



Curious by **nature.**

# Digitization Workflow Automation at BRIT

Jason Best  
Botanical Research Institute of Texas



BOTANICAL RESEARCH  
INSTITUTE OF TEXAS

Presented 13 August, 2016 at TORCH VIII 2016 at BRIT, Fort Worth, Texas

Plant to planet.



# The Vanderbilt Collection

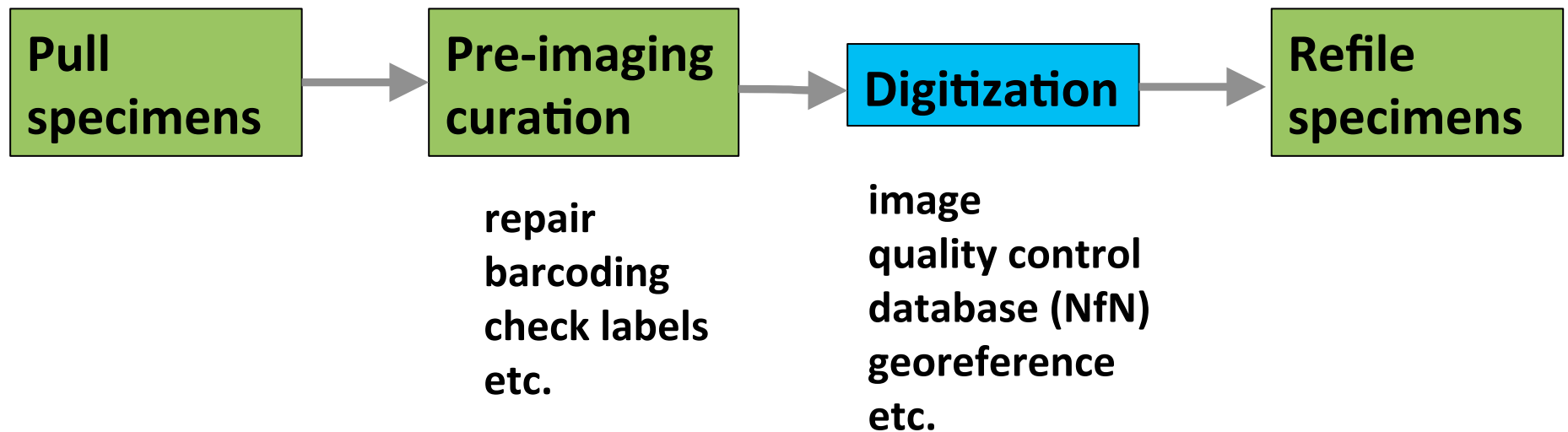
**In 1997, Vanderbilt University donated ~360k specimens to BRIT. The vast majority of these specimens are from the southeastern states of the US.**



Plant to planet.



# Workflow Overview



Plant to planet.



