Integrated Digitized Biocollections (iDigBio) Implementation Plan

Prepared By:
David Jennings, iDigBio Project Manager
Joanna McCaffrey, iDigBio Biodiversity Informatics Manager
Dr. Larry Page, iDigBio Principal Investigator
Dr. Jose Fortes, iDigBio Principal Investigator
Dr. Bruce MacFadden, iDigBio Principal Investigator
Dr. Greg Riccardi, iDigBio Principal Investigator
Dr. Pam Soltis, iDigBio Principal Investigator

Version:
3.0

Revision Date:
3/25/2013

REVISION HISTORY

Version	Description	Date
1.0	First draft	6/25/2012
2.0	Modified Digitization section and Roles & Responsibilities matrix	6/28/2012
3.0 DRAFT	Updated document formatting; Reviewed and updated milestones in each domain; Consolidated scope statement, communication plan, change management plan, risk management plan, quality management plan, and project closure plan into one document	3/25/2013

University of Florida • Florida Museum of Natural History • Dickinson Hall (Museum Rd. & Newell Dr.) • Gainesville, FL 32611 • 352-273-1906
iDigBio is funded by a grant from the National Science Foundation's Advancing Digitization of Biodiversity Collections Program (#EF1115210)

TABLE OF CONTENTS

1. PURPOSE

The Integrated Digitized Biocollections (iDigBio) implementation planning consists of this living document, which contains a high-level overview as well as more detailed milestones describing the project's implementation. This implementation plan synthesizes a variety of high-level actions within the project's five domains: (1) Project Administration & Management, (2) Education & Outreach, (3) Serving the Research Community, (4) Digitization, and (5) Cyberinfrastructure. This Implementation Plan will be modified as activities mature and information changes.

2. PROJECT SUMMARY

Integrated Digitized Biocollections (iDigBio) is the National Resource, or Home Uniting Biocollections (HUB), for the Advancing Digitization of Biodiversity Collections (ADBC) project funded by the National Science Foundation (NSF). Through ADBC, data and images for millions of biological specimens are being made available in electronic format for the research community, government agencies, students, educators, and the general public.

2.1. Vision

The vision for ADBC is a permanent database of digitized information from all biological collections in the U.S. that will lead to new discoveries through research and a better understanding and appreciation of biodiversity through improved education and outreach, which will result in improved environmental and economic policies. Creation of the digitized database will occur in four stages:

1) An initial stage in which the effort to digitize biological collections across the U.S. is catalyzed by funding from NSF and from effective iDigBio-driven activities that foster collaborations, identify priorities, and generate information on best practices related to standards, workflows, and data management for digitization of biological collections, as well as demonstrate the value of biodiversity and collections that document biodiversity.

2) An intermediate stage wherein digitization at Thematic Collections Networks (TCNs), Partners to Existing Networks (PENs), and other participating institutions improves methods and strategies, and demonstrates the scientific and societal benefits of validated and readily accessible data.

3) A third stage in which the vision for ADBC is realized through participation by all U.S. institutions with biological collections.

4) A fourth stage in which digitization is a routine and sustained practice in all institutions with biological collections, and the national database is easily accessible as an up-to-date source of information on biodiversity.

2.2. Mission

The mission of iDigBio is to develop a national infrastructure that supports the vision of ADBC by overseeing implementation of standards and best practices for digitization; building and deploying a customized cloud computing environment for collections; recruiting and training personnel, including underserved groups; engaging the research community, collections community, citizen scientists, and the public through education and outreach activities; and planning for long-term sustainability of the national digitization effort.

iDigBio will enable digitization of data from all U.S. biological collections and integrate those data to make them broadly available and useful with shared standards and formats. Ultimately, ADBC will further the discovery and understanding of biological diversity, and iDigBio will engage the research, collections, and education communities in a spirit of collaboration that will open biological research collections to new downstream user communities.

iDigBio involves the development of a permanent and powerful cloud computing infrastructure to link biological data from collections across the U.S. into a single unified web interface, overcoming the “data silos” that currently exist across the country. Search and analytical tools enable users to mine diverse data, such as taxonomy, geographic location, 2- and 3-dimensional images, vocalizations, and molecular resources tied to specimens in collections. These data promote integrative biological research on living and fossil species and provide an immense resource for agricultural science and land use management, human health, and assessing the impacts of climate change, invasive species, and other natural resource management issues.

Key partners in this effort are the TCNs, which form a national grid of institutions that are digitizing specimens and associated resources. Integration with the greater community of biocollections resources, tools and organizations is critical to accomplishing the grand challenge of digitizing and integrating data from all U.S. collections, large and small. For more details regarding the larger

community that encapsulates ADBC, please refer to the Network Integrated Biocollections Alliance (NIBA) strategic plan and implementation plan.

3. PROJECT SCOPE

iDigBio is the national resource for digitized information about vouchered natural history collections within the context established by the NIBA community strategic plan and is supported through funds from the NSF ADBC program. As such, iDigBio serves as the administrative home for the national digitization effort; fosters partnerships and innovations; facilitates the determination and dissemination of digitization practices and workflows; establishes integration and interconnectivity among the data generated by collection digitization projects; and promotes the uses of biological/paleontological collections data by the scientific community and stakeholders including government agencies, educational institutions, non-governmental organizations (NGOs), and other national and international entities to benefit science and society through enhanced research, educational, and outreach activities. iDigBio provides these services to all stakeholders with clarity, simplicity, transparency, intuitive methodology, and intuitive design.

3.1. In-Scope Activities

To accomplish these objectives, iDigBio is responsible for the following specific in-scope activities:

1) Implement a scalable and secure cloud-based infrastructure and web portal to enable the storage, integration, search, and retrieval of existing biological/paleontological specimen data, images, and other media files contributed by Thematic Collections Networks, other networks, resources, and collaborating institutions.
2) Deliver appliances that integrate and package existing digitization technologies in a manner that enhances and/or simplifies the user experience. Appliances are intended to improve the deployment and interoperability of digitization tools, and to simplify integration with the iDigBio specimen database and storage infrastructure.
3) Provide user services to support interaction with both specimen databases and with appliances. User services will support both data/appliance contributors and data/appliance consumers. User services are provided in the form of a ticket submission and tracking system for requests and problems, telephone support, email support, user documentation, and site visits.
4) Research, evaluate, benchmark, integrate, and disseminate digitization methodologies, end-to-end processes, tools, recommended standards, and workflows that improve the efficiency and scalability of digitization.

5) Provide user services to support efficient, scalable and effective digitization of specimen images, media, and specimen data. User services are provided in the form of a ticket submission and tracking system for requests and problems, telephone support, email support, user documentation, and site visits.

6) Coordinate and fund workshops and working groups to:

    a. Foster partnerships and collaboration within the collections community, as well as to connect to stakeholder organizations external to the collections community.

    b. Conduct training related to digitization, technology, workflows, and other applicable fields.

    c. Recommend standards, common practices, guidelines, workflows, and optimal digitization tools and software for use by ADBC participants.

    d. Foster innovation related to biological/paleontological collections digitization (including imaging). The outputs of these innovation workshop sessions may include:

        i. Specific application and/or hardware development requirements that are assigned to existing organizations funded for, and tasked with, tool development.

        ii. Documentation of challenges and proposed solutions that may lead to proposals to obtain funding for separate projects to deliver required technologies.

        iii. Creation or improvement of digitization and databasing tools resulting from “hackathons”. These sessions bring together skilled session participants to deliver a specific functional product during the workshop. Tools created in “hackathons” must be delivered with strategies for maintenance and sustainability.

7) Facilitate the development of standards for digitization, technology, and process training.

8) Coordinate and execute iDigBio Education and Outreach activities. Provide advice to and coordination with other digitization projects regarding the integration of outreach activities.

9) Provide opportunities and technologies that encourage communication, collaboration and status reporting among members of the ADBC community.

10) Oversee development of a plan to accomplish digitization of existing biodiversity collections in the US, and establish the long-term sustainability of the ADBC data and related infrastructure, and for iDigBio user services operations.

11) Establish an iDigBio Internal Advisory Committee (IAC) that meets regularly to report on progress in digitization efforts, share and identify best practices and standards, identify gaps in digitization areas and technology, and enhance training efforts. Also establish an External Advisory Board that meets annually to provide advice regarding project activities, the integration of digitization projects, research, education and outreach activities, strategic direction, and management policies.

12) Track research outcomes, the results of outreach activities, and innovative discoveries related to the project.

3.2. Out-of-Scope Activities

In order to reduce uncertainty in the scope of iDigBio's mission and to prevent scope creep as project requirements are evaluated, the following specific activities are defined as outside the scope of iDigBio:

1) Direct development of new tools (hardware or software), or improvements to existing software tools intended to enhance the digitization of existing, vouchered biological/paleontological collections. iDigBio is not funded or staffed to execute hardware or software development; exceptions are the creation and maintenance of the core iDigBio.org website, the portal/database designed to integrate digitized specimen data, and integration via appliances with existing tools that support digitization.

2) Collections in institutions physically located outside the United States will not be included in the iDigBio collections integration portal. Specimens collected outside the United States but housed within a US collection/location are within scope. Federally owned collections will be integrated with iDigBio through the federal data center when it is established.

3) Occurrence records and media not supported by vouchered specimens (e.g., bird sightings without a collected specimen) will not be included in the iDigBio collections integration portal. However, as other resources for these data are established, appropriate links will be created.

4) iDigBio is not responsible for the acquisition, data curation/management, and quality control of data provided by TCNs and other collaborating collections. However, as part of the execution of in-scope Activity #1 and Activity #2, iDigBio will endeavor to provide tools, features, error-checking, historical record tracking, and feedback mechanisms designed to simplify data curation/management, fitness for use tracking, and quality control by TCNs and other contributing institutions.

