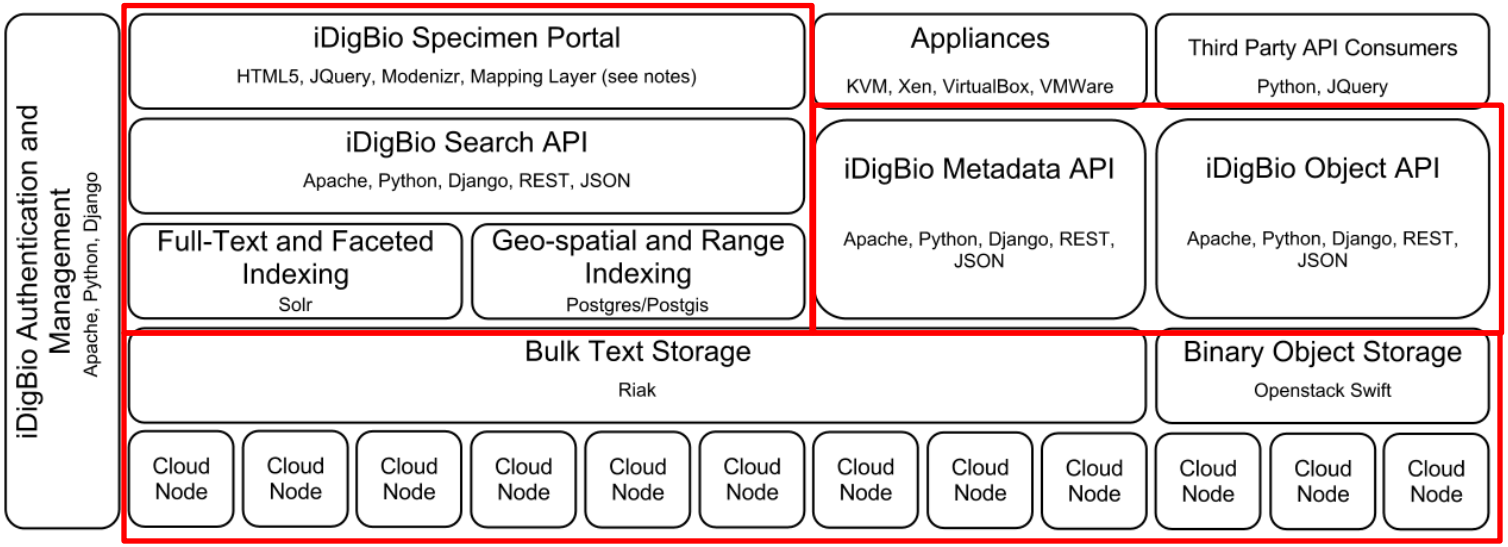
iDigBio Nuts and Bolts: Appliances Present and Future

Renato Figueiredo
(on behalf of the
iDigBio IT team)

