a "place" to store, search and share data from research on plant-microorganism-virus interactions.

Marta Vallino, *Piero Caciagli, Simona Abbà, Giancarlo Birello, Anna Perin, Stefano Ghignone*

WHY

WHO

WHAT

HOW

The problem…

Managing large amount of data is a challenge for many research institutes.

Data can be non-homogeneous (pictures, books, register, multimedia…) and not digitized.
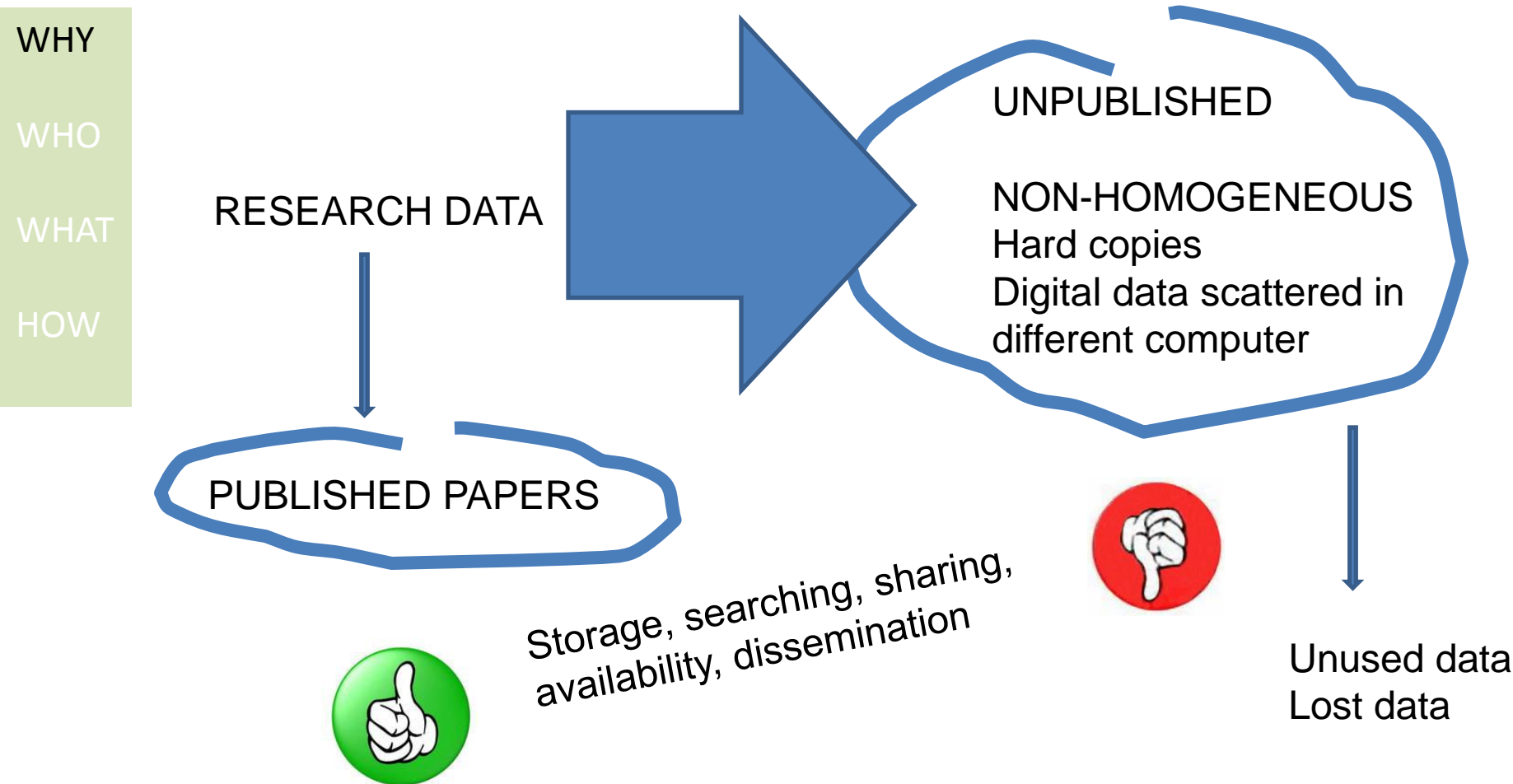
The problems are: storage
                        preservation
                        retrieving
                        searching
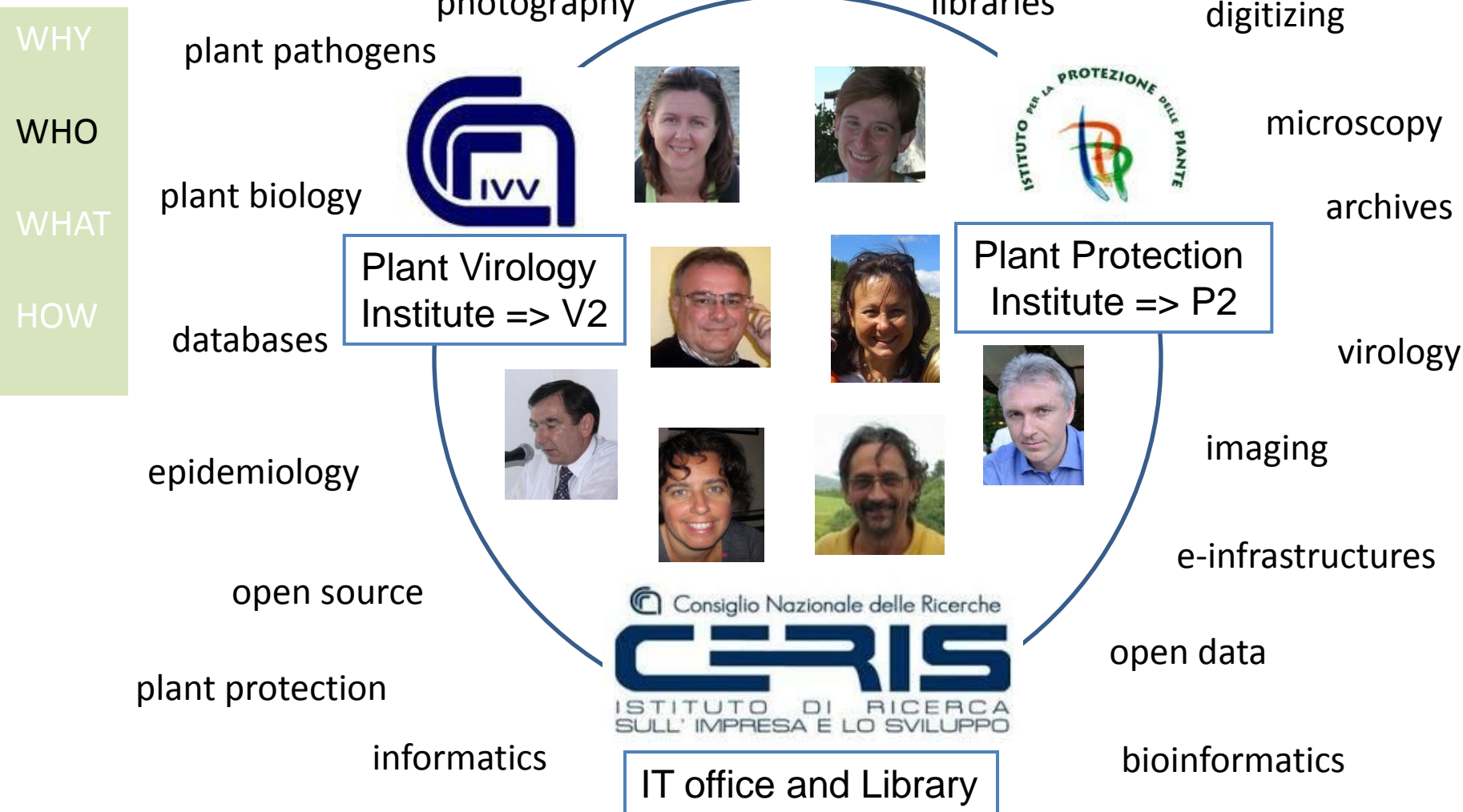   making available the data

