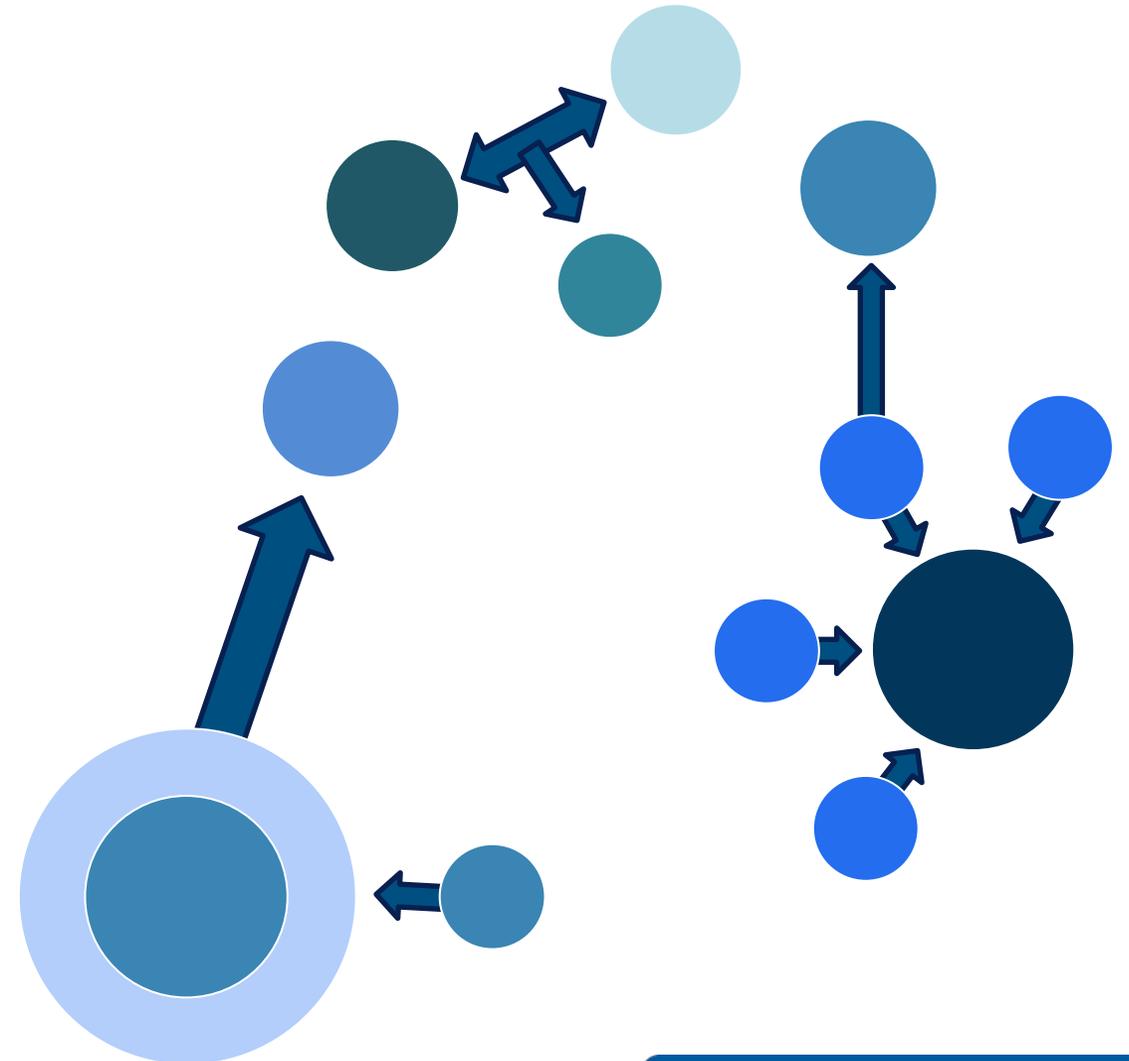
digital ecosystem

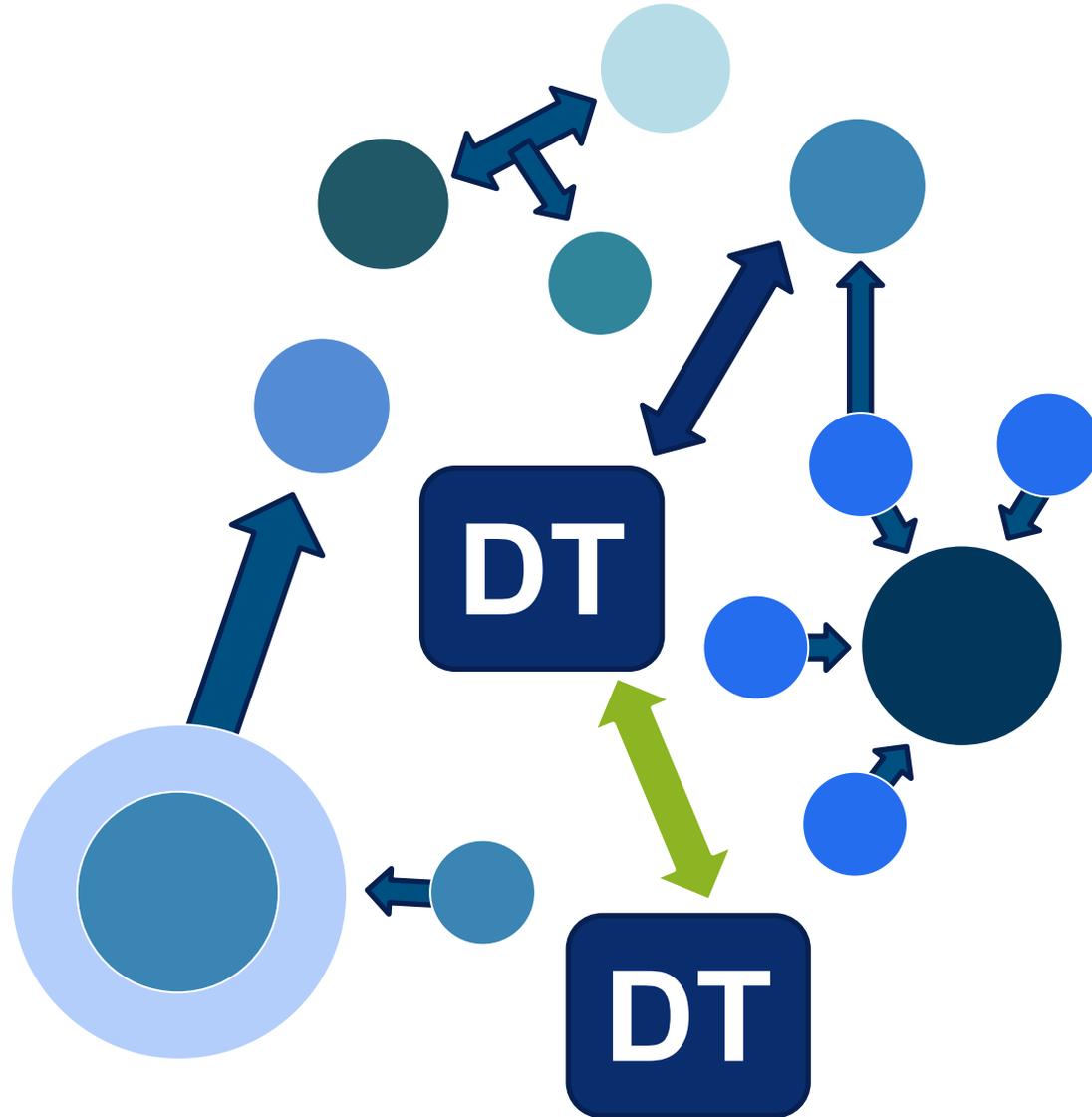
A digital ecosystem is a distributed, adaptive, open **socio-technical system** with properties of self-organisation, scalability and sustainability inspired from natural ecosystems.



