IPT Workbook

GBIF IPT version 2

The following workbook will take you through concrete examples that illustrate how to install, administer, and use the Integrated Publishing Toolkit (IPT) to publish biodiversity datasets. To be able to follow these examples, it is assumed the reader:

- Is familiar with the IPT User Manual
  - http://code.google.com/p/gbif-providertoolkit/wiki/IPT2ManualNotes
- Has installed Tomcat (6.0 or 7.0)
  - http://tomcat.apache.org/
- Has downloaded the workbook's "sample datasets" folder

The workbook has a total of 5 parts:

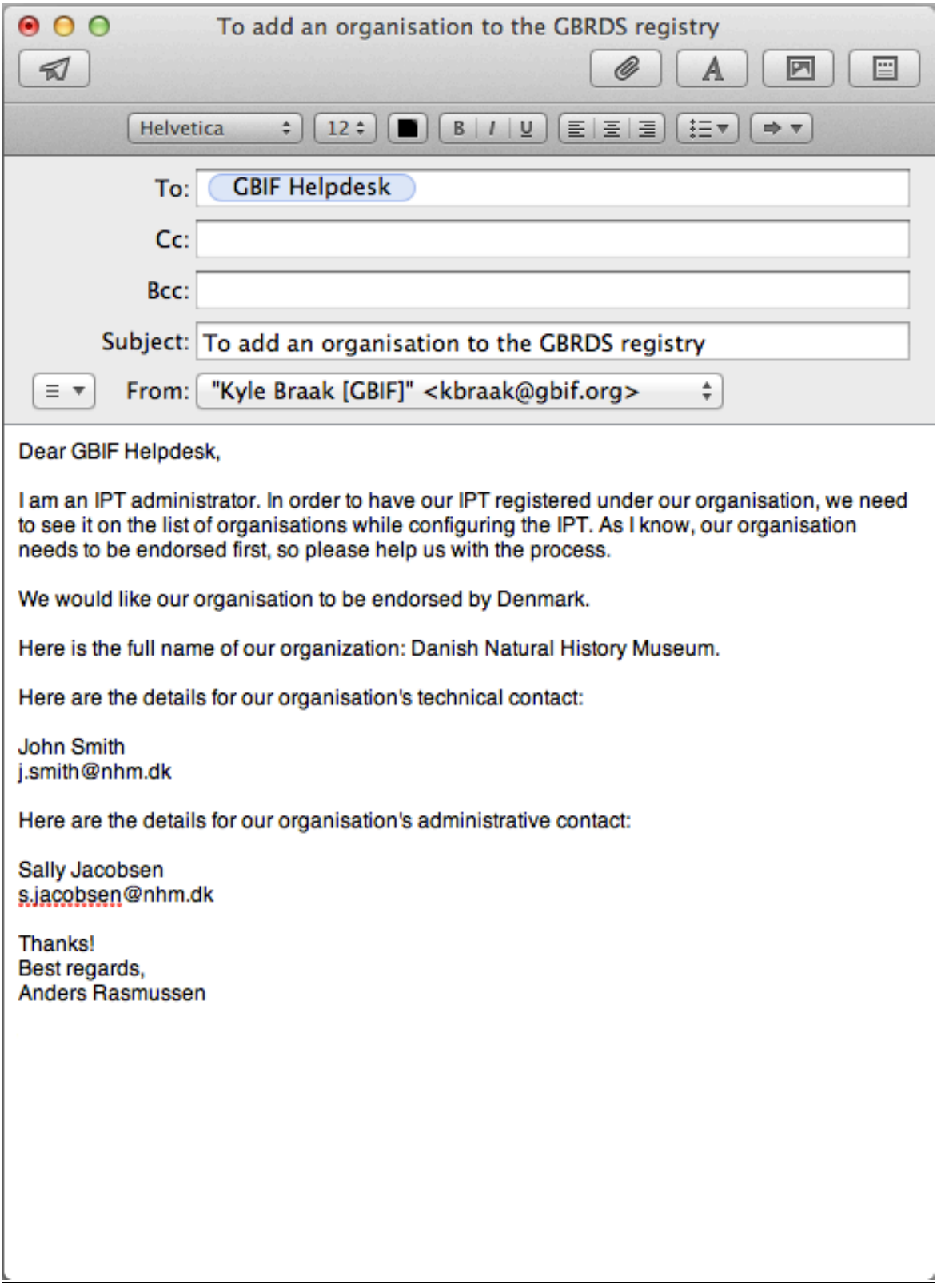
1) Installing and configuring the IPT
2) Reflecting different organizational structures into the IPT
3) Publishing basic resources in the IPT
4) Publishing resources with extensions
5) Extra practice

