DOCUMENTING THE OCCURRENCE THROUGH SPACE & TIME OF AQUATIC NON-INDIGENOUS FISH, MOLLUSKS, ALGAE, & PLANTS THREATENING NORTH AMERICA'S GREAT LAKES

North America's Great Lakes (Erie, Huron, Michigan, Ontario, and Superior) hold 21% of Earth's fresh water, and are an unparalleled natural resource of national security concern to both the United States and Canada. Just last year, on Sept 7, 2012, the two countries renewed the **Great Lakes Water Quality Agreement** meant to address serious environmental issues affecting the region. This bi-national treaty was first signed in 1972, again in 1987, and now has been updated once more with a focus on 21st Century threats. The overall mission of the Agreement is "to restore and maintain the chemical, physical and biological integrity of the waters of the Great Lakes." Among the key biological aspects of the treaty are: 1) Preventing environmental threats before they turn into actual problems; 2) Developing conservation strategies to protect native species and restore habitat; and 3) Preventing the introduction and spread of invasive species (http://www.epa.gov/glnpo/glwqa/).

The threat to the environmental health of the Great Lakes by invasive species, especially those considered to be non-indigenous to the region, has been well documented. Nearly every public boat launch in MI, WI, and other Great Lakes states has warning signs posted to educate citizens about the threat that species such as zebra mussel, Eurasian milfoil, and Asian carp pose to the ecological balance of the local waterways. These 'poster children' of aquatic Great Lakes invasives, of course, represent only the tip of the iceberg. Biologists today have documented more than 180 non-indigenous species known to be reproducing in the Great Lakes Basin **(Fig 1 below)**, i.e., the five primary lakes plus connecting channels and water bodies within their respective drainages (Mills et al. 1993, Ricciardi 2001, Ricciardi 2006). In most cases, the occurrence of these organisms within the region is well documented by voucher specimens held within herbaria and natural history museums – many of them small and with a strong regional focus. Unfortunately, these specimens, more often than not, are generally accessible only to a limited community of local scientists, are appreciated by only a few specialists who rarely consider species outside of their area of taxonomic expertise, and are inefficiently used as a primary source of data for constructing databases and other resources designed to track and monitor the spread of invasives.

We propose to form a Thematic Collections Network targeting the digitization of non-indigenous species and their congeneric taxa of: **vascular plants** (ferns & angiosperms), **fish** (agnaths & bony), **green algae** (charaphytes), and **mollusks** (bivalves & gastropods) documented to be present in the Great Lakes Basin. Because the introduction and spread of these species, their close relatives, and hybrids into the region is known to have occurred mostly from areas in North America outside of the Basin, our effort will include specimens collected from throughout North America. We will also collaborate with our Canadian colleagues who maintain the national http://*Canadensys.net* resource. To the best of our knowledge, this level of bi-national, cross-collection collaboration has rarely been attempted. The unique methods of three-dimensional liquid preservation of whole vertebrates by ichthyologists compared to 3D preservation of mollusk shells by malacologists, and mostly two-dimensional preservation of pressed plant specimens by botanists, will present a challenge to our team, but the value added by 'reaching across the museum' should be considerable, and will develop into an unprecedented data resource.

Target Taxa and Collections

Table 1 lists all genera that will serve as the focus of this effort. Each genus includes at least one species that is a Great Lakes non-indigenous taxon or on a "watchlist", meaning that it has not arrived in the Great Lakes Basin yet, but has the potential to do so, especially in light of human activity and climate change. These lists were generated by conducting a query in the database of **GLANSIS** – the Great Lakes Aquatic Nonindigenous Species Information System, http://www.glerl.noaa.gov – maintained by the National Oceanic and Atmospheric Administration (NOAA). Only aquatic non-indigenous species that are established in the Great Lakes Basin below the high water mark *are* included in the GLANSIS database. Species which have invaded only inland lakes within the Basin, but not the Great Lakes, *are not* included.

Within Table 1 we make note of 12 plant families also being targeted by the currently funded NSF ADBC "Tri-trophic" digitization project. Five of the 20 herbaria within our proposed network are also members of that TCN, but most are being funded only to digitize specimens collected within their state as a level one priority. This relatively minimal overlap was taken into account when calculating total specimens that will be targeted for digitization under this effort.

Impact of Non-indigenous Invasive Species

We emphasize our use of the term 'non-indigenous', rather than alien, foreign, exotic, or naturalized, as well as overlap with the term invasive. The latter, in particular, has different meanings to different constituents. While it is true that many aggressive 'invasive' species in the Great Lakes are not historically native to North America, others are (e.g., rainbow trout). Regardless of provenance, too many of these fish, algae, mollusks, and plants have wreaked havoc on our environment and our economy. Aquatic plants affect property values for homeowners, sea lamprey caused a near collapse of the fishing industry, zebra mussels force shipping companies to expend millions of dollars on preventative measures and maintenance, and algae blooms drain local communities of valuable tourist dollars.

Great Lakes invasives were even a topic of the 2012 Presidential campaign platforms and debate, with President Obama already having invested ca. $1billion in the **Great Lakes Restoration Initiative** (http://greatlakesrestoration.us), and GOP nominee Mitt Romney critically stating that "I am deeply concerned about the threat posed to the lakes by invasive species (such as Asian Carp) . . . I am outraged that, five years after Congress ordered the Army Corps of Engineers to identify a solution, we are still years away from a recommendation." (Milwaukee Journal Sentinel: Oct 11, 2012). Regardless of political persuasion, >78% of citizens in Wisconsin say that they are concerned about restoring and protecting the Great Lakes. The information derived from specimens held within our natural history collections undoubtedly has a role to play in contributing to these restoration and protective efforts.

PRIMARY OBJECTIVES

(1) To digitize >1.73 million specimens from 2,550 species in 101 genera
(2) To harvest & organize significant data associated with collections
(3) To share specimen images and data with the greater scientific community
(4) To promote cross-collection efforts in the study of aquatic/invasive species
(5) To promote the use of collections data by educators and the public

---

OBJECTIVE ONE: To digitize >1.73 million specimens from 2,550 species in 101 genera

PREMISE: Digitized specimens of Great Lakes non-indigenous species and their congeners will allow for more accurate identification of invasive species and hybrids from their non-invasive relatives by a wider audience of end users.

