

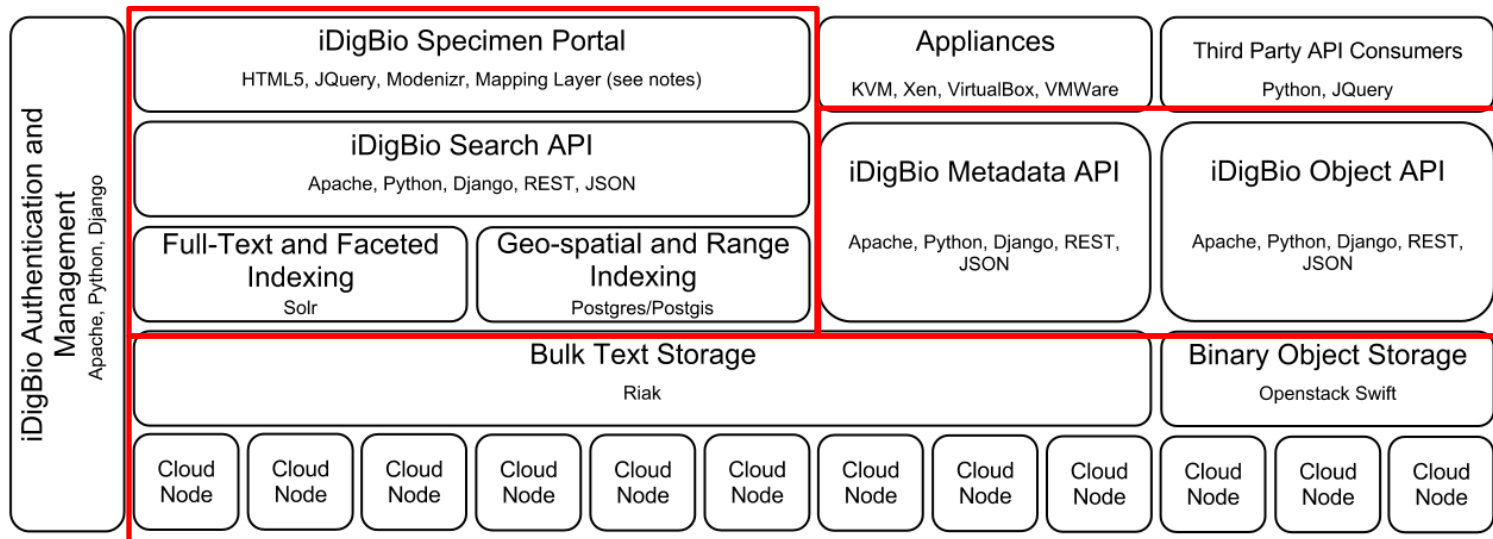
iDigBio Image Ingestion Appliance

Renato Figueiredo
(on behalf of the
iDigBio IT team)



Overview

- iDigBio cloud storage and specimen portal are at the core of our cyber-infrastructure
 - Cloud infrastructure provides scalable storage of records and images
 - Growing set of capabilities and services exposed through Web interfaces and REST APIs



Overview

- Appliances complement the cyber-infrastructure core
 - Functionality desired on the client; hide low-level iDigBio APIs, expose user-friendly interface (e.g. image ingestion)
 - Package tools of general interest to the community in virtual machines for ease of software deployment
 - On resources (desktops, servers) local to the institution
 - On cloud-provisioned resources (Amazon EC2, iDigBio cloud)

