

The GBIF Integrated Publishing Toolkit User Manual

GBIF Secretariat

May 2009

Published by:

Global Biodiversity Information Facility

<http://www.gbif.org/>

Copyright notice:

This user manual is released under a '*Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported License*'

<http://creativecommons.org/licenses/by-nc-sa/3.0/>

In general terms, that means that you are free to copy, distribute, transmit, remix and adapt the work, under the following conditions:

- you must attribute the work in the manner specified in this page (but not in a way that suggests that GBIF or the GBIF Secretariat endorses you or your use of the work);
- you may not use this work for commercial purposes;
- if you alter, transform, or build upon this work, you may distribute the resulting work only under the same or similar license to this one.

Recommended citation:

Réveillon, A. 2009. *The GBIF Integrated Publishing Toolkit User Manual*, version 1.0. Copenhagen: Global Biodiversity Information Facility. 37 pp.

Disclaimer:

While the editor, authors and the publisher have attempted to make this user manual as accurate and as thorough as possible, the information contained herein is provided on an "as is" basis, and without any warranties with respect to its accuracy or completeness. The editor, authors and the publisher shall have no liability to any person or entity for any loss or damage caused by using the information provided in this user manual.

Contents

1	Background Information	5
2	General specifications	6
3	Installation	6
3.1	Requirements	6
3.2	Brief overview of Java and Tomcat installation and configuration	7
3.2.1	Java	7
3.2.2	Application server	7
3.3	IPT Installation	8
3.4	Geoserver Installation	9
3.4.1	Geoserver and bundled plugin installation	9
3.4.2	Plugin installation	9
3.5	Geoserver Configuration	9
3.6	Mac OS X (TM) test installation	10
4	Configuration	10
4.1	Configuration	10
4.1.1	Settings	11
4.1.2	Geoserver Settings	12
4.1.3	Organisation	13
4.1.4	IPT Instance	15
4.2	Extensions	16
4.3	Thesaurus	17
4.4	Manage Users	18
4.4.1	User Profile	19
4.5	Current Users	20
5	Resources Management	21
5.1	Data Types	21
5.1.1	Taxon Primary Occurrence Data	21
5.1.2	Taxonomic Checklists	21
5.1.3	Resource Metadata	21
5.2	Resource Creation	21
5.3	Resource Metadata	21
5.3.1	Basic Metadata	22
5.3.2	Resource Originator	23
5.3.3	Geographic Coverage	25
5.3.4	Taxonomic Coverage	26
5.3.5	Temporal Coverage	27
5.3.6	General Keywords	27
5.3.7	IPRights	27

5.3.8	Research Project	29
5.3.9	Methods	30
5.4	Data Resource Configuration (occurences and checklists only)	30
5.4.1	Source Data	30
5.4.2	Mappings	33
5.5	Resource Publication	37

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.

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 it 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 or mapping services);
- the need to offer data quality assessment tools to data publishers and to 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 certain 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-providertoolkit/>.

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 tab 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 details 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 (TM) 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 higher is installed;

- a web server with a servlet container, such as Apache Tomcat (<http://tomcat.apache.org/>) or Jetty (<http://jetty.mortbay.com/>), is installed and connected to the internet.

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) includes a fully configured and ready-to-use Java runtime environment. For MS Windows (TM) machines, follow these steps to install:

1. Download the Java 2 Standard Edition Runtime Environment (JRE), release version 5.0 or later, from <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 require 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 (see general instructions below).

Tomcat installation

1. Download a binary distribution of Tomcat from <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 for your administrative password.
5. Go to the *bin* folder and execute startut.bat (Window) or startup.sh (Mac OS). 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 improve performance levels, this amount should be increased depending on the 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 this, see the following links:

- <http://explanatorygap.net/2005/03/06/configuring-tomcat-under-os-x-server-for-more-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>
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 startup.bat (windows) or startup.sh (Mac OS)
4. The IPT can then be accessed in any browser at <http://localhost:8080/ipt/>

3.4 Geoserver Installation

GeoServer (<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 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 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. This 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/>

3.6 Mac OS X (TM) test installation

A standalone version without any dependencies is provided for Mac OS X. 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. At the time of writing, the disk image can be downloaded from the following address: <ftp://ftp.gbif.org/projects/ipt/downloads/ipt-1.0b.8.dmg>