4. PROJECT TEAM

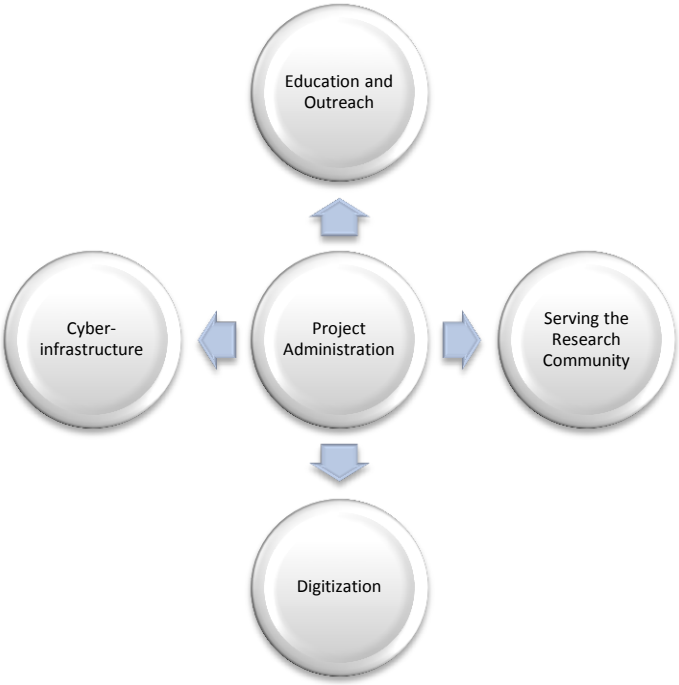
Project team members are responsible for the execution of major project activities, and project leaders serve as activity approvers. Any internal or external stakeholder may be involved in consultation or be informed about a major project activity. Core project team members are provided in the following table (students will be added and removed as they work on the project):

Name	Role / Title	Group	Phone	Email
Larry Page	Lead PI	Principal Investigator	352-273-1952	lpage@flmnh.ufl.edu
Bruce MacFadden	Co-PI (E&O)	Principal Investigator	352-273-1937	bmacfadd@flmnh.ufl.edu
Pam Soltis	Co-PI (Research)	Principal Investigator	352-273-1964	psoltis@flmnh.ufl.edu
José Fortes	Co-PI (Cyberinfrastructure)	Principal Investigator	352-392-0912	fortes@acis.ufl.edu
Greg Riccardi	Co-PI (Digitization)	Principal Investigator	850-644-2869	griccardi@fsu.edu
David Jennings	Project Manager	iDigBio Project Staff	352-273-1906	djennings@flmnh.ufl.edu
Joanna McCaffrey	Biodiversity Informatics Manager	iDigBio Project Staff	352-294-1948	jmccaffrey@flmnh.ufl.edu
Cathy Bester	Project Assistant	iDigBio Project Staff	352-294-1949	cbester@flmnh.ufl.edu
Kevin Love	User Support IT Expert	iDigBio Project Staff	352-294-1924	klove@flmnh.ufl.edu
Renato Figueiredo	Senior Personnel (Appliances)	Steering Committee	352-392-6430	renatof@ufl.edu
Shari Ellis	Project Evaluator	Steering Committee	352-273-2066	sellis@flmnh.ufl.edu
Betty Dunckel	Senior Personnel (Informal Education)	Steering Committee	352-273-2088	bdunckel@flmnh.ufl.edu
Reed Beaman	Senior Personnel (Informatics)	Steering Committee	352-273-1980	rbeaman@flmnh.ufl.edu
Austin Mast	Senior Personnel (Public Participation)	Steering Committee	850-645-1500	amast@bio.fsu.edu
Gil Nelson	Digitization Support	Steering Committee	850-766-2649	gnelson@bio.fsu.edu
Jill Holliday	Executive Editor	iDigBio Project Staff	229-327-4216	jaholliday@ufl.edu
Austin Hendy	Postdoctoral (E&O)	iDigBio Project Staff	352-274-1942	ahendy@flmnh.ufl.edu
Charlotte Germain-Aubrey	Postdoctoral (Research)	iDigBio Project Staff	352-273-1962	cgermain@ufl.edu
Grant Godden	Graduate Student (Research)	iDigBio Project Staff	352-273-1976	g0ddengr@ufl.edu
Ryan Moraski	Graduate Student (Research)	iDigBio Project Staff	352-273-1960	ryanmoraski@ufl.edu
Catherine Snyder	Undergraduate Student (E&O)	iDigBio Project Staff	352-273-2107	cssnyder@ufl.edu
Alex Thompson	Infrastructure IT Expert	ACIS Project Staff	352-226-1269	godfoder@acis.ufl.edu
Andréa Matsunaga	Research Scientist	ACIS Project Staff	352-846-2466	ammatsun@ufl.edu
Matt Collins	Infrastructure IT Expert	ACIS Project Staff	352-392-5414	mcollins@acis.ufl.edu
Greg Traub	IT Expert	ACIS Project Staff		gtraub@ufl.edu
Yonggang Liu	Graduate Student	ACIS Project Staff	352-328-5972	myidpt@gmail.com

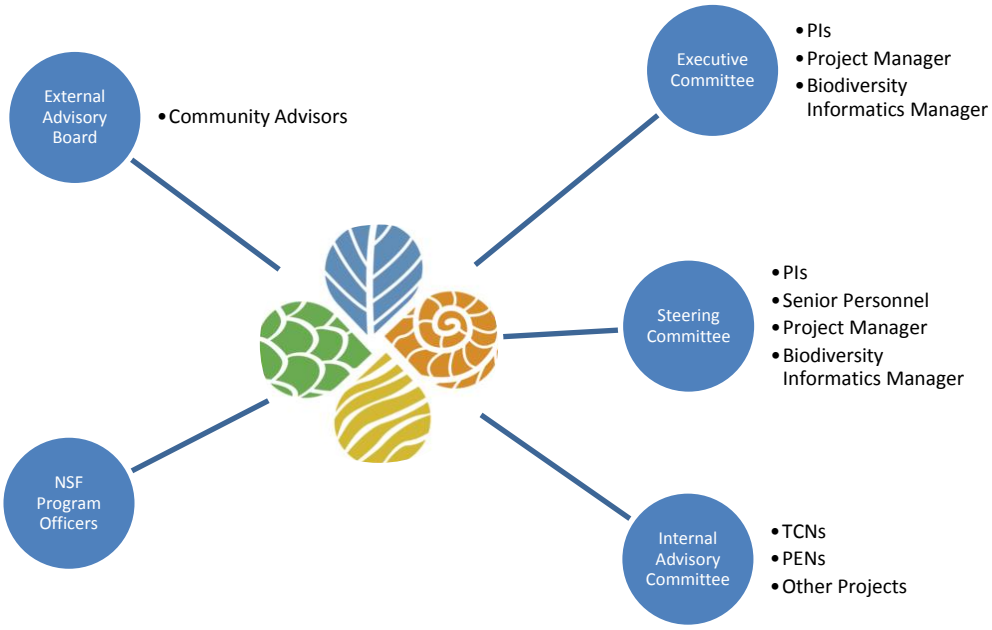
Name	Role / Title	Group	Phone	Email
Sarfaraz Soomro	Graduate Student	ACIS Project Staff	352-870-0753	sarfarazsoomro@ufl.edu
Guillaume Jimenez	Informatics IT Expert	FSU Project Staff	850-644-5775	gjimenez@fsu.edu
Deb Paul	Digitization Support	FSU Project Staff	850-644-6278	dpaul@fsu.edu
Marcia Mardis	Senior Personnel (Social Sciences)	FSU Project Staff	850-644-3392	marcia.mardis@cci.fsu.edu
Casey McLaughlin	IT Administrator	FSU Project Staff	850-645-5673	casey.mclaughlin@cci.fsu.edu
Jeremy Spinks	Web Designer	FSU Project Staff	850-228-6071	jspinks@fsu.edu
Karen Francl	External Advisory Board	Radford University	540-831-6537	kfrancl@radford.edu
Donald Hobern	External Advisory Board	GBIF	+45-3532-1471	dhobern@gbif.org
Stinger Guala	External Advisory Board	USGS	703-648-4311	gguala@usgs.gov
Vincent Smith	External Advisory Board	Coordinator for ViBRANT, NHM PI for e-Monocot	+44 (0) 207 942 5127	vince@vsmith.info
Stan Blum	External Advisory Board	California Academy of Sciences	415-379-5189	sblum@calacademy.org
Judy Skog	NSF Program Officer	NSF	703-292-7909	jskog@nsf.gov
Anne Maglia	NSF Program Officer	NSF	703-292-7380	amaglia@nsf.gov

5. PROJECT ORGANIZATION & GOVERNANCE

iDigBio is organized into five domains to effectively achieve the project goals: (1) Project Administration & Management, (2) Education & Outreach, (3) Serving the Research Community, (4) Digitization, and (5) Cyberinfrastructure. Details regarding each domain are discussed in the Implementation Plan section of this document.



The project also includes several oversight committees that provide governance, advice, and leadership concerning the project's activities, progress, goals, strategies, and implementation. A discussion of each oversight committee is provided in the following paragraphs.



5.1. Executive Committee

The Executive Committee (EC) is composed of the iDigBio Principal Investigators, the Project Manager, and the Biodiversity Informatics Manager. The Executive Committee is responsible for overall project management, general oversight of the iDigBio activities, and implementation of the strategic plan, including assurance that the digitization, research and educational missions of iDigBio are integrated and accomplished. The Executive Committee is also responsible for developing, executing, and monitoring a plan to manage conflicts of interest for the project, including the External Advisory Board.

5.2. Steering Committee

The Steering Committee (SC) is composed of the lead senior personnel from iDigBio, including the Principal Investigators, Project Manager, and Biodiversity Informatics Manager. The Steering Committee is responsible for reviewing progress and coordinating activities in digitization, computation, research, education and outreach at iDigBio, and for advising the Principal Investigators on overall resource allocation, strategic directions and management policies.

