



**Building an Integrated, Open, Findable, Accessible, Interoperable, and Reusable (BIOFAIR) Data Network**

## **BIOFAIR Data Network Listening Sessions: A Summary November 13, 2024**



# Housekeeping

- This session is being **recorded** and will be posted on the BCoN website.
- You may use the **chat** feature to share comments and to introduce yourself.
- Please use the **Zoom Q&A** feature to ask your questions.
- If you would like to speak, please use the **raise hand** function.
- Contact **AIBS IT Support** via chat if you are experiencing technical difficulties.
- **Agenda:**
  1. Introduction to BCoN and BIOFAIR Data Network
  2. The BIOFAIR Listening Sessions
  3. Cross-cutting Summary
  4. Next Steps: Workshop



**BIODIVERSITY  
COLLECTIONS NETWORK**

## **Vision**

An integrated 21st century biological collections community that strives to understand, promote and conserve the diversity of life.

## **Mission**

To promote the use and expand the accessibility of biodiversity collections and data, and to extend their impact as critical infrastructure for research and education for the benefit of life on earth.

<https://bcon.aibs.org/>

# BCoN Steering Committee



**John Bates**  
Field Museum



**Andrew Bentley**  
University of Kansas  
Biodiversity Institute



**Dori Contreras**  
Perot Museum of Nature  
and Science



**Elizabeth Ellwood**  
iDigBio



**Nico Franz**  
NEON Biorepository,  
Arizona State University



**Michael Lomas**  
National Center for  
Marine Algae



**Anna Monfils**  
Central Michigan University



**William Moser**  
National Museum of  
Natural History



**Gil Nelson**  
iDigBio



**Jyotsna Pandey**  
AIBS & NSCA



**Sinlan Poo**  
Memphis Zoo



**Barbara Thiers**  
New York Botanical  
Garden (Emerita)



**Gregory Watkins-Colwell**  
Yale Peabody Museum of  
Natural History



**Mike Webster**  
Macaulay Library, Cornell  
Lab of Ornithology



**Breda Zimkus**  
Harvard Museum of  
Comparative Zoology



**Nimanthi Abeyrathna**  
Clarkson University



**David Kunkel**  
Oklahoma State  
University



**Brooke Long-Fox**  
Phoenix Bioinformatics



**Cameron Pittman**  
Denver Museum of Nature &  
Science



**Julia Portmann**  
Yale University



**Matthew Sheik**  
Denver Botanic Gardens

# Extending U.S. Biodiversity Collections to Address National Challenges

Webster, M., ed. 2017 The Extended Specimen. Routledge.



Released March 2019

[bcon.aibs.org/resources/](https://bcon.aibs.org/resources/)



EXTENDED SPECIMEN  
NETWORK



JOURNAL ARTICLE

## The Extended Specimen Network: A Strategy to Enhance US Biodiversity Collections, Promote Research and Education

James Lendemer , Barbara Thiers, Anna K Monfils, Jennifer Zaspel, Elizabeth R Ellwood, Andrew Bentley, Katherine LeVan, John Bates, David Jennings, Dori Contreras ... [Show more](#)

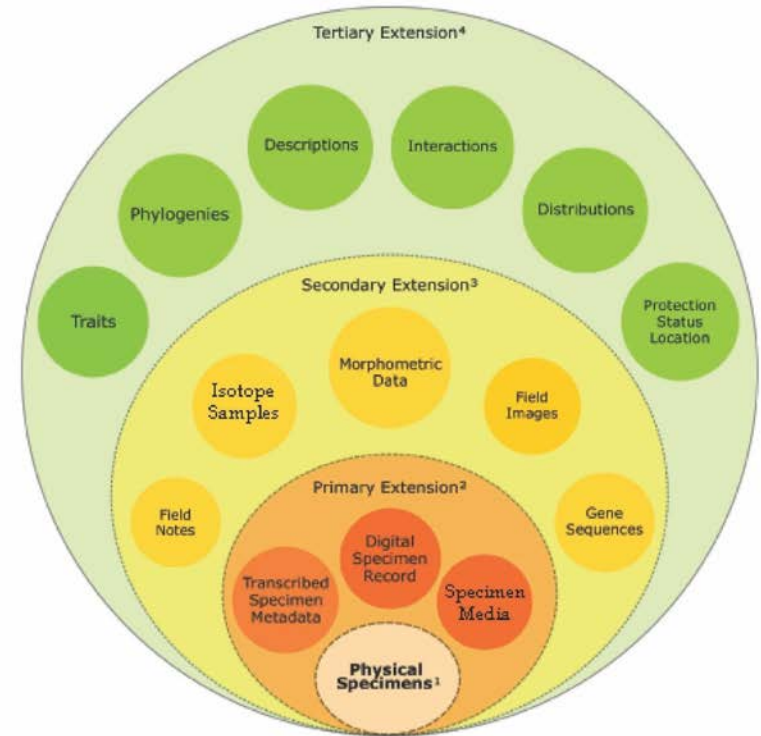
*BioScience*, Volume 70, Issue 1, January 2020, Pages 23–30,  
<https://doi.org/10.1093/biosci/biz140>

