Mobilizing New England vascular plant data to track environmental change

(NEVP)

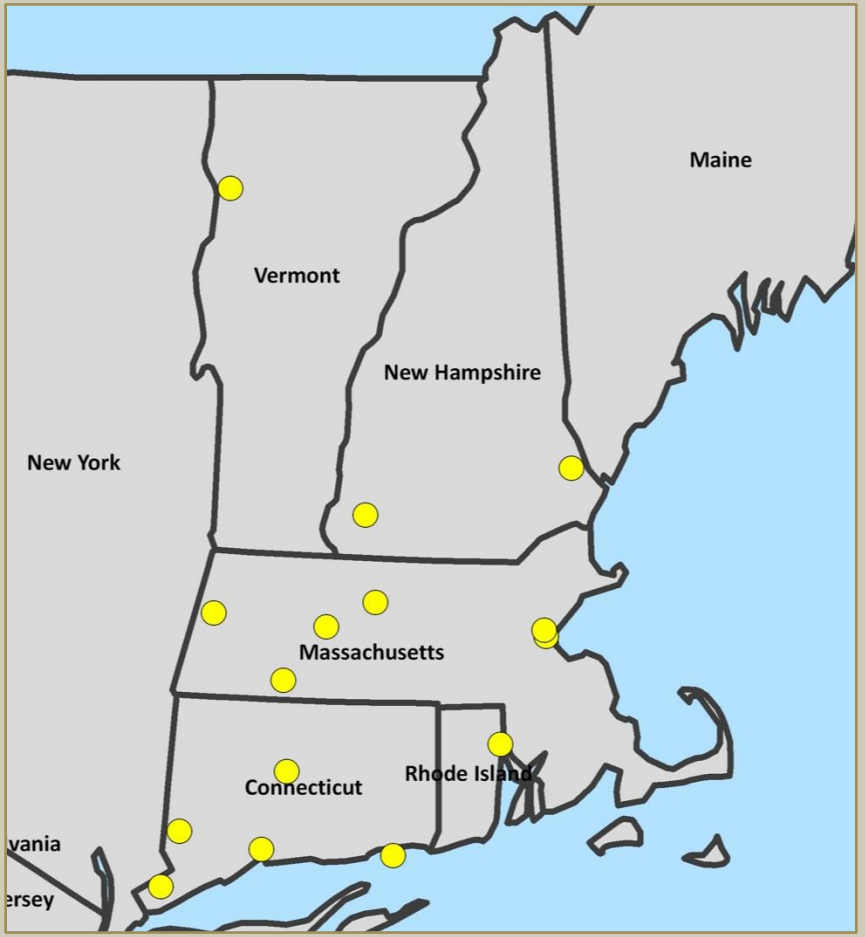
P.W. Sweeney

Yale Peabody Museum of Natural History

Mobilizing New England Vascular Plant Specimen Data to Track Environmental Changes

