

# A data management workflow of biodiversity data from the field to data users

Rachel A. Hackett<sup>1,2,5</sup> , Michael W. Belitz<sup>1,3</sup> , Edward E. Gilbert<sup>4</sup> , and Anna K. Monfils<sup>1</sup> 

Manuscript received 21 July 2019; revision accepted 21 October 2019.

<sup>1</sup>Department of Biology, Institute for Great Lakes Research, Central Michigan University, Bioscience Building 2100, 1455 Calumet Court, Mount Pleasant, Michigan 48859, USA

<sup>2</sup>Michigan Natural Features Inventory, Michigan State University Extension, P.O. Box 13036, Lansing, Michigan 48901-3036, USA

<sup>3</sup>Florida Museum of Natural History, University of Florida, Gainesville, Florida 32611, USA

<sup>4</sup>School of Life Sciences, Arizona State University, Tempe, Arizona 85287, USA

<sup>5</sup>Author for correspondence: hackett5@msu.edu

**Citation:** Hackett, R. A., M. W. Belitz, E. E. Gilbert, and A. K. Monfils. 2019. A data management workflow of biodiversity data from the field to data users. *Applications in Plant Sciences* 7(12): e11310.

doi:10.1002/aps.3.11310

**PREMISE:** Heterogeneity of biodiversity data from the collections, research, and management communities presents challenges for data findability, accessibility, interoperability, and reusability. Workflows designed with data collection, standards, dissemination, and reuse in mind will generate better information across geopolitical, administrative, and institutional boundaries. Here, we present our data workflow as a case study of how we collected, shared, and used data from multiple sources.

**METHODS:** In 2012, we initiated the collection of biodiversity data relating to Michigan prairie fens, including data on plant communities and the federally endangered Poweshiek skipperling (*Oarisma poweshiek*).

**RESULTS:** Over 23,000 occurrence records were compiled in a database following Darwin Core standards. The records were linked with media and biological, chemical, and geometric measurements. We published the data as Global Biodiversity Information Facility data sets and in Symbiota SEINet portals.

**DISCUSSION:** We highlight data collection techniques that optimized transcription time, including the use of predetermined and controlled vocabulary, Darwin Core terms, and data dictionaries. The validity and longevity of our data were supported by voucher specimens, metadata with measurement records, and published manuscripts detailing methods and data sets. Key to our data dissemination was cooperation among partners and the utilization of dynamic tools. To increase data interoperability, we need flexible and customizable data collection templates, coding, and enhanced communication among communities using biodiversity data.

**KEY WORDS** biodiversity data; FAIR data principles; natural history collections; plant diversity; research infrastructure.

The ability to digitally record, store, and retrieve vast quantities of biological, physical, chemical, ecological, and geographical data creates novel opportunities for understanding dynamic natural systems and guiding evidence-based resource management (Heidorn, 2008; Hampton et al., 2013; Morrison et al., 2017; Cheruvilil and Soranno, 2018). Accessible data from multiple sources have been integrated and used to answer broad-scope questions in biogeography, climatology, ecology, evolution, and phylogenomics (Chapman, 2005; Soltis et al., 2018), and have been used to refine species distribution models (Garcia-Rosello et al., 2015; Botella et al., 2018), track changes in morphology (MacLean et al., 2018), and explore spatial phylodiversity patterns (Allen et al., 2019).

Initiatives have been implemented by the natural history collection, scientific research, and natural resource management

communities to make their biodiversity data publicly available (Powers and Hampton, 2019). Citizen scientists are transcribing large volumes of data for collections (Chandler et al., 2017; Yost et al., 2018); journals are encouraging the publication of raw data and code online for scientific research (Heidorn, 2008; Biodiversity Collections Network, 2019; Powers and Hampton, 2019); and adaptive management projects are collecting, storing, and sharing data for national and regional management communities (Hunt et al., 2015; Great Lakes Phragmites Collaborative, 2017). The scale and resolution of biodiversity data being collected, stored, and made publicly accessible are rapidly expanding.

Despite the increased use of biodiversity informatics in conservation research and management, it was estimated that less than one percent of ecological data collected meet FAIR (findable, accessible,

interoperable, and reusable) data principles (Reichman et al., 2011). For many projects, data are organized and stored pragmatically for the initial goals of the data collector without considering the future reuse of the data by the scientific community at large (Heidorn, 2008; Hampton et al., 2015; Cheruvilil and Soranno, 2018; Lewis et al., 2018). The heterogeneity of biodiversity data presents challenges in using FAIR data principles for large-scale research (Wilkinson, 2016; Lewis et al., 2018; Powers and Hampton, 2019).

Although the collections, management, and research communities all recognize the value of FAIR data principles (Wilkinson, 2016; Mons et al., 2017), there is a need for greater guidance in data curation and archiving (Guralnick and Hill, 2009; Parr et al., 2012; Applegate, 2015; Wilkinson, 2016; Mons et al., 2017). Hindrances to a more consistent yet flexible standard of data curation and FAIR data principles among communities documenting biodiversity may stem from the individual-driven culture of different scientific disciplines. The individuality of the data can be seen in the inconsistent and, at times, confusing formats delivered to data users outside of the initial project. The lack of focus on long-term data utility and insufficient funding for the personnel and technological components of data management, such as script development and storage, exacerbate data interoperability challenges (Hunt et al., 2015; Culley, 2017); however, established data standards, online storage, and open-source tools exist to enable data sharing. Several biodiversity data guidelines and standards exist, with the most widespread being Darwin Core (DwC), DMPTool, the European Search Catalogue for Plant Genetic Resources (EURISCO) Descriptors, and the Federal Geographic Data Committee (FGDC; Guralnick and Hill, 2009; Giovanni et al., 2012; Wiczorek et al., 2012; Darwin Core Maintenance Group, 2014). Despite this, the use of these standards is not consistent across communities collecting biodiversity data.

The inconsistent use of biodiversity data guidelines and standards places a burden on the future data user. After finding the data, the user must then locate separate metadata files, reorganize and rename fields, integrate aggregated data, and dispose of inadequately documented or questionable data (Parr et al., 2012; Hunt et al., 2015; Culley, 2017). Abiding by FAIR data principles early in data collection projects will allow greater amounts of data to be retained and reduce the time spent cleaning the data by future users. The increased amount of retained data could then be used to better

inform conservation efforts and research programs across geopolitical, administrative, and institutional boundaries.

In 2012, we began targeted efforts to document the biodiversity associated with globally vulnerable prairie fen wetlands. Our research goals were to elucidate local- and landscape-level drivers of diversity in these highly speciose wetland ecosystems. We included plant diversity research and species-focused studies concerning the biology, ecology, and behavior of the federally endangered Poweshiek skipperling (*Oarisma poweshiek* Parker; Lepidoptera: Hesperidae) (U.S. Endangered Species Act [ESA 1973, as amended]; SARA, 2002; COSEWIC, 2014). Our research included the use of historical records to elucidate the factors leading to the sharp decline in the number and sizes of Poweshiek skipperling populations.

The workflow provided is a case study of how we collect, share, and use data from natural history collections, fieldwork, and outside contributing organizations (Fig. 1). We discuss the challenges of efficiently collecting and transcribing data, integrating data, and validating and documenting data to increase its longevity. Our workflow could be used to inform data collection and the curation of biodiversity data while abiding by FAIR data principles, enabling researchers, managers, and policymakers to address issues of global and future concern.