iDigBio Summit
Nov 20th, 2013

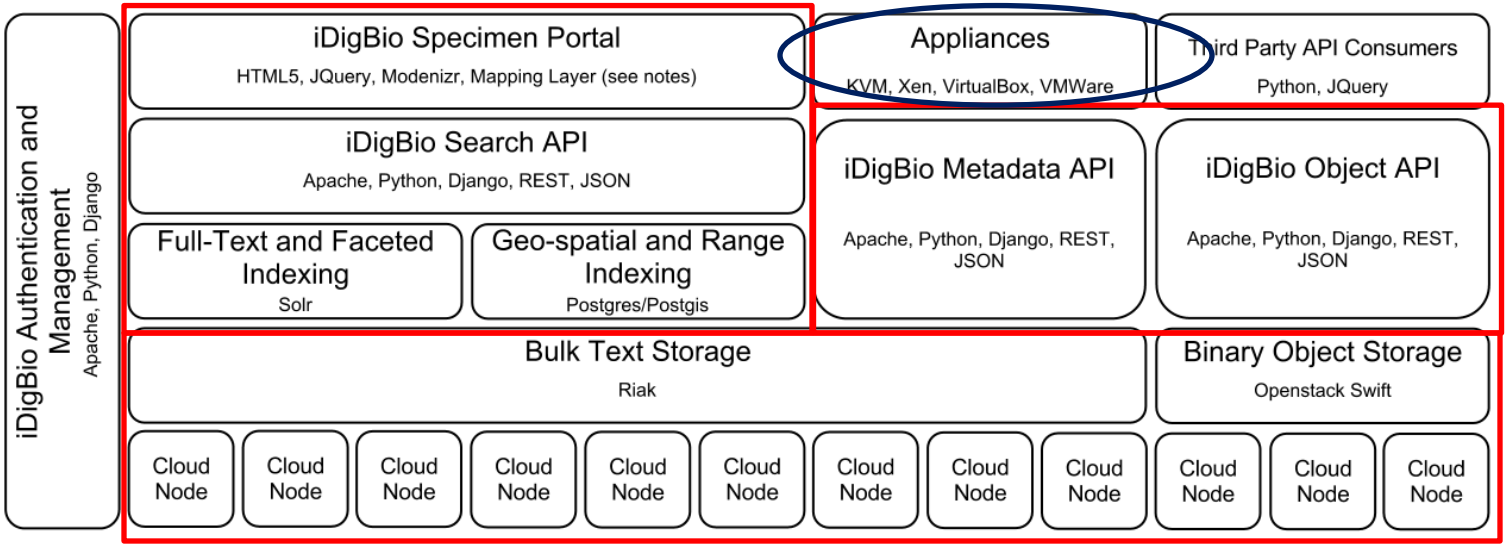
Overview

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



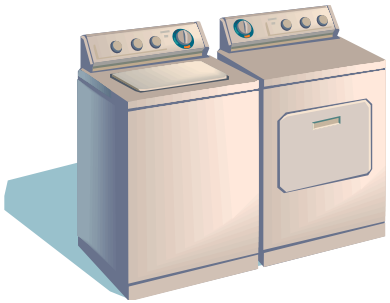
Overview

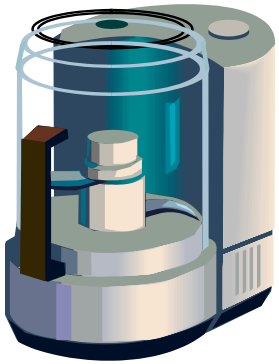
- Appliances complement the cyber-infrastructure core
  - Functionality desired on the client; hide low-level iDigBio APIs, expose user-friendly interface (e.g. image ingestion)
  - Package tools of general interest to the community in virtual machines for ease of software deployment



What is an appliance?

- Physical appliances
  - Webster – “an instrument or device designed for a particular use or function”





What is an appliance?

- Hardware/software appliances
  - DVR: TV receiver + computer + hard disk + Linux; channel browsing user interface

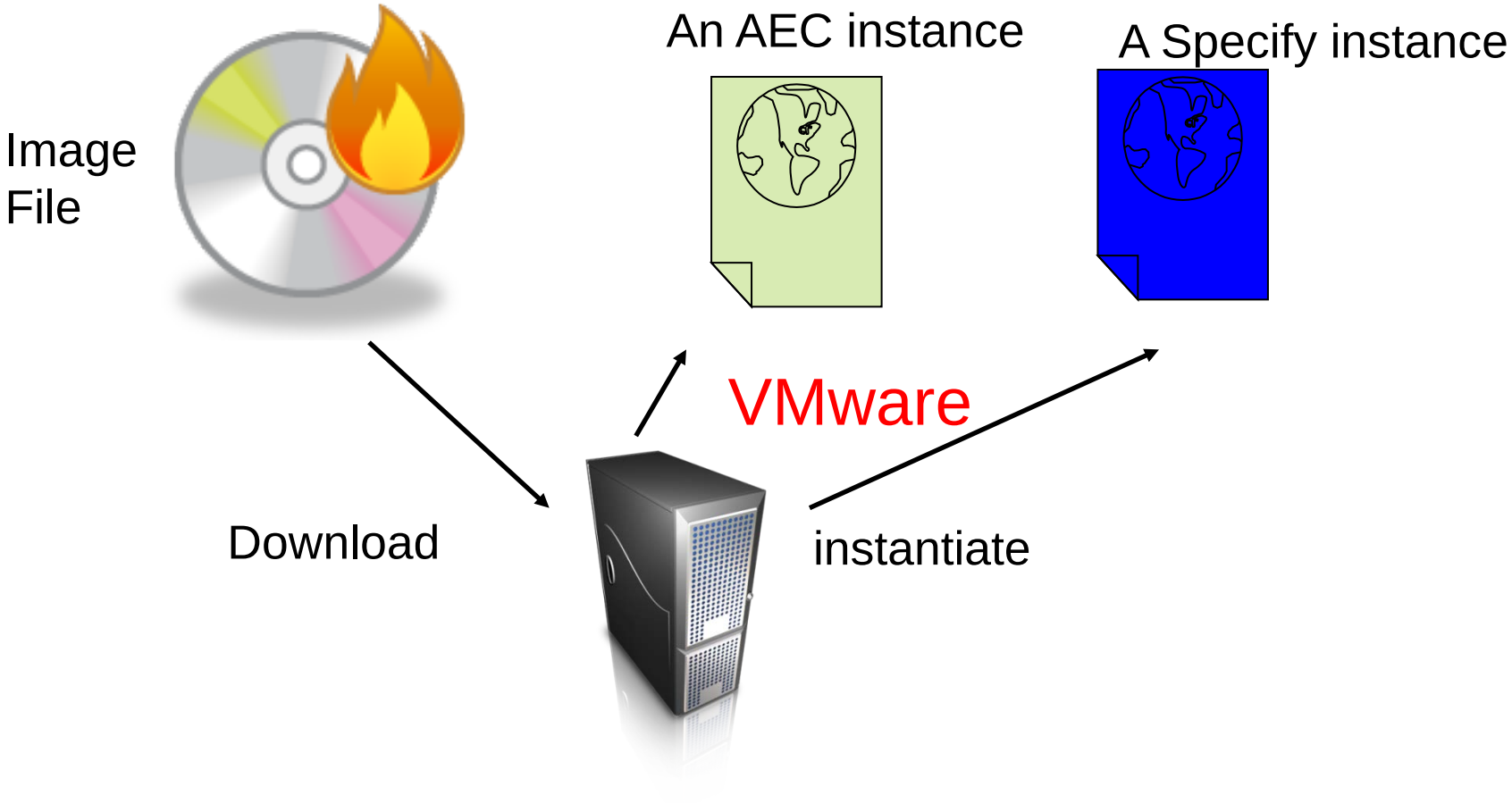
  - Wireless router: Computer + wireless radio + modem+ FreeBSD; Web configuration user interface

