CERNs image distribution system for the internal cloud

Hepix Fall 2010, Ithaca, NY

Romain Wartel
CERN IT - LXCLOUD

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

List of endorsers

CERN IT Department

CMS list of endorsers

Endorser 1:
- Real Name
- Digital Identity
- URL of VMIC

Endorser 2:
- Real Name
- Digital Identity
- URL of VMIC

...

RAL list of endorsers

Endorser 1:
- Real Name
- Digital Identity
- URL of VMIC

Endorser 2:
- Real Name
- Digital Identity
- URL of VMIC

RAL local endorsers:
- Real Name
- Digital Identity
- URL of VMIC

...

XYZ list of endorsers

Endorser X:
- Real Name
- Digital Identity
- URL of VMIC

Endorser Y:
- Real Name
- Digital Identity
- URL of VMIC

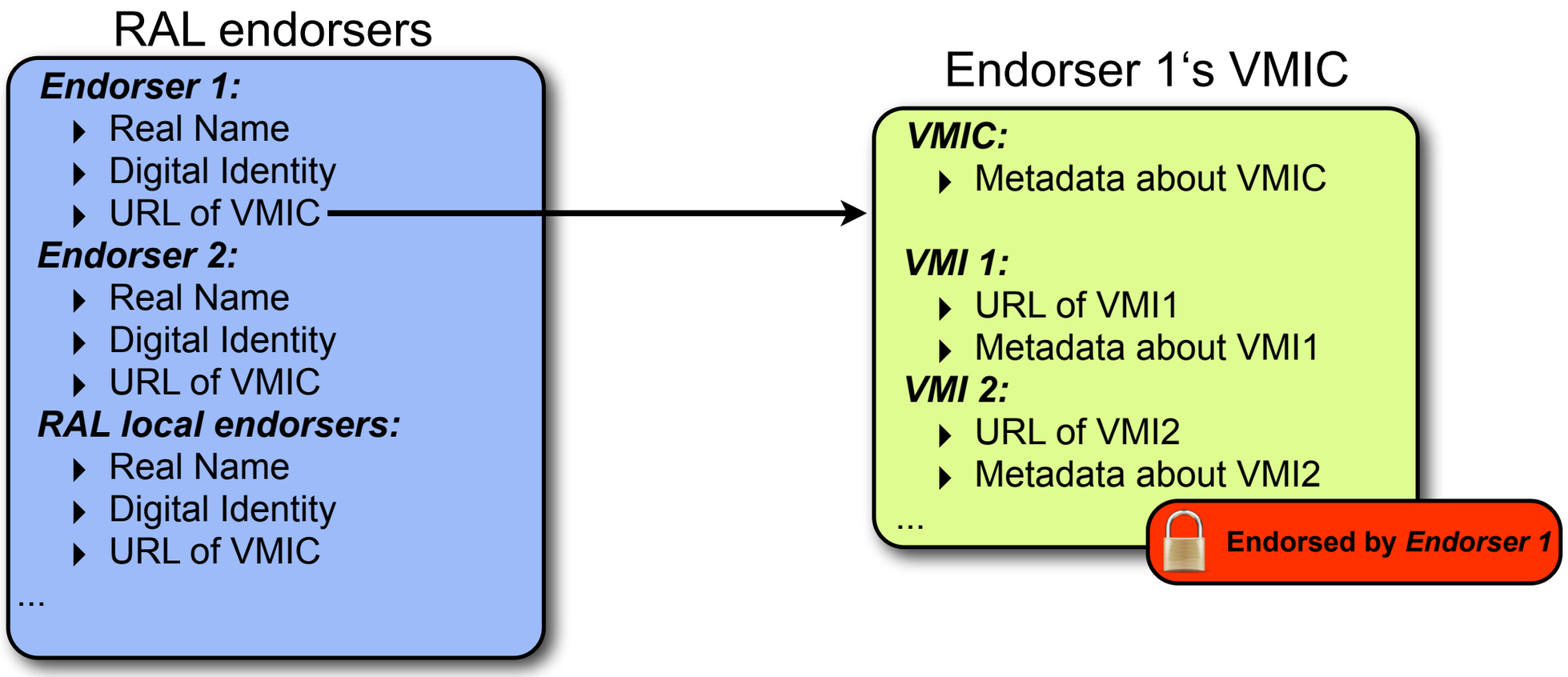
...

Different communities/sites/groups maintain a list of "valid endorsers"

- The list may built from scratch
- The list may also be based on existing list
  - With a mandatory synchronization mechanism to the original source
- Each community/site/group must document but is free to:
  - Complement the JSPG policy with more criteria to maintain the list
  - Decide how changes are managed/announced

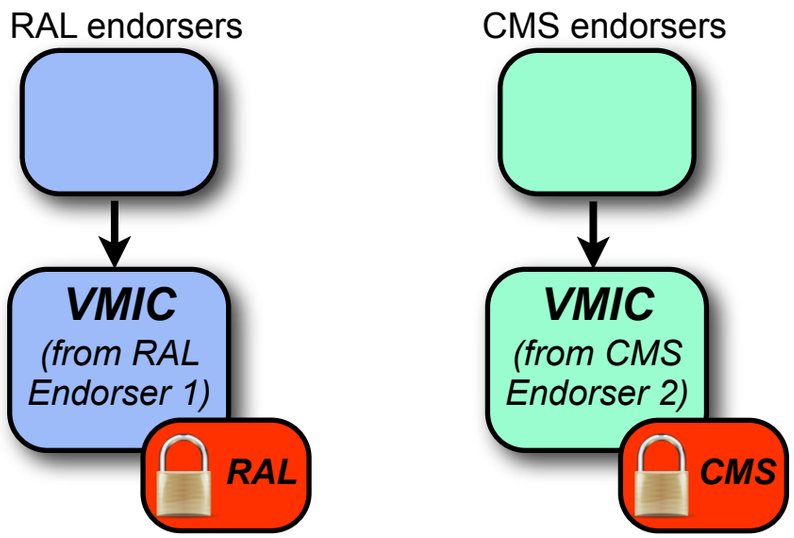
CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

