Transforming Data into Information for Societal Benefit: An Illustrated Overview of Interoperability Approaches

Limitations. This document was specifically prepared as a follow-up to the generous invitation by the PCAST to provide an overview of NEON. The materials herein, unless otherwise noted, have not undergone rigorous peer review, and does not necessarily represent the views of the National Science Foundation nor any of the other Federal agencies mentioned herein.

Introduction. The accelerating rate of large-scale environmental changes and the concomitant eco-socio-economic impacts places a premium on the ability to efficiently and effectively transform high-quality, credible scientific data into information and knowledge for societal benefit. There exists a class of civil earth observing systems, like those operated by NOAA, that were specifically commissioned for operational purposes. Other systems, built specifically for research purposes, produce high-quality environmental data that may be repurposed for operational needs.

Regardless of the original intent under which these infrastructure were commissioned, are there ways to leverage existing tax-payer investments to inform how the nation should manage its environmental capital? What steps can be taken to facilitate the repurposing of research data and information for operational purposes? How should we assess gaps in our environmental observation infrastructure?

This document is an compilation of one-page overviews that may provide parts of the answer to such questions. These overviews represent an emergent picture of possible solutions borne out of, and informed by, interactions between NEON and our Federal and academic constituents.

Figure 1 is the launching point for these overviews. One-pagers corresponding to selected elements of Figure 1 are indicated by the yellow circles with red numerals. For example, 2 is an indicator that more details can be found on page 2.