**Published:** 22 November 2019

# Extended Specimen Network (ESN)

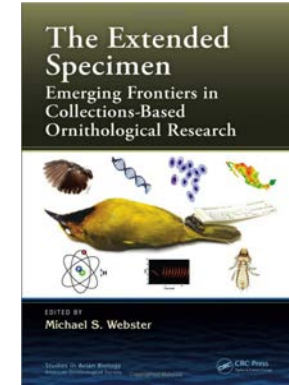
## Components:

- “Born digital” documentation of new specimens
- Continued retroactive specimen digitization and data editing
- Data integration and cyberinfrastructure
- Institution and community infrastructure
- Education and Workforce Training



## How can the ESN vision integrate data and expand the utility of collections and their associated data to broader communities?

- Integration at all levels from specimen creation to publication products
- Expanded the use of unique identifiers to facilitate congregation throughout the data lifecycle
- Reliant on published, complete, FAIR data



## Data sharing and integration needed across data domains to realize the Extended Specimen Network (ESN)



# Why should the collections community care about this?

## Data Infrastructure

- Provides more **accurate and cleaner data**, increasing the visibility and use of collections data, **broadening relevance** for novel research.
- Facilitates understanding of **gaps** and where **more biodiversity sampling** is needed
- Enables immediate **innovation and troubleshooting**
- Enhances **information exchange** among data providers and users

## Social Infrastructure

- Fosters shared expertise and “**extended**” **community support** for collections and collections data
- Expands and **diversifies the workforce** with transferable skills
- Ensures **engagement in larger discussions** with other data domains

# Building an Integrated, Open, Findable, Accessible, Interoperable, & Reusable (BIOFAIR) Data Network



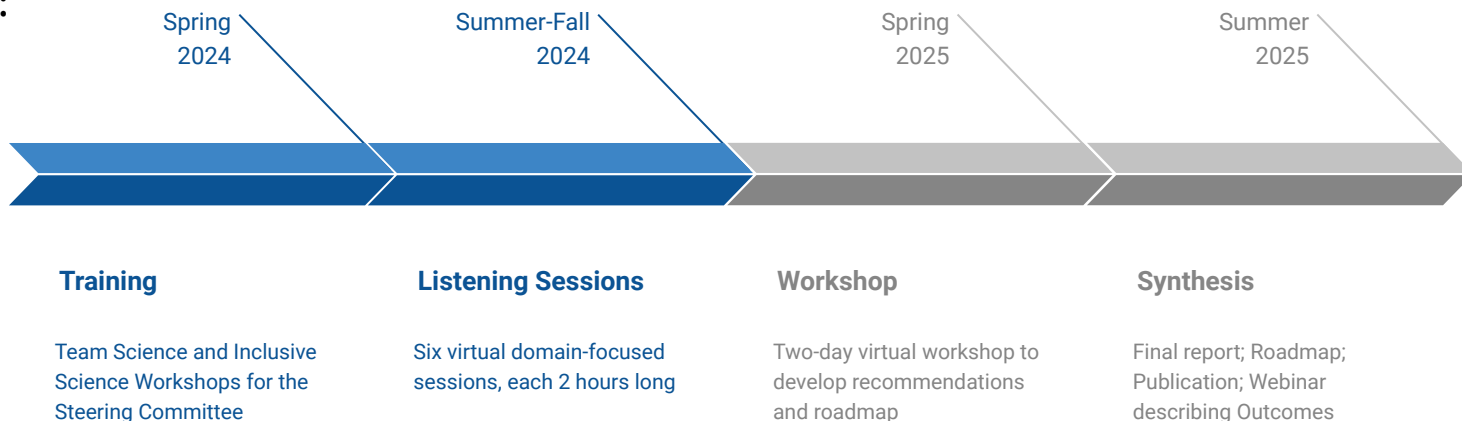
<https://bcon.aibs.org/biofair/>

*Catalyzing cross-domain discussions  
towards data integration*

NSF Award: [2303588](#)