## METHODS AND RESULTS

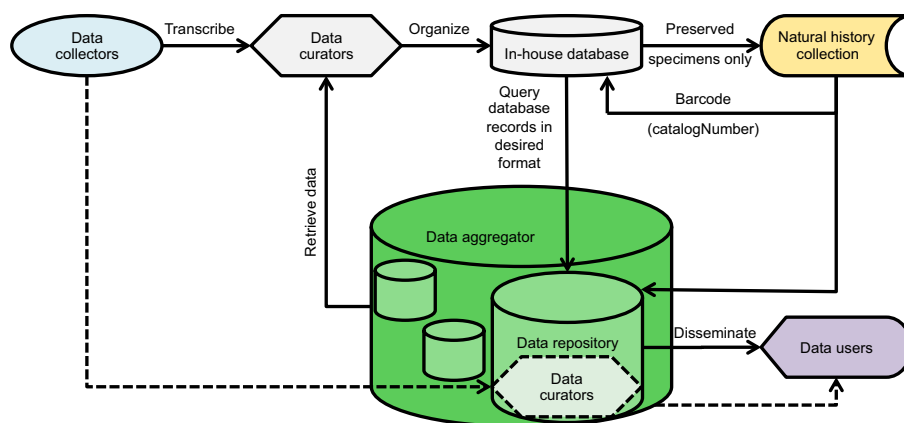
### Biodiversity data and measurements

The data records for both the plant diversity and butterfly surveys encompassed three levels of DwC terms: *location*, *event*, and *occurrence* (Table 1, Box 1). Field data and measurements were recorded in a written field journal, on a handheld GPS unit (Juno SB Handheld GPS; Trimble, Sunnyvale, California, USA) with a customized Trimble Data Dictionary, and/or in voice recordings (Digital Voice Recorders ICD-BX140; Sony, Tokyo, Japan). The controlled vocabulary was predetermined and used whenever possible for recording attributes and measurements.

**Plant community survey**—A prairie fen, as documented by the Michigan Natural Features Inventory (MNFI) (Michigan Natural Features Inventory, 2011), was considered to be a *location record*.

Prairie fens were selected for study based on ecological location (i.e., ecoregion), accessibility, and variability in quality and location-level measurements (see Hackett et al., 2016 and Pogue et al., 2019 for more detailed selection methods). Each of the 29 locations surveyed was assigned a location identifier and attributes describing its geospatial position (Box 1). Field measurements taken at the location level were documented in handwritten field journals. Other measurements were calculated using geospatial data, such as area and surrounding proportion of wetland (see Box 1).

Each location was sampled using an area-proportional random transect-quadrat method (see Hackett et al., 2016 for the detailed sampling method). Each 1-m<sup>2</sup> quadrat was considered a sampling *event record* with



**FIGURE 1.** Generalized workflow describing the people, places, and processes involved in the transfer of data from the field to users. Dashed lines represent alternative pathways offered by some online data repositories. Descriptions and examples of each group are provided in Table 1.

**TABLE 1.** Term table clarifying the terminology used in this paper (Wieczorek et al., 2012; Darwin Core Maintenance Group, 2014). Definitions are provided for terms with potential for variable interpretation.

Term	Definition	Example
Biodiversity data	Data related to knowledge about individual biological organisms and the ecosystems they shape	organismal, geographical, ecological, environmental data
Data	Unstructured quantitative and qualitative measurements and facts that are yet to be analyzed	count, length, life stage
Data aggregator	A virtual entity where information can be stored, searched, and queried for distributable information. Many data aggregators are also data repositories.	GBIF, iDigBio, DataONE
Data collector	Person or machine that collects primary data	citizen scientists, optical character resolution, technicians
Data curator	Person or organization that organizes, analyzes, and disseminates data into information	researchers, collections, management agencies
Data repository	A virtual entity where data curators can deposit, edit, and manipulate their data	LeptNet, Midwest Consortium of Herbaria
Data user	Person or organization that retrieves data or information from data curators, data repositories, and/or aggregators to clean, analyze, and use for their own purposes	land owners and managers, management agencies, non-profit organizations, researchers, students,
Event	A Darwin Core table/class tied to a location, date, and time. There can be many event records corresponding to the same location.	
Field	Categorization of a set of data values as a column in a table. Also referred to as an attribute, column, or term name.	locationID, eventID, eventDate, lifeStage
GPS data dictionary	A form created to record attribute data and measurements to accompany a record or shapefile generated by a GPS unit. Often customizable.	ArcCollector, Trimble Pathfinder data dictionary, ColectoR, iNaturalist
In-house database	Location in which organized data are stored, linked, searched, and queried for distributable information	BIOTICS World Heritage Database, SQL database, Specify
Information	Processed, organized, structured, or presented data that are given context so they are useful for answering a question	figure, map, mean, statistical analysis
Location	A Darwin Core table/class tied to a geographic location	Brandt Road Fen, Site 5
Natural history collection	An archived collection of preserved physical specimens	herbarium, museum
Occurrence	A Darwin Core table/class tied to the collection or observation of an organism or related material during an event. There can be many occurrence records corresponding to the same event and the same location.	
Record	Related data. Also referred to as a row.	
Table	A grouping of related data, with each row corresponding to one record and columns containing fields describing data values. Also referred to as a worksheet, spreadsheet, or class.	Location, Event, Occurrence, MeasurementOrFact

a unique combination of point coordinates, date, and time, and was assigned a unique identifier. A total of 1445 sampling events were recorded. Event-level measurements (e.g., species richness, percentage of water cover) and the filenames of quadrat photographs were recorded in a written field journal and in the handheld GPS unit (Box 1). Water and soil samples were sent to laboratories for analysis, and the returned measurements were recorded at the event level (Box 1). The filenames of the photographs of each quadrat were linked to the event record as associated media.

Each plant species rooted in a 1-m<sup>2</sup> quadrat was recorded as an *occurrence record*, with a unique alphanumeric field number. In the field, data collectors recorded taxonomic information and measured Daubenmire's percent cover class for each species (Daubenmire, 1959) (Box 1). Occurrence records were documented as either a preserved specimen or a human observation.

An occurrence was classified as a preserved specimen if it was the data collector's first observation of a species, or if the species could not be identified in the field. Preserved specimen occurrences were collected and deposited in the Central Michigan University Herbarium (CMC). Each specimen collected was given an identifying record number as per natural history collection protocol (Smith and Chinnappa, 2015). The filenames of the photographs were recorded.

An occurrence was classified as a human observation if the species occurred in a 1-m<sup>2</sup> quadrat and the lead data collector had

previously collected a voucher of that species. After the first observation was collected and vouchered, voucher specimens were subsequently linked with human observation occurrence records of that species.

A total of 21,213 plant occurrences were recorded, which consisted of 19,684 human observations and 1529 preserved specimens. A total of 575 voucher specimens were provided by the two lead data collectors (315 from Rachel Hackett, 260 from Clint Pogue). Another 948 preserved specimens were collected whose identities required herbarium confirmation or were a new species occurrence for the county (Hackett et al., 2016).

**Butterfly surveys**—Poweshiek skipperling are small butterflies with a wingspan of approximately 3 cm. They emerge as flying, reproductive adults from late June to mid July (i.e., the flight period) (Selby, 2005; Belitz et al., 2019). They were abundant until as recently as the mid-1990s, with hundreds of populations, each population consisting of hundreds of individuals, in prairies and prairie fens throughout the upper Midwest (Selby, 2005). Over the past 20 years, their numbers dropped dramatically to only six extant sites worldwide, and their combined abundance in four Michigan prairie fens was estimated to be fewer than 400 adult individuals (Belitz et al., 2019).

