Symposia Abstracts

S02: Biodiversity Data Quality – concepts, methods and tools

Organizers: *Arthur D. Chapman, Antonio M. Saraiva, Alexander M. Thompson*

Abstract

The goal of this 4th Symposium is to discuss concepts, problems, policies, metadata, methodologies and mechanisms related to Biodiversity Data Quality, which can be reused by the Biodiversity Informatics community collaboratively and incrementally. Data quality (DQ) is a major concern in Biodiversity Informatics. The distributed nature of data acquisition and digitization, the specific difficulties imposed by some of the data sub-domains, such as taxonomic data and geographic data, among other aspects, make it important to discuss DQ in biodiversity so that data made available in portals and other systems can be used for various purposes such as education, science, and decision-making. Although several initiatives in the Biodiversity Informatics community have been developing tools and best practices about DQ, there is no consensus related to concepts, metadata, policies, methodologies and tools about DQ. The size of DQ check pipelines has also posed challenges for existing methodologies and tools and may need to drive some of the discussion on concepts and policies. Previous symposia on DQ were held at the TDWG meetings in Florence, Italy/2013, Jönköping, Sweden/2014 and Santa Clara de San Carlos, Costa Rica/2016. As a result, a joint TDGW/GBIF Interest Group on DQ was proposed and approved in 2014. Subsequently three task groups were also created: TG1 – BDQ Framework, TG2 – BDQ Tools, services and workflows, and TG3 – BDQ Use cases. They tackle some of the most important issues identified by the attendants to the previous symposia.

The group has been able to meet between TDWG meetings to advance on its activities. In March, 2016, in São Paulo, Brazil, the IG and TGs conveners got together with members from two fitness-for-use groups supported by GBIF on species distribution modeling and agrobiodiversity, and have made good progress towards adopting a common conceptual basis and developing profiles for those two use cases. In October, 2016, in Melbourne, Australia, further advance was made focusing on another fitness-for-use group on alien and invasive species. Another meeting is scheduled to take place in Canberra, Australia, in May, 2017.

By the 2017 TDWG Meeting, in Ottawa, significant advances will have been made in all three Task Groups as well as in Vocabularies. In this 4rd Symposium we want to advance further on those discussions and increase participation of other stakeholders, keeping the same principles: discuss and share experiences on the ways we deal with DQ in the Biodiversity Informatics community, avoiding duplication of efforts and sharing knowledge.

S04: Advances in data accessibility and data management for marine species occurrence data

Organizers: *Andrew Sherin, Ward Appeltans, Lenore Bajona, Mary Kennedy*

Abstract

The symposium will include presentations on the practices of data accessibility and data management for marine biological occurrence data by members of the Canadian Ocean Data Management Community of Practice and others on the following topics:

* Making data accessible for reuse and data rescue initiatives
* Biodiversity plus environmental measurements: OBIS-ENV and beyond
* Encouraging a ‘Community of practice’ for marine biological occurrence data
* Development of vocabularies and tools for quality control for DwC terms
* Identification of the need for new DwC terms
* Biodiversity use cases demonstrating data integration with habitat (e.g. Coastal Web Atlases)
* Data management training curriculums
* Citizen science and volunteer coordination

The symposium will be structured into several mini-symposia of selected topics based upon the topics of received presentation proposals. Not all of the topics listed above may be addressed within the symposium.

Each mini-symposia will conclude with a panel discussion to identify next steps for collaboration and / or recommendations initially to the Canadian Ocean Data Management Community of Practice.

S10: 500 Years of Big Data from the Biodiversity Heritage Library

Organizers: *Martin R. Kalfatovic, Carolyn A. Sheffield*

Abstract

With 89.9 terabytes of data spanning over 500 years of data collection, the Biodiversity Heritage Library (BHL) is an important galaxy in the universe of biodiversity data. Embedded in those 89.9 terabytes are over 174 million species name occurrences and a currently unknown number of species occurrences, descriptions, identified traits, and related data--all locked in the over 500 years of data captured in the BHL. As a consumer of name occurrence data, BHL works closely with the GNA and data providers such as ITIS. Recent developments and activities of ITIS and how these enhance BHL will be discussed.

A current IMLS funded grant, 'Foundations to Actions,' has funded five post-grad residents to gather data for requirements of BHL Version 2.