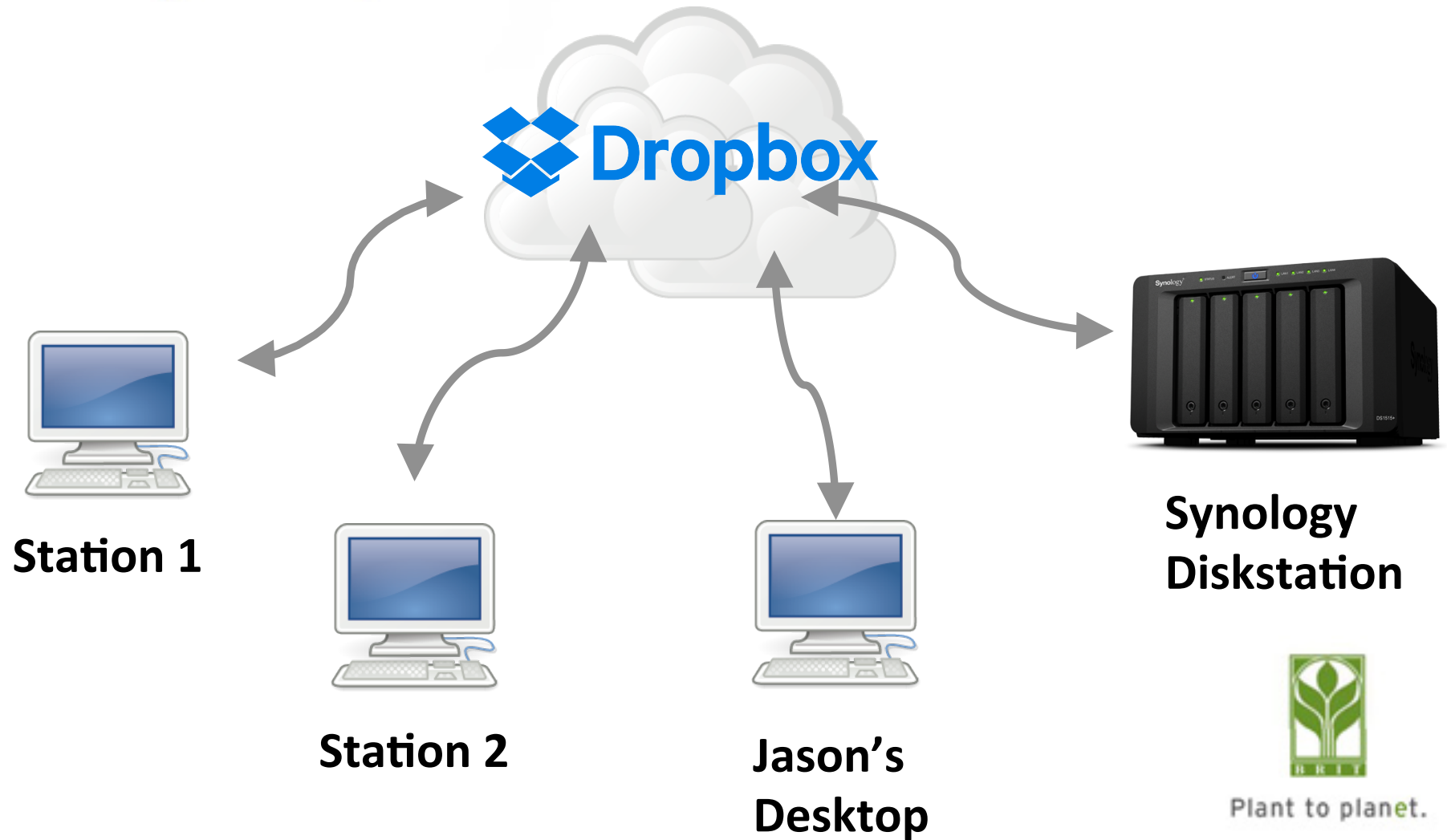
# Imaging Engine



Plant to planet.

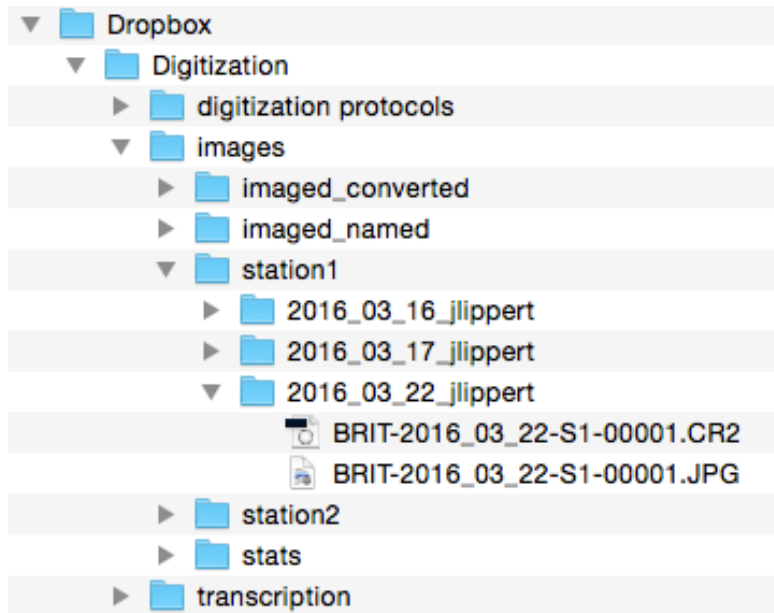


Explorers **by nature.**





Explorers by nature.



**Station 1**



**Station 2**



Plant to planet.