WHY

WHO

WHAT

HOW

RESEARCH DATA

UNPUBLISHED

NON-HOMOGENEOUS
Hard copies
Digital data scattered in
different computer

PUBLISHED PAPERS

Storage, searching, sharing, availability, dissemination

Unused data
Lost data

efsa
European Food Safety Authority

Joint EFSA-EPPO Workshop
"*Data collection and information sharing in plant health*"
Parma, 1-2-3 April 2014

oepp
eppo

The need/challenge…

WHY

WHO

WHAT

HOW

PRESERVE

SEARCH

SHARE

V2P2repository **has been designed to face this challenge.**

Joint EFSA-EPPO Workshop
"Data collection and information sharing in plant health"
Parma, 1-2-3 April 2014

efsa
European Food Safety Authority

oepp
eppo

WHY

WHO

WHAT

HOW

photography   libraries   digitizing

plant pathogens

microscopy

plant biology

archives

Plant Virology Institute => V2

Plant Protection Institute => P2

databases

virology

epidemiology

imaging

open source

e-infrastructures

open data

plant protection

informatics   IT office and Library   bioinformatics

V2P2repository: a "place" to store, search and share data from research on plant-microorganism-virus interactions.

Consiglio Nazionale delle Ricerche

CERIS
ISTITUTO DI RICERCA
SULL' IMPRESA E LO SVILUPPO

WHY

WHO

WHAT

HOW

Data from scientific research in the field of
plant-microorganisms-virus interactions:

- viral infections
- plant disease symptoms
- insect vectors
-viral particles
- biology, biodiversity and ecology of soil fungi
- mycorrhizal roots

Research activity from 1960s…

efsa
European Food Safety Authority

Joint EFSA-EPPO Workshop
"Data collection and information sharing in plant health"
Parma, 1-2-3 April 2014
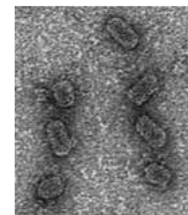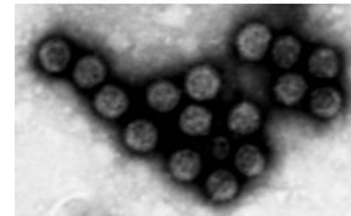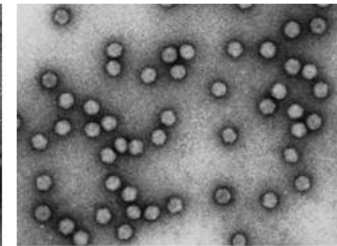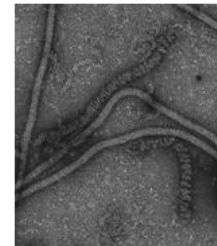
oepp
eppo

WHY

WHO

WHAT

HOW

14,600 glass plate negatives
33,000 film negatives
16,500 prints on high-definition photo paper
40,000 b/w and colour slides

Depicting plants showing viral infections, disease symptoms, insect vectors, viral particles
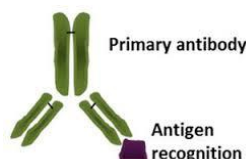
European Food Safety Authority