An international consortium, the BHL partners work to make biodiversity literature openly available to the world as part of a global biodiversity community. Through its extensive network of Members and Affiliates, over 52 million pages of biodiversity literature are now available through the BHL portal. Providing deeper access to the data contained in this literature is a goal of an ongoing project to provide full-text search capability to BHL. Technical details and use-case examples for this will be discussed.

BHL is not only data, but generates new data via wide-ranging and successful outreach activities. From tagging of images on Flickr, transcription of archival materials, and OCR correction from gaming activities, BHL has been able to leverage crowd-sourced input from users to better understand our collections. Outreach strategies designed to encourage citizen science activities will be discussed, as well as the importance of community engagement to sustain crowd-sourcing initiatives. Next steps and challenges related to incorporating crowd-sourced data into the BHL collection will also be addressed.

The Biodiversity Heritage Library has grown to be an important part of the infrastructure of biodiversity. In an attempt to solve the literature component of the taxonomic impediment, the BHL continues to provide access to legacy print publications and make this data widely available for reuse in collections support systems. This symposium will include a brief update on BHL activities since that 2016 TDWG. The final section of this symposium will be a guided discussion of desideratum and enhancements TDWG participants see as important for BHL.

Symposium participants will include representatives from the BHL Secretariat, BHL-related projects, ITIS, and other BHL partners.

S12: Citizen Science Contributions to Biodiversity Research and Standards

Organizer: *Robert D. Stevenson*

Abstract

Citizen science encompasses a diverse set of activities that contribute to biodiversity research including decoding and transcribing museum labels, collecting specimens and images, analyzing and classifying digital collections, setting out camera traps and doing field surveys. This symposium offers a venue for citizen science practitioners to describe their biodiversity projects and to learn about the TDWG community. Likewise, informatics specialists and database managers working on biodiversity projects are encouraged to present the issues they face when supporting citizen science programs. Aspects of a citizen science programs from project design, to data collection and data curation are welcome. Projects that address taxonomic or data quality issues are especially encouraged. Participants are invited to join in the work of the Citizen Science Interest Group in developing standards to support biodiversity research.

S15: Agricultural Informatics Contributions to Biodiversity Science and Biodiversity Assessments

Organizers: *Gail Kampmeier, Cyndy Parr, James Macklin*

Abstract

Agricultural biodiversity has long been ignored by the traditional biodiversity community and the aggregators of their data. The February 2016 GBIF 'Final Report of the Task Group on GBIF Data Fitness for Use in Agrobiodiversity,' provided recommendations primarily regarding crops and their wild relatives, but did not address wider issues of crop pests (plant diseases and their vectors, arthropods) and management systems that affect the greater biodiversity of those crops. The TDWG16 symposium, 'Agricultural Biodiversity Standards & Semantics,' highlighted the current status of agrobiodiversity data management and discussed major challenges in this field. This year's symposium will provide a progress report on addressing challenges such as crop management and experimental protocol standards and infraspecific taxonomic coverage. It will dive deeper into trends in semantics and data mining for agriculture, and in application of standards to biodiversity assessments. We aim to provide specific examples where shared data management standards and practices across both basic and applied biodiversity research communities can lead to improved outcomes for both science and society.

S16: Using Big Data Techniques to Cross Dataset Boundaries - Integration and Analysis of Multiple Datasets

Organizers: *Matthew Collins, Robert Guralnick, Martin R. Kalfatovic*

Abstract

For many historical reasons, much biodiversity data is cataloged and aggregated by type: occurrence, gene sequence, literature, etc. While this has advantages, interesting and new scientific questions require information from more than one source but the varied structure and axes of existing datasets don't make it easy to simply join them up.

In this symposium we will have presentations from researchers on techniques used to join, extract, and infer information across traditional occurrence, taxonomy, trait, phenology, genetic, environment, literature, image, climate, and other boundaries. We also will have limited presentations from data providers on the work they are doing and challenges they face in building and exposing linkages among their datasets. Our goal is to provide a forum to expose working techniques for addressing the disjoint nature of our current data ecosystem.

