

# Mobilizing New England vascular plant data to track environmental change

(NEVP)

P.W. Sweeney Yale Peabody Museum of Natural History





#### **OBJECTIVES**

Digitize 1.3 million
 N.E. vascular plant
 specimens from 15
 regional herbaria



## **DIGITIZATION PLAN: ORGANIZATION**

- Brown University (BRU)
- Harvard University (HUH)
- U. of New Hampshire (NHA)
- U. of Massachusetts Amherst (MASS)
- U. of Vermont (VT)
- Yale University (YU)
- Bartlett Arboretum (BART)
- Berkshire Museum (BERK)
- Boston University (BSN)
- Central Connecticut State U. (CCSU)

- Connecticut College (CCNL)
- Harvard Forest (HF)
- Keene State (KESC)
- Western Connecticut State U. (WCSU)
- Westfield State U. (WSCH)
- University of Oklahoma
- North Carolina State U.

#### **DIGITIZATION PLAN: WORKFLOW**



## **DIGITIZATION PLAN: WORKFLOW**



#### **WORKFLOW: PRE-CAPTURE**

 Capture data that reflects the physical storage structure of the collection before imaging and data basing individual specimens (e.g., Taxon & State)



#### **WORKFLOW: PRE-CAPTURE**



Cardamine concatenata (Michx.) Sw. Connecticut YU



#### {"m1p":"[NEVP

TCN]","m2v":"1.1","g":"Dryopteris","s":"palustris", "ir":"variety","i":"pubescens","sn":"Dryopteris palustris variety pubescens (Lawson) Nakai","a":"(Lawson)

Nakai","st":"Massachusetts","sr":"3.9","cc":"YU"}

#### **WORKFLOW: PRE-CAPTURE**

Pre-capture complete at three institutions, ongoing at two - approximately 700,000 specimens with at taxon captured,400,000 with state or lower geographic level

 Capture an image, a barcode number, a subset of label data & associate precapture data with specimen occurrence records



To increase the efficiency of capturing an image and specimenlevel data, we are developing a high throughput digitization apparatus



- Development complete, 1st installed at Harvard.
- Aug.-Oct. refinement and testing
- In beta-testing phase, with production runs
  - per specimen throughput rates are about 35s











 interchange standard using RDF/XML, the W3C
 OpenAnnotationOntology, and DarwinCore was
 defined, and code was
 added to Symbiota to
 support the ingest of
 these documents into
 Symbiota.