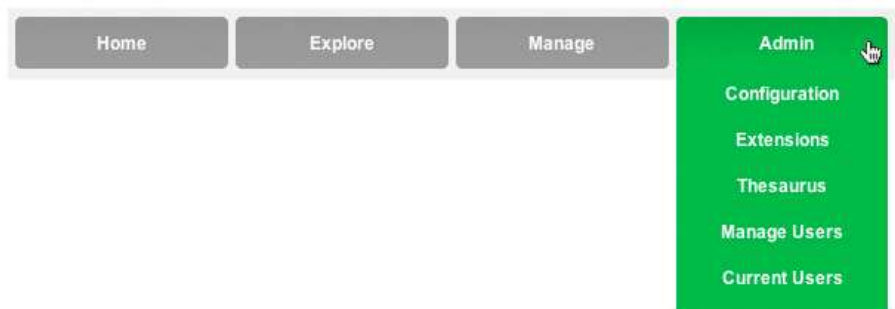
4 Configuration

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

To be able 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:

- Configuration
- Extensions
- Thesaurus
- Manage Users
- Current Users



4.1 Configuration

The Configuration section is divided into three configuration options available via the menu on the right hand side of the page:

- Settings

- Organisation
- IPT Instance

[Configuration](#)
[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.

Settings

IPT Base URL *

IPT Data Directory

Logging

Enable GBIF Usage Tracking

Google Maps API Key

[Get Google Maps API key](#)

Additional raw HTML headers

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 public URL (e.g. `http://67.233.45.122:8080/ipt` or `http://example.com:8080/ipt`).

IPT Data Directory

This field defines the location of the IPT data folder in the file system. It is filled in by default and does not need to be changed. It is followed by a

button that must be clicked 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.

Logging

This option is used for debugging purposes. It should be left on *production.xml* except in case of bugs and errors, in which case it can be moved to help developers in their debugging.

Before changing this setting, the administrator should contact the IPT developers.

Enable GBIF Usage Tracking

The GBIF usage tracking box is used to track the usage of all IPT instances.

Google Maps API Key

In order to use 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.

Additional raw HTML headers

This field allows the manager to add some HTML headers to all pages of the IPT instance. It can be useful to add keywords or usage tracking (e.g. Google Analytics).

The content must be XHTML compliant and will be included in the *HEAD* section of every single HTML page.

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>

Geoserver Settings

Base URL *	
<input type="text" value="http://ipt-geo.gbif.org"/>	
Data Directory *	
<input type="text" value="/usr/share/tomcat5/webapps/geoserver/data"/>	
Admin User *	Password *
<input type="text" value="admin"/>	<input type="text" value="geoserver"/>

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 a public URL (e.g. `http://67.233.45.122:8080/geoserver` or `http://example.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.

Organisation

Please select your registered organisation (the title form unlocks after all organisations have been downloaded from the GBIF registry), or enter a new organisation in case you dont find yourself. Please enter the contact in charge of maintaining installations. This contact will receive a password required for further configurations. If you have already selected an organisation but you want to enter a new one, please click the "clear form" link at the bottom.

Title *		
<input type="text" value="GBIF Secretariat"/>		
Organisation Identifier	GBIF Node *	Organisation Password
<input type="text" value="8D483710-2056-11DE-8710-F27EF57405F3"/>	<input type="text" value="Denmark"/>	<input type="text" value="xbFvCjYzm1Q"/>
Contact Name *	Contact Email *	
<input type="text" value="Markus Dring"/>	<input type="text" value="mdoering@gbif.org"/>	
Homepage		
<input type="text" value="http://www.gbif.org"/>		
Description		
<input type="text"/>		
<input type="button" value="Save"/> <input type="button" value="Cancel"/>		Clear form Resend Password

The following fields must filled in:

Title Full name of the organisation

Organisation Identifier 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 was just typed, save the information in the form or register the organisation with the GBIF. When doing so, an

Organisation Identifier will be sent that will automatically fill the right field. The *Register* button is only available if the organisation is not yet registered.

Clear form This link is only available if a registered organisation is already defined. It allows the administrator to clear the form in order to register a new organisation.


Resend Password This link is only available if a registered organisation is already defined. An e-mail with the organisation's password is sent to the organisation's contact person when the link is clicked.

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 feed the homepage of this IPT instance.

IPT Instance

This description will be presented on the homepage. HTML markup is allowed in the description. The description image url should be a valid image URL that is being display on the homepage too.