S18: Bridging Gaps between Biodiversity Informaticians and Collections Professionals

Organizers: *Holly Little, Deborah Paul, Gail Kampmeier*

Abstract

The biodiversity informatics standards, tools, and resources developed by TDWG and its members are vital components in the daily work of curators and collections professionals in natural history museum collections. However, opportunities for communication and collaboration between these groups are too often rare and siloed. Therefore, building a mutual understanding of and solid collaboration about how biodiversity informatics concepts are developed and then how they are applied to the mobilization of collections data can be difficult. In 2018, the TDWG meeting will be held in conjunction with the Society for the Preservation of Natural History Collection (SPNHC) meeting, providing an invaluable opportunity to address these issues. In order to enhance the effectiveness of the 2018 meeting it is important for each group to develop a better comprehension of the vernacular of the other. This symposium aims to bridge gaps in understanding between the biodiversity informaticians and the ways collections professionals use their work. It will provide examples of how collections are using TDWG standards, applicability statements, and best practices, as well as exploring where this community, along with discipline specific experts, can work to provide more controlled vocabularies for existing terms, participate in development activities of interest and task groups, and identify and contribute solutions to needs as yet unaddressed by current standards.

S20: Traits - Data models, sources, vocabulary, interoperability and their presentation and discoverability via species information pages

Organizers: *Jeff Gerbracht, Robert Guralnick, Jennifer Hammock, Deborah Paul, Pamela Soltis*

Abstract

With potentially 3-4 billion specimens globally, the world's natural history collections offer treasure troves of relevant information. This is complemented by a long historic record of citizen science observations, and recently by molecular sampling. New contributions in both these areas are born digital and are accelerating the growth of contemporary biodiversity occurrence data. Meanwhile, efforts such as the NSF's Advancing Digitization of Biodiversity Collections Program are increasing the rate of digitization of specimen records, filling in our historic biodiversity record just as rapidly. Occurrence records typically contain locality information, and these georeferenced occurrences have revolutionized our ability to model species distributions and project future distributions based on models of climate change. These records may also contain information, such as body size, phenology, environment, that is relevant to ecological and evolutionary research. Even literature sources, where traits measured are not associated with deposited specimens, contain detailed metadata describing location, environment, lifestage and concurrently measured traits. Textual descriptions, from regional floras, catalogs, monographs and other literature, also contain extensive ecological information, suitable for extraction using semantic analysis. This deluge of data is not without its challenges. Among these is the need to develop ontologies, which are crucial for big data solutions, so that trait data can be extracted and assembled informatically. This requires the development of ontologies, which are crucial for big data solutions.

There is an array of existing ontologies in the biodiversity space, many maintained by active communities like the OBO Foundry. Gaps remain in even the most mature ontologies, but as term adoption by outside projects increases, and new tools developed for terms nomination and other community curation activity, it becomes more practical to develop robust structured vocabulary coverage for the biodiversity space.

There is increasing convergence around traits of urgent concern for conservation and applied research. The Essential Biodiversity Variables, identified by GEO BON as essential for monitoring biodiversity change, highlight key traits needed for this stewardship.

In this symposium, we will explore the breadth of sources of biodiversity trait data, and techniques and consensus needed to support analysis over these diverse sources. With over 780 million occurrence records now available via iDigBio and GBIF, and many millions more appearing every year, now is the perfect time to consider which traits are most significant for ecological and evolutionary research, the ontologies required for meaningful communication and data assembly, and the tools needed for ensuring interoperability of these data.

S22: Biological Interaction Data - towards data standardization

Organizers: *Antonio M. Saraiva, Jennifer Hammock, Quentin Groom*

Abstract

Scientists use a variety of methods to collect, record, and store biological interaction data (predator-prey, parasite-host, pollinator-plant, etc.). Uses for these data are equally diverse. For example, they could play an important role in building decision support systems for conservation and sustainable use in agriculture. Numerous efforts are underway to aggregate, organize, and efficiently disseminate these data. However, we lack a formal data standard to support this work. The goal of this symposium is to provide an opportunity, for those involved or interested in the digitization of biological interaction data, to share their experiences and ideas so that we can move forward and propose a biological interaction data standard. During the 2016 TDWG Conference, in Santa Clara de San Carlos, Costa Rica, a group of people gathered to start discussing this issue under the TDWG umbrella. The objective is to propose the creation of an Interest Group on Biological Interactions Data where this topic can be discussed within the biodiversity informatics community. During the symposium, we will provide an update on the group's achievements as well as welcome other interested parties to present their work.