5.3. Internal Advisory Committee

The Internal Advisory Committee (IAC) is composed of representatives from the Thematic Collections Networks (TCNs), Partners to Existing Networks (PENs), and other digitization projects and collections working with iDigBio. The IAC meets regularly to report on progress in digitization efforts, share and identify best practices and standards, identify gaps in digitization areas and technology and enhance training efforts.

5.4. External Advisory Board

The External Advisory Board (EAB) is composed of members selected by iDigBio and approved by the cognizant NSF program official. The EAB meets at least once a year, and is responsible for providing written and verbal advice to iDigBio on its activities, including progress and integration of digitization projects, research, education and outreach activities among all funded institutions. In addition, the EAB is responsible for advising iDigBio leadership on strategic directions and management policies.

6. PROJECT IMPLEMENTATION PLAN

The following Project Implementation Plan delineates a high-level project progression as well as detailed milestones with target completion dates within each of the five project domains. This Implementation Plan will be modified as activities mature and information changes.

6.1. High-Level Project Progression

The high-level project progression is a schedule created based upon input from project team members, project leadership and key stakeholders. This high-level schedule contains the agreed upon set of key milestones and target completion dates that will be used to guide and monitor the project to successful completion. The high-level project progression is provided in Appendix A.

6.2. Roles & Responsibilities

A core group of project leaders, staff, students and internal stakeholders are responsible for delivering the high-level actions within each of the five project domains. Specific activities, roles, responsibilities, and target completion dates for delivery of these high-level actions are detailed in the following sections. Definitions for the roles and responsibilities in each section are:

- ***Responsible*** for execution of the activity and stimulating timely group progress
- ***Approver*** for the activity, confirms successful completion, and makes final decisions
- ***Consulted*** during execution of the activity, contributes information and provides input
- ***Informed*** about activity progress and decisions

6.3. Project Administration and Management

6.3.1. Key Objectives

iDigBio Project Administration and Management decisions and activities are executed under the leadership of Dr. Larry Page, Lead iDigBio PI, and David Jennings, iDigBio Project Manager. The Project Administration and Management domain is tasked with implementation of the following key objectives:

Action	Responsible	Approver	Consulted	Informed
a. Coordination of advisory boards, Steering Committee, and key meetings	David Jennings, Cathy Bester	Larry Page	Board/Committee /Meeting Participants	Principal Investigators
b. Planning and implementation of commitments to ensure project sustainability	Larry Page	NSF Program Officers	Community, General Public	Steering Committee
c. Complete quarterly, semiannual, and annual reporting	David Jennings	Larry Page	iDigBio Employees & Leadership	Principal Investigators
d. Comply with NSF requirements	David Jennings, Principal Investigators	Larry Page	iDigBio Employees & Leadership	Principal Investigators
e. Build and maintain strategic partnerships	David Jennings, Principal Investigators, Joanna McCaffrey	Larry Page	iDigBio Employees & Leadership	Principal Investigators, NSF
f. Budget planning, monitoring, and control	David Jennings, Cathy Bester	Larry Page	iDigBio Leadership	Principal Investigators
g. Oversee risk management	David Jennings	Larry Page	iDigBio Employees & Leadership, Community, NSF	Principal Investigators
h. Oversee change management	David Jennings	Larry Page	iDigBio Employees & Leadership, Community, NSF	Principal Investigators
i. Plan and manage project-level communication	David Jennings	Larry Page	iDigBio Employees & Leadership	iDigBio Employees & Leadership
j. Develop and maintain collaboration and communication capabilities (tools, resources)	Kevin Love	David Jennings	iDigBio Employees & Leadership	iDigBio Employees & Leadership
k. Overall project planning and coordination	David Jennings, Cathy Bester, Joanna McCaffrey	Larry Page, Principal Investigators	iDigBio Employees & Leadership	iDigBio Employees & Leadership, NSF
l. Participate in community implementation planning with NIBA	David Jennings, Principal Investigators	Larry Page	iDigBio Employees & Leadership	iDigBio Employees & Leadership

6.3.2. Methodology

Daily planning and decision-making fall within the level of authority of the Project Manager. Any project decisions impacting project scope, budget or schedule require a consensus opinion of the Steering Committee. iDigBio activities are coordinated within and across iDigBio domains, and with collaborative organizations. Scheduled monthly Steering Committee reviews provide regular validation of progress and harmonization of project activities.

6.3.3. Roadmap

Specific Activities

- 2012/Q2: Release a common set of slides about iDigBio for the project team to use in presentations to reinforce concise and common messaging regarding iDigBio's mission, including near and long term objectives; the slides will be posted to the iDigBio website and made available for downloading and use by workshop participants and others as needed
- 2012/Q3: Establish and maintain a NIBA/ADBC FAQ on the iDigBio Wiki
- 2012/Q3 – 2012/Q4: Review and improve current project branding, messaging, and marketing, including a website review and redesign
- 2012/Q4: Establish and maintain a Biology FAQ on the iDigBio Wiki
- 2012/Q4: Establish and maintain an Technology FAQ on the iDigBio Wiki
- 2012/Q4: Release an updated poster for distribution to collaborating institutions advertising the activities and value of iDigBio, TCNs, and interaction with strategic partners
- 2012/Q4: Conduct iDigBio Summit II to bring together all TCNs and iDigBio staff for information sharing and tactical planning/coordination among projects
- 2012/Q4: Participate in the NIBA Community Implementation Planning process
- 2013/Q1: Initiate a Sustainability Working Group, including support plans for "retired" TCNs and their data. Involve private enterprise (e.g., Microsoft, Oracle, Google); develop strategies to increase community buy-in; identify key partners in the public and private sectors in a position to support sustainability; assign resources and establish plans to foster relationships with these key partners.
- 2013/Q2: Release a pamphlet advertising the activities, roadmap, objectives and value of iDigBio, TCNs, and interaction with strategic partners
- 2013/Q2: Establish a Working Group focused on "Stakeholder Inclusion, Community Building and User Support". Identify, engage, and build relationships with the broader stakeholder community (i.e., groups outside the ADBC initiative). Acquire iDigBio functional requirements from this broad end-user group. Identify User Support and training needs. Enlist support from the broader stakeholder community to provide structured assistance to iDigBio and TCNs via production environment bug detection, feature requests, and user acceptance testing.

- 2013/Q3: Establish a Working Group focused on International activities related to iDigBio for information sharing and tactical planning/coordination among projects.
- 2013/Q4: Conduct iDigBio Summit III to bring together all TCNs and iDigBio staff for information sharing and tactical planning/coordination among projects

Ongoing Activities

- Maintain social collaboration infrastructure via www.idigbio.org Articles, Forums, Wiki, eNewsletter, Listservs, Video/Teleconferencing services, and social media presence. Continue to enable and encourage social networking and collaboration within the community, and publish content via the dissemination channels regarding project progress, digitization resources, issues important to the collections community, and pertinent Information Technology updates.
- Conduct regular Steering Committee meetings to evaluate progress, set strategic and tactical goals, coordinate project activity, and make high-level decisions that impact project scope, budget or schedule
- Conduct regular Internal Advisory Committee (IAC) meetings to enable communication and collaboration between and among iDigBio and the Thematic Collections Networks; share progress, discuss technical and procedural issues and potential solutions, identify opportunities for coordination of activities, answer questions, and identify TCN assistance needs that may be met by iDigBio or fellow TCNs.
- Conduct annual External Advisory Board meetings to evaluate high-level project progress, set strategic goals and future direction, provide recommendations for project-level improvement, and recommend enhancements or modifications to project objectives
- Conduct regular project budget reviews
- Conduct regular project schedule reviews, focused on attainment of project objectives and milestones
- Complete ongoing activities required for compliance with the NSF Cooperative Agreement
- Identify strategic partners within the community such as NSCA, SPNHC and biodiversity-oriented societies. Seek out synergistic alliances, communicate regarding progress, and maintain integrated strategic plans/roadmaps with these partners. Add identified actions to the implementation plan as they are developed.
- Track project risks, and conduct risk management to produce a known and monitored set of risk triggers and planned actions in response to the realization of a risk
- Follow the change management process for any project modifications that impact project schedule, budget, or scope
- Follow an internal and external project communication plan, monitor the effectiveness of the plan, and make revisions as needed

- Follow the established process to enable new, interested strategic partners to engage iDigBio via Workshop Proposals and Working Group Proposals (https://www.idigbio.org/content/workshop-or-working-group-proposal)

6.4. Education and Outreach

6.4.1. Key Objectives

iDigBio Education and Outreach (E&O) decisions and activities are executed under the leadership of Dr. Bruce MacFadden, iDigBio co-PI. The Education and Outreach domain is tasked with implementation of the following key objectives:

Action	Responsible	Approver	Consulted	Informed
a. Engage the general public through informational resources, compelling deliverables, and opportunities to participate that highlight the importance of biodiversity collections and digitization	Bruce MacFadden, Betty Dunckel	Steering Committee	iDigBio Employees & Leadership	General Public
b. Develop educational resources for K-12 students related to digitization and biodiversity	Bruce MacFadden, Betty Dunckel	Steering Committee	iDigBio Employees & Leadership	General Public
c. Foster project awareness within the professional community	Bruce MacFadden, Betty Dunckel	Steering Committee	iDigBio Employees & Leadership	Collections Community
d. Identify target audiences, including university students, downstream user groups and other stakeholders, and assess their needs	Bruce MacFadden, Betty Dunckel	Steering Committee	iDigBio Employees & Leadership	General Public
e. Identify downstream partners and stakeholders, document their needs, and execute outreach activities to meet those needs	Bruce MacFadden, Betty Dunckel	Steering Committee	Shari Ellis	General Public
f. Coordinate the “Fossils in the Cloud” Program	Bruce MacFadden	Steering Committee	Betty Dunckel	General Public
g. Measure the geographic distribution of the impact and success of intended learning outcomes	Kevin Love, Alex Thompson, Shari Ellis	Bruce MacFadden	Steering Committee	David Jennings, Principal Investigators, NSF
h. Identify and seek additional	Bruce MacFadden	Steering	Betty Dunckel,	David Jennings,

Action	Responsible	Approver	Consulted	Informed
funding sources for spin-off E&O projects that promote the mission of iDigBio		Committee	Austin Mast	Principal Investigators, NSF

6.4.2. Methodology

Many substantial E&O activities are dependent upon infrastructure prerequisites. The iDigBio Specimen Portal must be developed and populated with some amount of collections data before hands-on iDigBio E&O activities can occur. As such, E&O opportunities will grow as infrastructure develops and new stakeholders are involved. In the early project stages, E&O activities will focus on general digitization curricula development, stakeholder identification, public speaking engagements (to be recorded and published for national impact), and planning activities for future years.

