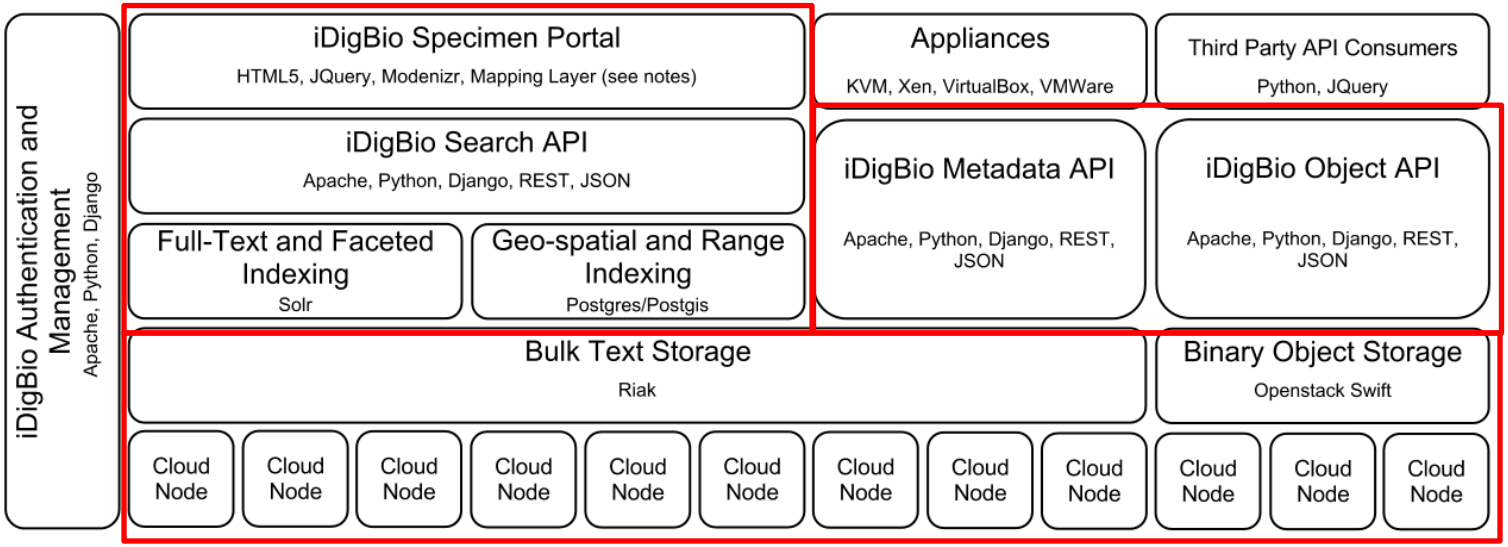
iDigBio Image Ingestion Appliance

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)

