

Project Report

White paper on the alignment and interoperability between the Distributed System of Scientific Collections (DiSSCo) and EU infrastructures - The case of the European Environment Agency (EEA)

Niels Raes^{‡,§}, Ana Casinol, Hilary Goodson[¶], Sharif Islam^{§,#}, Dimitrios Koureas^{§,α},
Edmund K. Schiller[«], Leif Schulman[»], Laura Tilley^l, Tim Robertson[¶]

‡ NLBIF, Leiden, Netherlands

§ Naturalis Biodiversity Center, Leiden, Netherlands

| CETAF, Brussels, Belgium

¶ Global Biodiversity Information Facility, Copenhagen, Denmark

DiSSCo, Leiden, Netherlands

α Distributed System of Scientific Collections - DiSSCo, Leiden, Netherlands

« Natural History Museum Wien, Vienna, Austria

» LUOMOS, Helsinki, Finland

Corresponding author: Niels Raes (nlbif@naturalis.nl)

Reviewable

v 1

Received: 22 Dec 2020 | Published: 30 Dec 2020

Citation: Raes N, Casin A, Goodson H, Islam S, Koureas D, Schiller EK, Schulman L, Tilley L, Robertson T (2020) White paper on the alignment and interoperability between the Distributed System of Scientific Collections (DiSSCo) and EU infrastructures - The case of the European Environment Agency (EEA). Research Ideas and Outcomes 6: e62361. <https://doi.org/10.3897/rio.6.e62361>

Abstract

The Distributed System of Scientific Collections (DiSSCo) Research Infrastructure (RI) is presently in its preparatory phase. DiSSCo is developing a new distributed RI to operate as a one-stop-shop for the envisaged European Natural Science Collection (NSC) and all its derived information. Through mass digitisation, DiSSCo will transform the fragmented landscape of NSCs, including an estimated 1.5 billion specimens, into an integrated knowledge base that will provide interconnected evidence of the natural world. Data derived from European NSCs underpin countless discoveries and innovations, including tens of thousands of scholarly publications and official reports annually (supporting legislative and regulatory processes on sustainability, environmental change, land use,

societal infrastructure, health, food, security, etc.); base-line biodiversity data; inventions and products essential to bio-economy; databases, maps and descriptions of scientific observations; educational material for students; and instructive and informative resources for the public. To expand the user community, DiSSCo will strengthen capacity building across Europe for maximum engagement of stakeholders in the biodiversity-related field and beyond, including industry and the private sector, but also policy-driving entities. Hence, it is opportune to reach out to relevant stakeholders in the European environmental policy domain represented by the European Environment Agency (EEA). The EEA aims to support sustainable development by helping to achieve significant and measurable improvement in Europe's environment, through the provision of timely, targeted, relevant and reliable information to policy-making agents and the public. The EEA provides information through the European Environment Information and Observation System (Eionet). The aim of this white paper is to open the discussion between DiSSCo and the EEA and identify the common service interests that are relevant for the European environmental policy domain. The first section describes the significance of (digital) Natural Science Collections (NHCs). Section two describes the DiSSCo programme with all DiSSCo aligned projects. Section three provides background information on the EEA and the biodiversity infrastructures that are developed and maintained by the EEA. The fourth section illustrates a number of use cases where the DiSSCo consortium sees opportunities for interaction between the DiSSCo RI and the Eionet portal of the EEA. Opening the discussion with the EEA in this phase of maturity of DiSSCo will ensure that the infrastructural design of DiSSCo and the development of e-Services accommodate the present and future needs of the EEA and assure data interoperability between the two infrastructures.

The aim of this white paper is to present benefits from identifying the common service interests of DiSSCo and the EEA. A brief introduction to natural science collections as well as the two actors is given to facilitate the understanding of the needs and possibilities in the alignment of DiSSCo with the EEA.

Keywords

Science policy interface, DiSSCo, EEA, interoperability, Research Infrastructure

1. The significance of (digital) Natural Science Collections

European natural science collections (NSCs) are an integral and important part of the global natural and cultural capital. They include an estimated 1.5 billion animals, plants, fossils, rocks, minerals and meteorites, which account for 55% of the NSCs globally (Ariño 2010, Wheeler et al. 2012), and represent an estimated 80% of the world bio- and geo-diversity (CETAF 2015, The DiSSCo team 2017). The NSCs represent the primary reference material supporting biodiversity and geo-diversity discovery spanning 4.5 billion years of the Earth's natural history which are constantly curated and updated through new

scientific expeditions and research projects by more than 5000 full-time European scientists.

The last two decades have seen a rapid growth in the digitisation of NSCs, as well as the mobilisation of existing digital NSC data stored in local collection management systems of Natural History Museums (NHMs) and herbaria to the public domain (Nelson and Ellis 2018). The digitisation initiatives of NSC data include the United States National Science Foundation's Advancing Digitization of Biodiversity Collections (ADBC) and related Integrated Digitized Biocollections (iDigBio^{*1}) data portal, Australia's Atlas of Living Australia (ALA^{*2}), Mexico's National Commission for the Knowledge and Use of Biodiversity (CONABIO^{*3}), Brazil's Centro de Referência em Informação Ambiental (CRIA^{*4}), and China's National Specimen Information Infrastructure (NSII^{*5}) (Nelson and Ellis 2018). These initiatives are now complemented by the European Distributed System of Scientific Collections – DiSSCo^{*6} (Section 2). The Global Biodiversity Information Facility (GBIF^{*7}) aggregates all available digital NSC data at a global level and now serves 1.6 billion records on the temporal and spatial distribution of biodiversity, of which 220 million records are derived from NHM specimens (Sept. 2020). This body of information is rapidly growing and contributes to numerous scientific (CODATA, the Committee on Data of the International Science Council' et al. 2020) and policy papers, notably the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES 2019) and the Global Biodiversity Outlook 5 (Secretariat of the Convention on Biological Diversity 2020).

Historically, the primary function of NSCs has been a record of biological and geological diversity and as references for taxonomists and geologists (Meineke et al. 2018b). Data derived from NSCs underpin all taxonomic species and mineral descriptions and countless innovations. These include many thousands of scientific publications and policy reports relating to environmental change, sustainability, health, and food security, inventions and products critical to our bio-economy, databases, maps and descriptions of scientific observations, instructional material for education, and informational material for the public (Suarez and Tsutsui 2004). The importance of NSCs relies on the variety of information they contain. NSCs not only represent the localities and dates of collection associated with the vouchered specimens and thereby provide large-scale, verifiable data on native and base-line distributions of organisms and how those distributions have changed through time (Page et al. 2015), the collections themselves constitute a far larger body of knowledge and likely include presently unknown knowledge (E.g. the scientific 'omics' domain is relatively recent). Physical specimens are an important source of information for many scientific disciplines including global change and conservation biology research (Meineke et al. 2018b), such as detecting shifts in morphological traits and trait expressions (e.g. body mass of animals and leaf shapes of plants; Yang et al. 2015, Gardner et al. 2011), phenological responses (e.g. flower and fruiting seasons of plants; Primack et al. 2004, Willis et al. 2017, Meineke et al. 2018a), shifts in nutrient compositions through time (McLauchlan et al. 2010), heavy metal accumulation (Weiss et al. 1999), mismatching of pollinator-plant interactions (Miller-Struttman et al. 2015, Molnár et al. 2015, Byers 2017), herbivore interactions (Lees et al. 2011, Schilthuizen et al. 2016,

Beauvais et al. 2017, Meineke et al. 2018a), pathogen discovery and pathobiology research (Cook et al. 2017, Cook et al. 2020, Dunnum et al. 2017, Murphy et al. 2020), and changes in physiological processes (Miller-Rushing et al. 2009), among many others. Next to information on the chemical composition and morphological expression of traits of species, NSCs are a key resource of genetic material, the very source of DNA barcodes (Hebert and Gregory 2005, Ratnasingham and Hebert 2007, Ratnasingham and Hebert 2013, Hollingsworth et al. 2009). DNA barcodes are instrumental to identify indicator species, invasive species or the community composition in biodiversity assessments using environmental or eDNA metabarcoding techniques (Corlett 2017, Deiner et al. 2017, Makiola et al. 2020). Major advantages of eDNA samples is that they are non-invasive and non-destructive sampling methods that have the capacity to detect the presence of species in the environment which otherwise would remain unnoticed, especially in aquatic (Carraro et al. 2018) and soil environments (Epp et al. 2012), but also in wildlife biology and biodiversity monitoring (Bohmann et al. 2014). These advances would not be possible without the vouchered specimens in NSCs that generated the DNA barcodes.

Together with the ongoing digitisation of NSC specimens and the growing body of digital specimen information, new techniques emerge that automate the extraction of morphological traits and measurements of size (especially for plants as herbarium vouchers are more easy to digitised than others NSC groups of organisms; Soltis 2017). These techniques include automated identification (Carranza-Rojas et al. 2017, Wäldchen and Mäder 2018, Wäldchen et al. 2018, Little et al. 2020), trait extraction (Corney et al. 2012, Ott et al. 2020, Weaver et al. 2020), long-term insect-plant interactions (Meineke et al. 2020) and phenological data (Goëau et al. 2020, Pearson et al. 2020). These advances transform biodiversity collections from merely physical specimens into suites of interconnected resources enriched through study over time (Page 2016, Nelson and Ellis 2018, Thiers et al. 2019, Lendemer et al. 2020). These interconnected resources extend biodiversity specimens to potentially limitless digital resources and additional physical preparations, e.g. tissue samples and DNA extractions (Lendemer et al. 2020). This concept was coined by Webster (2017) as the “extended specimen” and represents the next generation of NSCs (Schindel and Cook 2018). The extended specimen concept is further developed by Biodiversity Collections Network (BCoN⁸) in the US. Around the same time, within the DiSSCo aligned ICEDIG project (see Section 2), the concept of a Digital Specimen (DS) was developed (Hardisty et al. 2019). A DS is the representation of a digitised physical specimen and contains information about a single specimen with links to all derived supplementary information. Hardisty et al. (2019) propose a new openDS TDWG (Biodiversity Information Standards) standard that is a specification of a DS and other object types essential to mass digitisation of NSCs and their digital use. The openDS standard includes 1) the DS, 2) its storage container, 3) the collection it belongs to, 4) the organisation that houses the collection, and 5) interpretations of the specimen such as its determination of species, trait measurements, DNA sequence data, etc. The similarities between the extended specimen concept and openDS were recognized by all parties and has resulted in a ‘Letter of Intent for Collaboration - in a Global and Open Process - on Interoperable Enriched Specimen Information Model’⁹. Signatories of the letter express their intent to collaborate on converging the DS concept (initiated by DiSSCo) and

Extended Specimen concept (initiated by BCoN) towards a global specification or, if differences in use cases require so, multiple specifications which are interoperable. Furthermore, signatories aim to collaborate in a global process, open to participation from all stakeholders, under the umbrella of the GBIF led 'Alliance for Biodiversity Knowledge'¹⁰ and with stakeholders in the Earth Samples domain.