6.4.3. "Fossils in the Cloud" Program

The "Fossils in the Cloud" program is an iDigBio E&O activity that is focused on diverse target audiences interested in using digitized fossil specimens for research and education in U. S. natural history museum collections. During Year 1 of iDigBio, the "Fossils in the Cloud" project reached out to two Florida fossil clubs to present talks about digitizing fossil specimens in the U.S. During Years 2 through 5, the "Fossils in the Cloud" project will ramp up activities by hiring a postdoc, graduate students and undergraduates as well as engage interested faculty curators to further develop and promote this program to target audiences and other potential stakeholders. In addition to outreach activities, a proximal goal (years 2 through 4) of "Fossils in the Cloud" will be to conduct research (e.g., morphological, distributional) on digitization of specimens from Panama that have been collected as a result of recent paleontological activities along the Panama Canal conducted by UF and Smithsonian scientists.

6.4.4. Roadmap

Specific Activities

- 2012: Conduct planning for University Education materials to be developed and delivered through discrete releases in Y2, Y3, Y4 and Y5 related to "Fossils in the Cloud"
- 2012/Q3: iDigBio 2012 Visiting Scholar – Dr. Anna Monfils
    - Design and release interactive online lab experience introducing the purpose and procedures of specimen databasing to majors and non-majors in the biological sciences. Students "collect" data with camera phones, then "deposit" their collections into a state-wide, student-generated database, and then mine that database. Apply a problem-based learning format to help students understand strengths and challenges of online data, virtual collections, cyberinfrastructure, bioinformatics, and global communication.

    - Organize, present and record a collections digitization workshop for professional botanists, students and the public. Disseminate the recording via www.idigbio.org to a national audience.
- 2012/Q4: Conduct an Education & Outreach session at the iDigBio Summit II to review and compare detailed E&O plans among the TCNs, and to produce a plan to coordinate and consolidate E&O activities to reduce duplication of effort where commonalities exist between multiple ADBC participants.
- 2013/Q1: iDigBio 2013 Visiting Scholar – Dr. Corey Toler-Franklin
    - American Museum of Natural History (AMNH) in New York – Introduce non-invasive optical capture techniques originally developed in computer science research and adapted for art history (and archaeology) conservation and digital archiving to digitize recent and fossil primates in the AMNH Vertebrate Paleontology and Mammalogy collections.
    - Duke University Lemur Center in North Carolina – Continue to apply her digital capture methods in the Fossil Primates Division, taking into consideration the different issues faced by larger (AMNH) vs. smaller (Duke) collections in regard to digitization efforts and resources.
    - Florida Museum of Natural History in Gainesville – Understand the overall goals of ADBC program and learn about approaches used by collection digitizers in Florida and at other TCNs across the country.
    - As an ongoing part of the project, Dr. Toler-Franklin plans to present her research and findings in seminars at institutions including AMNH, CUNY, NYU and Duke. She also expects to publish on this research in fields that include Computer Science, Anthropology, and Vertebrate Paleontology. Lastly, Dr. Toler-Franklin plans to offer a workshop at the iDigBio Hub that will enable her to present results and lessons learned as well as introduce practitioners to her data capture methods.
- 2013-2014: Deliver short educational videos via YouTube, www.idigbio.org, TeacherTube and/or the Explore Research exhibit at the FLMNH focused on TCN-specific activities to engage the public in specific digitization projects, and to encourage participation and support in areas such as Crowdsourcing for each project.
- 2013/Q1: Establish an iDigBio Education & Outreach subcommittee, whose purpose is to plan and coordinate E&O activities that relate to the Broader Impact objectives of iDigBio, and to the extent possible (given resources), the TCNs.
- 2013/Q3: Conduct an Education & Outreach Workshop to enhance and expand E&O deliverables and objectives.

- 2013/Q2: Explore iDigBio participation in the US Science Exposition in 2014; with a focus on demonstrating the digitization portal, sparking interest in public participation, and demonstrating tools such as 3-D Printing and digitization stations
- 2013/Q2: Evaluate the Drexel University (via Richard McCourt) undergraduate course on bioinformatics for dissemination via www.idigbio.org
- 2013/Q3: Conduct planning for K-12 educational resources to be developed and delivered through discrete releases in Y3, Y4 and Y5 related to "Fossils in the Cloud"
- 2013/Q3: Disseminate museum-studies curricula via www.idigbio.org for implementation in institutions throughout the country
- 2013/Q3: Assuming that additional funding is secured, approximately 12 undergraduates will be trained in Paleo specimen digitization as part of the "Fossils in the Cloud" initiative
- 2013/Q3 – 2014/Q3: Identify partners and stakeholders via a front-end survey
- 2013/Q3 – 2015/Q3: Complete a comprehensive needs assessment for downstream users of iDigBio services (both technical and informational)
- 2014 – 2015: Deliver interactive content for Smartboards, accessible to schools and organizations with Smartboards across the county (including the Explore Research exhibit at the FLMNH), that enables interactive access through a simplified, K through 12-focused interface to the data available within the iDigBio database.
- 2014/Q3 – 2014/Q4: Conduct planning for Museum and Cyberexhibit education resources to be developed and delivered through discrete releases in Y4 and Y5
- 2015 – 2016: Expand and enhance the interactive Smartboard interface and the video content produced through the Explore Research program.
- 2015/Q3: Plan new activities and adjust plans based upon results of the 24-month downstream user needs assessment

Ongoing Activities

- Recruit and retain Postdoc, Grad Student, and Undergraduate Students to participate in planning and delivery of research and educational deliverables, in the form of published papers and learning materials for downstream users. Activities are part of the "Fossils in the Cloud" program.
- If funds are available, iDigBio will implement a Distinguished Speaker program in Years 2-5. One individual will be compensated for five speaking engagements per year. This individual will be trained to become a skilled ambassador for iDigBio, speaking at Universities and public forums. Key presentations will be recorded and disseminated via www.idigbio.org to a national audience.
- Facilitate the securement of funding for E&O projects.

- Identify target audiences and user groups for E&O, including 1) Collection researchers and other professionals; 2) TCNs and other non-federal collections networks; 3) University students and others in the community; 4) Government agencies and policy makers; 5) Industry and business innovators; 6) Downstream users and other stakeholders: 4H, interest groups, clubs, etc.
- Deploy iDigBio staff to disseminate information and increase visibility to the professional community by presenting posters, pamphlets, and talks at national meetings, conferences and workshops.
- Deploy iDigBio staff to disseminate information and increase visibility to downstream user groups. Record key talks and disseminate via www.idigbio.org.

6.5. Serving the Research Community

6.5.1. Key Objectives

iDigBio decisions and activities related to effectively serving the research community are executed under the leadership of Dr. Pamela Soltis, iDigBio co-PI. The Serving the Research Community domain is tasked with implementation of the following key objectives:

Action	Responsible	Approver	Consulted	Informed
a. Produce detailed Use Cases for research applications of specimen data, and provide these Use Cases to the Cyberinfrastructure team	Pam Soltis	Steering Committee	Collections Community	iDigBio Employees & Leadership, Collections Community, NSF
b. Engage the research community to market and build interest in utilizing iDigBio's services, including both data access services and collaboration tools	Pam Soltis, José Fortes	Steering Committee	Collections Community, iDigBio IT & Digitization Employees	iDigBio Employees & Leadership, Collections Community, NSF
c. Seek opportunities for integration of iDigBio specimen data and data access services with key data and research services from other projects and organizations	Pam Soltis, José Fortes	Steering Committee	Collections Community, iDigBio IT & Digitization Employees	iDigBio Employees & Leadership, Collections Community, NSF
d. Identify strategic partners in the research and collections community and develop synergistic relationships with those partners	Pam Soltis, David Jennings, Kevin Love	José Fortes/Greg Riccardi (IT Requirements Committee)	Principal Investigators, iDigBio IT & Digitization Employees	iDigBio IT & Digitization Employees

6.5.2. Methodology

Existing relationships between iDigBio PIs and the collections/research community provide an informal mechanism to accomplish the key objectives. Branding and messaging related to iDigBio will be delivered clearly and frequently in order to gain traction for community adoption of iDigBio's Specimen Portal, and also the adoption within the community of iDigBio's infrastructure for collaboration (e.g., Wiki, Forums, Articles, Listservs, Adobe Connect teleconferencing, Workshop coordination/funding, Working Group coordination/funding, Appliance development, Hosting of services). Through workshops and presentations at professional meetings, iDigBio personnel will highlight developments and opportunities for research using specimen data either alone or in conjunction with other types of data. Integration with other data repositories will be accomplished through ongoing discussions and collaborations. The creation of detailed Use Cases will be managed on an ongoing basis, with small focus groups and conversations with power users occurring following version releases of the Specimen Portal. As new functionality is provided by a release, the Research Community will be queried to validate the effectiveness of the newly delivered functionality, and to develop additional Use Cases for future development. This will enable the research community to contribute to the iterative software development cycle for the Specimen Portal and associated services. Furthermore, the development of Use Cases will identify external data sources that should be integrated with iDigBio.

