



Hewlett Packard
Enterprise

LA UCM como productora de BIGDATA

Cátedra Extraordinaria en Big Data Analytics, HPE- Universidad
Complutense

JM Udías, Javier Gutiérrez Puebla
'encargados'

An example from outside UCM



<http://spectrum.ieee.org/energy/environment/global-warming-to-environmental-activism>

<http://www.globalforestwatch.org/>

A new forest-mapping tool relies on unprecedented data crunching plus google-earth pictures of the regions under study to assess the degree of deforestation around the world

1 million CPU-core hours on 10 000 computers

Researchers from the World Resources Institute told GOOGLE staff what to look for, and they post-processed them with their own algorithms

Multiple activities participated by UCM
call for big data analysis

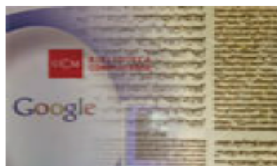
Colección Digital Complutense (CDC)

AVISOS

Por problemas técnicos y de forma temporal, el acceso a algunos recursos de la Colección Digital Complutense está limitado y no se pueden hacer búsquedas conjuntas en la colección.

Una extensa colección documentos en acceso abierto: artículos científicos, libros y grabados antiguos, tesis doctorales leídas en la UCM y materiales docentes constituyen esta Colección Digital Complutense

Búsqueda de Libros BUC-Google



Decenas de miles de libros a texto completo pueden ya consultarse gracias al trabajo que la Biblioteca de la Universidad Complutense y Google realizan conjuntamente con otras bibliotecas de prestigio y multitud de editoriales.

Archivo Institucional E-Prints Complutense



Más de 5.200 tesis complutenses, y materiales de apoyo a la docencia y a la investigación en acceso abierto.

Archivo Rubén Darío



2.221 documentos digitalizados, transcritos y clasificados, procedentes del Archivo Rubén Darío.

Colección de dibujos antiguos de Bellas Artes



Conjunto de 287 dibujos, entre 1752 y 1914.

Portal de Revistas Científicas Complutenses



Texto completo de los artículos publicados en las revistas científicas editadas por el Servicio de Publicaciones de la UCM y, asimismo, de aquellas otras revistas editadas por los departamentos de la UCM que quieren incorporarse a este proyecto de edición digital. El portal dispone de 77 revistas y casi 27.000 artículos.

Grabados de la Colección Dioscórides



47.000 grabados digitalizados del fondo antiguo complutense.

Archivo Histórico del PCE



Serie de 800 negativos digitalizados, procedentes del Archivo del Partido Comunista, y referentes a distintos momentos de la Guerra Civil española.