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
    xmlns:dwcFP="http://filteredpush.org/ontologies/oa/dwcFP.owl#"
    xmlns:foaf="http://xmlns.com/foaf/0.1/"
    xmlns:cnt="http://www.w3.org/2011/content#"
    xmlns:oa="http://www.w3.org/ns/oa#"
    xmlns:co="http://purl.org/ontology/co/core#"
    xmlns:oad="http://filteredpush.org/ontologies/oa/oad.rdf#"
    xmlns:dwc="http://rs.tdwg.org/dwc/terms/"
    xmlns:dc="http://purl.org/dc/elements/1.1/"
    xmlns:obo="http://purl.obolibrarv.org/obo/"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    <rdf:Description rdf:about="urn:uuid:b7196d9f-8395-f630-73dd-7c0900009a4e">
        <!-- Comment that follows records the svn commit of this document, it is only need
        <rdfs:comment xml:lang="en">$Id: NEVP_specimen_example.xml 2222 2013-03-31 23:38:5
        <rdfs:comment xml:lang="en">Example of a new occurrence (new specimen record) expr
        <!-- co:count can provide a counter for the number of annotations serialized in th
        <co:count xml:type="xsd:integer">1</co:count>
    </rdf:Description>
    <!-- UUIDs are almost always guids for instances in this document, exceptions are: (1)
    <oa:Annotation rdf:about="urn:uuid:5af81a7f-a1c3-dcf5-997a-ed5e00009a2b">
        <oa:hasTarget>
            <oa:SpecificResource rdf:about="urn:uuid:f6fa574a-a45d-5be5-0b1b-b0750000a237</pre>
                <oa:hasSelector>
                    <oad:KVPairQuerySelector rdf:about="urn:uuid:f16d4efb-e216-0e57-a3c6-0</pre>
                        <dwc:collectionCode>YU</dwc:collectionCode>
                        <dwc:institutionCode>Yale Peabody Museum</dwc:institutionCode>
                    </oad:KVPairQuerySelector>
                </oa:hasSelector>
                <oa:hasSource rdf:resource="http://filteredpush.org/ontologies/oa/oad.rdf#</pre>
            </oa:SpecificResource>
        </oa:hasTarget>
        <oa:hasBody>
            <dwcFP:0ccurrence rdf:about="urn:uuid:3b5c06ce-aa39-95e4-4f0b-d4890000a0e1">
                <dc:type>PhysicalObject</dc:type>
                <dwcFP:hasBasisOfRecord rdf:resource="http://rs.tdwg.org/dwc/dwctype/Prese</pre>
                <dwc:catalogNumber>YU.057764</dwc:catalogNumber>
                <dwcFP:hasCollectionByID rdf:resource="http://biocol.org/urn:lsid:biocol.c</pre>
                <dwc:collectionCode>YU</dwc:collectionCode>
                <dwcFP:hasIdentification>
                    <dwcFP:Identification rdf:about="urn:uuid:e6788bbb-01b9-59da-7d7f-fb99</pre>
                        <dwcFP:isFiledUnderNameInCollection>YU</dwcFP:isFiledUnderNameInCollection>YU
                        <dwc:scientificName>Acer rubrum var. tridens</dwc:scientificName>
                        <dwc:genus>Acer</dwc:genus>
                        <dwc:specificEpithet>rubrum</dwc:specificEpithet>
                        <dwcFP:infraspecificRank>var.</dwcFP:infraspecificRank>
                        <dwc:infraspecificEpithet>tridens</dwc:infraspecificEpithet>
                        <dwc:scientificNameAuthorship>Alph.Wood</dwc:scientificNameAuthors
                        <dwc:identificationOualifier></dwc:identificationOualifier>
```

# **TRAINING & OUTREACH**

Richard Primack's team at Boston **University (BU)** established the **New England** Leaf Out Project (NELOP). 200 volunteers last year

nature's , notebook Nature's Notebook Home USA-NPN Home Contact Us

Connecting People with Nature to Benefit Our Changing Planet

ABOUT US

/ M

OBSERVE

NN Home » Connect Regionally » New England Leaf Out Project

#### New England Leaf Out Project

PROJECT GOALS



Scientists at Boston University need your help monitoring leaf out times of trees in New England (Maine, New Hampshire, Vermont, Massachusetts, Rhode Island and Connecticut). They are interested in how rapidly trees are responding to warming temperatures across the region.

- Are trees leafing out earlier now than in the past?
- Do trees leaf out at different times across New England?

#### JOIN US! IF YOU WOULD LIKE TO PARTICIPATE ...

You are invited to gather observations of **leaf out times** of trees in the **spring of 2013**. You can join this effort by periodically checking individual trees in your area and reporting your observations via *Nature's Notebook* (or by sending your observations to Dr. Richard Primack).

#### HOW TO PARTICIPATE...

**1. Select your plants** - Identify one or more individual trees from the list below to track. The pages for these species include information on how to identify them and how to monitor them.

#### www.usanpn.org/nn/nelop

## **ACKNOWLEDGEMENTS**



National Science Foundation (EF1208829, EF1208835, EF1208972, EF1208973, EF1208975, EF1208989, EF1209149).



#### Symbiota Project

#### **FilteredPush**



iPlant Collaborative<sup>™</sup> Empowering A New Plant Biology



**Biota of North America** 

## TIMELINE

	Year 1	Year 2
Informatics Infrastructure	Set-up infrastructure and digitization workflows Develop & test digitization apparatuses at OU	development, testing, and fine- tuning
Pre-capture		
Primary Digitization		

## **WORKFLOW: ENHANCEMENT**

#### Clients Intermediaries FilteredPush Network Annotation Processor FP-Lite **Specify 6** AnnotationProcessor O O Neutral Disagree Collection Code Insert Catalog Number FP FP-Lite SparglMessaging Scientific Name Mapper FP-Core Scientific Name Message API SPARQL Fuseki API AX-WS Helpe Specify6 Driver TBD Push Determiner: Macklin, James Query 25/16/2012 RDF Handler Specify6-HUH Driver FP Message SQL Driver Domain Configuration Scientific Name Authorship E. J. Palmer 1 **FP** Network Node Workflow for cleaning lists of botanical names Identified By Paul Date Identified 2012/05/15 05-48-45 -----Kepler Kuration FP-Core Triage WSDL & Domain Configuration Schema FP Access Point **Client Tools** Domain FP-Core Configuration JAX-WS Helper Knowledge Analysis Messaging RDF Handler SparalPush Kepler MongoMessaging Knowledge Analysis TBD/Fuseki MongoDB **Client Tools** Morphbank Domain Kepler SparglKnowledge Configuration FP-PHP-Library SparglPushMessaging MonaoDB TDB/Fuseki Dubraillin Dubraillin Drosp FilteredPush instance SparalPush FP-Core Last Mudified: 20. Publish Date: 20. Description: 115 Edit Image: 9 Fedora/Mulagra Akka JMS-KVP/MySQL Analysis **Client Tools** MongoDB Akka FP-PHP-Library Harvester PHP SOAP Client Helper OAI/PMH **Symbiota** Key to Diagram FP-Core Domain Configuration Supporting Library RDF Handler Public/Private FP Component Keypair Current Schema Specific Sianed Implementation Component Encrypted Message Supporting OAI/PMH Provider Shared Software Configuration Symbiota Authorized Files Mapping Keys