



Virtual Earthquake and seismology Research Community in Europe e-science environment  
Project 283543 – FP7-INFRASTRUCTURES-2011-2  
[www.verce.eu](http://www.verce.eu)  
[info@verce.eu](mailto:info@verce.eu)

# Seismology Data Management in VERCE

Visakh Muraleedharan (CNRS-IPGP)  
Alessandro Spinuso (KNMI)  
and VERCE Team



Helsinki, 19th May 2014

# VERCE Project Partners

## Scientific Partners

Centre National de la Recherche Scientifique (CNRS-INSU), IPGP and ISTerre, France  
Royal Netherlands Meteorological Institute (KNMI-ORFEUS), Netherlands  
European-Mediterranean Seismological Centre (EMSC), France  
Istituto Nazionale di Geofisica e Vulcanologia (INGV), Italy  
Ludwig-Maximilians-Universität (LMU), Germany  
University of Liverpool (ULIV), United Kingdom

## Technology Partners

University of Edinburgh (UEDIN), United Kingdom  
Bayerische Akademie der Wissenschaften (BADW-LRZ), Germany  
Fraunhofer-Gesellschaft e.V. (SCAI), Germany  
Centro di Calcolo Interuniversitario (CINECA), Italy



Seismology Data Management in VERCE

Helsinki, 19th May 2014

# VERCE Project

VERCE supports seismology research by developing a data-intensive e-science environment

Goals:

- Combine computing infrastructures (EGI, PRACE, CLOUD) and local resources
- Access to European data archives and services
- Workflow tools and Registries
- Data Management and Provenance System
- Software as a service via the VERCE Science Gateway (<http://portal.verce.eu>)

# Two classification of use cases in VERCE

## HPC Use cases

- Generation of synthetic seismograms enabling evaluation and comparison of various Earth Models
- Data source: Configuration files, input data, mesh and models consist of roughly 300MB
- Intermediate data:~ 4GB of data produced after mesh processing.
- Results:Synthetic seismograms, plots, 3D images, Videos.  
(100 stations = 900 products and metadata )  
5-10 GB for a 1000 cores run

## DI Use cases

- Processing real data from stations and noise cross-correlation to analyse and study various Earth Models
- Typically:
- Data archive 382 GB
  - 1-day stack for 210 pairs, 1 filter 5.9 GB
  - REFs for 210 pairs 13 MB
  - Each moving-window stack for 210 pairs, 1 filter 6.0 GB

\* MSNoise <http://srl.geoscienceworld.org/content/85/3/715.full.pdf>

# Goals of a Data Management Platform

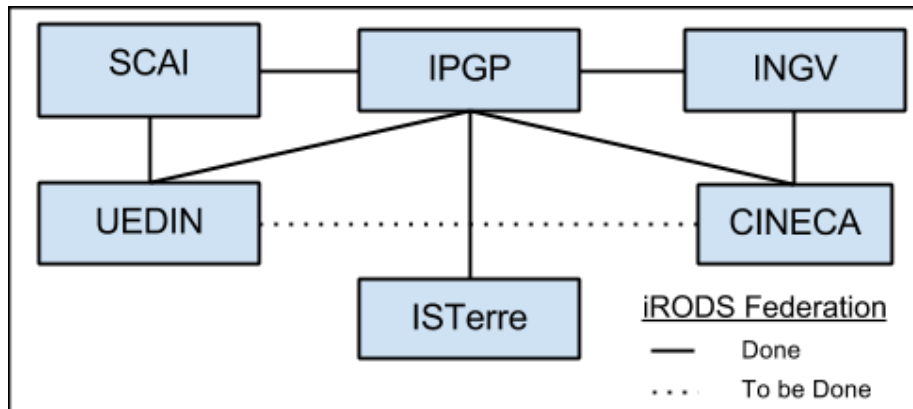
- Integrate resources available in different partner sites
- Preserve data policies of different partners
- Provide access based on scientific metadata
- Provide fast parallel data transfer capability to different applications
- Minimise the movement of data during processing

# iRODS in Partner sites

iRODS is the backbone of this data platform that

- Integrates iRODS installations at different partner sites
- Retains full data privacy and permission to administrators of each site
- Provides rules (triggers) and microservices to catalog/ingest data
- Includes interface to different types of data resources

VERCE has iRODS infrastructure setup and running in the following partner sites.