6.5.3. Roadmap

Specific Activities

- 2012/Q3: Conduct a Digitization Tools and Practices Workshop and Symposium at the Botany 2012 conference
- 2012/Q3: Gather user Research Community feedback regarding the specimen portal and high-level Cyberinfrastructure roadmap. Discuss, document, and prioritize advanced Use Cases for Research Applications of Specimen Data, in order to enable effective design decisions from the Cyberinfrastructure domain. Specifically address:
    - How would data sets be accessed?
    - How would data sets be downloaded?
    - What research-based queries would run against the database?
    - What are novel points of comparison between specimen records?
    - What specific data elements are required, including verbatim and derived data?
    - What big research needs loom that cannot be solved with current technologies?
    - Is there a requirement for in-tool mapping of locality data? What would that application look like? What mapping layers are needed?
    - What cross-species links are important?
    - Are in-tool links to external data sources required, and if so which sources (medical, climate, genetics, phylogenetic trees, etc.)?

    - Provide Use Cases that describe basic biological research questions.
    - What are the most compelling data fields, with some general ordering by importance?
- 2012/Q3: Establish a Working Group focused on "Serving the Research Community". This working group will provide a core group to assist with evaluation of product functionality and ongoing creation/refinement of Use Cases.
- 2013: Develop and prioritize Use Cases to extend iDigBio connections to databases inside systematics and outside systematics, from ecology to genomics (i.e., from NEON to GenBank).
- 2012/Q3: Conduct a "Linked Data Workshop" in collaboration with NESCent
- 2013/Q3: Conduct a workshop to engage members of the systematics community, resulting in specific Use Cases for the specimen data integrated by iDigBio and additional services that may be required within the toolset.

Ongoing Activities

- Research team, consisting of a post-doc and two graduate Research Assistants, will begin a project to investigate patterns of biodiversity in Florida, integrated with molecular-based phylogenies, using specimen data from FLMNH and other aggregated sources
- Begin construction of national network of genetic resources repositories, through follow-up from previous GRR surveys (RAs)
- Work with research communities to develop collections and research-related workshops and symposia at existing meetings and conferences.
- Work with research communities to identify opportunities for interfaces with data repositories and other services, to promote integrated research. Opportunities will be formalized as Use Cases that are vetted, prioritized, and scheduled for implementation with the Cyberinfrastructure domain.
- Promote iDigBio and identify opportunities for synergy and partnerships to botanical gardens, zoos, and culture collections.
- Conduct presentations and spark community interest and discussion at national meetings of professional societies (systematics, ecology, evolution, genomics).
- Form formal working relationships with various data repositories, and identify opportunities for synergy and partnerships that may be incorporated into the implementation plan. Repositories include: TreeBASE, Dryad, NEON, GenBank, iPlant

6.6. Digitization

6.6.1. Key Objectives

iDigBio decisions and activities related to optimization of digitization workflows and processes, digitization documentation, and efforts to share and improve digitization tools within the collections

community are executed under the leadership of Dr. Greg Riccardi, iDigBio co-PI. The Digitization domain is tasked with implementation of the following key objectives:

Action	Responsible	Approver	Consulted	Informed
a. Engage the collections community to market and build interest in utilizing iDigBio's services, including both data access services and collaboration tools	Deb Paul, Gil Nelson, Austin Mast, Joanna McCaffrey	Greg Riccardi	Steering Committee, Collections Community	iDigBio Employees & Leadership, Collections Community
b. Obtain preliminary data sets for ingestion, storage, and exposure via the iDigBio Specimen Portal	Deb Paul, Gil Nelson, Joanna McCaffrey	José Fortes/Greg Riccardi (IT Requirements Committee)	Collections Community	iDigBio IT & Digitization Employees
c. Establish Minimum Information Standards and data fitness for use parameters	Gil Nelson, MISC Working Group	José Fortes/Greg Riccardi (IT Requirements Committee)	Collections Community	iDigBio IT & Digitization Employees, Collections Community, NSF
d. Optimize digitization workflows	Deb Paul, Gil Nelson, Joanna McCaffrey	Greg Riccardi	Collections Community	Collections Community, NSF
e. Conduct digitization training and produce online training materials	Deb Paul, Gil Nelson	Greg Riccardi	Collections Community	Collections Community, NSF
f. Enhance and broaden exposure to digitization tools and resources including Georeferencing, Optical Character Recognition (OCR), and Natural Language Processing (NLP)	Deb Paul	Greg Riccardi	Steering Committee	Collections Community
g. Enhance and broaden exposure to digitization tools and resources including Authority Files and Optimized Digitization Workflows	Gil Nelson	Greg Riccardi	Steering Committee	Collections Community
h. Enhance and broaden exposure to digitization tools and resources including Crowdsourcing	Austin Mast	Greg Riccardi	Steering Committee	Collections Community
i. Evaluate, document and publish analysis related to digitization hardware and software tools	Gil Nelson	Greg Riccardi	Collections Community	Collections Community

Action	Responsible	Approver	Consulted	Informed
j. Identify significant technological gaps in digitization capabilities that require additional resource investment in order to ensure the success of ADBC	Deb Paul, Gil Nelson, Austin Mast, Greg Riccardi, David Jennings, Joanna McCaffrey	Steering Committee	Collections Community	NSF
k. Conduct activities as required to improve TCN efficiencies, resolve TCN problems, and remove roadblocks; specifically, provide User Services related to digitization questions from the community	Deb Paul, Gil Nelson, David Jennings, Joanna McCaffrey	Greg Riccardi	Requestor, Steering Committee	Collections Community

6.6.2. Methodology

Digitization experts from iDigBio will pursue a process of information gathering and documentation based upon both grounded theory and business process modeling/management. A high degree of community contact will occur as workflows, processes, and digitization practitioners are observed in the field. These observations and subsequent analysis will yield documentation, papers, training materials, and prepared presentations that will concisely present effective practices for digitization, including both qualitative and quantitative measures of success. Documentation of hardware and software tools available to the collections community will be developed and maintained. A virtual community dialog related to disseminating and discussing all aspects of digitization will be encouraged via the iDigBio website, including community-building tools such as the iDigBio Forum, Wiki, and Articles.

iDigBio digitization experts will work with the community to understand gaps related to process and tools that prevent effective and efficient digitization of specimens. Direct improvements will be made whenever possible via collaboration with experts within and outside of the biodiversity community. When direct improvements are not possible, recommendations for NIBA initiatives will be developed. In order to yield data that is scientifically beneficial within the integrated Specimen Portal produced by the Cyberinfrastructure domain, community engagement will be sought in the definition of Minimum Information Standards and parameters to determine the fitness of use for contributed data. All of these activities will be catalyzed by site visits, virtual and on-site workshops, working groups, industry research, contact with tool developers, and user contact.

6.6.3. Roadmap

Specific Activities

- 2012/Q2: Submit a publication to ZooKeys summarizing the information gathered from observations of digitization practices at numerous institutions, including both mature digitization processes and digitization activities that are being newly implemented
- 2012/Q2: Release the initial Collections Management Database software technology matrix to the www.idigbio.org Documentation page.
- 2012/Q3: Enable the community Georeferencing leaders to consolidate Georeferencing training and reference materials onto the iDigBio Wiki and Forums for public consumption
- 2012/Q3: Conduct Minimum Information Standards for Collections (MISC) meetings to define data fit for scientific purposes, and publish the results as a minimum data standard (for both completeness and quality) in order to satisfy fitness for scientific use
- 2012/Q3: Produce for public review and comment an optimized, modular digitization workflow for “Specimens on Flat Sheets” (Herbaria)
- 2012/Q3: TDWG to open Audubon Core and Audubon IPT Extension proposal submission and public comment. Upon approval work with GBIF to implement modifications.
- 2012/Q3: Conduct a review with each TCN to ensure they are gathering and have a data plan to share appropriate/useful information based upon the MISC publication
- 2012/Q3: Establish and promote an “Innovation HUB” on the iDigBio website, leveraging the Wiki and Forums. Publication of useful tools, tips and techniques from various institutions will occur on the Wiki, and a dialog related to these innovative ideas will be cross-referenced on the Forum.
- 2012/Q3: Produce for public review and comment an optimized, modular digitization workflow for “Pinned Specimens”
- 2012/Q3: Finalize the “Specimens on Flat Sheets” workflow and post to the www.idigbio.org Documents section, within the Workflows page
- 2012/Q3: Conduct the first of a four workshop series of Digitization Workshops for institutions planning or implementing new digitization programs. Workshop #1: Vascular and non-Vascular Plant Collections
- 2012/Q3: Conduct the “Public Participation in the Digitization of Biodiversity Specimens Workshop” (Citizen Science) and develop a plan to implement activities defined in the Workshop for engaging the public in the digitization process
- 2012/Q4: Finalize the “Pinned Specimens” workflow and post to the www.idigbio.org Documents section, within the Workflows page
- 2012/Q4: Compile the initial Digitization Workshop materials into online video, documentation and eLearning products distributed through www.idigbio.org