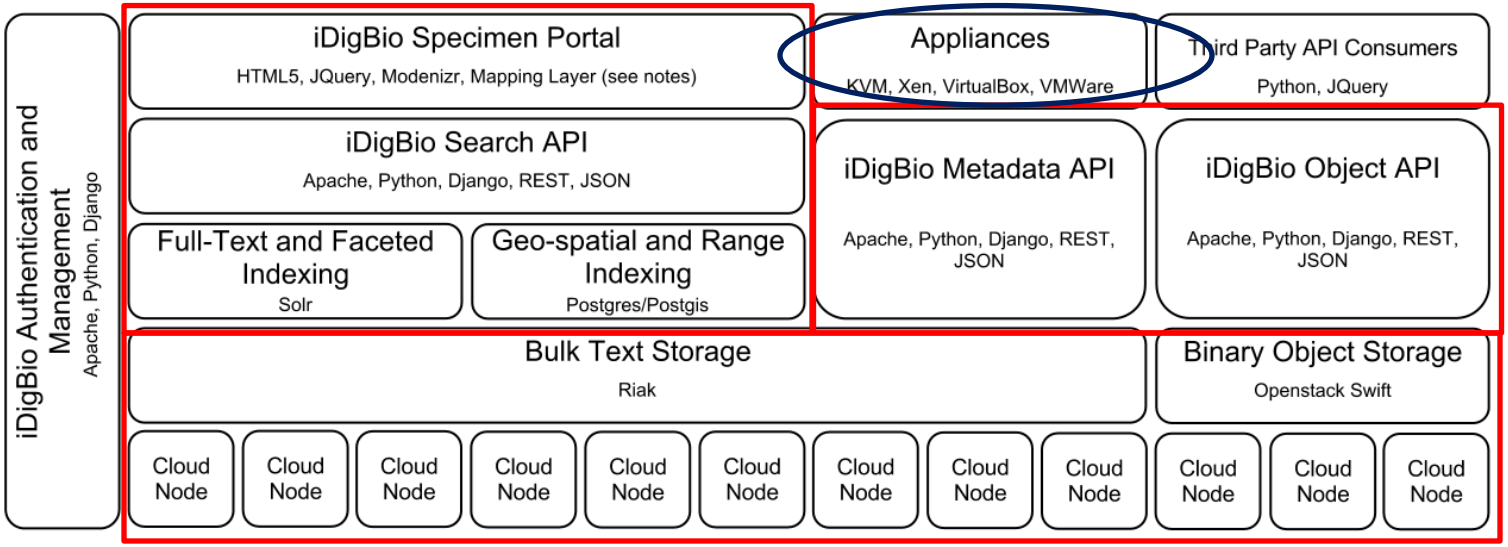


Image ingestion appliance use cases

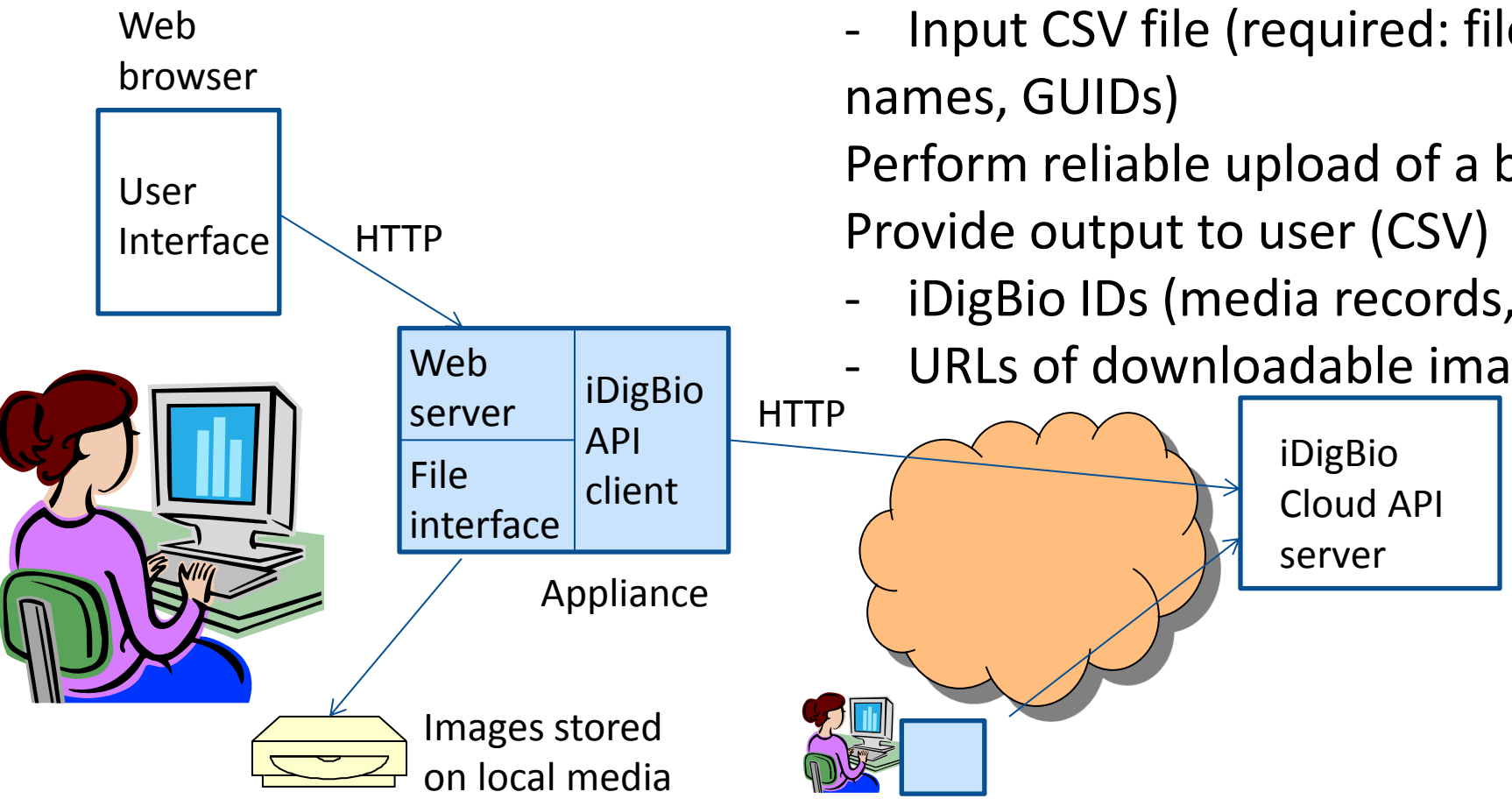
- Allow end-users to ingest images to the iDigBio cloud to provide a mechanism to support crowd-sourcing
  - Ingest batches of images
  - Assign GUIDs to media information
  - Image searchable on the portal through media information
  - Generate lists of media information for each batch including the 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

