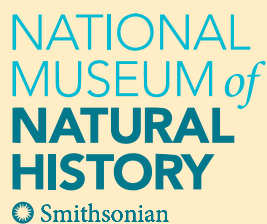




VIRTUAL EVENT

SEPTEMBER 22-23, 2021






DAY ONE: WEDNESDAY, 22 SEPTEMBER 2021


Time EDT	Session One 8:30 - 11:00 AM
-------------	--


8:30	<i>Opening</i> Gil Nelson, iDigBio & Rebecca Johnson, NMNH
------	---


8:40	 <i>Welcome</i> Kirk Johnson National Museum of Natural History, Smithsonian Institution
------	---

Oral Presentations Topic: Innovations: Strategy & Coordination

9:00	 <i>Increasing the Reach and Impact of Scientific Collections through 3D-imaging</i> David Blackburn iDigBio Florida Museum of Natural History, University of Florida
------	--

9:20	 <i>Progress in data mobilization: standards, coordination, and linking data</i> Holly Little Department of Paleobiology National Museum of Natural History, Smithsonian Institution
------	--

9:30	 <i>Creating the Building Blocks for Biodiversity Research</i> Diane Zorich Digitization Program Office Smithsonian Institution
------	---

9:40	 <i>Translating digitization and data mobilization capacity into the geosciences and other disciplines</i> Gary Motz Indiana Geological and Water Survey Indiana University
------	---

9:50	Session Q&A
------	------------------------

10:00



[Virtual NMNH Botany Conveyor Tour](#)
Sylvia Orli, Eric Schuettpelz, Victor Shield

10:20



Supporting regional capacity development to address challenges within biodiversity data mobilization and use
Maheva Bagard Laursen
Global Biodiversity Information Facility (GBIF)
Denmark

10:30



Digitizing large insect collections - approaches and prioritization for sharing meaningful data and media
Torsten Dikow
Department of Entomology
National Museum of Natural History, Smithsonian Institution

10:40



user experience (UX) for citizen science
Mia Ridge
The British Library

10:50

Session Q&A

Time
EDT

Session Two
1:00 - 3:00 PM

Poster Session

Topic: Digitization Spotlights (Poster Session)







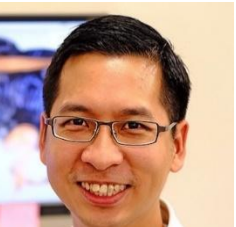
Visit kumospace.com/biodigi2021 for an opportunity to view posters, interact with poster presenters, and socialize with colleagues and friends.

Pre-recorded poster presentations are additionally available for asynchronous viewing on the [event wiki page](#).

Time
EDT

Session Three
6:00 - 8:00

Oral Presentations
Topic: Community

6:00		<p><i>Mobilizing the US Paleontological Collections Community</i> Talia Karim Invertebrate Paleontology CU Museum of Natural History, University of Colorado</p>
6:20		<p><i>South African Natural Science Collections Data</i> Michelle Hamer Natural Science Collections Facility, South African Biodiversity Institute, South Africa</p>
6:30		<p><i>DigiVol: Crowdsourcing Collection Digitisation</i> Paul Flemons Digital Collections and Citizen Science Australian Museum</p>
6:40		<p><i>WeDigBio - Next steps as the community blends a global data campaign, virtual science festival, and local outreach opportunities</i> Austin Mast Department of Biological Science, Florida State University</p>
6:50	Session Q&A	
7:00		<p><u>Virtual NMNH Tour: Digitizing Molluscs Through Space and Time</u> Kathy Hollis, Holly Little, John Pfeiffer, Ellen Strong, Bill Moser</p>
7:10		<p><i>We're in this Together: Arctos Community and the Shared Data Ecosystem</i> Emily Braker Vertebrate Zoology CU Museum of Natural History, University of Colorado</p>
7:20		<p><i>Automation helps accelerate data mobilization for small regional collections</i> Wei Song Hwang Lee Kong Chian Natural History Museum, National University of Singapore</p>

7:30



*International research (data) infrastructure frameworks as leverage for sustainable natural science collections funding;
The European example of the Distributed System of Scientific Collections*

Dimitris Koureas

Naturalis Biodiversity Center, Netherlands

7:40



Regional collaboration through GBIF: Latin America and the Caribbean

Anabela Plos

Global Biodiversity Information Facility (GBIF)

Argentina

7:50

Session Q&A



DAY TWO: THURSDAY, 23 SEPTEMBER 2021

Time
EDT

Session Four
8:30 – 11:00 AM

8:30



Welcome
Global Collaboration for Improved Biodiversity Data
Joe Miller
Global Biodiversity Information Facility (GBIF)

Oral Presentations

Topic: Natural History Collections in Grand Challenges

8:50



The grandest of challenges – why we need digitization, machine learning, and artificial intelligence in a time of global change
Rebecca Johnson
National Museum of Natural History, Smithsonian Institution

9:10



Connecting Natural History Collections and Genomics to Meet Grand Challenges
Pam Soltis
iDigBio
Florida Museum of Natural History, University of Florida

9:20



How can archive collections support pandemic preparedness?
Yvonne Linton
Walter Reed Biosystematics
National Museum of Natural History, Smithsonian Institution

9:30



Developments in valuing digital collections data
Helen Hardy
The Natural History Museum, London




9:40

Session Q&A

9:50

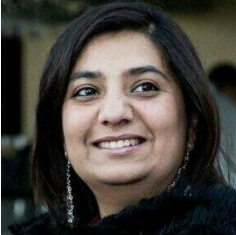




[Virtual NMNH Tour: Digitization and Replication of Cultural Objects](#)
Eric Hollinger, Vince Rossi

10:10		<p><i>Extended Digital Specimens on the Internet</i> Alex Hardisty School of Computer Science and Informatics Cardiff University, United Kingdom</p>
10:20		<p><i>Bird's Eye View of the Past 10 years of Collections Digitization in the U.S.</i> Barbara Thiers New York Botanical Garden</p>
10:30		<p><i>speciesLink's data cleaning strategies</i> Dora Ann Lange Canhos CRIA - Centro de Referência em Informação Ambiental, Brazil</p>
10:40	Session Q&A	

Time EDT	Session Five 1:00 – 3:30
-------------	---

Oral Presentations
Topic: What's Next

1:00		<p><i>The African Coordinating Mechanism for Biodiversity Information Management: Strengthening Science, Technology, and Innovation in Africa</i> Fatima Parker-Allie South African National Biodiversity Institute</p>
1:20		<p><i>Innovation with participation</i> Johannes Vogel Museum für Naturkunde, Germany</p>
1:30		<p><i>New Drivers for Digitization at NMNH.</i> Carol Butler Collections National Museum of Natural History, Smithsonian Institution</p>

1:40



Bee management: Establishment of the first reference collection of West African bees

Drissa Coulibaly

University Peleforo Gon Coulibaly of Korhogo

1:50

Session Q&A

2:00



[Virtual NMNH Tour: Entomology](#)

Floyd Shockley, Torsten Dikow

2:20



Paleo, μ CT, Deep Learning pipelines and research applications

Stewart Edie

Department of Paleobiology

National Museum of Natural History, Smithsonian Institution

2:30



Automated use of digitized specimens in plant science

Pierre Bonnet

CIRAD, France

2:40



Biological collections as research and development infrastructure

Scott Miller

Smithsonian Institution

2:50



Hopping from NEON specimens and samples into a web of linked data

Paula Mabee

National Ecological Observatory Network (NEON)

3:00



The community has spoken. How now, can we translate words into action?