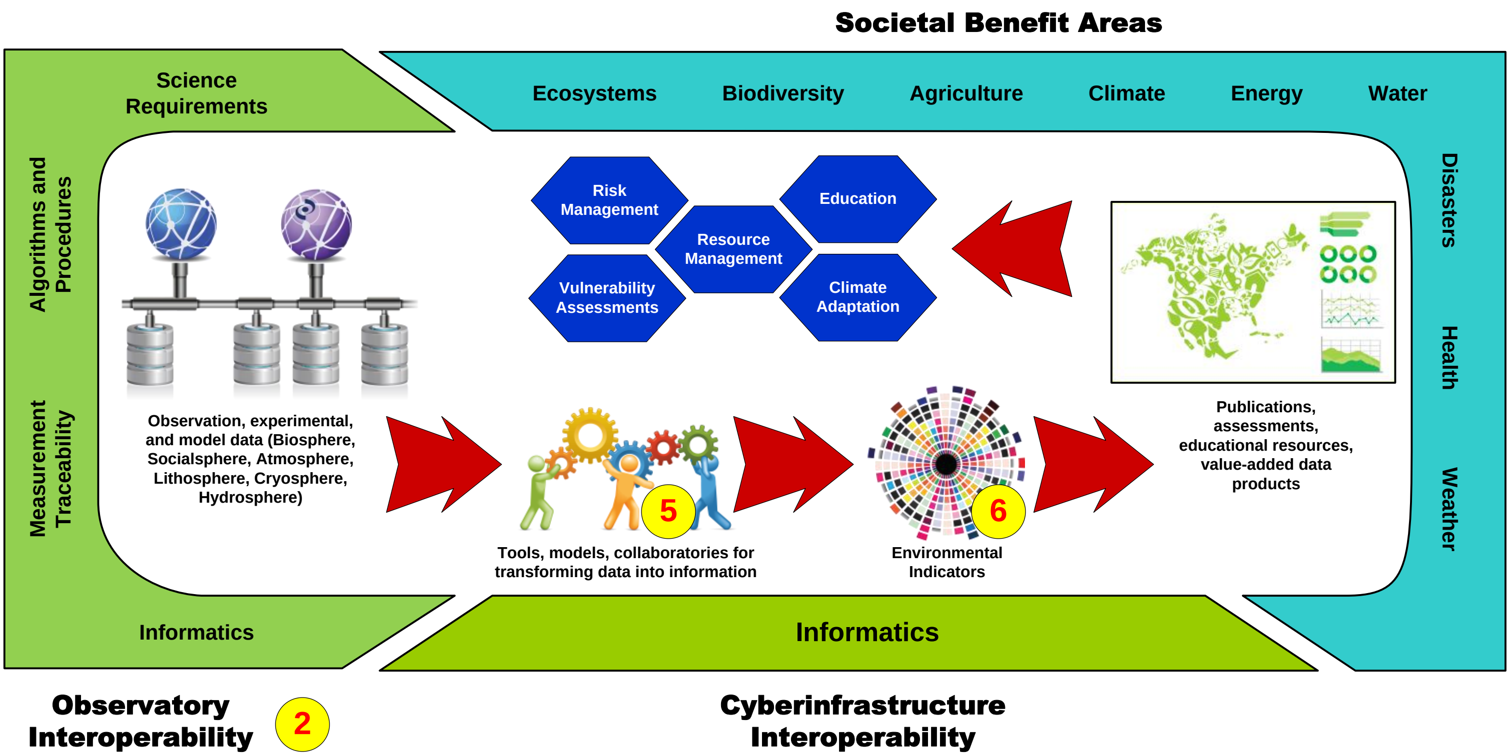


Figure 1: Transforming data into information for societal benefit

The National Ecological Observatory Network is a project sponsored by the National Science Foundation and managed under cooperative agreement by NEON, Inc. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation. For more information, please contact: Brian Wee (bwee@neoninc.org); or Hank Loescher (hloescher@neoninc.org).

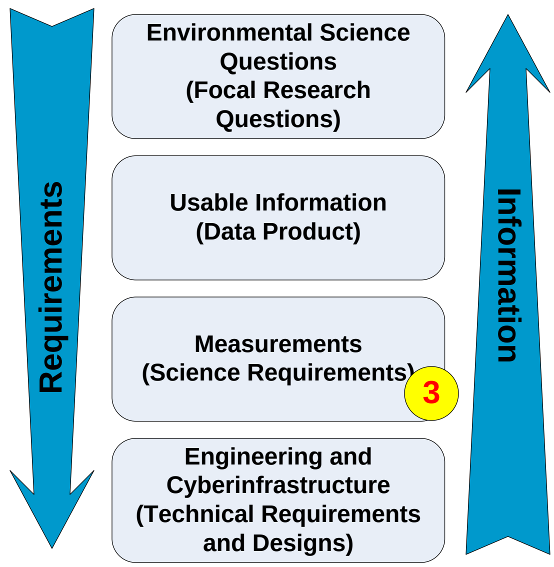
The Interoperability Fabric and Requirements Driven Design

The Interoperability "Fabric" that NEON uses to structure its interaction with its partners is modeled closely after the requirements driven approach (Figure 4) used to design major scientific infrastructure like NEON . The "fabric", if successfully implemented, binds participating environmental observatories along clearly defined interfaces to seamlessly deliver data and information to the public.

The components of the interoperability fabric are:

- **Distillation of Science Questions and Hypotheses into Requirements**
  - Mapping questions to 'what must be done'
  - Defining joint science scope
  - Defining interfaces between subsystems
- **Algorithms/Protocols**
  - Performing intercomparisons to quantify relative uncertainties
  - Understanding sources of biases
- **Traceability of Measurements**
  - Using recognized standards
  - Traceability to recognized standards, or first principles
  - Ascertaining signal to noise ratio
  - Managing QA/QC
  - Quantifying uncertainty budgets
- **Informatics**
  - Standards for data formats, metadata, web services, provenance, identifiers
  - Ontologies and controlled vocabularies
  - Data licensing, policies, legal constraints
  - Authentication, identity management, access management

A requirements-driven design starts with high-level science questions or criteria. The ensuing structured decomposition successively yields detailed science requirements, system and subsystem design and their implementation, and eventually measurements and data products.



This process facilitates traceability of design decisions, and is a useful tool to identify implementation priorities.

Figure 4: Requirements Driven Design Methodology

Strategies for Enabling Measurement Interoperability

Co-location of Environmental Observatories (EOs), as shown in Figure 5, is one way to ameliorate interoperability challenges. In this hypothetical example, NEON and the USDA Long-term Agro-Ecosystem Research (LTAR) network are co-located at sites A, B, and C. (NEON and LTAR are currently co-located at three locations.)

It is desirable to implement shared measurements between NEON and LTAR, regardless of whether those are at co-located sites or agro-ecologically distinct sites. Strategically selected shared measurements extend an EO's measurement footprint into another EO's physical footprint. As shown in Figure 5, an LTAR measurement may be estimated at a NEON site even though there is no actual LTAR measurement observed at that site.