HathiTrust

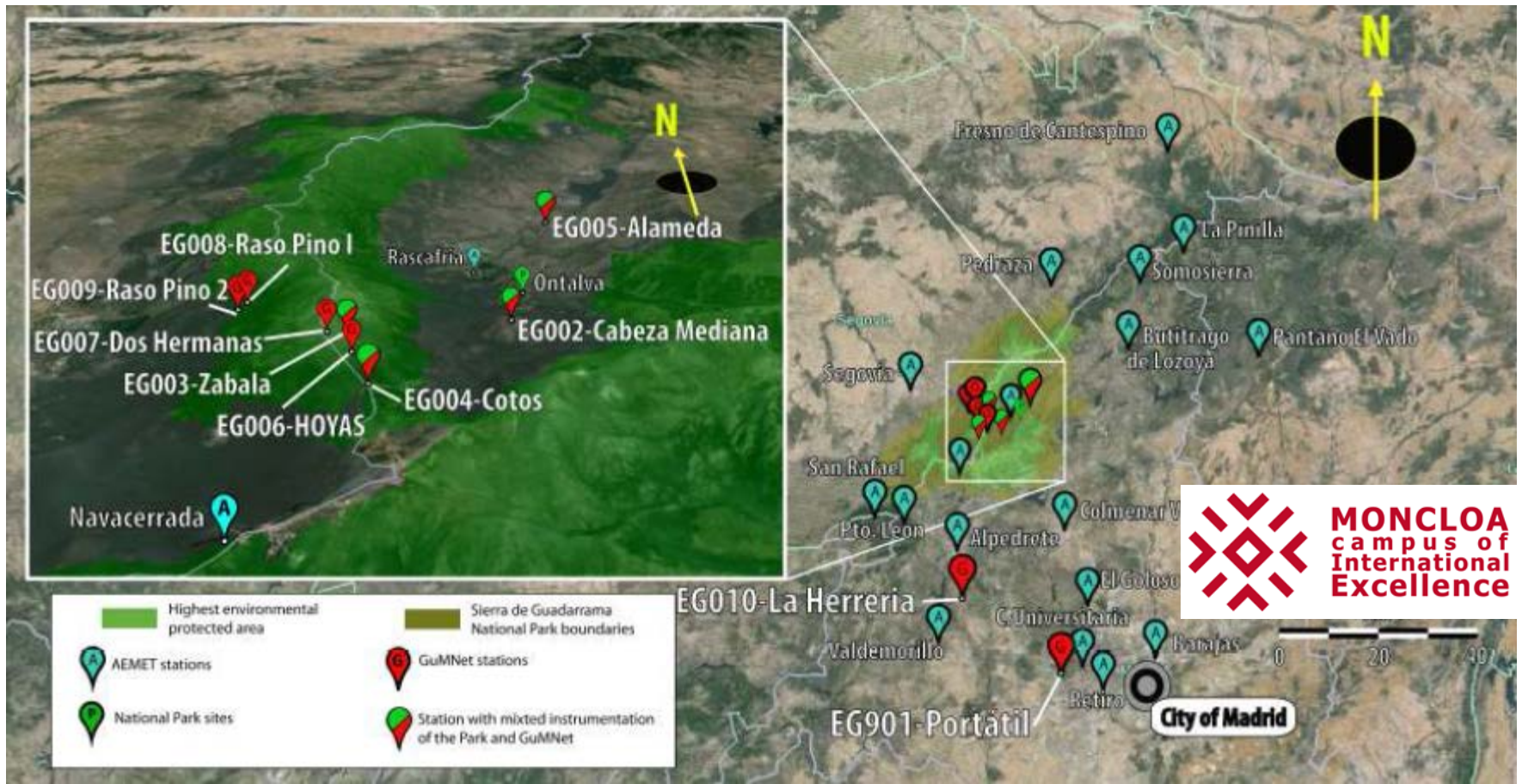


HATHI
TRUST

Colección de 10 millones de ejemplares digitalizados procedentes de los principales bibliotecas científicas y de investigación. La Universidad

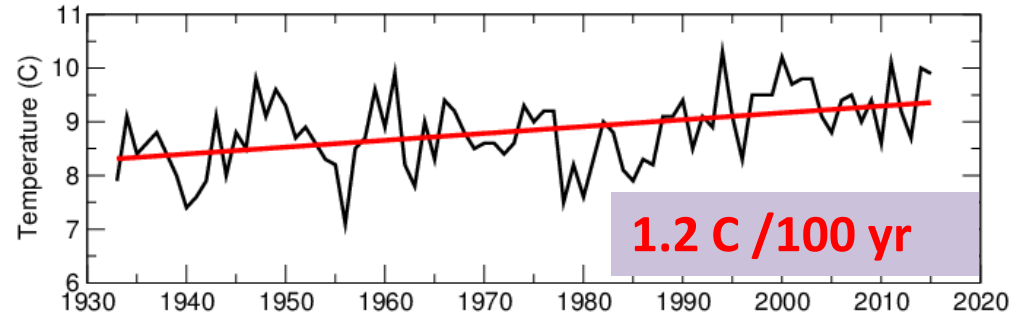
Complutense participa con más de 100.000 documentos libres de derechos de autor.