digital ecosystem

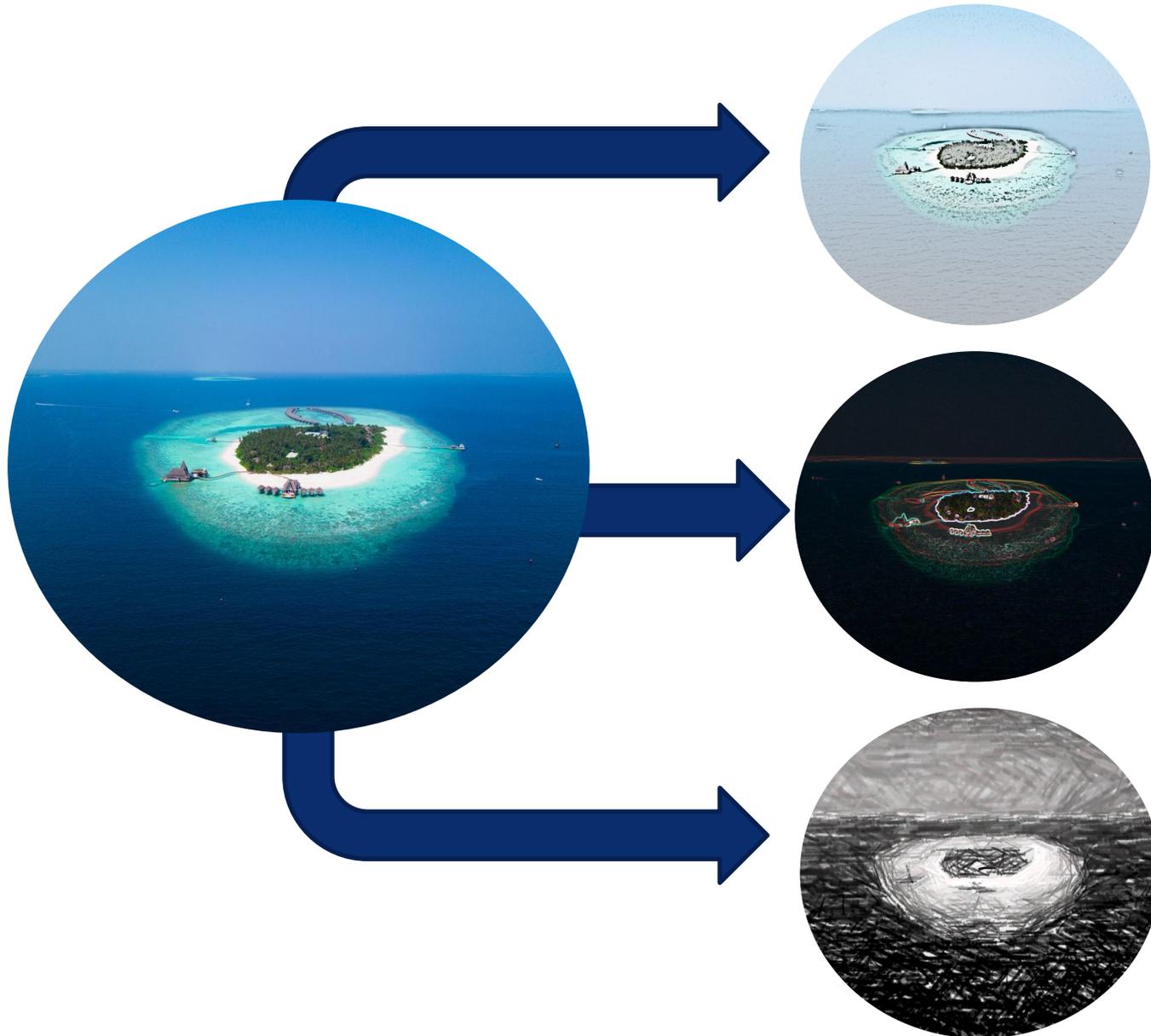
- The densest parts of a digital ecosystem which are capable of tracking a real-world entity through time are the most suited to twinning (autodetection potentials)
- Robust and sustainable twins will harvest their raw materials from *independent* (and secure) digital ecosystems
- Those digital ecosystems should be transparent and as open as possible to achieve the Decade's goals





- Digital twins will become new (hyper)nodes in digital ecosystems, managing the disruption (positive and negative) is key
- How they will communicate with other nodes will define novel data management norms
- Their robustness and sustainability will depend on that of the subsystems they rely on, and how well embedded their communities and resourcing plans are

Prepare for (more) diversity



- Different stakeholders will build twins/avatars emphasising their concerns
- Authoritative twins will need auditing routines, higher transparency, and rigorous QC
- Detecting and analysing (dis)agreements between interoperable twins can be the route to great things via ensemble approaches

Requirement /
need
specifications

Twin-focused
Interoperability
frameworks

Meta-indices fo
component
discovery

The socio-technical challenge

Cultivating a global digital culture – from strategy to implementation – to create robust digital ecosystems capable of supporting and integrating digital twins.

The technologies exist, but the global culture is just dawning

Harmonised
digital standards
(incl. regulatory)

Diagnostic
routines

AI-enhanced
cross-twin
discovery

Our global success will be defined by how stable and interchangeable these components are between independent DTs

**“...there ain't no journey what
don't change you some.”**

— David Mitchell, Cloud Atlas