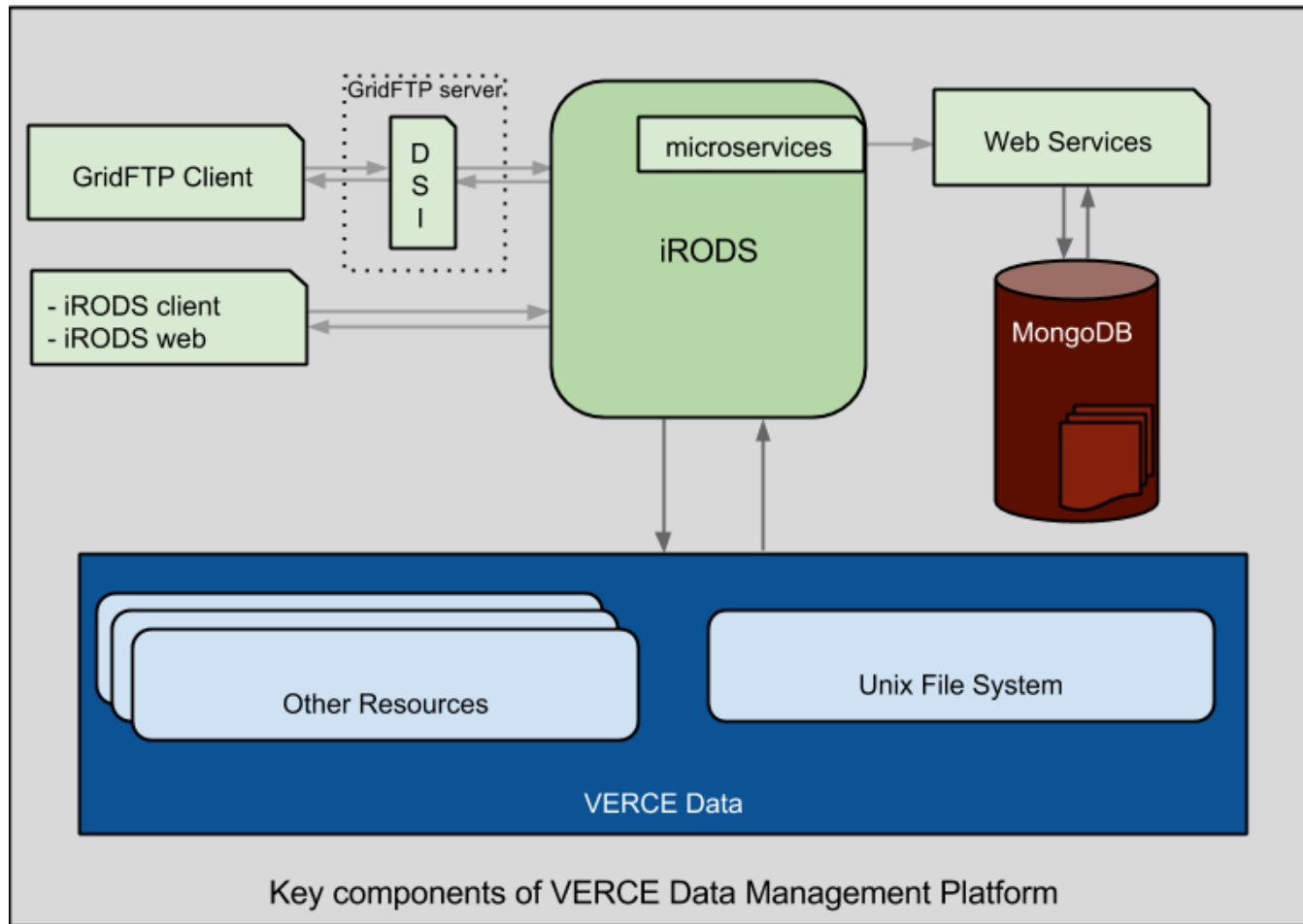
CINECA, INGV and ISTERre already use iRODS for managing user data in production environment.