- **GUADARRAMA MONITORING NETWORK Gumnet**, Fidel González Rouco, CC. Físicas and IGEO <http://www.ucm.es/gumnet/research>
<https://www.youtube.com/watch?v=Lsn0jRza55g>

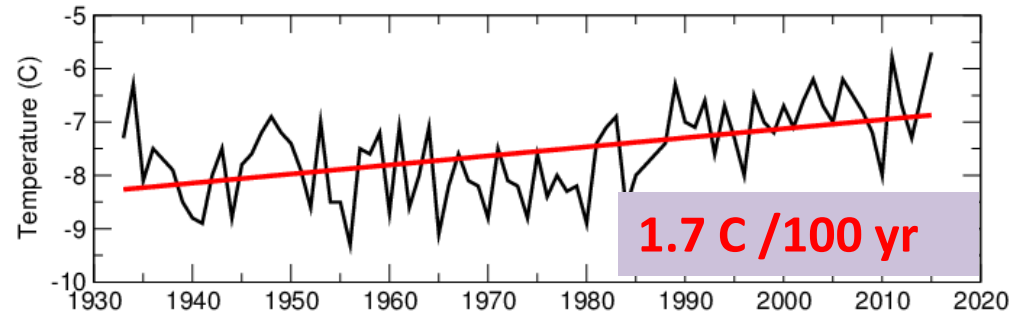


Climatology and Climate Change

Bern, 500 m a.s.l.



Top of Europe: Jungfrauoch 3466 m a.s.l.



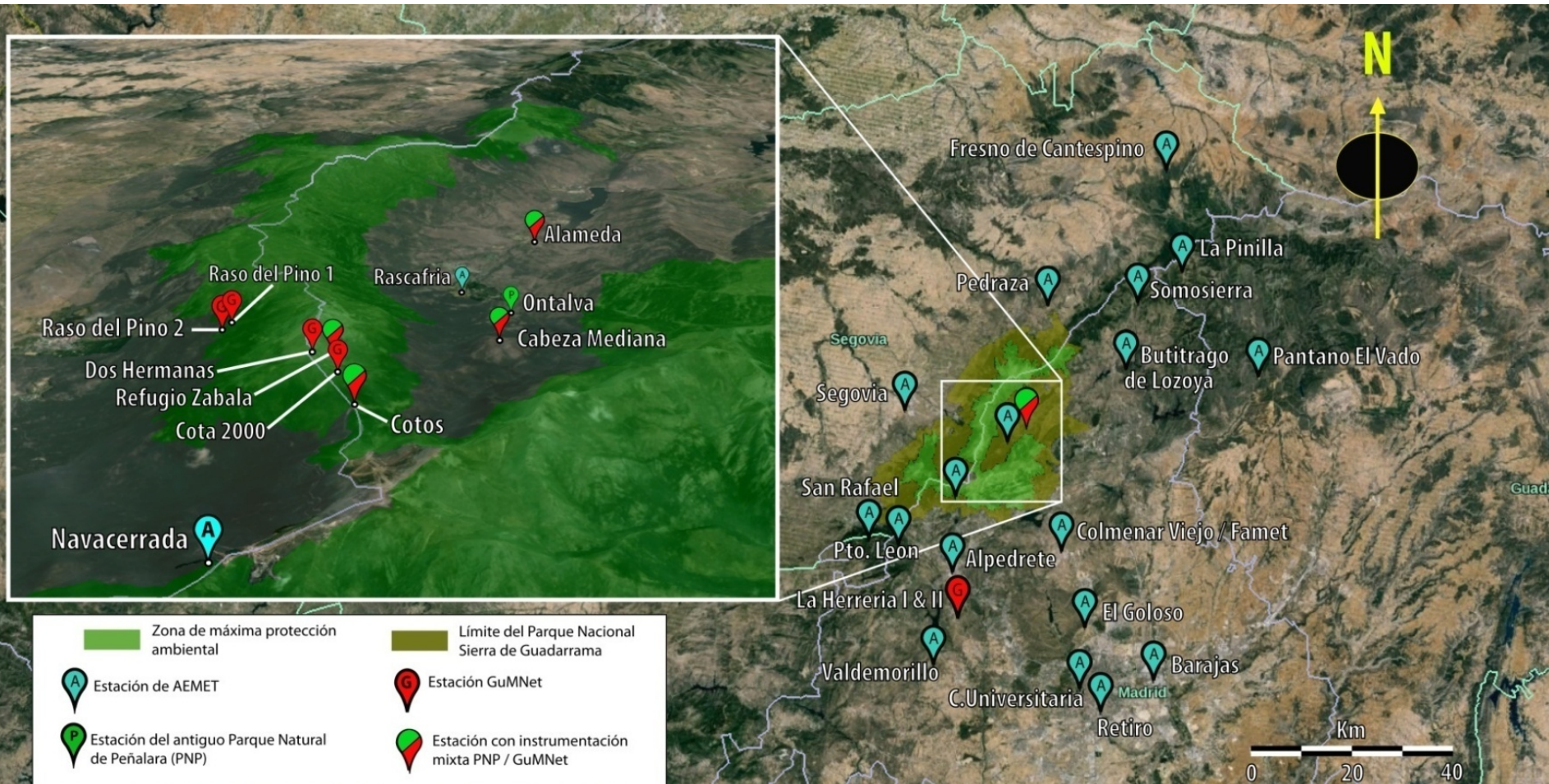
➡ Climate is more sensitive
in mountain regions!