The richness of information captured by NSCs makes them stand out in quality and potential from human and machine observations on biodiversity in space and time shared through data portals like GBIF, and urge the need for mass digitisation of NSCs. The potential of NSCs by persistently linking all derived information to the physical specimens (from where the information was obtained), along with the recognition of the urgent need for mass digitisation of NSCs (Blagoderov et al. 2012) to support science, policy and society, is at the core of the new European research infrastructure – DiSSCo.

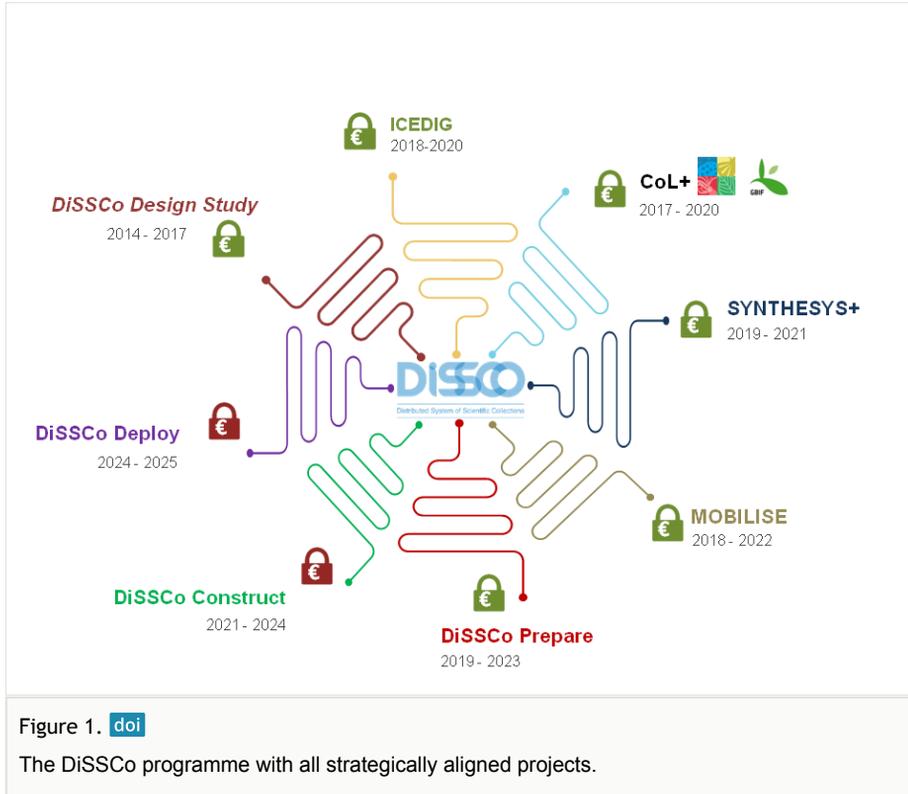
2. Distributed System of Scientific Collections - DiSSCo

DiSSCo⁶ – the Distributed System of Scientific Collections – is a new European Research Infrastructure (RI) for NSCs that entered the ESFRI (European Strategy Forum on Research Infrastructures) roadmap in 2018 in the environmental domain¹¹. The DiSSCo RI works for the digital unification of all European natural science assets under common curation and access policies and practices. These aim to make the data easily Findable, more Accessible, Interoperable and Reusable, a.k.a. FAIR (Wilkinson et al. 2016). As such, the DiSSCo RI enables the transformation of a fragmented landscape of the NSCs into an integrated knowledge base that provides interconnected hard evidence on the natural world.

The DiSSCo consortium represents the largest ever formal agreement between Natural History Museums (NHMs), botanical gardens and collection-holding universities from Europe, currently representing 120 institutions from 21 countries. The DiSSCo consortium is still growing and open for new institutes to join by signing the Memorandum of Understanding (MoU) of DiSSCo. Collectively the DiSSCo consortium hosts 80% of the world's known bio- and geodiversity represented by the 1.5 billion specimens that are preserved in its NSCs. By bringing together NSCs institutions at this scale and combining earlier investments in data interoperability practices with technological advancements in digitization, cloud services and semantic linking of all derived data, DiSSCo makes all data from NSCs available as one virtual data cloud in association with a wide range of end-user services. These include finding data, accessing data, using data, and improving and updating data (Hardisty et al. 2020). DiSSCo connects historical NSC data with data emerging from new techniques including DNA barcodes, partial and whole genome sequences, proteomics and metabolomics data, chemical data, (morphological) trait data, phenological data and imaging data (Computer-assisted Tomography (CT), Synchrotron, etc.) to name but a few (See Section 1; Hardisty 2020).

Although the final product upon completion will be the deployment of the DiSSCo RI, at present DiSSCo is a programme of strategically aligned EU-funded projects with

collaborations from relevant organisations, such as CETAF^{*12} (Consortium of European Taxonomic Facilities). CETAF represents the scientific institutions and community of experts that underpin the creation and further development of the DiSSCo RI. Furthermore, DiSSCo collaborates with the Catalogue of Life (CoL) through the CoL+ project^{*13} under the ‘Alliance of Biodiversity Knowledge’, led by GBIF (Hobern et al. 2019). All these projects and initiatives collectively contribute to the realisation of the DiSSCo RI (Fig. 1).



DiSSCo started in 2014 with the **DiSSCo design study**, which resulted in the submission of DiSSCo to the ESFRI roadmap where it was accepted in 2018^{*11}. In the meanwhile, from 2018-2020 the EU-funded and aligned **ICEDIG**^{*14} – “Innovation and consolidation for large scale digitisation of natural heritage” – project was operational, and it delivered the conceptual design blueprint for the DiSSCo digitization infrastructure addressing the technical, financial, policy and governance aspects necessary to operate a large distributed initiative for natural sciences collections across Europe (Hardisty et al. 2020) as a key output. Under ICEDIG, also the first design and version of a Collection Digitisation Dashboard was developed (van Egmond et al. 2019) that provides collection descriptions with their level of digitisation of the collection holdings of all DiSSCo members. Knowledge on the content of institutional holdings provides more detailed information on the European collections and can drive prioritized digitisation through Digitisation-on-Demand (DoD) requests of specimens of particularly topical scientific or policy relevance.

During the same period, CoL and GBIF embarked on the **CoL+** project which develops a new IT architecture to compile the most comprehensive and authoritative global index of species currently available, and which is going to serve as the backbone taxonomy for GBIF, DiSSCo, and others, among which the EEA.

In 2019, the **SYNTHESYS+**^{*15} project entered the fourth and final iteration of the “Synthesis of systematic resources” (SYNTHESYS) programme. A core element in SYNTHESYS is to provide funded researcher visits (Access), both physical and virtual, to the specimens housed by SYNTHESYS institutions. Alongside 'Access', Joint Research Activities (JRA) aim to improve the quality of, and increase access to, collections and data within natural history institutions by developing digital collections. Many activities under SYNTHESYS+ contribute to the development of infrastructural components of the DiSSCo RI, notably the European Loans and Visits System^{*16} (ELViS), the European Curation and Annotation System (ECAS) and the further development of Collection Digitisation Dashboards (piloted under ICEDIG) led by CETAF in collaboration with the TDWG Collection Description^{*17} (CD) group. To emphasize collaboration and alignment, SYNTHESYS+ has a strong focus on networking activities (NA) through which organizations like CETAF, GBIF, TDWG and GGBN^{*18} (Global Genome Biodiversity Network) lead specific actions towards the interoperability, alignment and embedment of efforts across Europe and beyond.

The next project under the DiSSCo programme is the **MOBILISE**^{*19} EU COST Action ‘Mobilising Data, Experts and Policies in Scientific Collections’, which is running from 2018-2022. The aim of MOBILISE is to build up a cooperative, inclusive, bottom-up and responsive network with active involvement of European stakeholders to support research for biodiversity and geo-diversity informatics, many of which represent DiSSCo members or future members. MOBILISE organizes events, workshops, short scientific missions and training to transfer knowledge and technology between researchers, domain specialists, data aggregators and industry. Furthermore, MOBILISE promotes the development of standards and best practices as well as innovative workflows and techniques to increase efficiency of large-scale collection digitisation and data mobilisation. Finally, MOBILISE raises awareness about the need that sustainable data access infrastructures are an integral component of biodiversity research.

