

HERBADROP

A pilot for storing and
analysing herbarium images

ISTC 2016 - Bratislava



MUSÉUM NATIONAL
D'HISTOIRE NATURELLE

Herbadrop - ISTC2016 - Bratislava



Royal
Botanic Garden
Edinburgh

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



tiff: 180MB

zip: 80MB

jpg: 1MB

Total: 161MB

Herbadrop - ISTC2016 - Bratislava

Storage

- Needs disk space
- Needs bandwidth
- Backup & recovery procedure
- File format checking



Image Analysis

- Useful information on labels
- The 1st step : OCR
- Tuning regognition
- Pattern matching



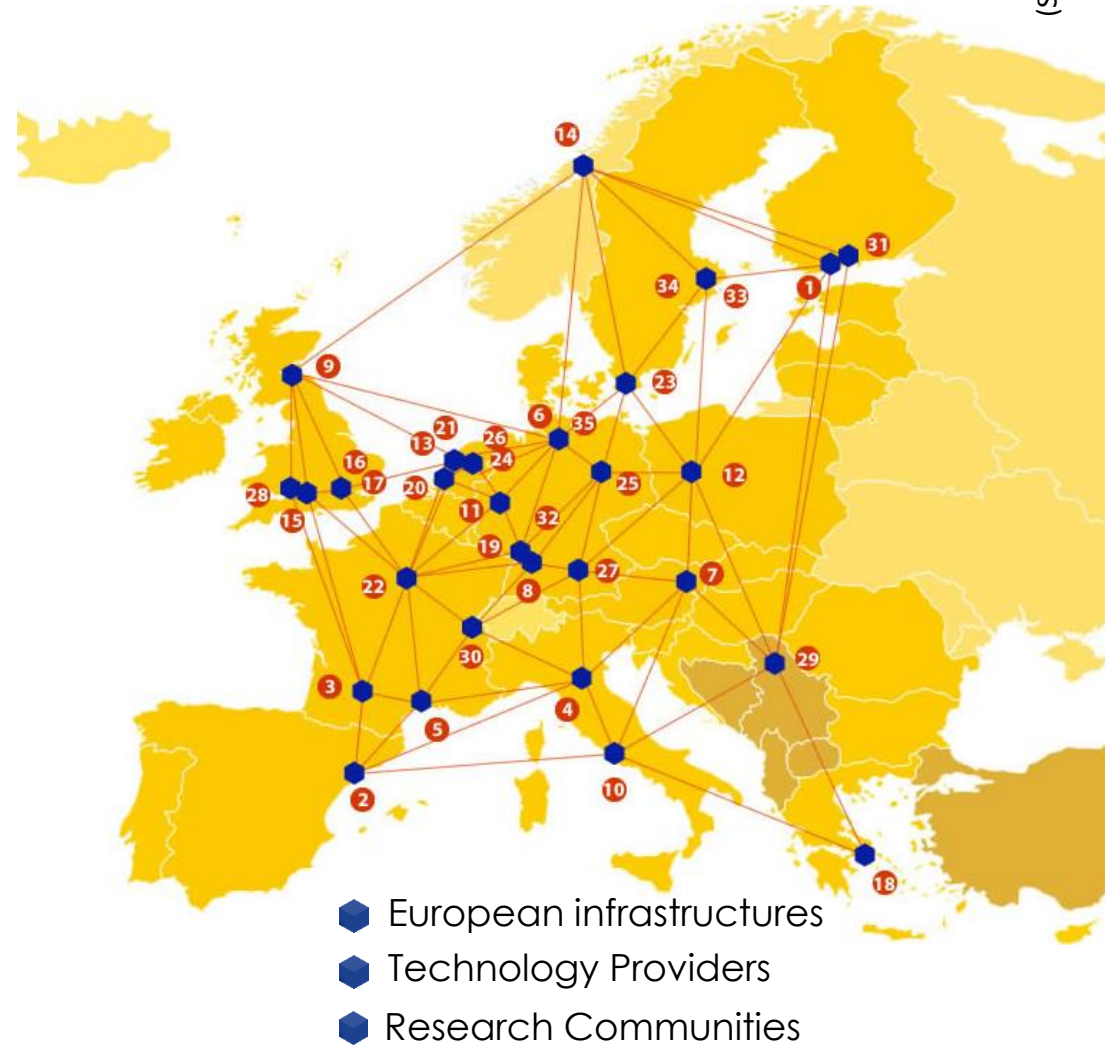
The CINES Data center



EUDAT: A pan-European Infrastructure

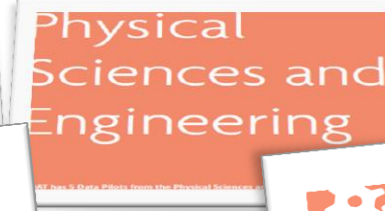
EUDAT offers **common data services** to **individuals, research infrastructures & communities and service providers** through a network of **35 European organisations**.

