

12th International Digital Curation Conference - 20<sup>th</sup> Feb 2017 - Edinburgh University  
Workshop on How EUDAT Services could support FAIR data.

# Two part presentation:

1. The EUDAT – Herbadrop project.

Lifted slide by slide from Pascal Dugénie, Herbadrop Project Officer - C.I.N.E.S (Centre Informatique National de l'Enseignement Supérieur)

2. The reality of being a data partner with HERBADROP

Robert Cubey – Plant Records Officer – RBGE (Royal Botanic Garden Edinburgh)



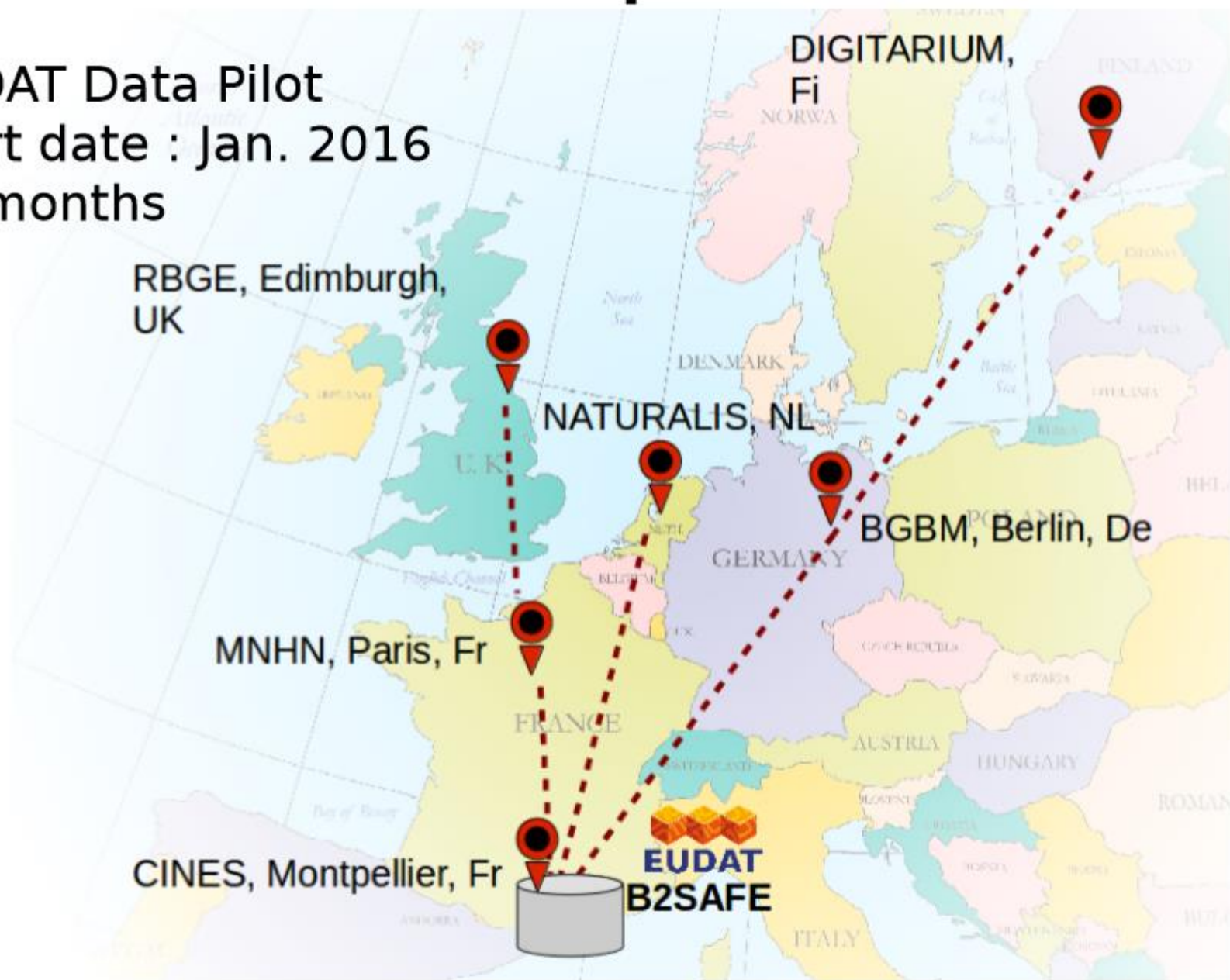
HERBADROP

EUDAT receives funding from the European Union's Horizon 2020 programme –  
DG CONNECT e-Infrastructures. Contract No. 654065



# The Herbadrop Data Pilot

- EUDAT Data Pilot
- Start date : Jan. 2016
- 18 months



# Herbarium Images

- Millions of specimens in herbaria all over the world
- Global trend to industrial digitizing
- Data difficult to handle even for medium size institutes
- Same challenges being faced by potentially hundreds of herbaria in Europe
- Makes sense to work together to develop a solution





## **Herbadrop challenges**

- Specimen images available online
  - cross institute & domains data analyses
- Large datasets of herbarium specimens
  - 500 million worldwide
  - 22 million in Germany
  - 20 million in France
- Extracting information from the specimen labels :
  - enhanced methods using state-of-the-art OCR technologies
  - particularly complex due to the presence of biological material in the image
- Discovering new links between objects of different collections

HERRIER, LEUHAU - DE PIGNAL  
Projet Plantes - France 2

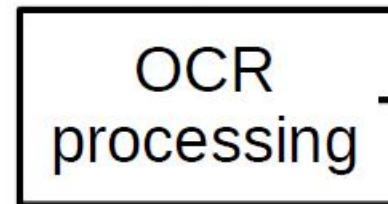
1965 11/01/65 20/01/65

GASTRODIUM

DINAGHAW MOUNTAIN PARAGUAY

1965 11/01/65 20/01/65

GASTRODIUM



BRESIL, Espin'lo Sanlu, Près  
ijemo do Mcio,  
Fazenda Michel Frey,

Forêt atlantique d'alliluda terrain humide

n°H319	31/07/1993	dupl. P, BAH
--------	------------	--------------

BRESIL, Espírito Santo, Près Riberão do Meio.  
Fazenda Michel Frey.  
20° 45' - 20° 30' S 41° 00' - 41° 15' W  
Forêt atlantique d'altitude, terrain humide.

†Values obtained were taken from the data of the first series. Mean  $\pm$  SD.