A prairie fen, as described in the plant community survey section, was also considered a *location record* for the butterfly surveys. Butterfly surveys were conducted at the four Michigan prairie fens

**BOX 1.** Biodiversity data recorded at each record level. Darwin Core terms were used for the attributes of each record class, regardless of survey type (Wieczorek et al., 2012; Darwin Core Maintenance Group, 2014). Measurements for each survey type were included in the Darwin Core MeasurementOrFact table. For definitions, see Table 1. \*Not a Darwin Core term †Categorical text measurements.

	Location	Event	Occurrence
Attributes for each record	<ul style="list-style-type: none"> <li>• locality</li> <li>• locationID</li> <li>• county</li> <li>• stateProvince</li> <li>• countryCode</li> <li>• decimalLatitude</li> <li>• decimalLongitude</li> <li>• minimumElevationInMeters</li> <li>• verbatimLocality</li> <li>• locationRemarks</li> <li>• habitat</li> </ul>	<ul style="list-style-type: none"> <li>• eventID</li> <li>• locationID</li> <li>• datasetID</li> <li>• datasetName</li> <li>• eventDate</li> <li>• eventTime</li> <li>• decimalLatitude</li> <li>• decimalLongitude</li> <li>• minimumElevationInMeters</li> <li>• samplingProtocol</li> <li>• sampleSizeValue</li> <li>• sampleSizeUnit</li> <li>• samplingEffort</li> <li>• habitat</li> <li>• fieldNotes</li> <li>• PhotoStart*</li> <li>• PhotoEnd*</li> </ul>	<ul style="list-style-type: none"> <li>• fieldNumber</li> <li>• recordNumber</li> <li>• eventID</li> <li>• locationID</li> <li>• datasetID</li> <li>• basisOfRecord</li> <li>• sex</li> <li>• fieldIdentification*</li> <li>• occurrenceRemarks</li> <li>• recordedBy</li> <li>• Journal*</li> <li>• PhotoStart*</li> <li>• PhotoEnd*</li> </ul>
Plant survey measurements	<u>Prairie fen</u> <ul style="list-style-type: none"> <li>• species richness</li> <li>• Shannon's Diversity Index</li> <li>• surrounding land cover proportions (eight classes, four scales)</li> <li>• area</li> <li>• perimeter: area ratio</li> <li>• accumulated least cost path</li> <li>• least cost path distance</li> <li>• mean near distance</li> </ul>	<u>Quadrat (1 m<sup>2</sup>)</u> <ul style="list-style-type: none"> <li>• species richness</li> <li>• porewater pH</li> <li>• porewater temperature</li> <li>• depth to water table</li> <li>• water sample analysis results (three measurements)</li> <li>• soil sample analysis results (10 measurements)</li> <li>• floristic zone category†</li> <li>• estimate of percent cover (bare ground, water, vegetation)</li> </ul>	<u>Plant species quadrat</u> (Preserved specimens and human observations) <ul style="list-style-type: none"> <li>• Daubenmire cover class (Daubenmire, 1959)</li> </ul>
Butterfly survey measurements	<u>Area within prairie fen with suitable habitat</u> <ul style="list-style-type: none"> <li>• area</li> <li>• perimeter</li> </ul>	<u>Point (revisited)</u> <ul style="list-style-type: none"> <li>• wind speed</li> <li>• cloud cover</li> <li>• DAFOR scale density ranking of nectar sources (Rich et al., 2005)†</li> <li>• temperature</li> <li>• relative humidity</li> </ul>	<u>Butterflies</u> <ul style="list-style-type: none"> <li>• distance from observer</li> <li>• wing wear class</li> <li>• nectar sources utilized†</li> <li>• behavior over 10 min†</li> </ul> <u>Flowering species between point events</u> (Human observation) <ul style="list-style-type: none"> <li>• count of each flowering unit</li> </ul>

with extant Poweshiek skipperling populations. Vegetative surveys were also conducted at these sites.

Perpendicular transects were drawn to intersect the areas within the sites with most concentrated Poweshiek skipperling observations. Points along the transects at each site were revisited at systematically different times each day throughout the Poweshiek skipperling flight period. Each visit to a transect point was considered a sampling *event record*. From 12 June to 21 July 2017, a total of 2355 event records were created. Measurements at the event level (e.g., cloud cover, nectar source density) were taken using voice recordings (see Box 1). For each event, temperature and relative humidity data were collected using HOBO weather data loggers (Onset, Bourne, Massachusetts, USA) placed at the site during the survey season, and these data were linked to the events by date and time.

Both the target butterfly species and possible nectar sources were documented as *occurrence records*. All these occurrences were human observations. Occurrence measurements (e.g., distance from observer, butterfly behavior) were recorded for up to 10 min

on a voice recorder (Box 1). A total of 2035 butterfly observations, 97 of which were Poweshiek skipperlings, were recorded, and 73 Poweshiek skipperling behavior recordings were made (see Belitz et al., 2019 for detailed survey methodology).

**Historical Poweshiek skipperling data set**—A total of 3676 Poweshiek skipperling occurrence records from 37 external data sources were integrated into the data set. These occurrence records were based on either preserved specimens or human observations. The occurrence records based on preserved specimens included data available online from natural history collections (e.g., Lepidoptera of North America Network's [LepNet] Symbiota portal [http://symbiota.org]), previously non-digitized specimen label data, or historical data sets from state and federal agencies. Additional occurrence records were based on human observations, including citizen scientist observations (e.g., iNaturalist [https://www.inaturalist.org/]), society publications (e.g., The Lepidopterists' Society Season Summary), and long-term and ongoing survey records from researchers at the MNFI.



For details of the complete, dynamic Poweshiek skipperling data set, aggregation, and cleaning methods, see Belitz et al. (2018a, b). These data went through several iterations of quality control (e.g., georeferencing, removing duplicate records) before being integrated as acceptable records.

## Data workflow

*Data collectors* were trained individuals that gathered, recorded, and measured data, or machines that directly recorded data (such as weather data loggers) (Table 1). The human data collectors had variable levels of engagement in the research, as some were field technicians who only collected data and others researchers who designed the field studies, wrote protocols and procedures, curated and analyzed data downstream, and published their methods and results. The recorded data were supplied to the data curator (Fig. 1). The written data sheets were also taken to later cross-reference with voice recordings during transcription.

*Data curators* were responsible for transcribing and organizing data, retrieving data recorded by outside sources, linking media (e.g., images) to appropriate records, and disseminating information online and to partners (Fig. 1, Table 1). Our data curators worked to develop and augment a relational in-house database (Fig. 1).