- 2012/Q4: Plan the remaining three Digitization Workshops, focusing on diverse preservation methods and applying lessons learned from the first Workshop
- 2012/Q4: Conduct a Georeferencing "Train the Trainer" workshop to expand Georeferencing expertise within the TCNs and beyond
- 2012/Q4: Finalize planning for graduates of the "Train the Trainer" workshop to conduct their initial Georeferencing training sessions for their key target audience, with onsite oversight and assistance from Georeferencing experts
- 2012/Q4: Finalize the "3-Dimensional Specimens" workflow and post to the www.idigbio.org Documents section, within the Workflows page
- 2012/Q4: Publish the initial "Collaborative Partner Network" list to www.idigbio.org (see Ongoing Activities section below)
- 2013/Q1: Generate and publish to www.idigbio.org a "Project Sphere of Influence and Interaction" diagram to include annotations of touch-points and interoperability (now and in the future) with other projects and collaborating organizations
- 2013/Q1: Conduct an Augmenting Optical Character Recognition "Hackathon" to introduce and evaluate new OCR tools, determine configuration for optimal results, refine OCR code to improve performance and accuracy, and benchmark/compare OCR tools using identical datasets
- 2013/Q1: Document and publish parameters to be considered if a collection is building an in-house Collections Management Database solution
- 2013/Q1: Release recorded Online Georeferencing Webinars
- 2013/Q2: Produce for public review and comment an optimized, modular digitization workflow for "3-Dimensional Specimens" (e.g., mammals, fossils)
- 2013/Q3: Release Online Georeferencing eLearning products
- 2013/Q3: Identify and aggregate internal (collection-maintained) Authority Files, and make them available for public use and maintenance on www.idigbio.org
    - Taxonomic authority sources
    - Collector names authority sources
    - Geographic names sources
    - OCR-training files for default recognition configuration of OCR software applications
- 2013/Q3 – 2013/Q4: Develop detailed Use Cases for data cleansing protocols
- 2013/Q4 – 2014/Q1: Develop detailed Use Cases for digitization and mobilization of associated specimen data, including field notes, illustrations, gene sequences, vocalizations, habitat data, food habits, climatic data, and reproductive data

Ongoing Activities

- Seek and obtain high-quality vouchered specimen data sets from U.S. collections.

- Continuously improve, update and expand published documentation related to digitization workflows, hardware tools, and software tools.
- Post to www.idigbio.org an annotated bibliography of external articles pertinent to the digitization of natural history collections, and useful to the collections community
- Maintain the "Collaborative Partner Network" list documenting relationships with collaborative partners (e.g., Zooniverse, EOL, VertNet). Publish the existence and nature of each relationship as iDigBio Documentation. Enable individual TCNs and other data contributors to work through iDigBio Digitization User Services to leverage these existing connections in a coordinated manner, rather than requiring each affiliated entity to seek out and contact common partners each time a new connection is required.
- Provide workflow analysis services for digitizing institutions.
- Assist with and recruit collections personnel to collaborate with the Cyberinfrastructure domain for user testing and beta use of new software tools.
- Identify, document, and share new innovations from the TCNs (i.e., breakthrough methods) and other digitizing institutions; include this information in the "Innovation HUB" on www.idigbio.org.
- Provide leadership for the Minimum Information Standards, Authority Files, & Semantics (MISC) Working Group.
- Provide leadership and resource guidance for the Georeferencing Working Group.
- Provide leadership for the Public Participation in Digitization Working Group.
- Provide leadership for the Augmenting Optical Character Recognition Working Group
- Respond to Cyberinfrastructure needs, particularly in relation to lessons learned and opportunities for enhancement based upon Morphbank expertise

6.7. Cyberinfrastructure

6.7.1. Key Objectives

iDigBio Cyberinfrastructure decisions and activities are the responsibility of the Advanced Computing and Information Systems Laboratory (ACIS) at the University of Florida, under the leadership of Dr. José Fortes, iDigBio co-PI. Ongoing project decisions are evaluated based upon the experience and expertise of ACIS personnel, industry research, biodiversity community & information technology community input, and internal project input from the iDigBio project team at Florida State University. The Cyberinfrastructure domain is tasked with implementation of the following key objectives:

Action	Responsible	Approver	Consulted	Informed

Action	Responsible	Approver	Consulted	Informed
a. Research, select, implement and maintain a horizontally scalable cloud infrastructure for object (media) storage	Alex Thompson, Matt Collins	José Fortes	Collections Community IT Resources	Steering Committee
b. Research, select, implement and maintain a horizontally scalable cloud infrastructure for text (data/metadata) storage	Alex Thompson, Matt Collins	José Fortes	Collections Community IT Resources	Steering Committee
c. Implement infrastructure to enable hosting for the web services/websites of strategic partners	Renato Figueiredo, Matt Collins	José Fortes	Requestor, Steering Committee	Requestor, General Public
d. Deploy iDigBio appliances and services via multiple channels (e.g., web services, locally-run virtual machines, Infrastructure-as-a-Service cloud implementations) to enhance, simplify and/or improve activities completed by data providers and data consumers	Renato Figueiredo	José Fortes	Requestor, Steering Committee	Requestor, General Public
e. Implement a comprehensive authentication and access control system to enable data tracking and a cohesive user experience among the systems listed above, as well as the iDigBio collaboration and communication website (comprised of Drupal, Redmine, and MediaWiki installations)	Alex Thompson, Kevin Love	José Fortes	iDigBio IT & Digitization Employees	Steering Committee
f. Develop, implement and maintain iDigBio APIs to access text and media data stores	Alex Thompson, Kevin Love	José Fortes	Steering Committee, Collections Community	General Public, Collections Community

Action		Responsible	Approver	Consulted	Informed
g.	Develop, implement and maintain a Graphical User Interface (GUI) to enable end-users, including data contributors and data consumers, access to search/visualize/download text and media data from the cloud infrastructure	Alex Thompson, Greg Traub	José Fortes	Steering Committee, General Public, Collections Community	General Public, Collections Community, NSF
h.	Review, select and apply data and media standards	Andréa Matsunaga, Gil Nelson, Deb Paul, Alex Thompson, Matt Collins, Joanna McCaffrey	José Fortes	Steering Committee, Collections Community	iDigBio Employees & Leadership, Collections Community
i.	Review, select and apply data transmission standards and protocols	Andréa Matsunaga, Gil Nelson, Deb Paul, Alex Thompson, Matt Collins, Joanna McCaffrey	José Fortes	Steering Committee, Collections Community	iDigBio Employees & Leadership, Collections Community
j.	Integrate iDigBio services and user portals with Morphbank and other collaborators	Andréa Matsunaga, Alex Thompson, Matt Collins, Joanna McCaffrey	José Fortes	Steering Committee, Collections Community	Steering Committee, Collections Community, NSF
k.	Secure additional resources for project infrastructure to maintain adequate performance and capacity	José Fortes	Steering Committee	NSF	Steering Committee, NSF

6.7.2. Development Methodology

The Cyberinfrastructure implementation process will follow a formula that balances strategic planning with the agility to meet new challenges, short-term project needs, and enhanced/clarified specifications. The process follows three key tenets:

- Following the initial implementation of core features in 2012/Q3 and 2012/Q4, releases (with the exception of bug fixes) will be delivered on a 6-month schedule, in June and December. Each release will plan for a minimum of three weeks of unit and user testing as well as allocated time for technical (code) and user (system functionality) documentation, at the end of each cycle. The developmental components (objectives and deliverables) of the upcoming 6-month

development/testing cycle will be documented and approved by project leadership at least one month prior to the initiation of the cycle.

- The activities within each 6-month cycle will be compartmentalized into 3-week Sprints to enable reasonable periods for resource planning, progress/milestone checks, and validation of the course of development activities.
- Future releases will be coarsely mapped out a minimum of two years into the future to produce a product roadmap. The roadmap will be continually refined in an iterative manner when new information is received and when development cycles are completed.

6.7.3. Requirements Review Process

1) All internal and external feature, bug fix, and functional requests are documented as Issue tickets in Redmine (www.idigbio.org/redmine).

2) The ticket is assigned to an appropriate Subject Matter Expert (SME) to review the request based on completeness, feasibility, applicability to the iDigBio Scope, etc. Items that require additional information will be delegated as appropriate, including consultation with the iDigBio PIs to discuss the scientific merits of requests.

2) Requests that pass the “sanity check” will be presented at the regular iDigBio IT meetings for discussion and recommendation for approval or disapproval.

   a) Disapproved requests will be closed in Redmine, with comment as to the reason for rejection.

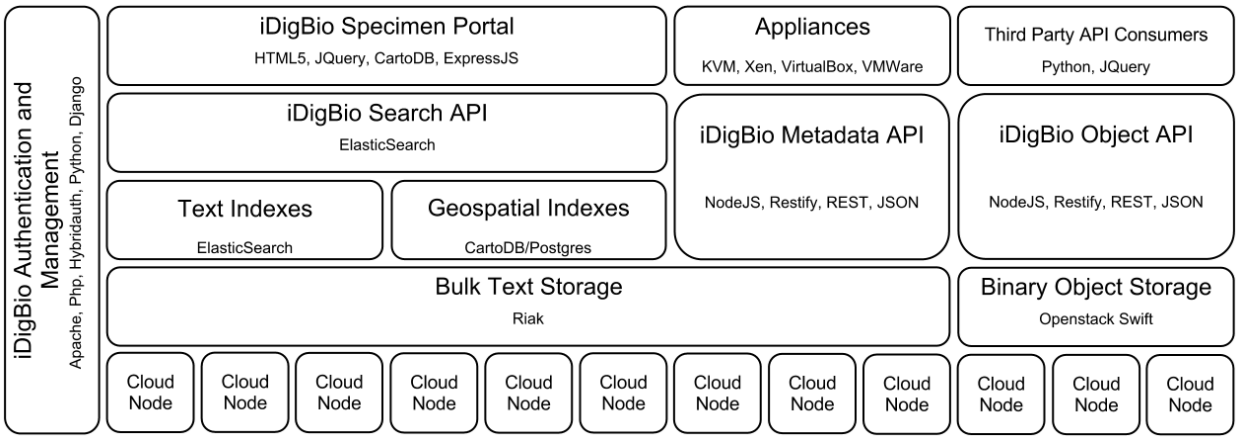
   b) Approved requests will be further developed as formal requirements and use cases. Community outreach and research community involvement, as well as additional input from the original requestor, will be required for this activity.