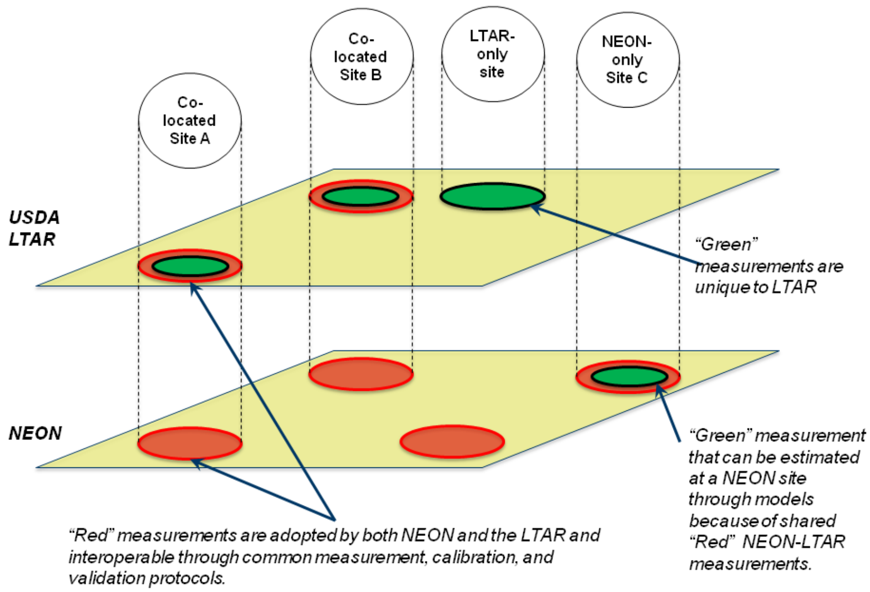


Figure 5: A hypothetical "stacked EO configuration" that enables an EO's measurement footprint to be extended into another EO.

Shared measurements do not necessarily mean identical sensors, sampling regimes, or protocols. For example, precipitation may be measured by different sensors in NEON and LTAR, but they are considered a "shared measurement" if the measurements are interoperable through common calibration, validation, and audit procedures such that their respective uncertain budgets can be quantified.

This enables variables between EOs to be fused ("data fusion", as opposed to "data integration") using statistical procedures, as a way to extend the observation footprint of either EOs.

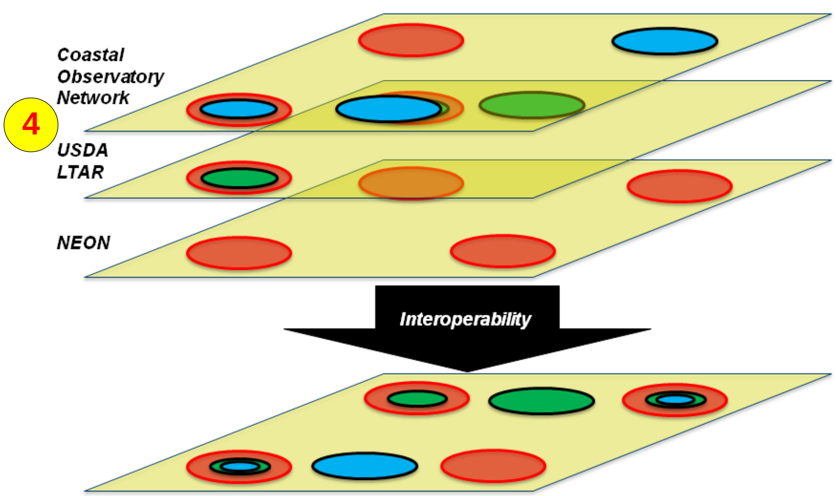


Figure 6: Seamless, virtual EO enabled through interoperability by collapsing the EO stack.

The same concept may be extended to other EOs, for example, a distributed coastal observatory. Figure 6 shows a "stacked EO" configuration, where concepts of shared measurements and site co-location are implemented through a requirements-driven process.

Depending on the degree of interoperability, collapsing the EO stack results in the seamless integration of data and information. The data consumer sees the portfolio of interoperable EOs as a single, seamless, virtual EO.