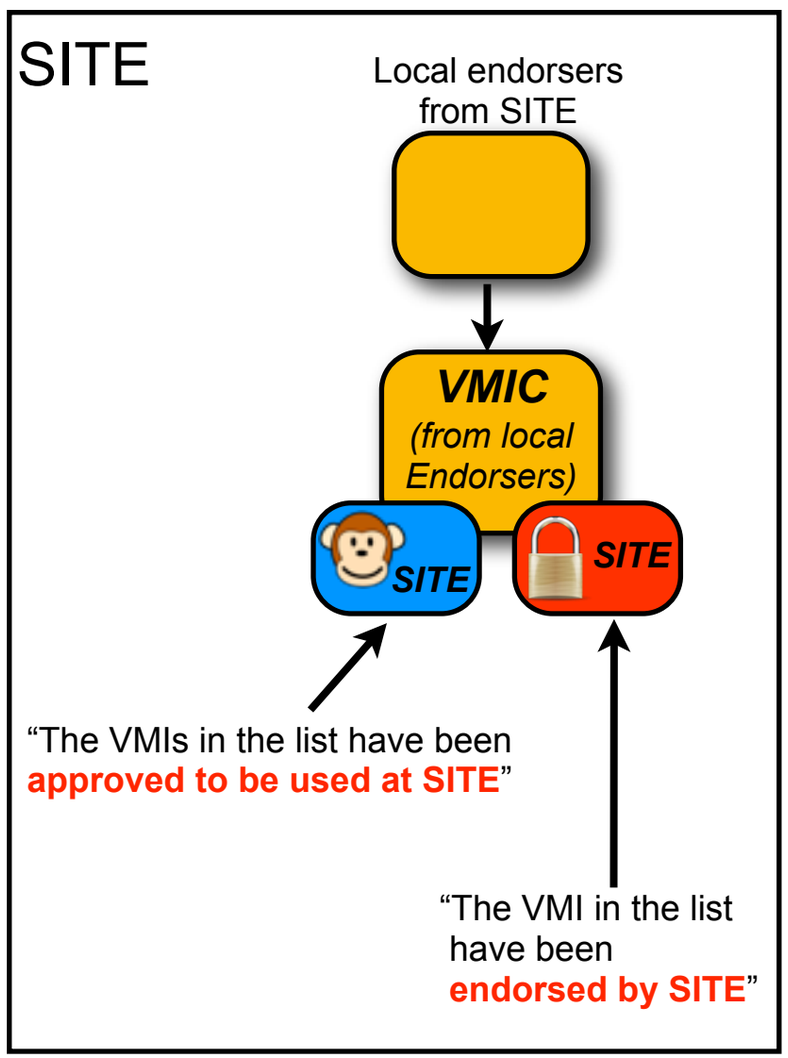
VM VMIs Catalog



- **Each endorsers publishes a VMIC**
  - The VMIC is signed by the digital identity of the endorser
- **Some metadata is included for the VMIC and for each VMI**
  - As defined in the VMIC design document:
    - VMIC: Real Name and digital identity of the endorser, max lifetime;
    - VMI: UUID, VMI checksum, OS version and 64/32bit status, hypervisor requirements, endorser's digital identity, date produced, date endorsed, VO tag;

Endorsed vs approved

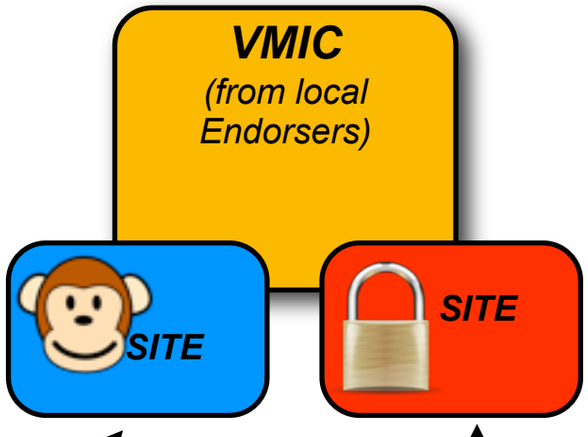




- **Distinction between:**
  - Endorsed (endorser decision):
    - Role defined in the policy document
    - Scope: VMI production & maintenance
  - Approved (site decision):
    - Marks the VMI "valid for use" by the site
    - Scope: operating the VMI
- **For a VMI to run, it must be both:**
  - Endorsed by an endorser (i.e. Part of the VMIC endorsed)
  - Approved by the local site
- **The VMI is run <u>only</u> when the two conditions are met**
  - The site has control over VMIs being run
  - The endorser has control over VMIs being produced/endorsed/published