***Warning**: For all the exercises in this workbook, the IPT must be setup in Test mode. In Test mode registrations go into the GBIF Test Registry (http://gbrdsdev.gbif.org) and resources will never be indexed.*

Part 1: Installing and configuring the IPT

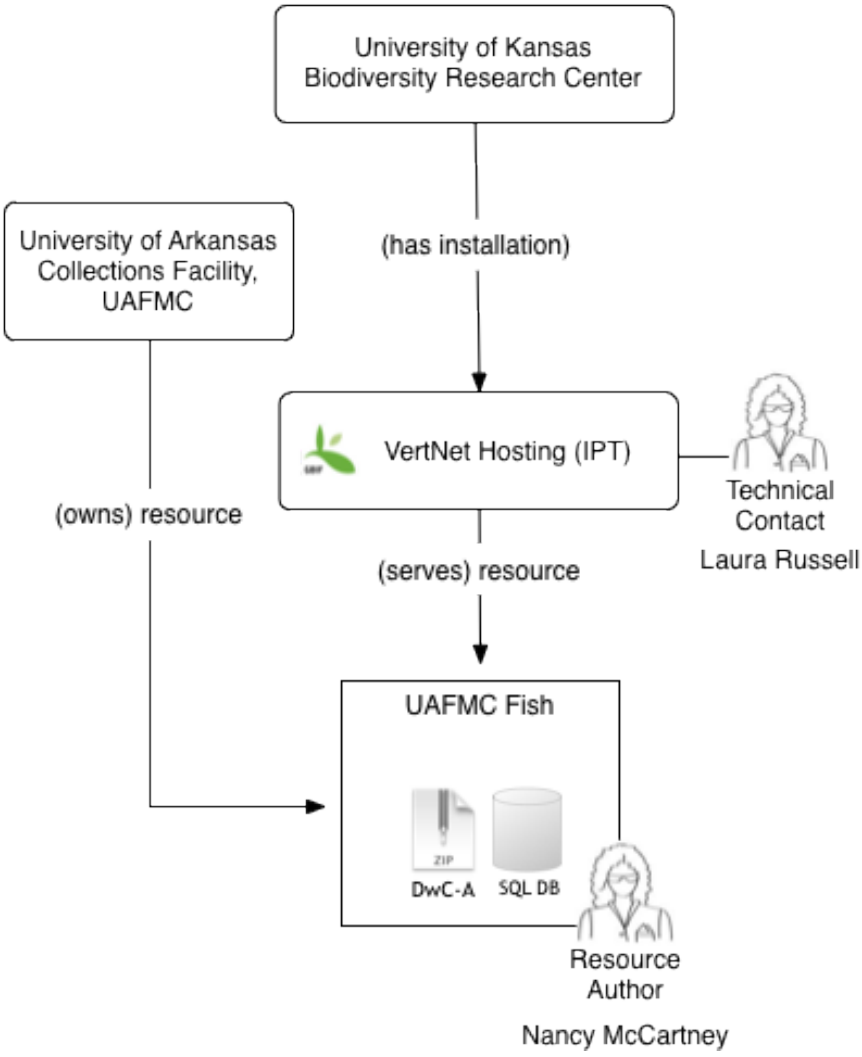
This session's exercises cover how to install the IPT and configure it for initial use. Upgrading an existing installation and changing the base URL are also covered. Additional exercises show how to select the desired port number, bypass a proxy, and use Apache to setup a virtual host name, which resembles a production environment.