Resource Identifier	
Not registered with GBIF	
Title *	
Welcome to IPT	
Contact Name *	Contact Email *
John Smith	john.smith@mailinator.com
Homepage	
Latitude	Longitude
55.662095	12.573853
	
Description *	
<pre><h2>You have succesfully installed the GBIF Integrated Publishing Toolkit</h2> <p>To finish the installation of the tool, you need to perform the following tasks:</p> <ol style="margin-left:20px"> Log in to your GBIF IPT instance with the administrator account using the button above (default u: admin p: admin), go to Admin &t; Settings and fill the information required. Make sure that you click on the Update Geoserver button. Register your IPT instance in the GBIF registry, selecting your national/regional/thematic GBIF Participant Node using the IPT registration feature. You can create a new organisation or select an existing one. Configure your IPT homepage to replace this default text. The IPT supports role based security. Administrators can grant manager roles to registered users who can then manage and publish resources. <h2>IPT Resources</h2> <p>If you need help regarding the GBIF IPT, try making use of the following resources:</p></pre>	
Image URL	
http://farm4.static.flickr.com/3411/3236810438_c882757150.jpg	
<input type="button" value="Save"/> <input type="button" value="Cancel"/> <input type="button" value="Register"/>	

Resource Identifier This is a code provided by the GBIF when the IPT instance is registered.

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

Homepage URL of the organisation's website

Latitude and Longitude The coordinates will be used to display the right portion of the Google Maps.

Description Description of the organisation

Image URL URL of the image used on the front page of the IPT instance. This URL must be public.

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.

Available Extension

Extensions allow for serving additional data types, such as multiple identifications or descriptive data about species. It is possible to define custom extensions, but this is not automated process with this version of the IPT. Please contact portal@gbif.org if you wish to define and register custom extensions.

[Check for all available extensions](#)

Extension Name	Properties	Installed	Documentation
Related Resources	22	<input type="radio"/> no	<input type="radio"/> unavailable
Event Attribute	8	<input type="radio"/> no	<input type="radio"/> unavailable
Identification History	9	<input type="radio"/> no	<input type="radio"/> unavailable
Alternative Identifier	2	<input type="radio"/> no	<input type="radio"/> unavailable
Resource Relationship	6	<input type="radio"/> no	<input type="radio"/> unavailable
Sample Attribute	8	<input type="radio"/> no	<input type="radio"/> unavailable
Multimedia	5	<input type="radio"/> yes	<input type="radio"/> unavailable
Vernacular Name	5	<input type="radio"/> yes	<input type="radio"/> unavailable
Darwin Core	124	<input type="radio"/> yes	<input type="radio"/> view info

By default, available extensions are:

- [ResourceRelationship](#)
- [SampleAttribute](#)
- [EventAttribute](#)
- [Identification](#)
- [Description](#)

- Multimedia
- CommonName
- Distribution
- MaterialExamined
- DarwinCore
- Taxon
- VernacularName

Clicking on an extension name will open a new page with information about the extension, including a list of properties. The administrator can get more information about a property by clicking on its name. It is also possible on this page to install or remove an extension.

EventAttribute Extension

Extension Name	EventAttribute
Table Name	dwc_event_attribute
Documentation	<ul style="list-style-type: none"> ○ unaviable
Installed	<ul style="list-style-type: none"> ● no
Properties	<ul style="list-style-type: none"> EventAttributeID EventAttributeType EventAttributeValue EventAttributeAccuracy EventAttributeUnit EventAttributeDeterminedDate

Qualified Name: <http://rs.tdwg.org/dwc/terms/EventAttributeDeterminedDate>
 Namespace: <http://rs.tdwg.org/dwc/terms/>
 Group:
 Name: EventAttributeDeterminedDate

EventAttributeDeterminedBy
EventAttributeRemarks

It is also possible to define custom extensions, but the process is not automated in version 1.0 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 thesauruses.

Thesaurus vocabularies

Thesaurus vocabularies allow for specifying controlled vocabularies to help standardise concepts. It is possible to define custom vocabularies, but this is not automated process with this version of the IPT. Please contact portal@gbif.org if you wish to define and register thesaurus vocabularies.

[Check for all available thesaurus vocabularies](#)

Vocabulary	URI
ISO 639-1 Languages	http://iso.org/639-1
ISO 3166 Country Codes	http://iso.org/iso3166
Darwin Core Type Vocabulary	http://rs.tdwg.org/dwc/dwctype/
TDWG Nomenclatural Codes	http://rs.tdwg.org/ontology/voc/TaxonName.rdf#NomenclaturalCodeTerm
Taxonomic Ranks	http://rs.tdwg.org/ontology/voc/TaxonRank

By default, the available thesauruses are:

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

Clicking on a thesaurus name will open a new page with information about the thesaurus, including a list of terms.