What is GuMNet?

a glimpse at the facility

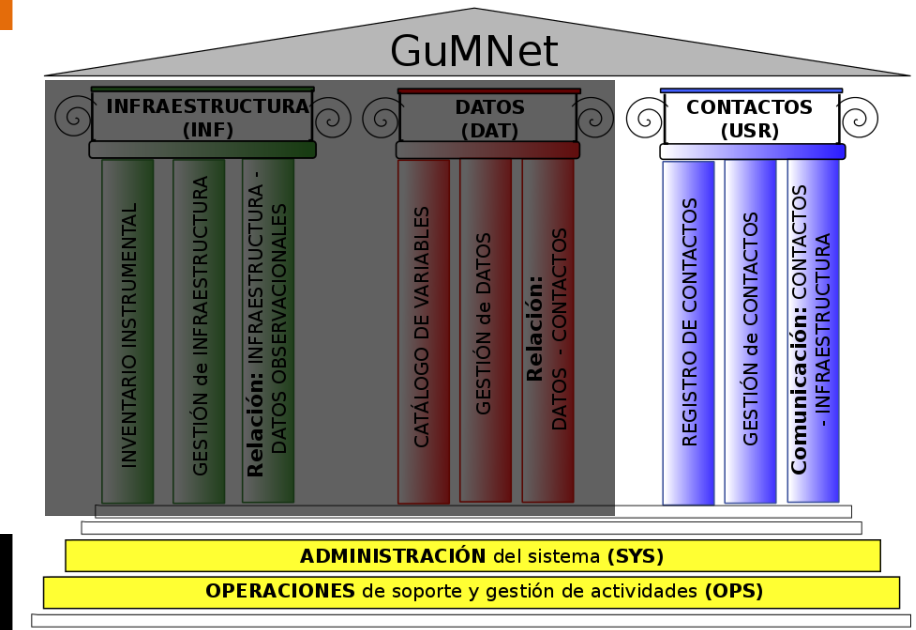
GuMNet is a new infrastructure of atmosphere, surface and subsurface observation
It leans on an existing facility promoted by the Parque Nacional de Peñalara
... updated, enlarged in instrumentations, sites.
It will merge with the meteorological network of AEMET



Miembros: InterMet (L. Durán, I Rodríguez, A. Montesinos), Jacinto (UCM), Edmundo (UCM) y Fidel (UCM)

Objectives:

- Coordinate development of a management tool for data, inventory and users/contacts.
- This tool should facilitate
... an inventory of users and contacts (researchers, students, companies, institutions) and their relations to data and materials.





What does GuMNet need?

A kick off on sustainability

An stable network needs continuity of management and maintenance.

Maintenance: calibrating, replacing, repairing, improving.