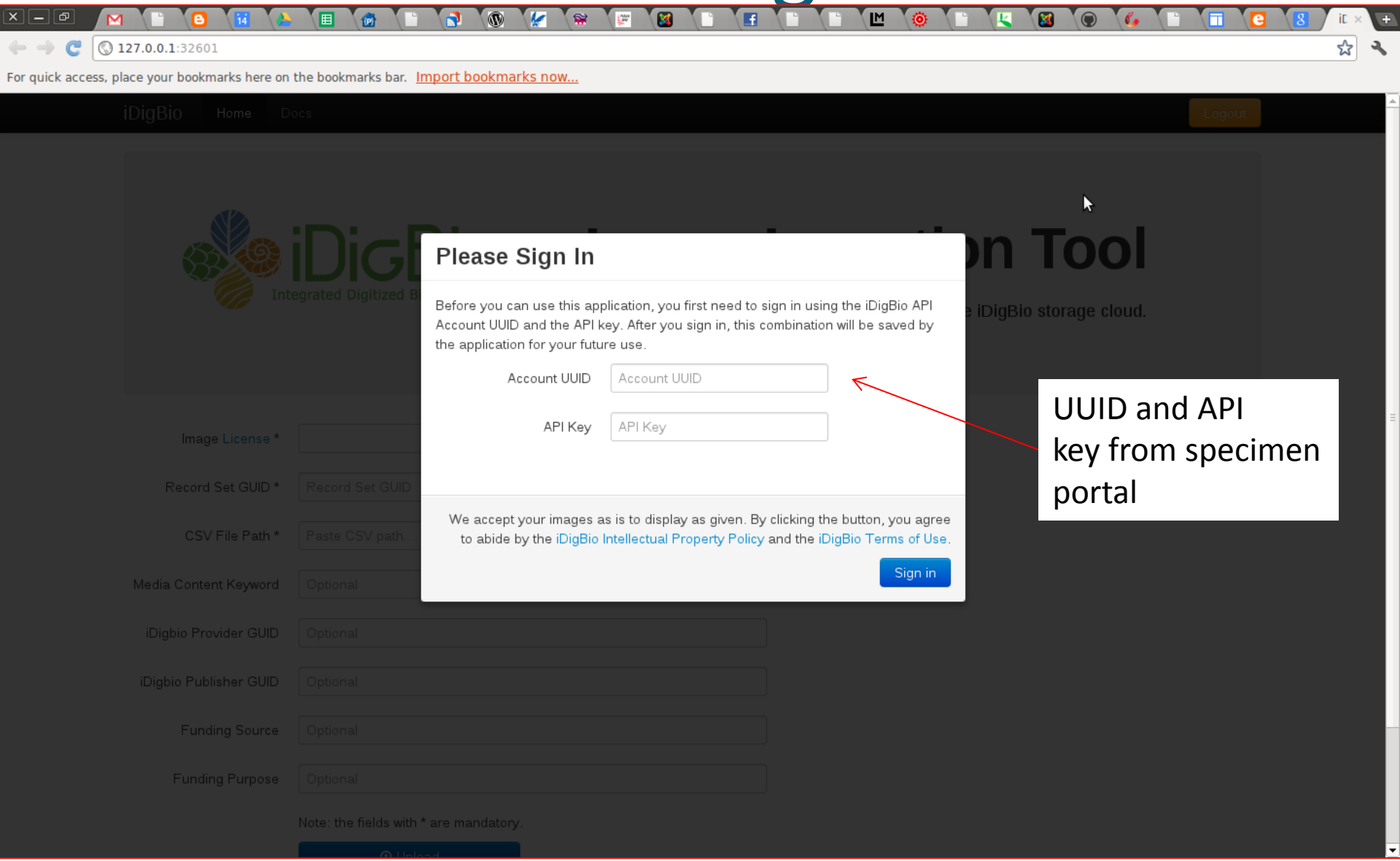
Obtain user inputs defining a batch:
- Web user interface
- Input CSV file (required: file names, GUIDs)