3) Formal requirements and use cases will be translated into technical requirements by ACIS and assigned an estimate of hours for completion. Task dependencies will also be identified (i.e., the feature cannot be implemented until another specific task has been finalized).

4) The formal requirements and use cases will be reviewed again to ensure that they have been developed as expected and remain within project scope. Implementation dates of approved requests are added to the Cyberinfrastructure roadmap for future development and product release. The project plan is updated accordingly.

6.7.4. iDigBio Cyberinfrastructure Architecture

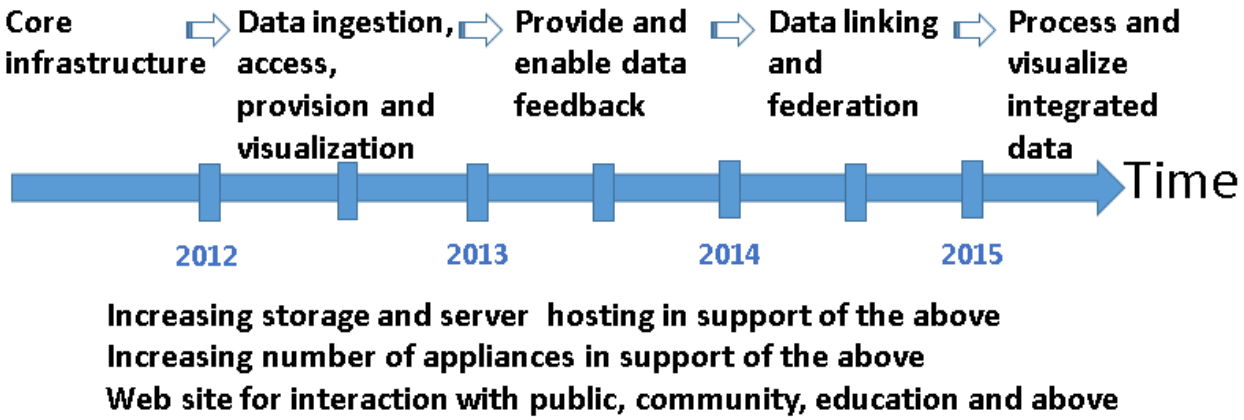
Roadmap deliverables can be traced to the architectural diagram displayed below:



Architectural diagram of the main iDigBio system components including examples of potential technology employed in each component.

6.7.5. Roadmap

Conceptual Progression



Specimen Portal Release Cycle Objectives

- **Version 0 (Jun 2012)** – Build the Core Infrastructure. Enable 'Binary Object Storage' (Media) and 'Bulk Text Storage' (Metadata) in a cloud-based infrastructure, with documented APIs to insert and retrieve data.
- **Version 1 (Dec 2012)** – Build simple GUI interfaces that allow users to query the textual data (simple and advanced search), and visualize images related to specimens.
- **Version 2 (Jun 2013)** – Data ingestion from TCNs, and implementation of iDigBio Authentication, Authorization and Access Control (AAA).

- **Version 3 (Dec 2013)** – Enable data access APIs that are useful for EOL, BiSciCol, and GBIF (push model).
- **Version 4 (Jun 2014)** – Enable data to flow back from data consumers to data providers, potentially considering integration with tools such as Filtered Push into data management applications (e.g., Specify, Arctos, Symbiota, Arthropod Easy Capture Database, KE EMu). Data flows back as generic "annotations".
- **Version 5 (Dec 2014)** – Enable data to be harmonized back into data provider's data management system. Accepted updates results in updates also in iDigBio with history tracking.
- **Version 6 (Jun 2015)** – Enable linking and federation of data, potentially through integration with tools such as BiSciCol.
- **Version 7 (Dec 2015)** – Enable complex integrated data retrieval and visualization, requiring MapReduce for dealing with computationally or data intensive tasks, and making use of high-quality data (e.g., Heat map generation).
- **Version 8 (Jun 2016)** – Expand complex integrated data retrieval and visualization, requiring MapReduce for dealing with computationally or data intensive tasks, high-quality data (e.g., for Heat maps).

Storage & APIs (Refer to the Architectural Map Above)

The following dates represent release dates when features and functionality are available for access:

- 2011/Q4: Provide temporary, traditional file system storage for use by TCNs
- 2012/Q1: Implement daily single-site backup and multi-disk redundancy
- 2012/Q2: Deploy permanent cloud storage infrastructure for the ‘Bulk Text Storage’
- 2012/Q2: Define requirements for an intuitive search User Interface (‘iDigBio Specimen Portal’)
- 2012/Q2: Deploy permanent cloud storage infrastructure for the ‘Binary Object Storage’
- 2012/Q2: Implement and deploy the ‘iDigBio Object API’
- 2012/Q3: Select and create a limited, useful set of indexes (‘Full-text and Faceted Indexing’)
- 2012/Q3: Implement search internally available to iDigBio staff via API (‘iDigBio Search API’)
- 2012/Q3: Enable web accessibility to the ‘Binary Object Storage’ (v0) via iDigBio APIs and a basic GUI (‘iDigBio Specimen Portal’)
- 2012/Q3: Gather end-user feedback regarding the v0 product. Document, prioritize and plan development activities for system expansion and refinement
- 2012/Q3: Implement text Get / Set APIs (‘iDigBio Metadata API’) to decompose a Darwin Core Archive on iDigBio systems and insert the data into the ‘Bulk Text Storage’. These APIs will not initially be released for public use due to data quality/consistency issues observed with the test data sets. Data sets will initially be received and inserted by iDigBio Cyberinfrastructure personnel.

- 2012/Q4: Deploy rich RESTful API for Riak text storage ('iDigBio Metadata API')
- 2012/Q4: Enable Full Text Search ('Full-text and Faceted Indexing')
- 2013/Q2: Implement and deploy authentication authorization and access control for all APIs ('iDigBio Authentication').
- 2013/Q2: Ingest textual data and media objects from TCNs and other data providers with data ready in DwC-A format.
- 2013/Q4: Provide robust data linking relations and federation
- 2013/Q4: Enable data feedback from end-users to data providers
- 2014/Q1: Implement a multi-site strategy with redundancy and outage fail-over to/from storage and infrastructure at a second geographical location
- 2014/Q2: Implement Tile Set Generators
- 2014/Q2: Web Feature Service
- 2015/Q2: Process and visualize integrated data for research-driven purposes
- 2015/Q2: Implement ability to execute MapReduce-like operations over the integrated data

Appliances

Appliances will be delivered as either:

1. Applications downloaded and installed by the end-user on a local desktop using a virtual machine (e.g., VMWare or Virtualbox) or, when applicable, as a packaged user application.

2. Applications that can be deployed by users on virtual machines that run in the iDigBio cloud (Infrastructure-as-a-Service) without the need for a local installation.

3. Applications that are deployed persistently in the iDigBio cloud and accessed as Web Services

The release of appliances will not be tied to the same release cycle of the iDigBio specimen portal, as appliance versions will be primarily dependent on the versions of tools integrated into each appliance. The roadmap outlined below focuses on the functionality needed for appliances in general; each appliance maintained as part of iDigBio may have its own application-specific roadmap.

There will be appliances maintained by community members external to iDigBio, and those integrated in collaboration with iDigBio personnel. We will establish a baseline appliance configuration and make available best practices and documentation on how to integrate tools with the baseline appliances to facilitate contributions from users in a manner that can automate the process of creating appliance images.

- 2012/Q1: Release of documentation and Web forms on the iDigBio portal outlining the process of requesting integration of tools into appliances by community members.

- 2012/Q2: Planning and roadmap construction for longer-term appliances involving both iDigBio development and collaboration with external application developers, to include OCR appliance(s) presented as external services, Kepler Kurator (based upon grant application status), batch object updates, GEOLocate standard and GEOLocate Crowdsourcing, and appliances to power advanced research.
- 2012/Q4: Roll-out implementation of hosted services, including Symbiota, Filtered Push, and GEOLocate.
- 2012/Q4: Documentation of baseline iDigBio appliance and best-practices on the development of packages and configuration for integration of tools with baseline appliance.
- 2013/Q1: SCAN drupal environment deployment as a hosted service.
- 2013/Q1: Release of an image ingestion appliance v1, enabling batch media object upload from TCN local storage to the 'Binary Object Storage'. V1 focuses on the use case of ingesting images that can then be accessed for crowd-source and linked to specimen records, and offers the ability to ingest a batch of images into iDigBio cloud through iDigBio REST APIs; retry uploads of individual files in a batch in the event of failures, within and across batch sessions; specify inputs for the batch upload through a combination of inputs provided through an interactive Web-based user interface, and through an input CSV file with terms defined by MISC; generate and retrieve iDigBio GUIDs and URLs through iDigBio API; generate CSV output file with MISC terms associated with each image ingested; and maintain and display history of batch uploads through the Web user interface.
- 2013/Q2: Gather additional image ingestion appliance functionality identified through usage and feedback by iDigBio users.
- 2013/Q4: Release of an image ingestion appliance v2, including user interface improvements (ability to automatically generate a properly-formed CSV input file template from images retrieved from the local file system including UUID-based GUIDs derived from a content hash of the image and optional global parameters that apply to all images; ability for the user to open and edit the CSV input file generated to manually add or change per-file parameters as needed before ingestion; ability of the user interface to work with a set of modern browsers available in typical desktop environments (Firefox, Chrome, Explorer, Safari, as appropriate for Windows, MacOS, and Linux); reliability improvements (improve logging information available in the code for debugging purposes, and provide ability for users to specify logging levels and to easily identify and send logging information to iDigBio developers); ingesting images linked to specimen records (determine upload-time checks that may needed for images that contain links to specimen records, and how to report back to users information regarding linking); and dealing with duplicates and updates.
- 2014/Q2: Define API exposed by upload appliance through interactions with bio-collection tools, define roadmap integration with representative tool (e.g., Specify), and develop transport and

protocol mechanisms (e.g. XML-RPC, JSON-RPC; provide inputs, upload, query status, retrieve outputs).
- 2015/Q2: Deploy the capability for end-users to create and share VM appliances on the iDigBio cloud.