Management: procedures to control maintenance of the network
promote networking

**Funding sources for
management &
maintenance :**

-In Campus: UCM-UPM

**-Additional Campus
institutions: CIEMAT,
CSIC/IGEO**

**-Interested institutions:
PNSdG, Patrimonio**

-Other: Canal, Councils

**-Public bids for facility
development**

-Private sponsors

knowledge oriented assessments to improve GuMNet

olicies of data use

provide accountability to institutions

promote use of data for research

promote educational dimension

interaction with management of SdG

promote research projects (public funding)

ate sponsoring

relationships with interested institutions

reach

ra-nationalization

er-nationalization

the sustainability of the network

A satellite with large solar panels is shown in orbit above the Earth's surface. The satellite is positioned diagonally across the frame, with its solar panels extended. The Earth's curvature and blue atmosphere are visible in the background.

José Fernández
CSIC

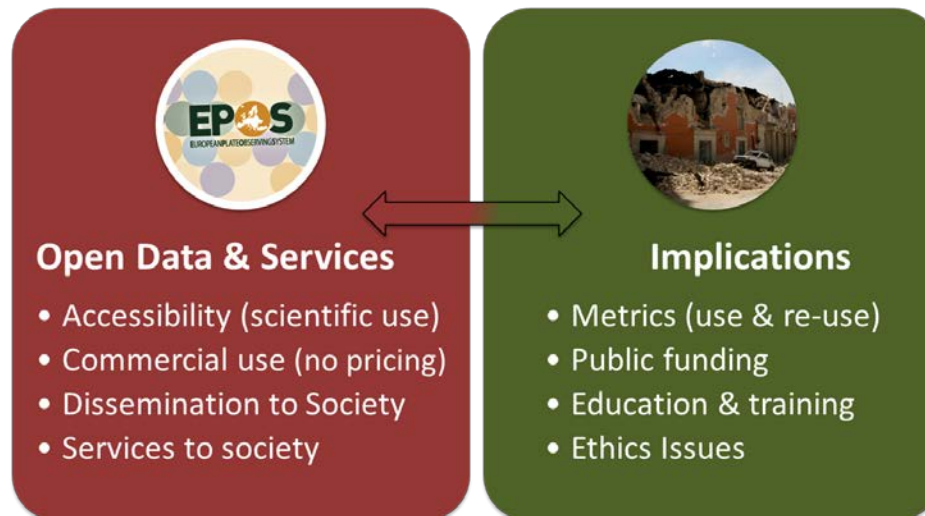
Instituto de Geociencias
(CSIC-UCM)

Multiple activities participated by UCM call for big data analysis

- **European Plate Observing System (EPOS).** José Fernández Torres, Matemáticas/CSIC e IGEO
<http://www.mat.ucm.es/jft/proyectos/european-plate-observing-system-epos/>
<https://www.epos-ip.org/>

The **European Plate Observing System (EPOS)** is the integrated solid Earth Sciences research infrastructure approved by the European Strategy Forum on Research Infrastructures (**ESFRI**) and included in the **ESFRI Roadmap** in December 2008

Open Access for Open Science



EPOS

Services to be implemented

Seismology

Seismic waveforms (ORFEUS)
Seismological products (EMSC)
Hazard & risk products (EFEHR)
Computational seismology

Near fault observatories

NFO multidisciplinary data & products
Borehole data
Virtual laboratory & early warning test beds

GNSS data and products

GNSS primary data & derived products
Processing and visualization tools

Satellite data

SAR interferograms
Integrated satellite products
On-line processing tools

Volcano observations

Multidisciplinary volcanic data & products
Hazard products
TNA to volcano observatories

Anthropogenic hazards

Data for AH episodes
Multi-hazard simulator - multi-risk assessment
AH data visualisation