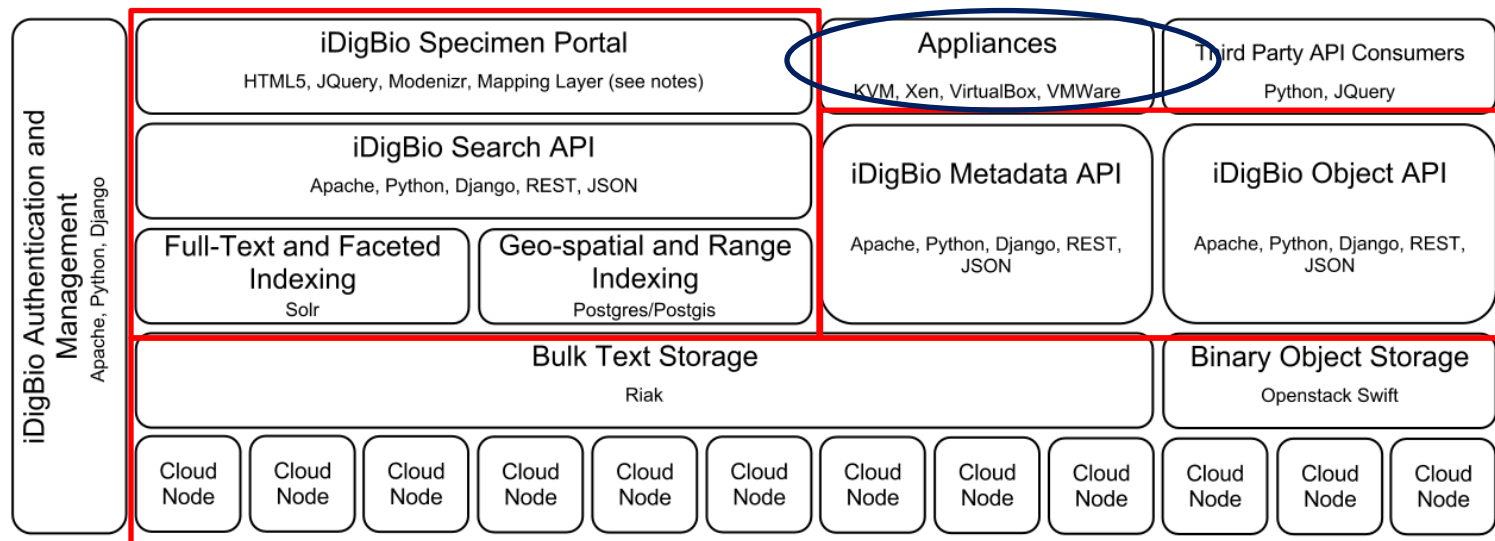


Image ingestion appliance use cases

- Allow end-users to ingest images to the iDigBio cloud to provide a mechanism to support crowd-sourcing
 - Ingest batches of images
 - Assign GUIDs to media information
 - Image searchable on the portal through media information
 - Generate lists of media information for each batch including the HTTP accessible endpoint at iDigBio for each ingested media object.
- Allow end users to ingest images that are linked to specimen records
- Provide a basis for integration with third-party bio-collections tools to create appliances that can automatically ingest images into the iDigBio cloud

Appliance functionality overview

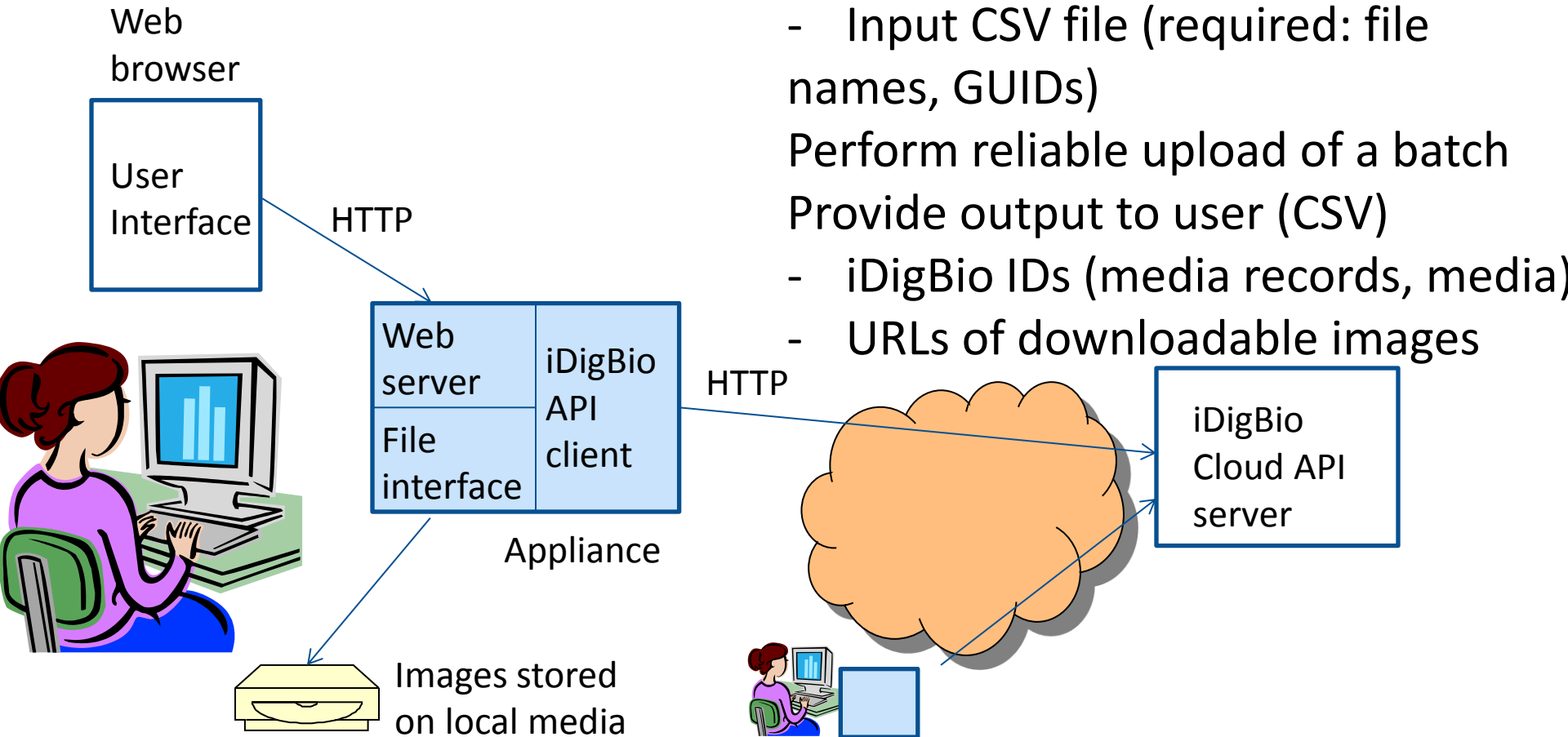
Obtain user inputs defining a batch:

- Web user interface
- Input CSV file (required: file names, GUIDs)

Perform reliable upload of a batch

Provide output to user (CSV)

- iDigBio IDs (media records, media)
- URLs of downloadable images



User interface - login

127.0.0.1:32601

For quick access, place your bookmarks here on the bookmarks bar. [Import bookmarks now...](#)

iDigBio Home Docs Logout

Please Sign In

Before you can use this application, you first need to sign in using the iDigBio API Account UUID and the API key. After you sign in, this combination will be saved by the application for your future use.

Account UUID

API Key

We accept your images as is to display as given. By clicking the button, you agree to abide by the [iDigBio Intellectual Property Policy](#) and the [iDigBio Terms of Use](#).

Image License *

Record Set GUID *

CSV File Path *

Media Content Keyword

iDigbio Provider GUID

iDigbio Publisher GUID

Funding Source

Funding Purpose

Note: the fields with * are mandatory.

Upload

UUID and API key from specimen portal

User interface – select batch

For quick access, place your bookmarks here on the bookmarks bar. [Import bookmarks now...](#)

iDigBio Home Docs Logout




Image Ingestion Tool

This is a tool that helps you ingest images into the iDigBio storage cloud.

Image License *

Record Set GUID *

CSV File Path *

Media Content Keyword

iDigbio Provider GUID

iDigbio Publisher GUID

Funding Source

Funding Purpose

Note: the fields with * are mandatory.

License
Record set GUID
Path to CSV input file

Start batch

User interface – input CSV

idigbio:OriginalFileName	idigbio:MediaGUID	idigbio:Description	idigbio>Title	idigbio:LanguageCode
C:\Users\renato\Desktop\upload2\AMNH_	www.fakeurl.edu/abc/def/image837226	Scanned herbarium s	Ilex glabra from FS lo	
C:\Users\renato\Desktop\upload2\AMNH_	www.fakeurl.edu/abc/def/image23734	Scanned herbarium s	Ilex glabra from FS lo	
C:\Users\renato\Desktop\upload2\AMNH_	www.fakeurl.edu/abc/def/image573382	Scanned herbarium s	Ilex glabra from FS lo	
C:\Users\renato\Desktop\upload2\AMNH_	www.fakeurl.edu/abc/def/image394976	Scanned herbarium s	Ilex glabra from FS lo	
C:\Users\renato\Desktop\upload2\AMNH_	www.fakeurl.edu/abc/def/image508595	Scanned herbarium s	Ilex glabra from FS lo	
C:\Users\renato\Desktop\upload2\AMNH_	www.fakeurl.edu/abc/def/image926987	Scanned herbarium s	Ilex glabra from FS lo	
C:\Users\renato\Desktop\upload2\AMNH_	www.fakeurl.edu/abc/def/image95308	Scanned herbarium s	Ilex glabra from FS lo	
C:\Users\renato\Desktop\upload2\AMNH_	www.fakeurl.edu/abc/def/image523643	Scanned herbarium s	Ilex glabra from FS lo	
C:\Users\renato\Desktop\upload2\AMNH_	www.fakeurl.edu/abc/def/image153637	Scanned herbarium s	Ilex glabra from FS lo	
C:\Users\renato\Desktop\upload2\AMNH_	www.fakeurl.edu/abc/def/image651953	Scanned herbarium s	Ilex glabra from FS lo	