TDWG Nomenclatural Codes Vocabulary

To participate in the discussion and definition of this vocabulary please visit:

Download vocabulary as [thesaurus xml](#)

Identifier	URI	Issued
icbn	http://rs.tdwg.org/dwc/dwctype/ICBN	Dec 17, 2008
iczcn	http://rs.tdwg.org/dwc/dwctype/ICZN	Dec 17, 2008
bc	http://rs.tdwg.org/dwc/dwctype/BC	Dec 17, 2008
icncp	http://rs.tdwg.org/dwc/dwctype/ICNCP	Dec 17, 2008
biocode	http://rs.tdwg.org/dwc/dwctype/BioCode	Dec 17, 2008

It is also possible to define custom vocabularies, but the process is not automated in version 1.0 of the IPT.

4.4 Manage Users

This page allows the administrator to manage users. It displays a list of users with their usernames, full names and emails. Clicking on the username edits the user profile. The list can also be exported using different formats (CSV, Excel, XML and PDF).

Users

16 users found, displaying all users.

Username	FullName	Email	
admin	Admiral Adminska	ipt-admin@mailinator.com	<input checked="" type="checkbox"/>
areve	Arnaud Réveillon	areve23@gmail.com	<input checked="" type="checkbox"/>
atalavan	Alberto González-Talaván	atalavan@gbif.org	<input checked="" type="checkbox"/>

Export options: CSV | Excel | XML | PDF

The administrator can also 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.

User Profile

The screenshot shows a web form titled "User Profile". It contains several input fields and sections:

- Username ***: A single-line text input field.
- Password *** and **Confirm Password ***: Two side-by-side text input fields.
- E-Mail ***: A single-line text input field.
- Password Hint ***: A single-line text input field.
- First Name *** and **Last Name ***: Two side-by-side text input fields.
- Account Settings**: A section containing four checkboxes: Enabled, Expired, Locked, and Password Expired.
- Assign Roles**: A section with two columns: "Available Roles" (containing ROLE_ADMIN, ROLE_MANAGER, ROLE_USER) and "Current Roles" (containing ROLE_USER). Between the columns are buttons: >>, All >>, <<, and All <<.
- At the bottom are **Save** and **Cancel** buttons.

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

Username (mandatory field)

Password (mandatory field) needs to be filled in twice for confirmation

E-mail (mandatory field)

Password hint (mandatory field) will be e-mailed to the user who forgot his/her password

First name (mandatory field)

Last name (mandatory field)

The administrator must also apply some settings to the account:

Enabled The account is enabled and ready to be used

Expired The account is no longer valid 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 Expired the password must be changed at next login.

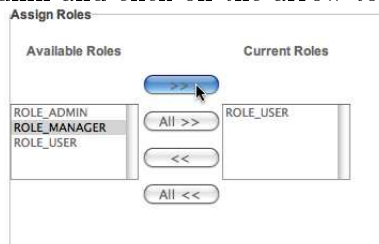
Finally, the administrator must assign one or several roles to the account. The following roles are available:

ROLE ADMIN the user is given administrative rights and has access to all the features of the IPT, in particular its technical settings. The administrator can also manage the data recorded in the IPT. This role 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, and 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, browse the published information and annotate resources.

To assign a user a role, the administrator must select it in the *Available Roles* column and click on the arrow to move it in the *Current Roles* column.



Conversely, to remove a role from a user, the administrator must select it in the *Current Roles* column and click on the arrow to move it back to the *Available Roles* column.

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 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.



5 Resources Management

5.1 Data Types

5.1.1 Taxon Primary Occurrence Data

Taxon Primary Occurrence Data (or simply, occurrence data) is defined in the scope of GBIF as digital text or multimedia data record detailing facts about the instance of occurrence of an organism, i.e. on the what, where, when, how and by whom of the occurrence and the recording.

In terms of connecting data to the IPT, an occurrence dataset is defined as a table or database storing information about the distribution of biological organisms in the form of points defined by a textual description and/or geographical coordinates.

5.1.2 Taxonomic Checklists

Taxonomic checklist data refers to structured data providing information about taxonomic names and their relationships.

5.1.3 Resource Metadata

In the context of GBIF, metadata records are those which provide information about the suppliers of biodiversity data and about the origins and purpose of those data.

As regards the IPT, a metadata record is composed of all the descriptive information that must be attached to a dataset in terms of origin (who produced the original data, who is connecting it to the IPT instance), contents (both taxonomically and geographically), intellectual property rights, etc.