WHY

WHO

WHAT

HOW

WHY

WHO

WHAT

HOW
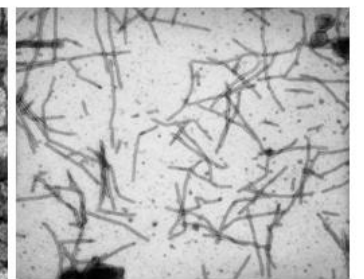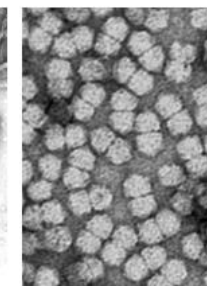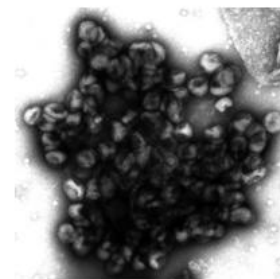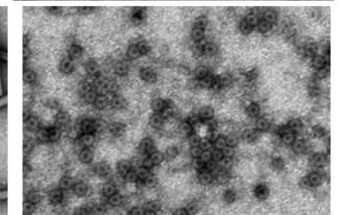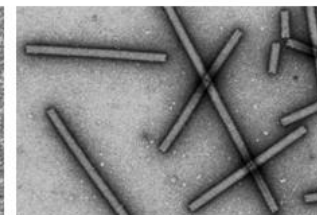
Collections of antibodies
400 antisera for rapid diagnosis of different viruses

Primary antibody

Antigen recognition

Technical/scientific reports
Grey literature

Collection of viruses

- registered with the World Federation of Culture Collections as ID #1057,
- the largest Italian plant virus and phytoplasma collection;
- started in 1960s, now has **3650 entries** (485 non-Italian)
- includes 134 different viral species and 28 phytoplasma isolates;
- stored at -20 C or in liquid nitrogen.
- Minimum infos: host plant, harvesting date, harvesting place, harvester.

WHY

WHO

WHAT

HOW

RESEARCH DATA

UNPUBLISHED

PUBLISHED PAPERS

Open source => containing costs

efsa
European Food Safety Authority

Joint EFSA-EPPO Workshop
"Data collection and information sharing in plant health"
Parma, 1-2-3 April 2014