Required terms

Optional terms:

Description

LanguageCode

Title

DigitizationDevice

NominalPixelResolution

Magnification

OcrTechnology

InformationWithheld

User interface – upload

The screenshot shows a web browser window displaying the iDigBio upload interface. The browser's address bar shows the URL 127.0.0.1:32601. The page has a dark header with the iDigBio logo, navigation links for Home and Docs, and a Logout button. The main content area contains a form with the following fields:

- Image License *: CC BY (Attribution)
- Record Set GUID *: RECORD_SET_GUID_1
- CSV File Path *: /home/renato/idigbio/testing_files/script3/test_input.csv
- Media Content Keyword: Optional
- iDigbio Provider GUID: Optional
- iDigbio Publisher GUID: Optional
- Funding Source: Optional
- Funding Purpose: Optional

Below the form, a note states: "Note: the fields with * are mandatory." A blue "Upload" button is positioned below the note. At the bottom of the form, a progress indicator is shown as a blue bar with diagonal stripes. The text next to the bar reads: "Progress: (Successful:3, Skipped: 0, Failed: 0, Total to upload: 10)". A red arrow points from the text "Progress indicator" to the progress bar.

© iDigBio 2012

User interface – output

Note: the fields with * are mandatory.

[Upload](#)

Progress: (Successful:10, Skipped: 0, Failed: 0, Total to upload: 10)

Show entries Search:

OriginalFileName	Online Path or Error Message
/home/renato/idigbio/testing_files/script3/AMNH_PBI 00087417.jpg	http://beta-api.idigbio.org/v1/mediaaps/73667a3c-9214-4f36-9827-bc1d2db9a0c7/media
/home/renato/idigbio/testing_files/script3/AMNH_PBI 00095610.jpg	http://beta-api.idigbio.org/v1/mediaaps/2bfb095a-81c8-4b46-487e-a9ec09adf252/media
/home/renato/idigbio/testing_files/script3/AMNH_PBI 00095734.jpg	http://beta-api.idigbio.org/v1/mediaaps/d76e5b79-8a2b-4c7e-95ac-dd8a8b5c0d6b/media
/home/renato/idigbio/testing_files/script3/AMNH_PBI 00099677.jpg	http://beta-api.idigbio.org/v1/mediaaps/333a28a1-06f6-49e2-a4bb-021c96a72269/media
/home/renato/idigbio/testing_files/script3/AMNH_PBI 00103095.jpg	http://beta-api.idigbio.org/v1/mediaaps/5e1dba83-db9c-4ee0-b2b3-7ca3289774e8/media
/home/renato/idigbio/testing_files/script3/AMNH_PBI 00142764_europiella_lividella_female.jpg	http://beta-api.idigbio.org/v1/mediaaps/7aaf0a3f-66da-4d13-8561-54410834ae12/media
/home/renato/idigbio/testing_files/script3/AMNH_PBI 00183850.jpg	http://beta-api.idigbio.org/v1/mediaaps/144a723a-cea5-4b31-92e0-b679ad599e58/media
/home/renato/idigbio/testing_files/script3/AMNH_PBI 00184007LM.jpg	http://beta-api.idigbio.org/v1/mediaaps/d4066dca-58ef-4b5f-b9d2-c254ad83ac61/media
/home/renato/idigbio/testing_files/script3/AMNH_PBI 00227349_campylomma_liebknechti_male.jpg	http://beta-api.idigbio.org/v1/mediaaps/dc66efbc-ade6-4212-a6dd-08893f43b66b/media
/home/renato/idigbio/testing_files/script3/AMNH_PBI 00327138.jpg	http://beta-api.idigbio.org/v1/mediaaps/2932d85d-ca6f-4992-ba4a-462ae7ca1705/media

