



# gbif-indexingtoolkit

The GBIF Harvesting and Indexing Toolkit (HIT)

[Project Home](#)[Downloads](#)[Wiki](#)[Issues](#)[Source](#)

Search  for

## ScreenShots

*Screenshots of the HIT.*

Updated Sep 17, 2010 by [kyle.br...@gmail.com](mailto:kyle.br...@gmail.com)

These screen shots are organised according to the page within the application where they would be found. Currently there are 5 primary pages:

- [BioDatasource\\_List](#)
- [Jobs](#)
- [Console](#)
- [Add/Edit\\_BioDatasources](#)
- [Registry](#)

## BioDatasource List



free and open access to biodiversity data

## GBIF HARVESTING AND INDEXING TOOLKIT <sup>(HIT)</sup>

Datasources

Jobs

Console

Registry

## BioDatasource List

An overview of all BioDatasources managed locally, divided into 2 categories: *metadata updaters* and *operators*. *Metadata updaters* gather information about the number of resources behind a given access point and create a new *operator* for each one. The *operators* are then used to manage and perform actions against that individual resource located at the given access point.

Data provider

A B C D E F G H I J K L M N O P Q R S T U V X Y Z

List: [All](#) [Deleted](#) Metadata Updaters: [ALL](#) [DiGIR](#) [BioCASE](#) [TAPIR](#) Operators: [ALL](#) [DiGIR](#) [BioCASE](#) [TAPIR](#)

Select: [All](#) [None](#)

Datasources									
	Datasource	Provider	URL	Target	Max Har	Harveste	Dropped	Harvesting Start	
<input type="checkbox"/>	<a href="#">Lichen DB (ww3.bgbm.org) - Lichen Herbarium B...</a>	Botanic Garden an	<a href="http://ww3.bgbm.org/biocase/pywrapper">http://ww3.bgbm.org/biocase/pywrapper</a>	84891	0	0	0	2009-07-21 11:17:03	
<input type="checkbox"/>	<a href="#">typecollection of fossil sponges and corals of Indo</a>	NLBIF	<a href="http://145.18.162.102/tapirlink/tapir.php/">http://145.18.162.102/tapirlink/tapir.php/</a>	3709	3709	3709	0		
<input type="checkbox"/>	<a href="#">Fungi (Äitte) - aitt-ain-fungi</a>	GBIF-Sweden	<a href="http://www.gbif.se/tapir/tapir.php/aitt-ain-">http://www.gbif.se/tapir/tapir.php/aitt-ain-</a>	251	248	248	3		
<input type="checkbox"/>	<a href="#">Birds (GBIF-SE:Artdatabanken) - sq-birds</a>	GBIF-Sweden	<a href="http://www.gbif.se/tapir/tapir.php/sq-bird-">http://www.gbif.se/tapir/tapir.php/sq-bird-</a>	0	1475958	1475958	-1475958		
<input type="checkbox"/>	<a href="#">Fishbase - fishbase</a>	FishBase	<a href="http://www.gbif.se/tapir/tapir.php/fishbas">http://www.gbif.se/tapir/tapir.php/fishbas</a>	1054983	971184	971184	83799		
<input type="checkbox"/>	<a href="#">GBIF-Spain (taray.csic.es) - Estacion Biologica Dc</a>	GBIF-Spain	<a href="http://taray.csic.es:6000/digir/DiGIR.php">http://taray.csic.es:6000/digir/DiGIR.php</a>	23902	0	0	23902		
<input type="checkbox"/>	<a href="#">Forster herbarium, Göttingen (GOET) (vweb-test2</a>	Universitatsherbari	<a href="http://vweb-test2.uni-goettingen.de/bioc">http://vweb-test2.uni-goettingen.de/bioc</a>	0	0	0	0		
<input type="checkbox"/>	<a href="#">Type herbarium, Göttingen (GOET) (vweb-test2.u</a>	Universitatsherbari	<a href="http://vweb-test2.uni-goettingen.de/bioc">http://vweb-test2.uni-goettingen.de/bioc</a>	0	0	0	0		
<input type="checkbox"/>	<a href="#">Israel Nature and Parks Authority (ww3.bgbm.org)</a>	Israel Nature and F	<a href="http://ww3.bgbm.org/biocase/pywrapper">http://ww3.bgbm.org/biocase/pywrapper</a>	433463	429852	429852	3611		
<input type="checkbox"/>	<a href="#">http://www.geo-artenvielfalt.de (www.geo-artenvie</a>	GEO-Tag der Arten	<a href="http://www.geo-artenvielfalt.de/biocase/">http://www.geo-artenvielfalt.de/biocase/</a>	0	0	0	0		
<input type="checkbox"/>	<a href="#">http://www.geo-artenvielfalt.de (www.geo-artenvie</a>	GEO-Tag der Arten	<a href="http://www.geo-artenvielfalt.de/biocase/">http://www.geo-artenvielfalt.de/biocase/</a>	0	0	0	0		
<input type="checkbox"/>	<a href="#">http://www.geo-artenvielfalt.de (www.geo-artenvie</a>	GEO-Tag der Arten	<a href="http://www.geo-artenvielfalt.de/biocase/">http://www.geo-artenvielfalt.de/biocase/</a>	21	21	21	0		
<input type="checkbox"/>	<a href="#">http://www.geo-artenvielfalt.de (www.geo-artenvie</a>	GEO-Tag der Arten	<a href="http://www.geo-artenvielfalt.de/biocase/">http://www.geo-artenvielfalt.de/biocase/</a>	0	0	0	0		

