1 Background Information

The Global Biodiversity Information Facility (GBIF) works to make digital biodiversity data openly and freely available on the Internet for everyone. GBIF provides access to scientific biodiversity data for decision-making, research endeavours and public use. GBIF endorses both open source software and open data access.

In terms of structure, GBIF is a dynamic, growing partnership of countries, organisations, institutions and individuals working together to mobilise scientific biodiversity data.

http://www.gbif.org/

GBIF is also a network of data publishers who retain ownership and control of the data they share. Linked datasets provide a more robust representation of biodiversity than any single dataset.

In this framework, GBIF develops software tools to make possible to easily integrate biodiversity data from heterogeneous sources using common standards and protocols. One of these tools is the GBIF Integrated Publishing Toolkit (IPT) that this user manual refers to.

The GBIF IPT is an open source, Java (TM) based web application that connects and serves three types of biodiversity data: taxon primary occurrence data, taxon checklists and general resource metadata. The data registered in a GBIF IPT instance is connected to the GBIF distributed network and made available for public consultation and use. Several factors have motivated the development of this new tool:

- the limitations of the existing systems to publish primary biodiversity occurrence records that use tools and protocols that were not designed to transfer big amounts of records (DiGIR, TAPIR, ABCD);
- the need to decentralise and speed up the process of indexing biodiversity occurrence datasets;
- the need to offer additional benefits and services to the data publishers to encourage data publication (i.e. public web interface to browse/expose (meta)data, mapping services);
- the need to offer data quality assessment tools to data publishers and generalise data cleaning habits;
- the need to increase data exchange using globally accepted formats (i.e. OGC web services);
- the need for easy ways to monitor data access and use for the different resources connected to the GBIF Network;
- the lack of appropriate tools to publish other types of biodiversity data, such as names checklists and dataset metadata;
existing requests to facilitate and automate dataset registration with the GBIF Network;

- the requirement to implement (and assign if needed) Universal Unique Identifiers (UUIDs) associated to all published elements to increase interoperability in the GBIF Network;

The GBIF IPT, as an open source project, is made freely available for both those who wish to use it to publish their data and those who wish to participate in its development, documentation and helpdesk. This is a list of the current online resources related to the GBIF IPT:

- the GBIF IPT mailing list: [http://lists.gbif.org/mailman/listinfo/ipt/]
- the GBIF IPT test installation (for testing purposes): [http://ipt.gbif.org/]
- the GBIF IPT code site: [http://code.google.com/p/gbif-provider-toolkit/]

2 General Specifications

The GBIF Integrated Publishing Toolkit (IPT) is an Open source Java based web application. It embeds its own database, is easily customisable and is multilingual. It has a user management feature based on roles, which allows for multiple data managers to share a common instance.

It connects and serves three types of data: taxon primary occurrence data, taxon checklists and general resource metadata. The IPT manages multiple data sources and has several upload options: relational database management systems or text files. It does not however accept manual entries (except for individual metadata descriptions).

Its public web interface allows for data browsing and full text search. It also offers customised detail pages for the three types of data, with specific functionality for each type (i.e. density maps, metadata edition).

The IPT manages Universally Unique Identifiers (UUID), recycling existing IDs or providing new ones, as appropriate. It also manages technical metadata about the installation and allows for an easy setup of Google Analytics for usage statistics.

3 Installation

3.1 Requirements

The IPT will run on any of the most widespread Operating Systems (Linux, Mac OS (TM), MS Windows (TM)) provided that:

- Java (TM) runtime environment ([http://java.sun.com/]) version 5 or more is installed;
3.2 Brief overview of Java and Tomcat installation and configuration

It is beyond the scope of this manual to provide detailed installation and configuration instructions for the virtual machine and the server. Here is a quick overview for MS Windows (TM) and Mac OS (TM) computers.

3.2.1 Java

Mac OS X (TM) already comes complete with a fully configured and ready-to-use Java runtime environment.

   For MS Windows (TM) machines, follow these steps to install it:

1. Download the Java 2 Standard Edition Runtime Environment (JRE), release version 5.0 or later, from [http://java.sun.com/j2se](http://java.sun.com/j2se).