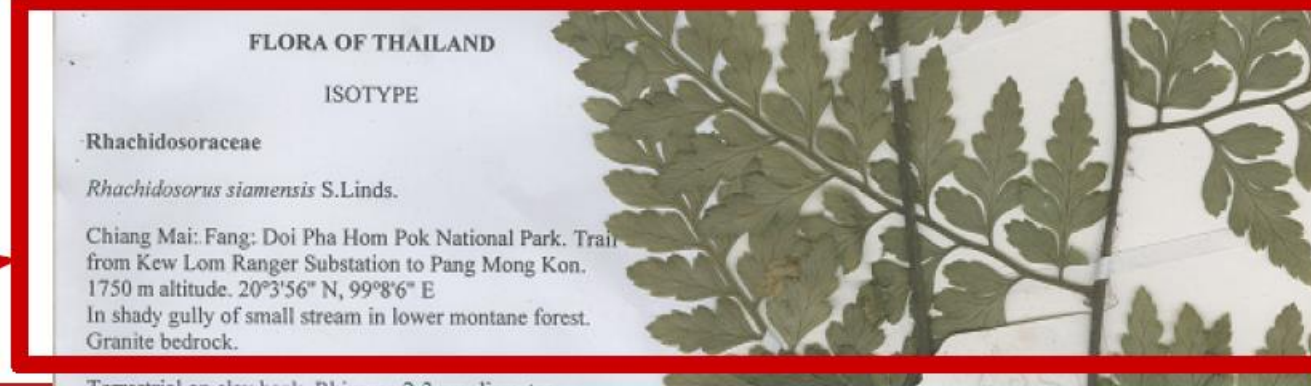
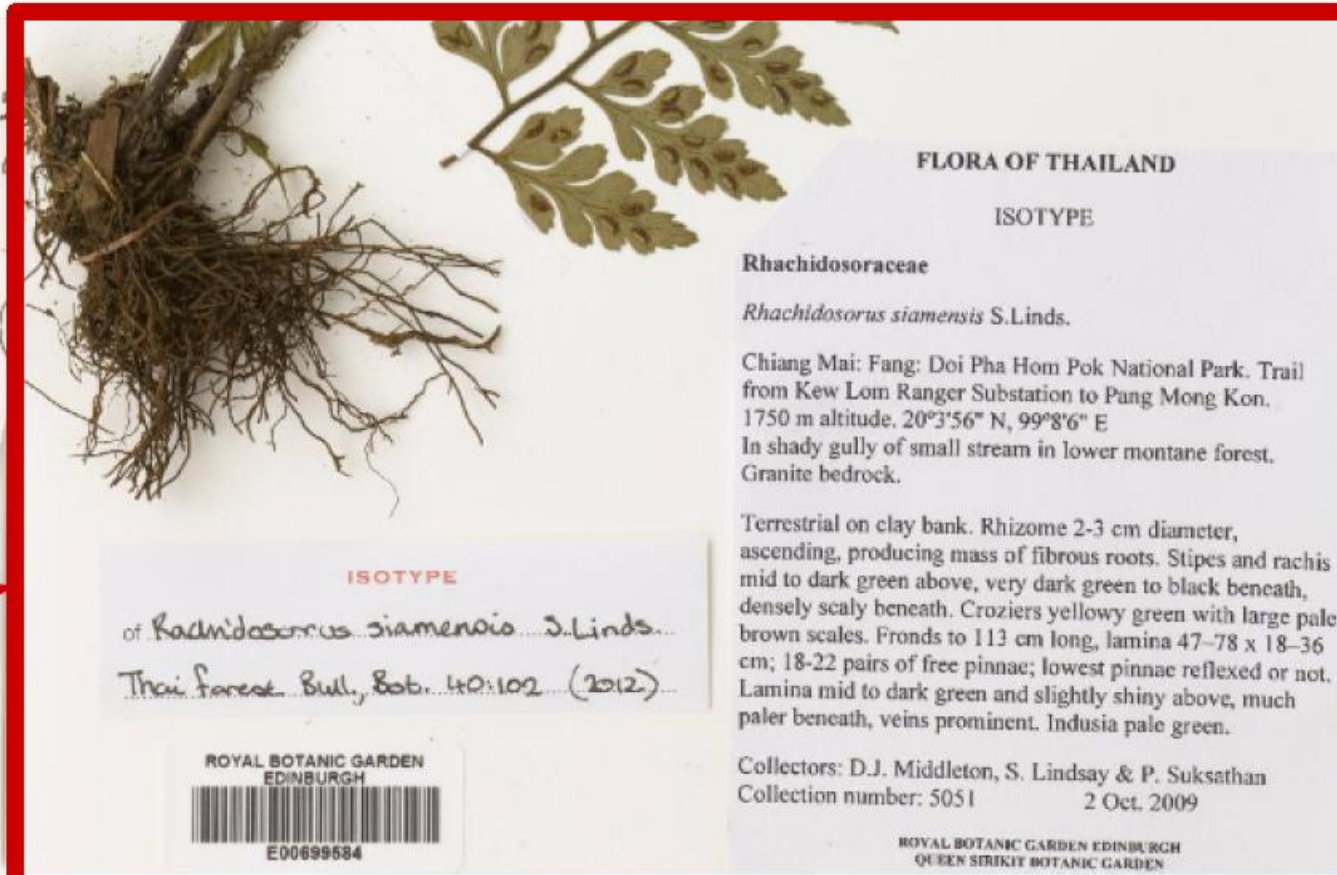