Above: a view of the BioDatasource List Page - taken from v. 0.1Alpha-SNAPSHOT

## Jobs



free and open access to biodiversity data

GBIF HARVESTING AND INDEXING TOOLKIT <sup>(HIT)</sup>

Datasources

Jobs

Console

Registry

## Job List

A view of all operations that have been scheduled and are awaiting execution. (The order of the queue going from top to bottom).

Jobs				
Name	Location	Created	Started	Next Fire
inventory	CENPAT ( <a href="http://argbif.cenpat.gov.ar">argbif.cenpat.gov.ar</a> ) - CNPBOT -> <a href="http://argbif.cenpat.gov.ar/digir/DiGIR.php">http://argbif.cenpat.gov.ar/digir/DiGIR.php</a>	2009-06-09T10:29:14	2009-06-09T10:29:19	
processInventoried	CENPAT ( <a href="http://argbif.cenpat.gov.ar">argbif.cenpat.gov.ar</a> ) - CNPBOT -> <a href="http://argbif.cenpat.gov.ar/digir/DiGIR.php">http://argbif.cenpat.gov.ar/digir/DiGIR.php</a>	2009-06-09T10:29:14		
search	CENPAT ( <a href="http://argbif.cenpat.gov.ar">argbif.cenpat.gov.ar</a> ) - CNPBOT -> <a href="http://argbif.cenpat.gov.ar/digir/DiGIR.php">http://argbif.cenpat.gov.ar/digir/DiGIR.php</a>	2009-06-09T10:29:14		
processHarvested	CENPAT ( <a href="http://argbif.cenpat.gov.ar">argbif.cenpat.gov.ar</a> ) - CNPBOT -> <a href="http://argbif.cenpat.gov.ar/digir/DiGIR.php">http://argbif.cenpat.gov.ar/digir/DiGIR.php</a>	2009-06-09T10:29:14		

Version 0.1Alpha-SNAPSHOT | [Project Home](#) | [Bug Report](#) | © 2009 [GBIF](#)

Above: a view of the Jobs Page - taken from v. 0.1Alpha-SNAPSHOT

## Console



free and open access to biodiversity data

## GBIF HARVESTING AND INDEXING TOOLKIT <sup>(HIT)</sup>

Datasources

Jobs

Console

Registry

## LogEvent List

A view of all log messages being generated by the application, automatically refreshed every few seconds.

Console mode: [verbose](#) | [normal](#) | [minimal](#)

pause

```
2009-06-09 10:29:59.0 Finished process harvested
2009-06-09 10:29:59.0 Header line of harvested tab file has been written successfully
2009-06-09 10:29:59.0 Start process harvested
2009-06-09 10:29:57.0 Finished harvest
2009-06-09 10:29:57.0 Success harvesting range [ Acaena - Vulpia megalura ]
2009-06-09 10:29:57.0 END_OF_RECORDS : true
2009-06-09 10:29:57.0 RECORD_COUNT : 859
2009-06-09 10:29:56.0 Writing to file: /tmp/hit/cenpat_argbif.cenpat.gov.ar/cnpbot/search_response.000
2009-06-09 10:29:39.0 Writing to file: /tmp/hit/cenpat_argbif.cenpat.gov.ar/cnpbot/search_request.000
2009-06-09 10:29:39.0 Start harvesting range [ Acaena - Vulpia megalura ]
2009-06-09 10:29:39.0 Start harvest
2009-06-09 10:29:29.0 Finished process inventoried
2009-06-09 10:29:29.0 Name ranges file has been written
2009-06-09 10:29:29.0 Start process inventoried
2009-06-09 10:29:22.0 Finished inventory
2009-06-09 10:29:22.0 END_OF_RECORDS : true
2009-06-09 10:29:22.0 RECORD_COUNT : 419
2009-06-09 10:29:22.0 Writing to file: /tmp/hit/cenpat_argbif.cenpat.gov.ar/cnpbot/inventory_response.000
2009-06-09 10:29:19.0 Writing to file: /tmp/hit/cenpat_argbif.cenpat.gov.ar/cnpbot/inventory_request.000
2009-06-09 10:29:19.0 Start inventory
```