2. Install the JRE according to the instructions included with the release.

3. Set an environment variable named JRE_HOME to the pathname of the directory into which you installed the JRE, e.g. `c:\jre5.0`:

   (a) Open the Control Panel and click the System icon.

   (b) Go to the Advanced pane and click the Environment variables button.

   (c) In the System variables section, click New. In the Variable name box, type JRE_HOME; in the Variable value box, type the path to the JRE e.g. `C:\Program Files\Java\jre6`

3.2.2 Application server

The IPT should run in any compliant java application server using Java 1.5. Tomcat 5.5 is recommended for the installation. It was also successfully tested on Tomcat 6 and Jetty.

The application server will need a rather large amount of memory allocated as the IPT runs a full database. It is recommended to increase the server’s available memory to at least 1GB (general instructions below).

Tomcat installation

1. Download a binary distribution of Tomcat from [http://tomcat.apache.org/download-55.cgi](http://tomcat.apache.org/download-55.cgi)
2. Unpack the binary distribution into a location of your choice (e.g. C:\apache-tomcat-5.5.27)

3. Open a command prompt window and navigate to the installation folder.

4. Mac OS only: some files need to have their permissions adapted: type `sudo chmod 775 *` at the command line. You will be asked your administrative password.

5. go to the `bin` folder and execute `startut.bat` (MS Windows (TM)) or `startup.sh` (Mac OS (TM)). A new Tomcat window will open. You need to keep this window open.

**Tomcat configuration**

Tomcat comes with very little allocated memory. In order to increase performance, this amount should be increased depending on the amount of physical memory on the server. If the server has at least 2GB of RAM, the available memory should be increased to 1GB. For more information on how to do it, see the following links:

- [http://explanatorygap.net/2005/03/06/configuring-tomcat-under-os-x-server-for-more-memory/](http://explanatorygap.net/2005/03/06/configuring-tomcat-under-os-x-server-for-more-memory/)
- [http://wiki.apache.org/tomcat/FAQ/Memory](http://wiki.apache.org/tomcat/FAQ/Memory)

### 3.3 IPT Installation

1. Download the web archive file (ipt-1.0.war at the time of writing) from the project downloads page: [http://code.google.com/p/gbif-providertoolkit/downloads/list](http://code.google.com/p/gbif-providertoolkit/downloads/list)

2. Rename it to `ipt.war` and copy it into the `webapps` folder of the Tomcat installation directory.

3. Launch Tomcat:

   (a) open a command prompt (MS Windows(TM)) or a terminal window (Mac OS (TM))

   (b) navigate to the Tomcat installation folder

   (c) go to the `bin` folder and execute `startut.bat` (MS Windows (TM)) or `startup.sh` (Mac OS (TM))

4. The IPT can then be accessed in any browser at http://localhost:8080/ipt/
3.4 Geoserver Installation

GeoServer ([http://geoserver.org/](http://geoserver.org/)) is an open source software server written in Java that allows users to share and edit geospatial data. It is used to run IPT geographic web services via a plugin that can access the IPT cache. If there is already a geoserver installation on the server, only the plugin needs to be installed. If there is no such installation, a geoserver archive bundled with the plugin is provided.

3.4.1 Geoserver and bundled plugin installation

This installation is required if the Geoserver is not yet installed on the machine.

1. Download the zipped archive from the IPT project downloads page;
2. Extract it into the Tomcat webapps folder and rename it as `geoserver`
3. Restart Tomcat (shutdown.bat then startup.bat on MS Windows (TM), shutdown.sh and startup.sh on Mac OS(TM)).

3.4.2 Plugin installation

If the Geoserver is already installed on the server, only the plugin needs to be installed.

1. Download the jar file from the IPT project downloads page;
2. Install it in your existing Geoserver installation, in the geoserver/WEB-INF/lib folder;
3. Restart Tomcat (shutdown.bat then startup.bat on MS Windows (TM), shutdown.sh and startup.sh on Mac OS (TM)).