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Endorsed vs approved

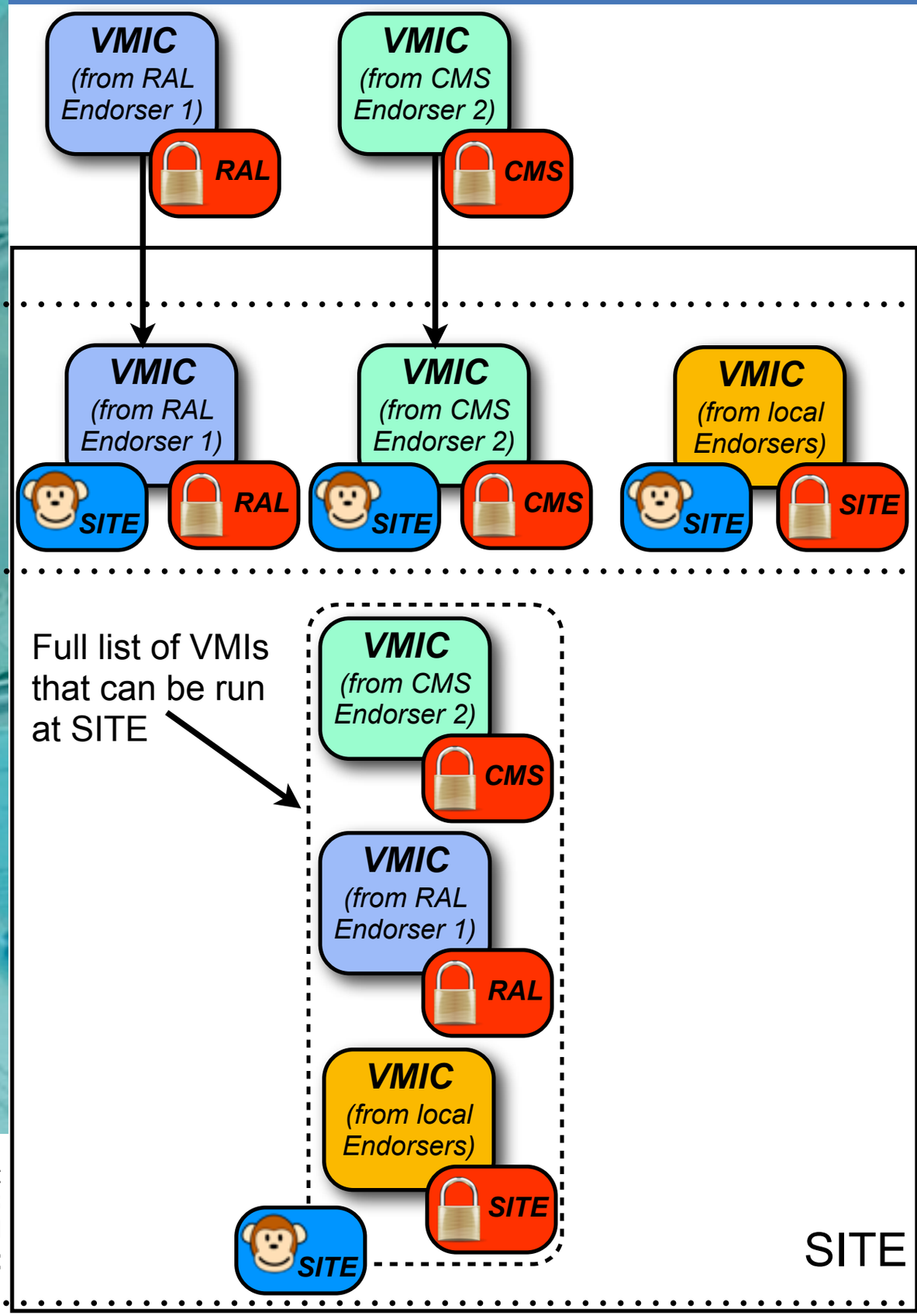


"The VMIs in the list have been **approved** to be used at SITE"

"The VMI in the list have been **endorsed** by SITE"

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Heterogenous example



1. SITE decides to approve VMIs endorsed by (XYZ) RAL and CMS

2. VMIs are approved(SITE)
Sites has fine-grained control over VMIs being approved (but can also approve them all)

3. The RAL and CMS VMIs are added to complement the VMIs produced locally

4. The resulting list of VMIs (endorsed by different entities) is approved by the local site

XYZ + SITE + SITE

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Implemention on LXCLOUD

- Environment:
  - 450+ hypervisors
  - Strong desire to use a solution compatible with the HEPiX WG
    - Collaboration essential, happy to adapt (even to learn Django...)
  - Now integrated in the production environment
- Focused on two aspects:
  - Transfer of the images to the nodes (LXCLOUD, production)
    - How to copy the images to all the hypervisors?
    - How to maintain a consistent set of images?
    - How to manage changes in the image set?
    - How could we optimise network usage?
  - Management of the virtual images (VMIC, HEPiX WG)
    - How can we establish a trusted model to share images?
    - What are the requirements on image producers?
    - How can we integrate our local image distribution with this?

