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

IVV - Istituto di Virologia Vegetale
Consiglio Nazionale delle Ricerche

Consiglio Nazionale delle Ricerche
CERIS
ISTITUTO DI RICERCA SULL' IMPRESA E LO SVILUPPO

DEVELOPMENT OF A NEW OPEN SOURCE RESEARCH INFRASTRUCTURE NETWORK FOR AGRICULTURAL DATA SHARING

Simona Abbà[1], Marta Vallino[1], Marina Ciuffo[1], Piero Caciagli[1], Giancarlo Birello[2], Anna Perin[2], Isabella Bianco[2], Ugo Finardi[2]
[1]Plant Virology Institute, IVV–CNR, Italy
[2]Institute for Economic Research on Firms and Growth, CERIS–CNR, Italy

INTRODUCTION AND AIMS

Background: sharing Open Research Data (ORD) improves the efficiency of the knowledge production processes and unlock new perspectives of data exploitation for commercial applications.

BUT

ORD are raising some issues:

1. (in general) costs and time: personnel, hardware and licensed software;

2. (in general) lack of guidelines for: a) associating metadata and relationships to agricultural data, b) building up and maintain the e-infrastructure;

3. (about plant health) unavailability of open access e-infrastructures that are user-oriented and meant for managing non-homogeneous agricultural data.

SO

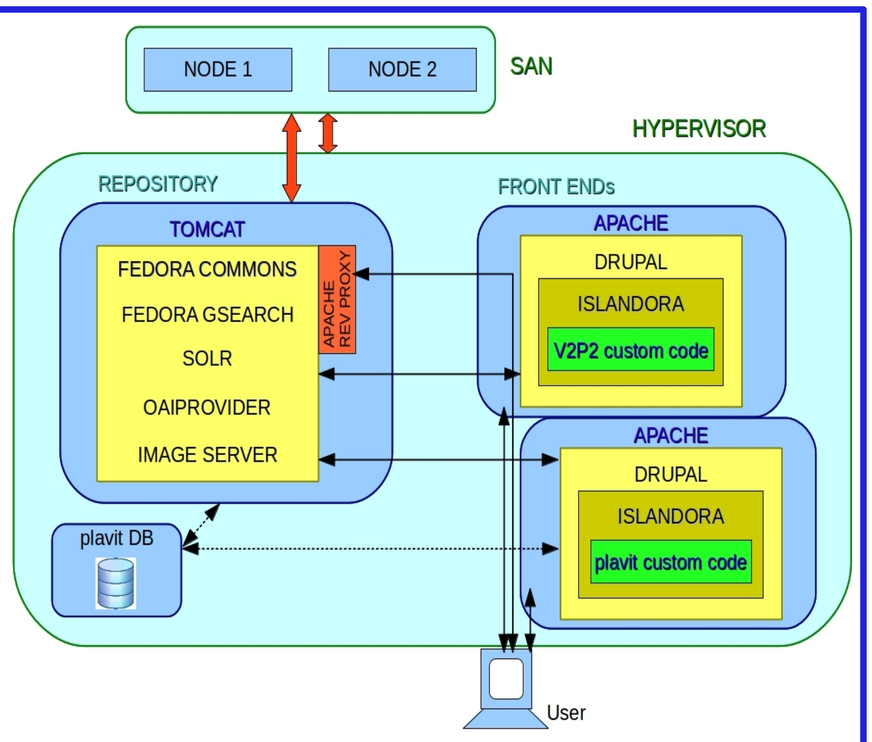
Our objectives are:

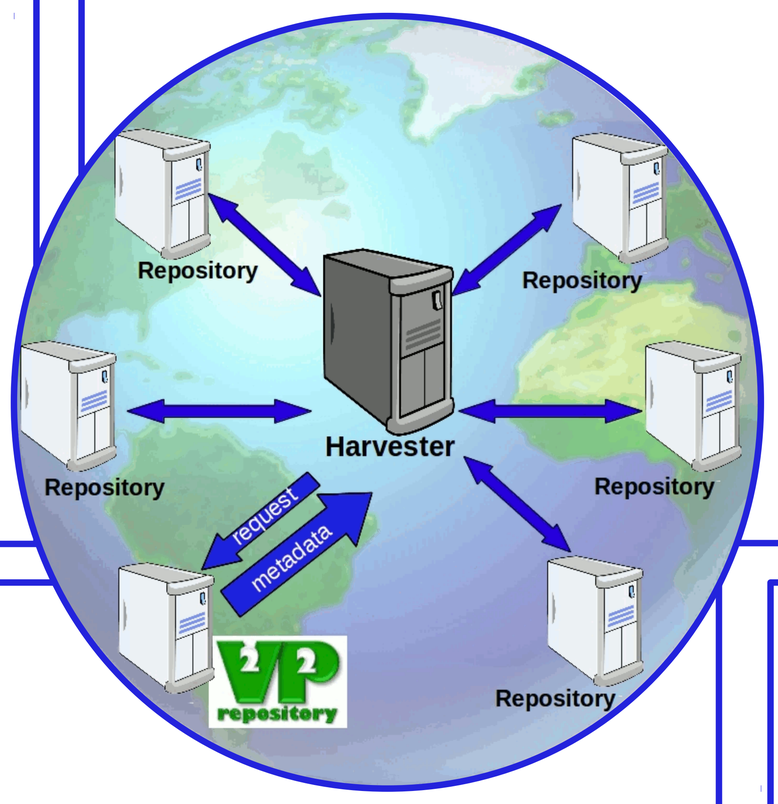
1. containing costs;
2. creating ready-to-use technical support tools and simple operating models for digitizing ORD and building up the e-infrastructure;
3. preserving and making accessible non-homogeneous agricultural ORD, involving data users in the web portal design.

METHODOLOGY
V2P2REPOSITORY: TECHNICAL SPECIFICATIONS

Why is a repository the best storage technology for our purposes?
- possibility to store different types of ORD;
- easy data mining and related research;
- preservation/durability.

Architecture characteristics and software in brief:
- 2-node active/passive open-source cluster;
- High Availability Hypervisor using cluster storage;
- customized scripting for ingesting;
- Islandora model and modules;

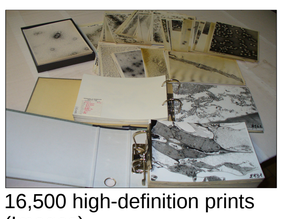




- Drupal front-end;
- Solr search platform from the Apache Lucene project;
- Fedora Commons repository;
- OAI-PMH harvesting;
- uploaded files: high resolution tiff, OCR txt file, pdf/a;
- metadata: Dublin Core;
- relationships: AGROVOC by FAO;
- full-text search.

CERIS Library and IT Office dev sites:
http://v2p2dev.to.cnr.it/doku.php
http://www.digibess.it/node/35
http://dev.digibess.it/doku.php

RESULTS
V2P2REPOSITORY: BIOLOGICAL DATA


16,500 high-definition prints (Images)


Grey literature (Texts)


14,600 glass plate negatives, 33,000 film negatives and 40,000 b/w and in color slides (Images)

PLAVIT, plant virus and phytoplasma collection (Database)

V2P2repository is a pilot project and will include only a very small part of our research data obtained in the field of plant virology and soil microbiology about:

- viral infections
- plant disease symptoms
- insect vectors
- viral particles
- biology, biodiversity and ecology of soil fungi
- roots, mycorrhized or not
- fruiting bodies and mushrooms

A demo version is available at http://v2p2demo2.to.cnr.it/

In evidence: PLAVIT, registered with the World Federation of Culture Collections as ID #1057,

- is the largest Italian plant virus and phytoplasma collection and is situated at IVV-CNR;
- started in 1960s, now has 3650 entries (485 non-Italian) stored at -20 °C or in liquid nitrogen;
- includes 134 different viral species and 28 phytoplasma isolates.

Minimum infos associated with each entry: host plant, harvesting date, harvesting place, harvester.

Stakeholders from the scientific community and industry (nurseries, regional phytopathological services, associations of farmers) have already expressed their interest in collaborating on the web portal design.

CONCLUSIONS

We are:

1. building up a repository using free open source software only;
2. providing technical support to create your own repository from scratch through our dev zones and dokuwikis;
3. providing guidelines to apply Dublin Core metadata and AGROVOC relationships to agricultural ORD;
4. designing and developing the first open access platform meant to share non-homogeneous agricultural ORD;
5. consulting and cooperating with stakeholders on the design of the V2P2repository web portal.

WORK IS STILL IN PROGRESS!

DO NOT MISS Marta Vallino's presentation on DAY 3 - Thursday 3rd April at 11.50: "V2P2repository: a "place" to store, search and share data from research on plant-microorganism-virus interaction"

ACKNOWLEDGMENTS

V2P2repository project is partially funded by "MIUR Progetti annuali, L.6/2000, D.D. 369/Ric. del 26/06/2012"

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

Thanks to the open-source community:

Islandora · Drupal · ubuntu · islandora · Apache Solr