Geomagnetic observations

Global and regional geomagnetic models
Magnetotelluric data

Geological information and modeling

Geological multi-scale data
Integrated geological maps
Borehole visualization

Multi-scale laboratories

Experimental & analogue data
TNA to experimental & micro-analytical facilities

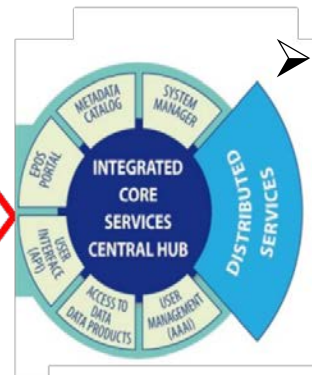
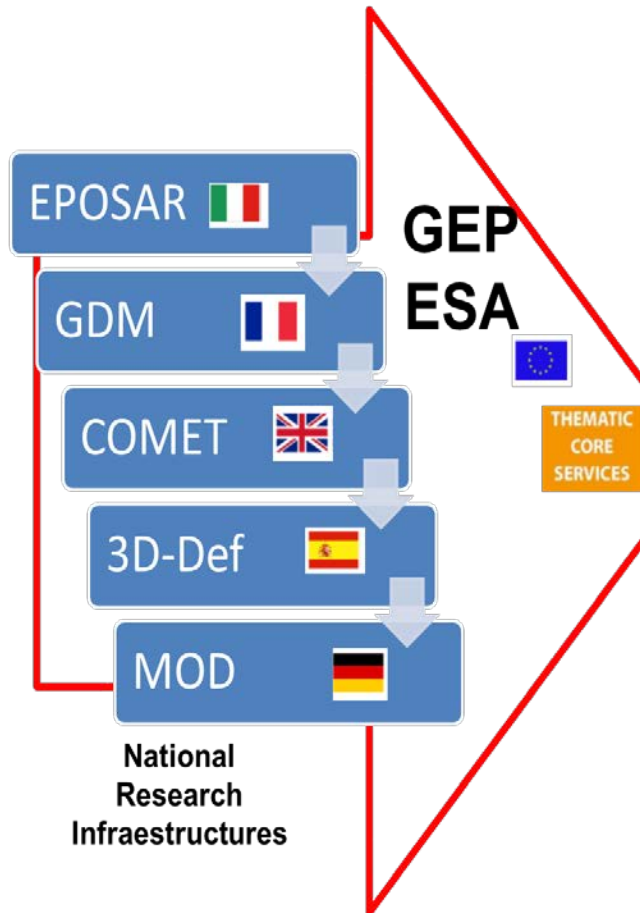
Geo energy test beds for low carbon energy

Geo energy test beds
Access to in-situ GETB experiments

3D-Def

3D-Def service consists of processing tools for determination of

- surface three dimensional, **3D**, **displacements** through combination of EO (both SAR and optical) and GNSS data.



- to couple deformation measurements with **models** and **inversion techniques**, thus estimating physical parameters such as masses anomalies or pressures sources.

EPOS en UCM

- Infraestructura de cálculo (financiada en la última convocatoria de infraestructura) .
107,000 euros.
- Máquina con 128 cores de cálculo (1/4 EOLO), 200 TB (1/2 EOLO) de disco (gran porcentaje SSD), 1 TB RAM (1/2 EOLO), estándar industrial abierto, 5 años de garantía y mantenimiento hardware/software
- Contará con personal técnico de apoyo (IGEO y proyectos EPOS). Se financiará con el pago por uso

EOLO

es el clúster de cálculo científico del Laboratorio de Cambio
Global del CEI MONCLOA

EOLO ha sido financiado por el Ministerio de Economía y Competitividad en el
marco del Programa del Campus de Excelencia Internacional

**(hay que mencionarlo en agradecimientos, en la web de EOLO, en la información
institucional, ...)**

- 1 nodo gestor
- 40 nodos de cálculo (12 x 40 cores, para hasta 24 x 40 = 960 hilos de cálculo con hyperthreading)
- 48 GB RAM por nodo, casi 2 TB RAM total
- Cabina de backup
- Sistema de almacenamiento, 6 nodos con IBRIX
- /home (58 TB)
- /pool (289 TB)
- Optimizado para múltiples accesos concurrentes a disco (multi-hilo, multi-nodo) y comunicaciones inter-hilo. No tanto para acceso serie de trabajos en modo granja. Hay otras máquinas más adecuadas para ello.
- **INSTALADO EN EL CPD DE LA UCM**

¿Qué se espera de EOLO?