Goals	Notes (& issues)
1) Make a brand new installation of the latest version of the IPT (2.0.4 at the time of writing) using Tomcat	1) Try using either the Tomcat manager, or putting the war file directly into webapps/. 2) The file name of the war file will be part of the URL for IPT. 3) Make sure Tomcat has the permission to access your data directory. 4) Do not select a data directory that is vulnerable to inadvertent changes or removal. E.g. /tmp 5) Make sure you understand the consequences of choosing "production" mode versus "test" mode.
2) Upgrade an existing installation (to newer version) preserving your original setup	1) Try deleting the old war file (and deployed war file) via the Tomcat manager or directly from webapps/. 2) Be sure to stop Tomcat before taking action. 3) Backup the data directory just in case something goes wrong
3) Send an email to the GBIF Helpdesk to register a new organization for you in the GBIF Registry (GBRDS).	1) If this is just for practice, please let Helpdesk know that in your email to *helpdesk@gbif.org*. 2) Use the email template provided for you in the IPT to help write your email. It will ensure you include all the information Helpdesk requires. 3) You will receive an email confirmation from Helpdesk later, depending on Helpdesk business hours in Copenhagen. 4) There is an email sample in the next page.


To add an organisation to the GBRDS registry
Helvetica
12
To:
GBIF Helpdesk
Cc:
Bcc:
Subject:
To add an organisation to the GBRDS registry
From:
"Kyle Braak [GBIF]" <kbraak@gbif.org>
Dear GBIF Helpdesk,
I am an IPT administrator. In order to have our IPT registered under our organisation, we need to see it on the list of organisations while configuring the IPT. As I know, our organisation needs to be endorsed first, so please help us with the process.
We would like our organisation to be endorsed by Denmark.
Here is the full name of our organization: Danish Natural History Museum.
Here are the details for our organisation's technical contact:
John Smith
j.smith@nhm.dk
Here are the details for our organisation's administrative contact:
Sally Jacobsen
s.jacobsen@nhm.dk
Thanks!
Best regards,
Anders Rasmussen

Part 2: Reflecting different organizational structures into the IPT

This session's exercises cover how to configure an IPT to meet different organizational requirements. Take the structure of VertNet Hosting below. In this example, the University of Kansas Biodiversity Research Center has an IPT that serves a resource owned by the University of Arkansas Collections Facility, UAFC.



Below are 2 goals. Each one presents a use case that you must use to configure your IPT. At the end, your IPT should:

1) have the proper user accounts with appropriate roles.
2) be registered against an organization.
3) have any organization(s) added
4) have one or more data resources registered against the right organization(s);

You can check the GBIF Registry to confirm the organizational structure has been well built.

Goals	Notes (& issues)
1) Use the use case below to configure an IPT using a single imaginary organization that serves and owns 2 resources.	1) Import Part2-dummy-occurrence.zip to create a new occurrence resource. 2) Import Part2-dummy-checklist.zip to create a new checklist resource. 3) To change the name of the resource, enter a new title in the basic metadata page and save. 4) Since Tanzania is a GBIF Node, TZM can be endorsed by Tanzania. Something else to think about: What if Tanzania wasn't a GBIF Node? Could the TZM still publish its own resources? 5) Re-publish the resource every time you want to update the information in the Registry.

Use case story:

Tanzania is a GBIF Node. Imagine Tanzania only has 1 organization responsible for all biodiversity related science called the Tanzanian Zoological Museum (TZM). The Tanzanian Node Manager has invited you to visit the TZM to assist them with publishing their 2 recently digitized resources: *Entomology Collection* and *Birds of Tanzania Checklist*.