Leveraging Federal Investments: Agricultural and Coastal Observatories

In light of the challenges facing agriculture over the next few decades, USDA and NEON leaders have been exchanging information on strategies for leveraging existing investments. Discussions have focused on the establishment of partnerships and the sharing of techniques, protocols, best practices, and physical infrastructure.

In late 2012, the USDA launched its **Long-Term Agro-Ecosystem Research (LTAR)** network with an initial configuration of ten sites, three of which are co-located with NEON. The objectives of the LTAR are to enable the better understanding of:

- How key agricultural system components interact at larger scales (e.g., watershed; landscape);
- How to forecast the environmental effects of shifting agricultural practices;
- How to improve the efficacy and information management of conservation programs;
- How to identify the broader societal benefits of modern agriculture (e.g., bio-energy production; carbon sequestration; improved water quality & water-use efficiency; wildlife habitat).

NEON and NOAA have been exchanging ideas on approaches to **integrate terrestrial and coastal observations**.

- In 2010, more than 39% of Americans live in coastal shoreline counties. These counties represent less than 10% of the US Land area, but are responsible for over half of the 2011 US GDP (State of the Coast, NOAA).
- Multiple stressors are already being experienced by coastal and near-shore ecosystems that will be exacerbated by climate change and ocean acidification (PCAST 2011, Burkett and Davidson 2012).

The connectivity between terrestrial and near-coastal systems is poorly understood and affect ecosystem services, transportation of nutrients, biodiversity, and ecosystem resilience. These ultimately impact the economic vitality of coastal communities.

Mapping the Transformation Process Against Products, Cyberinfrastructure, and Stakeholders

Figure 7 depicts a framework often used to describe the transformation of **D**ata into **I**nformation and **K**nowledge. Figure 8 utilizes this framework to demonstrate the relationship between data and value-added products that are often used for scientific, resource-management, and policy purposes.

The enabling cyberinfrastructure corresponds largely to the "Tools, Models, Collaboratories" depicted in Figure 1. The stakeholders represent the beneficiaries of the process.

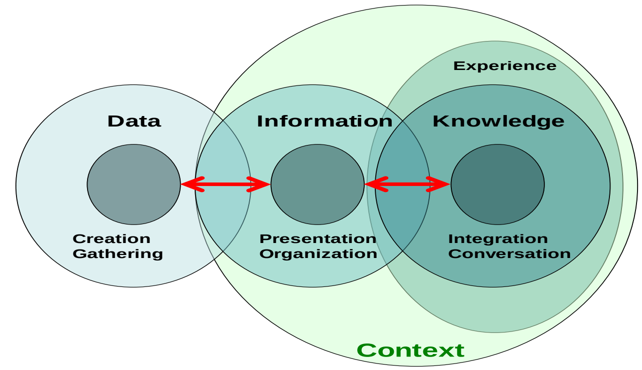


Figure 7: The DIK Framework. Courtesy of Peter Fox, Rensselaer Polytechnic Institute.