# Test environment

- Further modifications to support the workflow is tested by VERCE developers
- iRODS installation at University of Edinburgh is used for these tests
- Currently this setup supports the workflows for HPC use case
- This has all the elements setup to support VERCE platform
- On successful evaluation, this configuration will be implemented in partner sites

# Elements of VERCE data platform (1/3)



Test environment setup using OpenNebula Virtual Machines  
at University of Edinburgh (EDIM1)



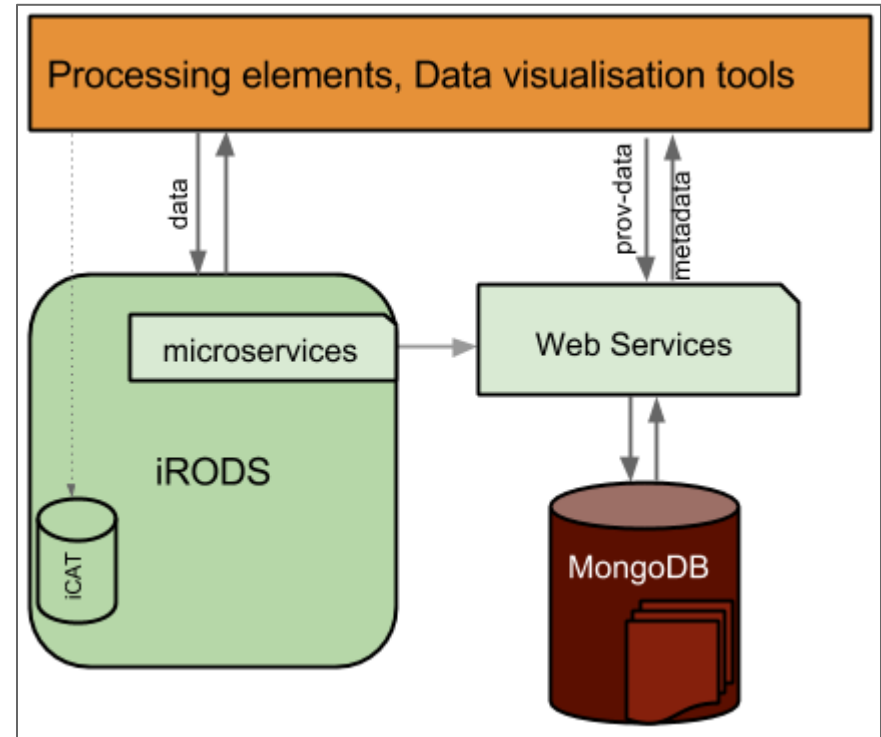
# Elements of data platform (2/3)

**MongoDB** catalog is used to catalog metadata and provenance data

During forward simulation the provenance data is stored and associated with results stored in iRODS

In case of raw data, iRODS **microservices** extract and store metadata from file **header** based on events or rules

Different processing elements and applications query the catalog to get the files based on metadata



iRODS and external catalog (EDIM1)