5.2 Resource Creation

New resources are created by a manager (or an administrator) on the *Manage* page.

The manager chooses which resource type to create by clicking on the corresponding link in the *Resource Actions* menu: *New Occurrence Resource*, *New Checklist Resource* or *New Metadata Resource*.

5.3 Resource Metadata

When the resource has been created, the manager can add its metadata.

5.3.1 Basic Metadata

Basic Metadata

Please describe the dataset you want to publish as a whole

Title *

Type *

Fossil Specimen

Contact Name * **Contact Email** * **Upload logo (68x68px)**

 no file selected

Description

The first step is to basically describe the dataset that will be published. The manager must fill in the following fields:

Title Title of the resource

Type The type of resource data. The list depends on the data type (*Occurrence*, *Checklist* or *Metadata*):

Checklist A resource describing a checklist.

Fossil Specimen A resource describing a fossilised specimen.

Legislative Checklist

Living Specimen A resource describing a living specimen.

Moving Image

Nomenclature Checklist A resource describing a nomenclatural checklist.

Organism Observation A resource describing an observation.

Organism Observation by Human A resource describing an observation made by one or more people without accompanying physical evidence.

Organism Observation by Machine A resource describing an observation made by a machine without accompanying physical evidence.

Preserved Specimen A resource describing a preserved specimen.

Regional Checklist A resource describing a taxonomic checklist.

Sound Recording

Species Descriptions A resource describing a general or detailed description of a species.

Species Distributions A resource describing the geographic distribution of a species.

Specimen

Still Image

Taxonomic Checklist A resource describing a taxonomic checklist.

Contact Name

Contact Email

Upload logo The manager can upload a logo associated with the resource. The logo must be a square image, 68 pixels by 68 pixels.

Description Description of the resource

Once the fields have been filled in and the form saved, the resource is created and can be accessed from the *Manage* section.

5.3.2 Resource Originator

Resource Originator

Resource Language *

First Name *	Last Name *
<input type="text"/>	<input type="text" value="Markus Döring"/>
Organisation	Position
<input type="text"/>	<input type="text"/>
Phone	Email *
<input type="text"/>	<input type="text" value="mdoering@gbif.org"/>
Homepage	
<input type="text"/>	
Address	
<input type="text"/>	
Postal Code	City
<input type="text"/>	<input type="text"/>
Province	Country *
<input type="text"/>	<input type="text"/>

This section describes the originator of the resource, i.e. the person responsible for the information entered in the IPT. The following fields are to be filled in:

Resource Language (mandatory field) The language of the resource data
(not the language of the metadata)

First Name (mandatory field)

Last Name (mandatory field)

Organisation

Position

Phone

Email (mandatory field)

Homepage

Address

Postal Code

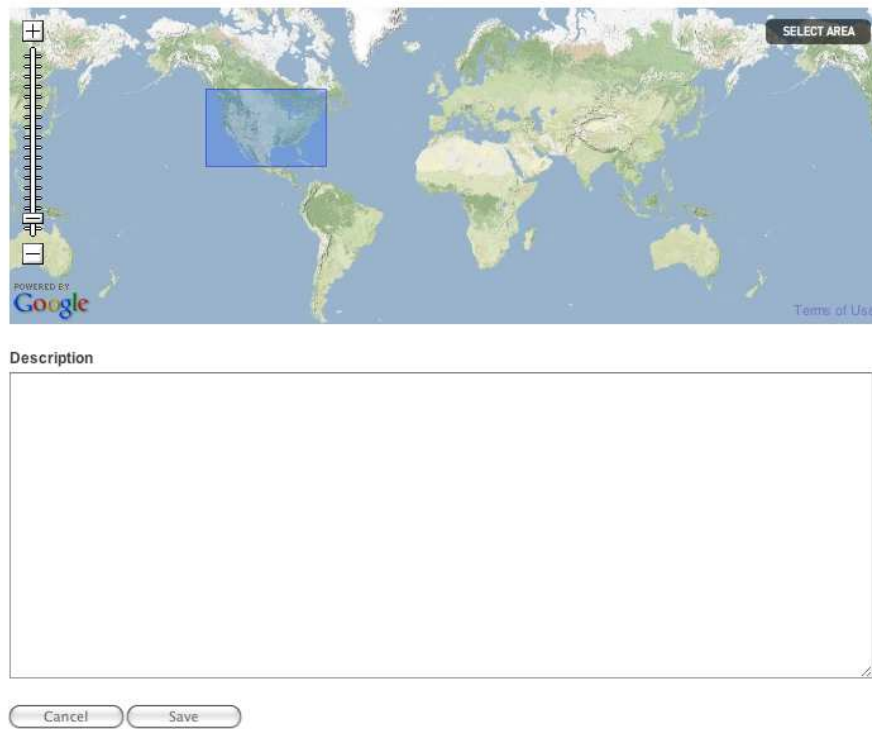
City

Province

Country (mandatory field)

5.3.3 Geographic Coverage

Geographic Coverage



In this section, the manager can describe the geographic coverage of the resource by drawing it on a map.