Reed Beaman

National Science Foundation

3:10 **Session Q&A**

Time **Session Six**
EDT **6:00 - 8:00 PM**

Panelist Discussion

Topic: Looking to the Future of Digitization



Conference Wrap-up

Oral Presentation Abstracts

Last Name	First Name	Affiliation	Title	Abstract
Bagard Laursen	Maheva	Global Biodiversity Information Facility (GBIF)	Supporting regional capacity development to address knowledge gaps within biodiversity data mobilization and use	<p>The Global Biodiversity Information Facility (GBIF) is an international network and research infrastructure that provides open access to more than 1.8 billion species occurrence records used in hundreds of peer-reviewed publications each year, including policy-relevant research on invasive species, climate change, and conservation.</p> <p>The need to integrate scientific knowledge about biodiversity into decision-making is widely recognized. However, important gaps remain in both the availability of biodiversity data and the capacity to use these data, impacting conservation efforts, resource management and biodiversity-related intergovernmental processes, among others.</p> <p>The GBIF-led Biodiversity Information Fund for Asia (BIFA) — funded by Japan Ministry of Environment — and Biodiversity Information for Development (BID) programme — funded by the European Commission — aim to enhance capacity for effective mobilization and use of biodiversity data in research and policy in the target regions of Asia, Africa, the Caribbean and the Pacific. This presentation will highlight how the BIFA and BID programmes support the development of regional capacity to empower local communities of practice that have to date mobilized over 1,600,000 species occurrence records contributing to filling important data gaps. We will share our experiences in supporting the development of regional capacity and describe the different mechanisms we developed for this. We will also illustrate how processes developed under the BIFA and BID programmes can be replicated in other contexts and regions</p>
Beaman	Reed	National Science Foundation	The community has spoken. How now, can we translate words into action?	<p>The biological collections community has articulated a vision and strategy for digitizing and mobilizing data associated with biological collections through recent reports from the National Academies of Sciences, Engineering, and Medicine (NASEM) and the Biodiversity Collections Network (BCoN), and outcomes of the Biodiversity Literacy in Undergraduate Education (BLUE) project. In the United States, acceleration of digitization initiatives traces back more than ten years to the Interagency Working Group on Scientific Collections “Green Report” and the Network Integrated Biological Alliance (NIBA) strategic plan. In the time since, we can cite progress that has transformed the biocollections community and the use of biological collections. This new generation of reports chart revised paths and priorities, and remind us that still, we have a long way to go. What has changed? First, the number and breadth of stakeholders has expanded. With new stakeholders, new priorities arise in an evolving community of practice. Second, recommendations from NASEM and BCoN expand the concepts of what and how we should be digitizing collections, to include ancillary data and research quality digital media, e.g., 3D images, and third-party resources, that enhance the research value of physical vouchers. Third, this all needs to be indexed, linked, and made accessible in a cyberinfrastructure environment that adds new capability, and challenges our ability to harness those capabilities, at every turn. Finally, the new digital landscape provides access well beyond those who have traditionally had access to and used biological collections. How do we enable the next generation, globally, to have the bandwidth to do even more?</p>
Blackburn	David	Florida Museum of Natural History	Increasing the Reach and Impact of Scientific Collections through 3D-imaging	<p>Digital representations of physical specimens and objects have made natural history collections more accessible and transformed research. Yet whereas information about specimens is accessible via a quick web-based search, access to the physical specimens remains much as it was for early naturalists. With the advent of new imaging technologies using visible light or x-rays to create digital, high-fidelity, three-dimensional representations of physical specimens, objects in natural history collection can now be more accessible than ever before to a broad audience. With examples from the NSF-funded openVertebrate (oVert) Thematic Collections Network, I will discuss how these digital 3D media (1) transform the scale and speed at which research questions can be addressed, (2) connect students and their teachers to both museums and scientists, and (3) create opportunities for reaching audiences that do not typically interact with scientific collections. The last will include discussion of how we might engage a broader audience in research using 3D data via on-line community science platforms. Last, I will discuss how usage data from repositories might inform our priorities for creating new 3D representations that are useful to specific audiences.</p>
Bonnet	Pierre	CIRAD, France	Automated use of digitized specimens in plant science	<p>Natural history collections are one of the richest sources of information in the world for identifying, describing and studying the diversity of living organisms. Their recent mass digitization offers new opportunities for their automated exploitation thanks to artificial intelligence, to meet new scientific challenges.</p> <p>Two of these challenges will be addressed in this presentation: (1) the first one is the development of new approaches for the field plant species identification, especially for tropical ones which are less well known. The international PlantCLEF challenge, which has been organised for the past ten years, contributes to progress in this field through the implementation of an annual evaluation campaign, which involves dozens of research teams. For several years now, this campaign has been focusing on the capacity of automated visual methods to correctly identify images of living plants, thanks to the use of a large number of digitized herbarium specimens. (2) The second challenge presented will be the experiment of deep learning techniques for the study of the plant phenology, which has shown in recent years that it is now possible to automatically predict the phenological status of plants on digitized plant specimens.</p> <p>This presentation will be an opportunity to examine the complementarity between field data and data from natural history collections, and to encourage the development of mixed approaches taking advantage of these two data sources.</p>

Oral Presentation Abstracts

Last Name	First Name	Affiliation	Title	Abstract
Braker	Emily	Arctos	We're in this Together: Arctos Community and the Shared Data Ecosystem	The Arctos Community is a consortium of museum collection and data managers from diverse disciplines and institutions that work in a collaborative collection management system. A shared data environment compels Arctos users to collectively manage code tables and authorities across several nodes of the database, such as geography, taxonomy, and agents, in an effort to promote data normalization and discovery. As a result, proposed vocabulary or changes to the functionality of Arctos undergoes a community decision-making process, ensuring that database developments are guided by Arctos users and reflect community needs. By integrating datasets across biological, geological, and cultural collections, Arctos brings together varied perspectives and data types that have led to innovative and broadly beneficial new features and capabilities. Learning Arctos is also collaborative, including a community-sourced documentation and tutorial wiki, a new member collection mentorship program, and regular virtual office hours. Active participation through GitHub, and Arctos Working Group meetings and webinars maintains a steady stream of dialogue among Arctos members, resulting in a transparent culture of sharing as well as collaborative grants and research. Though consensus-building presents its own challenges, the Arctos model encourages community-based solutions, workflow efficiencies, and data quality improvements, ultimately advancing best practices in museum data management and data utility.
Butler	Carol	National Museum of Natural History	New Drivers for Digitization at NMNH.	The Smithsonian National Museum of Natural History has been creating digital records of its collections for decades. These began as basic catalog records for inventory purposes and to serve as pointers into the collections. Over time, expectations of digital records have changed, with more data being included or linked to individual records. Today's records are accessible via websites and are harvested by aggregators such as GBIF. We strive to provide high quality, rich data representing as much of the total collection as possible. Users have been predominantly the scholarly community, and our projects and concepts have been tuned to represent the interests and perspectives of this group. Now, NMNH is broadening its practices to be accessible, welcoming and relevant to more diverse audiences and users. New drivers of diversity and inclusion are shaping current and upcoming projects, such as review of descriptive terms, use of new labeling, and inclusion of previously unavailable histories.
Canhos	Dora Ann Lange	CRIA (Centro de Referência em Informação Ambiental)	speciesLink's data cleaning strategies	speciesLink's development began in 2001. Based on existing experiences in Mexico and Australia, CRIA's team concluded that data providers should maintain full control over their data, retaining any sensitive data in their own platform. Another conclusion was that the system under development would not filter incoming data to eliminate errors or incomplete data. Our assumption was that having full control over all public data available online would be a way to keep a connection with the data providers and help them improve the quality of their data. Throughout the years, CRIA's team developed a data cleaning report online that is reprocessed, every time data providers update their records. The report highlights "suspect" records for taxonomic and geographic data. Another tool is the annotation system. Data users can provide their opinion as to the specimen's name, a new determination, or correct any other information included in the specimen occurrence record. The system automatically sends an email with this information to the curator and appends the comment to the record as an annotation. Accepting inconsistent and incomplete data could represent a problem for data users that, in most cases, wish to retrieve "clean" or high quality data for their analysis. To work around this potential problem, the search interface includes a number of filters that are able to eliminate (or include) suspect data.
Coulibaly	Drissa	Unité de Formation et de Recherche des Sciences Biologiques, Département de Biologie Animale, Université Peleforo Gon Coulibaly de Korhogo, Ivory Coast	Bee management: Establishment of the first reference collection of West African bees	The presentation at the conference will show some of the tools and strategies developed Agriculture is very important to the economic development for West African countries. Unfortunately, since a few years we are recording a yield decrease of many cash crops. If the variability of precipitation, caused by climate change effects, is one of factors contributing to this decrease, many scientific studies show that pollinators decline is the main cause. Moreover, ecosystem degradation, depletion of plant species, habitat fragmentation, use of insecticides and global warming constitute severe threats to these insects (e.g. bees) being clearly at risk of extinction. Indeed, bees are the most important pollinators of many crops and wild plant species. The ecosystem service of pollination provided by these insects is crucial to maintain overall biodiversity and to secure crop yields worldwide. Especially improving the livelihood of smallholders in developing countries through higher crop yields is essential for achieving global food security and poverty reduction. The knowledge of bees is still a crucial tool in the management of sustainable and sufficient agriculture. Increasingly, bee products such as honey, propolis, wax, are used by people in many parts of Africa and are an important asset for the commercial sector. Despite the great ecological and economic importance of bees as pollinators hardly anything is known about the bee species in West Africa. The establishment of a reference collection thus makes it possible to better understand and protect the bee community. The GBIF small grant made it possible to mobilize several data on bees. A total of 43358 bee specimens were caught with coloured pantraps in near-natural savannah habitats and in nearby fields in Burkina Faso and Cote d'Ivoire. Among these specimens, 115 bee species belonging to 34 genera and 4 families (Apidae, Megachilidae, Halictidae and Colletidae) were identified. All data were recorded in Darwin form and published on GBIF portal. The reference collection is accessible to all users with an even stronger demand for use in universities. It is therefore a real scientific database that we have today to convince policy makers and help primary users such as farmers, students and researchers