oepp
eppo
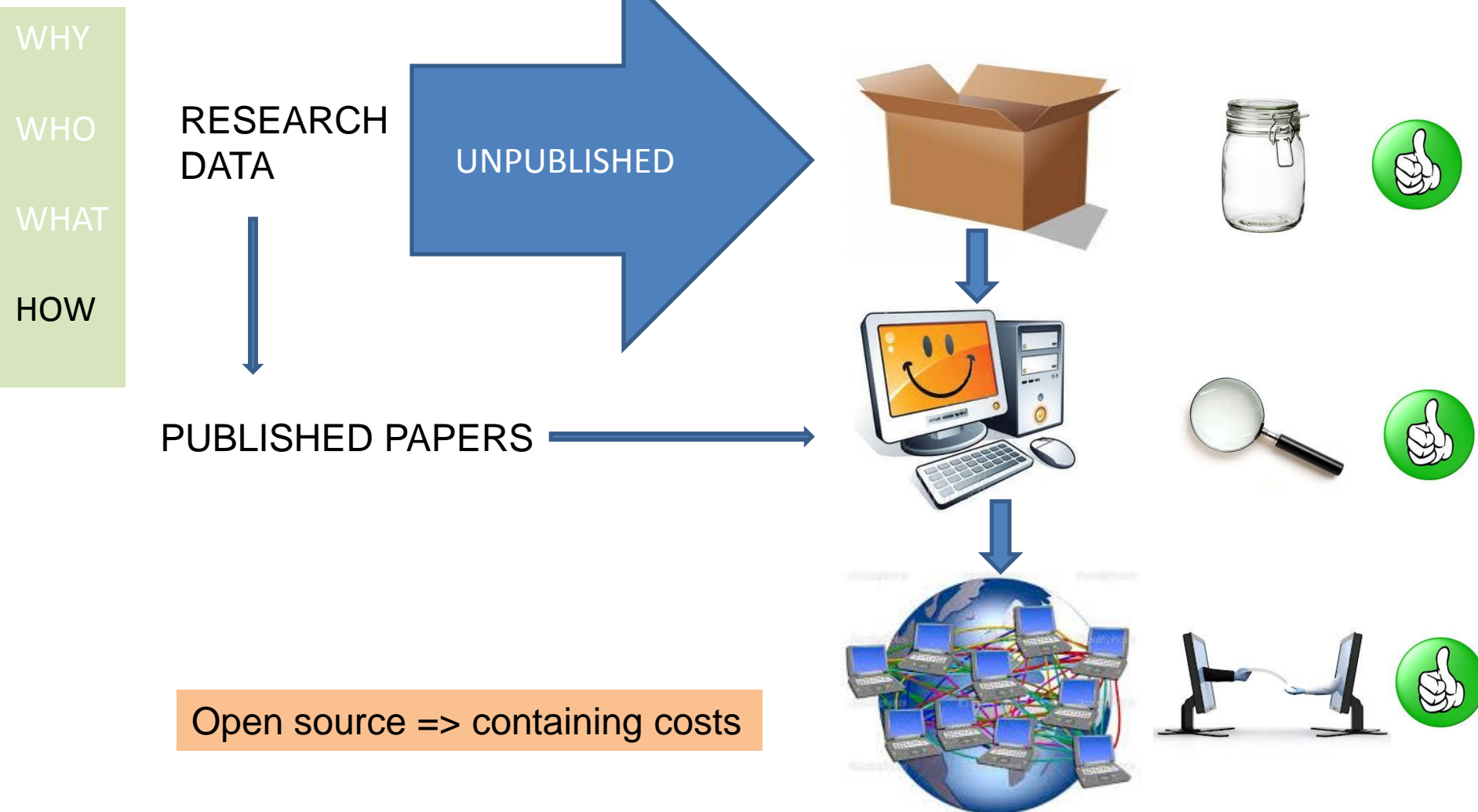
WHY

WHO

WHAT

HOW

A REPOSITORY as the best storage technology for our purposes…
…possibility to store, ingest, handle objects from different sources;
…easy data mining and interoperability;
…long-term preservation.

The repository and its front-end are based on OPEN SOURCE software and the project outcomes will be released as OPEN ACCESS.
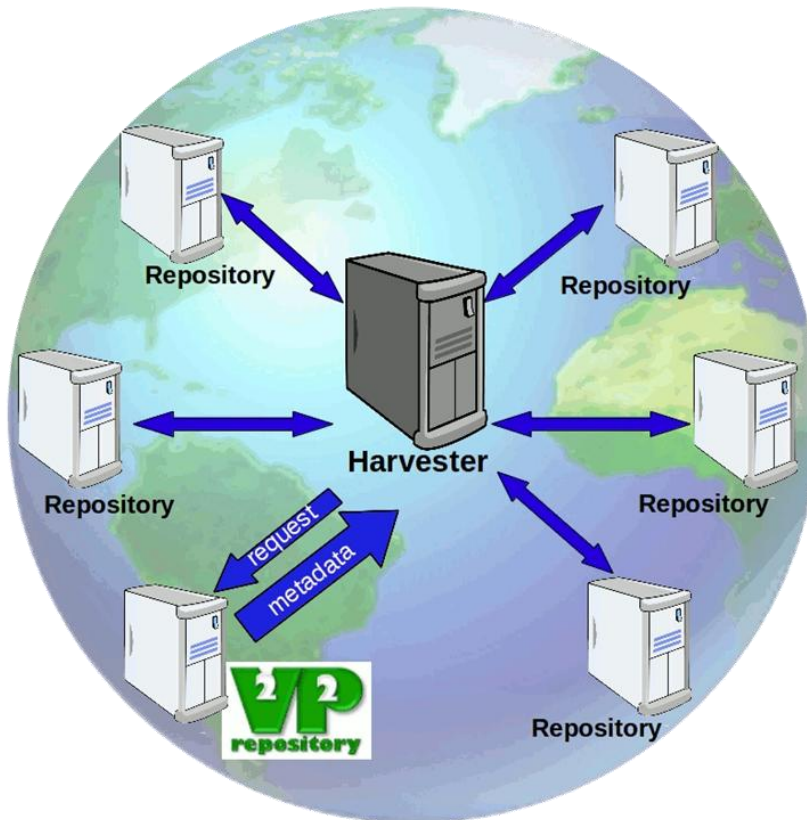

According to ROAR (Registry of Open Access Repositories) *V2P2* will be one of the first repositories to use Fedora Commons not only for text and publication archives.