# Elements of data platform (3/3)

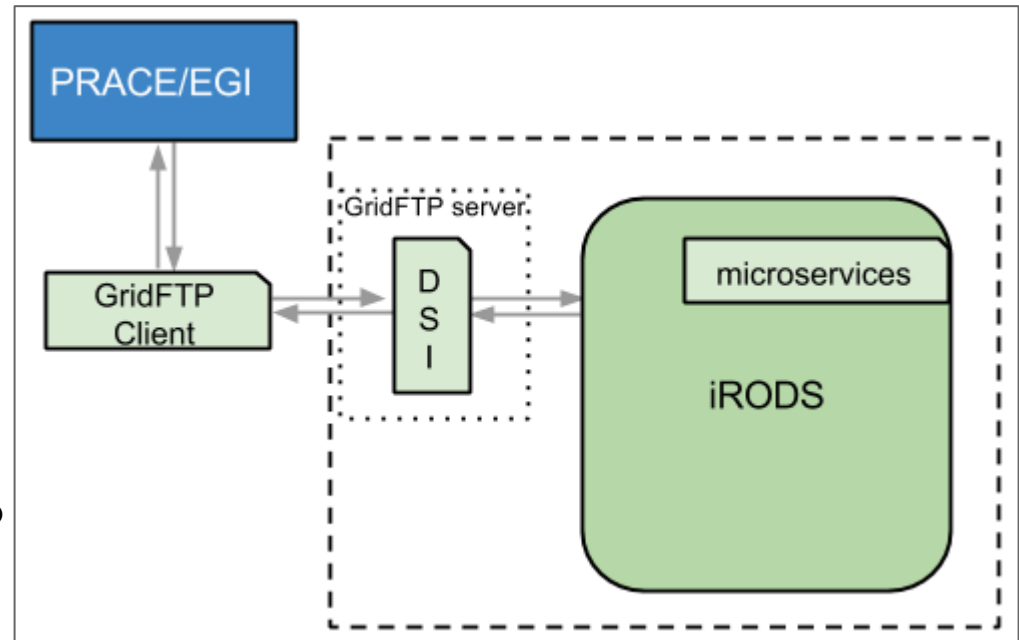
iRODS provides GSI authentication

Typically data is generated from HPC or Grid resources. Moving this results to the data platform requires high throughput parallel transfer

Even though iRODS provides native parallel transfer capability between iRODS server and its client, using a standard transfer protocol like GridFTP is required with PRACE and EGI resources

CINECA has developed a GridFTP iRODS DSI to provide a standard interface for iRODS

\* <https://hpc-forge.cineca.it/trac/iRODS-Tools>



GridFTP Interface for iRODS (EDIM1)

# Client tools

iDrop-web interface showing file management options (Refresh, New Folder, Info) and metadata for a folder. The metadata includes:

- Created: Sun Jan 28 16:00:00 GMT 2007
- Updated: Sun Jan 28 16:00:00 GMT 2007
- Owner: rods
- Owner Zone: UEDINZone
- Type: NORMAL
- Object Path: /
- Description: Basic information, including update of tags and a description

```
visakh@ui1:~$ voms-proxy-init --voms verce.eu
Enter GRID pass phrase:
Your identity: /O=GRID-FR/C=FR/O=CNRS/OU=IPGP/CN=Visakh Muraleedharan
Creating temporary proxy ..... Done
Contacting verce-voms.scai.fraunhofer.de:15000 [/C=DE/O=GermanGrid/OU=Fraunhofer
SCAI/CN=verce-voms.scai.fraunhofer.de] "verce.eu" Done
Creating proxy ..... Done

Your proxy is valid until Mon May 12 23:18:09 2014
[visakh@ui1 ~]$ iinit
Using GSI, attempting connection/authentication
[visakh@ui1 ~]$ ils
/IPGP05Zone/home/visakh:
MANAGE.txt
caneco
caneco234
caneco26
caneco7
client.py
out
provout.jsn
[visakh@ui1 ~]$
```

iCommands

```
visakh@ui1:~$ uberftp dir-irods.epcc.ed.ac.uk
220 localhost GridFTP Server 6.43 (gcc64, 1396993756-83) [Globus Toolkit 5.2.6rc
0] ready.
230 User visakh logged in.
UberFTP> ls ../visakh#IPGP05Zone
-rwxr-xr-x  0    root    root      14891 Mar 18 14:11 INSTALL.txt
UberFTP>
```