Oral Presentation Abstracts

Last Name	First Name	Affiliation	Title	Abstract
Dikow	Torsten	National Museum of Natural History	Digitizing large insect collections - approaches and prioritization for sharing meaningful data and media	The digitization of specimens in natural history collections poses a number of challenges no matter what resources are available because the sheer number of objects is astonishing. This is particularly so for (pinned) insect collections where even a small natural history collection at a museum or university will easily harbor several Million specimens and usually has an incredibly high specimen-to-staff ratio. How can curatorial staff digitize the collection in order to share research-grade data and media meaningfully? Digitization approaches and prioritizations based on the effort of a team of two (a museum specialist and a researcher/curator) overseeing a collection of 1.7 Million specimens including 15,000 primary types at the Smithsonian USNM are considered. Our experience shows that focusing on specific units streamlines the work today and will so in the future. Units can be a cabinet, a drawer, a unit tray (holds all specimens of a species or parts thereof) or a family/taxon and such units should always be digitized as a whole and completed. Digitization approaches include making specimen data and high-quality photographs of primary type specimens available, which has been a challenge for our collection. Specimen-level databasing was directed towards groups of organisms for which fewer than 100 specimens are in the collection. These are likely rare taxa that are seldom collected and uncommon in natural history collections and making them immediately accessible to the scientific community and the public at large at GBIF might have the biggest impact. Specimen-level databasing is also performed for research purposes, following field-work, and specimen handling during incoming or outgoing loans. This presentation will elaborate on these approaches and highlight others.
Eddie	Stewart	National Museum of Natural History	A few applications of μ CT and deep learning to morphological data in macroevolutionary research	Biodiversity science increasingly interrogates large, multi-dimensional datasets. Among those dimensions, morphology can be difficult to capture en masse at broad phylogenetic scales. Micro-CT has become a good option for high-throughput 3D imaging—especially for fossils that are often encased in matrix. Digital excavation of fossils using phase differentiation takes minutes instead of weeks commonly required for physical preparation. However, such segmentation struggles for fossils composed of similar materials to their surrounding matrix—i.e. many marine invertebrates. Deep learning can separate these fossils from their matrix based on textural differences, enabling analyses of fossil and modern specimens in the same 3D morphological framework. Beyond segmentation, deep learning can quantify and underlie morphospaces of complex shapes, such as the many spines, ribs, and flanges of bivalve sculpture that long-established methods struggle to characterize. Used together, micro-CT and deep learning can build largescale morphological datasets, helping us to examine how this dimension of biodiversity responds to gains or losses in taxonomic and/or ecological diversity across time and space.
Flemons	Paul	Australian Museum	DigiVol: Crowdsourcing Collection Digitisation	In 2011 the Australian Museum and the Atlas of Living Australia launched a prototype collection label transcription website – the Biodiversity Volunteer Portal. In the intervening ten years that prototype has developed into the hugely successful DigiVol website that has seen more than 70 institutions around the world use DigiVol to engage more than 10,000 volunteers to “transcribe” more than 3.8 million tasks across a broad variety of collections. These tasks include transcribing text from labels and field notes as well as tagging images with species identifications or traits. The Australian Museum has incorporated DigiVol into its digitising strategy, building an onsite lab that engages volunteers in capturing the images from which data is then transcribed through the DigiVol website. Running DigiVol and using it as a core component of a digitising strategy has many benefits and its fair share of risks and challenges, particularly in the age of Covid. As DigiVol continues to find new applications of crowdsourcing to collection digitisation the promise of a synergy with AI has promised much but as yet has delivered very little.
Hamer	Michelle	South African National Biodiversity Institute	South African Natural Science Collections Data	The Natural Science Collections Facility (NSCF) is a research infrastructure project funded by the South African government's Department of Science & Innovation and includes the 16 major museums, science councils and universities that together hold more than 30 million natural science collection objects. The network is co-ordinated by a Hub at the South African National Biodiversity Institute. In addition to securing the collections, there are objectives related to data mobilisation, upgrades and making data accessible for research and decision-making. The decision to standardise software was taken early in the establishment of the NSCF and Specify is used for most of the zoological and palaeontology collections, and Brahm's for the herbarium collections. Common standards and protocols for all aspects of collection data management have been developed and published in the NSCF Collection Management & Curation Manual. Current projects include ensuring that all the vertebrate collection data are as complete and accurate as possible, which includes dealing with unaccessioned material, verification of identifications, and checking data quality. A second project is aimed at completing data capture for herbarium specimens, with a common focus on specific families, scanning of specimens being carried out within the herbaria, with centralised data capture and distributed quality control. Training in the use of Specify has been run and relevant staff have been supported to attend a data management course. Technical support is provided by the Hub staff to institutions. While substantial progress has been made, mechanisms for increasing the pace of all activities without increasing cost is essential. The main challenges, however, are related to getting all institutions and staff to accept and adopt common standards and to make data openly accessible. While leadership development and transformation interventions have been implemented, individualist cultures are old and often deeply entrenched.

Oral Presentation Abstracts

Last Name	First Name	Affiliation	Title	Abstract
Hardisty	Alex	School of Computer Science and Informatics, Cardiff University, United Kingdom	Extended Digital Specimens on the Internet	<p>In an era of data-driven science, the heterogeneity of collection management systems and the systemic absence of linkages between digitized, vouchered specimens and between specimens and the secondary data derived from them, such as genetic data, sub-article level data embedded in literature, ecosystem data, chemical composition data, etc. makes it hard to adequately answer contemporary scientific questions for the challenges humankind faces on planet Earth. This difficulty is compounded by the geographically dispersed nature of c.3 billion specimens in collections around the world and the increasing difficulty to scientifically curate across taxonomic groups, especially for smaller institutions.</p> <p>Extending the digital representation of a specimen to attach all kinds of data to it and to link it to other specimens, making this representation actionable and FAIR i.e., readable by humans, processable by machines, and driving social transformation towards community curation whilst maintaining provenance and attribution for elevation of trust-levels are three ideas brought together by Extended Digital Specimens on the Internet as a new abstraction for the next generation of 'biodiversity data as infrastructure'.</p>
Hardy	Helen	The Natural History Museum London	Developments in valuing digital collections data	<p>Globally, the case for natural science collections digitisation has always drawn on case studies of how collections inform our responses to grand challenges for humanity and the planet. But valuing these contributions systematically has always been difficult, with a complex chain between digitisation, research and impact, and the potential for very large benefits from unpredictable or one-off events. Government funding has been very important to digitisation in many countries - to keep both public and philanthropic funding flowing at the necessary scale, we are likely to need further evidence of impacts, and monetised estimates of the expected return on investment. Recently, further efforts have been made to develop valuation methodologies, including the USA's Advisory Report on Economic Analyses of Federal Scientific Collections (Schindel, David E. et al https://doi.org/10.5479/si.13241612.v1); and to provide estimates of economic value, for example in the related area of taxonomic discovery in Australia (Deloitte 2020, Cost benefit analysis of a mission to discover and document Australia's species).</p> <p>The Natural History Museum London are now working with Frontier Economics to develop estimates of the economic impact of natural science collections digitisation in the UK. This presentation will discuss the theory of change model developed for this work, identifying key pathways to impact for collections data. Modelling is focusing both on a 'top down' approach considering the value of research, and on a thematic approach looking at the 'size of the prize' for particular grand challenges or areas of economic activity such as health or conservation. The study is not yet complete, but we aim to share emerging order of magnitude figures, as well as some of the reasoning and narrative behind them.</p>
Hwang	Wei Song	Lee Kong Chian Natural History Museum, National University of Singapore	Automation helps accelerate data mobilization for small regional collections	<p>Small regional collections outnumber globally representative mega-collections, but they often struggle to mobilize data effectively for various reasons, with limited resources being an often-cited challenge. Such collections bring value to current digitization initiatives by filling significant knowledge gaps in both spatial and temporal dimensions and provide finer resolution datasets not available elsewhere. We share our experience here from a GBIF BIFA-funded project to digitize 10,000 butterflies from Peninsular Malaysia and Singapore, which greatly benefited from the adoption of a few automation techniques that accelerated work processes and improved the deployment of precious staffing. When scaling up digitization efforts, routine tasks such as renaming of files and image editing becomes time-consuming affairs and prone to human errors. We introduced OCR reading capabilities and automated image editing into our digitization pipeline as a solution and recommend these practices for future digitization efforts.</p>
Johnson	Rebecca	National Museum of Natural History, Smithsonian Institution	The grandest of challenges – why we need digitization, machine learning, and artificial intelligence in a time of global change	<p>We are in a time of unprecedented global change – driven by human activity. At the Smithsonian's National Museum of Natural History, we are custodians of the world's largest biological collections. The collections support foundational discovery science and collectively represent a substantial baseline of biodiversity and ecosystem knowledge, including how these have changed over time. While NMNH holds the world's largest natural history collections, much of the information about these specimens is dark data – meaning that the data may not be discoverable or accessible. Creating accessible records will not merely illuminate our collections to understand the origins of changes to the Earth's biodiversity and ecosystems, but it will also be the key to understanding and mitigating the range of catastrophic global changes in our near future. Illuminating dark data at the scale of the NMNH collections is a challenge that we are tackling harnessing the power of mass digitization and new technologies.</p> <p>To scale up digitizing nearly 150 million specimens and their attendant metadata from our collections, NMNH is using Machine Learning and Artificial Intelligence tools to process these large data sets. Our goal is to develop ML and AI as tools that can be used by the research community and the public. While these tools accelerate the processing of mass data, they require the expertise and question-driven interrogation by human experts to draw conclusions and generate solutions. The baseline knowledge of NMNH's accessible collections has been used in numerous ways to provide answers to questions about origins and fate biodiversity in the context of global change over time. Our goal is to continue to make these data open and accessible – we believe they are essential for the research community, educators, policy makers, and the public to address the grand challenge of global change.</p>