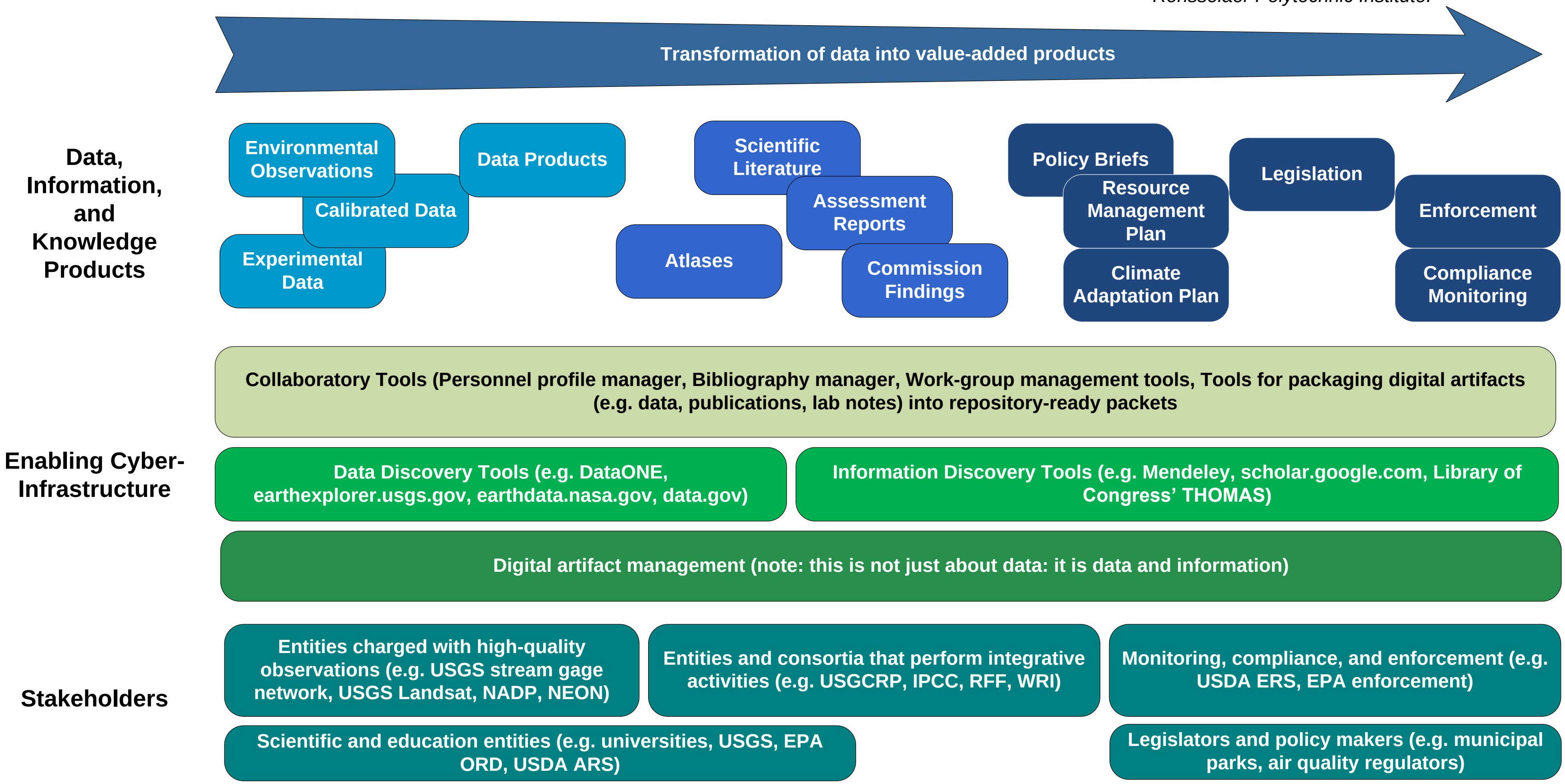


Figure 8: Information and knowledge products mapped against the cyberinfrastructure and stakeholders.

Environmental Indicators as a Means to Foster Coordination between Observatories

Indicators are designed to inform us quickly about something of importance, and represent important features of the status, trends, or performance of a system of interest (e.g. the economy, agriculture, air quality) (Janetos et. al. 2012). Monitoring them over time can help determine whether problems are developing and, if so, which interventions might effectively alleviate them (Orians G, oral presentation, SERDP conference).

The US Global Change Research Program's (USGCRP) third National Climate Assessment (NCA) will create a system of indicators to help inform policy-makers and citizens understand key impacts of the changing climate (USGCRP NCA website). USGCRP agricultural indicators may include workable field days during growing season, crop distribution maps, pest distribution maps, disaster and crop insurance payments (Janetos et al 2013).

Another such example of an indicator is phenology, which is a robust indicator of the impacts of climate change on biodiversity and ecosystem processes (Rosenzweig et al 2007, Karl et al 2009). The concept map below reflects a selected subset of relationships between domestic and international entities that are relevant to phenology observations. Indicators that are based on Essential Climate Variables (ECVs) or Essential Biodiversity Variables (EBVs) are a useful mechanism to prioritize and structure interoperability initiatives between Environmental Observatories.