Perform reliable upload of a batch

Provide output to user (CSV)
- iDigBio IDs (media records, media)
- URLs of downloadable images

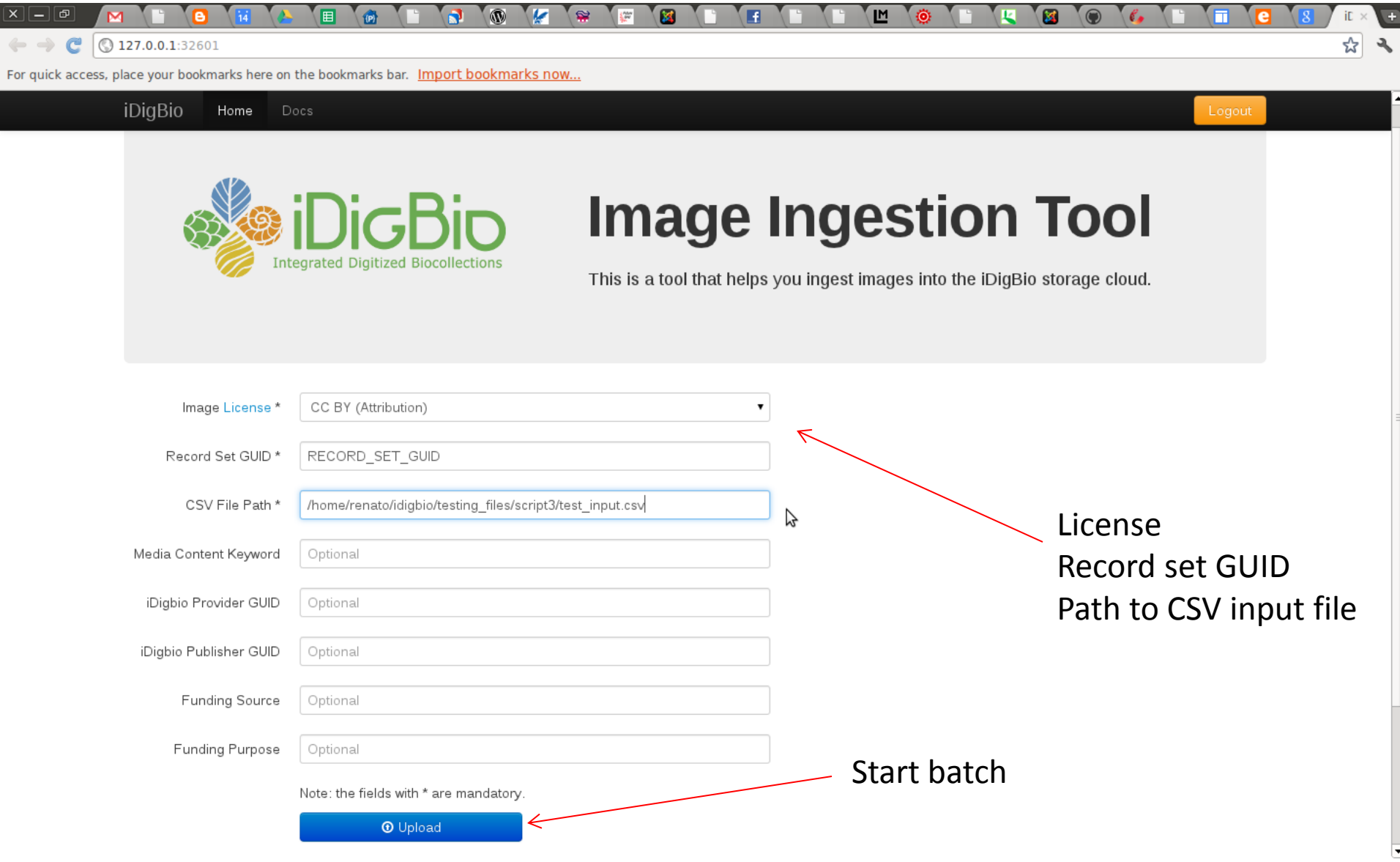


User interface - login



UUID and API key from specimen portal

User interface – select batch