El **Laboratorio de Cambio Climático** proporciona recursos (equipos y *expertise*) de simulación y computación avanzada para cambio climático y teledetección a **todos los investigadores del CEI MONCLOA. EOLO es uno de los equipos utilizados para proporcionar dichos servicios**

EOLO también proporciona servicios generales de cálculo científico a **cualquier investigador de la UCM que lo solicite**

Actuaciones realizadas

- 60 usuarios de 30 grupos dados de alta, incluidos huérfanos ABACO (14).
- Usuarios de las Facultades de Ciencias Físicas y Geológicas (26), Matemáticas (9), Informática (6), Económicas (5), Biológicas (4), Psicología (3), Escuela de Óptica (2), Químicas (1), de UCM
- Asignadas cuotas de 12 TB en /home y 50 TB en /pool
- Organizadas cuotas y grupos
- Reuniones individuales del técnico y JMU con los usuarios en el momento de abrirles cuenta y para resolverles dudas, ayudarles a portar códigos, a elegir las colas adecuadas, etc.
- Se da servicio **por igual** a todos los usuarios, sean CEI o UCM. Se han abierto todas las cuentas que los investigadores del CEI o el CPD ha pedido
- **Aproximadamente el 50% de todas las cuentas se han abierto por peticiones llegadas desde el CPD, de usuarios UCM al margen del CEI Moncloa**

Issues in Big Data analysis

- Big Data analysis is complex, and it is developing at exponential speeds
- Big Data can be useful in many different areas, natural sciences, social sciences, health sciences, etc. It will benefit from computing skills as much as from intense knowledge about particular fields of application
- There is the danger of creating a ***technological*** gap among researchers/countries/companies who may or may not be able to benefit from big data analysis

Issues?

social, legal and ethical issues of Big Data

- Modern data analytics have been touted as a "revolution that will transform how we live, work, and think" (Mayer-Schönberger and Cukier 2013)
- Politics: Increased emphasis on predictive analytics as a rationale and justification for *political* action and ***decision*** making
- Ethics: Big Data analysis bears the potential of ***profiling*** groups and even individuals

Big Personal Data?

- Big Data ***may involve personal data*** relating to an individual. A name, photo, an email address, bank details, ***posts on social networking websites***, web searches, medical information...
- Big Data and ensuring personal data protection becomes more challenging as information is shared widely around the world. Information regarding individual's health, location, electricity use, online activity and so forth can be publicised, raising concerns about profiling, discrimination, exclusion and loss of control

http://ec.europa.eu/justice/data-protection/files/data-protection-big-data_factsheet_web_en.pdf

New requirements and challenges in the ***governance*** of Big Data

- The general complexity and speed of development. Regulations become obsolete even before being passed
- The profound lack of technical skills and competencies in the regulatory bodies
- Insufficient access to data in both the government sector and academia

Big Data and Trust

- 81% of Europeans feel that they do not have complete control over their personal data online - Eurobarometer, 2015.
- A large majority of Europeans (69%) would like to give their explicit approval before the collection and processing of their personal data – Eurobarometer, 2015.
- Only 24% of Europeans have trust in online businesses such as search engines, social networking sites and e-mail services. – Eurobarometer, 2015.

Summary

- Big Data analysis bears the potential to transform the way we obtain information and derive new knowledge
- As important as the technical tools (software, hardware) is a deep understanding of the field of application
- Big Data analysis also raises ethical, social, regulatory concerns
- The extraordinary chair HPE-UCM would take advantage of the vast and unique capabilities of UCM to foster Big Data applications and studies

- ¿Máster en applied data science? ¿big data y sociedad? ¿Conjunto con UPM?
¿Internacional? ¿coordinación con otros másters para TFMs, etc?
- Esponsorización, patrocinios, financiación de empresas
- ¿Relación con los expertos técnicos?
- ¿Servicios de computación científica?
¿asesoramiento para el CLOUD? ¿APIS?
- ¿Hardware, software o asesoramiento?
- <https://rc.fas.harvard.edu/about/people/#staff>