After acceptance of DiSSCo on the ESFRI roadmap in 2018, the **DiSSCo Prepare** project^{*20} was submitted and got funded in 2019. DiSSCo Prepare runs from 2019-2023 and spans most of the preparatory phase of DiSSCo. DiSSCo Prepare builds on the 104 recommendations of the conceptual design blueprint for the DiSSCo digitisation infrastructure (Hardisty et al. 2020) from the ICEDIG project. DiSSCo Prepare has two main aims:

1. **Raise** DiSSCo's Implementation readiness level (IRL) across the five dimensions of relevance as identified by the ICEDIG project, namely the technical, scientific, data, financial and organisational dimensions (Fig. 2). This will enhance DiSSCo's ability to successfully execute the construction phase and effect related actions based on clear, actionable guidelines with minimum risks for the RI partners.

2. **Deliver** DiSSCo's Construction Masterplan. This comprehensive and integrated Masterplan will be the product of the outputs of all of its content-related tasks and will be the project's final output. It will effectively serve as the organisational, scientific, financial and technical blueprint for the construction of the DiSSCo RI including establishing it as a legal entity.

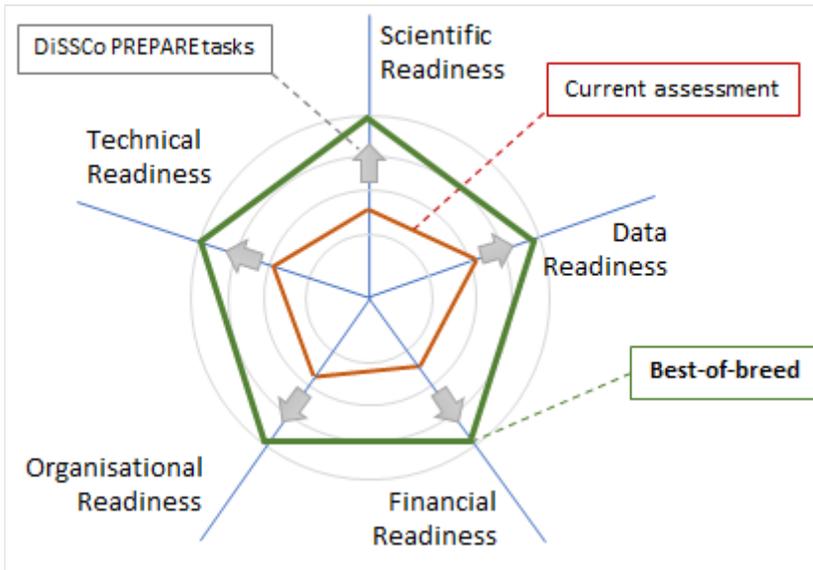


Figure 2. [doi](#)

The current and required implementation readiness levels for DiSSCo to enter the construction phase of the DiSSCo RI. The grey arrows indicate the tasks for the DiSSCo Prepare project.

Under the aligned projects and in the initial phase of the DiSSCo Prepare project, major advances have been made regarding a) the definition of digital specimens through the openDS standard (Hardisty et al. 2019), the equivalent of the extended specimen (Webster 2017, Lendemer et al. 2020), and b) the infrastructural requirements needed to serve the integrated information of digital specimens. Currently, explorations of the possibilities of a Digital Object Architecture (DOA), which offers a way of grouping, managing and processing fragments of information related to digital specimens, is in an advanced stage and included in the proposed openDS standard (Kahn and Wilensky 2006, Hardisty et al. 2019, Hardisty 2020). A first example of a simple digital specimen is depicted in the light-green-shaded box of Fig. 3 with links to all derived information. This particular specimen has a name represented by its name identifier from CoL, has its holotype deposited at the Natural History Museum of London, and has related sequence data stored in GenBank, etc. The rectangular darker green box illustrates how the digital specimen content could be serialized in JSON format for transfer between systems, as well as for application processing and entry into databases. Digital specimens are aggregations of widely distributed and heterogeneous data about biological and geological specimens. The use of the proposed DOA data model and components act as an approach to solving the

EEA and the aligned Knowledge Centre on Biodiversity to assure that a) the service portfolio of DiSSCo e-services providing unified discovery, access, interpretation and analysis of complex linked biodiversity data are interoperable with the services and data portals provided by the EEA, and b) that the complex and integrated biodiversity data served by DiSSCo's e-services are compliant with the present and future needs of the European Union.

3. European Environment Agency - EEA

The European Environment Agency^{*21} (EEA) is an agency of the European Union whose core task is to provide sound, independent information on the environment. The EEA aims to support sustainable development by helping to achieve significant and measurable improvement in Europe's environment, through the provision of timely, targeted, relevant and reliable information to policy making agents and the public. The EEA's key goals are to be the prime source of environmental knowledge at European level, play a leading role in supporting long-term transition to a sustainable society, and to be a lead organisation for environmental knowledge-sharing and capacity-building (European Environment Agency 2015). Its mandate is twofold: a) to help the European community and its member and cooperating countries make informed decisions about improving the environment, integrating environmental considerations into economic policies and moving towards sustainability; and b) to coordinate the European environment information and observation network - Eionet.

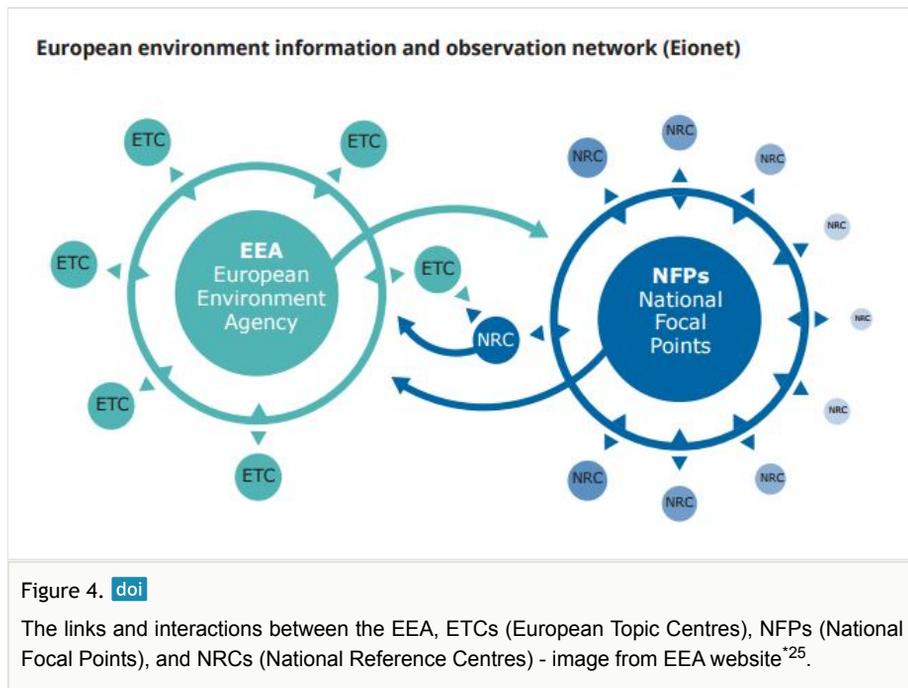
3.1 Eionet - European environment information and observation network

Eionet^{*22} is a partnership network of the EEA and its 32 member and 6 cooperating countries. The EEA is responsible for developing Eionet and coordinating its activities since 1994. To do so, the EEA works closely with National Focal Points (NFPs^{*23}) which are the main contact points for the EEA in the member and cooperating countries, typically based in national environment agencies or environment ministries. NFPs in turn are responsible for coordinating networks of 24 National Reference Centres (NRCs^{*24}). NRCs are located in organisations which are regular collectors or suppliers of environmental data at the national level and/or possess relevant knowledge regarding various environmental issues, monitoring or modelling (Fig. 4). Apart from the NFPs and NRCs, Eionet currently covers seven European Topic Centres (ETCs). ETCs are consortia of institutions across EEA member countries dealing with specific environmental topics. Institutions that are part of an ETC may also represent NRCs (Fig. 4).

ETCs^{*26} are centres of thematic expertise contracted by the EEA to carry out specific tasks identified in the EEA Multi-annual Work Programme and the annual work programmes. They are designated by the EEA Management Board following a Europe-wide competitive selection process and work as extensions of the EEA in specific topic areas. Each ETC consists of a lead organisation and specialist partner organisations from the environmental research and information community, which combine their resources in their particular areas of expertise. The ETCs, working together with Eionet countries, facilitate the

provision of data and information from the countries and deliver reports and other services to the EEA and Eionet. At present, there are seven ETCs covering the following environmental domains, 'Air and Climate', 'Nature', 'Sustainability and well-being', and 'Economic sectors':

1. Air Pollution, Transport, Noise and Industrial Pollution
2. Biological Diversity
3. Climate Change Impacts, Vulnerability and Adaptation
4. Climate Change Mitigation and Energy
5. Inland, Coastal and Marine Waters
6. Urban, Land and Soil Systems
7. Waste and Materials in Green Economy



3.2 European Topic Centre on Biological Diversity - ETC/BD

From the above seven ETCs that provide input to the EEA, the ETC on Biological Diversity (ETC/BD) is most closely related to the activities the DiSSCo RI and data it will be serving. The ETC/BD has the tasks of:

- a) assisting the EEA in reporting on Europe's environment by addressing the state and trends of biodiversity in Europe, b) providing relevant information to support the implementation of environmental and sustainable development policies in Europe in particular for EU nature and biodiversity policies (DG Environment: Nature and

Biodiversity^{*27}), c) build capacity for reporting on biodiversity in Europe, mainly through Eionet.

The activities of the ETC/BD are instrumental for the implementation of the EU nature directives, Natura 2000 areas and protected areas in general via the common database on designated areas (CDDA) and the EMERALD network (areas beyond EU27), and all related data flows and reporting formats. The ETC/BD is responsible for the reporting cycle of the Nature Directives including its assessment resulting in the 'State of Nature reports' (European Environment Agency 2019) and reporting under the Invasive Alien Species (IAS) Regulation. This information has fed into the preparation of the EEA support to the EU Biodiversity Strategy 2030 (European Commission 2020) and into the support to European Commission-coordinated MAES (Mapping and Assessment of Ecosystems and their Services; Maes et al. 2020) activities and work on Green Infrastructures^{*28}. The ETC/BD is responsible for two main data portals to deliver data to Eionet, the Biodiversity System for Europe (BISE^{*29}) and the European Nature Information System (EUNIS^{*30}).