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

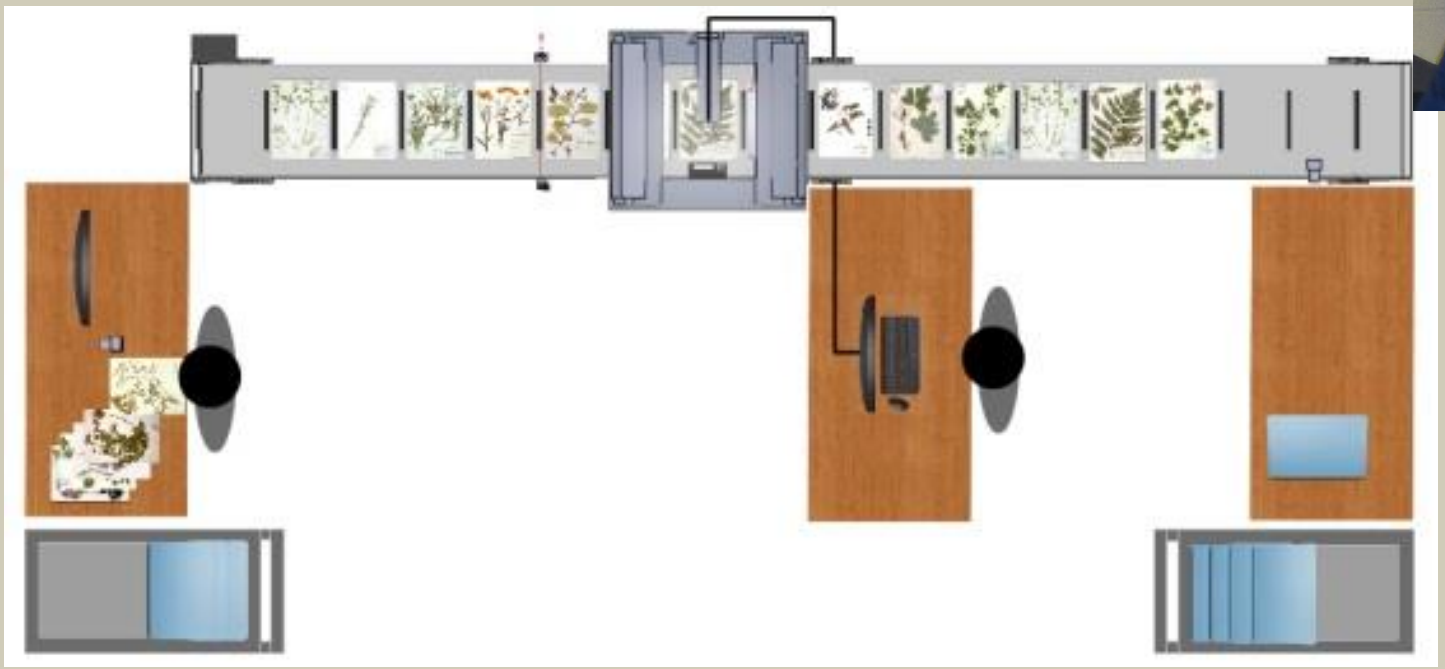
- Pre-capture complete at three institutions, ongoing at two - approximately **700,000** specimens with at taxon captured,400,000 with state or lower geographic level

WORKFLOW: PRIMARY DIGITIZATION