3.5 Geoserver Configuration

In order to allow the IPT and the Geoserver to communicate to one another, it is necessary to define the IPT Base Directory. It is done by:

- navigating with your web browser to http://localhost:8080/ipt/;
- logging in with username `admin` and password `admin`;
- navigating to `Settings > Settings > IPT Base Directory`;
- pressing the `Update Geoserver` button.

This procedure will be explained with more details in the `Configuration` section of this document.

For more information about the Geoserver, please refer to its own manual: [http://geoserver.org/](http://geoserver.org/)
3.6 Mac OS X (TM) test installation

A standalone version without any dependencies is provided for Mac OS X (TM). It is very easy to install as it only requires a simple click and is very convenient to test the IPT. It should not however be used for production purposes. The disk image can be downloaded at the time of writing from the following address: ftp://ftp.gbif.org/projects/ipt/downloads/ipt-1.0b.8.dmg

4 Configuration

Once the IPT has been installed, it is accessible in a web browser at the following URL: http://localhost:8080/ipt/

To configure the software, one must login as administrator by clicking on the Login button in the main menu and entering a username and password. By default, the administrative username is admin and the password is also admin.

Once logged in, a new Admin section appears in the menu. The subsequent drop-down menu offers the following choices:

- Settings
- Extensions
- Thesaurus
- View Users
- Current Users
- Active Datasources
- Background Tasks

4.1 Settings

The Settings section is divided in three configuration options:

- Settings
- Organisation
- IPT Instance
4.1.1 Settings

The Settings page allows the administrator to define a number of parameters that are essential for the IPT to function correctly.

**IPT Base URL**

This is the URL of the IPT. It is filled in by default with the local URL (http://localhost:8080/ipt). In order to make the IPT instance accessible by distant users on a network, the parameter should be changed to the networked URL (e.g. http://67.233.45.122:8080/ipt or http://exemple.com:8080/ipt).

**IPT Base Directory**

This field defines the location of the IPT in the file system. It is filled in by default and does not need to be changed. It is followed by a button that needs to be clicked in order for the Geoserver to know where to access the IPT. If the IPT is moved or renamed, it is necessary to renew the location by pressing the *Update Geoserver* button.

**Google Maps API Key**

In order to use the the Google Maps (TM) feature embedded in the IPT pages, it is necessary to own a Google Maps API Key. This free key is specific to a server and needs to be requested for each public installation. The administrator can get such a key for free by clicking on the *Get Google Maps API key* link and following instructions provided by Google.

4.1.2 Geoserver Settings

The Settings page also allows the administrator to define some parameters for the Geoserver. For more information about the Geoserver, see its users guide: [http://geoserver.org/display/GEOSDOC/Users+Guide](http://geoserver.org/display/GEOSDOC/Users+Guide)

**Base URL**

The Base URL is the URL of the Geoserver. It is filled in by default with the local Geoserver URL (http://localhost:8080/geoserver). In order to make the Geoserver accessible by distant users on a network, the parameter should be changed to the networked URL (e.g. http://67.233.45.122:8080/geoserver or http://exemple.com:8080/geoserver).

**Data Directory**

The Data Directory is the location in the server file system of the directory where data reside. It is filled in by default but can be modified manually.

**Admin User and password**

By default, the administrative username is *admin* and the administrative password is *geoserver*. These parameters *must* be changed if the server is to be used for production and/or accessible to the public.
4.1.3 Organisation

This page allows the administrator to define the GBIF organisation hosting the IPT installation. It is necessary to define the organisation in order to register the IPT. The GBIF keeps a directory of organisations, resources and services and a password is necessary to register the organisation or to access its data.

The following fields must be filled in:

**Title:** this is the full name of the organisation

**Organisation Identifier:** this is a code provided automatically by the GBIF once the organisation is registered filling and sending this form

**GBIF Node:** GBIF Participant Node that endorses the publication of the data in this installation to be connected to the GBIF Network

**Organisation password:** this password can be used by other members of the organisation to retrieve the information

**Contact:** Name: Name of the contact person

**Contact:** Email: Email of the contact person

**Homepage:** URL of the organisation’s website

**Description:** Description of the organisation

**Cancel / Save / Register:** The three buttons allow the administrator to respectively delete what he just typed, saved the information in the form or register the organisation with the GBIF. When does so, he will receive an Organisation Identifier that will be automatically filled in the right field.