BISE - Biodiversity System for Europe - is a single entry point for data and information on biodiversity supporting the implementation of the EU biodiversity strategy (European Commission 2011, European Commission 2020) and the Aichi biodiversity targets (Secretariat of the Convention on Biological Diversity 2013) in Europe. To deliver progress reporting on the tasks and activities of the ETC/BD, it has developed the 'Streamlined European Biodiversity Indicators' (Appendix 1 - SEBI^{*31}) indicator set as the main tool providing input to BISE.

EUNIS - European Nature Information System - brings together European data from several databases and organisations into three interlinked modules on:

a) sites, b) species, and c) habitat types.

EUNIS is a reference information system for anyone working in ecology and conservation, or those with an interest in the natural world. It is also used for assistance to the Natura 2000 process (EU Nature Directives including the EU Birds and Habitats Directives) and is coordinated with the related EMERALD Network of the Bern Convention (Council Of Europe 1979), the development of indicators, and environmental reporting connected to EEA reporting activities.

4. Connecting the EEA and DiSSCo

This white paper serves the purpose to characterise DiSSCo and the EEA to facilitate mutual understanding and identify connections between the data portal Eionet and all its components and DiSSCo, an integrated knowledge base that provides interconnected hard evidence of the natural world. As DiSSCo represents the largest ever formal agreement between natural history museums, botanic gardens and collection-holding universities in the world, it is an ideal platform for transdisciplinary interaction and dialogue that can provide input to the European community to make informed decisions about improving the environment, integrating environmental considerations into economic policies and moving

towards sustainability. DiSSCo's goals and vision also should align with broader regional and global strategic initiatives such as the recently published European Biodiversity Strategy 2030 (European Commission 2020), which has defined a number of specific commitments and actions to be delivered by 2030, including:

- Establishing a larger **EU-wide network of protected areas on land and at sea**, building upon existing Natura 2000 areas, with strict protection for areas of very high biodiversity and climate value.
- An **EU Nature Restoration Plan** - a series of concrete commitments and actions to restore degraded ecosystems across the EU by 2030, and manage them sustainably, addressing the key drivers of biodiversity loss.
- A set of measures to **enable the necessary transformative change**: setting in motion a new, strengthened governance framework to ensure better implementation and track progress, improving knowledge, financing and investments and better respecting nature in public and business decision-making.
- Measures to tackle the **global biodiversity challenge**, demonstrating that the EU is ready to lead by example towards the successful adoption of an ambitious global biodiversity framework under the Convention on Biological Diversity.
- Establish a new **Knowledge Centre for Biodiversity**, managed by the JRC (Joint Research Center) in close cooperation with the EEA.

DiSSCo, at its present level of maturity, has to make design and infrastructural decisions to facilitate the data infrastructure where all information generated through mass-digitisation of the European NSCs will be findable, accessible, interoperable and reusable (FAIR). As DiSSCo is still in the preparatory phase, modifications and adjustments to the infrastructural design are still possible to assure that the future needs of the European environmental policy domain are accommodated. Below we provide examples for a number of scientific domains with high policy relevance to the EEA. By providing these examples we aim to open further discussions and identify additional use cases of the DiSSCo infrastructure that can support the aims of the European Union set out in the European Biodiversity Strategy 2030 and its environmental policy.

4.1 Species information

Both EUNIS and DiSSCo (along with GBIF, Encyclopedia of Life (EOL), Biodiversity Heritage Library (BHL), LifeWatch, and the Barcode of Life) aim to adopt the new Catalogue of Life (CoL) as their backbone taxonomy. Through the shared use of the taxonomic name identifiers, EUNIS will gain access to a wealth of information related to specimens stored in the European NSCs. In figure 3, the CoL ID is depicted in red. The example of a (simple) Digital Specimen demonstrates how linked information on the type specimen (A single specimen selected by a scientist to define in future what distinguishes e.g. a certain species from all other), synonyms, spatial distribution, environmental conditions, phenology, morphology and traits (from images or measurements stored in linked data infrastructures like TRY (Kattge et al. 2011), EOL Traitbank (Parr et al. 2016), OpenTraits (Gallagher et al. 2020), DNA barcodes (e.g. BOLD; Ratnasingham and Hebert

2007), DNA sequence data (INSDC; Cochrane et al. 2016), chemical composition, etc. (See section 1) will become available through the DiSSCo RI.

Linking EUNIS with DiSSCo through CoL provides hard evidence on the past and present distributions of species that are protected under the European Nature directives and potentially on the spread of Invasive Alien Species (IAS). Combination of specimen data with spatial environmental data through niche modelling techniques will allow assessing the invasive potential of IAS or the chance of survival in certain habitats under future climate change conditions. The analysis of phenological information might indicate their suitability as a nectar source for the pollinator community of crops, e.g. their suitability for use in flower borders of pollinator-dependent crop lands (Blaauw and Isaacs 2014). Analysis of specimens might indicate populations that are resistant to certain pathogens (Malmstrom et al. 2007, Lavoie 2013). Many more and possibly unforeseen applications can be found and will emerge when more digital specimen data becomes available.

4.2 Restoration ecology & Biodiversity monitoring

To assess the effectiveness of the EU nature restoration plan outlined in the EU biodiversity strategy 2030 (European Commission 2020) requires information on the occurrence of species prior to the degraded state. The historical European collections can provide that information. Not only do the collections provide information on the community composition in a particular region or Natura 2000 designated area, functional traits represented by the community also provide information on ecosystem processes and the services the community provides (de Bello et al. 2010, Hanisch et al. 2020, Perez et al. 2020).

Restoration action may involve planting or releasing new individuals. The choice on which genetic resources are used for this purpose may define the restoration success. The collections provide information on the environmental boundaries or ecological niche conditions under which species occur through their georeferences. In anticipation of predicted effects of climate change, genetic material should be obtained from populations that are optimally adapted to (near) future climatic conditions in the area of restoration. This is especially relevant for sessile and long-lived organisms like trees (Hsu et al. 2012). The application of ecological niche models or species distribution models combining occurrence records from collections (and observations) with spatial information on present and future bioclimatic and edaphic or aquatic conditions is instrumental in this respect (Elith et al. 2010).

The use of environmental DNA or eDNA in future biodiversity monitoring is expected to rapidly increase, especially for aquatic and soil communities (Bohmann et al. 2014, Deiner et al. 2017, Yan et al. 2018, Ruppert et al. 2019). Links from DNA barcodes to specimens in DiSSCo to the related functional traits will provide information on different trophic levels and the functional diversity of the communities. The trait distribution of the community identified from the eDNA sample is indicative for ecosystem services that can be provided by the community (Díaz et al. 2007, de Bello et al. 2010, Hanisch et al. 2020). The same can be applied to conventional species surveys.

4.3 Climate change mitigation

With the ongoing changes in the global climate, it is likely that species will shift their distribution range to track suitable climatic conditions. Predicting range shifts of species under climate change requires understanding of the bioclimatic conditions and biotic interactions that contribute to the success of the species. Information from NSC specimens is instrumental to assess the required conditions through niche modelling or species distribution modelling. Individuals at the trailing edges of their distribution are likely to disappear from local communities. With that they may leave gaps in the functional trait distribution of these communities. Information on the traits that these species represent might allow selecting other species adapted to climate change, which can fill this gap and restore the ecosystem services provided by this local community (Bochet and García-Fayos 2015, Carlucci et al. 2020). At the other side of the species distribution range, the moving front, the dispersal of species might be hampered by large scale land use change and/or limited dispersal capacity. At this side of the range, assisted migration may be required to secure the survival of species (McLachlan et al. 2007, Seddon et al. 2014). The projection of species distribution models allows the identification of habitats with suitable future climatic conditions. Data from genetic data portals like European Nucleotide Archive (ENA; Leinonen et al. 2011, Harrison et al. 2019) linked to the specimens can assure that the future genetic variation within the population is maintained (Hoffmann et al. 2015).

4.4 Capacity building

DiSSCo is built on three pillars, 1) development of the IT architecture that accommodates the storage and linking of all derived information, 2) mass digitisation of the European NSCs, and 3) capacitation of the user community of the DiSSCo-RI. Opening the communication channels with the EEA in this phase of development of the DiSSCo RI is instrumental to accommodate the future use of the DiSSCo RI and its service catalogue by users from the environmental policy domain. Tailor-made documentation and helpdesk support on the DiSSCo RI requires information about the background knowledge and IT skills of users in the environmental policy domain represented by the EEA. We hope that this white paper contributes to open the discussion to identify the needs of users in the environmental policy domain so that the infrastructural design of the DiSSCo RI can accommodate future user demands.