License
Record set GUID
Path to CSV input file

Start batch

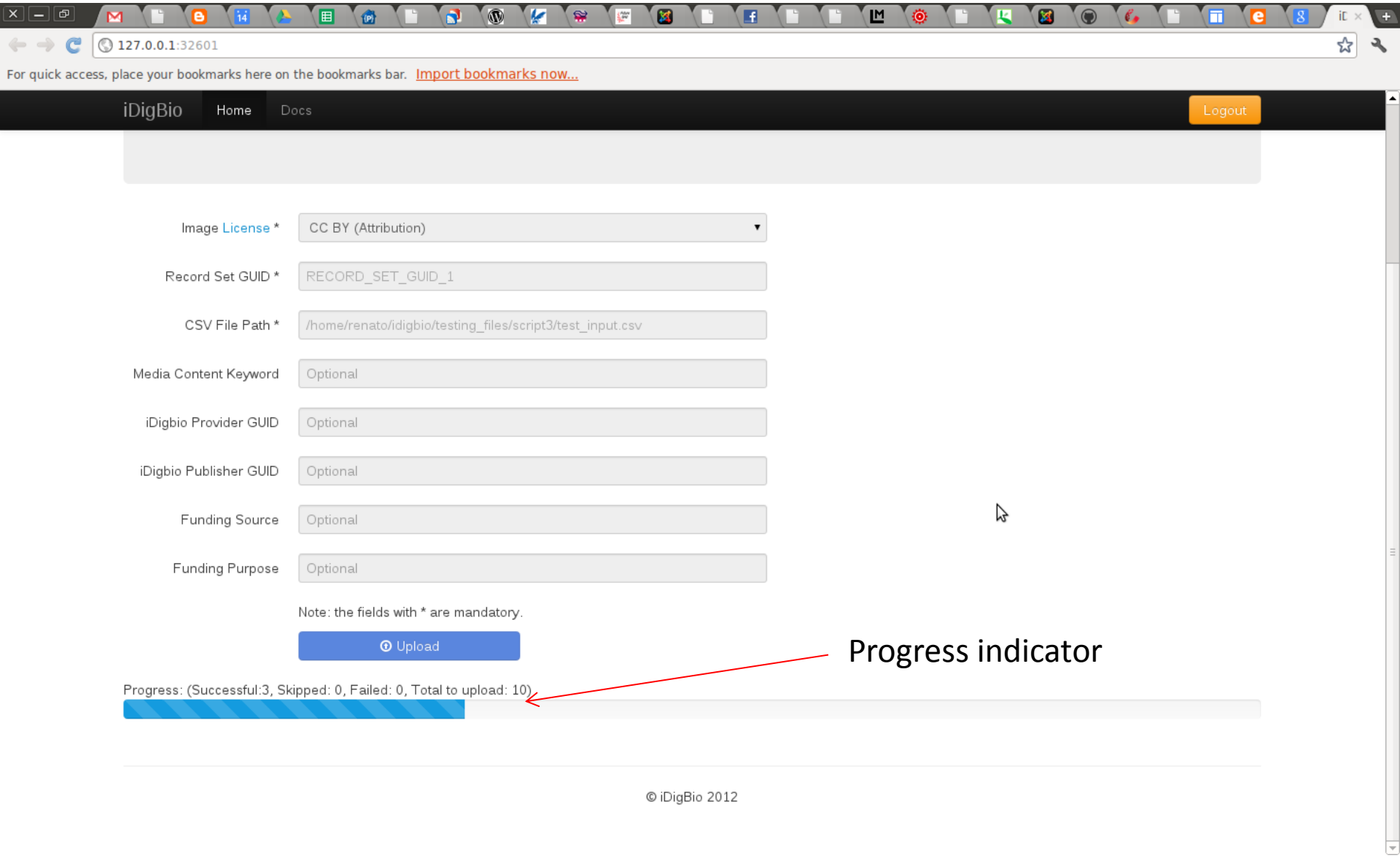
User interface – input CSV

idigbio:OriginalFileName	idigbio:MediaGUID	idigbio:Description	idigbio:Title	id
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