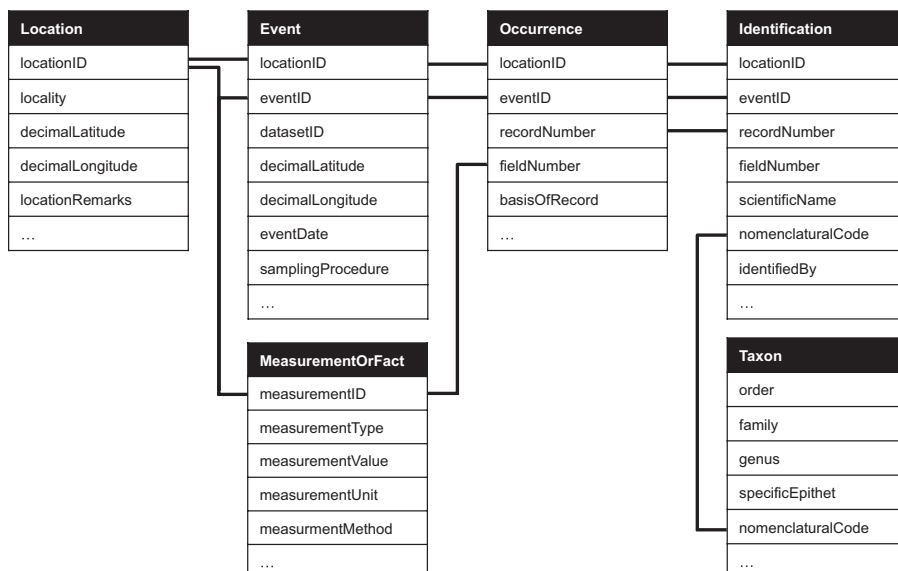
Within the *in-house database*, the transcribed data were uploaded or appended into a relational database management system. The organization was guided by DwC documentation to ensure all data could be easily queried and exported to the DwC data standard (Fig. 2) (Darwin Core Maintenance Group, 2014). Although DwC was developed with natural history collections in mind, basic DwC principles were applied to the plot-based vegetation and butterfly surveys because it is a versatile standard that was supported by our online data repository.

In DwC, all media are contained in an associated media field that lists the identifiers of the media (e.g., globally unique identifier [GUID], publication, URL). The data curators wrote an R script to intuitively rename field image filenames and write tables of image URLs to append to the in-house database ([https://github.com/hackelra/data\\_workflow\\_paper.git](https://github.com/hackelra/data_workflow_paper.git)). The renamed images were uploaded to our data server.

The data curators created queries for quality control in the in-house database (Sutter et al., 2015; Hunt et al., 2016; Yost et al., 2018). These queries could locate records without links to other tables and flag records to be checked for transcription error. Queries designed to export requested data for various uses or repositories were also created, including herbarium labels for preserved specimens, lists of threatened and endangered species occurrences for permit renewal, and records to upload to our online data repository.

A *natural history collection* (i.e., CMC Herbarium) housed preserved specimens of plant occurrence records (Table 1, Fig. 1). The data curators submitted a single spreadsheet file exported from our database query to the CMC Herbarium. From that spreadsheet, the labels were generated in the format desired and printed on archival paper. CMC Herbarium followed normal procedure for curating and digitizing preserved specimens. The data curator later added a barcode containing a unique catalog number for each preserved specimen record in the identification table in our relational database. The spreadsheet was also used for digitized records in online data repositories to reduce legacy data and the need to transcribe from the printed label (Fig. 2).

The Consortium of Midwest Herbaria portal (<http://midwestherbaria.org>) was used as a *data repository/aggregator* to manage and publish the plant occurrence data. This biodiversity data portal was established using the Symbiota open-source software platform (Gries et al., 2014). After exporting a query of occurrence records from our in-house database, human observation occurrence records were batch-uploaded to a new collection: the Prairie Fen Biodiversity Project (PFBP) Observation Collection. Preserved specimen occurrence records were batch-uploaded to the online CMC Collection using the label query we provided (Fig. 1). As a regional member of the greater SEINet Portal Network, publishing data within the Consortium of Midwest Herbaria interface ensured the extended distribution of data: all plant occurrence data were published to global data aggregators Global Biodiversity Information Facility (GBIF; <https://www.gbif.org/>) and iDigBio (<https://www.idigbio.org/>) through the portal's DwC publishing module (<http://midwestherbaria.org/portal/collections/data-sets/datapublisher.php>) (Central Michigan University Herbarium, 2018, 2019). *Data users* were invited to retrieve our checklists and data sets housed in the online repositories (Table 1).



**FIGURE 2.** Relational in-house database. Each box with a black header represents a table included in the database. Each row in the boxes represents a field/attribute. Three dots (“...”) indicate that additional fields were not included for ease of viewing. Black lines represent common fields between tables that were linked for query capabilities. Darwin Core Standards were used for the table and field names (Wieczorek et al., 2012; Darwin Core Maintenance Group, 2014).


**Dissemination of data**—During the project initiation stage, we were cognizant of FAIR data practices within our lab and in relation to partner needs. We worked closely with partners both in the field and when creating data

products to ensure data accessibility for them. We partnered with non-governmental organizations, governmental agencies, businesses, and private landowners. The non-governmental organizations included the Livingston Land Conservancy, Michigan Natural Features Inventory, Michigan Nature Association, North Oakland Headwaters Land Conservancy, and The Nature Conservancy. The governmental agencies were the Michigan Department of Natural Resources, Springfield Charter Township, Jackson County, Washtenaw County, and the U.S. Fish and Wildlife Service. The businesses were Consumers Energy and the Oak Pointe Country Club.

For each of the 29 locations, a plant checklist was created in the Consortium of Midwest Herbaria portal that contained all our preserved specimens and human observations (<http://midwestherbaria.org/portal/projects/index.php?pid=113>) (Fig. 3). Thirty-four historical specimens were added to checklists from the University of Michigan Herbarium (MICH) and CMC Herbarium. Using our checklists, data users could download occurrence records, download lists of species names, view associated media, view interactive maps of occurrence records, access species information, use polyclave identification keys, or test their identification skills.

Data users who wanted access to geographical information about listed/protected plant species registered for a free account with the data aggregator and contacted the collection via the online portal. We granted partners permissions as “Rare Species Readers” for the CMC and PFPB collections. A “parent” checklist, Prairie Fens of Southeast Michigan, was created, into which all our individual prairie fen plant checklists were fed. Our plant community data sets were published in GBIF and iDigBio using the Darwin Core Archive Publishing tool in the SEINet portal for potential use by those outside of our partner network.

For the Poweshiek skipperling, data users can access our data set of historic Poweshiek skipperling occurrence records in GBIF’s repository through the Integrated Publishing Toolkit (Robertson et al., 2014; Belitz et al., 2018a, b). The results of ongoing butterfly surveys were regularly reported to research and conservation partners working to address the complexities and challenges facing Poweshiek skipperling populations. After publication of the data set, additional feedback and data verification were received from lepidopterists with extensive knowledge of historical butterfly collections, locations, and individual collectors. We were able to enhance the data set using this feedback, creating new data set versions as

Home > Prairie Fen Research Collaborative > **Waterloo Recreation Area - Mount Hope Road Fen/Glenn Fen** 

### Waterloo Recreation Area - Mount Hope Road Fen/Glenn Fen

**Authors:** Hackett, RA; Karbowski, HM; Peters, S; Monfils, AK;  
**Citation:** Hackett RA, Monfils MJ, Monfils AK (2016) Evaluating a sampling protocol for assessing plant diversity in prairie fens. *Wetland Ecology and Management* 24:609–622. doi: 10.1007/s11273-016-9491-1  
[Less Details](#)  
**Locality:** Mt Hope Road Fen (42.310100, -84.208600)  
**Abstract:**

Prairie fen dominated by *Carex*, *Asteraceae*, and grasses but encroachment by shrubs (*Rubus*, *Cornus*, *Betula*, *Toxicodendron vernix*). A few pockets and walls of shrub-carr and wooded fen. Evidence of some management. Hummocky in areas from *Carex stricta*, other areas dominated by other *Carex* and *Dasiphora fruticosa*. State owned.