Appendix 1. SEBI - Streamlined European Biodiversity Indicators

- **SEBI 001** Abundance and distribution of selected European species: Trends in bird and butterfly populations from 1990-2017 <<https://www.eea.europa.eu/data-and-maps/indicators/abundance-and-distribution-of-selected-species-8/assessment-1>>
- **SEBI 002** Red List Index for European species: To date, the Red List Index has been calculated only for bird species at a European level and was **not updated** after 2004 <<https://www.eea.europa.eu/data-and-maps/indicators/red-list-index-for-european-species/red-list-index-for-european>>

- **SEBI 003** Species of European interest: Status and trends of species and habitats under the Habitats and Birds Directives (Art. 17 & 12) <<https://www.eea.europa.eu/data-and-maps/indicators/species-of-european-interest-2/assessment>>
- SEBI 004 Ecosystem coverage: The coverage of ecosystem classes under the EU 'Mapping and Assessment of Ecosystems and their Services' (MAES) framework <<https://www.eea.europa.eu/data-and-maps/indicators/ecosystem-coverage-3/assessment>>
- SEBI 005 Habitats of European interest: Progress towards the 2020 target of improving the conservation status of habitats covered by the EU Habitats Directive <<https://www.eea.europa.eu/data-and-maps/indicators/habitats-of-european-interest-1/assessment>>
- SEBI 006 Livestock genetic diversity: **Not updated** since 2005 <<https://www.eea.europa.eu/data-and-maps/indicators/livestock-genetic-diversity/livestock-genetic-diversity-assessment-published>>
- SEBI 007 Nationally designated protected areas: The total area of nationally designated protected areas in Europe <<https://www.eea.europa.eu/data-and-maps/indicators/nationally-designated-protected-areas-10/assessment>>
- SEBI 008 Natura 2000 sites designated under the EU Habitats and Birds Directives <<https://www.eea.europa.eu/data-and-maps/indicators/sites-designated-under-the-eu-3/assessment>>
- SEBI 009 Critical load exceedance for nitrogen: **Not updated** since 2006 <<https://www.eea.europa.eu/data-and-maps/indicators/critical-load-exceedance-for-nitrogen/critical-load-exceedance-for-nitrogen>>
- **SEBI 010** Invasive alien species in Europe: The cumulative number of alien species introduced. **Not updated** since 2008 <<https://www.eea.europa.eu/data-and-maps/indicators/invasive-alien-species-in-europe/invasive-alien-species-in-europe>>
- SEBI 011 Impact of climate change on bird populations <<https://www.eea.europa.eu/data-and-maps/indicators/impact-of-climate-change-on/impact-of-climate-change-on>>
- SEBI 012 Marine trophic index of European seas: populations of predatory fishes decline to the benefit of smaller fish and invertebrates. **Not updated** since 2015 <<https://www.eea.europa.eu/data-and-maps/indicators/marine-trophic-index-of-european-seas/marine-trophic-index-of-european>>
- SEBI 013 Fragmentation of natural and semi-natural areas <<https://www.eea.europa.eu/data-and-maps/indicators/fragmentation-of-natural-and-semi-1/assessment-1>>
- SEBI 014 Fragmentation of river systems (in preparation)
- SEBI 015 Nutrients in transitional, coastal and marine waters <<https://www.eea.europa.eu/data-and-maps/indicators/nutrients-in-transitional-coastal-and-1/nutrients-in-transitional-coastal-and>>
- SEBI 016 Freshwater quality <<https://www.eea.europa.eu/data-and-maps/indicators/freshwater-quality/freshwater-quality-assessment-published-may-2>>

- SEBI 017 Forest: growing stock, increment and fellings <<https://www.eea.europa.eu/data-and-maps/indicators/forest-growing-stock-increment-and-fellings-3/assessment>>
- SEBI 018 Forest: deadwood <<https://www.eea.europa.eu/data-and-maps/indicators/forest-deadwood-1/assessment-1>>
- SEBI 019 Agriculture: nitrogen balance <<https://www.eea.europa.eu/data-and-maps/indicators/agriculture-nitrogen-balance-1/assessment>>
- SEBI 020 Agriculture: area under management practices potentially supporting biodiversity <<https://www.eea.europa.eu/data-and-maps/indicators/agriculture-area-under-management-practices/agriculture-area-under-management-practices-2>>
- SEBI 021 Status of marine fish and shellfish stocks in European seas <<https://www.eea.europa.eu/data-and-maps/indicators/status-of-marine-fish-stocks-4/assessment>>
- SEBI 022 Aquaculture: effluent water quality from finfish farms. **Not updated** since 2006. <<https://www.eea.europa.eu/data-and-maps/indicators/aquaculture-effluent-water-quality-from/aquaculture-effluent-water-quality-from>>
- SEBI 023 Ecological footprint of European countries <<https://www.eea.europa.eu/data-and-maps/indicators/ecological-footprint-of-european-countries-2/assessment>>
- SEBI 024 Patent applications based on genetic resources. **Not updated** since 2015 <<https://www.eea.europa.eu/data-and-maps/indicators/patent-applications-based-on-genetic-resources/patent-applications-based-on-genetic>>
- SEBI 025 Financing biodiversity management. **Not updated** since 2006 <<https://www.eea.europa.eu/data-and-maps/indicators/financing-biodiversity-management/financing-biodiversity-management-assessment-published>>
- SEBI 026 Public awareness of biodiversity in Europe <<https://www.eea.europa.eu/data-and-maps/indicators/public-awareness-2/assessment>>

Funding program

Horizon 2020

Grant title

H2020-INFRAIA-2018-2020 | SYNTHESYS PLUS | Grant Agreement Number 823827

Conflicts of interest

None

References

- Ariño A (2010) Approaches to estimating the universe of natural history collections data. *Biodiversity Informatics* 7 (2): 81-92. <https://doi.org/10.17161/bi.v7i2.3991>
- Beauvais M, Pellerin S, Dubé J, Lavoie C (2017) Herbarium specimens as tools to assess the impact of large herbivores on plant species. *Botany* 95 (2): 153-162. <https://doi.org/10.1139/cjb-2016-0206>
- Blaauw B, Isaacs R (2014) Flower plantings increase wild bee abundance and the pollination services provided to a pollination-dependent crop. *Journal of Applied Ecology* 51 (4): 890-898. <https://doi.org/10.1111/1365-2664.12257>
- Blagoderov V, Kitching I, Livermore L, Simonsen T, Smith V (2012) No specimen left behind: industrial scale digitization of natural history collections. *ZooKeys* 133-146. <https://doi.org/10.3897/zookeys.209.3178>
- Bochet E, García-Fayos P (2015) Identifying plant traits: A key aspect for species selection in restoration of eroded roadsides in semiarid environments. *Ecological Engineering* 83: 444-451. <https://doi.org/10.1016/j.ecoleng.2015.06.019>
- Bohmann K, Evans A, Gilbert MTP, Carvalho G, Creer S, Knapp M, Yu D, de Bruyn M (2014) Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution* 29 (6): 358-367. <https://doi.org/10.1016/j.tree.2014.04.003>
- Byers D (2017) Studying plant–pollinator interactions in a changing climate: A review of approaches. *Applications in Plant Sciences* 5 (6). <https://doi.org/10.3732/apps.1700012>
- Carlucci M, Brancalion PS, Rodrigues R, Loyola R, Cianciaruso M (2020) Functional traits and ecosystem services in ecological restoration. *Restoration Ecology* n/a (n/a). <https://doi.org/10.1111/rec.13279>
- Carranza-Rojas J, Goeau H, Bonnet P, Mata-Montero E, Joly A (2017) Going deeper in the automated identification of Herbarium specimens. *BMC Evolutionary Biology* 17 (1). <https://doi.org/10.1186/s12862-017-1014-z>
- Carraro L, Hartikainen H, Jokela J, Bertuzzo E, Rinaldo A (2018) Estimating species distribution and abundance in river networks using environmental DNA. *Proceedings of the National Academy of Sciences* 115 (46): 11724-11729. <https://doi.org/10.1073/pnas.1813843115>
- CETAF (2015) Strategy & strategic development plan 2015-2025. URL: https://cetaf.org/sites/default/files/final_strategy_and_strategic_development_plan.pdf
- Cochrane G, Karsch-Mizrachi I, Takagi T, Sequence Database Collaboration IN (2016) The International Nucleotide Sequence Database Collaboration. *Nucleic Acids Research* 44 (D1). <https://doi.org/10.1093/nar/gkv1323>
- CODATA, the Committee on Data of the International Science Council', Pfeiffenberger H, Uhlir P (2020) Twenty-Year Review of GBIF. <https://doi.org/10.35035/ctzm-hz97>
- Cook J, Galbreath K, Bell K, Campbell M, Carrière S, Colella J, Dawson N, Dunnum J, Eckerlin R, Fedorov V, Greiman S, Haas GS, Haukialmi V, Henttonen H, Hope A, Jackson D, Jung T, Koehler A, Kinsella J, Krejsa D, Kutz S, Liphardt S, MacDonald SO, Malaney J, Makarikov A, Martin J, McLean B, Mulders R, Nyamsuren B, Talbot S, Tkach V, Tsvetkova A, Toman H, Waltari E, Whitman J, Hoberg E (2017) The Beringian Coevolution Project: holistic collections of mammals and associated parasites reveal novel perspectives on evolutionary and environmental change in the North. *Arctic Science* 3 (3): 585-617. <https://doi.org/10.1139/as-2016-0042>