# Mining for duplicates

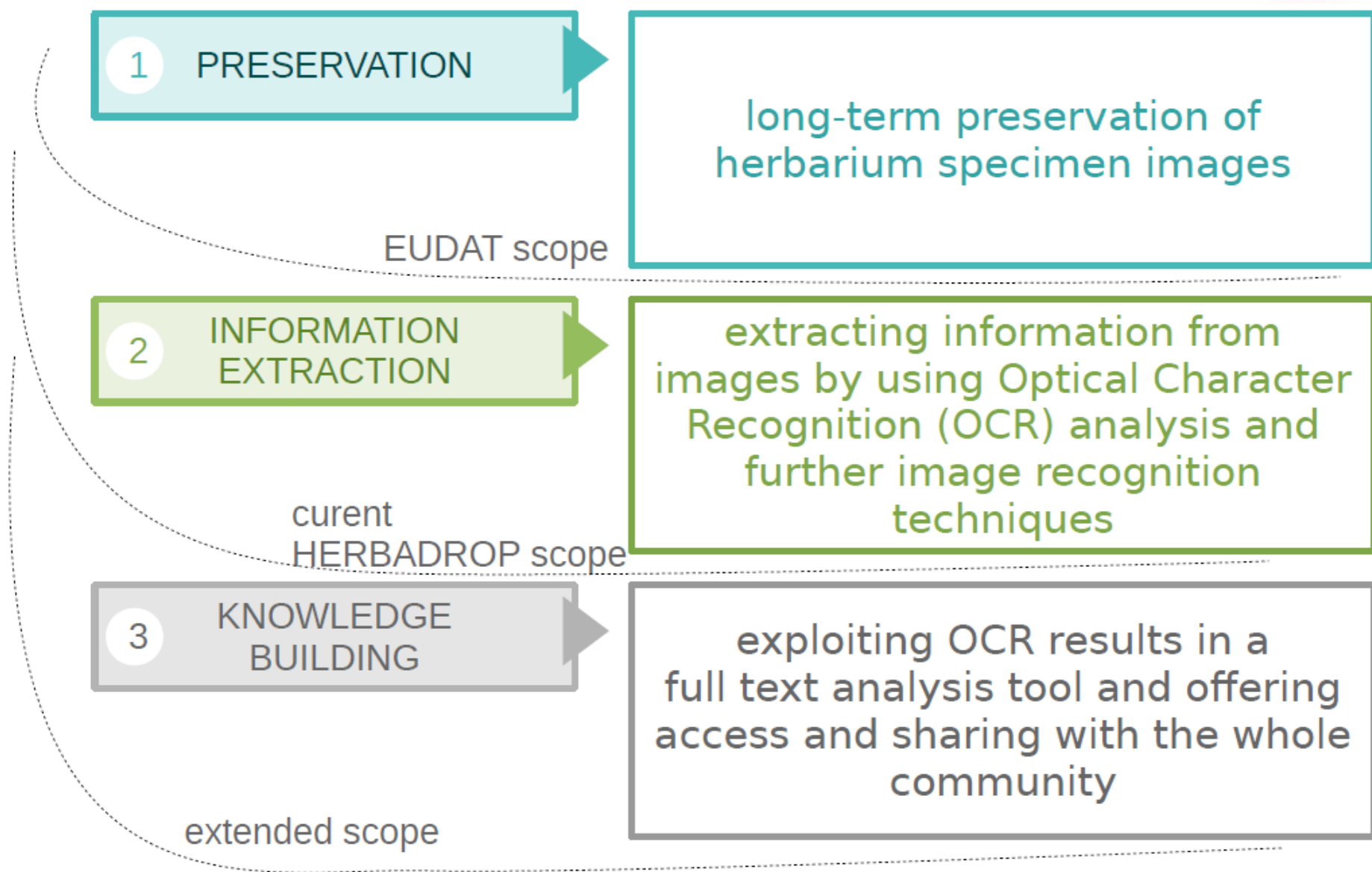
E00699584



P02142145

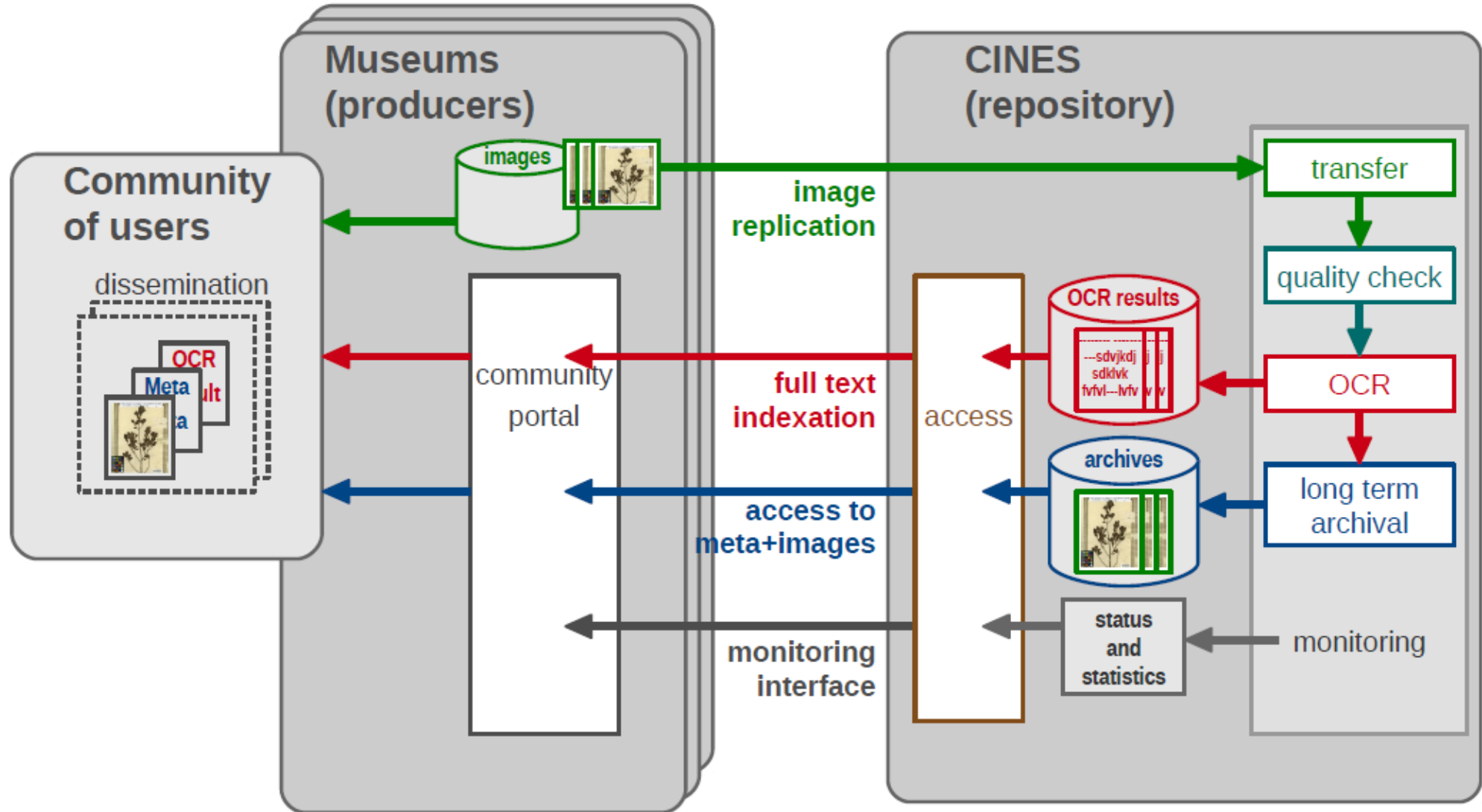


# Herbadrop objectives

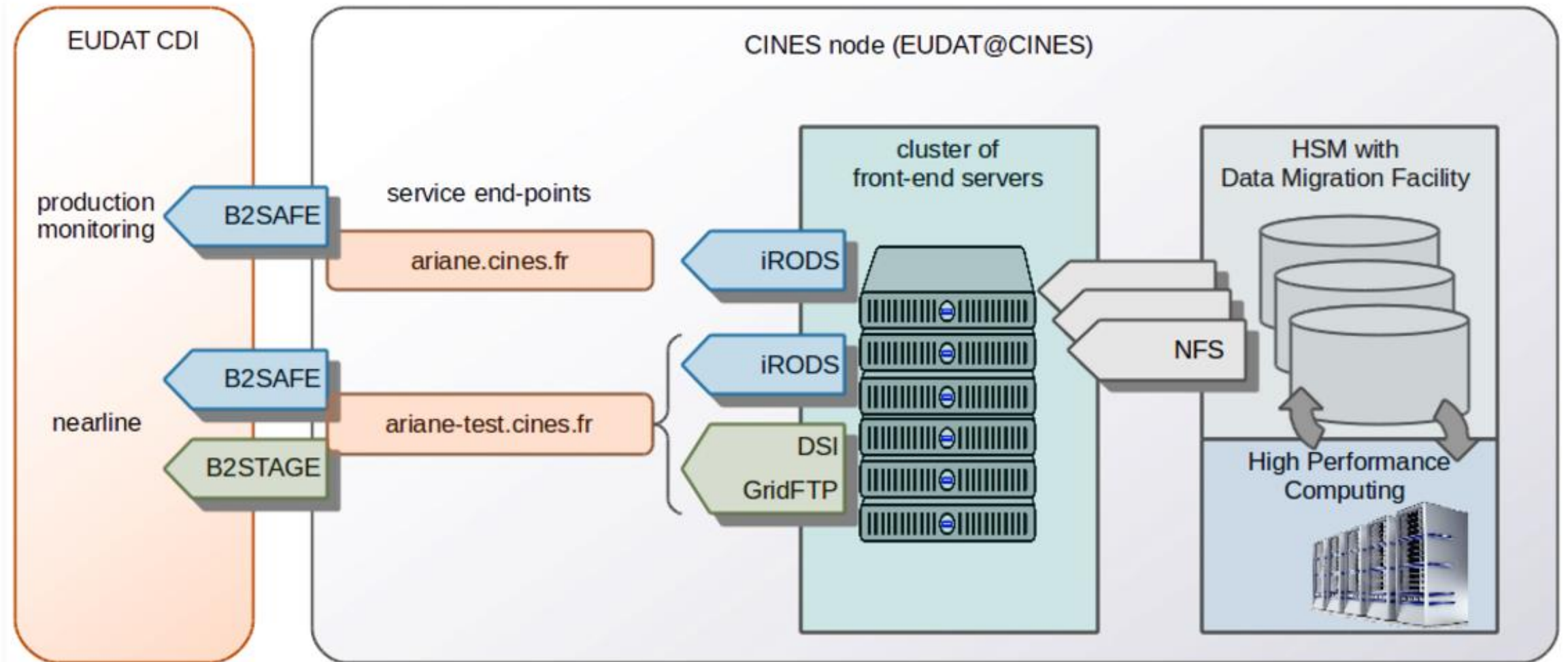




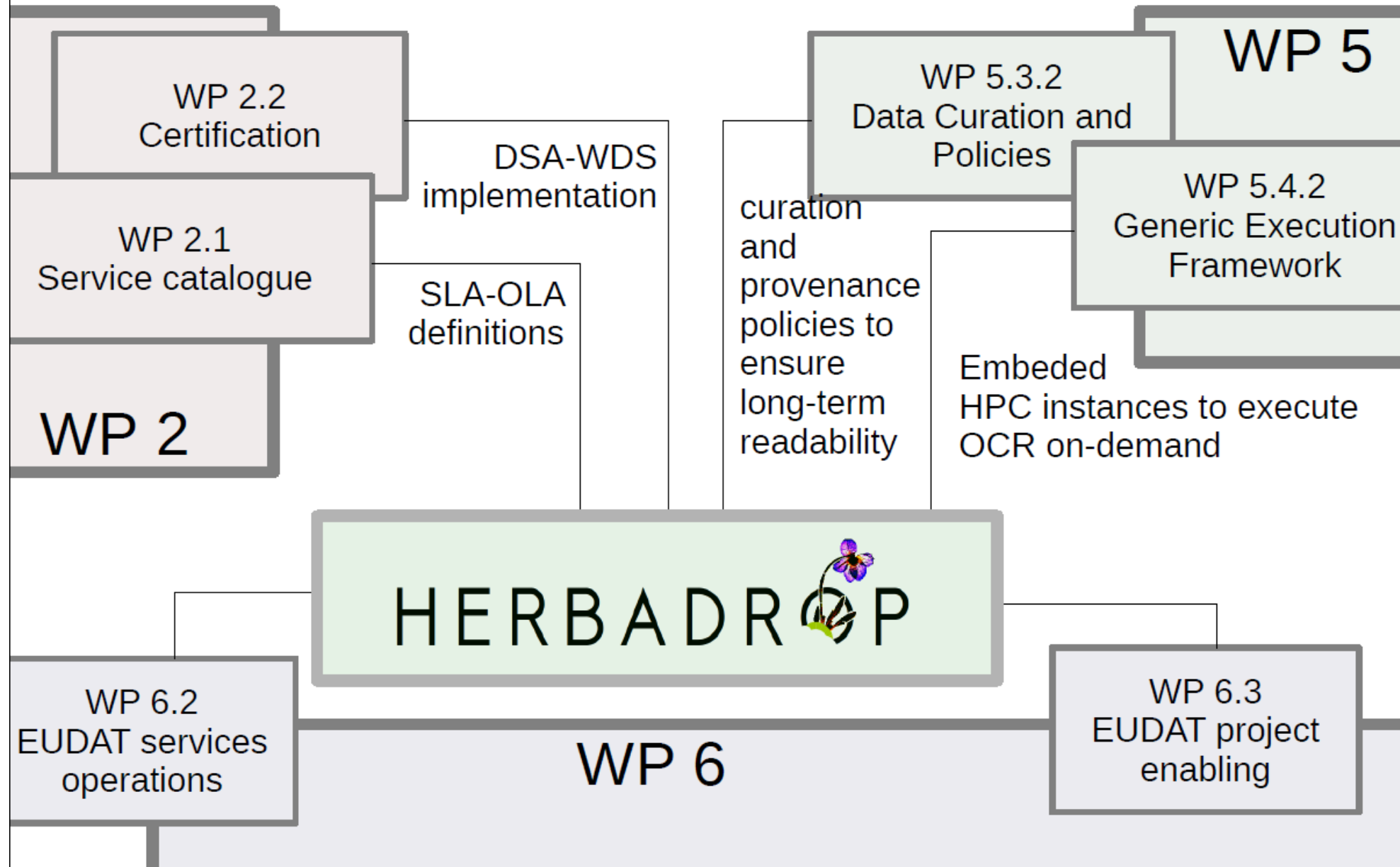
# Technical implementation



# Service hosting



# Collaborations with other EUDAT WP





## Current status

- Achievements
  - service operation and workflows between the B2SAFE submission and the process of OCR data
  - ingestion of more than 3 Millions of images representing 11 M OCR files and 1,78 M of images processed
  - more than 22 TB of volume and 200 000 hours of computation power
- What still needs to be done

## Current status

- Achievements
- What still needs to be done
  - finalize the archive function, harmonizing the Metadata model
  - defining a common interface for harvesting specimen catalog including common set of metadata attributes based on Darwin Core standard.
  - improving recognition of hand written text in herbarium images
  - data curation as part of the preservation process (identifying duplicates, or inducing new taxonomic relations, etc.).
  - developing an OAI-PMH endpoint to publish the content the database.