4.1.4 IPT Instance

This page allows the administrator to define some parameters regarding the IPT instance that is being run. The information filled in on this page will be used to populate the homepage of this IPT instance.

**Organisation Identifier:** this is the code provided automatically by GBIF during the organisation registration process

**Title:** this is the name given to the IPT instance. It can be the name of an organisation, a project, a natural history collection, etc.

**Contact** Name: Name of the contact person

**Contact** Email: Email of the contact person
4.2 Extensions

Extensions allow for serving additional data types, such as multiple identifications or descriptive data about species.

The page provides a list of available extensions and shows whether they are installed or not. Each extension can be installed or uninstalled by clicking on its name.

By default, available extensions are:

- ResourceRelationship
- SampleAttribute
- EventAttribute
- Identification
- Description
- Multimedia
- CommonName
- Distribution
- MaterialExamined
- DarwinCore
- Taxon
- VernacularName

The user can get more information about an extension and install it by clicking on its name.

It is possible to define custom extensions, but this is not automated process with this the 1.0 version of the IPT.
4.3 Thesaurus

Thesaurus vocabularies allow for specifying controlled vocabularies to help standardise concepts. This page shows a list of the available thesauri. More information can be obtained by clicking on the names under the Vocabulary column.

By default, available thesauri are:

- ISO 639-1 Languages
- Country Codes
- Language Vocabulary
- Darwin Core Type Vocabulary
- Nomenclatural Codes
- Taxonomic Ranks

It is possible to define custom vocabularies, but this is not an automated process with version 1.0 of the IPT.

4.4 View Users

This page allows the administrator to manage users. There are three user roles: administrator, manager and user.

The page displays the list of users their usernames, full names and emails. By clicking on the username, one can edit the user profile. One can also export the list in different formats (CSV, MS Excel (TM), XML and PDF).

The page also gives the opportunity to add a new user through the Add button at the bottom of the page.

4.4.1 User Profile

The User Profile page allows the administrator to add a new user or to edit or delete an existing user.

To add a new user, the following fields must be filled in:

username: user’s username
password: needs to be filled in twice for confirmation
e-mail
password hint: will be e-mailed to the user who forgot his password
first name: user’s first name
last name: user’s last name

The administrator must also apply some settings to the account:

**Enabled**: the account is enabled and ready to be used

**Expired**: the account is not valid anymore and cannot be used to login to the IPT

**Locked**: the account is locked and cannot be used to login to the IPT

**Password Expired**: the user must change his password next time he logs in.

Finally, the administrator must assign one or several roles to the account. The following roles are available:

**ROLE_ADMIN** the user is given administrative rights and has access to all the features of the IPT, in particular its technical settings. He can also manage the data recorded in the IPT. The administrator supersedes all the other roles and can do anything any other role can do.

**ROLE_MANAGER** the user is given manager rights and can manage the data recorded in the IPT. He can create, edit, delete, or publish data.

**ROLE_USER** the user role is the most basic role of the IPT. The user can explore the site and browse the published information. He can also annotate resources.

A visitor that is not logged in is considered to be an *anonymous user*. Such a user can explore the site and browse the published information.

### 4.5 Resources

This page displays a list of all resources that have been uploaded to the IPT.

In order to find the right resource, the administrator can order the list by clicking on the table headers. Resources can be ordered by title, date of modification, UUID or Type. A second click on the header will invert the order.

The administrator can get more details about each resource by clicking on its title. The administrator can then modify the UUID, the contact name and contact email. He also has access to the registered services.

### 4.6 Current Users

This page displays a list of the currently logged in users on the IPT. The administrator is offered the possibility to send them an e-mail.
4.7  Active Datasources
This page displays a list of all the active datasources in the IPT.

5  Primary Occurences
This section is not available for this version of the manual

6  Name datasheets
This section is not available for this version of the manual

7  Metadata
This section is not available for this version of the manual