For more information about the Prairie Fen Research Collaborative, visit <https://www.researchgate.net/project/Prairie-Fen-Research-Collaborative>


**Notes:** Also known as Glenn Fen. Direct questions about data and data usage to Anna K Monfils ([monf1ak@cmich.edu](mailto:monf1ak@cmich.edu)).

---


**Families:** 48  
**Genera:** 99  
**Species:** 148  
**Total Taxa (details):** 149

**ADOXACEAE**


*Viburnum dentatum*   
 Nicole Schmidt 1 [CMC]

*Viburnum lentago*   
 Rachel Hackett RHHK223 [PFBP], Rachel Hackett RHHK223 [PFBP]

**ALISMACEAE**

*Sagittaria latifolia*   
 Rachel Hackett RHHK538 [PFBP], Rachel Hackett RHHK538 [PFBP], Rachel Hackett RHHK538 [PFBP], Rachel Hackett RHHK538 [PFBP], more...

**ANACARDIACEAE**

*Toxicodendron vernix*   
 Rachel Hackett TOXVER [PFBP], Rachel Hackett... TOXVER [PFBP], Rachel Hackett TOXVER [PFBP], Rachel Hackett TOXVER [PFBP], more...

**Options**




**Search:**

☐ Common Names  
☒ Synonyms

**Filter:**

**Original Checklist**

☐ Display Synonyms  
☐ Common Names  
☐ Display as Images  
☒ Notes & Vouchers  
☐ Taxon Authors  
☐ Show Taxa Alphabetically

**Rebuild List**   

**FIGURE 3.** Screenshot of the Waterloo Recreation Area–Mount Hope Road Fen/Glenn Fen checklist of one prairie fen location (<http://midwestherbaria.org/portal/checklists/checklist.php?pid=113&clid=4393>). Specimens not collected for this project but geolocated to this location were added through the portal (e.g., Nicole Schmidt 1 [CMC]). This dynamic checklist is a “child” checklist contributing all human observations and preserved specimen plant occurrences to the Prairie Fens of Southwest Michigan (<http://midwestherbaria.org/portal/checklists/checklist.php?clid=4362&pid=113>), which includes species occurrences from all 29 surveyed prairie fens.

the new information was provided (<http://ipt.idigbio.org/resource?r=cmc>).

## DISCUSSION

The dynamic prairie fen biodiversity data set we developed bridges taxonomic boundaries (e.g., plants, butterflies), abiotic and biotic measurements at varying scales (e.g., location, occurrence), sampling methodologies, and lead researchers. The development of a relational database using DwC class and term organization allowed us to easily customize our data to the needs of our partners, data repositories, permitting agencies, and publications (Wieczorek et al., 2012; Darwin Core Maintenance Group, 2014). To communicate with our partners, we developed mechanisms of data curation that facilitated open data exchange, dissemination, and utility.

### Data recording

**GPS data dictionary**—GPS data dictionaries allowed us to collect geospatially linked data and record measurements using a controlled vocabulary. These devices are easy to carry in the field and reduced data curator transcription time and error. The main limitations of the GPS data dictionaries are that (1) the data are recorded in flat tables requiring subsequent parsing by the data curator, (2) the text fields have a limited number of characters, and (3) when moving between the recorded data sets on the handheld unit, the files can easily become corrupted.

These limitations could be addressed by implementing the specific programming of the data dictionaries to record and/or export information in multiple linked tables in a one-to-many fashion and with expanded text fields. Emerging GPS survey tools (e.g., ArcCollector, Survey123) have the capability to customize and relate linked tables to a collected shapefile feature (e.g., polygons, polylines, points). These apps are flexible and allow the user to modify data form templates or design forms from scratch. The forms can be designed to collect attributes and measurements at the location, event, and occurrence levels on the same form. According to Hardisty et al. (2019), if customizable digital data collection form templates following common data standards such as DwC were available, data curator time could be reduced and data interoperability improved. Balance would be needed to avoid overcomplicating digital data collection forms for data collectors, which would increase field collection time.

Projects collecting field data only at the occurrence level might find apps such as iNaturalist or ColectoR (Maya-Lastra, 2016) sufficient, but these were not flexible enough to accommodate the field measurements we collected at the location, event, and occurrence levels. Other projects and classrooms involving occurrences have developed efficient protocols to use apps to voucher and validate specimens while retaining linked metadata and media (Heberling and Isaac, 2018).

**Voice recordings**—Voice recordings of butterfly observations reduced the number and weight of items that were carried in the field. Voice recording allowed the data collector to reliably track rapidly moving butterflies, but placed an additional burden on the data curator to transcribe the recordings into digital tables. Lewis et al. (2018) used voice-to-text software to dictate observations, but our voice records required the transcription of observations

into digital data using appropriate fields and controlled vocabulary instead of a single, catch-all text field a voice-to-text software would provide.

**Written field journals**—Plant data were recorded in field journals and included descriptive data and field-based sketches that could be easily cross-referenced for field identifications. Some GPS data dictionaries and apps can be difficult to cross-reference in the field because of corrupted files or a lack of cellular service. The written records were not easily corrupted or lost, but the transcription of data and observations from more than 50 field journals was time consuming.

Butterfly data were also recorded in written data sheets for cross-referencing with voice recordings and backup data in case of voice recording file loss or corruption. To prevent data loss, these sheets were scanned and saved as PDFs using the CamScanner app (CC Intelligence Corporation, Milpitas, California, USA), and then uploaded to a central cloud storage drive. In addition, maintaining field journals, specimens, and digital data requires a long-term commitment to curation.

### Integrating media files

Augmenting the event and occurrence records led to an optimized specimen collection, data integration, and reuse. Event and occurrence records containing associated media have expanded the value of the data to other users, because observations, validation, and measurements can be made outside the original purpose of the data (Vellend et al., 2013; LaFrankie and Chua, 2015; Schindel and Cook, 2018; Thiers et al., 2019). Our field cameras lacked a naming customization feature, which necessitated later renaming of the files. Renaming the photograph files, uploading the files to the data server, and integrating the URLs of the files into the associated media field of the event or occurrence were time-intensive activities. With the improvement of cameras on GPS handheld units, tablets, and cell phones and the increased affordability of memory cards, the intuitive renaming of the photographs can now be bypassed to a degree. GPS survey apps for use with these devices typically have naming customization options.

### Incorporating measurements

Our projects involved collecting data beyond the attributes described in DwC. We classified these data as measurements, each with its own metadata (e.g., measurementType, measurementUnit), in the DwC MeasurementOrFact table. This project organization was flexible; it was adapted over time by three different graduate student researchers, and additional taxa were added without losing the readability of the data. Keeping the metadata for each integrated measurement allowed for quality control and consistency of the measurements and units from year to year. Measurement metadata increased the long-term utility of the data and are crucial for research transparency and data set interoperability (Hampton et al., 2015; Griffin et al., 2018; Powers and Hampton, 2019). Within the MeasurementOrFact table, we could cite our published manuscripts, theses, and other documents that provided details of our study designs and data-gathering methodologies. Future data users could see the measurement, look up the cited work, and determine its compatibility with other data sets.

For the DwC MeasurementOrFact table, the data are required to be in a text format. Our measurement values were either text or

numeric (e.g., species richness, wing wear class; Box 1). When converting our numeric measurement values to text, we lost many of the utilities offered by the database, such as the direct performance of calculations in queries; it was therefore necessary to adjust the column format after exportation before performing the analysis. One way to address this issue is to convert all text measurements into categorical integers. This adjustment worked in the case of wing wear in Poweshiek skipperlings, but was not ideal for other measurements. Categorical integers can confound the downstream use of the data, requiring additional keys or references.