Fortunately, the TZM has good server capacity and a systems administrator that has already installed the IPT on Tomcat for you. It is waiting for you to do the initial IPT registration, and publish these 2 resources now. Your job is to:

1. Register the IPT so that it is seen as hosted by the TZM with primary administrative contact Angus Young (ayoung@tzm.org). The password is "j5CIZBUFa04".
2. Name your IPT instance "TZM_IPT *[your name]*"
3. Create an administrator account for their technical contact Bruce Springsteen who will take over once you leave.
4. Create an occurrence resource named *Entomology Collection [your name]*, publish it, and then register it so that it is owned by TZM.
5. Create a checklist resource named *Birds of Tanzania Checklist [your name]*, publish it, and then register it so that it is owned by TZM.
6. Create one manager (with registration rights) created to manage the *Entomology Collection* called Tina Turner, with her email as tturner@tzm.org.
7. Create one manager (with registration rights) created to manage the *Birds of Tanzania Checklist* called Lenny Kravitz, with his email as lkravitz@tzm.org.

Goals	Notes (& issues)
2) Use the use case below to configure an IPT that is registered against an imaginary organization, which serves and owns 1 resource, and which serves 1 other resource for another organization. Use case story: Argentina is a GBIF Node. They have at present 2 organizations responsible for all biodiversity related science. The 1st is called the Argentinian Zoological Museum (AZM). The 2nd is called the Argentinian Herbarium (AH). The Argentinian Node Manager has invited you to visit the AZM to assist them with setting up their IPT and the following 2 resources: • 1 occurrence resource owned by AZM: I*chthyology Collection*. • 1 occurrence resource owned by AH: M*agnoliidae Checklist.* Fortunately, the AZM has good server capacity and a systems administrator. The AH has neither of those. Therefore, AZM has kindly offered to host AH's resource for them, and AH readily accepted the offer. The systems administrator at AZM has already installed the IPT on Tomcat for you. It is waiting for you to do the initial IPT registration, and publish these 2 resources now. Your job is to: • Register the IPT so that it is seen as being hosted by the AZM with primary administrative contact Jimi Hendrix (jhedrix@azm.org). The password is "AuH9SVFA0bPZ". • Name your IPT instance "AZM IPT *[your name]*" • Add the organization AH to the IPT. The password is "NAUf8wdne4L". • Create an administrator account for their technical contact Chris Cornell who will take over once you leave. • Create an occurrence resource named I*chthyology Collection [your name]*, publish it, and then register it so that it is owned by TZM. • Create a checklist resource named M*agnoliidae Checklist [your name]*, publish it, and then register it so that it is owned by AH. • Create one manager (with registration rights) created to manage the I*chthyology Collection* called Shania Twain, with her email as stwain@azm.org. • Create one manager (with registration rights) created to manage the M*agnoliidae Checklist* called Frank Sinatra, with his email as fsinatra@azm.org.	1) Import Part2-dummy-occurrence.zip to create a new occurrence resource. 2) Import Part2-dummy-checklist.zip to create a new checklist resource. 3) To change the name of the resource, enter a new title in the basic metadata page and save. 4) Since Argentina is a GBIF Node, AZH and AH can be endorsed by Argentina. Something else to think about: Could the ARZ IPT also host resources for organizations located outside Argentina? 5) Re-publish the resource every time you want to update the information in the Registry.

Part 3: Publishing basic resources in the IPT

This session's exercises cover how to publish simple resources. The resource can be as simple as having only metadata, so publishers can expose their resource first, and then complete their resource by adding source data when it's appropriate.

Exercises involving data demonstrate how resources can be derived from data originating in different formats, for example in GBIF Spreadsheet Templates, MS Excel, and a MySQL database.