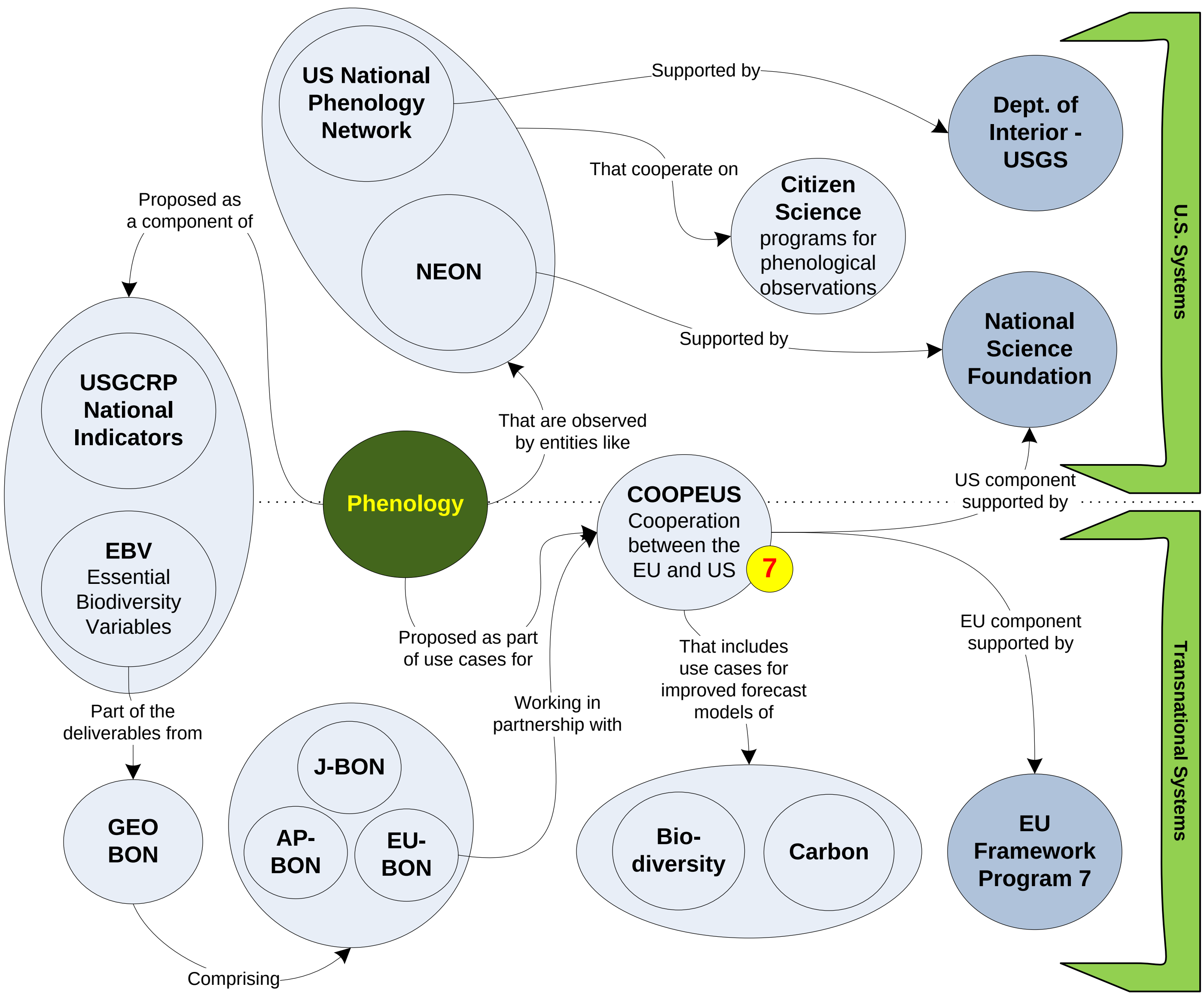


Figure 9: Phenology as an exemplar indicator for coordinating interoperability initiatives between environmental observatories.

COOPEUS – Connecting Environmental Observatories between the EU and the US

The COOPEUS project is a US-EU funded project to strengthen the cooperation between the EU and the US in the field of environmental research infrastructures. Europe's major environmental related research infrastructure projects involved include EISCAT (space weather), EPOS (solid earth dynamics), EMSO (ocean observatories), LifeWATCH (biodiversity), and ICOS (carbon observatories), with their corresponding US counterparts that are responsible for the NSF funded projects AMISR, EARTHSCOPE, Oceans Observatories Initiative, and NEON.