- Cook J, Arai S, Armien B, Bates J, Bonilla CC, Cortez MBdS, Dunnum J, Ferguson A, Johnson K, Khan FAA, Paul D, Reeder D, Revelez M, Simmons N, Thiers B, Thompson C, Upham N, Vanhove MM, Webala P, Weksler M, Yanagihara R, Soltis P (2020) Integrating Biodiversity Infrastructure into Pathogen Discovery and Mitigation of Emerging Infectious Diseases. *BioScience* 70 (7): 531-534. <https://doi.org/10.1093/biosci/biaa064>
- Corlett R (2017) A Bigger Toolbox: Biotechnology in Biodiversity Conservation. *Trends in Biotechnology* 35 (1): 55-65. <https://doi.org/10.1016/j.tibtech.2016.06.009>
- Corney DA, Clark J, Tang HL, Wilkin P (2012) Automatic extraction of leaf characters from herbarium specimens. *Taxon* 61 (1): 231-244. <https://doi.org/10.1002/tax.6111016>
- Council Of Europe (1979) Explanatory Report to the Convention on the Conservation of European Wildlife and Natural Habitats.
- de Bello F, Lavorel S, Díaz S, Harrington R, Cornelissen JC, Bardgett R, Berg M, Cipriotti P, Feld C, Hering D, Martins da Silva P, Potts S, Sandin L, Sousa JP, Storkey J, Wardle D, Harrison P (2010) Towards an assessment of multiple ecosystem processes and services via functional traits. *Biodiversity and Conservation* 19 (10): 2873-2893. <https://doi.org/10.1007/s10531-010-9850-9>
- Deiner K, Bik H, Mächler E, Seymour M, Lacoursière-Roussel A, Altermatt F, Creer S, Bista I, Lodge D, Vere Nd, Pfrender M, Bernatchez L (2017) Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology* 26 (21): 5872-5895. <https://doi.org/10.1111/mec.14350>
- Díaz S, Lavorel S, Bello Fd, Quétier F, Grigulis K, Robson TM (2007) Incorporating plant functional diversity effects in ecosystem service assessments. *Proceedings of the National Academy of Sciences* 104 (52): 20684-20689. <https://doi.org/10.1073/pnas.0704716104>
- Dunnum J, Yanagihara R, Johnson K, Armien B, Batsaikhan N, Morgan L, Cook J (2017) Biospecimen Repositories and Integrated Databases as Critical Infrastructure for Pathogen Discovery and Pathobiology Research. *PLOS Neglected Tropical Diseases* 11 (1). <https://doi.org/10.1371/journal.pntd.0005133>
- Elith J, Kearney M, Phillips S (2010) The art of modelling range-shifting species. *Methods in Ecology and Evolution* 1 (4): 330-342. <https://doi.org/10.1111/j.2041-210X.2010.00036.x>
- Epp L, Boessenkool S, Bellemain E, Haile J, Esposito A, Riaz T, Erséus C, Gusarov V, Edwards M, Johnsen A, Stenøien H, Hassel K, Kauserud H, Yoccoz N, Bråthen KA, Willerslev E, Taberlet P, Coissac E, Brochmann C (2012) New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems. *Molecular Ecology* 21 (8): 1821-1833. <https://doi.org/10.1111/j.1365-294X.2012.05537.x>
- European Commission (2011) Our life insurance, our natural capital: an EU biodiversity strategy to 2020.
- European Commission (2018) Turning FAIR into reality : final report and action plan from the European Commission expert group on FAIR data. ISBN: 9789279965470 9789279965463 Publisher: Publications Office of the European Union. <https://doi.org/10.2777/1524>
- European Commission (2020) EU Biodiversity Strategy for 2030; Bringing nature back into our lives.

- European Environment Agency (2015) EEA general brochure - who we are, what we do, how we do it. URL: <https://www.eea.europa.eu/publications/eea-general-brochure>
- European Environment Agency (2019) The European environment — state and outlook 2020: Knowledge for transition to a sustainable Europe. European Environment Agency.
- Gallagher R, Falster D, Maitner B, Salguero-Gómez R, Vandvik V, Pearse W, Schneider F, Kattge J, Poelen J, Madin J, Ankenbrand M, Penone C, Feng X, Adams V, Alroy J, Andrew S, Balk M, Bland L, Boyle B, Bravo-Avila C, Brennan I, Carthey AR, Catullo R, Cavazos B, Conde D, Chown S, Fadrique B, Gibb H, Halbritter A, Hammock J, Hogan JA, Holewa H, Hope M, Iversen C, Jochum M, Kearney M, Keller A, Mabee P, Manning P, McCormack L, Michaletz S, Park D, Perez T, Pineda-Munoz S, Ray C, Rossetto M, Sauquet H, Sparrow B, Spasojevic M, Telford R, Tobias J, Violle C, Walls R, Weiss KB, Westoby M, Wright I, Enquist B (2020) Open Science principles for accelerating trait-based science across the Tree of Life. *Nature Ecology & Evolution* 4 (3): 294-303. <https://doi.org/10.1038/s41559-020-1109-6>
- Gardner J, Peters A, Kearney M, Joseph L, Heinsohn R (2011) Declining body size: a third universal response to warming? *Trends in Ecology & Evolution* 26 (6): 285-291. <https://doi.org/10.1016/j.tree.2011.03.005>
- Goëau H, Mora-Fallas A, Champ J, Love NR, Mazer S, Mata-Montero E, Joly A, Bonnet P (2020) A new fine-grained method for automated visual analysis of herbarium specimens: A case study for phenological data extraction. *Applications in Plant Sciences* 8 (6). <https://doi.org/10.1002/aps3.11368>
- Hanisch M, Schweiger O, Cord A, Volk M, Knapp S (2020) Plant functional traits shape multiple ecosystem services, their trade-offs and synergies in grasslands. *Journal of Applied Ecology* 57 (8): 1535-1550. <https://doi.org/10.1111/1365-2664.13644>
- Hardisty A, Ma K, Nelson G, Fortes J (2019) 'openDS' – A New Standard for Digital Specimens and Other Natural Science Digital Object Types. *Biodiversity Information Science and Standards* 3 <https://doi.org/10.3897/biss.3.37033>
- Hardisty A (2020) What is a Digital Specimen? URL: <https://dissco.tech/2020/03/31/what-is-a-digital-specimen/>
- Hardisty A, Saarenmaa H, Casino A, Dillen M, Gödderz K, Groom Q, Hardy H, Koureas D, Nieva de la Hidalga A, Paul D, Runnel V, Vermeersch X, van Walsum M, Willemsse L (2020) Conceptual design blueprint for the DiSSCo digitization infrastructure - DELIVERABLE D8.1. *Research Ideas and Outcomes* 6 <https://doi.org/10.3897/rio.6.e54280>
- Harrison P, Alako B, Amid C, Cerdeño-Tárraga A, Cleland I, Holt S, Hussein A, Jayathilaka S, Kay S, Keane T, Leinonen R, Liu X, Martínez-Villacorta J, Milano A, Pakseresht N, Rajan J, Reddy K, Richards E, Rosello M, Silvester N, Smirnov D, Toribio A, Vijayaraja S, Cochrane G (2019) The European Nucleotide Archive in 2018. *Nucleic Acids Research* 47 (D1). <https://doi.org/10.1093/nar/gky1078>
- Hebert PN, Gregory TR (2005) The Promise of DNA Barcoding for Taxonomy. *Systematic Biology* 54 (5): 852-859. <https://doi.org/10.1080/10635150500354886>
- Hobern D, Baptiste B, Copas K, Guralnick R, Hahn A, van Huis E, Kim E, McGeoch M, Naicker I, Navarro L, Noesgaard D, Price M, Rodrigues A, Schigel D, Sheffield C, Wiczorek J (2019) Connecting data and expertise: a new alliance for biodiversity knowledge. *Biodiversity Data Journal* 7 <https://doi.org/10.3897/BDJ.7.e33679>
- Hoffmann A, Griffin P, Dillon S, Catullo R, Rane R, Byrne M, Jordan R, Oakeshott J, Weeks A, Joseph L, Lockhart P, Borevitz J, Sgrò C (2015) A framework for incorporating

- evolutionary genomics into biodiversity conservation and management. *Climate Change Responses* 2 (1). <https://doi.org/10.1186/s40665-014-0009-x>
- Hollingsworth P, Forrest L, Spouge J, Hajibabaei M, Ratnasingham S, van der Bank M, Chase M, Cowan R, Erickson D, Fazekas A, Graham S, James K, Kim K, Kress WJ, Schneider H, van AlphenStahl J, Barrett SH, van den Berg C, Bogarin D, Burgess K, Cameron K, Carine M, Chacón J, Clark A, Clarkson J, Conrad F, Devey D, Ford C, Hedderson TJ, Hollingsworth M, Husband B, Kelly L, Kesanakurti P, Kim JS, Kim Y, Lahaye R, Lee H, Long D, Madriñán S, Maurin O, Meusnier I, Newmaster S, Park C, Percy D, Petersen G, Richardson J, Salazar G, Savolainen V, Seberg O, Wilkinson M, Yi D, Little D (2009) A DNA barcode for land plants. *Proceedings of the National Academy of Sciences* 106 (31). <https://doi.org/10.1073/pnas.0905845106>
 - Hsu RC, Tamis WM, Raes N, Snoo Gd, Wolf JD, Oostermeijer G, Lin S (2012) Simulating climate change impacts on forests and associated vascular epiphytes in a subtropical island of East Asia. *Diversity and Distributions* 18 (4): 334-347. <https://doi.org/10.1111/j.1472-4642.2011.00819.x>
 - IPBES (2019) Summary for policymakers of the global assessment report on biodiversity and ecosystem services. Zenodo. DOI: 10.5281/zenodo.3553579. URL: <https://zenodo.org/record/3553579#.X2hllGgzaHs>
 - Kahn R, Wilensky R (2006) A framework for distributed digital object services. *International Journal on Digital Libraries* 6 (2): 115-123. <https://doi.org/10.1007/s00799-005-0128-x>
 - Kattge J, Díaz S, Lavorel S, Prentice IC, Leadley P, Bönisch G, Garnier E, Westoby M, Reich PB, Wright IJ, Cornelissen JHC, Violle C, Harrison SP, van Bodegom PM, Reichstein M, Enquist BJ, Soudzilovskaia NA, Ackerly DD, Anand M, Atkin O, Bahn M, Baker TR, Baldocchi D, Bekker R, Blanco CC, Blonder B, Bond WJ, Bradstock R, Bunker DE, Casanoves F, Cavender-Bares J, Chambers JQ, Chapin Iii FS, Chave J, Coomes D, Cornwell WK, Craine JM, Dobrin BH, Duarte L, Durka W, Elser J, Esser G, Estiarte M, Fagan WF, Fang J, Fernández-Méndez F, Fidelis A, Finegan B, Flores O, Ford H, Frank D, Freschet GT, Fyllas NM, Gallagher RV, Green WA, Gutierrez AG, Hickler T, Higgins SI, Hodgson JG, Jalili A, Jansen S, Joly CA, Kerkhoff AJ, Kirkup D, Kitajima K, Kleyer M, Klotz S, Knops JMH, Kramer K, Kühn I, Kurokawa H, Laughlin D, Lee TD, Leishman M, Lens F, Lenz T, Lewis SL, Lloyd J, Llusià J, Louault F, Ma S, Mahecha MD, Manning P, Massad T, Medlyn BE, Messier J, Moles AT, Müller SC, Nadrowski K, Naeem S, Niinemets Ü, Nöllert S, Nüske A, Ogaya R, Oleksyn J, Onipchenko VG, Onoda Y, Ordoñez J, Overbeck G, Ozinga WA, Patiño S, Paula S, Pausas JG, Peñuelas J, Phillips OL, Pillar V, Poorter H, Poorter L, Poschlod P, Prinzing A, Proulx R, Rammig A, Reinsch S, Reu B, Sack L, Salgado-Negret B, Sardans J, Shiodera S, Shipley B, Siefert A, Sosinski E, Soussana JF, Swaine E, Swenson N, Thompson K, Thornton P, Waldram M, Weiher E, White M, White S, Wright SJ, Yguel B, Zaehle S, Zanne AE, Wirth C (2011) TRY – a global database of plant traits. *Global Change Biology* 17 (9): 2905-2935. <https://doi.org/10.1111/j.1365-2486.2011.02451.x>
 - Lannom L, Koureas D, Hardisty A (2019) FAIR Data and Services in Biodiversity Science and Geoscience. *Data Intelligence* 2 (1-2): 122-130. https://doi.org/10.1162/dint_a_00034
 - Lavoie C (2013) Biological collections in an ever changing world: Herbaria as tools for biogeographical and environmental studies. *Perspectives in Plant Ecology, Evolution and Systematics* 15 (1): 68-76. <https://doi.org/10.1016/j.ppees.2012.10.002>