One of the greatest challenges to documenting the presence and spread of non-indigenous and potentially invasive or harmful species is being able to identify them. Examples are numerous for animals (e.g., gypsy moth – tent caterpillar larvae) and plants (e.g., American bittersweet – Asian bittersweet) that are notoriously difficult to distinguish unless compared side-by-side. Documenting diversity, morphology, and phenotypic variation is, of course, one of the primary functions of natural history museum collections, and so it is entirely logical and necessary for specimen-based museum databases to include images. Common sense informs us that specimen morphology, not only label data, is critically important when making identification. When considering the potential impact of aquatic invasives, moreover, it is imperative. Both non-indigenous carp and mussels offer obvious examples of taxa for which galleries of specimen images could serve to assist biologists and the public in making proper identification of species.

Table 1. Genera (and their corresponding family) targeted by this effort. Lists were generated by querying NOAA's GLANSIS database http://www.glerl.noaa.gov. Genera with species not yet in the Great Lakes Basin, but on the 'Watchlist' are in blue. **Our TCN will target ca. 2,550 species in 101 genera** (parantheses are numbers of North American species within each genus).

PLANTS		PLANTS (continued)		FISH (continued)	
Genus (2147)	**Family**	**Genus (2147)**	**Family**	**Genus (290)**	**Family**
Agrostis (36)	Poaceae*	***Potamogeton (63)***	Potamogetonaceae	***Morone (4)***	Moronidae
Alnus (14)	Betulaceae	***Puccinellia (31)***	Poaceae*	***Neogobius (1)***	Gobiidae
Alopecurus (16)	Poaceae*	***Rorippa (28)***	Brassicaceae	***Notropis (91)***	Cyprinidae
Butomus (1)	Butomaceae	***Rumex (55)***	Polygonaceae*	***Noturus (29)***	Ictaluridae
Cabomba (4)	Cabombaceae	***Salix (170)***	Salicaceae*	***Oncorhynchus (11)***	Salmonidae
Carex (593)	Cyperaceae*	***Solanum (104)***	Solanaceae*	***Osmerus (1)***	Osmeridae
Chenopodium (51)	Chenopodiaceae*	***Solidago (77)***	Asteraceae*	***Perca (1)***	Percidae
Cirsium (95)	Asteraceae*	***Sparganium (10)***	Sparganiaceae	***Perccottus (1)***	Odontobutidae
Conium (1)	Apiaceae*	***Trapa (2)***	Trapaceae	***Petromyzon (1)***	Petromyzontida
Echinochloa (20)	Poaceae*	***Typha (4)***	Typhaceae	***Phenacobius (5)***	Cyprinidae
Egeria (1)	Hydrocharitaceae	***Veronica (34)***	Scrophulariaceae*	***Phoxinus (6)***	Cyprinidae
Eichhornia (4)	Pontederiaceae			***Proterorhinus (1)***	Gobiidae
Epilobium (45)	Onagraceae			***Rutilus (1)***	Cyprinidae
Frangula (8)	Rhamnaceae*	**FISH**		***Salmo (2)***	Salmonidae
Glyceria (18)	Poaceae*	**Genus (290)**	**Family**	***Scardinius (1)***	Cyprinidae
Hydrilla (1)	Hydrocharitaceae	***Alburnus (1)***	Cyprinidae		
Hydrocharis (1)	Hydrocharitaceae	***Alosa (6)***	Clupeidae		
Hygrophila (6)	Acanthaceae	***Apeltes (1)***	Gasterosteidae	**MOLLUSKS**	
Impatiens (11)	Balsaminaceae	***Atherina (1)***	Atherinidae	**Genus (113)**	**Family**
Iris (52)	Iridaceae	***Babka (1)***	Gobiidae	***Bithynia (1)***	Bithyniidae
Juncus (123)	Juncaceae	***Benthophilus (1)***	Gobiidae	***Cipangopaludina (2)***	Viviparidae
Lupinus (165)	Fabaceae*	***Carassius (1)***	Cyprinidae	***Corbicula (1)***	Corbiculidae
Lycopus (10)	Lamiaceae*	***Channa (2)***	Channidae	***Dreissena (2)***	Dreissenidae
Lysimachia (42)	Primulaceae	***Clupeonella (1)***	Clupeidae	***Elimia (50)***	Pleuroceridae
Lythrum (13)	Lythraceae	***Cottus (33)***	Cottidae	***Gillia (1)***	Hydrobiidae
Marsilea (12)	Marsileaceae	***Ctenopharyngodon (1)***	Cyprinidae	***Lasmigona (9)***	Unionidae
Mentha (13)	Lamiaceae*	***Cyprinella (30)***	Cyprinidae	***Monodacna(1)***	Cardiidae
Myosotis (12)	Boraginaceae	***Cyprinus (1)***	Cyprinidae	***Pisidium (13)***	Sphaeriidae
Myosoton (1)	Caryophyllaceae	***Enneacanthus (3)***	Centrarchidae	***Potamopyrgus (1)***	Hydrobiidae
Myriophyllum (14)	Haloragaceae	***Esox (4)***	Esocidae	***Radix (1)***	Lymnaeidae
Najas (8)	Najadaceae	***Gambusia (24)***	Poeciliidae	***Sphaerium (20)***	Pisidiidae
Nasturtium (5)	Brassicaceae	***Gymnocephalus (1)***	Percidae	***Valvata (8)***	Valvatidae
Nitellopsis (3)	Characeae (algae)	***Hypophthalmichthys (2)***	Cyprinidae	***Viviparus (3)***	Viviparidae
Nymphoides (7)	Menyanthaceae	***Knipowitschia (1)***	Gobiidae		
Pistia (1)	Araceae	***Lepisosteus (4)***	Lepisosteidae		
Pluchea (11)	Asteraceae*	***Lepomis (13)***	Centrarchidae		
Poa (96)	Poaceae*	***Leuciscus (1)***	Cyprinidae	*** = Plant family originally targeted by "Tri-trophic" TCN**	
Polygonum (80)	Polygonaceae*	***Misgurnus (1)***	Cobitidae		