The intention is that by interlinking these activities new synergies are generated that will stimulate the creation of a truly global integration of existing infrastructures. The key of this integration process will be the efficient access to and the open sharing of data and information produced by the environmental research infrastructures. Trends in this area include growing collaborations between computer and environmental scientists, leading to the emergence of a new class of scientific activity structured around networked access to observational information.

Figure 10 depicts the project organization of the COOPEUS project. Each work package is defined by detailed subtasks. Progress reports are submitted regularly to the EU and US funding agencies. COOPEUS project personnel met most recently in late September 2013 at the NEON headquarters in Boulder, CO.

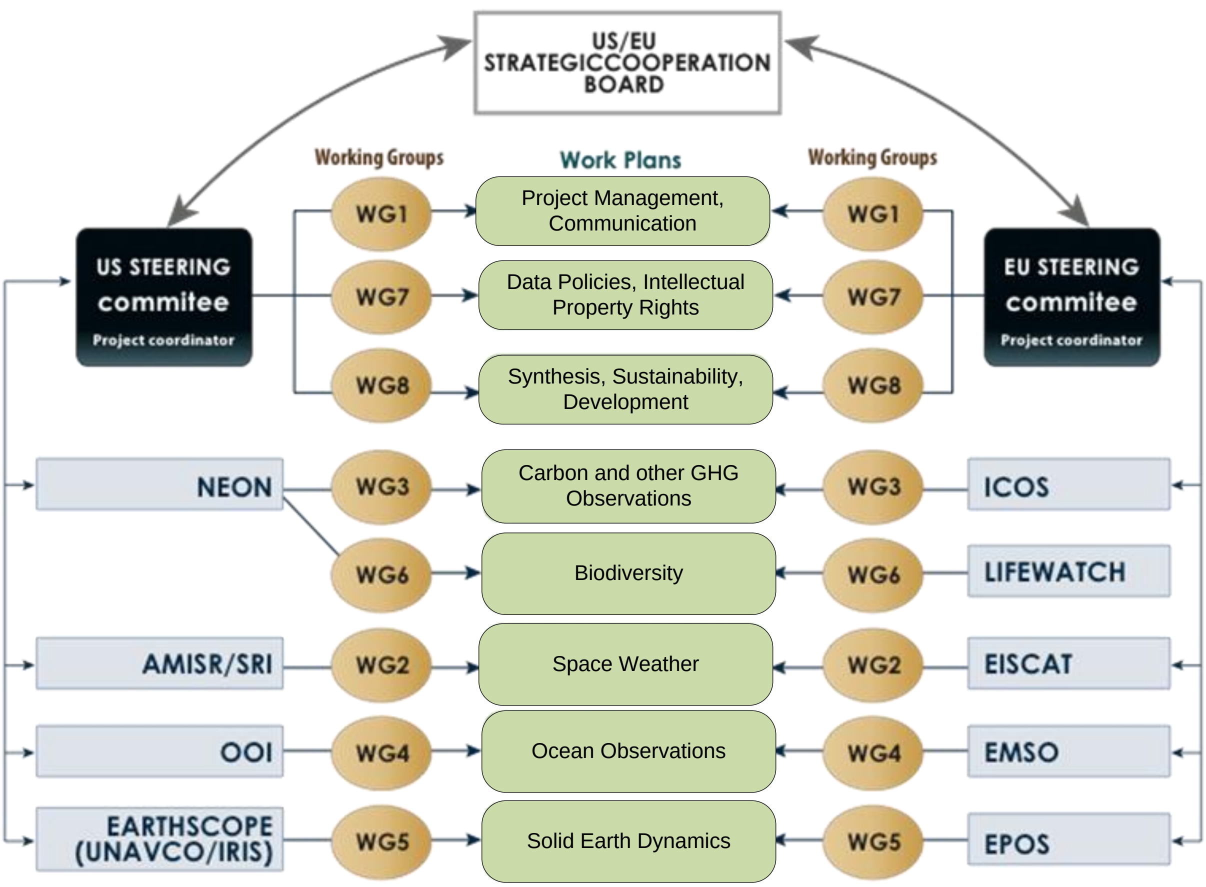


Figure 10: COOPEUS project organization.