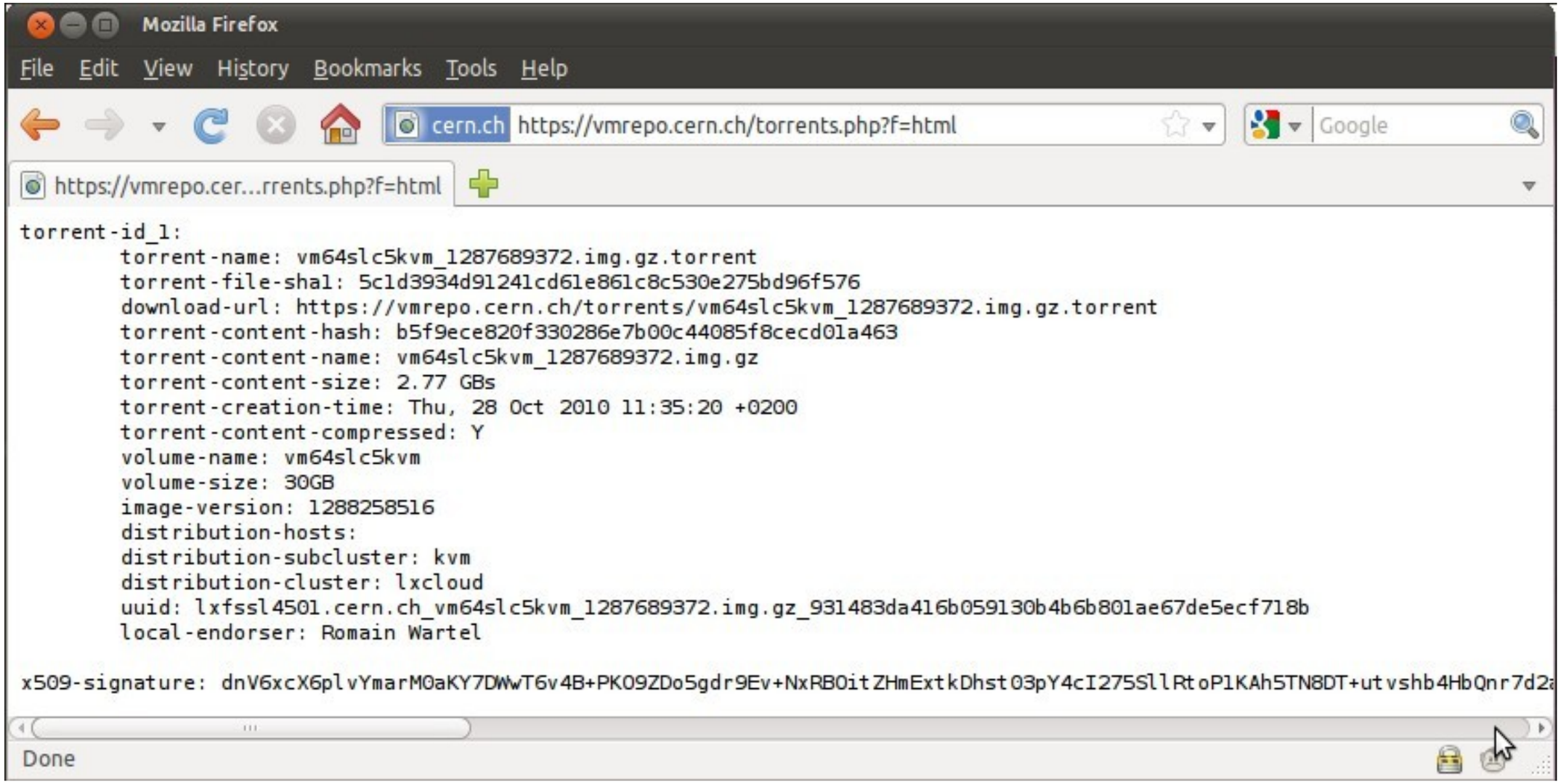
Transferring the images

CERN IT Department

- Transfer of the images using Bittorrent
  - Central torrent index of trusted images
    - Signed list of trusted torrents (in YAML and HTML)
    - Contains metadata (including hash) of valid torrents
  - All hypervisors:
    - Run a local rtorrent instance
    - Download the torrent index on a regular basis, verify its signature
    - Select the relevant torrents to be downloaded (might be all)
    - Use the YAML data to download the torrent files, and check their signatures
    - Feed the torrent files to rtorrent to download the actual image
  - Opentracker used as a “booster”
    - DHT-only is not sufficient for bootstrapping the P2P network
    - DHT and Peer Exchange enabled on all nodes
  - Tunning:
    - Hypervisors can upload only from max 5 peers, and seed to max 10 peers
    - More throttling is possible but not implemented (yet?)

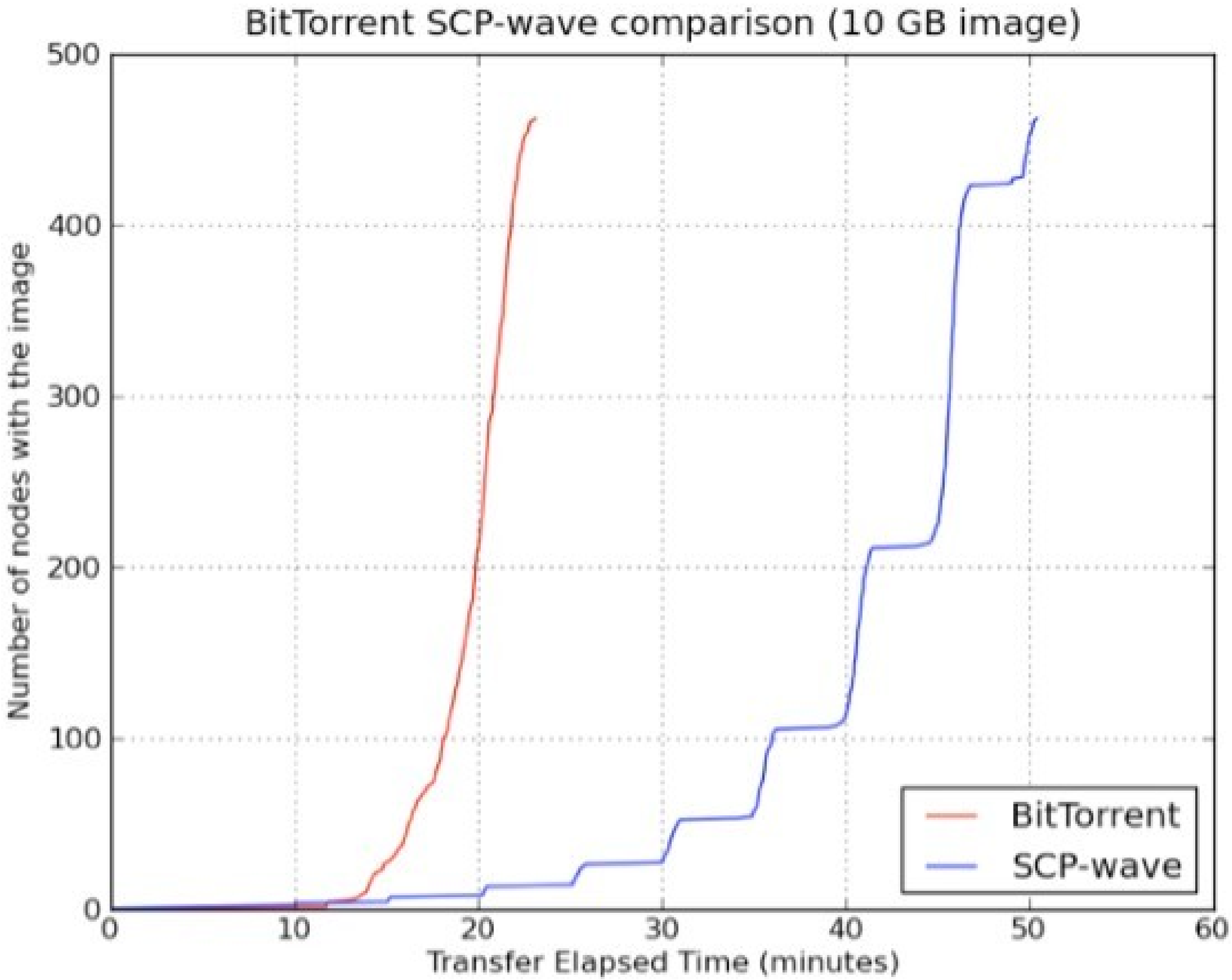
https://twiki.cern.ch/twiki/bin/view/FIOgroup/VMImageDistribution