Explorers **by nature.**

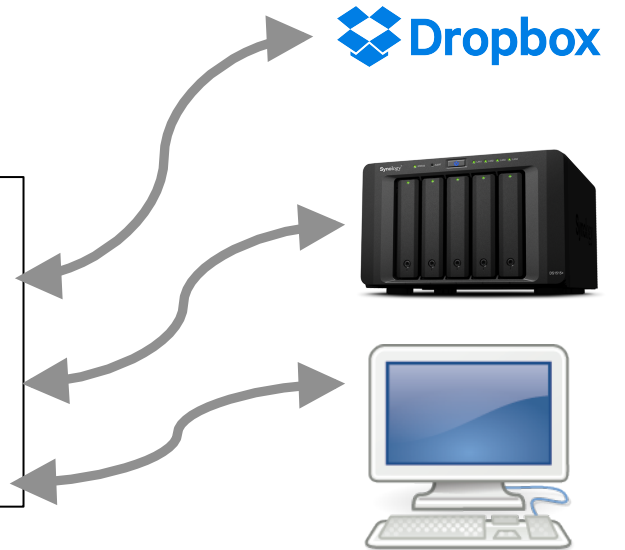


**station1**

**2016\_03\_16\_jlippert**

**BRIT-2016\_03\_22-S1-00001.CR2**

**BRIT-2016\_03\_22-S1-00001.JPG**



Plant to planet.





Explorers **by nature.**

# Summary thus far

**Off the shelf:**

**Products**

**Services**

**General infrastructure**

**Provides:**

**File management and automated sync**

**Local redundancy**

**Automated cloud backup**



Plant to planet.





# Daily imaging session

**Typical process:**  
**camera adjustments**  
**software check**  
**etc.**  
**create session folder...**





Explorers by nature.

# Folder and file names

station[#]

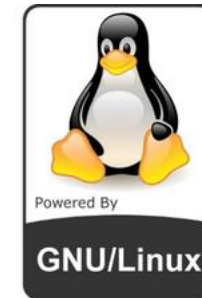
[YYYY]\_[MM]\_[DD]\_[user\_id] (Daily session folder)

BRIT-[YYYY]\_[MM]\_[DD]-S[#]-[nnnnnn].CR2

BRIT-[YYYY]\_[MM]\_[DD]-S[#]-[nnnnnn].JPG



Plant to planet.



# Automation

**Open source Python scripts supporting  
workflow automation:**

[github.com/jbest/digitization\\_tools](https://github.com/jbest/digitization_tools)



Plant to planet.