Oral Presentation Abstracts

Last Name	First Name	Affiliation	Title	Abstract
Karim	Talia	Invertebrate Paleontology, CU Museum of Natural History, University of Colorado	Mobilizing the US Paleontological Collections Community	The paleontological community in the United States has a history of coming together to work on large collaborative projects (e.g. Paleobiology Database, PaleoPortal) and the fossil collections managed by this community have historically been well cataloged and were some of the earliest to be electronically databased. When the most recent push for mass digitization of collections started about a decade ago, the Paleontological Community in the US was well positioned due to these previous efforts. Financial and logistical support from the NSF ADBC Program and iDigBio played a key role in the recent mobilization of the Paleontological Collections Community by supporting numerous professional development opportunities. It would be an understatement to say that the sustainability of these mobilization efforts would not have been possible without numerous dedicated people involved who have shown a strong commitment and desire for making fossil collections data more accessible. The results of mobilizing this community include publications, identification of problems with paleontological data integration by large aggregation portals, discussions of issues with existing data standards and solutions put forward, and the establishment of a virtual discussion group focused on paleontological data issues and creating shared online documentation. In closing, the past decade of digitization has been a rallying cause that has increased the cohesion of the Paleontological Collections Community and reinforced its ability to identify issues and propose solutions related to fossil collections data together, forming a strong foundation that will carry it forward into the next decade of digitization.
Koureas	Dimitris	Naturalis Biodiversity Center Distributed System of Scientific Collections	International research (data) infrastructure frameworks as leverage for sustainable natural science collections funding; The European example of the Distributed System of Scientific Collections	The Distributed System of Scientific Collections (DiSSCo) is a newly established priority Research Infrastructure (RI), being developed in the context of the European Strategy Forum on RIs (ESFRI). DiSSCo aims at virtually uniting natural science collections in Europe. It focuses on substantially lowering the access threshold to collections for a multitude of users. To achieve that it puts forward a continental scale programme that delivers organisational, financial, and operational shared capacity across 130+ natural science collections. From the user perspective, DiSSCo introduces a framework for a step-change by massively improving the capacity of scientists to discover, access and analyse complex and previously disjointed information deriving from the study of the vast European Natural Science Collections (NSC). Besides the opportunities linked to the common European future for our collections, the inclusion of DiSSCo in the European roadmap of priority RIs has also led to a transformational shift of the perception of national funders regarding the role and future of NSCs at the national level. Despite the fact that NSCs have been essential physical infrastructures for science for more than half of a millennium, they have not been perceived as such by most national-level science policy and funding agencies. Instead, their value had been predominantly seen in their function as cultural and historical objects. In this presentation, we discuss the European landscape of infrastructures and describe how the inclusion of collections in the European priority roadmap of infrastructures, has enabled national NSCs to tap into previously unknown sources of long term and sustainable new funding.
Linton	Yvonne-Marie	Walter Reed Biosystematics Unit (WRBU), Smithsonian Institution Museum Support Center	How can archive collections support pandemic preparedness?	Pandemic preparedness is reliant on accurate pathogen prediction, rapid pathogen detection and successful pathogen mitigation strategies. Today, as never before, the World understands the devastating impact of zoonoses – where known (or novel) pathogens, normally restricted to wildlife spillover into humans with devastating consequences. With humans, livestock and wildlife sharing increasingly limited resources, the opportunity and severity of infectious and vector-borne diseases are increasing in severity and frequency and the ability to predict risk is critical to government, military and civilian health stakeholders alike. In her presentation, Dr. Linton discusses the historical role of the U.S. military in building the U.S. National Mosquito Collection, and the value of those comprehensive historical archive specimens in the fight against today's global emerging infectious disease threats.
Little	Holly	National Museum of Natural History	Progress in data mobilization: standards, coordination, and linking data	What does it actually mean to mobilize data in an increasingly interconnected, global, digital ecosystem? Within the past decade of digitization there have been major shifts in our practices and perspectives on mobilizing and sharing specimen information. As we work to ensure our data is interoperable, our local practices are increasingly impacted by global needs. This talk will share how approaches to data mobilization have evolved, the impact of standards development and community coordination in that change, and the important steps ahead of us to reach new possibilities for expanding our web of natural history data.
Mabee	Paula	The National Ecological Observatory Network (NEON)	Hopping from NEON specimens and samples into a web of linked data	The National Ecological Observatory Network (NEON) is a continental-scale Observatory, operated by Battelle for the National Science Foundation, and designed to collect standardized, long-term, open access ecological data, specimens, and samples to better understand how U.S. ecosystems are changing. Over its 30-year period, 100,000 biological, genomic, and geological samples and specimens from terrestrial and aquatic sites will be ingested by the NEON Biorepository annually, for a total of over 3 million samples of 70 different types related to 181 NEON data products. Unlike most natural history collections with long historical legacies, collections focused exploration sampling for organisms specific to geographic regions or taxa, used primarily by systematists, and under gradual programs of digitization – the collections of the NEON Biorepository were initiated recently, amassed through strategic and regular resampling using standardized protocols for ecoclimatic coverage, and tailored to address grand challenge questions. NEON specimens and samples are born digital in an extended web of data that includes a rich set of standardized contextual data and collected through highly coordinated efforts of humans and technology. Developed using over 11,000 sensors in diverse instruments including lidar, imaging spectrometers, phenocams, soil moisture sensors, flux towers, and aquatic buoy mounted water quality sensors, NEON's related environmental data products capture atmospheric dynamics, biogeochemistry, ecophysiology, land cover and processes and are connected to the specimens in the NEON Biorepository. New visualizations and tools are needed for researchers to traverse and fully exploit this linked open data aggregate for creative and interdisciplinary research.

Oral Presentation Abstracts

Last Name	First Name	Affiliation	Title	Abstract
Mast	Austin	iDigBio, Florida State University	WeDigBio—Next steps as the community blends a global data campaign, virtual science festival, and local outreach opportunities	The Worldwide Engagement for Digitizing Biocollections (WeDigBio) project provides resources for biodiversity research collections to effectively and creatively engage their local communities and global internet users in digitization of specimens and associated content, especially during a pair of annual WeDigBio events, with an emphasis on increasing the science literacy of participants, growing the diversity of science contributors, and expanding the relevance of collections to solving humanity's grand challenges. WeDigBio has just completed a strategic planning process for 2021–3, including the convening of its first board. The talk will provide an overview of plans for the three-year period, including a vision for the WeDigBio Science Festival—a blended virtual and face-to-face experience celebrating biodiversity collections. All biodiversity collections are welcome to participate!
Miller	Joe	Global Biodiversity Information Facility (GBIF)	Global Collaboration for Improved Biodiversity Data	The integrated challenges of climate change and biodiversity loss have increased the urgency for more and better data for research, policy and decisions from local to global scales. This can be best achieved through global cooperation to use the limited funding in our community as we cannot afford unnecessarily duplicate efforts. The alliance for biodiversity knowledge, managed from the GBIF Secretariat, is a framework to build a community to better to align efforts to deliver current, accurate and comprehensive data, information and knowledge on the world's biodiversity. This introduction to Day 2 will include some background on the alliance, describe some exemplar alliance projects of global collaboration and suggest some next steps to better coordinate global biodiversity knowledge.
Miller	Scott	Smithsonian Institution	Biological collections as research and development infrastructure	The US Government's Interagency Working Group on Scientific Collections has helped raise awareness of collections, recognize collections as infrastructure, raise standards and documentation for collections management, increase discoverability of collections (https://www.gbif.org/grscicoll), and increase coordination across government owned and funded collections. Recent activities include (https://iwgsc.nal.usda.gov): a workshop on the relevance of traditional biodiversity collections to biosecurity, a study of economic models for describing the value of collections, and an updated general report on the uses and impacts of collections. Future opportunities and challenges include addressing long-term stability of collections, persistent identifiers and linked data, access and benefits sharing in biological collections, and diversity, inclusion and workforce development. A recent study for the National Invasive Species Council (doi: 10.1007/s10530-019-02147-x) on the government's capacity to identify species is relevant to future digitization priorities, including the need for taxonomic authority files, digitization of type specimens, and DNA reference libraries.
Motz	Gary	Indiana State University	Translating digitization and data mobilization capacity into the geosciences and other disciplines	The global pandemic has illustrated the need for remotely accessible natural science collections. Object-oriented research, sampling, project completion, and other parts of the research data lifecycle were interrupted by an abrupt shift to work-from-home. Curators, collections managers, students, and visitors were immediately denied access to specimen and sample data that exists in formats limited, in many cases, exclusively to in-person evaluation. Geological surveys and geosciences departments, in particular, were forced to reckon with discrepancies in data available electronically as collections-based data is often made available only in generalities or collections-level inventories. Few US-based institutions have the capacity to deliver item-level catalogues, metadata, and quality georeferenced sample information for geological collections objects, much less associated datasets that might include mineralogical composition or determination, geochemical characterizations, or other analytical parameters useful in relevant research programs. Interconnected organizational efforts and intentional, highly-collaborative structures that derive utility from strategies employed by parallel disciplinary practices will strengthen human capacity development efforts by building upon shared solutions to common problems. Increased reliance on scalable cyberinfrastructure and a lasting commitment to FAIR data standards (promoting equity in the findability, accessibility, interoperability, and reusability of collections data) will be the pathway to success for disciplines like the geosciences that aim
Parker-Allie	Fatima	South African National Biodiversity Institute	The African Coordinating Mechanism for Biodiversity Information Management: Strengthening Science, Technology, and Innovation in Africa	Africa is one of the most megadiverse continents in the world. Here biodiversity plays a critical role in sustainable development, provides vital ecosystem services, and is one of our greatest regional assets. The benefits of biodiversity are crucial to key economic sectors i.e., forestry, agriculture, fisheries, tourism, health, and energy, and to providing solutions to sustainable development and poverty alleviation on the continent. The management of our natural assets and the information related to this are crucial. Ultimately, we cannot manage what we cannot measure. To date South Africa has mobilised more than 30 million primary biodiversity records to GBIF, with a key focus on data mobilisation, capacity enhancement, developing and supporting networks and supporting the data-science and policy landscape. SANBI-GBIF has also actively been driving strategic efforts to support biodiversity informatics endeavours on the African continent, through leading and supporting engagements through the African Coordinating Mechanism. The ACM is a regional structure to support biodiversity information management on the continent and falls within the framework of the GBIF-Africa Nodes activities. These efforts also connect to the African Union's Science, Technology, and Innovation Strategy for Africa 2024, with an aim to accelerate Africa's transition to an innovation led, knowledge-based economy. STI has a fundamental role to play in achieving the Sustainable Development Goals, with mega-science initiatives such as GBIF, its knowledge management platform, the big-data landscape, and the large network of players having huge potential towards driving STI impact. South Africa's future is inextricably linked to that of the rest of the African continent, and therefore the potential of STI for African development and continental integration needs to be fully leveraged and aligned to the African Union 2063 Agenda.