globus-url-copy

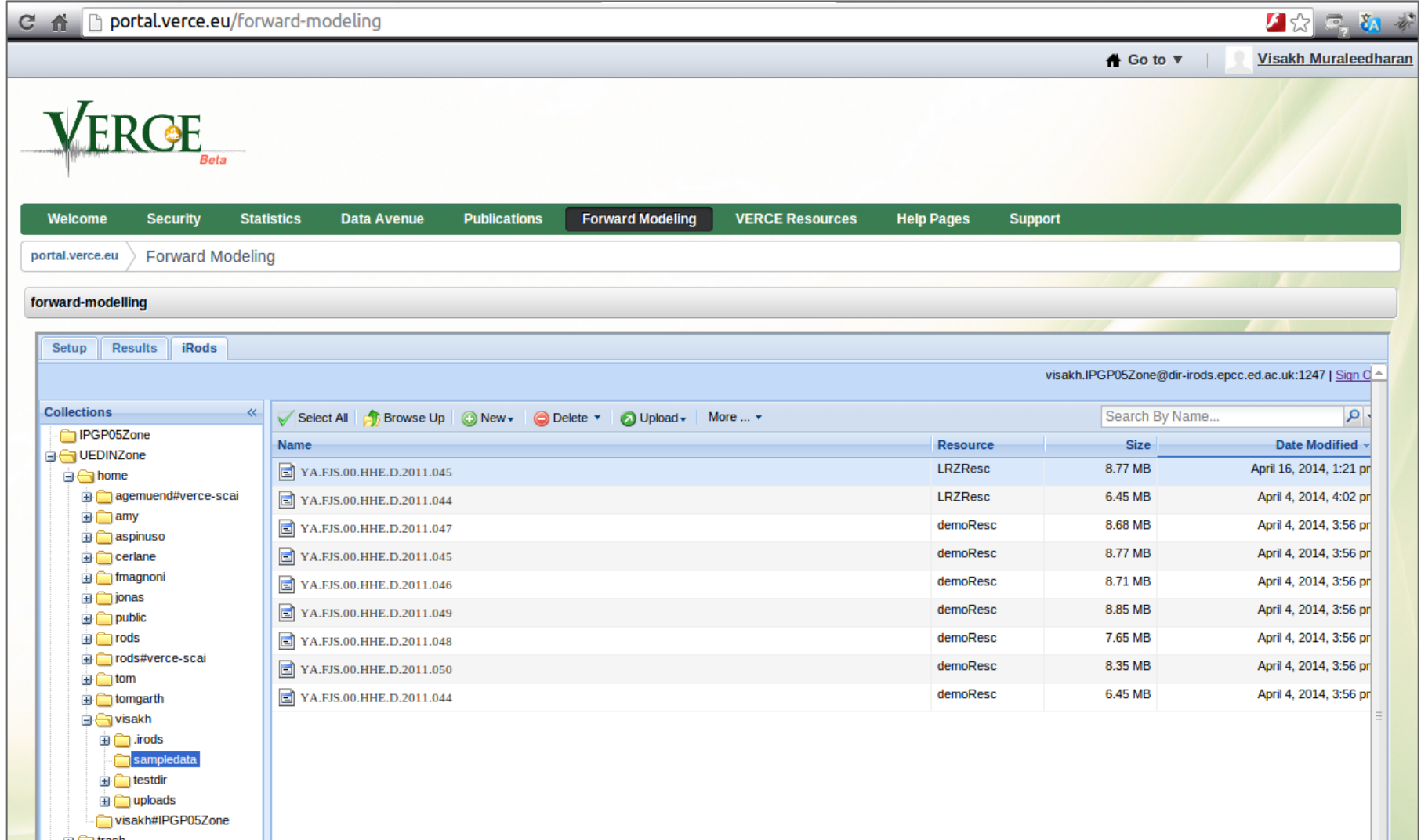
iDrop-Desktop interface showing a file system view with columns for File System, size, and last modified. The file system view includes:

File System	size	last modified
/		
IPGP05Zone		Mon Mar 17 12:39:50 CET 2014
UEDINZone		Fri Mar 07 13:30:49 CET 2014
home		Fri Mar 07 13:30:49 CET 2014
trash		Fri Mar 07 13:30:49 CET 2014
verce		Fri Apr 04 12:51:58 CEST 2014
verce-scai		Mon Mar 17 13:35:00 CET 2014

Upload Options: Manage Pause



# Web Interface and portal integration



The screenshot displays the VERCE web interface for forward modeling. The browser address bar shows `portal.verce.eu/forward-modeling`. The user is logged in as Visakh Muraleedharan. The navigation menu includes: Welcome, Security, Statistics, Data Avenue, Publications, Forward Modeling (active), VERCE Resources, Help Pages, and Support. The breadcrumb trail is `portal.verce.eu > Forward Modeling`. The main content area is titled `forward-modelling` and contains tabs for Setup, Results, and iRods. The iRods tab is active, showing a file management interface with a sidebar for Collections and a main table of files.

**Collections**

- IPGP05Zone
- UEDINZone
  - home
    - agemuend#verce-scai
    - amy
    - aspinuso
    - cerlane
    - fmagnoni
    - jonas
    - public
    - rods
    - rods#verce-scai
    - tom
    - tomgarth
    - visakh
      - .irods
      - sampledata
      - testdir
      - uploads
      - visakh#IPGP05Zone

**File Management Table**

Name	Resource	Size	Date Modified
YA.FJS.00.HHE.D.2011.045	LRZResc	8.77 MB	April 16, 2014, 1:21 pm
YA.FJS.00.HHE.D.2011.044	LRZResc	6.45 MB	April 4, 2014, 4:02 pm
YA.FJS.00.HHE.D.2011.047	demoResc	8.68 MB	April 4, 2014, 3:56 pm
YA.FJS.00.HHE.D.2011.045	demoResc	8.77 MB	April 4, 2014, 3:56 pm
YA.FJS.00.HHE.D.2011.046	demoResc	8.71 MB	April 4, 2014, 3:56 pm
YA.FJS.00.HHE.D.2011.049	demoResc	8.85 MB	April 4, 2014, 3:56 pm
YA.FJS.00.HHE.D.2011.048	demoResc	7.65 MB	April 4, 2014, 3:56 pm
YA.FJS.00.HHE.D.2011.050	demoResc	8.35 MB	April 4, 2014, 3:56 pm
YA.FJS.00.HHE.D.2011.044	demoResc	6.45 MB	April 4, 2014, 3:56 pm

# To the future...

- Development of query services for this platform is in progress
- Investigating possibilities of pre-processing/downsampling data before shipping
- Distributed data preparation in data nodes triggered by user defined rules
- Workflow integration

# Summary

- VERCE data platform allows integration of different partner resources
- Each partner retains the full access to their user data
- Better data access provided through metadata and provenance catalog
- GridFTP interface provides faster data transfer to compute resources
- Investigating ways to minimise data transfer during data processing

Beta version of portal available at:

<http://portal.verce.eu/home>

Demo:

<https://www.youtube.com/watch?v=Tkr36KWowAA>

Support:

<http://portal.verce.eu/support>

# Thank you!

## Questions?

Beta version of portal available at:

<http://portal.verce.eu/home>

Demo:

<https://www.youtube.com/watch?v=Tkr36KWowAA>

Support:

<http://portal.verce.eu/support>

## Connect with us

Website:

[www.verce.eu](http://www.verce.eu)

Email:

[info@verce.eu](mailto:info@verce.eu)