Using the map

The manager can navigate on the map by clicking and dragging the cursor around and can zoom using the scale on the left of the map. Double-clicking somewhere on the map will zoom and centre the map on that point.

Creating a new area

The manager must click on the grey select area button in the top right corner of the map. Clicking and dragging with the mouse will define the geographic coverage area. Only one area can be created per map.



Description

The manager can describe the geographic coverage in the box below the map.

5.3.4 Taxonomic Coverage

Taxonomic Metadata

General Taxonomic Coverage Description

Taxonomic Keywords

Cancel

Save

Lowest Common Taxon

Scientific Name *

Common Name

Rank

Cancel

Save

Taxonomic Metadata

General Taxonomic Coverage Description

Taxonomic Keywords

Lowest Common Taxon

Scientific Name (mandatory field)

Common Name

Rank (mandatory field)

5.3.5 Temporal Coverage

Temporal Coverage

Start Date * End Date * Single Date
e.g. 1999-07-21

Cancel Save

The manager must fill in the temporal coverage of the collection of information related to the resource: start and end date. The dates must be filled in using the format "YYYY/MM/DD" (e.g. 2003/04/22). If the resource covers a single day, the manager can enter a single date and check the corresponding checkbox.

5.3.6 General Keywords

General Keywords

Keywords separated by comma

Cancel Save

In this section, the manager can indicate a series of keywords defining/describing the resource. The different keywords must be separated by commas.

5.3.7 IPRights

On this page, the manager can indicate the information related to intellectual property rights (IPRights) attached to the resource.

IPRights

Purpose

Maintenance

IPRights

Considering using a [Creative Commons](#) license?

Cancel

Save

Purpose A description of the purpose of this dataset.

Maintenance A description of the maintenance of this data resource. This includes information about the frequency of update, and whether there is ongoing data collection.

IPRights Typically, an intellectual Rights element will contain a rights management statement for the resource, or reference a service providing such information. Rights information encompasses Intellectual Property Rights (IPR), Copyright, and various Property Rights. In the case of a data set, rights might include requirements for use, requirements for attribution, or other requirements the owner would like to impose.

Considering using a Creative Commons license? The GBIF promotes the use of the Creative Commons licenses, which allow the originator to keep his/her copyright and at the same time allow to copy and distribute the information under the conditions specified by the originator. For more information, see <http://creativecommons.org/license/>.

The link below the IPRights box 'Considering using a Creative Commons license?' automates the process of creating a Creative Commons licence so users can copy the output straight into the IPRights box.

5.3.8 Research Project

Research Project

Title

Project Lead Organisation

Abstract

Funding

Study Area Description

Design Description

If the resource is linked to a specific research project, the manager can describe it in this section.

Title

Project Lead Organisation

Abstract

Funding

Study Area Description

Design Description The field Design Description contains general textual descriptions of research design. It can include detailed accounts of goals, motivations, theory, hypotheses, strategy, statistical design, and actual work. Literature citations may also be used to describe the research design.

5.3.9 Methods

Methods

Methods

Sampling Description

Quality Control

Cancel Done

Methods

Sampling Description

Quality Control

5.4 Data Resource Configuration (occurrences and checklists only)

Once the resource has been created, the manager must define a data source and map properties.

5.4.1 Source Data

Data sources can be of two kinds in the IPT: File Data Sources or SQL Data Sources (connections to existing database servers). The manager can define as many data sources as wanted, but it is necessary to define at least one. As the main one will be mapped to the Darwin Core terms, it is best for the source data to be adjusted accordingly (i.e. column names should be Darwin Core terms). Others will be used to complement a core Darwin Core record. For this to work, additional sources need to refer to core identifiers, i.e. have a foreign key, but do not need their own identifier.

File Data Sources



In order to define a file data source, the manager must upload a text file with a single header row and with tab separated values. Such *tab delimited text files* are easily created from any spreadsheet software or database management system.

Zip-compressed can be used and will be automatically unpacked on the server. Uploaded files (compressed or not) cannot be larger than 100MB.