## Part two:

### 2. The reality of being a data partner with HERBADROP





At the RBGE – Plant collection: This is what we are very good at doing.





Plant identification: good at this.





Plant curation: again, good this.





Digitisation: We have learnt in the last 10 years to become good at this.



# Royal Botanic Garden Edinburgh Herbarium Catalogue

Hits matching Genus=**primula** Species=**scotica** Filing region=**1A (Britain and Ireland)** Images only=**Yes**

Order results by then by then by [Sort results](#)

Page 2 of 3 showing Go to page: 2

Hits 31 to 60 of 85

[New query](#) [Edit query](#)

Download selection options:

You have 5 items selected for download

[Download](#) [Select all on page](#) [Deselect all on page](#) [Add all results](#) [Empty basket](#)

<p>Family: PRIMULACEAE Collection number: s.n.</p> <p><b><i>Primula scotica</i> Hook.</b></p> <p>Collection date: July 1832</p> <p>Filing region: Britain and Ireland</p> <p>Country of origin: United Kingdom</p> <p>Barcode: E00052924</p> <p>Cite as: <a href="http://data.rbge.org.uk/herb/E00052924">http://data.rbge.org.uk/herb/E00052924</a></p> <p>Label data partially captured</p>	<p>Select this record for download <input type="checkbox"/></p> <p><a href="#">Name details &gt;&gt;</a></p> <p><a href="#">Full specimen details &gt;&gt;</a></p>	
<p>Family: PRIMULACEAE Collection number: s.n.</p> <p><b><i>Primula scotica</i> Hook.</b></p> <p>Collection date: June 1887</p> <p>Filing region: Britain and Ireland</p> <p>Country of origin: United Kingdom</p> <p>Barcode: E00052984</p> <p>Cite as: <a href="http://data.rbge.org.uk/herb/E00052984">http://data.rbge.org.uk/herb/E00052984</a></p> <p>Label data partially captured</p>	<p>Select this record for download <input type="checkbox"/></p> <p><a href="#">Name details &gt;&gt;</a></p> <p><a href="#">Full specimen details &gt;&gt;</a></p>	
<p>Family: PRIMULACEAE</p> <p><b><i>Primula scotica</i> Hook.</b></p> <p>Collector/Expedition: J.C. Crawford</p> <p>Collector: J.C. Crawford</p> <p>Collection number: s.n.</p> <p>Collection date: 16 August 1907</p> <p>Filing region: Britain and Ireland</p> <p>Country of origin: Not recorded</p> <p>Barcode: E00019587</p> <p>Cite as: <a href="http://data.rbge.org.uk/herb/E00019587">http://data.rbge.org.uk/herb/E00019587</a></p> <p>Label data partially captured</p>	<p>Select this record for download <input type="checkbox"/></p> <p><a href="#">Name details &gt;&gt;</a></p> <p><a href="#">Full specimen details &gt;&gt;</a></p>	



<http://data.rbge.org.uk/herb>

Online dissemination: An ongoing process, but we have become good at doing this.





[Back to results](#)



To download a TIFF version of this image, select the image resolution you require and then click the Get Tiff button

6117 x 10034 ☒ 3058 x 5017 ☐ 1529 x 2508 ☐ 764 x 1254 ☐ 382 x 627 ☐ 191 x 313 ☐ 95 x 156 ☐ [Get Tiff](#)

[Back to Top](#)

Image size, quality and usability: With 600 dpi TIFF and 7 options of image size we are now good at this.