Torrent index



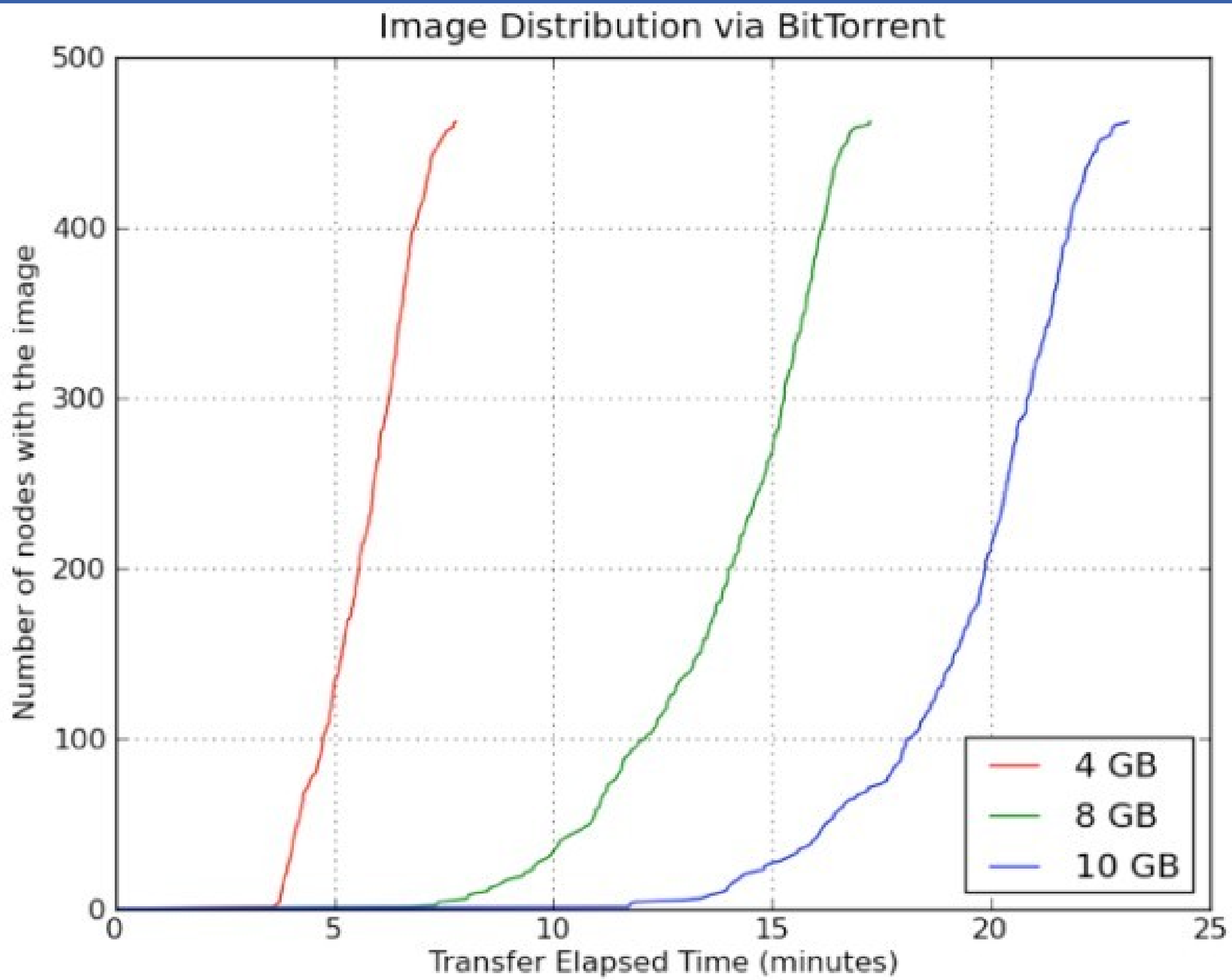
CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Bittorrent performance



CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

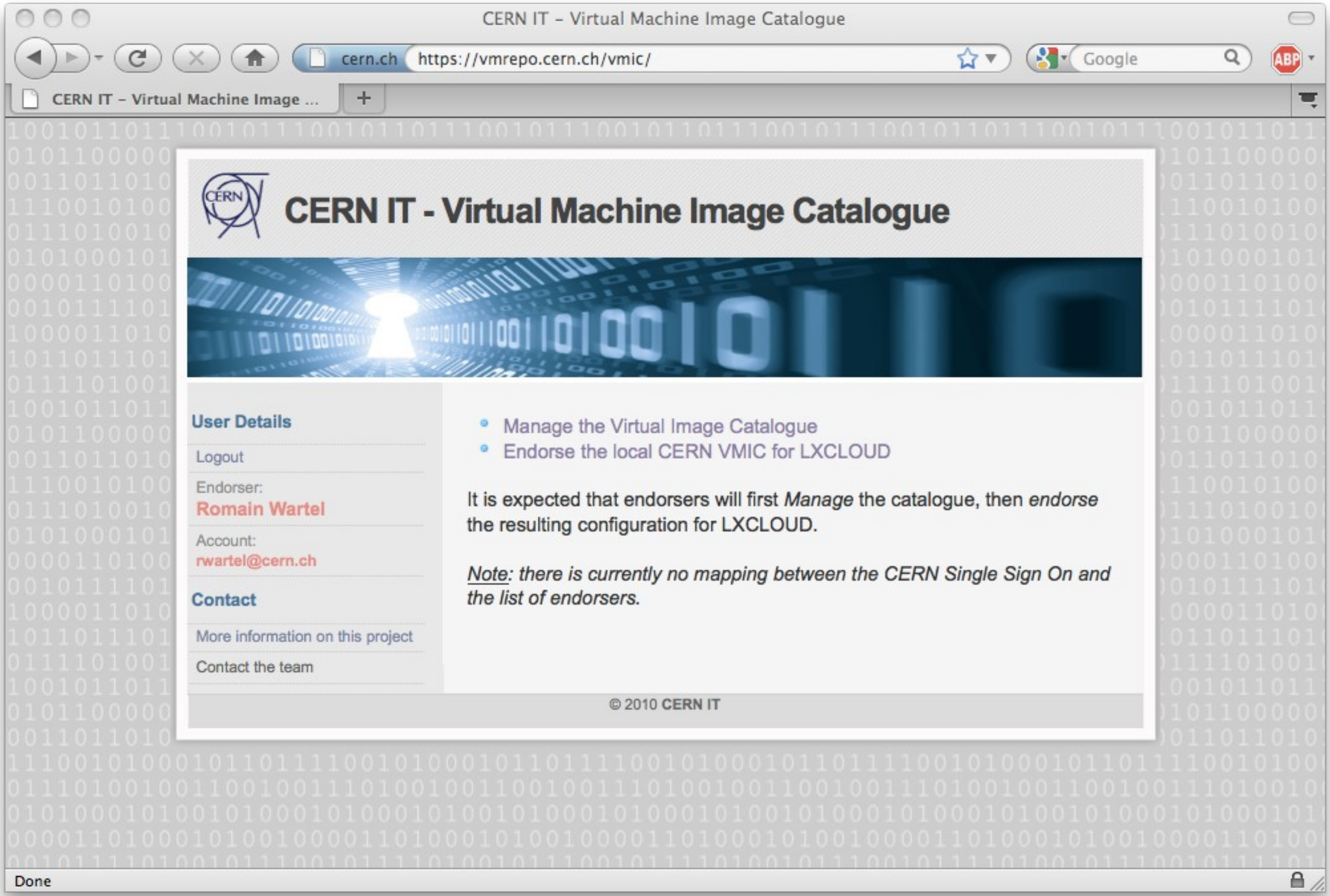
Bittorrent performance



CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

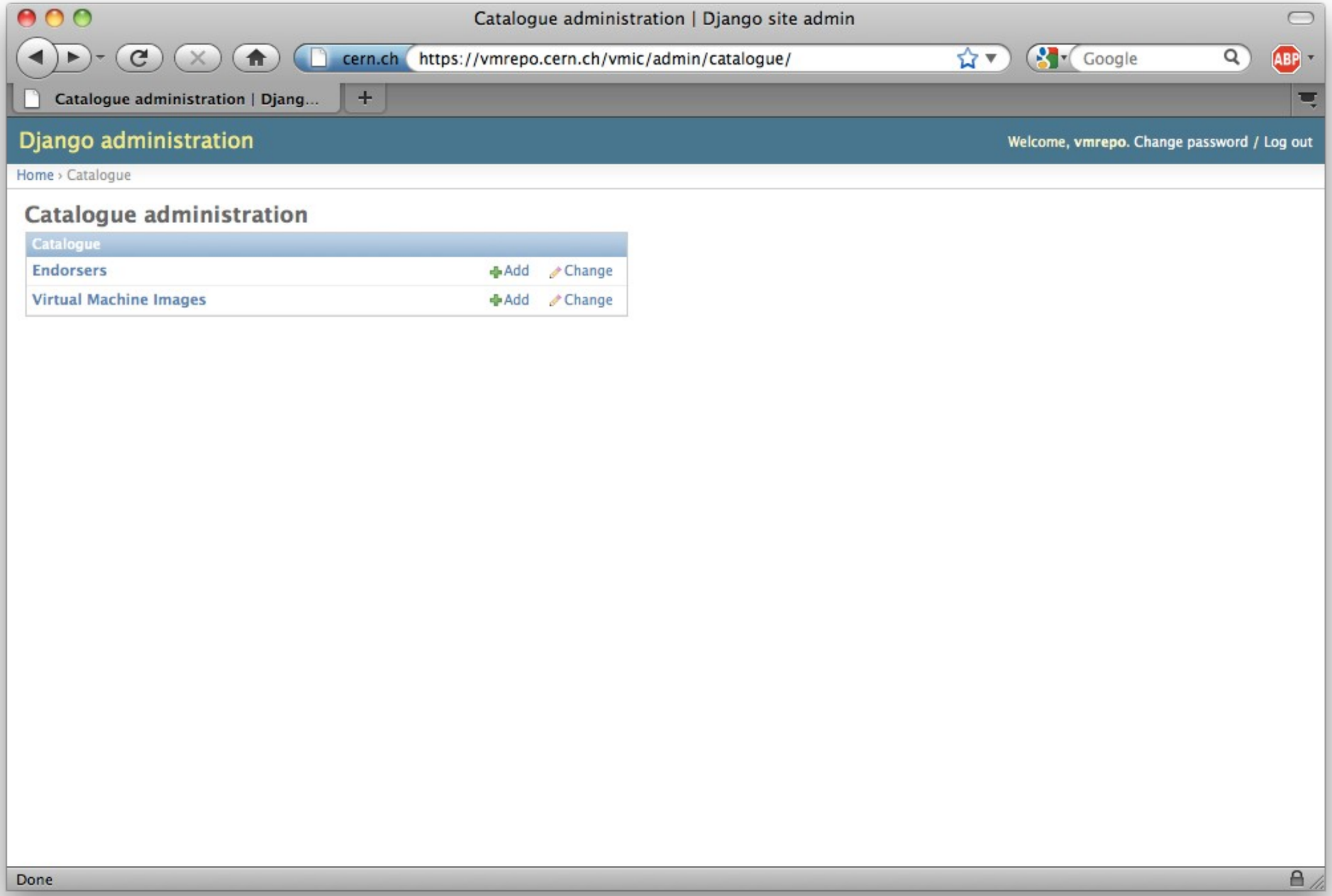
VMIC

CERN IT Department

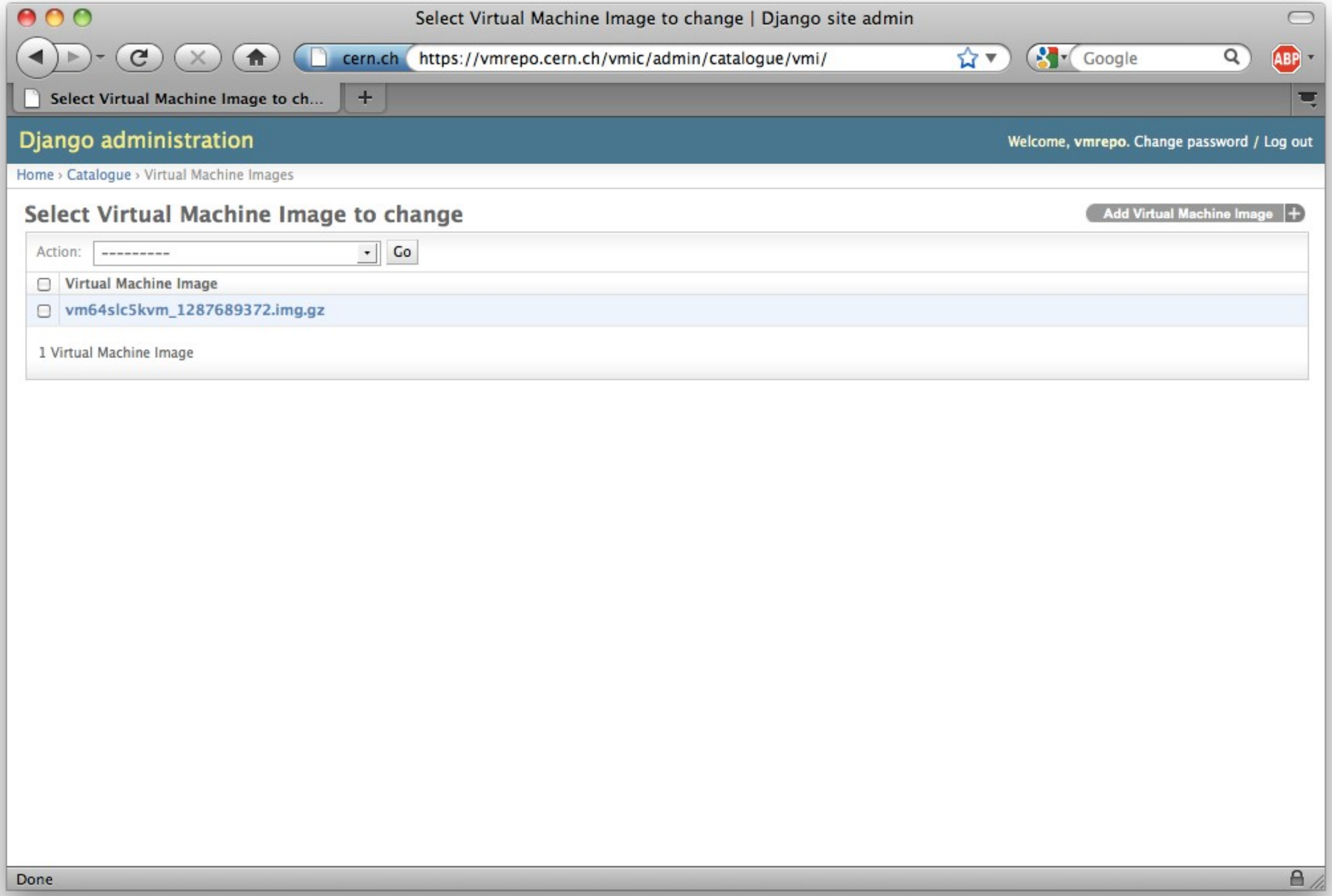


CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

VMIC


Catalogue administration | Django site admin
cern.ch
https://vmrepo.cern.ch/vmic/admin/catalogue/
Google
Catalogue administration | Djang...
Django administration
Welcome, vmrepo. Change password / Log out
Home › Catalogue
Catalogue administration
Catalogue
Endorsers
Add
Change
Virtual Machine Images
Add
Change
Done

VMIC



CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

VMIC

CERN IT Department

Change Virtual Machine Image | Django site admin

cern.ch https://vmrepo.cern.ch/vmic/admin/catalogue/vmi/2/