We addressed the text/numerical issue by creating two MeasurementOrFact tables, one in numeric and one in text format. Using the numerical MeasurementOrFact table, we could perform calculations upon querying if desired, and negate the need for reformatting before the data analysis.

### Vouchering

With over 20,000 individual plant species observations, an at-risk prairie fen habitat, and restrictions on the collection of state- and federally listed plant and animal species, it was not feasible or possible to voucher every species at each locality. Instead, our lead data collectors collected at least one preserved specimen of each species they encountered to validate their identification ability, adding cogency to their human observations. This reduced time spent collecting and preparing preserved specimens and increased data longevity through the generation of a related, physical record. The time needed for the collection, preparation, and submission of preserved specimens was balanced by the increased likelihood of the long-term utility and versatility of the data (Beaman and Cellinese, 2012; Turney et al., 2015; Antunes and Schamp, 2017). The herbarium had a prompt mounting and digitization period, allowing the digital occurrence records to be FAIR within a year of collection (Wilkinson, 2016; Mons et al., 2017).

Because of restricted permitting, vouchers were not possible for the federally listed Poweshiek skipperling. We accessed vouchered museum specimens to confirm the identification of this butterfly and photo-documented its presence. Photo- or video-documenting of each observation was not possible due to the distances between the researchers and the butterflies and the rapid movement of these animals. An emphasis on the photo documentation of occurrences would have severely constrained our behavioral and ecological study.

### Documenting and integrating occurrence data sets

Data curators documented our methods and promoted them as a means to improve data access and longevity in peer-reviewed methods papers (Hackett et al., 2016; Belitz et al., 2018a; Pogue et al., 2019) and online data sets (Belitz et al., 2018b; Central Michigan University, 2018, 2019). Because our data were organized using standards similar to the data repositories/aggregators that we selected, we benefited in reduced time and effort required for dissemination. Although there were no direct incentives for us as data curators to include additional metadata and media outside of our published research in our data sets, these additions were essential in the data-cleaning and quality-control processes performed by us as data users.

As users of the Poweshiek skipperling occurrences data set (Belitz et al., 2018a, b), we recognized the need for data standardization and the linkage of supporting media and metadata. Occurrence

data from curators, repositories, and aggregators using DwC terms were easily parsed into tables and linked to our field-recorded data. Non-standardized data had to be converted into a compatible and standardized format. Standardization helped with the integration process; however, occurrences from both standardized and non-standardized sources were removed during the quality-control process if they did not include where or when the occurrence was generated. The removal of occurrences restricted our knowledge of historical Poweshiek skipperling populations. Despite this, adequate supporting data enabled the confirmation of a new historical population of Poweshiek skipperlings in Nebraska, expanding the known states that once had Poweshiek skipperlings (Belitz et al., 2018a). Supporting data were also used to correct locations and review the identification of occurrences by obtaining images of Poweshiek skipperlings and having an expert, who has worked with this species for over 20 years, confirm the specimen identification.

### Data access for users

By including partners and stakeholders at all possible stages of the project development and data archiving, we strengthened our interdisciplinary collaborations. We strived to bring partners and stakeholders into the field to connect them with our study system and immerse them in the challenges and opportunities of data collection. These collaborations ultimately enhanced the utility of the data to both researchers and managers by improving understanding and clarifying the expectations among the involved groups.

Our partners in prairie fen and Poweshiek skipperling research consisted of non-governmental organizations, state and federal agencies, businesses, and private landowners. Each of these groups had their own stakeholders and information needs. We were deliberate in our efforts to alert our partners to where our data were housed online. Many of our partners were unaware of the online tools available, which could be attributed to obstacles such as access, technological proficiency, and a lack of cross-community inclusion or communication during tool development (Polk, 2015; Baumber et al., 2018); however, no formal partner feedback survey was conducted on the topic of online tools and obstacles to their use.

We found that land managers, businesses, and private landowners were interested in obtaining plant species lists specific to different properties. Each checklist could be augmented on an ongoing basis with new occurrences for that location if the occurrence is recorded in the greater portal. For state and federal partners, we could provide lists of threatened and endangered species occurrences for permit renewals or decision making. For researchers, we published a methods paper (Hackett et al., 2016) and data sets (Central Michigan University, 2018, 2019). These peer-reviewed sources and online data sets complied with FAIR data principles, providing data access and promoting data longevity through the inclusion of metadata.

Our Poweshiek skipperling research partners, land managers, and decision makers in ongoing conservation efforts are using our peer-reviewed publication and data set (Belitz et al., 2018a, b). Their uses include efforts to understand the Poweshiek skipperling decline and biology, and to research potential reintroduction locations.

### Conclusions

The long-term utility of data lies in its versatility and documentation. To save resources but still meet interoperability requirements,



data curators must continue to put forethought into the minimum record standards and must control the vocabulary prior to establishing protocols and training data collectors (Lewis et al., 2018; Hardisty et al., 2019), as was shown here. The documentation of methods and metadata tied to the data, preferably in the same format as other similar data (e.g., DwC), will contribute to their long-term utility (Giovanni et al., 2012; Guralnick et al., 2018). Vouchered specimens and associated media not only support the longevity of the data, but also provide materials from which future users can verify the existing data and derive new data.

The interoperability of the data will permit larger, currently unconceived questions to be answered. The Biodiversity Information Standards (TDWG), which develop and maintain DwC among other standard biodiversity practices, already provide standards in vocabulary and organization that are versatile for new data types. TDWG is actively adapting to meet the needs of new communities and data types. Their Biodiversity Services and Clients Interest Group is tasked specifically with expanding TDWG's understanding of biodiversity data collectors and users in order to reach new data contributors and users. Publicly available digital data forms and/or templates that are generalized, yet flexible or customizable, would increase the acceptance and practice of data standards in new groups (Hardisty et al., 2019). Another route would be to develop and distribute flexible code or scripts to reorganize and append data collected in one flat sheet into a DwC relational database. The best practice would be for data curators to publish their methods and reference them in the metadata, making it easier for data users to combine data sets for stronger analyses and to answer large-scale questions.

The future of FAIR biodiversity data lies in the collaboration of data curators, repositories, aggregators, and users. Stakeholders in these groups should be included in conversations on how to achieve FAIR data principles to meet the long-term needs of the collections, research, and management communities (Hampton et al., 2015; Hunt et al., 2015; Biodiversity Collections Network, 2019). Conservation managers and environmental assessors are already major users of natural history collection data (Cantrill, 2018). The inclusion and integration of data collected by those users would improve future data usage, standards, and organization to promote the long-term utility of data sets. Adaptive management programs record and share the data needed to assess management success and provide a quantitative basis for management decisions (Williams and Brown, 2012). Such decisions could be informed by data from the collections or research communities and, likewise, research and collection could inform adaptive management practices. All communities could strive to work to understand, meet, and forecast their collaborative data needs.

## ACKNOWLEDGMENTS

The authors thank G. Nelson and M. Monfils for providing their expertise at multiple steps throughout the project, and Blake Cahill for reviewing the initial stages of this work. Support for these projects was provided by the Botanical Society of America, Central Michigan University, Hanes Trust Foundation, Michigan Garden Club Inc., NatureServe, and The Nature Conservancy. Additional funding was provided by the U.S. National Science Foundation (award DBI-1730526) to develop educational materials for biodiversity data literacy (<https://www.biodiversityliteracy.com/poweshiek-skipperling>).