Once the file has been uploaded, the manager can define whether the first row of the file is a header or not, by checking or unchecking the *column headers* box and pressing the *save* button. All the data fields will then be named after this header. If not set, generic names (col001, col002, col003 and so on) will be used.

When done adding data sources, the manager can press the *Next* button to go to the *Mappings* page.

SQL Data Sources

In order to define a SQL data source, the manager must first configure a database connection, then add some SQL select statements to create views on the data.

Pressing the *Edit* button on the *Source Data* page will bring the manager to the *SQL Settings* page. This is the page where the database connection will be defined. To this effect, the following information is needed: the database type, the database connection URL, the database user and the database password. The right Java DataBase Connection (JDBC) driver must also be installed. Many of them are already bundled with the IPT. It is possible to add new ones, by downloading them and adding the *jar* file to the ITP *lib* folder and restarting the application server.

The *SQL Settings* page also gives a list of important parameters for the most used databases.

SQL Settings



Database Type *
MySQL

Database Connection URL *
jdbc:mysql://localhost/YOUR_DATABASE

Database User *

Database Password *

Save Cancel

Database Type The type of the database holding the data must be defined. If this particular system is not in the list, the manager can use a generic ODBC or choose the *other* option and to define a custom driver class.

Database Connection URL The manager must fill in the URL of the database connection. The format to be used is explained in the second part of the *SQL Settings* page. The manager will generally need to know the host or server (*localhost* if the database is on the same machine as the IPT), the port (default ports are given for most used databases), and the name of the database (to be asked to the database administrator). For instance, for a local MySQL database called *test*, the database connection URL will be `jdbc:mysql://localhost:3306/test`.

Database User The username of a database user with SELECT rights

Database Password The corresponding password

After defining the SQL settings, the manager must add some SQL *SELECT* statements to create views on the data. This is done by clicking on the *ADD* button and filling in the form.

SQL Data Sources

Database Connection URL
jdbc:mysql://localhost/pontaurus

Edit

! Please configure at least one sql view to start using this datasource

Add

Several of these queries can be created, e.g. one for each Darwin Core extension in use.

Source Name *
specimen

SQL Statement *
select * from specimen

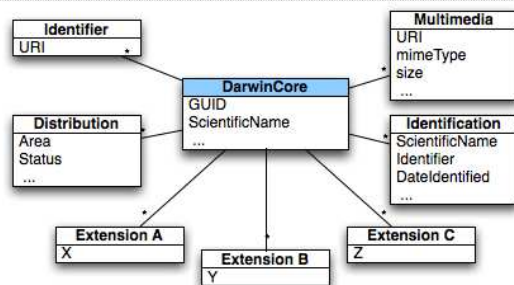
Save Cancel Delete

Once all data sources have been defined, pressing the *Next* button will bring the manager to the *Mappings* page. The SQL statements will be mapped the same way as uploaded text files.

5.4.2 Mappings

Data in the IPT is organised along a star schema: a fixed table holds Darwin Core terms and is linked to a number of extension tables (in a one-to-many relationship) that can be optionally configured and defined by the IPT administrator (see *Extensions* section above). Each extension can hold multiple records for a single core record, e.g. there can be many images for a species or occurrence.

Darwin Core Star Schema



The *Property Mappings* page displays a list of the existing mappings, that can be edited, gives the possibility to add new mappings and allows to actually import or re-import the data.

Your Existing Mappings

DarwinCore 24 properties mapped to source *pontaurus.txt* Edit

Add New Mapping

Extension to map: Source of data: Add

Cache Database

The IPT caches all data being served. You can update the cache at any time and start a data import from your sources based on the transformations and mappings you have configured. Depending upon the amount of data this process may take a long time during which this resource is blocked.

Last Upload: Mar 31, 2009 10:13:26 PM ([logs](#))
 Number of Core Records 1,534

Import

Your Existing Mappings

When defining new data sources, the IPT will automatically map those fields whose name fits a Darwin Core term or the field name of an installed extension to the furthest extent possible. These mappings can be viewed or modified by pressing the *Edit* button.

The page is split in three parts:

ID Properties A list of all properties used as identifiers

Property Mappings A list of all the selected extension or core properties that have been mapped to a data source field.

Available Properties A list of all the selected extension or core properties that have not been mapped to a data source field.

ID Properties

In this section, the manager can define which fields are used as identifiers (IDs).