S28: Financial Models for Sustaining Biodiversity Informatics Products

Organizer: *James H. Beach*

Workshop Abstracts

W01: Establishing a global registration system of algal names and types

Organizers: *Andreas Kohlbecker*

Abstract

Establishing an on-line registration system for algal names and their nomenclatural types is one of the main objectives of a DFG (Deutsche Forschungs Gemeinschaft) project, which started in 2016. This registration system will become part of the international network of biodiversity data and one of the operating global registration systems for names of organisms. Once established, PhycoBank will remove a major obstacle to research in the field of phycology. The envisaged workflows include manual registration of new algae names, as well as the automatic registration of new algal names published in scientific journals that already have a functioning electronic publishing workflow. An important component of the registration system will be a comprehensive index of algae names. Its main purposes are to cover all algae names known to sources external to PhycoBank, and to support the registration workflow by providing these names in a harmonized and consistent way. The index could be furthermore a valuable source of names for other data systems. It may contribute to the Global Names Architecture (GNA) in the medium term. In the short term, it aims to provide names to the backbone of biodiversity networks such as GBIF and the Biological Collection Access Service (BioCASe). The workshop will review and discuss interoperability aspects of the system with the global network of biodiversity information systems, including topics like persistent identifiers for names and typifications, data flows to name catalogs, and other services, as well as the harvesting and indexing of names and taxa for the algae name index. Another major focus of the workshop will be the integration of name registration systems in the workflow of digital publishers.

W03: Towards robust interoperability in multi-omic approaches to biodiversity monitoring

Organizers: *Robert Hanner, Pier Luigi Buttigieg, Luke Thompson*

Abstract

The routine use of (meta)genomics, (meta)transcriptomics, and the targeted sequencing of taxonomic marker genes is enabling monitoring of both micro and macro-organismal assemblages across a wide range of environments. Such monitoring of phylogenetic and functional diversity offers novel and valuable insight into how the biosphere is responding to environmental change. However, there are considerable obstacles to unifying multi-omic observation and monitoring strategies, which span environments from the deep marine subsurface to the urban atmosphere. Perhaps foremost, rapid development within the field often outpaces attempts to harmonise its techniques, despite progress by initiatives such as the Earth Microbiome Project, TARA Oceans, and Ocean Sampling Day. Overcoming these obstacles will be essential in realising an informatics infrastructure suited to a global, omically-enabled biodiversity monitoring system and incorporating omics-driven insights into the biodiversity metrics which inform the conservation and policy domain.

This workshop will begin with a collection of talks that will outline the scope of the challenge and possible solutions, followed by facilitated discussion towards delivering a draft data standard for reporting species occurrences as detected through varied environmental DNA (eDNA) methods and community meta-omics approaches. This issue is nontrivial and involves the intersection of geospatial, genomic and biodiversity informatics perspectives, among others. We will examine efforts such as long-term, omically enabled observatories as use cases in need of "future proof", flexible standards and propose strategies for existing standards to contend with their demands. We also welcome contributions identifying aspects of omic data which have yet to be acknowledged in the standards landscape. Our activities aim to catalyse the robust interoperability and convergence of community-endorsed/best practices, metadata checklists, standard terminologies, knowledge representations, and persistent identifiers for the environments, samples, processes, information artifacts, equipment, and agents involved in this emerging, multidisciplinary domain.

We welcome submissions on topics directly related to the above, including:

Community-endorsed, best practices for 'omic observatories and/or observation campaigns, such as:

Design and/or governance - emphasising interoperability

Standardised sampling, sample management, and archiving, including the use of persistent sample identifiers and FIMS/LIMS technologies
Standardised generation, processing, analysis, and archiving of ‘omics data
Acquisition, standardisation/structuring, processing, management, and exchange of metadata (e.g. use of checklists, structured formats, controlled vocabularies, ontologies, and FIMS/LIMS technologies)
Efforts to translate omics outputs to diverse stakeholders, such as those in the policy, conservation, commercial, and public spheres

W06: Data Quality Workshop

Organizers: *Arthur D. Chapman, Antonio M. Saraiva*

Abstract

The goal of the workshop is to present the advances and the status of the three Task Groups of this [Interest Group] TG1 – BDQ Framework, TG2 – BDQ Tools, services and workflows, and TG3 – BDQ Use cases. We will also work on specific issues raised during the year, which need to be addressed by the group, and plan next steps, including how to increase participation of other stakeholders and plans toward a TDWG standard on data quality tests and assertions.