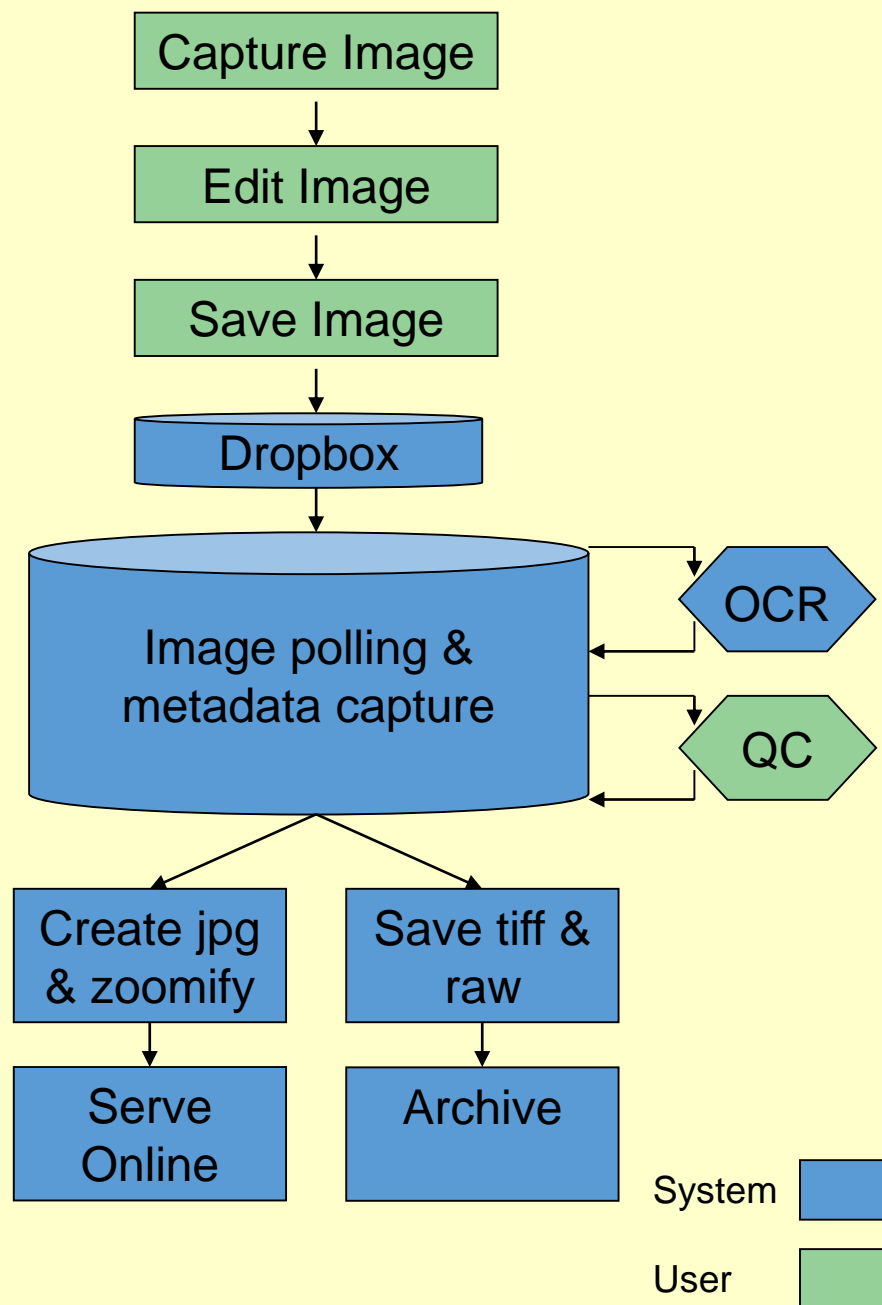


Image workflow developed at RBGE incorporating:

- 📷 image capture
- 📷 automated image processing
- 📷 metadata recording
- 📷 optical character recognition
- 📷 quality control
- 📷 image streaming online
- 📷 archiving

Image workflows: We had to get good at this, it has taken while but we are there now.

Internal collection  
management  
databases



# Herbarium Data



Internally  
produced  
resources

## RBGE project websites

ADIAIC Diatom Image Database  
*Apiales* Resource Centre  
*Dipterocarpaceae* Database  
Robert Wight Catalogue  
*Sapotaceae* Resource Centre  
SE Asian *Begonia* Database  
*Zingiberaceae* Resource Centre

Flora of Nepal  
Webcyta Genetic Flora  
Dryflora  
Cerrado  
*Gesneriaceae* Resource Centre  
CMEP Plant & Birds Apps.  
Plants & You (Nepal)  
Botanics Stories

## RBGE online catalogues

Searchable &  
downloadable data  
RBGE online catalogues

External  
aggregators

 GeneBank	 BOLD	 BioCase	 EOL	 Global Plants	 iDigBio	 GBIF	 Europeana	 CRIS Species link	 NBN	 ALS
Collection and management of specimen, distributional, and molecular data as well as analytical tools to support their validation.	A genetic sequence database, an annotated collection of all publicly available DNA sequences	An international network of European biological collections. Using open-source, system-independent software and open data standards and protocols.	Online collaborative encyclopaedia to document all of the 1.9 million species of life on earth.	The world's largest database of digitized plant specimens & a locus for international scientific research and collaboration	New Tools for Increasing the Use of Natural History Collections for Research, Education and Informed Decisionmaking	International Biodiversity data on distribution plants, animals, fungi, microbes and scientific names data	Multi-lingual online collection of millions of digitized items from European museums, libraries, archives and multi-media collections.	The goal of the speciesLink network is to integrate species and specimen data available in natural history museums.	Information network of UK biodiversity data from nature agencies, wildlife trusts & museums	Provides a platform for the collection, aggregation, analysis and use of biological data in Scotland.

External  
end  
users

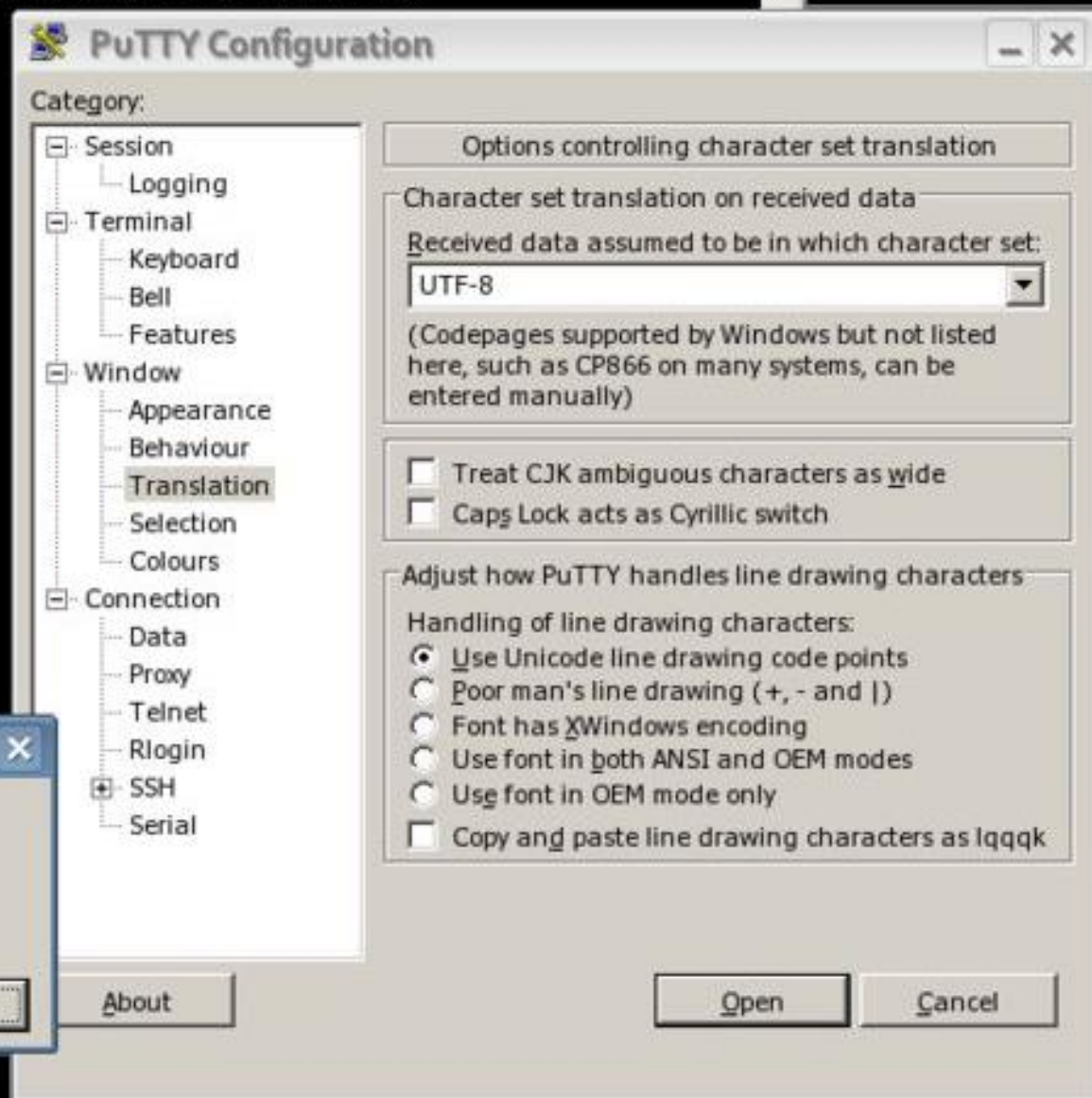
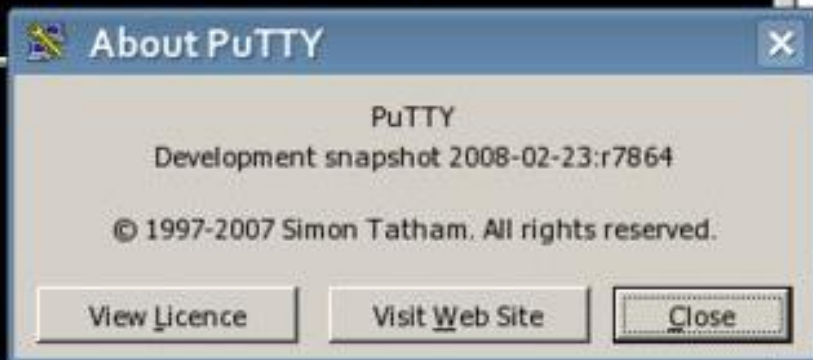