What is a virtual appliance?

- A virtual appliance packages all software and configuration needed for a particular purpose into a virtual machine “image”
- The “image” is just a file (a large file)
  - Download, copy, remove
- The image can be *instantiated* to run on computer hardware using virtual machine software
  - VMware, VirtualBox

Virtual appliance example

- Linux + Apache + MySQL + PHP + (tool)



Virtual appliance use cases

- Technology dissemination and evaluation
  - Package your nifty tool(set) as an appliance (we can help)
  - Users can quickly download and try it out
  - No need to worry about installing other software
  - Can remove easily
- Training, workshops
  - Avoid problems with software installation, dependences
  - Speed up setup time
  - Count on consistent environment for all attendees (even post-workshop)
- Turnkey toolsets for small collections
  - Deploy on local servers/desktops (or cloud resources)

Current virtual appliances

- AEC (Arthropod Easy Capture)
  - Web server, database, PHP, AEC
- Specify
  - Thin and thick client
  - Web server, database, Java, Specify
- Geo-referencing calculator
  - Web server, Java, Web browser

https://www.idigbio.org/wiki/index.php/IDigBio_Virtual_Appliances

short link: http://bit.ly/1e6ExvF

What do I need to run an appliance?

- Virtual machine software
  - VMware or VirtualBox
  - Install once to run any appliance
- Download and unzip appliance .zip file
- Double-click to run
- Access tool interface
  - Through Web browser
  - Through VNC (remote desktop)

Demo time

Under the hood: VM software

- Two major desktop-class options:
  - VMware – leading commercial technology
    - simplest to install and run
    - Player (Windows, Linux) – free
    - Fusion (MacOS) – not free; 30-day trial
  - VirtualBox – open-source technology
    - Windows, Linux and MacOS
- Appliances:

Under the hood: VM software

- VM software starts and runs appliance(s) on a computer
  - Each appliance looks like yet another window; but it's a full-featured "sub-computer"
- What runs within an appliance is completely independent from what runs in the appliance's host
  - And from what runs within another appliance
- Key feature that allows customization
  - Package once; run in many different environments
    - (Some differences in user interface of the VM software)
- Example:
  - MySQL on host, Specify appliance, AEC appliance
  - Different versions, tables