Above: a view of the Console Page - taken from v. 0.1Alpha-SNAPSHOT

```
2009-05-27 13:14:29.0
mappingFile=indexMapping_mainis_dwc_1_0.gbif.biodatasourceid=440}}
Inventory request could not be executed: No route to host
java.net.PlainSocketImpl.socketConnect ( PlainSocketImpl.java: -2 )
java.net.PlainSocketImpl.doConnect ( PlainSocketImpl.java: 333 )
java.net.PlainSocketImpl.connectToAddress ( PlainSocketImpl.java: 199 )
java.net.PlainSocketImpl.connect ( PlainSocketImpl.java: 182 )
java.net.SocksSocketImpl.connect ( SocksSocketImpl.java: 432 )
java.net.Socket.connect ( Socket.java: 520 )
java.net.Socket.connect ( Socket.java: 470 )
java.net.Socket. ( Socket.java: 367 )
java.net.Socket. ( Socket.java: 240 )
org.apache.commons.httpclient.protocol.DefaultProtocolSocketFactory.createSocket ( DefaultProtocolSocketFactory.java: 60 )
org.apache.commons.httpclient.protocol.DefaultProtocolSocketFactory.createSocket ( DefaultProtocolSocketFactory.java: 122 )
org.apache.commons.httpclient.HttpConnection.open ( HttpConnection.java: 707 )
org.apache.commons.httpclient.HttpMethodDirector.executeWithRetry ( HttpMethodDirector.java: 387 )
org.apache.commons.httpclient.HttpMethodDirector.executeMethod ( HttpMethodDirector.java: 171 )
org.apache.commons.httpclient.HttpClient.executeMethod ( HttpClient.java: 397 )
org.apache.commons.httpclient.HttpClient.executeMethod ( HttpClient.java: 323 )
org.gbif.harvest.util.RequestUtils.executeGetRequest ( RequestUtils.java: 60 )
org.gbif.harvest.digit.DigitHarvester.fireInventory ( DigitHarvester.java: 252 )
org.gbif.harvest.digit.DigitHarvester.inventory ( DigitHarvester.java: 199 )
org.gbif.harvest.digit.DigitHarvester.inventory ( DigitHarvester.java: 1065 )
sun.reflect.NativeMethodAccessorImpl.invoke0 ( NativeMethodAccessorImpl.java: -2 )
sun.reflect.NativeMethodAccessorImpl.invoke ( NativeMethodAccessorImpl.java: 39 )
sun.reflect.DelegatingMethodAccessorImpl.invoke ( DelegatingMethodAccessorImpl.java: 25 )
java.lang.reflect.Method.invoke ( Method.java: 585 )
org.apache.commons.beanutils.MethodUtils.invokeExactMethod ( MethodUtils.java: 334 )
org.gbif.harvest.scheduler.BiodatasourceOperationLauncher.launch ( BiodatasourceOperationLauncher.java: 184 )
org.gbif.harvest.scheduler.Worker$Runner.run ( Worker.java: 86 )
java.lang.Thread.run ( Thread.java: 613 )
Writing to file: /tmp/hit/academy_of_natural_sciences_digit_acznatool_herpetology/inventory_request.000
Start inventory
2009-05-27 13:14:29.0
Launching harvesting job using factory{org.gbif.harvest.digit.DigitHarvesterFactory} operation{inventory} args{[directory=academy of natural
```

Above: a view of a log message from the console, showing an error and in expanded form showing the full exception stacktrace - taken from v. 0.1Alpha-SNAPSHOT

## Add/Edit BioDatasource



free and open access to biodiversity data

GBIF HARVESTING AND INDEXING TOOLKIT (HIT)

Datasources

Jobs

Console

Registry

## BioDatasource Detail

### Configure your BioDatasource

Save Cancel

BioDatasource name\*: CENPAT (argbif.cenpat.gov.ar)

Provider Name\*: Centro Nacional Patagonico - CONICET

URL\*: http://argbif.cenpat.gov.ar/digir/DiGIR.php

Factory class\*: DiGIR Metadata Factory

► [Sign in](#) to add a comment

Version 0.1Alpha-SNAPSHOT [Terms](#) [Privacy](#) [Project Hosting](#) [Help](#) © 2009 [GBIF](#)

Powered by [Google Project Hosting](#)

Above: a view of the Add/Edit BioDatasource Page - taken from v. 0.1Alpha-SNAPSHOT

## Registry

flickr



Cette photo n'est plus disponible

Above: a view of the Registry Page - taken from v. 0.1Alpha-SNAPSHOT