Change Virtual Machine Image | D...

Change Virtual Machine Image

History

VMI endorsement

Endorser: LXCLOUD endorser

VMI download location

VMI full path: /vmrepo/shared/torrents/vm64slc5kvm_

Status of the VMI

☑ This VMI is APPROVED to be run locally ☐ This VMI can be shared with other sites

Metadata about the VMI

Production date: Date: 2010-10-21 Today | Time: 22:05:32 Now |

Endorsement date: Date: 2010-10-21 Today | Time: 22:05:35 Now |

Hypervisor: lxbsq0908

Metadata about the VM

OS version: SLC5

Architecture: x86_64

VO tags: all

CERN metadata about the VM

☑ Cern torrent content compressed

Volume name: vm64slc5kvm

Volume size: 30GB

Image version:

Distribution hosts:

Distribution subcluster: kvm

Distribution cluster: lxcloud

Delete

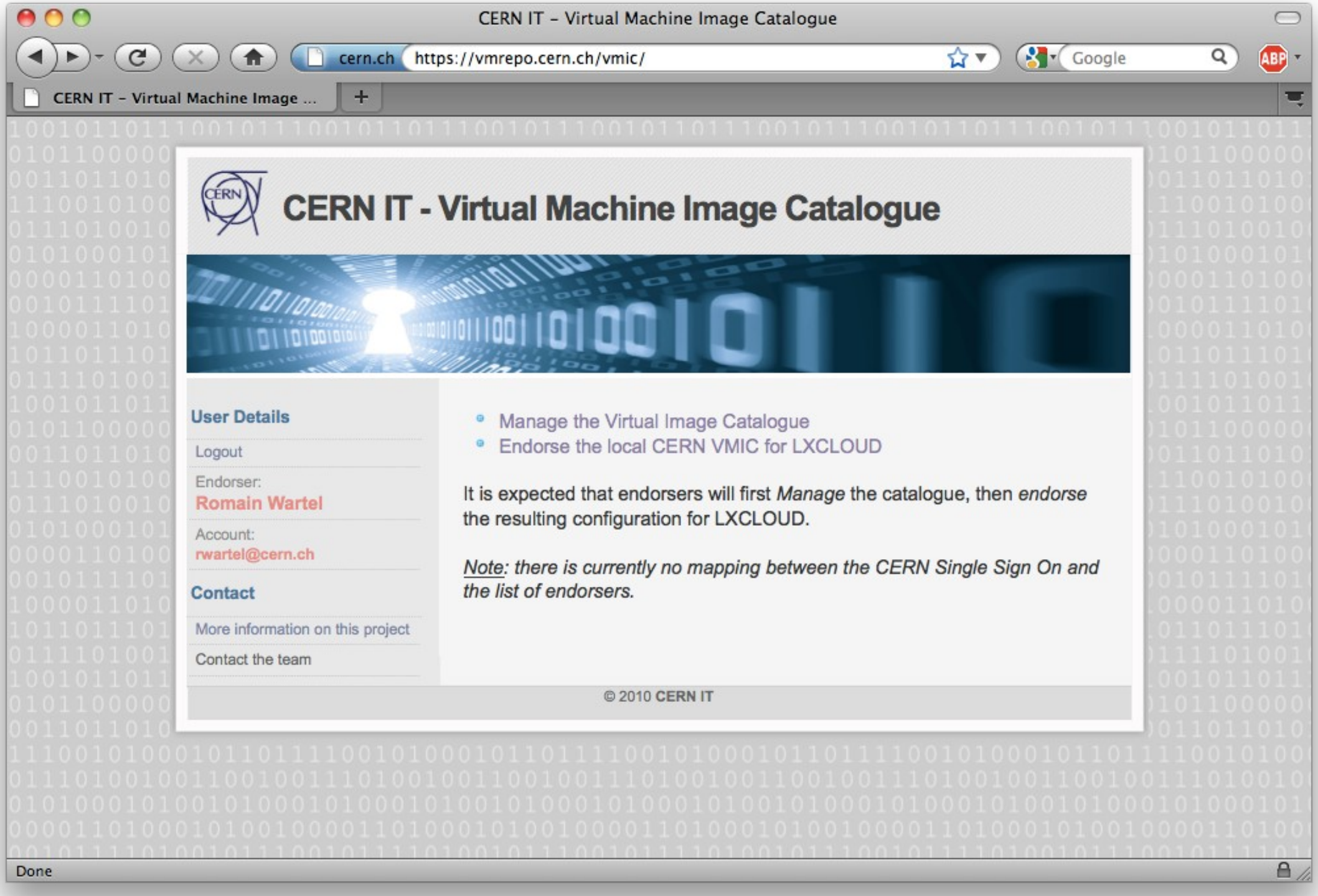
Save and add another | Save and continue editing | Save