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

WORKFLOW: PRIMARY DIGITIZATION

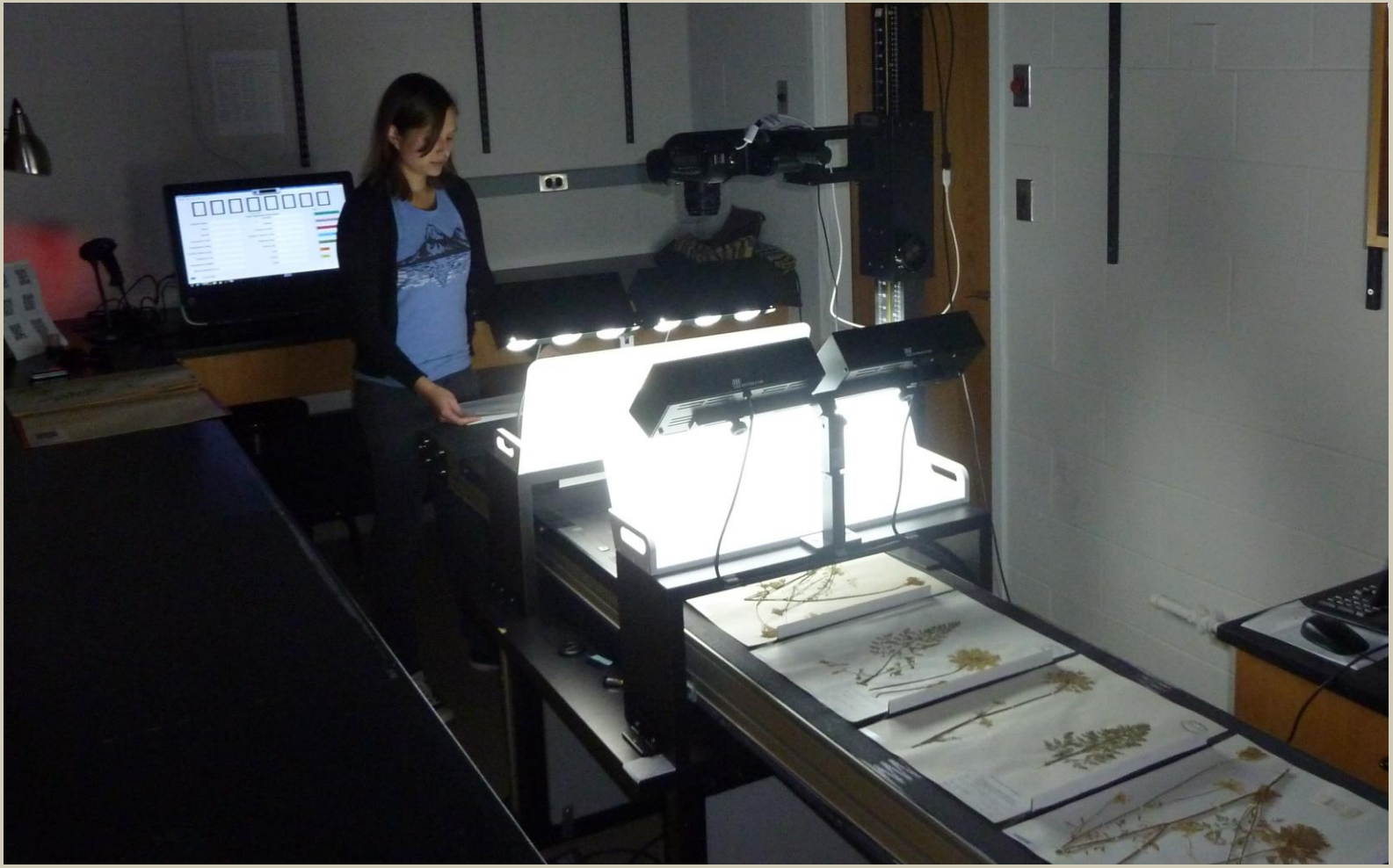
- To increase the efficiency of capturing an image and specimen-level data, we are developing a high throughput digitization apparatus