Showing 1 to 10 of 10 entries 1

beta-api.idigbio.org/v1/mediaaps/d76e5b79-8a2b-4c7e-95ac-dd8a8b5c0d6b/media

© iDigBio 2012

Clickable
URLs of images

User interface – output

	A	B	C	D	E	F	G
1	OriginalFileName	MediaE	MediaGUID	MediaRecordUUID	MediaAccessUUID	Cont	UploadTime
2	/home/renato/idigbio/testing_files/script2/AMNH_PBI_00085633.jpg	true	www.fakeurl.edu/abc/def/image907219	27d0a524-9cec-4770-8a94-bd6ff4aa68e6	735d984f-bed2-44dd-aa37-d9fa1481c2e		2013-02-12 18:36:18.817865
3	/home/renato/idigbio/testing_files/script2/AMNH_PBI_00087166.jpg	true	www.fakeurl.edu/abc/def/image18226	bba3e46e-cabb-40c4-97d0-d754722af155	86a44f3d-5ed1-46c3-8e6d-91e060ab72d0		2013-02-12 18:36:26.264883
4	/home/renato/idigbio/testing_files/script2/AMNH_PBI_00094286.jpg	true	www.fakeurl.edu/abc/def/image154501	b0bdc183-7f40-41c3-9d45-5485830605f4	bd623868-3927-4341-a14d-ec4f45543cb0		2013-02-12 18:36:21.413399
5	/home/renato/idigbio/testing_files/script2/AMNH_PBI_00095361.jpg	true	www.fakeurl.edu/abc/def/image245925	ccf1ba03-5480-47ab-a374-f5662c3e2d18	8509b003-6e73-43b0-be4d-a81471a7587d		2013-02-12 18:36:23.940936
6	/home/renato/idigbio/testing_files/script2/AMNH_PBI_00095417.jpg	true	www.fakeurl.edu/abc/def/image910412	9f50e46b-1c3d-4c07-8aab-3851d95132db	8b2c42d0-2ffc-49b7-8f6a-3cbfcb70e11		2013-02-12 18:36:31.297378
7	/home/renato/idigbio/testing_files/script2/AMNH_PBI_00095700.jpg	true	www.fakeurl.edu/abc/def/image85575	88b38451-e478-4398-bef6-52bafd3a7d20	c27bb675-0404-46fc-b547-b7582750abc5		2013-02-12 18:36:11.379107
8	/home/renato/idigbio/testing_files/script2/AMNH_PBI_00102296_male	true	www.fakeurl.edu/abc/def/image467010	7f9a568a-20f9-44c9-941a-75a271df19ce	26e081dd-a844-404f-b106-c1c2eed06b80		2013-02-12 18:36:34.473037
9	/home/renato/idigbio/testing_files/script2/AMNH_PBI_00183940.jpg	true	www.fakeurl.edu/abc/def/image607215	dbb40a5e-f549-4854-a6fb-03549b6fb27b	f8591802-3325-414e-947b-3498efe8a6e2		2013-02-12 18:36:16.451055
10	/home/renato/idigbio/testing_files/script2/AMNH_PBI_00184158.jpg	true	www.fakeurl.edu/abc/def/image461514	36dd1e0c-d18d-4442-bae4-41a04821dc0d	46684431-de23-4b9f-9e2a-19f5ba31bb63		2013-02-12 18:36:13.894334
11	/home/renato/idigbio/testing_files/script2/AMNH_PBI_00184192L.jpg	true	www.fakeurl.edu/abc/def/image596688	26bc4345-15e4-489e-b902-230b2057a74b	8863a60d-7258-4997-9887-82ccbc46e056		2013-02-12 18:36:28.804474
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							
23							
24							
25							
26							
27							
28							
29							
30							
31							
32							
33							
34							
35							
36							
37							

Local image path

MediaGUID

MediaURL

CSV output, MISC terms

OriginalFileName	MediaExists	<u>MediaGUID</u>	
MediaRecordUUID			
MediaAccessUUID	Comments	UploadTime	<u>MediaURL</u>
Description	LanguageCode	Title	DigitalizationDevice
NominalPixelResolution	Magnification	OcrOutput	OcrTechnology
InformationWithheld	MediaMD5	<u>MimeType</u>	
MediaSizeInBytes	ProviderCreatedTimeStamp		
providerCreatedByGUID			
MediaRecordEtag	RecordSetUUID	iDigbioProvidedByGUID	
<u>MediaContentKeyword</u>	FundingSource	FundingPurpose	
iDigbioPublisherGUID	RightsLicenseStatementUrl		
RightsLicenseLogoUrl	iDigbioProviderGUID	RightsLicense	
CSVfilePath	RecordSetGUID		

User interface – output

MediaURL:

<http://beta-api.idigbio.org/v1/mediaaps/73667a3c-9214-4f36-9827-bc1d2db9a0c7/media>



Implementation overview

- Written in Python
 - CherryPy Web framework
 - cx_freeze - generate standalone application (Windows, Mac)
- Possible to bundle with virtual machine image
 - Advantage of using application is local access to media

Directions

- Near-term (weeks)
 - Work with early adopters
 - Fine-tune use case/requirements
 - Improvements in user interface and functionality
 - Taking AudubonCore inputs
- Medium-term (months)
 - Collaborate on road-map for integration with bio-collection tools
 - Work closely with adopted tools, generalize
 - Packaging of virtual appliances