Oral Presentation Abstracts

Last Name	First Name	Affiliation	Title	Abstract
Plos	Anabela	Global Biodiversity Information Facility - GBIF, Argentina	Regional collaboration through GBIF: Latin America and the Caribbean	<p>The LAC region, within the GBIF community, is particularly active. It has the particularity of sharing, in a large part of its territory, the same language. This does not restrict activities and initiatives with non-Spanish-speaking countries such as Brazil, Suriname or Jamaica, among others.</p> <p>GBIF offers a series of activities, facilities and lines of financing that favor interactions and community work within LAC, as well as with other regions:</p> <ul style="list-style-type: none"> * Participate within the community of translators, mentors, trainers and open data ambassadors, facilitating content in Spanish and supporting nodes and potential new nodes to be trained, transferring capacities and disseminating GBIF activities; * GBIF facilitates an IPT installation for LAC, for providers that do not yet have a national node or for nodes that require it and, thanks to an order managed by Argentina and supported by the nodes in the region, a regional portal will be available shortly, a GBIF window into the region; * Financing for nodes, such as the Capacity Enhancement Support Program (CESP) and the Biodiversity Information for Development (BID) program for the Caribbean, which in addition to financing projects has, for the period 2021-2023, a specialist hired to provide support to projects financed in that period.
Ridge	Mia	British Library	user experience (UX) for citizen science	<p>In the early days of citizen science, each project had to come up with its own interfaces, interaction design, data management and communication practices. Over time, design patterns started to emerge as new and updated projects applied insights from academic research and practical experience. As platforms for building projects such as Zooniverse (and in the humanities, Transkribus, Scribe and FromThePage) developed, the work of designing a new project was simplified, with most of the interface elements provided as part of the platform. New projects still have to design task instructions, provide tutorials and explain how their task will enhance research, but they no longer have to start from a blank slate.</p> <p>However, UX - or user experience design - encompasses all aspects of the end-user's interaction with a project, its services, and its products (NNGroup). User interface elements - graphic design, content design, features and functions - might be the most visible parts of the overall user experience, but marketing and communication, community management, and data quality and verification systems all have a part to play in creating good UX for citizen science projects.</p> <p>When so many elements of the user experience are provided by the platforms commonly used for citizen science projects, thinking about UX might seem less important. However, it's still worth investing time and effort in understanding what barriers potential participants might face to contributing to citizen science projects, and doing what you can to design high quality experiences for participants that ensure richer rewards for participation and contribute to data quality. This talk will present some concepts from UX / human computer interaction that are important for citizen science projects by looking at interfaces from past and present citizen science and humanities projects.</p>
Soltis	Pamela	iDigBio, Florida Museum of Natural History	Connecting Natural History Collections and Genomics to Meet Grand Challenges	<p>Natural history collections – the best documentation of life on Earth – are foundational to all of biology and are increasingly applied to solving societal Grand Challenges related to the environment. Likewise, biodiversity genomics emphasizes the power of the Tree of Life in driving discovery, conservation, and the bioeconomy. Where do natural history collections and genomics intersect? In this short talk, I will address three primary themes: (1) the importance of vouchering all materials used in genomic research and steps toward improving the process and outcomes, with an example on metadata standards and GenBank; (2) natural history collections as sources of materials for genomic analysis; and (3) potential connections to living collections, especially zoos, aquaria, and botanical gardens. Within this framework, I will also briefly consider the role of vouchering in environmental DNA research and possible synergies between natural history collections and microbial collections.</p>
Thiers	Barbara	New York Botanical Garden	Bird's Eye View of the Past 10 years of Collections Digitization in the U.S.	<p>In 2009, representatives from a wide range of collections met to imagine how we might digitize all U.S. natural history collections. These discussions resulted in the NIBA strategic plan, to which NSF responded with the first call for proposals for the ADBC program. As the digitization hub, iDigBio quickly brought together individuals from a wide range of collections types, creating for the first time a national collections community around specimen digitization. This effort led to a wide range of new research uses for digital specimen data and educational initiatives, and broadened the membership and programs of professional societies such as SPNHC and TDWG. The Biodiversity Collections Network Research Coordination Network was funded to address the non-ADBC components of the NIBA plan, hosting a variety of workshops on topics such as how to publicize the importance of collections, data aggregation, and what sort of program should follow ADBC to build on digitization efforts. Comparing the NIBA plans with the ADBC results, we underestimated the technology needed to revolutionize the digitization process. Although the stakeholder base for digitized data did expand significantly during the decade, we did not see the degree of expansion of the funding base for digitization among a wider range of government or private agencies that was predicted in the NIBA plan.</p>
Vogel	Johannes	Museum für Naturkunde, Germany	Innovation with participation	<p>Collection based institutions, and the science associated with it, are facing 'Deep and rapid change' at an unprecedented scale. Two trends collide: huge technological advances and societal change. Very often it appears that our community is at the fringes of such developments rather than engaging pro-actively with stakeholders and drivers. On the technological side we are witnessing rapid advance in automation, AI, genomics, quantum computing to name a few, but all of these will present new and unexpected opportunities and challenges for us. At the same time, society is demanding greater say and participation in science and key challenges come from challenges and demands such as greater openness (in relation to stakeholder and user needs), equality, diversity, inclusion, access and restitution. In this talk I will briefly outline how we are addressing these opportunities with our Zukunftsplan and utilize the huge additional resources we have secured for the next ten years to deliver deep change to our organisation.</p>

Oral Presentation Abstracts

Last Name	First Name	Affiliation	Title	Abstract
Zorich	Diane	Digitization Program Office, Smithsonian Institution	Creating the Building Blocks for Biodiversity Research	<p>The Smithsonian's Digitization Program Office (DPO) was established in 2009 to increase the quality, quantity, and impact of the Smithsonian's digitized collections. Our mass digitization and 3D digitization programs have distinct workflows and processes but share a common purpose: to make Smithsonian collections available to the largest number of people for the widest array of uses.</p> <p>This presentation reviews our projects in the context of natural history collections (specifically those at the National Museum of Natural History (NMNH)), including new digitization projects undertaken during the pandemic. For our 3D digitization program, these efforts include building and honing infrastructure to scale up, preserve, and improve delivery and access to 3D assets. Our mass digitization program is exploring methods for scaling up digital record creation and enhancing specimen metadata, in addition to digital imaging projects.</p> <p>The impact of having large quantities of high-quality digital assets for natural history collections cannot be overstated: in addition to making collections more accessible, they offer better quantitative and visual data than can be collected or detected by physical means, enable computational research methodologies that tackle novel and complex research questions, and provide new opportunities for cross-disciplinary research. Although our work is grounded in digitization structures, workflows, processes, and experimentation, we are aware of how foundational these assets are to biodiversity research, and continue to explore ways to better align our work with the needs of the community.</p>

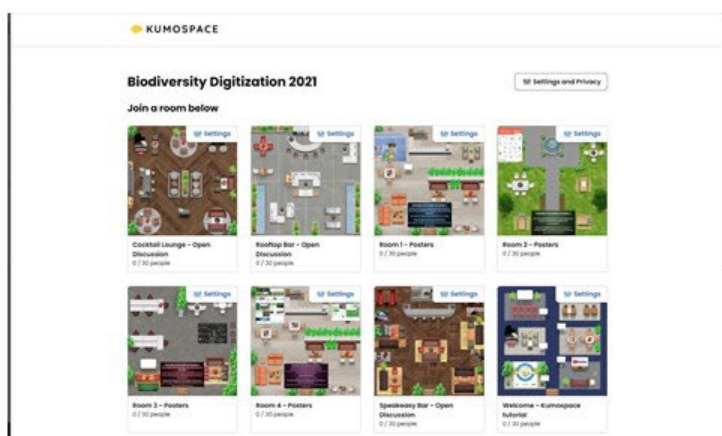
Poster Presentations

Pre-recorded poster presentations are available for asynchronous viewing on the [conference wiki page](#).

Live Poster Session in Kumospace:
Tuesday, September 22 1:00 - 3:00pm EDT

Visit Kumospace for an opportunity to view posters, interact with poster presenters, and socialize with colleagues and friends. The space will be open for the duration of the event