This is contribution 133 of the Central Michigan University Institute for Great Lakes Research.

## CONFLICT OF INTEREST

Edward Gilbert is a developer and manager of the SEINet and LepNet online data repositories and aggregators and the associated Symbiota software. In this paper, both portals are used and discussed as being useful tools for disseminating data to our partners.

## LITERATURE CITED

- Allen, J. M., C. C. Charlotte, N. Barve, K. M. Neubig, L. C. Majure, S. W. Laffan, B. D. Mishler, et al. 2019. Spatial phylogenetics of Florida vascular plants: The effects of calibration and uncertainty on diversity estimates. *Science* 11: 57–70.
- Antunes, P. M., and B. Schamp. 2017. Constructing standard invasion curves from herbarium data: Toward increased predictability of plant invasions. *Invasive Plant Science and Management* 10: 293–303.
- Applegate, R. D. 2015. The importance of data management in wildlife conservation. *Wildlife Society Bulletin* 39: 449–450.
- Baumber, A., G. Metternicht, P. Ampt, R. Cross, and E. Berry. 2018. From importing innovations to co-producing them: Transdisciplinary approaches to the development of online land management tools. *Technology Innovation Management Review* 8: 16–27.
- Beaman, R. S., and N. Cellinese. 2012. Mass digitization of scientific collections: New opportunities to transform the use of biological specimens and underwrite biodiversity science. *ZooKeys* 209: 7–17.
- Belitz, M., L. K. Hendrick, M. J. Monfils, D. L. Cuthrell, C. J. Marshall, A. Y. Kawahara, N. S. Cobb, et al. 2018a. Aggregated occurrence records of the federally endangered Poweshiek skipperling (*Oarisma poweshiek*). *Biodiversity Data Journal* 6: e29081.
- Belitz, M. W., L. K. Hendrick, M. J. Monfils, D. L. Cuthrell, C. J. Marshall, A. Y. Kawahara, N. S. Cobb, et al. 2018b. Dataset: Aggregated occurrence records of the federally endangered Poweshiek skipperling (*Oarisma poweshiek*). Available at <http://ipt.idigbio.org/resource?r=cmc> [accessed May 2019].
- Belitz, M. W., M. J. Monfils, D. L. Cuthrell, and A. K. Monfils. 2019. Life history and ecology of the endangered Poweshiek skipperling *Oarisma poweshiek* in Michigan prairie fens. *Journal of Insect Conservation* 23: 635–649.
- Biodiversity Collections Network. 2019. Extending U.S. biodiversity collections to promote research and education. American Institute of Biological Sciences, Washington, D.C., USA.
- Botella, C., A. Joly, P. Bonnet, P. Monestiez, and F. Munoz. 2018. Species distribution modeling based on the automated identification of citizen observations. *Applications in Plant Sciences* 6: e1029.
- Cantrill, D. J. 2018. The Australasian Virtual Herbarium: Tracking data usage and benefits for biological collections. *Applications in Plant Sciences* 6: e01026.
- Central Michigan University Herbarium. 2018. Central Michigan University. Occurrence dataset. <https://doi.org/10.15468/iykbez> [accessed via GBIF.org 2 May 2019].
- Central Michigan University Herbarium. 2019. Prairie Fen Biodiversity Project. Occurrence dataset. <https://doi.org/10.15468/jyjqgq> [accessed via GBIF.org 2 May 2019].
- Chandler, M., L. See, K. Copas, A. M. Z. Bonde, B. C. Lopez, F. Danielsen, J. K. Legind, et al. 2017. Contribution of citizen science towards international biodiversity monitoring. *Biological Conservation* 213: 280–294.
- Chapman, A. D. 2005. Uses of primary species-occurrence data. In Report for the Global Biodiversity Information Facility, Copenhagen. Available at <http://www.niobioinformatics.in/books/> [accessed October 2017].
- Cheruvilil, K. S., and P. A. Soranno. 2018. Data-intensive ecological research is catalyzed by open science and team science. *BioScience* 68: 813–822.