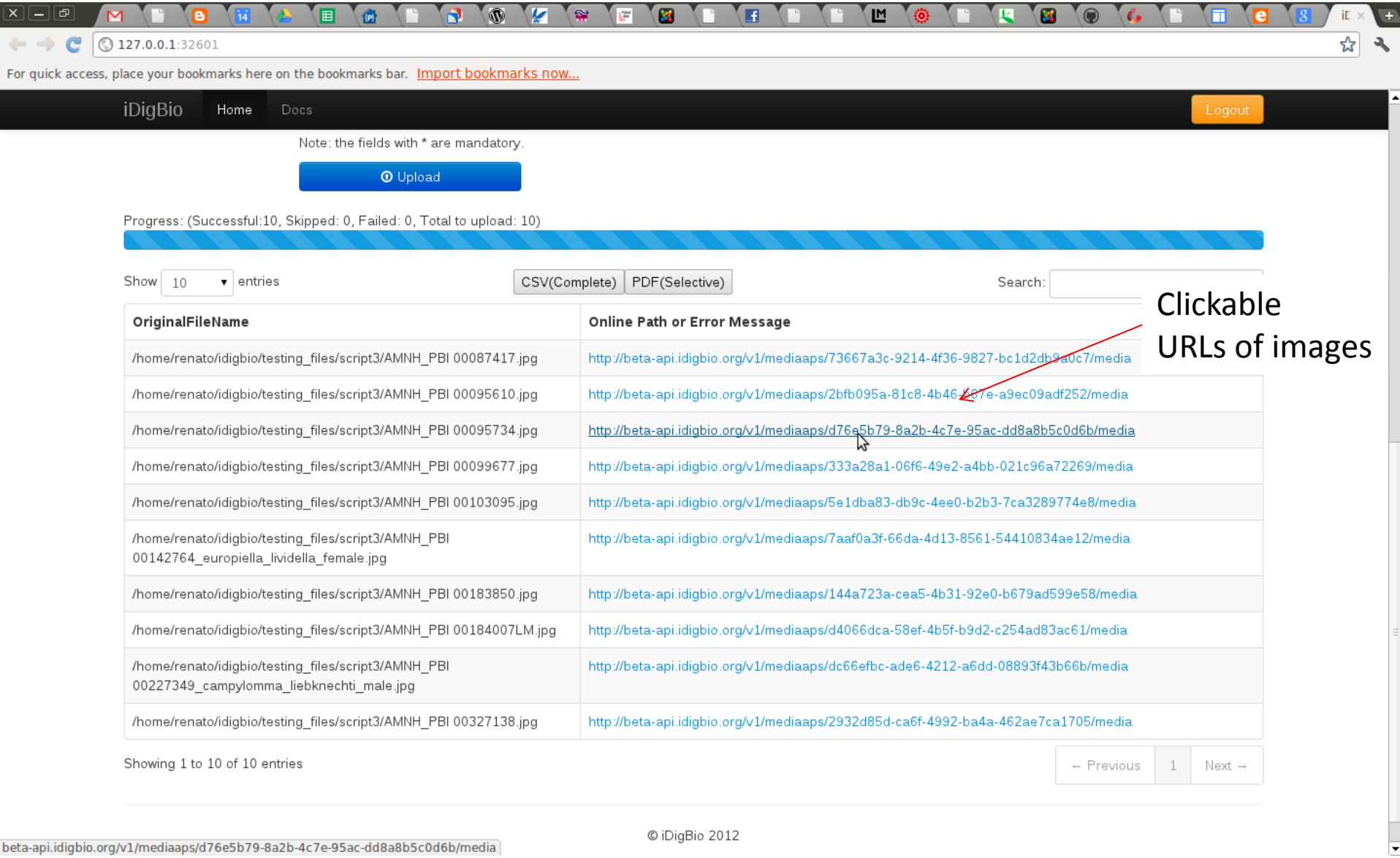
Required terms

Optional terms:
Description
LanguageCode
Title
DigitizationDevice
NominalPixelResolution
Magnification
OcrTechnology
InformationWithheld