www.kumospace.com/biodigi2021



Poster Presentation Abstracts

Title	Abstract	Authors	Room
<p>Biodiversity Literacy in Undergraduate Education: Data Literacy, Workforce Training, and the Extended Specimen</p>	<p>Biodiversity scientists must be fluent across disciplines; possess quantitative, computational, and data skills for working with large, complex datasets; and have foundational skills and content knowledge from ecology, evolution, systematics and environmental science. As scientists and educators, we can embrace the changing landscape of biodiversity science and leverage the foundational skills that collections have fostered for centuries to help engage, inspire, and build the next generation of biodiversity data scientists. Biodiversity Literacy in Undergraduate Education (BLUE) is a National Science Foundation Research Coordination Network focused on developing strategies and materials to infuse biodiversity data into the core of the undergraduate science curriculum, facilitating broad-scale adoption of biodiversity data literacy competencies, and improving undergraduate biology training to meet increasing workforce demands in data and biodiversity sciences. The BLUE Data Network has four major goals: 1) Cultivate a diverse and inclusive network of biodiversity researchers, data scientists, and biology educators focused on undergraduate data-centric biodiversity education; 2) build community consensus on core biodiversity data literacy competencies; 3) develop strategies and exemplar materials to guide the integration of biodiversity data literacy competencies into introductory undergraduate biology curricula; and 4) extend the network to engage a broader community of undergraduate educators in biodiversity data literacy efforts. We will present on our recent national and international efforts to collaborate across the education, data, and biodiversity science communities; showcase our most recent Open Education Resources; and discuss our on-going efforts to address workforce training needs highlighted in the Extended Specimen and Biological Collections reports from the Biodiversity Collections Network and National Academies of Science, Engineering and Medicine.</p>	<p>Monfils, Anna; Linton, Debra; Ellwood, Libby; and White, Lisa</p>	<p>1</p>
<p>Natural History Education DemoCamp</p>	<p>Natural History Collections offer a wealth of resources for educators and learners. To this end, digitization makes specimens, and the data and images associated with them, more accessible to a wide audience. However, educators and learners must know that resources exist, where to get those materials and how to use them. We, as collections professionals, must promote our outreach materials, make them readily available, and provide the background knowledge for accurate and safe use of specimens, data and supporting resources. The Society for the Preservation of Natural History Collections Education Committee formed in 2020 to help develop, support, and promote formal (K-12, Undergraduate and Graduate) and informal education, and outreach relative to natural history collections and biodiversity science as informed by natural history collections. One of the first premiere initiatives, Qubeshub, established as a site to aggregate natural history educational resources, publishes Open Educational Resources, or OER, through the Natural History Education Portal. Anyone can add an OER resource to this portal. Sharing a resource as an OER on QUBESHUB means your resource will be assigned a DOI and you will have access to usage metrics. The SPNHC Education Committee also hosted the first ever virtual Natural History Education (NHE) DemoCamp Building upon the previous work of the "Share Fair," the NHE DemoCamp strives to share, discover, and discuss educational materials that support a framework in natural history. This year's NHE DemoCamp had over 300 registrants and 21 different live demonstrations that took place over the two days. Educational materials shared varied widely in scope, audience, format, and topic, from how to use R to analyze biodiversity data in the classroom to how to make a compelling outreach video for a general audience. Our poster highlights collections based educational resources and ways to make them accessible to the educational community.</p>	<p>Anna Monfils, Jennifer Bauer, Elizabeth Leith, Molly Phillips, Julia Robinson, Jessa Waters</p>	<p>1</p>

<p>A decade of digitisation to secure, enhance & mobilise CSIRO's biological research collections.</p>	<p>The CSIRO's National Research Collections Australia (NRCA) is a world class science ready biological collections research facility. Its 15+ million specimens collected over 240 years include all major biological groups and cover the entire Australian continent and marine zones.</p> <p>In the last decade, we have made significant advances in bringing our collections together, moving from six collection databases to a single integrated collection management system, and addressed digital challenges in genomic research and associated data, specimen digitisation in image capture and crowdsourcing methods of databasing.</p> <p>While we digitally move to secure and mobilise CSIRO's biological collections, we are also working to further advance digitisation approaches and how that information gets utilised in research and application to real-world problems. Our projects centre around automation, machine learning and artificial intelligence to accelerate digitisation and utilise new data analytics in the future.</p> <p>This presentation will outline our achievements to date to mobilise, secure & make accessible our digitised collections as well as some of the work we are doing to help create a 21st century biological research collection.</p>	<p>Nicole Fisher and Pete Thrall</p>	<p>2</p>
<p>The Atlas of Living Australia. 10 years supporting Biodiversity Research and Decision Making</p>	<p>This poster provides an overview of activities and major accomplishments of the Atlas Of Living Australia (ALA) over the last 10 years that has supported digitisation of Australia biodiversity records. The ALA is an National Research Infrastructure Collaboration Strategy (NCRIS) Australian Government funded initiative. It is Australia's most comprehensive source of biodiversity data through partnerships with over 900 data providers including herbariums and museums, government agencies, research and citizen science programs. It currently holds over 100 million occurrence records and supports a wide range of digitisation platforms including DigiVol, Biodiversity Heritage Library Australia, iNaturalist Australia and BioCollect.</p>	<p>Atlas of Living Australia: www.ala.org.au Hamish Holewa, Deputy Director ALA Hamish.holewa@csiro.au</p>	<p>2</p>
<p>The iDigBio US Collections List: Now Powered By GBIF</p>	<p>iDigBio publishes a list of US Collections that is intended to be a comprehensive list of natural history collections in the United States of America. This list aims to provide access to information and metadata about natural history collections in the United States, including but not limited to collections descriptions, contact information, taxonomic scope of collections, and links to existing recordsets within iDigBio (if applicable). Previously, this list was maintained as a JSON endpoint via GitHub, with updates maintained manually, requiring substantial human involvement and reliance on third party services (e.g. TravisCI) to publish new collections entries or updates to existing collections metadata.</p> <p>In 2020, the iDigBio US Collections List was successfully merged with GrSciColl, the Registry of Scientific Collections at the Global Biodiversity Information Facility (GBIF). GrSciColl and the US Collections List fundamentally share the same goal: enhancing access to information about natural history collections, associated digitized recordsets, and personnel involved with these collections. The US Collections List is now maintained directly on GrSciColl by GBIF and iDigBio staff, and the US Collections List hosted on iDigBio.org is now populated via the GrSciColl Application Programming Interface (API). This merger has resulted in a more streamlined experience for both those maintaining the list and users of the list; changes submitted to US entries on GrSciColl now appear instantaneously on the US Collections List at iDigBio. Engaging the broader community is fundamental for data integrity; GrSciColl has implemented functionality for transparent requests for metadata changes (e.g. change in contact information) from GrSciColl users. These changes are evaluated by GrSciColl maintainers before publishing; approved changes are visible immediately.</p> <p>We hope that this unified global index of natural history collections will continue to enhance access to information about biodiversity collections and the people and data involved.</p>	<p>Caitlin Chapman, Ronald Canepa, Chris Wilson, Nicholas Rejack, Marie Grosjean, Morten Høfft, Marcos Gonzalez, Tim Roberston</p>	<p>3</p>

<p>GRSciColl: Registry of Scientific Collections</p>	<p>GRSciColl, the Registry of Scientific Collections, is a comprehensive, community-curated clearinghouse of collections information originally developed by the Consortium of the Barcode of Life (CBOL) and hosted by the Smithsonian Institution until 2019. It is now hosted and maintained in the Global Biodiversity Information Facility (GBIF) registry.</p> <p>Anyone can use GRSciColl to search for physical collections based on their attributes (country, preservation type, etc.) as well as their codes and identifiers. These users will find information on what the collections contain, where they are located, who manages them and how to get into contact.</p> <p>The current permission model aims to facilitate community curation. Anyone can suggest updates, and those changes can be applied or discarded by the appropriate reviewers: institution editors, country mediators, or administrators.</p> <p>More than half of the GRSciColl are updated during weekly synchronization with Index Herbariorum.</p> <p>More than 1,500 collection records were imported from the Integrated Digitized Biocollections (iDigBio). The iDigBio data managers are part of the GRSciColl editing team and review and maintain all U.S. collections in the GBIF registry. The collection information displayed on the iDigBio portal via the GRSciColl Application Programming Interface (API).</p> <p>The collection API also includes a lookup service to find GRSciColl records based on institution and collection codes and identifiers. This lookup service is used to link Specimen-related occurrences published on GBIF.org to GRSciColl entries whenever possible (see this example). This procedure allows aggregation of specimen-related occurrences under their GRSciColl-registered collections and institutions, regardless of the way they were published on GBIF. We hope to make GRSciColl a tool to help the community find each other and work together.</p>	<p>Marie Grosjean, Caitlin Chapman, Morten Høfft Gonzalez, Marcos Lopez, Tim Roberston</p>	<p>3</p>
<p>Hosted Portal Pilots Fuel Collaboration and Creativity Across North America</p>	<p>In late 2020, GBIF launched a pilot program to explore simple, customizable biodiversity data portals to lower the technical threshold for GBIF Participant nodes. This joint presentation by several North American Nodes will demonstrate the diversity of the portal concept across regional, national, institutional, and thematic areas. The portals represented in this presentation include GBIF North America, GBIF-US, BioMob, and VertNet. One exciting outcome of the pilot portal program has been the collaboration among nodes to develop a variety of customized portals. Benefits of the hosted portal solution include the use of a common platform, a shared data index maintained by GBIF, and a diversity of user interfaces tailored to specific purposes.</p>	<p>David Jennings, Carole Sinou, Abby Benson, David Bloom, Sharon Grant, James Macklin</p>	<p>4</p>

Legume Data Portal	<p>The Legume Phylogeny Working Group (LPWG) was established in 2010 to facilitate collaboration and to advance knowledge on the taxonomy, phylogenetics and systematics of the Leguminosae (Fabaceae). In 2020, four working groups were established: a taxonomy group focused on establishing a comprehensive species checklist with synonymy, plus groups on phylogenomics, traits and occurrence data.</p> <p>Communicating about and giving access to relevant and curated information about Legumes has been part of the objectives of the LPWG since the beginning.</p> <p>In the late 1980's, legume systematists developed the International Legume Data Information System (ILDIS). When launched ILDIS was pioneering in biodiversity informatics and rapidly became important and useful. The species list was later used in The Plant List and Catalogue of Life, but the ILDIS taxonomy has not been updated for 20 years and a new legume information system was long overdue.</p> <p>The need for a new portal about Legumes was first discussed in 2018, during a Legume conference in Japan. The hosted portal developed by GBIF is a great opportunity to start to build the portal the community envisioned during that meeting and is described in a paper in 2019 (https://doi.org/10.1071/SB19025).</p> <p>The Legume Data Portal gives access to the latest checklist built by the taxonomy working group, to all occurrence data available on GBIF, as well as information about the different working groups, the Bean Bag (LPWG newsletter) and news from the community. It is currently in a staging state (https://hp-legume.gbif-staging.org/) but will be officially launched this autumn. The LPWG plans to continue implementing new features relevant to the community even after the end of the hosted portal pilot phase.</p>	Carole Sinou, Anne Bruneau, Joe Miller	4
Piloting hosted portals in the GBIF network	<p>The GBIF community expressed interest in GBIF developing simple, customizable biodiversity data portals that participants can use to share news updates and create an identity for their work to help build a community of data publishers and users. To meet this need, GBIF developed a pilot program for fully hosted portals focusing on either a single museum, a country, a region or a thematic area, such as a taxonomic group. Each portal shows a specific view of a subset of GBIF-mediated data. For example, the US BISON pilot portal will serve all US data currently mediated through GBIF via our APIs.</p> <p>The portals do not have the advanced functionality of portals developed by large museums or the Living Atlases community, but the program seeks to fill an empty niche for simple, easy to use portals that showcase data shared with GBIF. The portals are hosted by GBIF but managed by the user group who add text, images and branding. The portals have an advanced occurrence data search, similar to that used on GBIF.org, and use the GBIF DOI download system. The first of 17 pilot portals are now coming online. GBIF seeks to learn from this pilot project to determine community interest and how to proceed past the pilot phase.</p>	Mélianie Raymond, Morten Høfft, Thomas Stjernegaard Jeppesen, Matthew Blissett, Tim Robertson, Joe Miller GBIF Secretariat	4