- COSEWIC [Committee on the Status of Endangered Wildlife in Canada]. 2014. Poweshiek skipperling (*Oarisma poweshiek*): COSEWIC assessment and status report 2014. Available at <https://www.canada.ca/en/environment-climate-change/services/species-risk-public-registry/cosewic-assessments-status-reports/poweshiek-skipperling-2014.html> [accessed November 2015].
- Culley, T. M. 2017. The frontier of data discoverability: Why we need to share our data. *Applications in Plant Sciences* 5: 1700111.
- Darwin Core Maintenance Group. 2014. Darwin Core, Biodiversity Information Standards (TDWG). Available at Zenodo repository. <https://doi.org/10.5281/zenodo.592792> [published 8 November 2014; accessed 10 May 2019].
- Daubenmire, R. 1959. A canopy-coverage method of vegetational analysis. *Northwest Science* 33: 43–64.
- Garcia-Rosello, E., C. Guisande, A. Manjarres-Hernandez, J. Gonzalez-Dacosta, J. Heine, P. Pelayo-Villamil, L. Gonzalez-Vilas, et al. 2015. Can we derive macroecological patterns from primary Global Biodiversity Information Facility data? *Global Ecology and Biogeography* 3: 335–347.
- Giovanni, R., J. Wiczorek, D. Bloom, R. Guralnick, S. Blum, M. Do, T. Robertson, and D. Vieglais. 2012. Darwin Core: An evolving community-developed biodiversity data standard. *PLoS ONE* 7: e29715.
- Great Lakes Phragmites Collaborative. 2017. Phragmites Adaptive Management Framework. Great Lakes Commission. Website <https://www.greatlakesphragmites.net/pamf/> [accessed 6 December 2019].
- Gries, C., E. Gilbert, and N. Franz. 2014. Symbiota – A virtual platform for creating voucher-based biodiversity information communities. *Biodiversity Data Journal* 2: e1114.
- Griffin, P. C., J. Khadake, K. S. Lemay, S. Orchard, S. E. Lewis, A. Pask, B. Pope, et al. 2018. Best practice data life cycle approaches for the life sciences [version 2; referees: 1 approved, 1 approved with reservations]. *F1000 Research* 6: 1618.
- Guralnick, R., and A. Hill. 2009. Biodiversity informatics: Automated approaches for documenting global biodiversity patterns and processes. *Bioinformatics* 25: 421–428.
- Guralnick, R., R. Walls, and W. Jetz. 2018. Humboldt Core – Toward a standardized capture of biological inventories for biodiversity monitoring, modeling and assessment. *Ecography* 41: 713–725.
- Hackett, R. A., M. J. Monfils, and A. K. Monfils. 2016. Evaluating a sampling protocol for assessing plant diversity in prairie fens. *Wetlands Ecology and Management* 24: 609–622.
- Hampton, S. E., C. A. Strasser, J. J. Tewksbury, W. K. Gram, A. E. Budden, A. L. Batcheller, C. S. Duke, and J. H. Porter. 2013. Big data and the future of ecology. *Frontiers in Ecology and the Environment* 11: 156–162.
- Hampton, S. E., S. S. Anderson, S. C. Bagby, C. Gries, X. Han, E. M. Hart, M. B. Jones, et al. 2015. The Tao of open science for ecology. *Ecosphere* 6: 120.
- Hardisty, A. R., W. K. Michener, D. Agosti, E. A. García, L. Bastin, L. Belbin, A. Bowser, et al. 2019. The Bari Manifesto: An interoperability framework for essential biodiversity variables. *Ecological Informatics* 49: 22–31.
- Heberling, J. M., and B. L. Isaac. 2018. iNaturalist as a tool to expand the research value of museum specimens. *Applications in Plant Sciences* 6: e1193.
- Heidorn, P. B. 2008. Shedding light on the dark data in the long tail of science. *Library Trends* 57: 280–299.
- Hunt, V. M., S. K. Jacobi, M. G. Knutson, E. V. Lonsdorf, S. Papon, and J. Zorn. 2015. A data management system for long-term natural resource monitoring and management projects with multiple cooperators. *Wildlife Society Bulletin* 39: 464–471.
- Hunt, V. M., S. K. Jacobi, J. J. Gannon, J. E. Zorn, C. T. Moore, and E. V. Lonsdorf. 2016. A decision support tool for adaptive management of native prairie ecosystems. *Interfaces* 46: 334–344.
- LaFrankie, J. V., and A. I. Chua. 2015. Application of digital field photographs as documents for tropical plant inventory. *Applications in Plant Sciences* 3: 1400116.
- Lewis, K. P., E. Vander Wal, and D. A. Fifield. 2018. Wildlife biology, big data, and reproducible research. *Wildlife Society Bulletin* 42: 172–179.
- MacLean, H. J., M. E. Nielsen, J. G. Kingsolver, and L. B. Buckley. 2018. Using museum specimens to track morphological shifts through climate change. *Philosophical Transactions of the Royal Society B* 374: 20170404.
- Maya-Lastra, C. A. 2016. ColectoR, a digital field notebook for voucher specimen collection for smartphones. *Applications in Plant Sciences* 4(7): 1600035.
- Michigan Natural Features Inventory. 2011. Biotics 4 – Michigan's Natural Heritage Database. Lansing, Michigan, USA. [https://msu.edu/course/rd/415/EIA\\_GIS/Tri%20County/biot\\_p.htm](https://msu.edu/course/rd/415/EIA_GIS/Tri%20County/biot_p.htm) [accessed 14 November 2011].
- Mons, B., C. Neylon, J. Velterop, and M. Dumontier. 2017. Cloudy, increasingly FAIR: Revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Information Services & Use* 37: 49–56.
- Morrison, S. A., T. S. Sillett, W. C. Funk, C. K. Ghalambor, and T. C. Rick. 2017. Equipping the 22nd-century historical ecologist. *Trends in Ecology & Evolution* 32: 578–588.
- Parr, C. S., R. Guralnick, N. Cellinese, and R. D. M. Page. 2012. Evolutionary informatics: Unifying knowledge about the diversity of life. *Trends in Ecology & Evolution* 27: 94–103.
- Pogue, C. D., M. J. Monfils, D. L. Cuthrell, R. A. Hackett, R. Zionce, and A. K. Monfils. 2019. Local and landscape level variables related to Poweshiek skipperling presence in Michigan, USA, prairie fens. *Journal of Fish and Wildlife Management* <https://www.fwspubs.org/doi/abs/10.3996/122018-JFWM-117>.
- Polk, M. 2015. Transdisciplinary co-production: Designing and testing a transdisciplinary research framework for societal problem solving. *Futures* 65: 110–122.
- Powers, S. M., and S. E. Hampton. 2019. Open science, reproducibility, and transparency in ecology. *Ecological Applications* 29: e01822.
- Reichman, O. J., M. B. Jones, and M. P. Schildhauer. 2011. Challenges and opportunities of open data in ecology. *Science* 331: 703–705.
- Rich, T., M. Rabane, M. Fasham, F. McMeechan, and D. Dobson. 2005. Ground and shrub vegetation. In D. Hill, M. Fasham, G. Tucker, M. Shewry, and P. Shaw [eds.], *Handbook of biodiversity methods: Survey, evaluation and monitoring*, 201–221. Cambridge University Press, Cambridge, United Kingdom.
- Robertson, T., M. Döring, R. Guralnick, D. Bloom, J. Wiczorek, K. Braak, J. Otegui, et al. 2014. The GBIF integrated publishing toolkit: Facilitating the efficient publishing of biodiversity data on the internet. *PLoS ONE* 9: e102623.
- SARA [Species at Risk Act, Revised Statutes of Canada]. 2002. Species at Risk Act, S.C. 2002, c. 29. Available at <http://laws-lois.justice.gc.ca/PDF/S-15.3.pdf> [accessed November 2015].
- Schindel, D. E., and J. A. Cook. 2018. The next generation of natural history collections. *PLoS Biology* 16: e2006125.
- Selby, G. 2005. Status assessment and conservation guidelines: Poweshiek skipperling, *Oarisma poweshiek* (Parker) (Lepidoptera: Hesperidae). Prepared for Twin Cities Field Office, U.S. Fish and Wildlife Service, Bloomington, Minnesota, USA. <https://doi.org/10.3996/052015-jfw-049.s8> [accessed 24 May 2018].
- Smith, B., and C. C. Chinnappa. 2015. Plant collection, identification, and herbarium procedures. In E. Yeung, C. Stasolla, M. Summer, and B. Huang [eds.], *Plant microtechniques and protocols*, 541–572. Springer, Cham, Switzerland.
- Soltis, P. S., G. Nelson, and S. A. James. 2018. Green digitization: Online botanical collections data answering real-world questions. *Applications in Plant Sciences* 6: e1028.
- Sutter, R. D., S. B. Wainscott, J. R. Boetsch, C. J. Palmer, and D. J. Rugg. 2015. Practical guidance for integrating data management into long-term ecological monitoring projects. *Wildlife Society Bulletin* 39: 451–463.
- Thiers, B., A. K. Monfils, J. Zaspel, E. Ellwood, A. Bentley, K. Levan, J. Bates, et al. 2019. Extending U.S. biodiversity collections to promote research and education. National Science Foundation, Washington, D.C., USA.
- Turney, S., E. R. Cameron, C. A. Cloutier, and C. M. Buddle. 2015. Non-repeatable science: Assessing the frequency of voucher specimen deposition reveals that most arthropod research cannot be verified. *PeerJ* 3: e1168.
- Vellend, M., C. D. Brown, H. M. Kharouba, J. L. McCune, and I. H. Myers-Smith. 2013. Historical ecology: Using unconventional data sources to test for effects of global environmental change. *American Journal of Botany* 100: 1294–1305.

- Wieczorek, J., D. Bloom, R. Guralnick, S. Blum, M. Döring, R. Giovanni, T. Robertson, and D. Vieglais. 2012. Darwin Core: An evolving community-developed biodiversity data standard. *PLoS ONE* 7(1): e29715.
- Wilkinson, M. D. 2016. Comment: The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3: 160018.
- Williams, B. K., and E. D. Brown. 2012. Adaptive management: The U.S. Department of the Interior applications guide. Adaptive Management Working Group, U.S. Department of the Interior, Washington, D.C., USA. Website <https://www2.usgs.gov/sdc/doc/DOI-%20Adaptive%20ManagementTechGuide.pdf> [accessed 19 March 2018].
- Yost, J. M., P. W. Sweeney, E. Gilbert, G. Nelson, R. Guralnick, A. S. Gallinat, E. R. Ellwood, et al. 2018. Digitization protocol for scoring reproductive phenology from herbarium specimens of seed plants. *Applications in Plant Sciences* 6: e1022.