**Goal:** *To engage an expansive set of current and prospective partners towards Building an Integrated, Open, Findable, Accessible, Interoperable, and Reusable (BIOFAIR) Data Network*

## Timeline:



# Virtual Listening Sessions



Six domain-focused virtual listening sessions aimed at building collaborations among disparate data communities.

1. Federal agencies (June 14)
2. Genetic and genomic data (June 26)
3. One Health (July 2)
4. Ecological data (July 12)
5. Climate and environmental data (July 26)
6. Biodiversity informatics (August 26)

## Discussions within each session explored:

- Current status of and plans for data resource(s)
- Challenges, gaps, needs, and opportunities for data integration
- Existing conceptual frameworks for data integration
- Path towards integration of diverse datasets

# 1. Federal Agencies

**Session leads:** Gil Nelson, William Moser, Nico Franz, Brooke Long-Fox

**Key Domain Partner:** Interagency Working Group on Scientific Collections (IWGSC)

## How are Federal Scientific Collections Different?

- Tied to Department and Agency Missions
  - Research is primarily mission-driven, not curiosity- or profit-driven
- Long-term goals and support derived from Congressional legislation
  - Sampling can extend over decades, creating unique longitudinal time-series for trend analyses
- Often used for legal and regulatory enforcement
  - Sampling and collection management may require chain-of-custody sampling, more secure storage



**Agricultural  
Research  
Service**



**NATIONAL  
CANCER  
INSTITUTE**



**NIST**

NATIONAL INSTITUTE OF  
STANDARDS AND TECHNOLOGY  
U.S. DEPARTMENT OF COMMERCE



**BOEM**  
Bureau of Ocean Energy  
Management



Smithsonian



**USGS**  
science for a changing world



# 1. Federal Agencies

## Challenges:

- **Cybersecurity requirements** differ across and within agencies and change continuously.
- Challenges associated with giving **broad access to federal data** that are linked to certain restricted and proprietary information
- Data stored in aggregation systems can get **'stale,'** i.e. data is deposited and then never updated, which leads to misrepresentative and inaccurate or outdated data.

## What's needed:

- **Harmonized data standards and IT/cybersecurity policies** across agencies can enable better data sharing and integration.
- Need to address challenges related to **legacy data representations** and **permissions** for data access and integration.

Read the  
summary  
report!



## 2. Genetic and Genomic Data

**Session leads:** Breda Zimkus, Andrew Bentley, John Bates, Mike Lomas, Nimanthi Abeyrathna

**Key Domain Partners:** Conrad Schoch, Kevin Kerr

### Considerations for Genetic and Genomic collections

- Key stakeholders include NCBI (including Genbank, Sequence Read Archive, BioProject, BioSample), International Barcode of Life (iBOL), Barcode of Life Data System (BOLD)
- Integration of sequence information with specimen records
- Benefit-sharing from use of digital sequence information (DSI) on genetic resources (Nagoya Protocol)



## 2. Genetic and Genomic Data

### Challenges:

- Lack of **metadata and persistent identifiers** associated with sequence records
- Lack of **consistency and accuracy** of voucher information by submitters
- Inability to **embargo data** for regulation compliance (e.g., Nagoya Protocol)

### What's needed:

- Development of **integration tools**
- **Data curation tools** that allow annotation by institutions and communities

Read the  
summary  
report!



# 3. One Health

**Session leads:** John Bates, Sinlan Poo, Greg Watkins-Colwell, Julia Portmann

**Key Domain Partners:** Kelly Speer, Marcia Revelez, Joseph Cook

## Considerations for One Health collections

- Key stakeholders include zoos, museums, agencies, and disease-focused non-profits
- Sample collection should emphasize locally-anchored infrastructure for long term sustainability
- Specimen data need to be accessible outside of the museum community



## 3. One Health

### Challenges:

- Handling different **biosafety level** specimens
- **Limited funding** and focus on **specific pathogens** rather than broad surveys
- **Lack of communication** between museum and veterinary communities

### What's needed:

- **Central organization** that can host all (or most) levels
- A **forecasting and real-time assessment service** for emerging zoonotic pathogens
- **Improved communication** around the import and export of samples

Read the  
summary  
report!