- Lees D, Lack HW, Rougerie R, Hernandez-Lopez A, Raus T, Avtzis N, Augustin S, Lopez-Vaamonde C (2011) Tracking origins of invasive herbivores through herbaria and archival DNA: the case of the horse-chestnut leaf miner. *Frontiers in Ecology and the Environment* 9 (6): 322-328. <https://doi.org/10.1890/100098>
- Leinonen R, Akhtar R, Birney E, Bower L, Cerdeno-Tárraga A, Cheng Y, Cleland I, Faruque N, Goodgame N, Gibson R, Hoad G, Jang M, Pakseresht N, Plaister S, Radhakrishnan R, Reddy K, Sobhany S, Ten Hoopen P, Vaughan R, Zalunin V, Cochrane G (2011) The European Nucleotide Archive. *Nucleic Acids Research* 39 (suppl_1). <https://doi.org/10.1093/nar/gkq967>
- Lendemer J, Thiers B, Monfils A, Zaspel J, Ellwood E, Bentley A, LeVan K, Bates J, Jennings D, Contreras D, Lagomarsino L, Mabee P, Ford L, Guralnick R, Gropp R, Revelez M, Cobb N, Seltmann K, Aime MC (2020) The Extended Specimen Network: A Strategy to Enhance US Biodiversity Collections, Promote Research and Education. *BioScience* 70 (1): 23-30. <https://doi.org/10.1093/biosci/biz140>
- Little D, Tulig M, Tan KC, Liu Y, Belongie S, Kaeser-Chen C, Michelangeli F, Panesar K, Guha RV, Ambrose B (2020) An algorithm competition for automatic species identification from herbarium specimens. *Applications in Plant Sciences* 8 (6). <https://doi.org/10.1002/aps3.11365>
- Maes J, Teller A, Erhard M, Conde S, Vallecillo Rodriguez S, Barredo Cano Jose I, Paracchini M, Abdul Malak D, Trombetti M, Vigiak O, Zulian G, Addamo A, Grizzetti B, Somma F, Hagyo A, Vogt P, Polce C, Jones A, Marin A, Ivits E, Mauri A, Rega C, Czucz B, Ceccherini G, Pisoni E, Ceglar A, De Palma P, Cerrani I, Meroni M, Caudullo G, Lugato E, Vogt J, Spinoni J, Cammalleri C, Bastrup-Birk A, San-Miguel-Ayanz J, San Román S, Kristensen P, Christiansen T, Zal N, De Roo A, De Jesus Cardoso A, Pistocchi A, Del Barrio Alvarellós I, Tsiamis K, Gervasini E, Deriu I, La Notte A, Abad Viñas R, Vizzarri M, Camia A, Robert N, Kakoulaki G, Garcia Bendito E, Panagos P, Ballabio C, Scarpa S, Montanarella L, Orgiazzi A, Fernandez Ugalde O, Santos-Martín F (2020) Mapping and Assessment of Ecosystems and their Services: An EU ecosystem assessment. Publications Office of the European Union. Accepted: 2020-10-13T00:03:19Z Artwork Medium: Online ISBN: 9789276178330 9789276229544 ISSN: 1831-9424 (online), 1831-9424 (supplement) Interview Medium: Online DOI: 10.2760/757183. URL: <https://publications.jrc.ec.europa.eu/repository/handle/111111111/61318>
- Makiola A, Compson ZG, Baird DJ, Barnes MA, Boerlijst SP, Bouchez A, Brennan G, Bush A, Canard E, Cordier T, Creer S, Curry RA, David P, Dumbrell AJ, Gravel D, Hajibabaei M, Hayden B, van der Hoorn B, Jarne P, Jones JI, Karimi B, Keck F, Kelly M, Knot IE, Krol L, Massol F, Monk WA, Murphy J, Pawlowski J, Poisot T, Porter TM, Randall KC, Ransome E, Ravigné V, Raybould A, Robin S, Schrama M, Schatz B, Tamaddoni-Nezhad A, Trimbos KB, Vacher C, Vasselon V, Wood S, Woodward G, Bohan DA (2020) Key Questions for Next-Generation Biomonitoring. *Frontiers in Environmental Science* 7 <https://doi.org/10.3389/fenvs.2019.00197>
- Malmstrom C, Shu R, Linton E, Newton L, Cook M (2007) Barley yellow dwarf viruses (BYDVs) preserved in herbarium specimens illuminate historical disease ecology of invasive and native grasses. *Journal of Ecology* 95 (6): 1153-1166. <https://doi.org/10.1111/j.1365-2745.2007.01307.x>

- McLachlan J, Hellmann J, Schwartz M (2007) A Framework for Debate of Assisted Migration in an Era of Climate Change. *Conservation Biology* 21 (2): 297-302. <https://doi.org/10.1111/j.1523-1739.2007.00676.x>
- McLachlan K, Ferguson C, Wilson I, Ocheltree T, Craine J (2010) Thirteen decades of foliar isotopes indicate declining nitrogen availability in central North American grasslands. *New Phytologist* 187 (4): 1135-1145. <https://doi.org/10.1111/j.1469-8137.2010.03322.x>
- Meineke E, Davis C, Davies TJ (2018a) The unrealized potential of herbaria for global change biology. *Ecological Monographs* 88 (4): 505-525. <https://doi.org/10.1002/ecm.1307>
- Meineke E, T. Jonathan Davies, Barnabas H. Daru, Charles C. Davis (2018b) Biological collections for understanding biodiversity in the Anthropocene. *Philosophical Transactions of the Royal Society B: Biological Sciences* 374 (1763). <https://doi.org/10.1098/rstb.2017.0386>
- Meineke E, Tomasi C, Yuan S, Pryer K (2020) Applying machine learning to investigate long-term insect–plant interactions preserved on digitized herbarium specimens. *Applications in Plant Sciences* 8 (6). <https://doi.org/10.1002/aps3.11369>
- Miller-Rushing A, Primack R, Templer P, Rathbone S, Mukunda S (2009) Long-term relationships among atmospheric CO₂, stomata, and intrinsic water use efficiency in individual trees. *American Journal of Botany* 96 (10): 1779-1786. <https://doi.org/10.3732/ajb.0800410>
- Miller-Struttman N, Geib J, Franklin J, Kevan P, Holdo R, Ebert-May D, Lynn A, Kettenbach J, Hedrick E, Galen C (2015) Functional mismatch in a bumble bee pollination mutualism under climate change. *Science* 349 (6255): 1541-1544. <https://doi.org/10.1126/science.aab0868>
- Molnár VA, Löki V, Takács A, Sschmidt J, Tökölyi J, Bódis J, Sramkó G (2015) No evidence for historical declines in pollination success in Hungarian orchids. *Appl. Ecol. Environ. Res* 13: 1097-1108. https://doi.org/10.15666/aeer/1304_10971108
- Murphy C, Gerth W, Pauk K, Konstantinidis P, Arismendi I (2020) Hiding in Plain Sight: Historical Fish Collections Aid Contemporary Parasite Research. *Fisheries* 45 (5): 263-270. <https://doi.org/10.1002/fsh.10411>
- Nelson G, Ellis S (2018) The history and impact of digitization and digital data mobilization on biodiversity research. *Philosophical Transactions of the Royal Society B: Biological Sciences* 374 (1763). <https://doi.org/10.1098/rstb.2017.0391>
- Ott T, Palm C, Vogt R, Oberprieler C (2020) GinJinn: An object-detection pipeline for automated feature extraction from herbarium specimens. *Applications in Plant Sciences* 8 (6). <https://doi.org/10.1002/aps3.11351>
- Page L, MacFadden B, Fortes J, Soltis P, Riccardi G (2015) Digitization of Biodiversity Collections Reveals Biggest Data on Biodiversity. *BioScience* 65 (9): 841-842. <https://doi.org/10.1093/biosci/biv104>
- Page R (2016) Towards a biodiversity knowledge graph. *Research Ideas and Outcomes* 2 <https://doi.org/10.3897/rio.2.e8767>
- Parr C, Schulz K, Hammock J, Wilson N, Leary P, Rice J, Corrigan Jr. R (2016) TraitBank: Practical semantics for organism attribute data. *Semantic Web* 7 (6): 577-588. <https://doi.org/10.3233/SW-150190>
- Pearson K, Nelson G, Aronson MJ, Bonnet P, Brenskelle L, Davis C, Denny E, Ellwood E, Goëau H, Heberling JM, Joly A, Lorieul T, Mazer S, Meineke E, Stucky B, Sweeney