Goals	Notes (& issues)
1) Publish a metadata-only resource (DwC-A).	
2) Publish a metadata-only resource (DwC-A) by first importing an existing EML file.	1) Use file Part3-Goal2-eml.xml 2) Use the IPT's built-in metadata editor to validate the metadata. Hint: open the Basic Metadata page and click save. 3) To be importable, the XML file must use the GBIF Metadata Profile.
3) Publish a taxonomic resource (DwC-A) from a GBIF Spreadsheet Templates. See: http://tools.gbif.org/spreadsheet-processor/	1) Spreadsheet templates need to be processed into a DwC-A before being imported into the IPT 2) Try with a different template for a species checklist, Part3-Goal3-taxonomy.xls 3) Hint: pick the right core mapping carefully!
4) Publish an occurrence resource (DwC-A) from a tab delimited source file that has been exported from MS Excel using a tab delimiter, ensuring tab spaces have been removed from all fields)	1) Use excel file Part3-Goal4-tab-hast.xls 2) It is advisable to use tab delimiters vs comma
5) Publish an occurrence resource (DwC-A) from a database, such as MySQL.	1) Use your own database if you can (if the IPT supports it of course). 2) Otherwise if you have MySQL installed, follow the instructions inside folder MySQL sample data to create a test database.
6) Publish a taxonomic resource (DwC-A) from multiple source files all having the identical set of columns.	1) Use the 3 files zipped inside Part3-Goal6-multiple.zip 2) Advanced question: Have all the columns been auto mapped? Check the logs displayed. 3) Advanced question: How could you give all records in the dataset the same basis of record?

Part 4: Publishing resources with extensions

This session's exercises cover how to publish more advanced resources in accordance with the star schema by making use of extensions. The powerful features of filtering, static values, and translations are also introduced.

Goals	Notes (& issues)
1) Publish a taxonomic resource (DwC-A) using the vernacular name extension.	1) Use files in folder Part4-Goal1 2) Make sure you have the vernacular name extension installed 3) Make sure you set the taxonID, so the IPT knows which vernacular name record belongs to which core taxon record.
2) Publish a taxonomic resource (DwC-A) using the vernacular name extension, but exclude all English names.	1) Use same files as in PS4-Goal1 2) Use the filtering feature on the mapping page to exclude all English names
3) Publish a taxonomic resource (DwC-A) using the vernacular name extension, but this time using multiple source files mapped to the vernacular name extension: one has English vernacular names, and the other has Spanish vernacular names.	1) Use vernacular name source files in folder PS4-Goal3
4) Publish a taxonomic resource (DwC-A) with a single source file, assigning static/global values for kingdom = "Animalia", nomenclaturalCode = "ICZN", accessRights = "not-for-profit use only", taxonomicStatus change "valid" to "accepted" using a translation	1) Use files in folder PS4-Goal4
5) Publish an occurrence resource (DwC-A) that ensures the mandatory fields required by GBIF (to index an occurrence resource) are actually mapped: the holy triplet (collectioinCode, institutionCode and catalogNumber), basisOfRecord, scientificName.	1) Use file PS4-Goal5 2) Hint: the file contains all required fields, but not all auto-mapped!
6) See what happens when trying to create a new resource by importing a zipped DwC-A that uses improper delimiters (semicolons used but also appearing in the data).	1) Use file PS4-Goal6-delimiter.zip 2) Use the Darwin Core Archive Validator to check if the archive is valid

Part 5: Extra practice

This session is a continuation of exercises from PS4. It gives you the chance to continue playing with the IPT, and learn about supplementary tools.

Goals	Notes (& issues)
1) Debug a DwC-A using the online validator: http://tools.gbif.org/dwca-validator/	1) Try validating a DwC-A, or just an eml.xml file from previous exercises
2) Use the GBIF Spreadsheet Templates and DwC-A Spreadsheet Processor to create a new *occurrence* resource this time http://tools.gbif.org/spreadsheet-processor/	1) Take a look at all the other templates available from the Spreadsheet Processor website
3) Change the port number of your Tomcat. Do not reinstall Tomcat. See notes for help. For example, use port 8090 instead of 8080.	1) Modify the $tomcat/conf/server.xml and make sure your Tomcat is restarted for the change to take effect. 2) This change affects the IPT's base URL, so be sure to save the new base URL in the administration settings. Remember your IPT is now at the new port!