## 4. Ecological Data

**Session leads:** Matthew Sheik, Mike Webster, Nico Franz, Jyotsna Pandey, Libby Ellwood

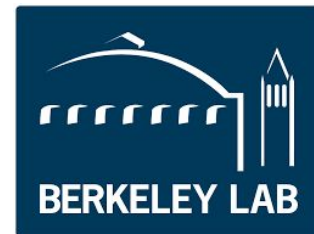
**Key Domain Partners:** Gregory Maurer, Stephanie Parker

### Current Considerations

- Data integration among ecological datasets as well as between ecological data and other biotic and abiotic data
- Cross-scale and cross-platform integration is challenging
- Potential for increased investment in standards, ontology development, and the adoption of translation tools to operate among services.



National Ecological Observatory Network



## 4. Ecological Data

### Challenges:

- **Emerging data types**, e.g., bioacoustic data and camera trap data, may not be easy to integrate into existing systems
- Considerations for things like **sample types and participatory science-derived data**

### What's needed:

- **Flexibility in data models**
- Importance of **dedicated managers** for species data
- **Translators across different repositories** to manage data heterogeneity and connect networks of ecological data

Read the  
summary  
report!



## 5. Climate and Environmental Data

**Session leads:** Barbara Thiers, David Kunkel, Dori Contreras, Anna Monfils

**Key Domain Partners:** Ben Halpern (NCEAS), Ty Tuff (ESIIL)

### Questions under consideration:

- Is species occurrence data derived from biological specimens included within the definition and scope of environmental data?
- What changes are needed to improve the interoperability of species occurrence data derived from specimens with other types of environmental data?
- What outcomes can we envision from greater integration of specimen-derived and other types of environmental data?



**Climate  
Hazards  
Center**  
UC SANTA BARBARA



## 5. Climate and Environmental Data

### Challenges:

- **Finding** relevant datasets
- **Missing or incomplete metadata** for existing datasets
- How to **document the loss or absence** of species

### What's needed:

- A **continuously updated catalog** of international environmental data resources
- **Global gap analysis** for environmental data, including occurrence data
- An enhanced **culture of data sharing** across environmental data types



Read the  
summary  
report!

# 6. Biodiversity Informatics



**Session leads:** Andrew Bentley, Mike Webster, Dori Contreras, Cameron Pittman

**Key Domain Partners:** GBIF (Tim Robertson), iDigBio (Jose Fortes), DiSSCo (Sharif Islam)

## Considerations:

- What existing tools and infrastructure are already being developed?
- What is missing?
- How we can synergize these disparate efforts while also identifying gaps that need filling and how best to achieve these needs?
- What are the major challenges to realizing data integration?
- How do we accommodate data sovereignty and traditional knowledge?



# 6. Biodiversity Informatics

## Challenges:

- **Lack of consistent unique identifier system**
- **Lack of attribution mechanisms**
- **Protections for data sovereignty and traditional knowledge vs FAIR data**



Read the  
summary  
report!

## What's needed:

- Explore **potential collaborations and integrations** between existing data infrastructure projects (e.g., GBIF, DiSSCo, iSamples).
- Investigate **legal and ethical considerations** for data sharing and integration across different jurisdictions.
- **Develop strategies for balancing FAIR data principles with necessary data protections.**

# Cross-cutting summary report

- 199 people representing 139 projects and institutions participated in the six listening sessions
- Participants indicated strong support for the vision of a globally accessible data network
- Data integration requires technical infrastructure solutions and development of social norms and contracts not yet realized
- Significant barriers exist - data access, discovery, standardization, insufficient training, and insecure data sources.
- Overcoming these barriers will require a collective impact model for FAIR data

# The Five Conditions of Collective Impact

## Common Agenda

All participants have a **shared vision for change** including a common understanding of the problem and a joint approach to solving it through agreed upon actions.

---

## Shared Measurement

**Collecting data and measuring results consistently** across all participants ensures efforts remain aligned and participants hold each other accountable.