DarwinCore mapping

Source ID for record *

GUID for record

Link to record details

Link template with <ID> placeholder

Save Delete Done

If the source is mapped to DarwinCore, the following fields are to be filled in:

Source ID for record (mandatory field) The sourceID is the local ID for each core record that extensions can link to. It is also used to compare an existing record when a new version is imported. It is a required field as the IPT currently doesn't support sources without identifiers.

GUID for record The GUID (Globally Unique Identifier) is optional but will be generated automatically if not mapped. It is used as a public ID for published records.

Link to record details This is a link to the HTML details page for a single record. It can link to any public webpage (i.e. not in the IPT) holding the data. If this link is not available in the database, the manager can use the *Link template with <ID> placeholder* (see below).

Link template with <ID> placeholder The manager can enter a direct link to an external webpage. The manager can use <ID> as a placeholder that will be dynamically replaced for every record. Each record can then be linked to a specific webpage.

NB: The *save* button will save the information that was entered in the fields. The *Done* button brings the manager back to the main *Property Mappings* page **without saving the information**.

Extension mapping



If the source is mapped to an extension, only one field is to be filled in:

Source ID for record (mandatory field) The source ID is the local ID for an extension record that the core can link to.

Property Mappings

Each Darwin Core or extension property mapped to a data source field is displayed in the list, followed by the data source field. This field can be changed as the drop-down menu displays the full list of fields of a given data source. It is also possible to assign a fixed value (a common value for all records, i.e. the project/collection code) by entering it in the text field (and choosing a blank value in the drop-down).

Property Mappings

For a single property that you want to map, select a column from your source or enter a fixed value into the text field. If the property has a vocabulary associated you can also select a term from the dropdown.

To add more properties please select them from the available properties just above (clicking the arrow opens menu).

InstitutionCode
InstitutionCode

CollectionCode
CollectionCode

BasisOfRecord
BasisOfRecord Terms or select a static value:

CatalogNumber
CatalogNumber

EarliestDateCollected
EarliestDateCollected

Collector
Collector

Continent
Asia

Country
TR

If controlled vocabulary has been defined for the Darwin Core or the extension, it is not possible to manually enter a fixed value. In this case, a *Terms* button and a new drop down are available.

BasisOfRecord
BasisOfRecord Terms or select a static value:

- Preserved Specimen
- Fossil Specimen
- Living Specimen
- Human Observation
- Machine Observation
- Still Image
- Moving Image
- Sound
- Physical Object

Managers have to use one of the controlled terms either as static values by selecting one from the drop down, or by selecting a source column that contains such values. As the source often uses different terms than the ones defined in the controlled vocabulary, the manager will be able to transform/map each value from the source to a controlled term in the vocabulary.

Available Properties

This section displays a list of the selected extension or core properties not yet mapped to a data source field. They are grouped in subsections that can be expanded by clicking on the green arrow. To map an available

property to a data source field, the manager must click on the said property that will then move to the *property mappings* section. The manager can then choose which data source field should be mapped to it (by selecting it in the drop-down menu).

Available Properties

- Identification
- SamplingLocation
- DublinCore
 - Language
 - Rights
 - RightsHolder
- Dataset
- Taxon
- SamplingEvent
- Sample

Add New Mapping

To add a new mapping, the manager must first choose what extension will be mapped to which data source, then press the *Add* button.

The procedure is then the same as working with *existing mappings* (see above).

Cache Database

The cache page allows the manager to actually import the data into the IPT. This is done by pressing the *Import* button. Depending on the amount of data and the network performance, the operation can take a long time. Once done, the IPT displays a confirmation message. If errors occur during the import operation, they are attached as *annotations* to the resource. These annotations are available through the *explore* menu.

The cache can be rebuilt at any time when the manager wants to update the information with new or updated sources.

5.5 Resource Publication

Once a Resource has been configured and its metadata filled in, it can be published on the IPT, via the *Publish* button.

Publishing a resource also registers it with GBIF and updates its metadata in the registry. A resource therefore has three states, indicated by a coloured box in the upper right corner of the page:

unpublished, unregistered This is the default for new resources. It is neither registered nor should it be visible in the public portal part of the IPT.



modified, not registered The resource is intended to be public, is present in the IPT portal and if settings allow, should be registered with GBIF. The resource metadata is modified though and differs from a) GBIFs registry metadata or b) the versioned EML document which is generated each time the manager publishes a resource (NB: the geoserver WFS/WMS services' own metadata are then also out of date).



published, registered Same as above, but the metadata in the registry/ EML&geoserver should reflect your latest state.



It is also possible to delete a resource by using the *Delete* button on its *Basic Metadata* page.