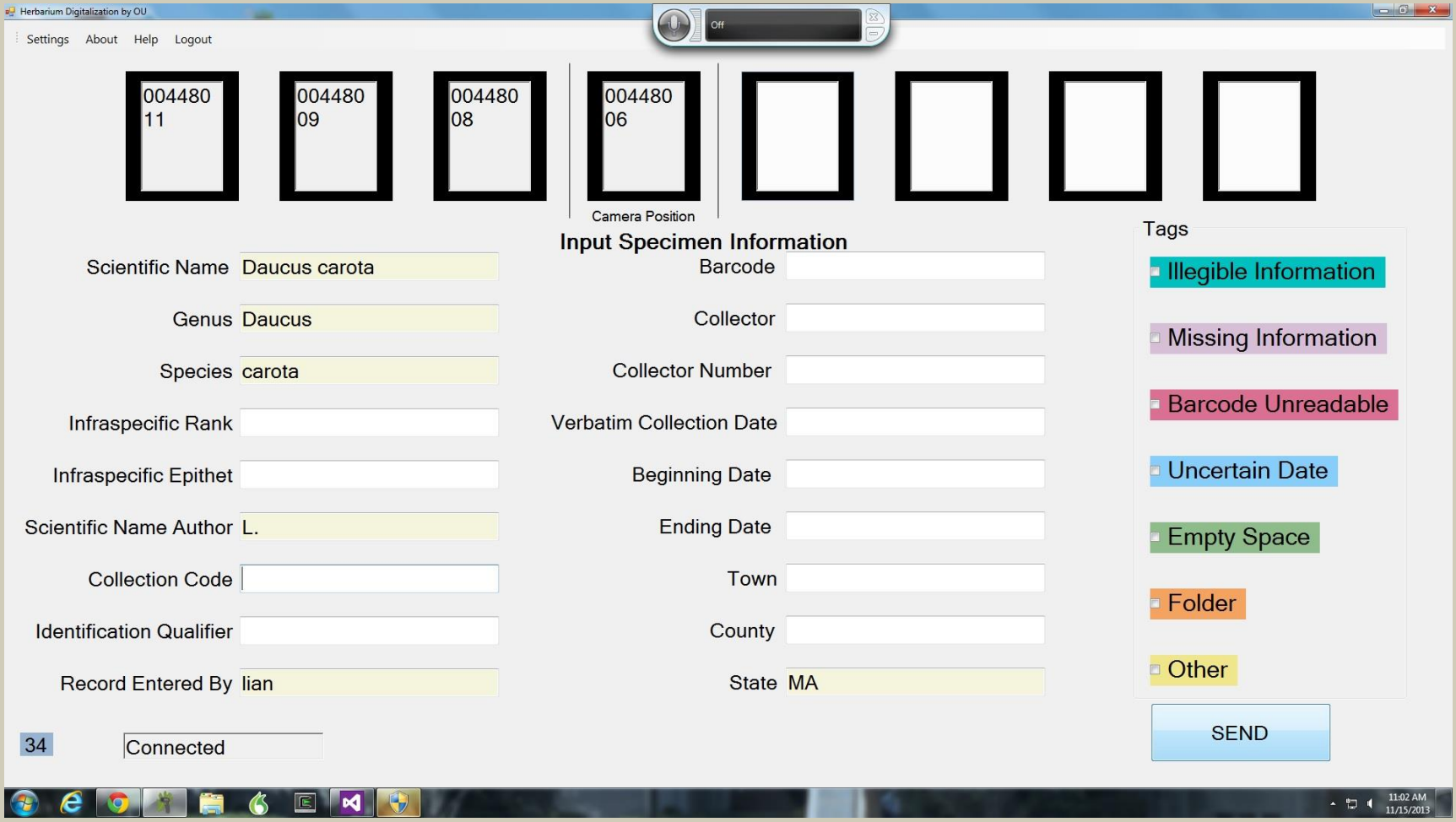
WORKFLOW: PRIMARY DIGITIZATION

- 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

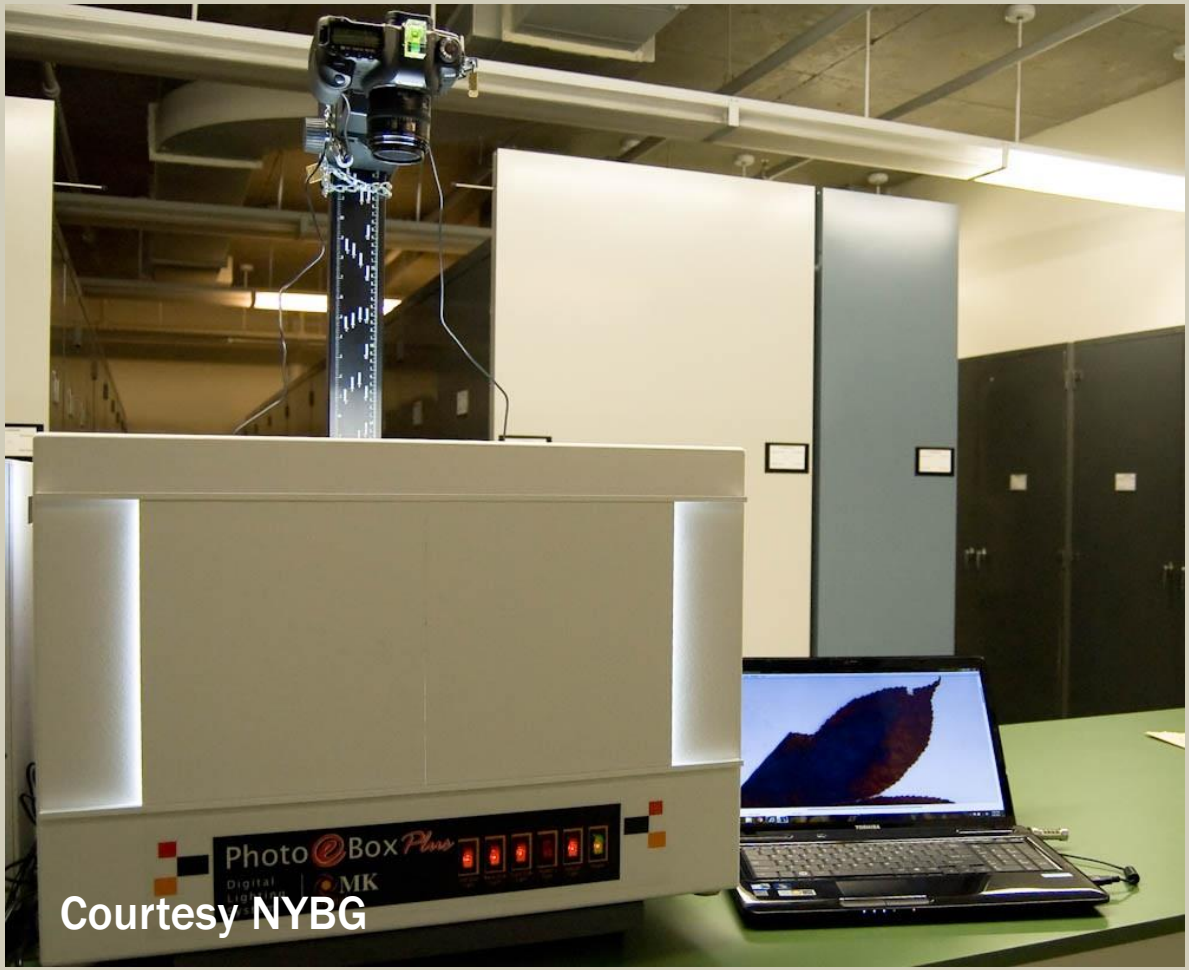
WORKFLOW: PRIMARY DIGITIZATION



WORKFLOW: PRIMARY DIGITIZATION

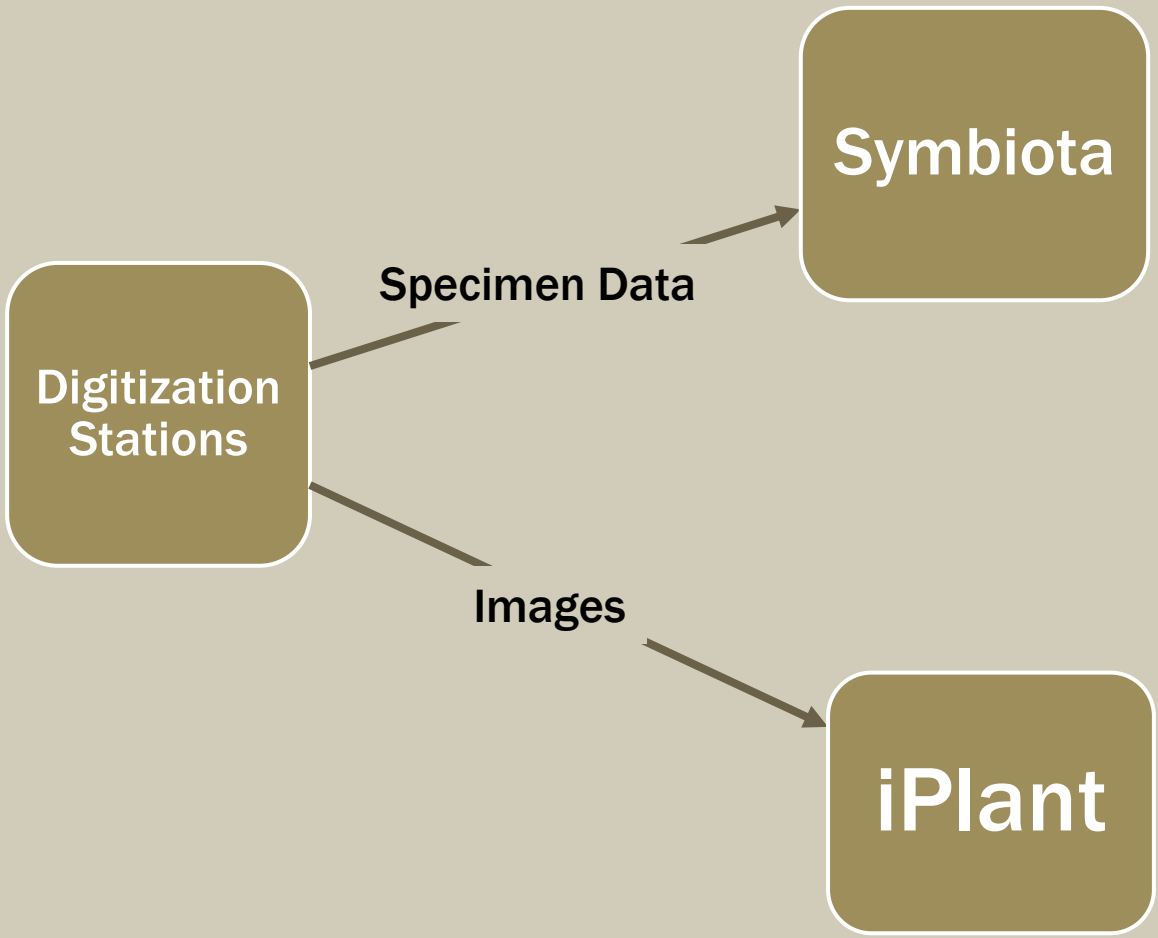


WORKFLOW: PRIMARY DIGITIZATION



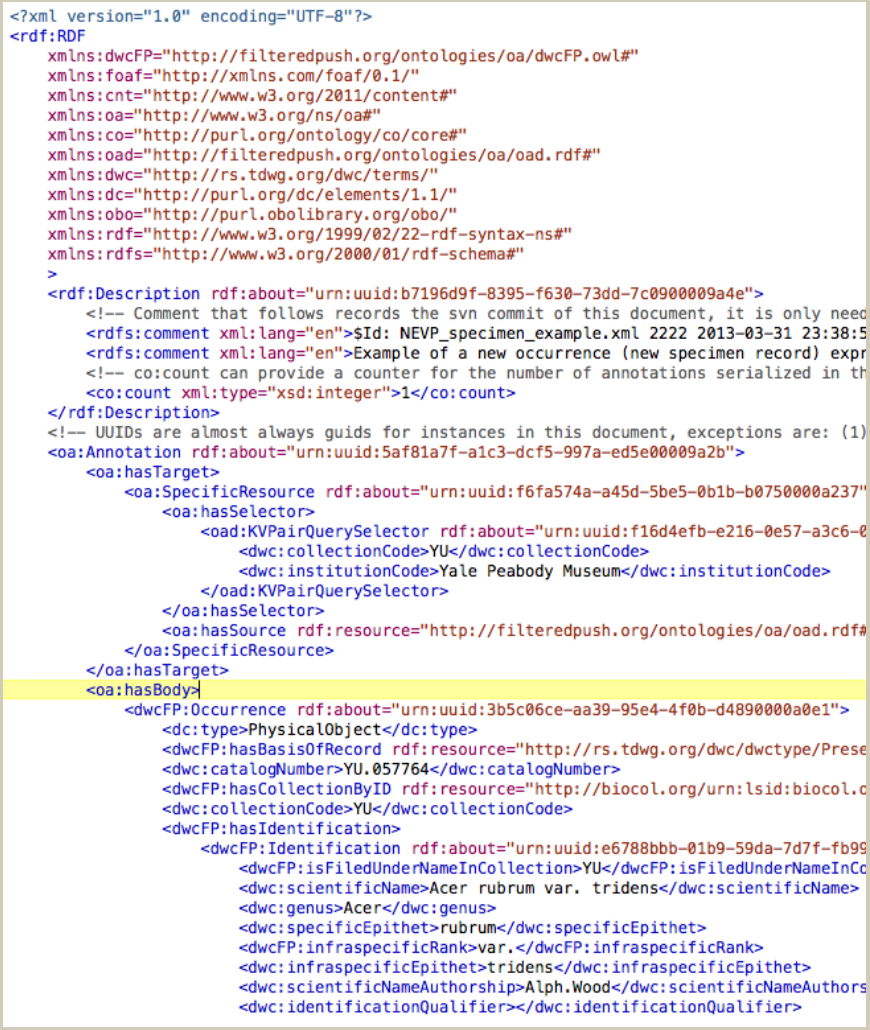
Courtesy NYBG

WORKFLOW: PRIMARY DIGITIZATION



WORKFLOW: PRIMARY DIGITIZATION