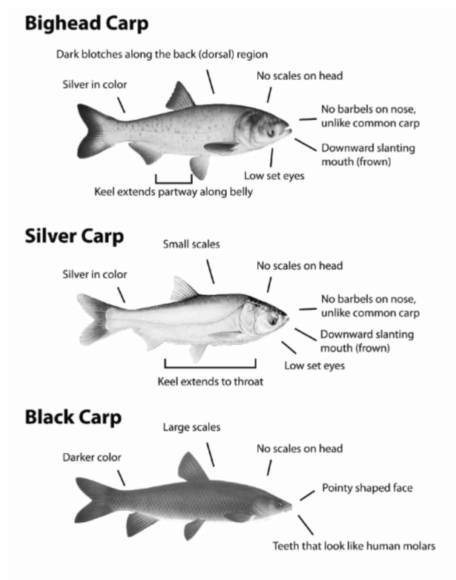

Figure 2 (right) is taken from a poster used to help citizens in determining whether they have spotted an Asian carp in a local body of water. The poster emphasizes "Make sure you have the right carp! . . . If in doubt, call your state personnel!" and "Look for hybrids, fish that might exhibit characteristics of both the silver and bighead species." Our proposal includes digitization of all species within five genera of Cyprinidae (carp and minnows), and the resulting images made available through our online portal will be an invaluable resource for the public wishing to make their own identifications rather that depending on busy "state personnel" to verify sightings. The poster's mention of potentially aggressive hybrids also indicates the value in being able to quickly examine images of thousands of individuals curated in scientific collections.

Identification of mussel species can be equally vexing, especially for non-specialists. An example is provided by the genus *Dreissena*, which contains just seven species (Rosenberg and Ludyanskiy, 1994). Among these are two infamous taxa introduced into North America - the notorious zebra mussel (*D. polymorpha*) and the more inconspicuous quagga mussel (*D. rostriformis bugensis*). The two can be challenging to recognize from each other and from non-invasive native species. DNA barcoding methods are being used more routinely to verify the identity of mussel species, but the two dreissenids can, in fact, be differentiated by morphological differences of the shell. When examined side by side in **Figure 3 (left)**, the zebra mussel has a conspicuous angle between the ventral and dorsal surfaces, whereas the quagga has a rounded carina (May and Marsden, 1992). The ventral side of the zebra mussel shell is flattened, but the quagga is convex. For the most part quaggas are rounder in shape than zebras, which are more triangular. Color patterns vary widely with black, cream, or white bands. These subtle variations are best appreciated when many specimens are examined at the same time. Again, we propose to make available hundreds of images of *Dreissena* specimens from the most important mollusk collections in the Great Lakes region so that a wide audience of end users will be provided an identification tool never before at their disposal.

OBJECTIVE TWO: To harvest & organize significant data associated with collections

PREMISE: Data derived from digitized specimens of Great Lakes non-indigenous species and their congeners will help biologists to track, monitor, and predict the spread of invasives through space and time, especially in the face of a more rapidly changing climate in the upper Midwest.

Specimens held within herbaria and zoology museums are unique forms of data in that they document an organism's distribution in both historical time and widespread geographic space. For studies of non-indigenous Great Lakes species, both variables are of paramount importance, especially with regard to understanding past invasions and using these data to model future expansions. An appropriate regional example of how our networks' metadata could be used is offered in a study of wetland plants by Delisle et al, (2003) in which six invasive species were compared to five native, non-expanding hydrophytes. The Saint Lawrence Seaway, an entry point into the Great Lakes, was the area of interest, and both the temporal and spatial distribution of herbarium specimens from seven Canadian museums was considered. They plotted the cumulative number of collection locations against time to construct invasion curves, and demonstrated that deviation of invasive species' collections from the native species invasion curve can be used to describe the spread of invasive species (see Fig 4), but also warn that sampling biases associated museum specimens must be taken into account in order to delineate periods of invasiveness for non-indigenous species.

Freshwater mussels provide a second example – not of one only documenting a historical invasion, but one concerning the timing and place of a recent point of entry as well. First discovered in the USA in 1988, the zebra mussel (*Dreissena polymorpha*) just a few years later, in 1990, managed to invade all of the Great Lakes (Hebert *et al.* 1989, Benson *et al.* 2007). In contrast, the quagga mussel (*D. rostriformis bugensis*) was first collected in 1989, but has only been slowly colonizing the lower Great Lakes (Mills *et al.* 1993, 1999; Benson *et al.* 2007). Museum specimens document this invasion. Its appearance in Lake Michigan was recorded a decade later (Mills *et al.* 1993, Nalepa *et al.* 2001), but fears began to grow that it would eventually enter Lake Superior. Four years ago Grigorovich et al. (2008) provided the first genetic and morphological evidence documenting the presence of both dreissenid mussels in a major river-embayment of Lake Superior between Duluth, MN and Superior, WI from collections made in 2005. The largest quagga mussel in the collection was 26.5 mm long, suggesting that the initial introduction to the Lake Superior Harbor occurred no later than 2003. These baseline data, documented by voucher specimens, are invaluable for establishing the spatial and temporal points of invasion.

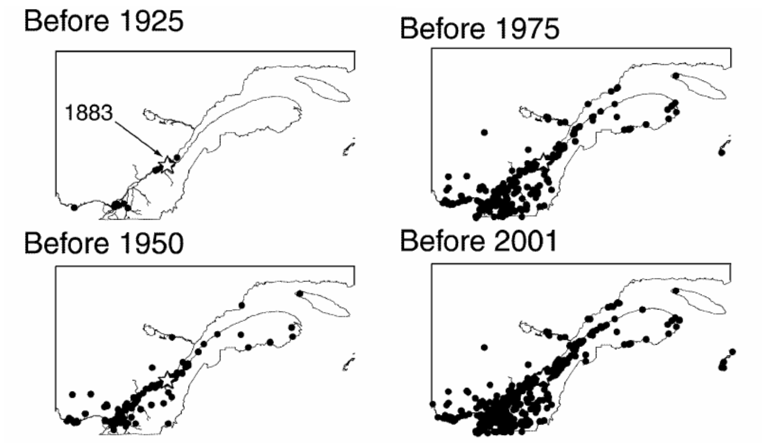

Fig 4. Temporal & spatial invasion of the St Lawrence by *Lythrum salicaria* [purple loosestrife] using herbarium specimen metadata (Delisle et al, 2003).