The distributed nature of data acquisition and digitization, the specific difficulties imposed by some of the data sub-domains, such as taxonomic data and geographic data, among other aspects, make it important to discuss DQ in biodiversity so that data made available in portals and other systems can be used for various purposes such as education, science, and decision-making. Although several initiatives in the Biodiversity Informatics community have been developing tools and best practices about DQ, there is no consensus related to concepts, metadata, policies, methodologies and tools about DQ. The size of DQ check pipelines has also posed challenges for existing methodologies and tools and may need to drive some of the discussion on concepts and policies.

A joint TDGW/GBIF Interest Group on DQ was proposed and approved in 2014. Subsequently three task groups were also created: TG1 – BDQ Framework, TG2 – BDQ Tools, services and workflows, and TG3 – BDQ Use cases. They tackle some of the most important issues identified by the attendants to the previous symposia, held at the TDWG meetings in Florence, Italy/2013, Jönköping, Sweden/2014 and Santa Clara de San Carlos, Costa Rica/2016.

The group has been able to meet between TDWG meetings to advance on its activities. In March, 2016, in São Paulo, Brazil, the IG and TGs conveners got together with members from two fitness-for-use groups supported by GBIF on species distribution modeling and agrobiodiversity, and have made good progress towards adopting a common conceptual basis and developing profiles for those two use cases. In October, 2016, in Melbourne, Australia, further advance was made focusing on another fitness-for-use group on alien and invasive species. Another meeting is scheduled to take place in Canberra, Australia, in May, 2017.

By the 2017 TDWG Meeting, in Ottawa, significant advances will have been made in all three Task Groups as well as in Vocabularies.

W07: Reusing an open source platform in order to create a community: example of the Living Atlases community

Organizers: *Marie-Elise Lecoq, David Martin, Christian Gendreau, Federico Mendez, Jeremy Goimard, Santiago Martínez de la Riva, Anders Telenius, Manash Shah*

Abstract

Since 2013, a community has grown around the Atlas of Living Australia (ALA) open source platform, mostly but not exclusively around GBIF nodes. Indeed, since 2014 five international workshops have been organized around the world and eight data portals have been released into production using this tool (ALA, Brazil, GBIF Argentina, GBIF Costa Rica, GBIF France, NBN, GBIF Spain, GBIF Portugal) and several others are currently in development (Canadensys, GBIF Sweden, etc.). Here you can find a map representing all countries with an interest in implementing ALA as their national data portal.

ALA modules work with standards defined by the TDWG community. Data shown on maps are loaded through Darwin Core Archive files previously published from IPT or directly downloaded on the GBIF.org. Users can filter and download occurrences and can also interact with data using API proposed by the

platform. We keep the data authority and highlight data publishers, institutions, collections and datasets by showing their metadata. Moreover, using this ALA technology as national data portals is the solution recommended in the GBIF Implementation Plan 2017 - 2021.

One of the main objectives of this workshop will be to present the community of Living Atlases by showing examples already in production (Atlas of Living Australia, GBIF France, GBIF Spain and NBN Atlas Scotland), and past & future projects involving the community. In addition, we expect to train participants in the basic ALA modules (Collectory and Biocache-hub) that focus on occurrence research tools, data visualization and metadata portal. This technical training will also include the installation of an ALA demo version in order to give the possibility to users to make their first configurations and developments of their new tool.

W08: Biodiversity data annotations – state of the play and perspectives

Organizers: *Lutz Suhrbier, Okka Tschöpe, Anton Güntsch, Walter Berendsohn*

Abstract

With the growing number of digitised specimen metadata and images on the world wide web, on-line annotation systems for correcting or enriching label information become increasingly important. The annotation system AnnoSys is one of the systems that provide functional infrastructure for annotating collection information as well as web service interfaces for searching existing annotations and their integration into research workflows. The recent stable release of AnnoSys has been integrated with the GBIF data portal, the World Flora Online Specimen Explorer, the Global Genome Biodiversity Network (GGBN) and others and stores all annotations conforming with the W3C Open Annotation Data Model in a centralised RDF repository together with the original specimen data.