WHY

WHO

WHAT

HOW
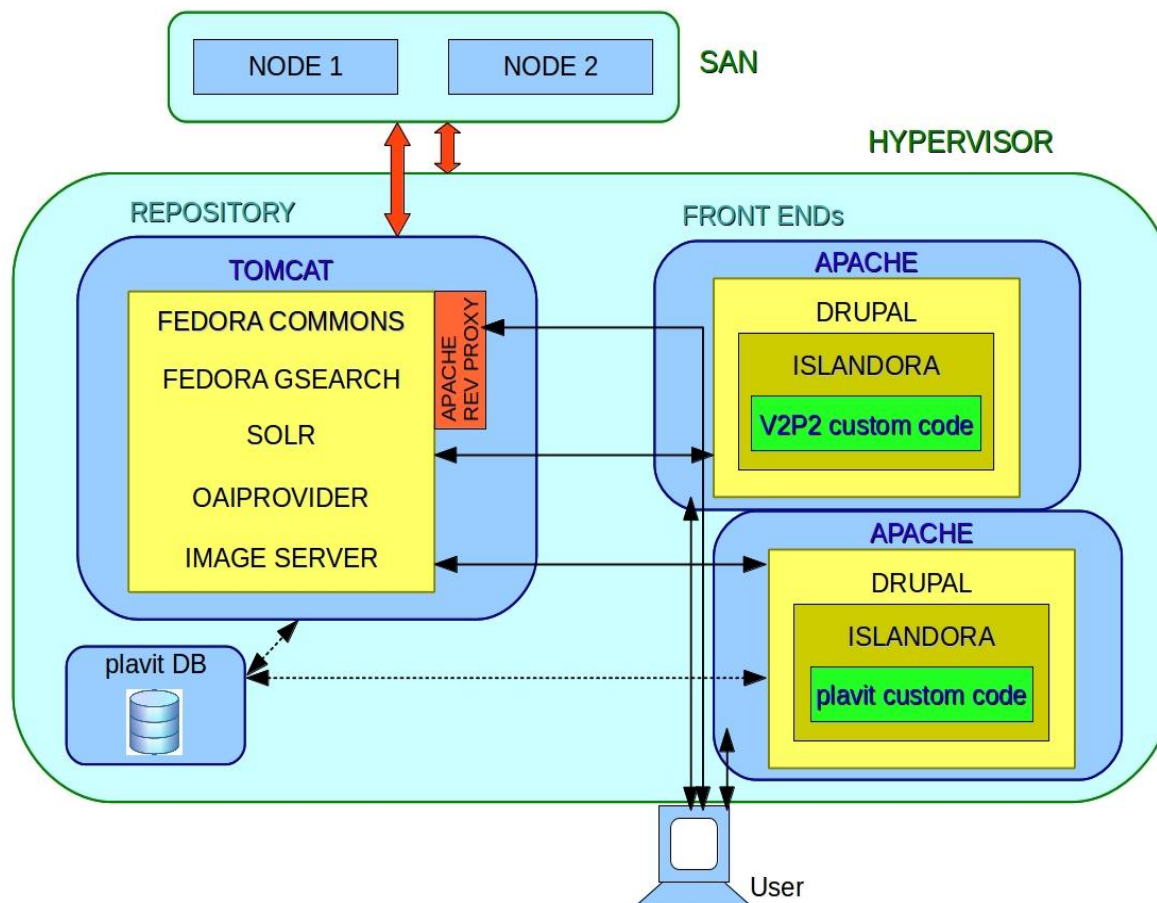
- 2-node active/passive open-source cluster
- High Availabilty Hypervisor using cluster storage
- Customized scripting for ingesting
- Islandora model and modules
- Drupal front-end
- Solr search platform from the Apache Lucene project
- Fedora Commons repository
- uploaded files: high resolution tiff, OCR txt file, pdf/a
- Metadata: Dublin Core
- Ontology and relationships: Agricultural Ontology Services (AOS-FAO)
- Full-text search

Joint EFSA-EPPO Workshop
"Data collection and information sharing in plant health"
Parma, 1-2-3 April 2014

efsa
European Food Safety Authority

oepp eppo

WHY

WHO

WHAT

HOW

OAI-PMH harvesting

Our protocols and architecture respect international standards to expose metadata to harvesters.

