Becoming an NBII/GBIF Data Provider

What is GBIF's role? What is NBII's role?

❖ The Global Biodiversity Information Facility, or GBIF, (http://www.gbif.org) is an international scientific co-operative project, based on a multilateral agreement (MoU), between countries, economies and international organizations, dedicated to:
  o establish an distributed information infrastructure containing primary biodiversity data,
  o with initial focus on species- and specimen-level data,
  o with links to molecular, genetic and ecosystems levels
  o in order to make the world's scientific biodiversity information freely and universally available to all
❖ The National Biological Information Infrastructure (NBII) (http://www.nbii.gov), as the designated U.S. Node to GBIF, is seeking educational and research institutions across the United States interested in becoming data providers in the US GBIF network. NBII seeks to provide access and highlight U.S. collections, in order to represent them and their needs to GBIF and the international community.
  o To facilitate this, NBII has established a "GBIF portal" for U.S. data providers, displaying data from participating U.S. institutions/organizations.
  o The portal can be found at http://gbif.nbii.gov

What is meant by becoming a Data Provider?

❖ Data providers are the institutions/organizations responsible for providing, through standard web exchange interfaces, metadata describing themselves and the data services they offer, as well as free access to biodiversity data. They also:
  o Agree to use a common data exchange format with a fixed structure which clearly defines how the information is to be shared.
  o Agree to exchange data in a way which makes it as simple as possible to compare and merge information from different resources.

How is data exchanged? What is DiGIR?

❖ Distributed Generic Information Retrieval (DiGIR), is a client/server protocol for retrieving information from distributed resources.
  o Uses HTTP as the transport mechanism and XML for encoding messages sent between client and server.
  o Three type of messages:
    ▪ Metadata: get metadata information of the provider and the resource(s) that serves.
    ▪ Search: find specimen and observation records based on search criteria, for example: the name of a species and/or a rectangle defining an area on the earth's surface and/or …
    ▪ Inventory: get the set of distinct values associated with a single concept, for instance: Species.
  o Maps database models of collections to Darwin Core2 (suitable for exchange of specimen and observation data).

❖ Advantages of DiGIR:
  o Provides a single point of access to one to many distributed information resources.
    ▪ Resource: a collection of data objects that conform to a common schema.
  o Enables search & retrieval of structured data.
    ▪ Search for data values in context (semantics)
    ▪ Results as structured data set
  o Makes location and technical characteristics of native resource transparent to the user.

Becoming an NBII/GBIF Data Provider

What is the primary data exchange format?

- ❖ Darwin Core2 is a model that allows data on individual specimens or observations to be structured and shared as XML documents that can be transmitted across the Internet.
    - o Suitable for collections and observations data.
    - o Only five required elements:
        - ▪ DateLastModified
        - ▪ InstitutionCode
        - ▪ CollectionCode
        - ▪ CatalogNumber
        - ▪ ScientificName
    - o Resources for more information:
        - ▪ http://digir.net/schema/conceptual/darwin/2003/1.0/darwin2.xsd
        - ▪ http://tsadev.speciesanalyst.net/documentation/ow.asp?DarwinCoreV2
- ❖ 48 Elements:

DateLastModified	InstitutionCode	CollectionCode	CatalogNumber
ScientificName	BasisOfRecord	Kingdom	Phylum
Class	Order	Family	Genus
Species	Subspecies	ScientificNameAuthor	IdentifiedBy
YearIdentified	MonthIdentified	DayIdentified	TypeStatus
CollectorNumber	FieldNumber	Collector	YearCollected
MonthCollected	DayCollected	JulianDay	TimeOfDay
ContinentOcean	Country	StateProvince	County
Locality	Longitude	Latitude	CoordinatePrecision
BoundingBox	MinimumElevation	MaximumElevation	MinimumDepth
MaximumDepth	Sex	PreparationType	IndividualCount
PreviousCatalogNumber	RelationshipType	RelatedCatalogItem	Notes

Components of the DiGIR Provider Package

- ❖ DiGIR Provider software, Apache2 WEB server and PHP libraries.
    - o Requires from the user only basic knowledge of the operating system.
- ❖ Two available releases:
    - o Linux (RedHat 7.3, 8, 9)
    - o MS Windows (2000, XP)
- ❖ Supported databases:
    - o MySQL
    - o PostgreSQL
    - o MS SQL Server
    - o MS Access (only the MS Windows package)

Becoming an NBII/GBIF Data Provider

Basic System Requirements

❖ At least 150 Mbytes free disk space on the drive you are performing the installation to. This is the minimum requirement for the installation. Extra disk space should be allocated for the application log files.

❖ Assuming a dedicated server, the system should have at least 256 Mbytes main memory (minimum installation requirement).

❖ WinZip 8.0 or higher

❖ It is recommended that the machine running the DiGIR Provider service has a fully qualified domain name (FQDN). If you do not have a FQDN for your server then you can use one of the many Dynamic DNS services to register a name for your server regardless of whether you have a static or dynamic IP address. One free service that has proven to be particularly stable is DynDNS.

❖ For A Proxy Host Name and Proxy HTTP port. Use the same values as the host name and the TCP/IP port number for the Apache WEB Server.

❖ An e-mail address for the administrator of the WEB server (e.g. helpdesk@stateu.edu).

❖ A password for the user admin of the WEB server. Using these credentials (admin/<password> you will be able to access the administrative part of the DiGIR Provider.

❖ For complete requirements, please see the "DiGIR Provider Package User's Guide" for either Windows or Linux.

Installation Process in a Nutshell

❖ Acquire a normal PC and place it safely outside your institution's firewall or make arrangements to use your institution's web server.

❖ Download a DiGIR provider package from GBIF.
   o http://www.gbif.org/serv/gbif-tools

❖ Install the package, and define your institution's contact details. (This can be done in mere minutes.)

❖ Register the new provider with NBII's UDDI registry.

❖ Link your database(s) to the provider. (This may require a little work - it depends on the shape of your databases. You can start small and add more data and several databases to the provider later.)

NBII Contact

❖ Mike Frame
   o Phone: 703-648-4164
   o Email: mike_frame@usgs.gov