OBJECTIVE THREE: To share images & data with the greater scientific community

PREMISE: National databases (USA & Canada) and projects documenting the location of Great Lakes non-indigenous species are incomplete, and will be significantly improved with the addition of specimen images and data from existing natural history collections made widely available by this TCN.

Wetland plants and invasive species, in general, and aquatic invasives, in particular, are the focus of several online efforts to document, describe, provide *in situ* photos, and monitor the distribution of taxa. For example, the USDA's National Invasive Species Information Center (NISIC) describes itself as a gateway to invasive species information covering federal, state, local, & international sources http://www.invasivespeciesinfo.gov. The site is information-rich and provides links to other online resources that target particular regions of the country, taxonomic groups, or ecological communities.

The GLANSIS database maintained by NOAA was described earlier. Unfortunately, that database provides little information other than lists and tables of non-indigenous species, without distribution maps or information about relative abundance based on collections. The managers have expressed enthusiastic support for this proposal and asked for our data to be shared with them.

A third outstanding resource is the US Geological Survey's Nonindigenous Aquatic Species (NAS) information resource at http://nas.er.usgs.gov. Based in Gainesville, this site was established as a central repository for biogeographic accounts of introduced aquatic species in the USA. The program provides scientific reports, online/realtime queries, spatial data sets, regional contact lists, and general information. Last year, unfortunately, visitors to NAS' site were met with the following disclaimer: ***The NAS site will no longer serve data or track aquatic plants or marine invertebrates. Plants are being discontinued due to budgets cuts.*** Today the site only tracks select vertebrate and invertebrate groups.

Other projects that focus on tracking invasive species are numerous, but many do not take full advantage of primary museum collections. Those that do, e.g., GBIF, BISON, Biota of North America (http://www.bonap.org/), and the US Army Corp of Engineer's new wetland plants portal are only as complete as the data available to or discovered by them. More data being shared is better for everyone!

Clearly a cross collections project like this requires careful consideration of data exchange with other existing groups. Having all collection records for invasive species in the Great Lakes region available in one place has major advantages for addressing the research and management questions posed in this proposal. However, the data gathered here will be very useful in other contexts as well and need to be shared with the appropriate collections networks. Symbiota has been chosen as the means for publishing the data digitized here and already has protocols established to synchronize data with the large national and international collections networks (GBIF http://www.gbif.org/ and iDigBio Portal http://portal.idigbio.org) and Webservices for a generic Darwin Core Data Package exchange are available. However, interchange protocols will have to be explored with the smaller thematic networks to provide new data to them and harvest pertinent existing data as these continue alongside the project proposed here. Specifically, VertNet (http://vertnet.org), Musselp (http://mussel-project.uwsp.edu/), FishBase **(**http://www.fishbase.org**)**, Canadensys (http://www.canadensys.net/), vPlants (http://www.vplants.org/), WisFlora (http://www.botany.wisc.edu/wisflora/), and other projects related to our theme are information-rich and valuable. Some of these networks are already Symbiota nodes and automatic exchange will be easily achievable. .

OBJECTIVE FOUR: To promote cross-collection efforts in the study of invasive species

PREMISE: Points of access, migration routes, lag times, and speed of colonization may be correlated among different lineages of non-indigenous organisms now established in the Great Lakes. Discovery of correlated patterns could help to prevent future invasions.

Correlation studies of different plant species have documented the phenomenon of lag phases in plant invasions, whereby a particular non-indigenous species appears to sit 'dormant' for a long period of time before undergoing rapid population growth and range expansion. Whether or not this phenomenon occurs within other organisms or follows the same pattern is unclear. Larkin et al. (2012) provide an example for the Great Lakes region. Mining data from the Wisconsin State Herbarium's WISFlora database of >350,000 specimen records, they found sufficient data to test for lags in 76 northern Wisconsin species, 90 for southern Wisconsin, and 91 for the southern Lake Michigan region; see Figure 6. Lags were identified in 77% (197) of these 257 datasets and ranged from 3–140 years with a mean of 47.3 ± 34.6 years. The importance of identifying lags and possible correlations in lag time among unrelated organisms within a community has relevance for the practice of early eradication since it can be unclear whether or not a non-indigenous species is benign or a potentially harmful ''sleeper weed.'' Statistical studies such as this are only objective and significant if large amounts of specimen data, accurately identified, georeferenced, and with precise dates of collection, are available. Our network will provide these data at an unprecedented scale.

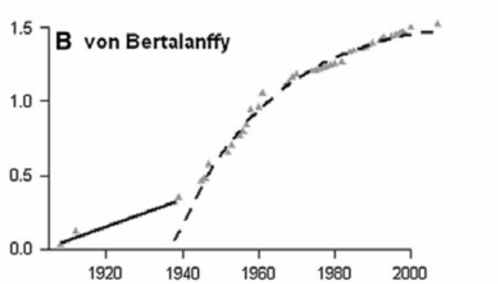

Fig 6. Invasive lag model fit for *Frangula alnus* [buckthorn] in the S Lake Michigan region of WI & IL for two-piece models with von Bertalanffy increase phases (Larkin et al., 2012).

OBJECTIVE FIVE: To promote the use of collections data by educators and the public

PREMISE: Collections-based outreach and education by herbaria and museums in both formal and informal education settings can increase the depth of public understanding about the ecological and economic threats posed by a wide diversity of non-indigenous fish, mollusks, plants, and algae.