- P, White A, Soltis P (2020) Machine Learning Using Digitized Herbarium Specimens to Advance Phenological Research. *BioScience* 70 (7): 610-620. <https://doi.org/10.1093/biosci/biaa044>
- Perez T, Rodriguez J, Heberling JM (2020) Herbarium-based measurements reliably estimate three functional traits. *American Journal of Botany* n/a (n/a). <https://doi.org/10.1002/ajb2.1535>
 - Primack D, Imbress C, Primack RB, Miller-Rushing AJ, Del Tredici P (2004) Herbarium specimens demonstrate earlier flowering times in response to warming in Boston. *Am. J. Bot.* 91 (8): 1260-1264. <https://doi.org/10.3732/ajb.91.8.1260>
 - Ratnasingham S, Hebert PD (2007) BOLD: The Barcode of Life Data System (<http://www.barcodinglife.org>). *Molecular Ecology Notes* 7 (3): 355-364. <https://doi.org/10.1111/j.1471-8286.2007.01678.x>
 - Ratnasingham S, Hebert PD (2013) A DNA-Based Registry for All Animal Species: The Barcode Index Number (BIN) System. *PLoS ONE* 8 (7). <https://doi.org/10.1371/journal.pone.0066213>
 - Ruppert K, Kline R, Rahman MS (2019) Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. *Global Ecology and Conservation* 17 <https://doi.org/10.1016/j.gecco.2019.e00547>
 - Schilthuizen M, Santos Pimenta L, Lammers Y, Steenbergen P, Flohil M, Beveridge NP, van Duijn P, Meulblok M, Sosef N, van de Ven R, Werring R, Beentjes K, Meijer K, Vos R, Vrieling K, Gravendeel B, Choi Y, Verpoorte R, Smit C, Beukeboom L (2016) Incorporation of an invasive plant into a native insect herbivore food web. *PeerJ* 4 <https://doi.org/10.7717/peerj.1954>
 - Schindel D, Cook J (2018) The next generation of natural history collections. *PLoS Biology* 16 (7). <https://doi.org/10.1371/journal.pbio.2006125>
 - Secretariat of the Convention on Biological Diversity (2013) Quick guides to the Aichi Biodiversity Targets.
 - Secretariat of the Convention on Biological Diversity (2020) Global Biodiversity Outlook 5.
 - Seddon P, Griffiths C, Soorae P, Armstrong D (2014) Reversing defaunation: Restoring species in a changing world. *Science* 345 (6195): 406-412. <https://doi.org/10.1126/science.1251818>
 - Soltis P (2017) Digitization of herbaria enables novel research. *American Journal of Botany* 104 (9): 1281-1284. <https://doi.org/10.3732/ajb.1700281>
 - Suarez A, Tsutsui N (2004) The Value of Museum Collections for Research and Society. *BioScience* 54 (1): 66-74. [https://doi.org/10.1641/0006-3568\(2004\)054\[0066:TVOMCF\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2004)054[0066:TVOMCF]2.0.CO;2)
 - The DiSSCo team (2017) DiSSCo application package.
 - Thiers B, Anna Monfils, Jennifer Zaspel, Elizabeth Ellwood, Andrew Bentley, Katherine Levan, John Bates, David Jennings, Dori Contreras, Laura Lagomarsino, Paula Mabee, Linda Ford, Robert Guralnick, Robert Gropp, Marcy Revelez, Neil Cobb, James Lendemer, Katja Seltsmann, Mary Catherine Aime (2019) Extending U.S. Biodiversity Collections to Promote Research and Education.
 - van Egmond E, Willemse L, Paul D, Woodburn M, Casino A, Gödderz K, Vermeersch X, Bloothoofd J, Wijers A, Raes N (2019) Design of a Collection Digitisation Dashboard.

Zenodo. DOI: 10.5281/zenodo.2621055. URL: https://zenodo.org/record/2621055#.X2iP_2gzaHs

- Wäldchen J, Mäder P (2018) Machine learning for image based species identification. *Methods in Ecology and Evolution* 9 (11): 2216-2225. <https://doi.org/10.1111/2041-210X.13075>
- Wäldchen J, Rzanny M, Seeland M, Mäder P (2018) Automated plant species identification—Trends and future directions. *PLoS Computational Biology* 14 (4). <https://doi.org/10.1371/journal.pcbi.1005993>
- Weaver W, Ng J, Laport R (2020) LeafMachine: Using machine learning to automate leaf trait extraction from digitized herbarium specimens. *Applications in Plant Sciences* 8 (6). <https://doi.org/10.1002/aps3.11367>
- Webster M (2017) *The Extended Specimen: Emerging Frontiers in Collections-Based Ornithological Research*. CRC Press
- Weiss D, Shotyk W, Kramers J, Gloor M (1999) Sphagnum mosses as archives of recent and past atmospheric lead deposition in Switzerland. *Atmospheric Environment* 33 (23): 3751-3763. [https://doi.org/10.1016/S1352-2310\(99\)00093-X](https://doi.org/10.1016/S1352-2310(99)00093-X)
- Wheeler QD, Knapp S, Stevenson DW, Stevenson J, Blum SD, Boom BM, Borisy GG, Buizer JL, Carvalho MRD, Cibrian A, Donoghue MJ, Doyle V, Gerson EM, Graham CH, Graves P, Graves SJ, Guralnick RP, Hamilton AL, Hanken J, Law W, Lipscomb DL, Lovejoy TE, Miller H, Miller JS, Naeem S, Novacek MJ, Page LM, Platnick NI, Porter-Morgan H, Raven PH, Solis MA, Valdecasas AG, Leeuw SVD, Vasco A, Vermeulen N, Vogel J, Walls RL, Wilson EO, Woolley JB (2012) Mapping the biosphere: exploring species to understand the origin, organization and sustainability of biodiversity. *Systematics and Biodiversity* 10 (1): 1-20. <https://doi.org/10.1080/14772000.2012.665095>
- Wilkinson M, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J, da Silva Santos LB, Bourne P, Bouwman J, Brookes A, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo C, Finkers R, Gonzalez-Beltran A, Gray AG, Groth P, Goble C, Grethe J, Heringa J, 't Hoen PC, Hooft R, Kuhn T, Kok R, Kok J, Lusher S, Martone M, Mons A, Packer A, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S, Schultes E, Sengstag T, Slater T, Strawn G, Swertz M, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3 <https://doi.org/10.1038/sdata.2016.18>
- Willis C, Ellwood E, Primack R, Davis C, Pearson K, Gallinat A, Yost J, Nelson G, Mazer S, Rossington N, Sparks T, Soltis P (2017) Old Plants, New Tricks: Phenological Research Using Herbarium Specimens. *Trends in Ecology & Evolution* 32 (7): 531-546. <https://doi.org/10.1016/j.tree.2017.03.015>
- Yan D, Mills J, Gellie NC, Bissett A, Lowe A, Breed M (2018) High-throughput eDNA monitoring of fungi to track functional recovery in ecological restoration. *Biological Conservation* 217: 113-120. <https://doi.org/10.1016/j.biocon.2017.10.035>
- Yang J, Spicer R, Spicer TV, Arens NC, Jacques FB, Su T, Kennedy E, Herman A, Steart D, Srivastava G, Mehrotra R, Valdes P, Mehrotra N, Zhou Z, Lai J (2015) Leaf form–climate relationships on the global stage: an ensemble of characters. *Global Ecology and Biogeography* <https://doi.org/10.1111/geb.12334>

Endnotes

- *1 <https://www.idigbio.org/>
- *2 <https://www.ala.org.au/>
- *3 <https://www.gob.mx/conabio>
- *4 <http://www.cria.org.br/>
- *5 <http://nsii.org.cn/2017/home.php>
- *6 <https://www.dissco.eu/>
- *7 <https://www.gbif.org/>
- *8 <https://bcon.aibs.org/>
- *9 <http://bit.ly/LolSpModel>
- *10 <https://www.allianceforbio.org/>
- *11 <http://roadmap2018.esfri.eu/projects-and-landmarks/browse-the-catalogue/dissco/>
- *12 <https://cetaf.org/>
- *13 <https://github.com/CatalogueOfLife/general>
- *14 <https://icedig.eu/>
- *15 <https://www.synthesys.info/about-synthesys.html>
- *16 <https://elvis.dissco.eu/welcome>
- *17 <https://www.tdwg.org/community/cd/>
- *18 <http://www.ggbn.org>
- *19 <https://www.mobilise-action.eu/>
- *20 <https://www.dissco.eu/what-is-prepare/>
- *21 <https://www.eea.europa.eu/>
- *22 <https://www.eionet.europa.eu/>
- *23 <https://www.eea.europa.eu/about-us/countries-and-eionet/list-of-members-of-the-nfp-eionet-group>
- *24 <https://www.eionet.europa.eu/countries/national-reference-centres>
- *25 <https://www.eea.europa.eu/about-us>
- *26 <https://www.eionet.europa.eu/etcs>
- *27 http://ec.europa.eu/environment/nature/index_en.htm
- *28 <https://ec.europa.eu/environment/nature/ecosystems/background.htm>
- *29 <https://biodiversity.europa.eu/>
- *30 <https://eunis.eea.europa.eu>
- *31 <https://biodiversity.europa.eu/topics/sebi-indicators>