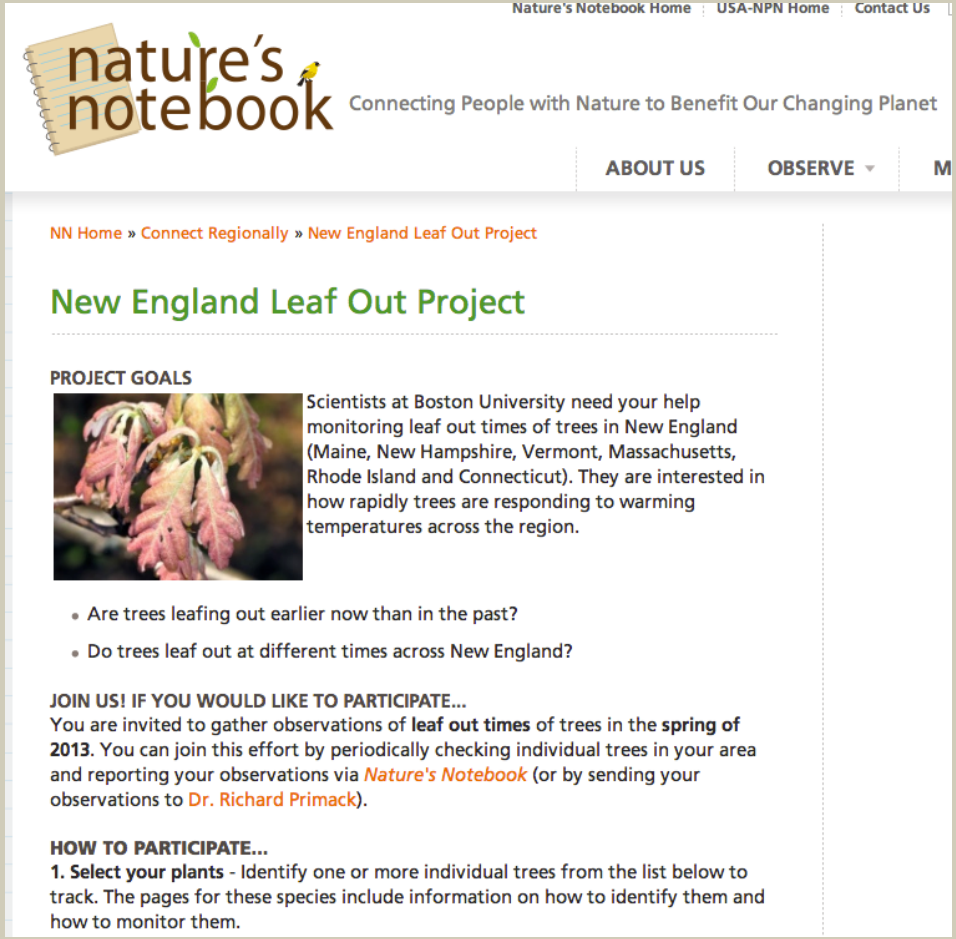
- interchange standard using RDF/XML, the W3C OpenAnnotationOntology, and DarwinCore was defined, and code was added to Symbiota to support the ingest of these documents into Symbiota.



```
<?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.obolibrary.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 neec
        <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"
                <oa:hasSelector>
                    <oad:KVPairQuerySelector rdf:about="urn:uuid:f16d4efb-e216-0e57-a3c6-0
                        <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#
            </oa:SpecificResource>
        </oa:hasTarget>
        <oa:hasBody>
            <dwcFP:Occurrence 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
                <dwc:catalogNumber>YU.057764</dwc:catalogNumber>
                <dwcFP:hasCollectionByID rdf:resource="http://biocol.org/urn:lsid:biocol.c
                <dwc:collectionCode>YU</dwc:collectionCode>
                <dwcFP:hasIdentification>
                    <dwcFP:Identification rdf:about="urn:uuid:e6788bbb-01b9-59da-7d7f-fb99
                        <dwcFP:isFiledUnderNameInCollection>YU</dwcFP:isFiledUnderNameInCo
                        <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:identificationQualifier></dwc:identificationQualifier>
```