Education and outreach in this proposal aim at increasing knowledge, understanding, and use of aquatic invasive collections among land managers, educators, and the general public. Our outreach and education efforts address three goals: (1) Providing online tools for identifying invasive species in the Great Lakes region and using collections data; (2) Providing high-quality data on distribution of invasives, and training our professional user base to utilize the resource; and (3) Building scientific capacity through education of K–12 teachers and development of circulating educational resources. We have been in conversation with leaders in invasive species outreach at U of MN Sea Grant (D.A. Jensen), NOAA Great Lakes Environmental Research Laboratory (R. Sturtevant), WI DNR Endangered Resources (S.K. Kearns), and the U of GA Center for Invasive Species and Ecosystem Health (G.K. Douce) to identify needs within the community that our portal could address. Resoundingly, these groups identify three needs that our data could address: helping people identify (1) what the threats are, (2) what to look for, and (3) where to look. Our outreach plans are consequently designed around these needs, and will be implemented in collaboration with these groups to ensure that we reach the relevant professional, public, and educational audiences. All components will be evaluated in-house, through collaboration with education departments at The Morton Arboretum and the Field Museum.

Goal 1: Provide online tools Static content is often both the primary target of visitation to a data portal and the magnet that draws users into the portal. For invasive species, substantial information is already available online on the basic biology of aquatics (e.g., http://www.miseagrant.umich.edu; http://cisr.ucr.edu/invasive_species_faqs.html; and numerous others. However, resource managers working with invasive species want detailed identification guides that include paired photos of diagnostic identification features in both the invasive species and similar species. We are budgeting time in year 1 for creating these pages in our Symbiota portal (at MOR) based directly on collection images, and in years 2-3 for external reviewing. Priorities for writing will be developed in collaboration with our entire advisory group and with our collaborators in the field. Colleagues at U of MN Sea Grant and WI DNR are eager to work with us in publicizing these identification resources and either republishing or linking to them. This goal represents several challenges for which Symbiota is uniquely positioned. Valuable species descriptions are the purview of the Encyclopedia of Life (http://eol.org) and Symbiota has developed protocols to synchronize these data efficiently with EOL – providing descriptions and using descriptions already in EOL. Another challenge unique to this cross kingdom identification system arises due to the lack of globally unique identifiers for taxonomic names, which has led to non-unique taxon names across different kingdoms. Character and character states are also not unique (i.e. the same term is used for different concepts in different taxonomic groups) and need to be treated within the taxonomic context to successfully handle this challenging project. Symbiota is built on principles incorporating taxonomic hierarchy and inheritance comparable to modern object oriented programming. Hence, not only can an overall keying system for all invasive species be built, Symbiota provides unique features to ease this task. This will provide valuable additions to existing more focused identification applications (e.g. WI fishID http://www.seagrant.wisc.edu/home/Default.aspx?tabid=604). ***Evaluation.*** All materials will be workshopped for readability and utility by three target audiences: a focus group of public garden users; a group of K–12 teachers who already incorporate invasives in their teaching; and land managers engaged in invasive species control. We will be evaluating online materials through in-person evaluation as well as by an online survey / comment system linked from the website.

Goal 2: Training professional users Land managers and ecologists are among the most active users of our herbaria and museums, and portals that monitor invasive species distributions depend on both observations and museum records. As part of this project, we will be providing data to GLANSIS; and EDDMapS (http://www.eddmaps.org/), a widely used mapping database and application, optimized for early detection of emerging invasive species. Both resources are eager to utilize our data to fill in gaps in sampling (for example, Illinois is missing altogether in the distribution of several aquatic invaders for EDDMapS due to a lack of observational data) and provide baseline data for observations going forward. Our outreach coordinator will assist with data-checking and data export; both EDDMapS and GLANSIS have agreed to import these data and make them available through their data portal as resources allow.

Whereas many resource managers provide specimens to regional museums, many others lack the experience, confidence, and/or time to work as actively with museums as they might like. We are planning four workshops for regional land managers and ecologists aimed at providing experience in using our portal. These workshops will be coordinated through regional meetings of professionals – we will be working with K. Kearns and D. Jensen to organize one for the Upper Midwest Invasive Species Conference and with M. Murray to organize one for the UW Madison Arboretum Native Gardening Conference, and we have budgeted to bring Paul Skawinski, Central WI Regional AIS Education Specialist and author of *Aquatic Plants of the Upper Midwest*, to teach one of these workshops in collaboration. ***Evaluation:*** We will be identifying training needs ahead of these workshops through online survey directed at regional participants, and effectiveness of the workshops through online survey and follow-up interviews conducted by staff not directly involved in the project.

Goal 3: Training students and teachers. Next generation teaching standards emphasize the importance of project-based teaching and the development and use of models in combination with analysis of data (http://www.nextgenscience.org). Yet, the greatest barriers to teaching biodiversity is neither interest nor willingness, but knowledge and confidence in the material (Schwartz et al. 2004; Griffith and Brem, 2004; Silverstein et al., 2009). The Morton Arboretum herbarium has successfully partnered with numerous area school teachers and administrators in the greater Chicago area over the past four years to create curriculum tools that integrate biodiversity science into the classroom through research collaborations (http://tinyurl.com/carexMorph), plant identification tools and lesson plans (http://tinyurl.com/herrickTrees), and other curricula. In this project, we propose to partner with and provide both training and support to K–12 teachers to utilize collections as a tool for teaching, and collections portals such as our as a gateway into studying biodiversity past and present.

In this proposal, we will be working with a cohort of ten teachers in a week-long summer workshop at the Arboretum in year three of the project. As there are already numerous curricula and educational aids in place for invasive species (http://www.weedcenter.org), our efforts will focus on broadening the use of collections in the classroom. Teachers will be selected by application, with the criteria that (1) they come with the support of their administrators to implement what they learn in the workshop and (2) they demonstrate potential room for collections data in their curriculum. All teachers will be provided a modest stipend. Education departments at the Field Museum and The Morton Arboretum work with large and complementary networks of city and suburban schools across the greater Chicago region, so that our candidate pool of teachers will draw from literally of schools. During the workshop, teachers will receive lectures and hands-on training in making and databasing collections (day 1); developing hypotheses and projects that can be addressed with georeferenced collections data (day 2); searching, managing, analyzing and visualizing data (days 2, 3); and understanding the issues of analysis and bias that are inherent to using collections data to map species distributions and test hypotheses (day 3). In the last two days of the week, teachers will work with curriculum materials we have developed to either improve those materials or modify them / develop their own, in collaboration with Co-PI Hipp and the Outreach and Education Coordinator (OEC). The PI and OEC will provide follow-up support to teachers in the form of classroom visits and consultation.