Digitization and Management of the National Museum of Natural History, Invertebrate Zoology Collections: Challenges, Workflows, and Solutions	<p>At approximately 50 million specimens, the Smithsonian Institution, National Museum of Natural History, Department of Invertebrate Zoology has one of the largest collections of invertebrates in the world. The department was founded in 1856 to house collections from the North Pacific Exploring Expedition. In a normal year, staff host over 200 visitors and process approximately 600 transactions a year; manage four US Federal Government affiliated agency partnerships; and concurrently operate numerous collection improvement and digitization projects. Recent large projects include inventory of the National Mollusk and Brachyura (crab) collections; curation and reorganization of the US National Parasite Collection; processing of large voucher and genomics collections; digitizing analog station data; uncataloged collections resolution (Backlog Inventory and Exemplar Data Capture); planning for the Dry Mollusk Reorganization and Move; and MSC glass slide reorganization and move. Key challenges include organizing and accounting for all of the resources (knowledge, staff, equipment, time, space, collection supplies, funding) required for the project. It is important to include all of the stakeholders (communication!), maintain flexibility (including remote digitization), and plan carefully and thoroughly. Effective solutions include organization and tailoring of workflows to specific projects with constant reevaluation and documentation to minimize cost and maximize efficiency.</p>	William Moser, Katie Ahlfeld, and Karen Reed Smithsonian Institution, National Museum of Natural History, Department of Invertebrate Zoology	5
Digitizing the U.S. National Insect Collection	<p>The Department of Entomology has taken strides to increase digitization of the U.S. National Insect Collection. Currently we maintain 3 separate inventories and utilize different digitization techniques to make our data available to the public and research community.</p>	Jessica Bird, Acting Collections Information Manager, Department of Entomology, National Museum of Natural History; Erin Kolski, Museum Specialist, Department of Entomology, National Museum of Natural History	5
Making Parasite-Host Associations Visible using Global Biotic Interactions (GloBI)	<p>The Terrestrial Parasite Tracker (TPT) Network digitizes and aggregates arthropod parasite collections to build an easily accessible, comprehensive database of parasite-host associations and vector distributions. TPT is working to digitally provide information on parasite collections by providing research-ready data and images from 1.2+ million parasite specimens, which will be accessible to scientists, educators, wildlife managers, and policy makers worldwide.</p> <p>Our network is providing needed baseline information for research and management of the ecological interactions among parasites, pathogens, and their hosts in North America (including the U.S. & territories) through Global Biotic Interactions (GloBI) data integration, and data review services. Global Biotic Interactions (GloBI) is an open data integration platform that continually indexes existing openly available species interaction datasets, literature, and specimen records using open source software. Since its inception in 2013, GloBI has grown to be the largest registry of biotic interaction claims, indexing over 8 million interaction claims sourced from hundreds of data sources and citing millions of references. For TPT, GloBI is used to keep track of availability of valuable evidence-based parasite-host interaction claims sourced from participating Natural History Collections as well as facilitating in-depth data reviews to help improve, and sometimes align, data exchange protocols (aka, integration profiles).</p> <p>TPT has contributed 500,000 biotic interaction records to GloBI, which will organize, standardize and integrate our records into existing species interaction datasets accessible via web tools and machines. Here, we provide examples of various parasite specimens with label data, demonstrate how biotic association data will be captured in different platforms, and provide a standard glossary of association terms. Over the course of the project, TPT and GloBI will continue to work with data providers to discover biotic interactions in their collections and refine standards specific to their data sharing workflows.</p>	Sullivan K, Poelen JH, Seltmann KC, Zaspel J	6





<p>Capturing California's flowers: Mobilizing phenological data from herbarium specimens</p>	<p>The timing of when plants flower is important to science, society, and biodiversity. Herbarium specimens can provide rich data on how flowering times vary across time and space and with changes in climate. The California Phenology Thematic Collections Network (CAP TCN) is an ADBC-NSF-funded project that aims to image over one million herbarium specimens and capture flowering (i.e., phenological) data from these specimens. To date, we have created over 1.6 million phenological scorings on the 3 million records housed in our data portal, CCH2 (CCH2.org). Our workflows and protocols are well-documented on our website (capturingcaliforniasflowers.org) and we have developed extensive educational and outreach resources—including lab materials, a course-based undergraduate research experience, and a workshop—to promote research using the data. Here, we describe how we capture phenological data using two tools we developed in our data portal to harvest phenological data from specimens, and we show how these data are mobilized using Darwin Core MeasurementOrFact files. Scoring phenology using these tools was fast and easy and could be implemented by other portals. In the next phase of our project, we are developing more robust data standards for efficient data mobilization through a TDWG Task Force, in which we invite all interested stakeholders to participate.</p>	<p>Katie Pearson, Jenn Yost</p>	<p>6</p>
<p>Accelerating Global Engagement Through Collaboration of Online Biodiversity Data Networks</p>	<p>Online networks are helping to make global biodiversity data readily available to researchers, educators and policy makers. Networks such as iDigBio (Integrated Digitized Biocollections; idigbio.org), GBIF (Global Biodiversity Information Facility; gbif.org), ALA (Atlas of Living Australia; ala.org.au), and DiSSCo (Distributed System of Scientific Collections; discco.eu) have fostered innumerable analyses, projects, and publications with the hundreds of millions of biodiversity occurrence records, e.g., natural history specimen records and observational datasets, available via their searchable databases. Currently, each network maintains independent databases, that best serve their respective data users. Collaboration among networks, i.e., into a "Network of Networks", could facilitate an accelerated pace of data discovery and biodiversity research. Working more closely together would allow for: 1) comprehensive access to all biodiversity data from a single resource, 2) consistent and predictable data interpretation and transformation, 3) stable and persistent references to all data records, 4) unification of perspectives among research communities, and 5) sustainable development and maintenance of tools and services. Progress in all five of these areas maximizes the benefits expected from all biodiversity monitoring, digitization, data mobilization, research, and curation efforts, and supports multitaxa studies that span domains. Here we show some of our collective efforts to streamline access to biodiversity occurrence records in support of research and actions that address today's biggest threats to biodiversity as well as the importance of data availability for ensuring food security, combatting vector-borne disease, encouraging bio-inspired design, and harnessing the power of artificial intelligence and machine learning in systematics and ecology.</p>	<p>Elizabeth Ellwood, Gil Nelson, Hamish Holewa, Dimitris Koureas, Joe Miller</p>	<p>7</p>