EUDAT enables **European researchers from any discipline and any geographic location** to **preserve, find, access, and process data** in a trusted environment.



EUDAT Data Pilot

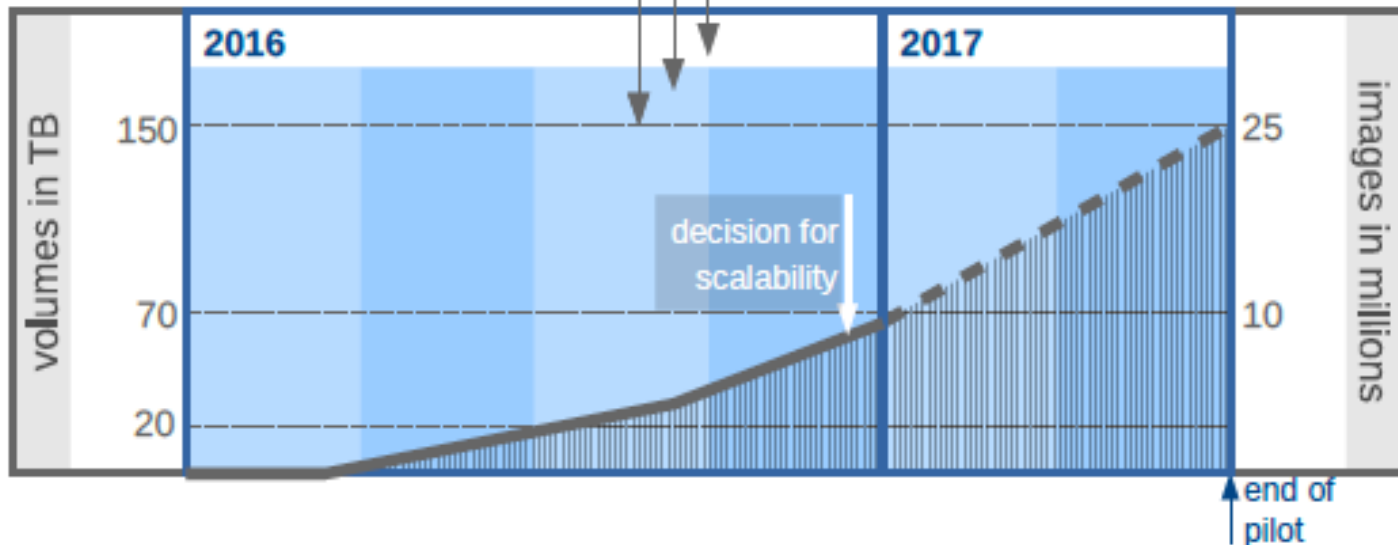
- Call for Data Pilots
 - Projects selected in December 2015
- Explain the EUDAT services area
- Encourage networking and cross-fertilisation of knowledge in smaller groups



A 18 Months pilot

	images	size/format	%	volumes
MNHN	20,00 M	5 Mo jpg	* 50	50,0 To
RBGE	0,30 M	150 Mo tiff	100	40,0 To
BGBM	0,01 M	250 Mo tiff	100	0,5 To
Naturalis	4,10 M	* 10 Mo	* 100	* 50,0 To
Digitalium	* 1,00 M	* 10 Mo	* 100	* 10,0 To

*hypothesis



The expected future impact



- Service extended beyond the five pilot institutes
- Specimens discoverable by the entire scientific community
- Solution will be of interest to similar collections in other parts of the world as well as by other collections such as the wider museum community and libraries

Thank you