To broaden impact among K–12 students and educators, we have added funding to support the Field Museum's *N. W. Harris Learning Collection* in developing a new Experience Box titled "Invasive Aquatic Species," that will include plant and animal specimens invasive in the region as well as their look-alikes. This customized Experience Box has the potential to reach a wide array of educators. For example, in 2012-13 no fewer than 2,135 items were borrowed from the *N.W. Harris Learning Collection* by the program's 315 active members. The lessons developed for the Activity Guide in this Experience Box would both highlight and support classroom use of the *Aquatic Invasives* web portal developed through the grant. ***Evaluation.*** Teacher and student awareness, interest in, and understanding of how scientists use collections data to understand the distribution, ecology, and evolution of organisms will be assessed using pre- and post surveys and by follow-up interviews.

PROJECT MANAGEMENT

Institutional Participation & Integration of the Collections

The institutions collaborating in our network are listed below, along with estimates of specimen numbers to be digitized by this effort. Herbaria codes are those assigned by *Index Herbariorum* (Thiers, 2012). These institutions come from Canada & the 7 states with sizeable shoreline along the Great Lakes (MN, WI, IL, IN, MI, OH, NY), and represent small teaching colleges, large research universities, public and private museums, and botanical gardens. Four of the 10 largest herbaria in North America (NY, F, MICH, WIS), each with a collection of >1 million specimens, are included and will serve as regional data processing centers. The fish collections held at eight of these institutions, and mollusk collections at six, will be participating as well. Most of the animal collections are curated in zoological museums separate from the institution's herbarium; several participants have commented that this cross-collection collaboration will be a first for their institution. We hope to introduce it as a model for others to follow.

Institution	State	TCN funding?	Plants & Algae	Fish Lots (indivdls)	Mollusk Lots (individls)	Data Center
Univ of WI-Madison (WIS)	WI	T, L, I	60,000	2,600 (20K)	500	WIS
Univ of WI-Steven's Point	WI		15,000	2,200		WIS
Univ of WI-Milwaukee	WI		5,400			WIS
Univ of WI-LaCrosse	WI		7,000			WIS
University of Minnesota	MN	T, L, I	65,000	11,000	1,100	WIS
Field Museum (F / FMNH)	IL	L, I	63,000	7,500 (100K)		F
University of Illinois	IL	T, L, I, M	102,000	30,000	7,800	F
Morton Arboretum	IL		18,000			F
University of Notre Dame	IN		14,000	140	150	F
Butler University	IN		11,000			F
Univ of Michigan (MICH)	MI	T, L, M	48,000	31,660 (490K)	25,000 (390K)	MICH
Michigan State University	MI	L, I	41,000			MICH
MI Small Herbarium Initv	MI		11,000			MICH
Central Michigan Univ	MI		3,000			MICH
Miami University	OH	T	18,000			MICH
Ohio State University	OH	L, I	10,000	28,000	9,000	MICH
Ohio University	OH		4,000			MICH
NY Botanical Garden (NY)	NY	T, L, M	103,000			NY
New York State Museum	NY		39,000			NY
Arizona State Univ	AZ	SYMBIOTA Database/Portal Development				
Université de Montréal	Canada	Data Sharing with *Canadensys.net*				
TOTALS: >637K plants + >681K fish (in >102K lots) + >408K mollusks (in >44K lots) = >1.73M						

T: Tritrophic Plants & Insects, L: Lichens & Bryophytes, I: Arthropod InvertNet, M: Macrofungi

Personnel & Project Administration

Our proposed network of 35 collections at 20 institutions will involve at least 100 individuals, each making an important and unique contribution to the total effort. These include >32 faculty-equivalent Co-PIs and Collaborators, ca. the same number of non-PI curators and museum support staff, four Digitization Project Managers, an Outreach Specialist, two graduate students, and dozens of undergraduate digitizers. The organizational structure of this integrated team consists of the following:

Steering Committee - Administration and oversight of the collaborative network will fall under the

responsibility of a steering committee consisting of the Co-PIs at the four subregional data processing centers (WIS, F, MICH, NY), the four Digitization Managers at those centers, the IT Specialists (Gries. Gilbert, Franz), the Outreach Coordinator (Hipp), *Canadensys* representative, and at least one ichthyologist and one malacologist to represent those collections. This committee will meet three times in Madison: 1) at the beginning of year one to establish policies, set goals for phase one, & review protocols; 2) at the midpoint of the project in year two to review progress, adjust strategy if necessary, and set goals for phase two; and 3) at the end of year three to review accomplishments, verify that goals were met, and address unfinished business. At each of these meetings the committee will attempt to identify other institutions who should be invited into the Network via the ADBC PEN proposal mechanism.
Network Subregional Partners & Curators - The Co-PIs at each of the four subregional data processing centers will work closely with their subregional Co-PI partners and Collaborators (e.g., WIS with UWL, UWM, UWSP, & MIN) to be sure that they receive training, understand goals and policies set forth by the Steering Committee, and achieve predetermined goals. Shortly after the first organizational meeting of the Steering Committee, in Fall 2014, four subregional planning meetings of the partners will take place: one each in Madison, Chicago, Ann Arbor, & New York. A partial List of Participants is provided in the Supplementary Materials of this proposal. Only faculty-equivalent personnel are listed, and these are the individuals who will be responsible for hiring and supervising students, overseeing digitization of their collections, transferring and archiving data, but we wish to recognize that most of these institutions also employ non-faculty collections personnel/curators who often go unrecognized for their efforts in determining, preparing, repairing, and filing specimens. These critically important staff members will be encouraged to attend their respective subregional planning meeting as well.
Graduate & Undergraduate Students – The students employed through this project are not simply a source of inexpensive labor. They represent the future of natural history museum curation, and will be trained in contemporary practices. A Facebook page will be created as a tool for social networking, experience sharing, troubleshooting, and promoting best practices / alternatives among the students and others. Opportunities for advanced training of the graduate students at Wisconsin and Ohio State, in particular, will be made available. For example, the graduate student at UW will participate in the 15week Natural Sciences Museum Studies Course (Zoology405) offered each fall semester.