<p>Digitization of specimen records in Canada: how far are we?</p>	<p>Digitization of specimen records hosted in Canadian collections started in 2001, with the deployment of the Canadian Biodiversity Information Facility (CBIF), which led to a first wave of publication of biodiversity data in Canada. With funding from the Canadian Foundation for Innovation, several projects have followed the steps of CBIF, allowing for a better access and understanding of biodiversity in Canada: the Beatty Museum at the University of British Columbia, the Barcode of Life initiative at University of Guelph, and Canadensys led by the Université de Montréal.</p> <p>Canadensys was officially launched in 2008, and was both a web infrastructure and a network of people in 10 Canadian universities representing about 30 collections, equipped with all the necessary tools to start digitizing specimens (cameras, scanner, informatics equipment, databases, minimal funding for salaries).</p> <p>Canadensys has now facilitated the publication of nearly 2.7M specimen records from 35 institutions, and has played an important role in training biodiversity informatics in order to build a network of skilled researchers, professionals and students in the country. For herbaria, for example, a recent overview of the state of digitization revealed that around 33% of specimens in Canadian collections have been digitized to date. In order to accelerate the rate of digitization in Canada, mass digitization projects using automated workflows would be necessary, but none is currently in place. Digitization in herbaria, and in most natural history collections in the country, still relies on manual digitization and volunteers, leading to highly accurate and complete records, but a slow growth rate. Federal collections are currently in a blitz of digitization, supported by funding of the BioMob project, and will be published in a near future. Similar efforts are required for mobilizing data from Canadian university collections to ensure rapid and accurate digitization efforts.</p>	<p>Carole Sinou, Anne Bruneau</p>	<p>7</p>
<p>Beyond biodiversity: Digitizing rocks, minerals, and meteorites at NMNH</p>	<p>In addition to bio- and paleontological specimens, many natural history collections also include rocks, minerals, and meteorites used to study the history of Earth and the Solar System. This poster will explore digitization of such specimens in the Department of Mineral Sciences at NMNH.</p>	<p>Adam Mansur, National Museum of Natural History</p>	<p>8</p>
<p>NMNH Paleobiology Department Fossil Data: Volunteering for Better Collection</p>	<p>As 60-70% of the workforce, the volunteers at the Smithsonian National Museum of Natural History Department of Paleobiology have an enormous impact. The department needed to find solutions to expand our capacity to create accessible, high quality digital collections records to support research. We determined that integration of a volunteer program for collections digitization was necessary. Volunteers create and enhance digital collections records through transcriptions, imaging, data collection, and identifications. Specimen information and digital assets are generated by individual volunteers under the supervision of their volunteer manager. Raw data is then cleaned, analysed and ingested into systems of record by a mix of museum staff and contractors. This system has greatly increased the digitization capacity for the Department of Paleobiology.</p>	<p>Miller, Matthew T.; Little, Holly; Millhouse, Amanda; Telfer, Abby</p>	<p>8</p>
<p>A Decade at a Glance: The Role and Impact of the Collections Program Technicians on NMNH Collections in an Ever-Increasing Digital World</p>	<p>In support of the Smithsonian Institution's mission for the increase and diffusion of knowledge, the National Museum of Natural History (NMNH) established a specialized group of museum technicians under the management of the Associate Director of Collections in 2012. The Collections Program technicians (CPTs) provide core services to all collecting departments at the museum. Some of these services include public outreach, physical collections care and management, and digitization. Over the years, the NMNH has made concerted efforts and investments to support and accelerate digital advances in museum collections and its metadata. As a result, the CPTs have made significant contributions and improvements in efficiency and output of various digitization projects. This poster not only highlights these achievements but also compares them to the overall goals and accomplishments of the NMNH over the past decade. Furthermore, with the impact and growing demand for more accessible collections in a digital world, the NMNH and CPT team have in turn evolved. This poster features the shift in priorities and types of CPT projects over the years and how improved technology, communication, and resources have benefited the museum, despite recent events affecting our nation and the world.</p>	<p>Teresa Hsu, Greg Polley, Alice Fornari, Katie Roberts</p>	<p>9</p>

<p>“Deep Time” of Digitization: History of Biodiversity Digitization at National Museum of Natural History</p>	<p>This poster provides an overview of digitization efforts at the Smithsonian’s National Museum of Natural History (NMNH) from inception to the present. Join us as we look at the evolution of the museum’s tools and methods, as well as its future goals and aspirations.</p>	<p>Katie Roberts, NMNH Collections Program Kasia Ahern, NMNH Collections Program Jessica Bird, NMNH Entomology David Bridge, Smithsonian Libraries & Archives volunteer Cailin Meyer, NMMH Collections Program Chris Milensky, NMNH Vertebrate Zoology Ducky Nguyen, NMNH Informatics Sylvia Orii, NMNH Botany</p>	<p>9</p>
<p>Exploring multiple approaches to engage underrepresented students in research with natural history collections.</p>	<p>Natural history collections contain records of life on earth including unique and rare specimens of extinct species and temporal information on changes in the distributions of native and introduced species. Associated with these collections are personnel that manage and care for these resources and their associated data, and these people serve as both resources and mentors. The collections (objects and data) and the collection personnel are an excellent conduit for introducing and preparing undergraduate students from underrepresented backgrounds for a myriad of careers in STEAM while equipping them with 21st century skills to successfully matriculate as the next STEAM workforce. Programs designed to serve underrepresented students exist on a continuum of how many people you can reach versus how deeply you can support everyone. Increasing reach is often to the detriment of impact and vice versa. Building from the undergraduate Introduction to Natural History Collections course model created by Flemming et al., the iDigBio team has leveraged natural history collections in a 3-pronged approach designed to engage undergraduates from underrepresented groups in collections-based research at multiple levels and create synergy within our institution around the common goal of broadening representation in the biological sciences. The three programs include a two-day biological career conference and fair created in collaboration with the University of Florida TRIO programs (large reach, shallow impact), a semester-long introduction to natural history collections course (intermediate reach, higher impact), and a paid, collections-based research summer internship program (limited reach, highest impact). Each program has been created to function independently, but also work synergistically for students that participate in multiple activities. During this session we will elaborate on the benefits and drawbacks for each program and explain how these programs can be adopted at other institutions to leverage natural history collections to help provide undergraduate students with necessary research experience.</p>	<p>David Blackburn, Alnycea Blackwell, Adania Flemming, Jeanette Pirlo, and Molly Phillips (Florida Museum, University of Florida)</p>	<p>10</p>
<p>BCEENET: Connecting digitized natural history collections with undergraduate research experiences</p>	<p>Biological Collections in Ecology and Evolution Network (BCEENET) brings together undergraduate educators, natural history collections professionals, researchers, and data experts to support the development and implementation of Course-based Undergraduate Research Experiences (CUREs) using digitized natural history collections. CUREs increase the number of undergraduate research opportunities, expand access to more students, and provide the same benefits to students as traditional research experiences, with many of the greatest gains in minority and first-generation college student populations. Collaborating with BCEENET can connect you and your collections to a vibrant research and education community, and increase use of your digital data through undergraduate research.</p>	<p>Janice L. Krumm, Carly N. Jordan, Cecily D. Bronson, Elizabeth K. Shea, Jean L. Woods Widener University, Chester, PA, The George Washington University, Washington DC, Portland State University, Portland, OR, Delaware Museum of Natural History, Wilmington, DE</p>	<p>10</p>

<p>Smithsonian's 3D Digitization Pipeline: Management, Automation, and Delivery</p>	<p>The Smithsonian has the ambitious goal of high-throughput 3D digitization of collection objects. The challenge of realizing this goal goes well beyond just the act of scanning itself. The meat of the problem rests in creating the enterprise systems that support the management, processing, and publication of 3D data created during scanning activities. The Smithsonian's Digitization Program Office, part of the Office of the Chief Information Officer, has been working to tackle this sprawling, complex infrastructure requirement, developing a suite of open source tools that can be deployed by any organization to support their 3D digitization activities. This presentation will cover the fruit of those efforts to date which includes Voyager, our content authoring and 3D scene publication platform; Cook, our automated framework for data processing; and Packrat, our system for managing 3D data, workflows, and 3D asset publication.</p>	<p>Jon Blundell</p>	<p>11</p>
<p>Symbiota: Managing and Mobilizing Biodiversity Data and Supporting Data Providers</p>	<p>The history of Symbiota spans nearly two decades of development and bio-collaboration. Initially designed as a simple online search engine featuring a handful of Arizona-based herbaria, Symbiota has since matured into a distributed network of theme-based research portals incorporating data from over 1,800 collections. More than 40 portals publish over 9 million images and 70 million occurrences, 12 million of which are published to the global data aggregator GBIF. Symbiota is the primary collection management system for over 700 collections, and Symbiota portals are actively used by over 3,000 registered users and countless more researchers, educators, and other members of the public. Through many collaborative efforts, the software developed into both a robust content management system (CMS) and a tool for biodiversity data exploration. As a CMS, Symbiota is specifically designed toward efficient, collaborative digitization with features including data entry from label images, data harvesting from specimen duplicates, batch georeferencing (even across collections), data validation and cleaning, generating progress reports, and additional tools. As a data exploration tool, Symbiota includes species inventories, interactive identification keys, integrated specimen and field images, taxonomic information, species distribution maps, and taxonomic descriptions.</p> <p>Symbiota has achieved such success because it fills a particular niche in the global biodiversity data network; Symbiota portals serve as low- to mid-level aggregators that can help cultivate communities of practice—often focused taxonomically, geographically, or both—that jointly improve data quality and promote data use. These communities also benefit from collaborative digitization and active data management enabled by Symbiota tools. Symbiota is open source, accessible, and is committed to maximum data interoperability by aligning with Darwin Core standards and enabling efficient data import and export.</p> <p>iDigBio 3, funded for the period of 2021-2026, newly includes the Symbiota Support Hub (SSH) as an integrated domain and service team to strengthen Symbiota portal user, manager, and software developer communities. Our mission includes help desk support for portal needs, scalable development, user training, and promotion of greater data sharing, such as publishing to global aggregators. The SSH launch will focus on developing and promoting community-supported, structured, and interactive documentation and training for Symbiota users. For more information or if you have questions, email us at symbiota@asu.edu.</p>	<p>Edward Gilbert¹, Nico Franz¹, Jenn Yost², Katie Pearson^{1,2}, Laura Rocha Prado¹, Samanta Orellana¹ ¹ Arizona State University, School of Life Sciences, Tempe, Arizona ² Cal Poly – San Luis Obispo</p>	<p>11</p>

NMNH Virtual Behind the Scenes Collection Tours

	<p><i>NMNH Botany Conveyor Tour</i> <i>Sylvia Orli, Eric Schuettpelez, Victor Shields</i> September 22 @ 10:00 am EDT</p>
	<p><i>Digitizing Molluscs Through Space and Time</i> <i>Kathy Hollis, Holly Little, John Pfeiffer, Ellen Strong, Bill Moser</i> September 22 @ 7:00 pm EDT</p>
	<p><i>Digitization and Replication of Cultural Objects</i> <i>Eric Hollinger, Vince Rossi</i> September 23 @ 9:50 am EDT</p>
	<p><i>Entomology</i> <i>Floyd Shockley, Torsten Dikow</i> September 23 @ 2:00 pm EDT</p>

Tour leaders will be available for Q&A after the live viewing.

If you cannot join live, recordings can be accessed on the [conference wiki page](#) or in the [Kumospace Viewing Room](#).