## Biodiversity scientists

Taxonomists, habitat restoration, species checklists, geneticists etc.

Providing data aggregators: Once again had to get good at this, if we wanted people to use our data.

```
Using username "xbarin02".
Using keyboard-interactive authentication.
Password:
Last login: Sat Feb 23 04:39:56 2008 from c03-414a.kn.vutbr.cz
xbarin02@c03-414a ~ $ uname
Linux
xbarin02@c03-414a ~ $ locale
LANG=cs_CZ.UTF-8
LC_CTYPE="cs_CZ.UTF-8"
LC_NUMERIC="cs_CZ.UTF-8"
LC_TIME="cs_CZ.UTF-8"
LC_COLLATE="cs_CZ.UTF-8"
LC_MONETARY="cs_CZ.UTF-8"
LC_MESSAGES="cs_CZ.UTF-8"
LC_PAPER="cs_CZ.UTF-8"
LC_NAME="cs_CZ.UTF-8"
LC_ADDRESS="cs_CZ.UTF-8"
LC_TELEPHONE="cs_CZ.UTF-8"
LC_MEASUREMENT="cs_CZ.UTF-8"
LC_IDENTIFICATION="cs_CZ.UTF-8"
LC_ALL=
xbarin02@c03-414a ~ $
```



This is where we are less strong – coding and command line interaction.





Preserving this – a high resolution image of a herbarium sheet is great, but we want more.



## FLORA OF BHUTAN.

Locality Phayudin Zimpu  
Date 4/8/14  
Altitude 12000'  
Remarks No. 2491

*Potentilla fruticosa* Linn.  
var.

R. E. COOPER, Collector for  
*det. H.B.G.* A. K. BULLEY, Esq., CHESHIRE.

Georeference the collection site, database the collector and collection date; as here is physical proof that this material was at a site on a particular date.

*P. fruticosa* var. *arbuscula*

*Daphnora unifoliolata*  
(Ludlow) Sojak  
rev. Jiri Sojak 19...  
2007

*Potentilla davurica* Nestl.  
var. *arbuscula* /D.Don/  
  
Revised by H. Rhodes 1956  
Department of Agriculture, Ottawa.  
SB 253  
12812-PSM-249

*Potentilla arbuscula* D. Don  
  
Det. A. J. C. Grierson & D. G. Long  
Bhutan Flora Project May 1978

Dets: (determinations) taxonomic opinions on the sheet, all with different levels of validity – OCR is never going to the complete answer for our material.



As a community we use unique identifiers for our material, we need these identifiers to connect material.

GBIF Global Biodiversity Information Facility

Login or Create a new account

Data - News - Community - About -

## Search occurrences

cooper+2491+bhutan

Use the filters to customize search results

2 Occurrences

Downloads using free text search are not supported

2 results

[\[View results as table\]](#)

The map view is currently experimental.

▼ Add a filter

COLLECTOR: cooper

BASIS OF RECORD: Specimen

COUNTRY: Bhutan

Map of Bhutan showing the location of the specimen. A red square indicates the location.

Refine your search

reset

BASIS OF RECORD

☒ Specimen (2)

DATASET

☐ Herbaria of the University... (1)

☐ Natural History Museum (Lo... (1)

COUNTRY

☒ Bhutan (2)

LICENSE

☐ CC0 1.0 (1)

☐ CC-BY-NC 4.0 (1)

MONTH

☐ October (1)

YEAR

☐ 1914 (1)

# Search occurrences

Potentilla arbuscula |



Use the filters to customize search results

35

Occurrences

Download

35 results

[\[View results as map\]](#)

[Configure](#) [Add a filter](#)

SCIENTIFIC NAME: **Potentilla arbuscula D. Don**

	LOCATION	BASE OF RECORD	DATE
1257803998 - Cat. 100457135 <b>Potentilla arbuscula D. Don</b> Published in Tropicos Specimen Data	Bhutan 27.80N, 89.36E Elevation: 4,922m	Specimen	8 / 2010
1257803738 - Cat. 100456882 <b>Potentilla arbuscula D. Don</b> Published in Tropicos Specimen Data	Bhutan 27.79N, 89.34E Elevation: 4,141m	Specimen	8 / 2010
727330575 - Cat. barcode-00195533 <b>Potentilla arbuscula D. Don</b> Published in Harvard University Herbaria	China 28.66N, 98.66E	Specimen	8 / 2001
1258228008 - Cat. 11594973 <b>Potentilla arbuscula D. Don</b> Published in Tropicos Specimen Data	China 28.65N, 98.66E Elevation: 4,400m	Specimen	8 / 2001
111311626 - Cat. 53365-1 <b>Potentilla arbuscula D. Don</b> Published in Aranzadi Zientziak Elkartea	Spain 43.26N, 2.16W Elevation: 200m	Specimen	6 / 1993
574660437 - Cat. E00073430 <b>Potentilla arbuscula D. Don</b> Published in Royal Botanic Garden Edinburgh Herbarium (E)	China N/A Elevation: 2,900m	Specimen	6 / 1992
1339167673 - Cat. WAG.1877668 <b>Potentilla arbuscula D. Don</b> Published in Naturalis Biodiversity Center (NL) - Botany Leiden	unknown or invalid N/A	Specimen	7 / 1981

## Refine your search

reset

BASE OF RECORD

- ☐ Specimen (31)
- ☐ Living Specimen (4)

TYPE STATUS

- ☐ Holotype (1)
- ☐ Type (1)