A harvester is a means for collecting metadata from different repositories and create useful aggregations of metadata

WORKFLOW

WHY

WHO

WHAT

HOW

Choosing material

Defining a protocol for digitizing

Digitizing

Defining metadata (Dublin Core)

**Fill in metadata sheet**

Transfer to IT office

Ingesting

Creating relationships

TITLE:
CREATOR:
CONTRIBUTOR:
SUBJECT:
DESCRIPTION:
DATE:
TYPE:
FORMAT:
IDENTIFIER:
COVERAGE:
LANGUAGE:
PUBLISHER:
RIGHTS:
SOURCE:
RELATION:

efsa
European Food Safety Authority

Joint EFSA-EPPO Workshop
"Data collection and information sharing in plant health"
Parma, 1-2-3 April 2014

oepp
eppo

Books (Handwritten registers, Proceedings, Grey literature…)
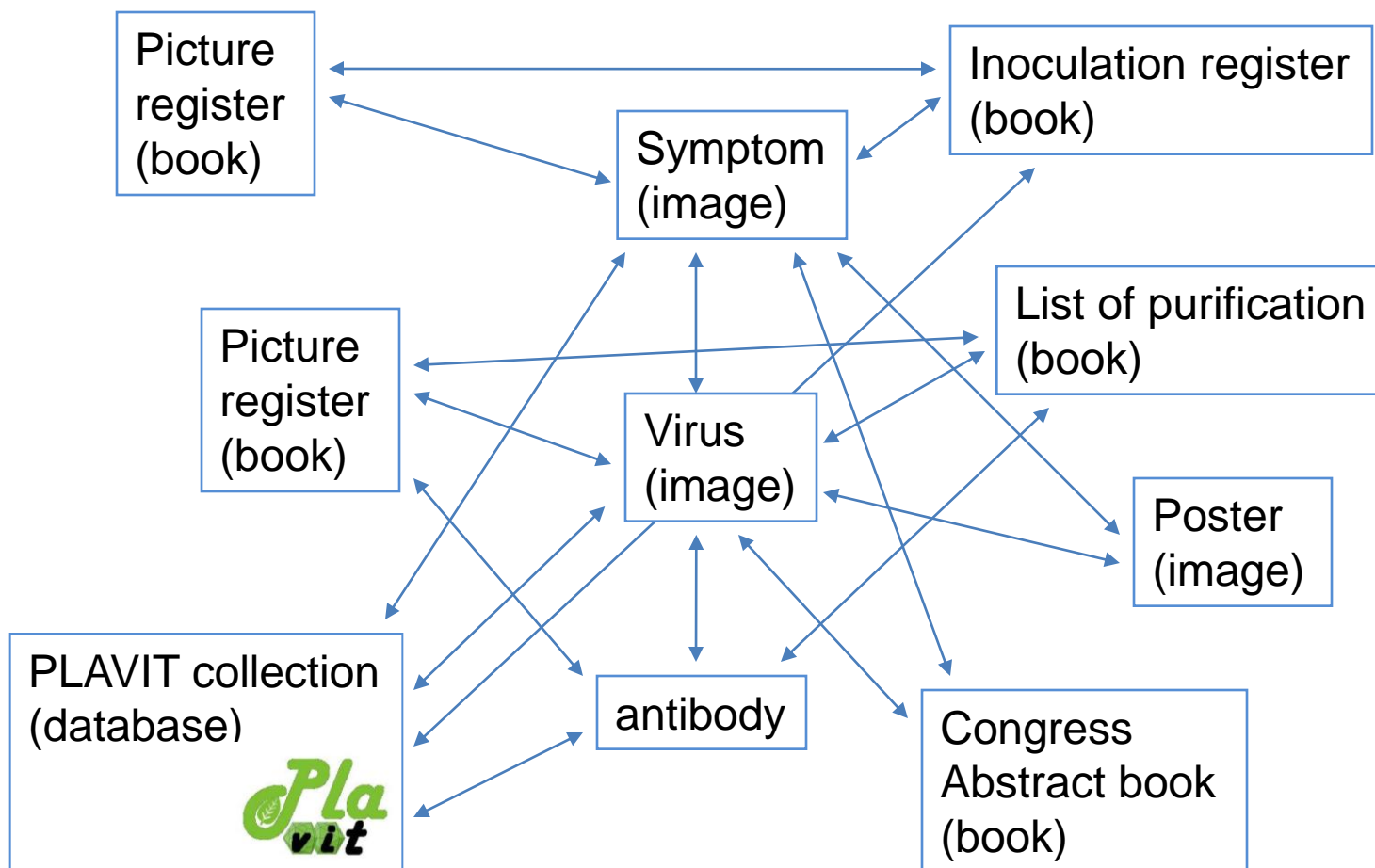
WHY