# **productivity.py**

**Extracts username from session**

**Logs all images**

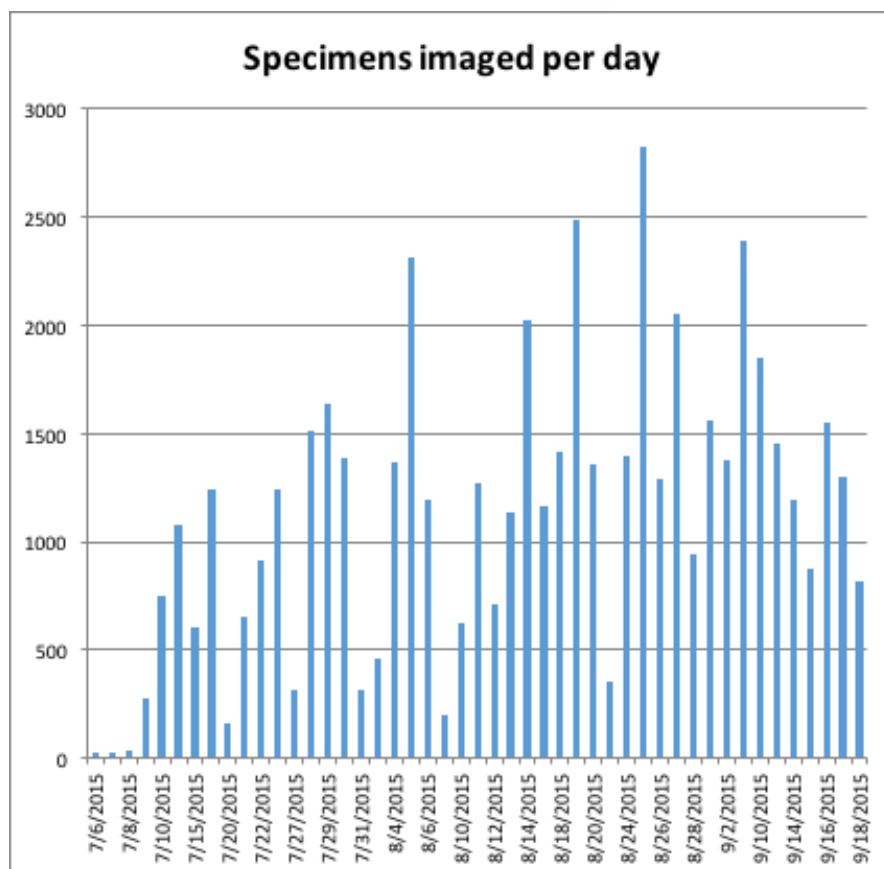
**Reads file creation timestamp**

**Calculates:**

- **time per specimen**
- **moving average**
- **cumulative time**



Plant to planet.



#### All Periods

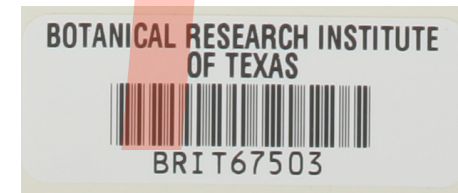
Total specimens (est.)	200,000
Image count	173,857
Percent completed this period	86.93
Average seconds/specimen	21.66
Average specimens/min	2.77

#### This Period

Total specimens (est.)	200,000
Image count	4,396
Percent completed this period	2.20
Average seconds/specimen	19.24
Average specimens/min	3.12



Plant to planet.



# Image Naming

JPEGs are scanned using ZBar module

Extracted barcode value is used to name JPGs and CR2 files

This step is an additional QC phase, confirming barcode is readable



Plant to planet.



Explorers **by nature.**

# **File Conversion**

**Convert raw Canon CR2 files to DNG**

**Many ways to read DNG files, but few to write**

**Adobe DNG Converter only on Mac and Win**

**But...**

**It can be made to work with WINE on Linux**



Plant to planet.





# Metadata

Write additional metadata into each file:  
Imager name, copyright, etc.

station[#]

[YYYY]\_[MM]\_[DD]\_[user\_id] (Daily session folder)

BRIT-[YYYY]\_[MM]\_[DD]-S[#]-[nnnnnnn].CR2

BRIT-[YYYY]\_[MM]\_[DD]-S[#]-[nnnnnnn].JPG



Plant to planet.



Explorers **by nature.**

# TODO

## Version 1.0

**Refine, test, and document a few scripts and release:**

**barcode.py**

**metadata.py**

**convert.py**

**Create a simple GUI**

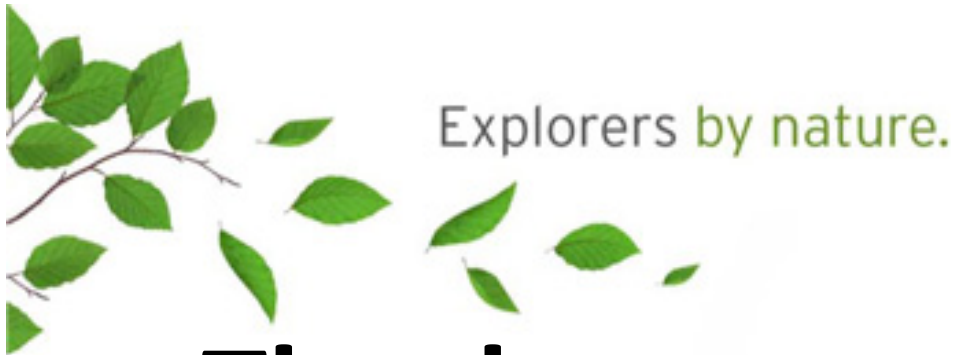
## Version 2?

**Full workflow management through a web interface**

**Plug and play digitization workflow appliance**



Plant to planet.



# Thanks

**National Science Foundation**

**SERNEC TCN:**

**Zack Murrell, Joey Shaw, Dwayne Estes, Ashley Morris, Michael Denslow,  
Herrick Brown**

**BRIT:**

**Tiana Rehman, Miranda Madrid, Peter Fristch**

**Joe Lippert, Jani Okai**

**BRIT Volunteers and Volunteer Coordinator Julie Donovan**



**Plant to planet.**