DATABASE

- ☐ Naturalis Biodiversity Center (8)
- ☐ Plant Specimen from Herbar... (6)
- ☐ The vascular plants collec... (4)
- ☐ Royal Botanic Garden Edinb... (3)
- ☐ Tropicos Specimen Data (3)
- ☐ Botany (USP) (3)
- ☐ Royal Botanic Gardens, Kew... (2)
- ☐ Natural History Museum (L... (1)
- ☐ Aranzadi Zientziak Elkartea (1)
- ☐ Harvard University Herbaria (1)



COUNTRY

- ☐ China (10)
- ☐ Nepal (10)
- ☐ India (3)
- ☐ Unknown or Invalid (3)
- ☐ Bhutan (2)
- ☐ Spain (1)

LICENSE

- ☐ CC-BY 4.0 (22)
- ☐ CC0 1.0 (8)

# Botanics Stories

RBGE PERSONAL & PROJECT STORIES

[Home](#) [Science](#) [Gardens](#) [Communities](#) [Exhibitions](#) [WW1](#) [Newsletter](#)

Jul  
12  
2013

## Roland Edgar Cooper (1890-1967)

[Other News, Science](#)



*Roland Edgar Cooper. RBGE Library Archive*

Roland Edgar Cooper was born in 1890 and orphaned at an early age. Once he turned sixteen he came under the guardianship of his aunt Emma Smith, his mother's half-sister, and her husband, the botanist and future Regius Keeper at RBGE, William Wright Smith.

In 1907 Cooper went to India with his uncle when Smith became Keeper of the Herbarium at the Royal Botanic Garden Calcutta. While in Calcutta, and then later in Lloyd Botanic Garden Darjeeling, Cooper studied botany and horticulture under the tutelage of George Thomas Lane, the then Curator of the Royal Botanic Garden Calcutta. While in India he accompanied his uncle on collecting trips to Sikkim and the borders of Nepal, Bhutan and Tibet, gaining early experience of the diversity of the Himalaya.

Cooper returned to Scotland in 1910 when Smith was offered a post at RBGE to describe the extensive collections then being received from George Forrest from Yunnan. Cooper's botanical and horticultural

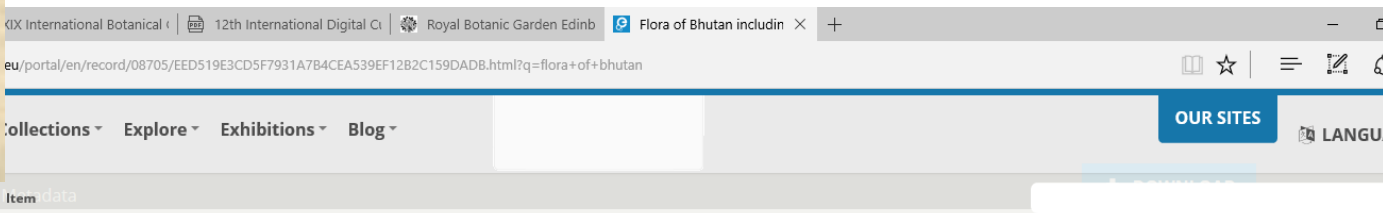
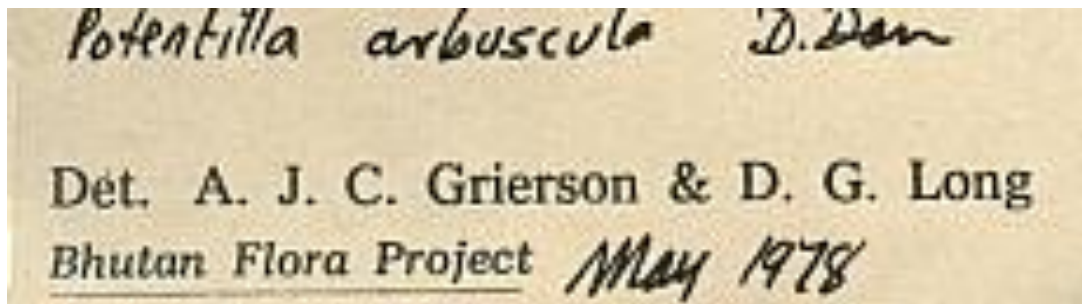
education continued when he signed up to be tutored at the Garden by Isaac Bayley Balfour with the intention of becoming a fruit farmer in British Columbia.

Before finishing his studies he took up a post leading a collecting expedition to Sikkim sponsored by A.K. Bulley of Ness Gardens, with encouragement and support from Smith and Balfour. From his early expeditions to Sikkim, and his studies at RBGE, Cooper was well qualified for this task. The reference that Balfour wrote for Cooper to Bulley sang his praises, citing that he had a good grasp of local Northern Indian dialects and an aptitude for learning languages.

## Collecting Trips

We need to have stable identifiers to link out to other sources





**Title** Flora of Bhutan including a record of plants from Sikkim and Darjeeling: Volume 2, Part 1 (part number)

**People** Contributor:  
Grierson, Andrew John Charles 1929-1990 , Long, David Geoffrey , Bates, M. (Illus) , Rodrigues, G. (Illus)

**Classifications** Type: Text  
Subject: FH\qrh , Floras

**Physical Description** Size: 1-426p. ill., figs., map, bibl., indices, PB, 20cm

**Date:** [http://semium.org/time/19xx\\_2\\_half](http://semium.org/time/19xx_2_half)  
**Creation Date:** 1991 , <http://semium.org/time/1991>  
**Period:** 1991  
**First Issued:** 1991 , <http://semium.org/time/1991>  
**Temporal:** <http://semium.org/time/19xx>

**Publisher:** Royal Botanic Garden, Edinburgh (Edinburgh)  
**Identifier:** 9h30r3h7-e6r1-4fra-82e4-66h3e62024fh 1872291023 51159

#### FIND OUT MORE

View at Royal Botanic Garden Edinburgh:  
[United Kingdom](#)

#### CAN I USE IT?

No Re-use  
[In Copyright](#)

#### SHARE



## East Himalaya Gazetteer of plant collecting localities

### Search results from the east himalaya gazetteer

page 1 of 1 showing

Hits 1 to 1 of 1 matching **Phadjuding**

1 **Phadjuding** in Thimphu District , Long. 089 34 E , Lat. 27 29 N



No more hits

We need to have stable identifiers to link out to other sources



# Error

Exception: "0" Status: Request failed: request timed out in *DrupalApacheSolrService->checkResponse()* (line 360 of */opt/portal/drupal/www/sites/all/modules/additional/apachesolr/Drupal\_Apache\_Solr\_Service.php*).

The website encountered an unexpected error. Please try again later.

We need these other sources to be funded and stable

# END

Robert Cubey  
Plant Records Officer  
Royal Botanic Garden Edinburgh

[r.cubey@rbge.org.uk](mailto:r.cubey@rbge.org.uk)

@RBGE\_plant\_rec  
+44 (0) 131 248 2948