WHO

WHAT

HOW

Choosing material
Defining a protocol for digitizing
Digitizing
Defining metadata (Dublin Core)
Fill in metadata sheet
Transfer to IT office
Ingesting

Experience from DigiBESS (CERIS)
(http://www.digibess.it/)

efsa
European Food Safety Authority

Joint EFSA-EPPO Workshop
"Data collection and information sharing in plant health"
Parma, 1-2-3 April 2014

oepp
eppo

Images (Negatives, Positives, Printed copies)

Choosing material
Defining a protocol for digitizing
Digitizing
Defining metadata (Dublin Core)
Fill in metadata sheet
Transfer to IT office
Ingesting
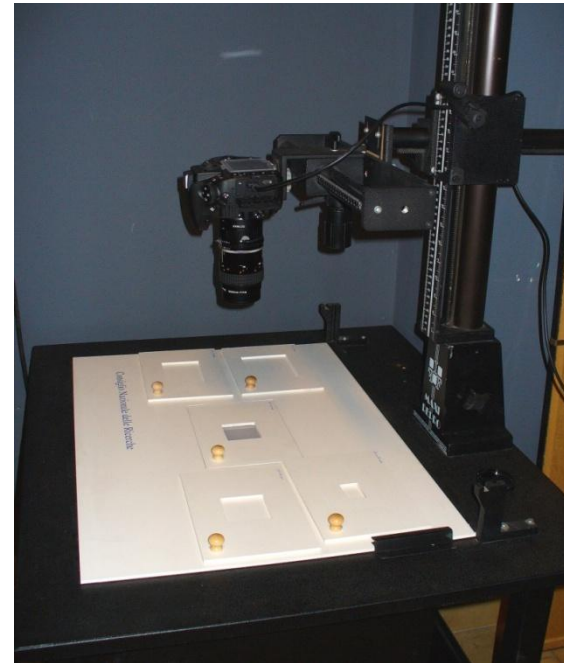
lack of guidelines for associating metadata and relationships to biological data, for managing non-homogeneous biological data.

Creating relationships: FAO-Agricultural Ontology Service

WHY

WHO

WHAT

HOW

WHY

WHO

WHAT

HOW

http://v2p2demo2.to.cnr.it/islandora/object/islandora%3Aroot

WHY

WHO

WHAT

HOW

This work is partially funded by MIUR (Ministero dell'Istruzione, dell'Università e della Ricerca), Progetti annuali (L.6/2000) D.D. 369/Ric. del 26/06/2012.

V2P2repository is a pilot project and will include only a very small part of our research data

We are looking for other funding and grants to expand our pilot project!

THE FUTURE BELONGS TO THOSE WHO UNDERSTAND HOW TO COLLECT AND USE THEIR DATA SUCCESSFULLY (BIG DATA WORLD FORUM, BEIJING 2012)