Demo revisited

Image ingestion appliance

- 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

Image ingestion appliance

Packaged as native application;
can go inside VM as well

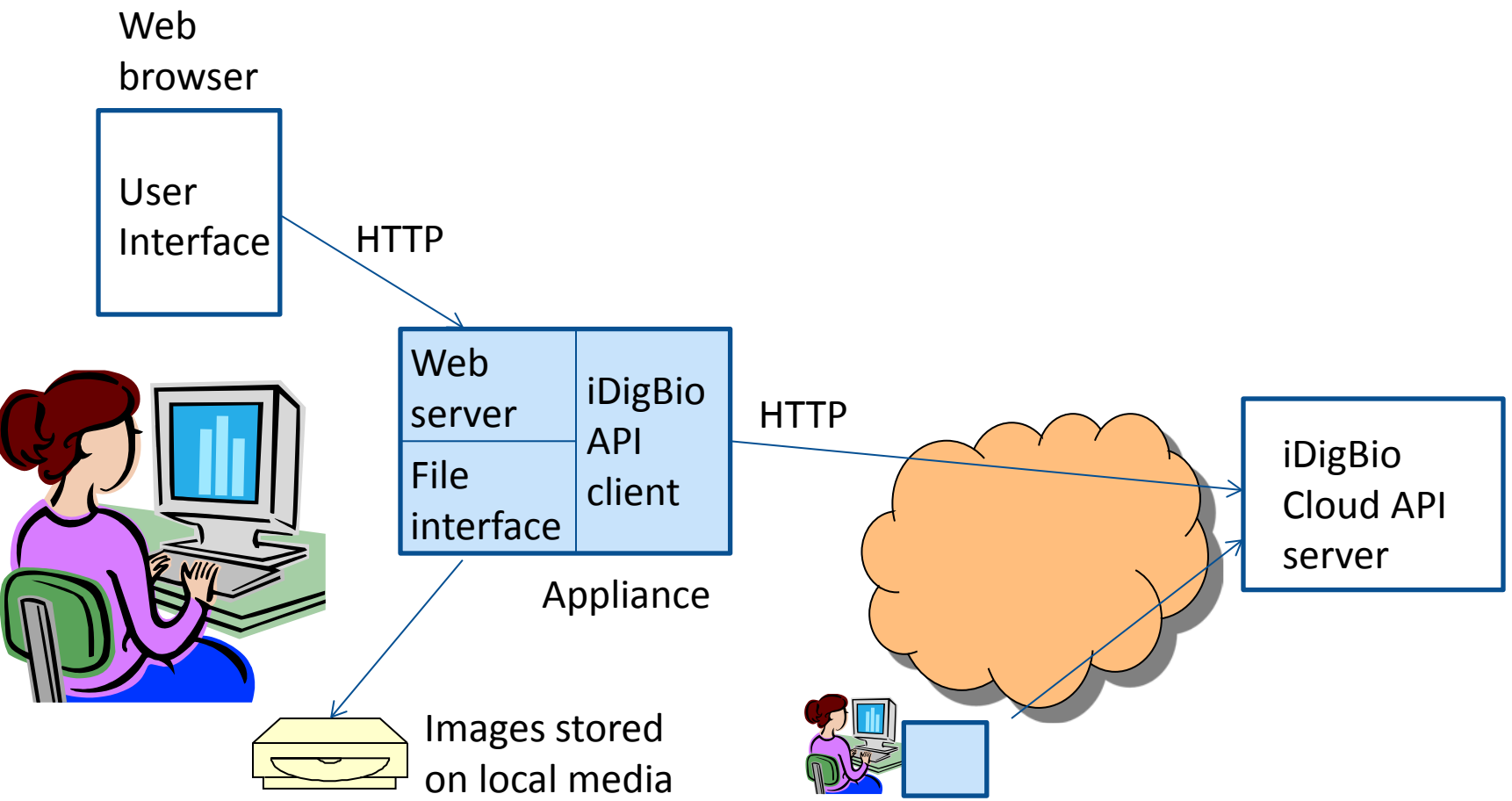


Image ingestion appliance demo

https://www.idigbio.org/wiki/index.php/CYWG_iDigBio_Image_Ingestion_Appliance

Short link: http://bit.ly/11SHQiO

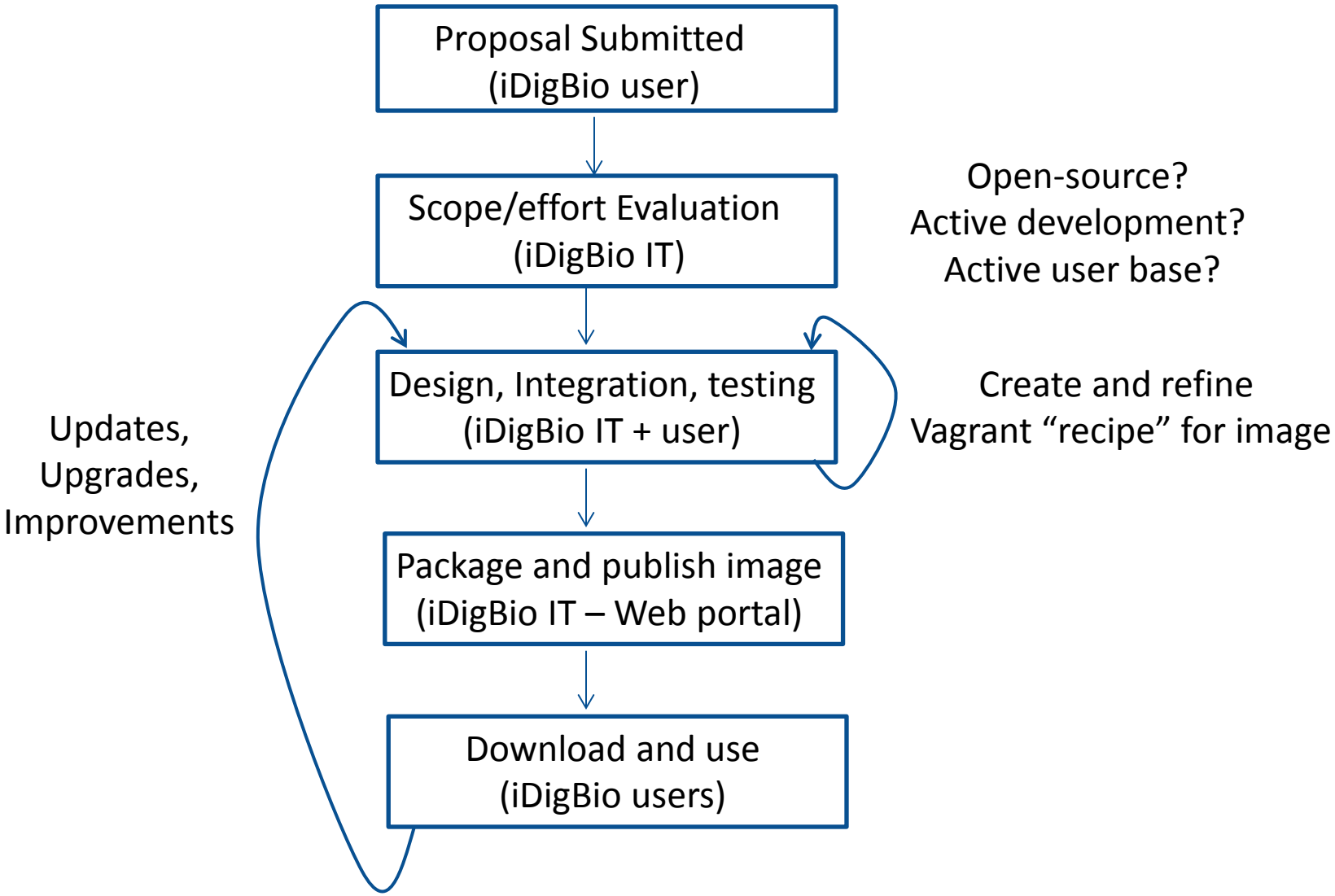
Virtual appliance futures

- Future appliances – those there are of interest to, and brought to us by the community

- We have a process in place to collaborate with the community to integrate, package, distribute appliances
  - Note: scope is to integrate existing tools in appliances – not develop new software

- It starts here:
- https://www.idigbio.org/content/appliance-proposal

Appliance production workflow



Discussion

- What are use cases of most interest to you?
  - To the community at large?
- What specific tools would be of most interest to you?
  - To the community at large?
- How to effectively disseminate the availability of appliances and the process to create them?
- Perceived barriers to adoption and how to address them?
- Individual tools vs. toolsets?
- …

Acknowledgments

- iDigBio IT team
  - Jose Fortes, Andrea Matsunaga, Matt Collins, Joanna McCaffrey, Reed Beaman, Debbie Paul: use cases, requirements, testing and feedback
  - Appliance developers: Kyuho Jeong, Yonggang Liu, Alex Thompson

This material is based upon work supported by the National Science Foundation under Cooperative Agreement EF-1115210. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.