Done

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

VMIC

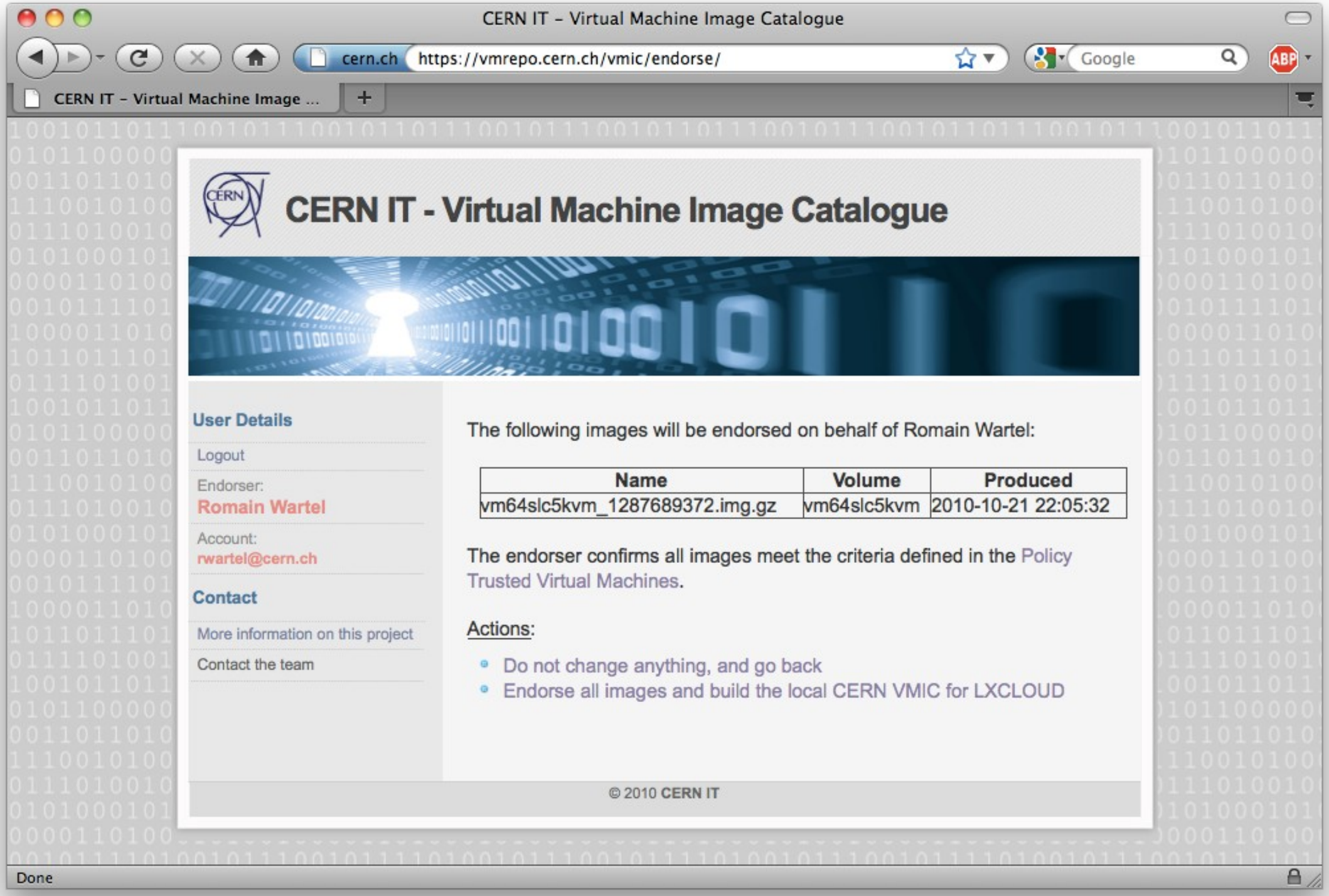
CERN IT Department



CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

VMIC

CERN IT Department



CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

VMIC

Mozilla Firefox

cern.ch https://vmrepo.cern.ch/vmic/build/

https://vmrepo.cern.ch/vmic/build/

Building local VMIC

Archiving old images

Nothing to archive.

Endorsing and registering the Virtual Machine Images

This VMIC release will be endorsed on behalf of **Romain Wartel**.

Running /usr/local/bin/vmic_add_torrent --volumename vm64slc5kvm --volumesize 30GB --subcluster kvm --cluster lxcloud --compressed Y --uuid lxfssl4501.cern.ch_vm64slc5kvm_1287689372.img.gz_931483da416b059130b4b6b801ae67de5ecf718b --endorsed-on-behalf 'Romain Wartel' /vmrepo/shared/torrents /vm64slc5kvm_1287689372.img.gz

Resulting file: vm64slc5kvm_1287689372.img.gz.torrent

Done

Future work

- Need to gain experience from production
  - Cleanup, rtorrent optimisations, etc.
- Need a better local user management on the VMIC
  - Abusing Django's admin interface is not nice
    - All VMIC endorsers authenticated via Single Sign On (SSO)
    - All VMIC endorsers the share the Django admin account
    - Their SSO uid is added to the metadata
  - Perhaps a good idea to introduce multifactor?
    - Important to provide a generic (i.e. non CERN-specific) solution
    - It seems Yubikeys can be used to be map Django users...
- Volunteers are welcome:
  - To test the current software in another environment
  - To implement missing export/import functions
  - To test image sharing