Workflow

Figure 7 summarizes our workflow strategy. An attempt was made to follow Nelson et al.'s (2012) recommendations for efficient and effective digitization of biological collections. Each institution (= Digitization Center [DC]) has been assigned to one of four Great Lakes subregions, for which there is a single Data Processing Center (DPC). Digitization will follow protocols and workflows well established by existing TCNs. At most of the DCs, student employees will be hired to photograph specimens *in situ*. At this step a skeletal label data record is established with barcode number and the taxonomic name as a minimum, but other information may be added (e.g. state or lake, collector and collection number). Images and the text file will be uploaded to the central server where processing scripts will establish records in the Symbiota portal, link the images and prepare the label information for editing. This last step will involve automated Optical Character Recognition and Natural Language Parsing, both of which have been applied successfully to well typed labels. Handwritten and poorly typed labels may have to be transcribed manually.
The Symbiota Portal. As mentioned throughout the proposal Symbiota (http://symbiota.org) is the software of choice for publishing the data within the context of invasives in the Great Lakes. However, it also supports data entry, label image transcription and editing via a sophisticated data entry interface implementing many data quality controls and access to GEOLocate for geo-referencing. Hence, it will serve as a commonly accessed project workbench at each of the DPCs. Full-time Regional Digitization Managers and one Central Data Manager will be responsible for this process supervising students and collaborating with the general public through Symbiota's crowd sourcing application. Once approved the records are available to download in Darwin Core format to integrate with local databases, however, advanced synchronization support is currently being developed between Symbiota and Specify. Protocols

are already in place to publish project data to the central iDigBio portal, which Symbiota can handle for all participating collections.

Meetings and Training Workshops. A series of workshops and coordination meetings for participants have been built into our proposal (discussed above). These will be important, especially, for those institutions that are currently not part of an existing TCN so that they can quickly adopt best practices sanctioned by iDigBio. **Training, education, and assessment** are critical to the success of this project, and also have been built into the outreach objective (see above). At the same time, we will encourage participants to register for workshops offered through iDigBio and national meetings. Already, several of our co-PIs have attended "Train the Trainers in Georeferencing", and "Citizen Scientist" workshops.

Pre-existing data. Although we have emphasized in this proposal that our project has only **minimal overlap with other efforts** (e.g., Tritrophic TCN [select families only] and VertNet [fish data only; no images]), several of the participating museums already have databases for some of the target taxa specimens. Early in our project management we will ask them to extract those data in Darwin Core format and the project data manager will be responsible for integration into our portal as a means of most efficiently populating our database. Although the specimens will not be imaged, a similar data transfer will take place with Canadensys to enhance our data set of Canadian vascular plant occurrences for these Great Lakes species. In return this project will make its data available to these networks and protocols will be developed to assure optimal synchronizations between all networks during this phase of active digitization. ***Digitization stations***. Most of the herbaria already own digitization stations. We have planned our workflow such that those stations will be freed up in phase 2 or 3 of our digitization effort. The large herbaria (e.g., F and MSC) that lack the standard digitization station will purchase a similar model to be retained by them. The same will be true for the zoological museums. Only the smallest museums will be loaned a temporary-use mobile digitization station. At 115 lb (52 kg), these mobile copy stand kits collapse into a protective watertight case (42"x22"x15") and include everything needed to quickly assemble a digitization station. Video instructions are included. Detailed descriptions of both station types are offered in our individual Budget Justifications. Scheduling and shipping these mobile units will be covered by the four regional Data Processing Centers. At the end of the project they will be available through inter-herbarium loan to other regional institutions who wish to digitize their collections for whatever purpose. Through the use of these mobile units, we feel that we can achieve **maximum efficiency of digitization** and maintain a **low cost per specimen ratio**.

Specimen imaging in situ. This aspect of the project will take the greatest time and involve the greatest number of people. For 2D plant specimens, protocols have been established by other TCNs. RAW files are obtained via a digital camera mounted on a copy stand, white balance is checked regularly, a scale ruler and color card are included with each image. Sixty sheets/hour should easily be attained. Protocols for fish and mollusks will require fine tuning, especially for those preserved in liquid and/or with hundred or more individuals in a 'lot'. Several of the malacologists already have experience with digitization via the mussel project (http://mussel-project.uwsp.edu). The zoologists have agreed, in advance, that they will image all individuals within a lot where there are <10 indiv/lot, but will probably take a "group" photo of lots with higher numbers (e.g., thousands of snails). Mollusks will be imaged dorsally and ventrally. Staging the fish in shallow dishes will be more time consuming than imaging either herbarium sheets or mollusks. In some cases, rubber gaskets will be damaged and require replacement in the process of opening jars of fish. We have adjusted budgets accordingly with these considerations in mind.

Globally unique identifiers. We will use archival quality barcodes generated in a standardized format to insure that each specimen digitized can be located through the use of an identifier that is unique within the project. For most collections these will have a "triplet" format: institutional code, followed by a letter to represent the specific collection, and a barcode number. The exact format will be addressed by the Executive Steering Committee at their first organizational meeting, with approved examples offered up as requirements for participants. Symbiota will track these identifiers and assign a globally unique identifier which follows the standards set by iDigBio while the iDigBio portal provides its own GUID management. Hence, the electronic records can be traced to the actual physical specimen, while unique identification is possible.