TRAINING & OUTREACH

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



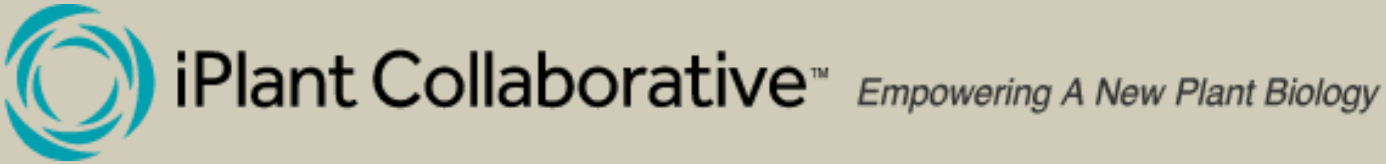
www.usanpn.org/nn/nelop

ACKNOWLEDGEMENTS

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

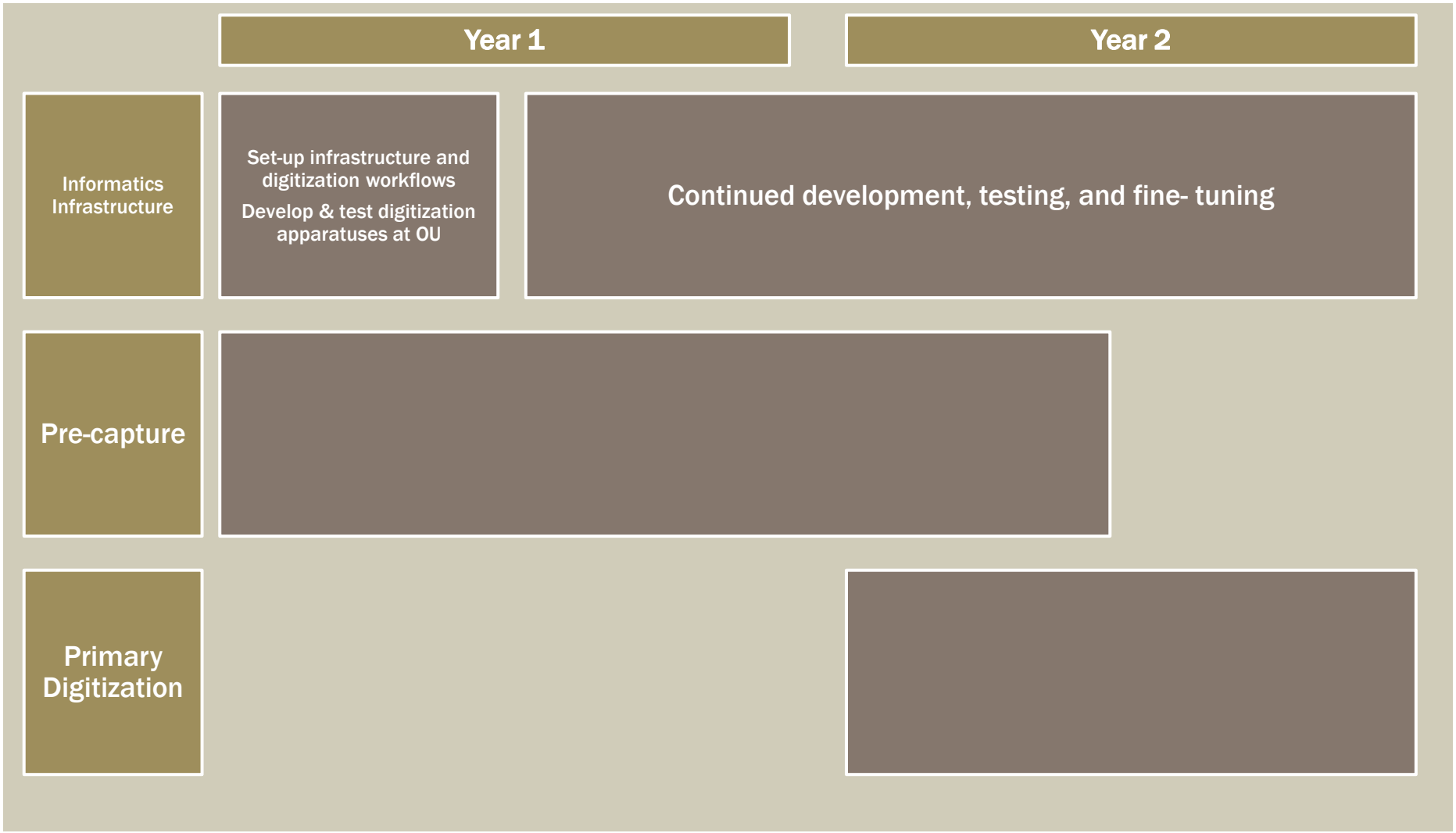
Symbiota Project

FilteredPush



Biota of North America

TIMELINE


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

WORKFLOW: ENHANCEMENT