User interface – upload

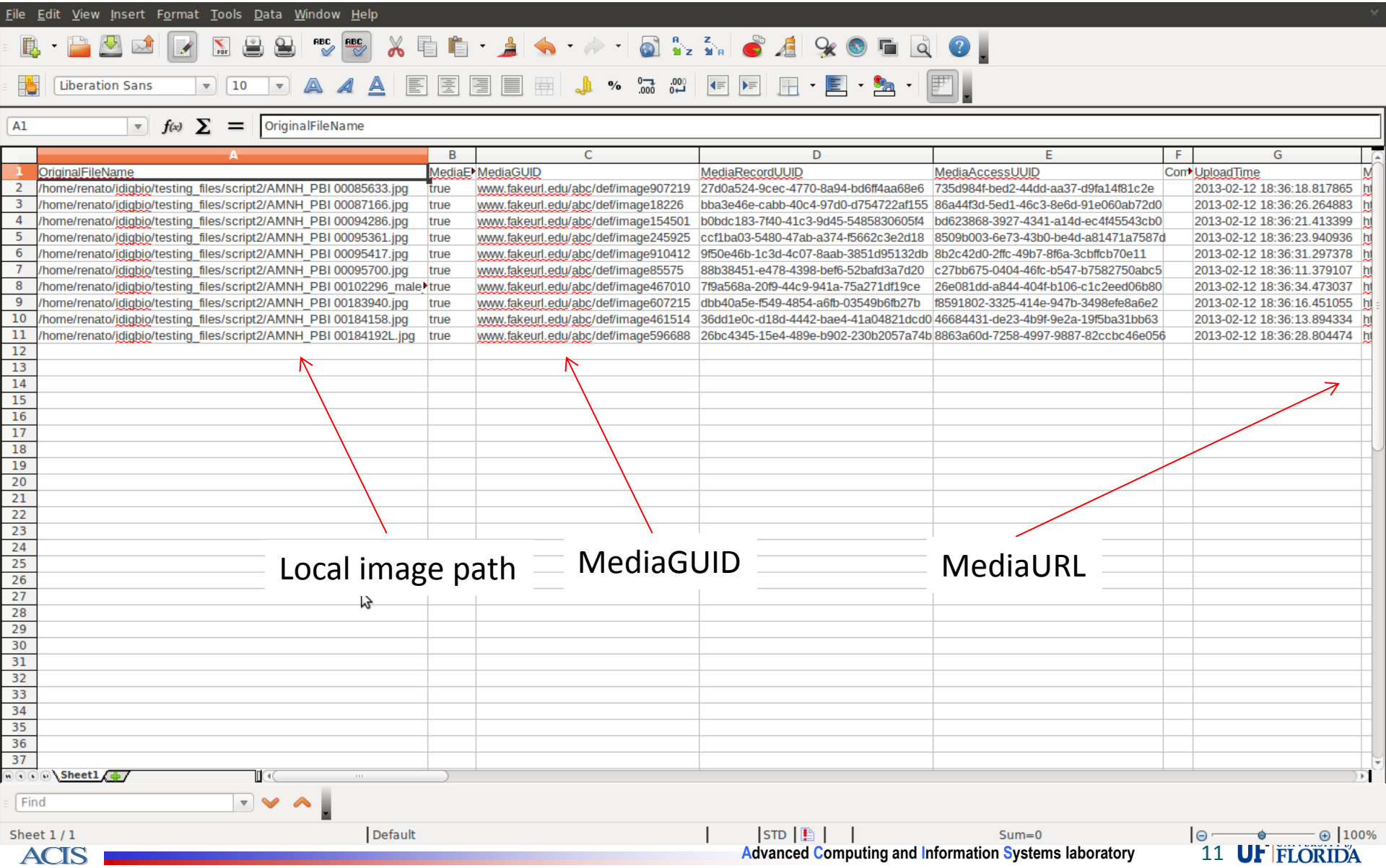


User interface – output



Clickable URLs of images

User interface – output



	A	B	C	D	E	F	G
1	OriginalFileName	MediaE	MediaGUID	MediaRecordUUID	MediaAccessUUID	Com	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-d9fa14f81c2e		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	b0bdc183-7f40-41c3-9d45-548583060 5f4	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-3cbffcb70e11		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-bae4-41a04821dcd0	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

CSV output, MISC terms

OriginalFileName MediaExists MediaGUID
MediaRecordUUID
MediaAccessUUID Comments UploadTime MediaURL
Description LanguageCode Title DigitalizationDevice
NominalPixelResolution Magnification OcrOutput OcrTechnology
InformationWithheld MediaMD5 MimeType
MediaSizeInBytes ProviderCreatedTimeStamp
providerCreatedByGUID
MediaRecordEtag RecordSetUUID iDigbioProvidedByGUID
MediaContentKeyword 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