The workshop focusses on practical demonstration of available annotation systems and their integration into the international biodiversity informatics landscape. We will discuss potential new features and future developments with a focus on semantic interoperability and synchronisation or aggregation mechanisms for distributed annotation repositories.

W09: Implementing Biodiversity Standards for Paleobiology

Organizers: *Denné Reed, Falko Glöckler, Mareike Petersen, Jana Hoffmann*

Abstract

The Paleobiology Interest Group (Paleo) was established in 2015 to broaden the application of existing standards such as Darwin Core and the Access to Biodiversity Collections Data Extended for Geosciences (ABCDEFG) to accommodate paleobiological data. This will represent the second meeting of the interest group at a TDWG conference.

The group seeks to extend existing standards to meet the needs of paleobiology and to foster greater integration between neontology and paleontology in the study of biodiversity across space and deep time. Understanding long-term temporal patterns in biodiversity provides the context for interpreting modern changes in biodiversity and understanding the process responsible for these changes.

Employing biodiversity information standards, such as Darwin Core and ABCDEFG to paleobiology data requires addressing a broader range of metadata requirements and adapting best practices. The role of the Paleo group is to provide guidelines for deploying existing biodiversity standards to paleobiology and to propose extensions and modifications to existing standards to make them more amenable (and generalizable) for paleobiology.

The primary goals for the 2017 meeting of the Paleobiology Interest Group are: 1) to update TDWG members on the group's history and progress, 2) to conduct in-depth discussions and debate on common use cases for deploying standards in paleobiology, 3) to discuss and produce examples of Darwin Core and ABCDEFG entries for those use cases.

The annual meetings offer the best opportunity for bringing biodiversity information specialists together with disciplinary specialists in paleobiology to resolve questions and make concrete advances in broadening the application of biodiversity information standards in a new disciplinary domain. The 2017 meeting of the TDWG Paleo Interest Group will include a focus on the relationship between Darwin Core and ABCDEFG and how they can work together.

W14: Standards for Citizen Science Biodiversity Studies

Organizers: *Robert D. Stevenson, Libby Ellwood*

Abstract

The TDWG Citizen Science Interest Group was chartered at the beginning of 2017. This workshop represents the first time the group can meet together and discuss community development and standards needs. Topics will include policies and standards for common name, symbols for non literate observers and data quality. An important aspect of the meeting will be to establish links to other TDWG working groups and to learn about efforts by groups outside the TDWG organization such as the Citizen Science Association Meta Data working group, the European Citizen Science Association, CoData, RDA Small Unmanned Aircraft Systems, and the Camera Trap Federated Minimum Data Standard about their efforts to draft standards. The PPSR_CORE meta standard has been defined. We will review this standard and its implementation.

W17: The EDIT Platform for Cybertaxonomy, Current State and Envisaged Integration into the Biodiversity Informatics Infrastructure

Organizers: *Andreas Müller, Walter Berendsohn, Anton Güntsch*

Abstract

The EDIT Platform for Cybertaxonomy has been developed over the past years to support the full taxonomic workflow and to serve as an information broker for all types of biodiversity data. Though a focus has been set on taxon level data it covers all relevant data types from literature to DNA alignment . In the early phase focusing on core functionality it now aims for a deeper integration into existing biodiversity infrastructure landscape by making use of its service oriented architecture.

The workshop will present the current state of the Platform and intends to discuss new features to further integrate the Platform.

W19: Construction of biodiversity knowledge graphs driven by federated text mining tools

Organizers: *Riza Batista-Navarro, Sophia Ananiadou, Teodor Georgiev, Evangelos Pafilis, Lyubomir Penev, Viktor Senderov, Axel J. Soto, William Ulate*

Abstract

Knowledge graphs - knowledge bases which encode information using graph structures - have recently demonstrated their power in supporting knowledge discovery tasks such as data aggregation and semantic search. Well-known knowledge graphs such as DBpedia and Google's Knowledge Graph have been built to store general-domain facts, and are now widely used in many knowledge discovery applications. We should also be able to construct large-scale graphs capturing biodiversity knowledge, which will enable linking and consolidation of information from multiple complementary sources - databases such as the Encyclopedia of Life (EoL), Global Biodiversity Information Facility (GBIF), Biodiversity Heritage Library (BHL) and Pensoft.