Ongoing Activities

- 2-3 Petabytes of storage will ultimately be required. ACIS has the lead role in seeking funding, collaborators, and resources to provide access to this quantity of storage.
- Consumption of available storage capacity will be monitored and reported monthly by ACIS to ensure that capacity is not exceeded due to unexpected usage.
- Network bandwidth usage will be monitored and reported monthly by ACIS to ensure that resources are not exceeded, causing a degradation in the user experience.
- Evaluation and implementation of selected Hosting and Appliance Development/Delivery requests.

7. COMMUNICATION MANAGEMENT

7.1. Strategy

Effective and open communication is essential to the success of the iDigBio project. Communication planning provides a framework for managing and coordinating the wide variety of communications that take place during the project. Communication management conveys clear, consistent, and timely information to stakeholders who can affect or may be affected by the objectives and outcome of the iDigBio project. The key communication objectives for the iDigBio project are:

- Engender transparency, consistency, and seamless collaboration
- Foster project awareness; promote iDigBio's mission and objectives
- Provide accurate and timely information about the project
- Ensure a consistent message

The key audiences of iDigBio communications are the project team, project leadership and management, and all internal and external stakeholders.

7.2. Resources

iDigBio maintains licenses for integrated collaboration and teleconferencing software to achieve the communication goals. iDigBio currently utilizes Adobe Connect (idigbio.adobeconnect.com) and MeetingOne (www.meetingone.com). iDigBio communication resources are periodically reviewed to ensure they continue to meet the needs of the project.

7.3. Internal Communication

Audience	Message	Method	Frequency	Communicator
Core Team	• Coordinate activities • Establish priorities • Support operations	• Meeting	Bi-weekly	Project Manager
IT Team	• Data integration • Computational needs • Assessment of new technologies/platforms to facilitate digitization efforts	• Meeting	Weekly	Cyberinfrastructure PI
Executive Committee (PIs)	• Project management and oversight • Accomplishment of mission and goals	• Meeting • Teleconference		Project Manager, Lead PI
Steering Committee	• Review progress & coordinate activities • Resource allocation • Strategic directions and management policies	• Meeting • Teleconference • Published minutes	Monthly	Project Manager
Internal Advisory Committee (IAC)	• Report on progress of digitization efforts • Share and identify best practices and standards • Identify gaps in digitization areas and technology • Enhance training efforts • Foster collaboration	• Meeting • Teleconference • Published minutes	Bi-monthly	Project Manager

7.4. External Communication

Audience	Message	Method	Frequency	Communicator
External Advisory Board (EAB)	• Advice on progress and integration of digitization projects, research, education, and outreach activities among all funded institutions • Advice on strategic directions and management policies	• Meeting • Teleconference • Published minutes	Yearly	Project Manager, PIs
NSF Program Officers	• Progress report • Future planning	• Teleconference	Bi-weekly	Lead PI
NSF	• Quarterly report • Semiannual report • Annual report	• research.gov	Quarterly, Semiannual, and Annual	Project Manager, PIs

8. CHANGE MANAGEMENT

8.1. Strategy

Change management planning provides a roadmap for decision making, approving, and reporting change when it occurs. Change occurs when the project's resources, costs, project plan, timeline, deliverables, specifications, and/or quality are affected by a particular course of action. Managing change requires planning, discipline, and effective communication among the project team, project leadership and management, and internal and external stakeholders. The key change management objectives for the iDigBio project are:

- Establish standardized methods, processes and procedures for all project changes
- Facilitate efficient and prompt handling of changes
- Balance the benefit of a change with the risk or impact of the change on the project

8.2. Process

1) Potential changes are identified, typically during meetings with team members, management, or other internal and/or external stakeholders. Sources of change include, but are not limited to:
    - Responses to internal project issues (e.g., bug fixes, budget reductions)
    - System enhancement requests
    - Changes in project requirements or strategy
    - Proactive changes to improve product performance, scope, or quality
2) The change is documented in the form of an Issue ticket in Redmine (www.idigbio.org/redmine).
3) The ticket is assigned to an appropriate Subject Matter Expert (SME) to analyze the impact of the change request on resources, schedule, budget and quality.
4) The SME analyzes the request and documents the impact of the change including feasibility of making the change, any impacts and/or risks associated with making (or not making) the change, and any impacts to the project schedule and budget.
5) The SME recommends whether to authorize, reject, or defer the change request.
    a. If approved, the Project Manager determines if the change is of significant magnitude based on the impact to project scope, resources, costs and/or schedule impact. Significant changes require additional review:
        i. Significant changes that are within the current project scope require assessment and approval by the Executive Committee (PIs).

ii. Significant changes that expand or modify the project scope require assessment and approval by the Steering Committee and NSF Program Officer.

b. If approved following Project Manager and higher-level review (if required), project documentation, schedule, and budget are updated as required to reflect the change. Impacted stakeholders are notified. The ticket is monitored and updated until the work is complete.

c. If rejected or deferred, the ticket is updated with the reasons for rejection or deferral and the ticket is closed.

9. RISK MANAGEMENT

9.1. Strategy

Risk management planning provides a systematic process of deciding how to approach, plan, and execute risk management activities throughout the life of the project, which serve to maximize opportunities and minimize threats to the project objectives. The project will use a qualitative approach to risk management including:

- Periodic monitoring for the realization of risks previously identified
- Notification of key stakeholders of any impacts of realized risks
- Continual management of stakeholder expectations

9.2. Definitions

Term	Definition	Measure or Strategy
Occurrence Likelihood	Qualitative assessment of the probability of the risk occurring (i.e., the risk being realized)	High, Medium, or Low
Cost Impact	Qualitative assessment of the impact to overall project budget	High, Medium, or Low
Schedule Impact	Qualitative assessment of the impact to overall project schedule	High, Medium, or Low
Trigger	Warning signs that indicate the risk is likely to occur or will occur soon;	Used to determine when to implement mitigation strategies
Mitigation Strategy	Specific actions intended to reduce the impact of the identified risk	• **Avoidance** seeks to eliminate the possible deviation by changing the project deliverables against which the deviation is defined. For a negative risk, this could mean deciding not to undertake the deliverable. For a positive risk or opportunity, this could mean exploiting the opportunity by incorporating it into the project

Term	Definition	Measure or Strategy
		as a planned deliverable. • **Mitigation** seeks to alter the likelihood or impact of the risk. For negative risks, steps may be taken to reduce the probability that risk factors will cause a deviation from the project plan or to reduce the amount of deviation. • **Acceptance** merely acknowledges the risk, but does not specify any action to take in response to the risk.

9.3. Process

1) Potential risks are identified, typically during meetings with team members, management, or other internal and/or external stakeholders. Risks are encountered in every aspect of project execution. Some factors that may affect the development of a project include

    - **Budget risks:** Risk that budget elements will deviate from the estimate
    - **Event risks:** Risk of internal or external events that affect the ability of the project team to meet the project deliverables on schedule
    - **Scope risks:** Risk of significant changes to the project scope due to external influences

2) Potential risks are documented in the form of an Issue ticket in Redmine (www.idigbio.org/redmine) with a "Risk" tracker.

3) The ticket is assigned to the Project Manager or an appropriate Subject Matter Expert (SME) to analyze the risk.

4) The Project Manager or SME analyzes the risk in terms of likelihood, cost impact, and schedule impact, and then documents the recommended mitigation strategy and anticipated risk triggers.

5) The Project Manager and/or Subject Matter Expert (SME) will periodically review documented risks for triggers that may cause a risk to be realized. This process is informal, but will be executed throughout the duration of the project.

6) For any identified trigger, the Project Manager will determine if the associated risk is realized and make a preliminary determination if mitigation strategies should be implemented.

    a. The Project Manager will consult with the Executive Committee and/or Steering Committee for risks with a significant impact to the project

    b. The Project Manager will ensure that appropriate action is taken to implement the mitigation strategy agreed upon for the realized risk.

c. The ticket will be updated with information regarding the identified trigger and the mitigation actions taken.

7) Key stakeholders and decision makers will be notified of realized risks through regularly communication.

10. QUALITY MANAGEMENT

10.1. Strategy

The purpose for managing quality is to validate that the project deliverables are completed with an acceptable level of quality. Quality management assures the quality of the project deliverables and the quality of the processes used to manage and create the deliverables. The key quality management objectives for the iDigBio project are:

- Project deliverables meet their requirements
- Project management processes are appropriately followed

10.2. Metrics

Quality is the degree to which the project deliverables and project processes fulfill their requirements. The entire project team shares responsibility for the quality of the deliverables and processes.

Objects of Quality Review	Quality Measure	Quality Evaluation Activities
Project Deliverables	• Deliverable Quality Standards • Customer Satisfaction	• Core Team meetings • Executive Committee meetings • Steering Committee meetings • Internal Advisory Committee meetings • Project surveys • NSF reporting
Project Processes	• Process Quality Standards • Stakeholder Expectations	• Develop/refine Implementation Plan • Periodic project reviews • Project surveys

11. PROJECT CLOSURE

The purpose of project closure is to ensure that all pertinent project records are identified, labeled, and properly maintained for easy retrieval at a later date. Project closure activities should be executed throughout the course of the work, not just at the end. Examples of closure activities include:

- Project evaluation, reward, and recognition
- Document lessons learned that can help improve future projects
- Archiving of project materials and documentation

APPENDIX A – IDIGBIO PROJECT PROGRESSION