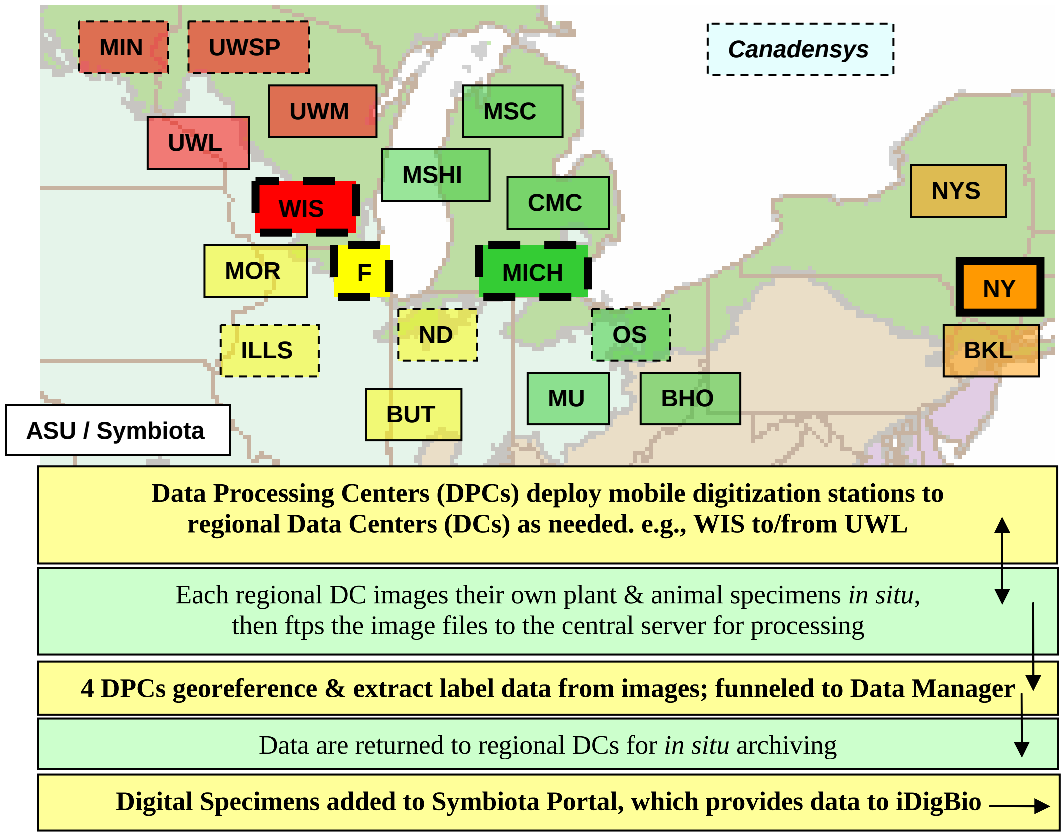

Fig. 7. Workflow Summary. The four Data Processing Centers (DPCs) are highlighted with thick lines. Subregional Digitization Centers (DCs) are color coded. Institutions processing both plant and animal specimens are indicated by dotted lines.

Specimen label data extraction from images. The technical Data Processors will use Optical Character Recognition (OCR) tools such as Tessaract built into the Symbiota portal to maximize the efficiency of **semi-automated methods** for label metadata extracting and parsing. A certain level of manual editing is anticipated for handwritten labels. We are aware of the fact that iDigBio has an OCR/data parsing taskforce in place, whose work hopefully will further streamline the label data extraction process.
Georeferencing. With images in hand, the Digitization Managers at each of the four DPCs will apply efficient and accurate georeferencing techniques using Geolocate (built into Symbiota) and/or Biogeomancer to specimens that lack latitude-longitude coordinates. All lat-long data will be entered into the specimen record indicating the source of the processed data and the level of **accuracy**. We expect that errors will be revealed occasionally through mapping functions built into our Symbiota portal.
Data deposition with iDigBio ex situ. If funded, we will work intimately with iDigBio to organize our images and data in a manner acceptable for **inclusion within the national resource** (iDigBio Specimen Portal - currently in beta production). We have reviewed their guidelines, which indicate that only images in JPEG/ JPEG2000 format with fully preserved metadata and lossless compression will be accepted.
Archiving in situ. Our project follows best practices guidelines established by iDigBio so that the issue of **data sustainability** will not be overlooked. Among these are recommendations that all data be made publicly available through one or more national data portals, but that each institution also develop a plan for archiving their digital records on site. Many of the larger institutions take advantage of in-house data servers offered through their campus IT department or through their campus library system. For those that do not, we have budgeted for one or more 2TB external hard drives for each institution for the purpose of data archiving. The recommended archival digital format for images is uncompressed DNG.

TIMETABLE OF ACTIVITIES

	2014		2015				2016				2017	
	Sr	Fl	Wr	Sn	Sr	Fl	Wr	Sn	Sr	Fl	Wr	Sn
Purchase equipment & supplies	■											
Deploy mobile digitization stations from WIS to UWL, UWM, BUT, & from NY to NYS for plant image capture		■	■	■	■							
Deploy mobile digitization stations from WIS to UWSP & ND for plant image capture						■	■	■	■			
Plant image capture by F, MOR, MSC, OS, BKL		■	■	■	■	■						
Plant image capture by WIS, MIN, MICH, NY, ILL, MU						■	■	■	■	■	■	■
Fish image capture by UWZM, MIN, MICH		■	■	■	■	■						
Fish image capture by F, ILNH, OSU							■	■	■	■	■	
Fish image capture using mobile digitization stations by UWSP, ND								■	■			
Mollusk image capture by FMNH, ILNH, OSU, UMZM		■	■	■	■	■						
Mollusk image capture by UWZM, MIN							■	■	■	■	■	
Mollusk image capture using mobile digitization station by ND									■			
Establish Symbiota portal	■											
Adapt Symbiota for lot based data entry		■	■									
Import existing data		■	■	■								
Establish synchronization protocols						■	■					
Data processing at 4 regional data centers		■	■	■	■	■	■	■	■	■	■	■
Meeting of Steering Committee in Madison	■					■						■
Subregional meetings of partners: 1 each in Madison, Chicago, Ann Arbor, & New York		■										
Outreach activity 1: Draft context and background text (MOR)	■	■	■	■			■	■			■	
Outreach activity 2: Professional workshops and presentations at practitioners' conferences (MOR)					■	■			■	■		
Outreach activity 3: Teacher training					■	■			■	■		
Outreach evaluation: data analysis					■	■	■	■	■	■	■	■
Present results at national conferences					■				■			■