This workshop is focussed on the demonstration and use of technologies that facilitate the construction of biodiversity knowledge graphs based on various data sources. Firstly, we shall showcase tools for automatically analysing secondary data, i.e., biodiversity literature. These range from tools or services that can automatically recognise mentions of pertinent concepts (e.g., taxa, environments) such as EXTRACT, to those that disambiguate and link such mentions with controlled vocabularies and primary (occurrence) data, through to ones which identify semantically related names (e.g., using distributional semantics). Importantly, we will demonstrate how a graph database - a Neo4j instance - can be populated with the automatically extracted information. Emphasis will also be given to the benefits of representing information using a graph database. Furthermore, we shall demonstrate how these diverse tools - as well as those contributed by other members of the community - can be federated, i.e., integrated into unified pipelines, using Argo: a graphical Web-based workbench for user-interactive construction of processing workflows. In this way, the workshop will provide the know-how for building knowledge graphs through bespoke processing workflows that do not

require any programming effort.
Applications of the above technologies in the context of use cases will then be presented. A specific example is the graph representation of knowledge stored in World Flora Online (WFO) which would: (1) integrate WFO content with data from other platforms, e.g., GBIF, EoL, and (2) enable searching for more descriptive plant species information, e.g., related common or vernacular names, location, habitat, reproductive state. An open discussion of further use cases that are of interest to the community will be held.

W21: Biological Interaction Data Workgroup

Organizers: *Antonio M. Saraiva, Jennifer Hammock, Quentin Groom*

Abstract
The objective of this workshop is to present the advances of the workgroup, since its creation at the 2016 TDWG Conference, work on specific issues raised during the year, and plan next steps, which may include the creation of an Interest Group on Biological Interactions Data within TDWG.
Scientists use a variety of methods to collect, record, and store biological interaction data (predator-prey, parasite-host, pollinator-plant, etc.). Uses for these data are equally diverse. For example, they could play an important role in building decision support systems for conservation and sustainable use in agriculture. Numerous efforts are underway to aggregate, organize, and efficiently disseminate these data. However, we lack a formal data standard to support this work. The goal of this workshop is to provide an opportunity, for those involved or interested in the digitization of biological interaction data, to share their experiences and ideas so that we can move forward and propose a biological interaction data standard.

W23: Towards an Online Atlas of Phenology

Organizers: *Zoe Panchen, Joel Sachs, Mélanie Bélisle-Leclerc, Scott Chamberlain, Jonathan Davies, Pamela Soltis, Rob Guralnick*

Abstract
An on-line phenology atlas would be a platform for integrating phenology data from the many individual researchers, institutions, and citizen science programs that are willing to make their phenology records available. It would provide capabilities for analyzing and visualizing the data with a species-based, location-based, or phenophase-based focus. This workshop will explore possible visions for such an atlas, and possible steps for making it a reality. Goals of the workshop include:

A survey of existing initiatives aimed at standardizing and integrating phenology data.
Identifying barriers to standardization and integration.
Generating use cases.
Generating competency questions - i.e. What queries would we want an on-line atlas of phenology to support?
Identifying possibilities for future development.
Submissions are invited relating to any of the above. As well, the organizers plan a number of small experiments in phenological data integration prior to the workshop, so that discussion is informed by what's currently easy, and what's currently hard. If you would like to participate in these experiments (by contributing data, use cases, or technical capacity), please contact the organizers.

Interest Group Abstracts

IG01: Data Quality

Organizers: *Arthur D. Chapman, Antonio M. Saraiva*

IG02: Technical Architecture Meeting

Organizers: *TBA*

IG03: Species Information

Organizers: *Francisco Pando*

IG04: Collections Description Meeting

Organizers: *Alexander Thompson, Deborah Paul*

Abstract
The Collections Description Interest Group is dedicated to developing and supporting the Natural Collections Description (NCD) data standard for describing entire collections of natural history materials. Examples include collections of specimens, observation data, original artwork, photographs, and materials from the many voyag ... https://doi.org/10.3897/tdwgproceedings.1.20322