---

## Mutually Reinforcing Activities

Participant activities must be **differentiated while still being coordinated** through a mutually reinforcing plan of action.

---

## Continuous Communication

**Consistent and open communication** is needed across the many players to build trust, assure mutual objectives, and appreciate common motivation.

---

## Backbone Support

Creating and managing collective impact requires a dedicated staff and a specific set of skills to **serve as the backbone for the entire initiative and coordinate participating organizations and agencies.**

*Reprinted with the permission of FSG and the Stanford Social Innovation Review*

# Cross-cutting summary report

- Summaries from individual listening sessions were used to generate comprehensive list of outcomes, challenges and recommendations
- A number of cross-cutting themes were identified while some unique challenges for individual groups were also highlighted.
- These were thematically collated and arranged in a summary report that will be used to inform the final workshop discussions
- The aim is to discuss possible solutions to the major impediments of a fully integrated BIOFAIR data network.

<https://bcon.aibs.org/2024/10/09/summary-report-listening-sessions/>



# Cross-cutting summary report

Broad themes:

- Address equity in data access
- Establish standards and protocols for ethical use of data
- Incentivize best practices
- Increase data availability
- Improve data integration
- Provide adequate training
- Maintain existing data resources

# Cross-cutting summary report

Broad themes:

- Address equity in data access
- Establish standards and protocols for ethical use of data
- Incentivize best practices
- Increase data availability
- Improve data integration
- Provide adequate training
- Maintain existing data resources

# Cross-cutting summary report

Broad themes:

- Address equity in data access
- Establish standards and protocols for ethical use of data
- **Incentivize best practices**
- Increase data availability
- Improve data integration
- Provide adequate training
- Maintain existing data resources

# Cross-cutting summary report

Broad themes:

- Address equity in data access
- Establish standards and protocols for ethical use of data
- Incentivize best practices
- **Increase data availability**
- Improve data integration
- Provide adequate training
- Maintain existing data resources

# Cross-cutting summary report

Broad themes:

- Address equity in data access
- Establish standards and protocols for ethical use of data
- Incentivize best practices
- Increase data availability
- **Improve data integration**
- Provide adequate training
- Maintain existing data resources

# Cross-cutting summary report

Broad themes:

- Address equity in data access
- Establish standards and protocols for ethical use of data
- Incentivize best practices
- Increase data availability
- Improve data integration
- **Provide adequate training**
- Maintain existing data resources

# Cross-cutting summary report

Broad themes:

- Address equity in data access
- Establish standards and protocols for ethical use of data
- Incentivize best practices
- Increase data availability
- Improve data integration
- Provide adequate training
- **Maintain existing data resources**

# Final Workshop (February 13, 2025)



An **interdisciplinary virtual workshop** will bring together an expansive configuration of stakeholders to work together to develop recommendations and create a roadmap that outlines the gaps, needs, and actions needed to build a FAIR, open, integrated data network.

## Workshop Participants:

- Designated participants from Virtual Listening Sessions
- Representatives from international participants, including GBIF, DISSCO, and others.
- BIOFAIR Steering Committee Members
- 50-60 total participants anticipated

## Workshop Outcomes:

- **Summary of principles** that should govern establishment of a collaborative network of biological/biodiversity related data and how such a collaborative would ideally function.
- **Use cases** for a integrated biological data network (for research, education, increasing diversity in scientific research).
- **Key challenges and priorities** for the implementation of the network.

# Output from the BIOFAIR Project:

01

A final report describing proceedings and outcomes from the Listening Sessions and Final Workshop

02

A “Statement of Engagement,” which articulates a set of agreed upon principles and a commitment to work collaboratively to develop an integrated data network

03

A roadmap that outlines the gaps, needs, and actions for creating a BIOFAIR Data Network

04

One or more publications summarizing outcomes

## Questions and Comments?

- What is missing from the discussion so far?
- What other data initiatives exist that we should be aware of?
- What concerns do you have?
- What negative or positive connotations are there for an effort like this in your domain?
- Ideas for future engagement to build this network?

# Get involved!

Share your thoughts...

<https://bcon.aibs.org/2024/10/09/summary-report-listening-sessions/>



Express your interest in attending the workshop...

<https://aibs.wufoo.com/forms/biofair-data-network-workshop-interest-form>

