Structural Biology in the Clouds: The HADDOCK/WeNMR-EOSC Ecosystem

Brian Jiménez-García

Utrecht University - Bonvin Lab

The Netherlands

b.jimenezgarcia@uu.nl

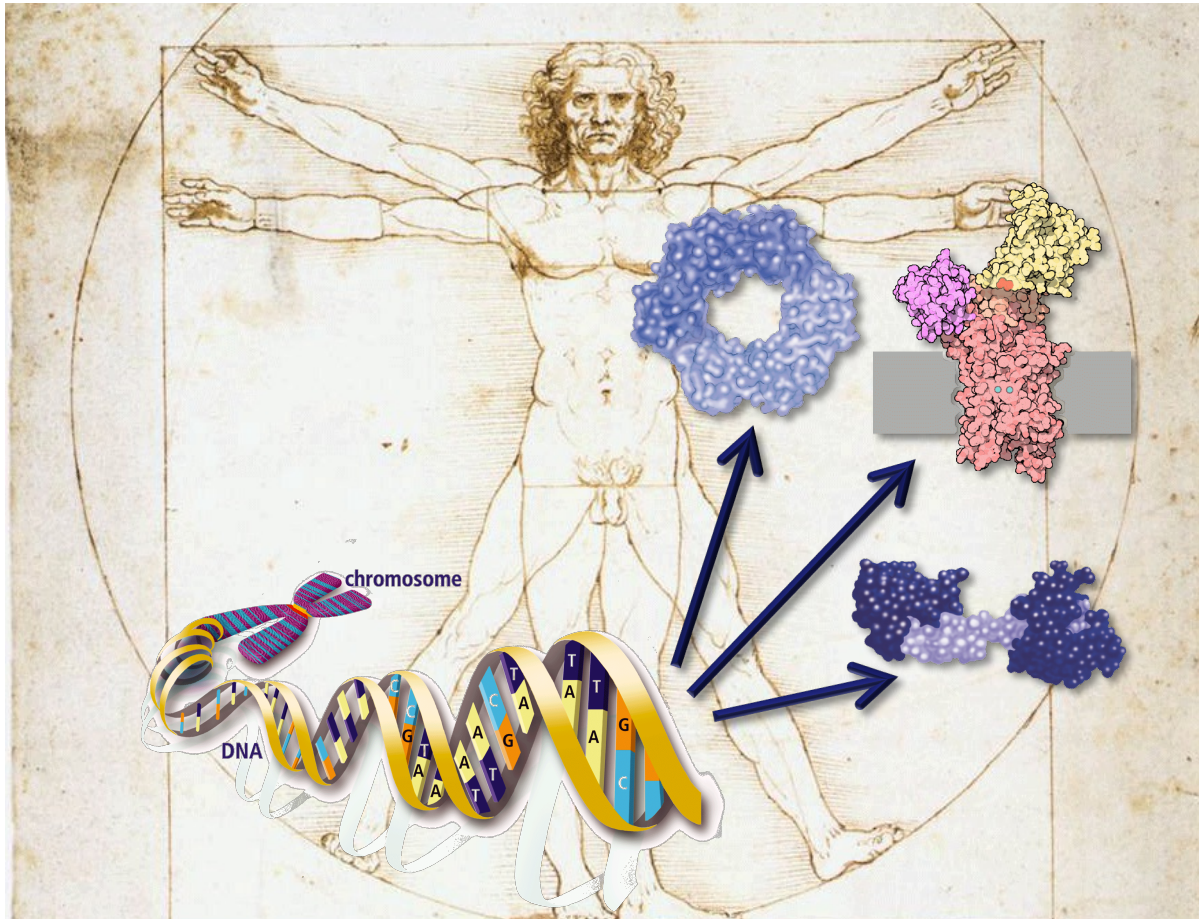
@br_jimenez

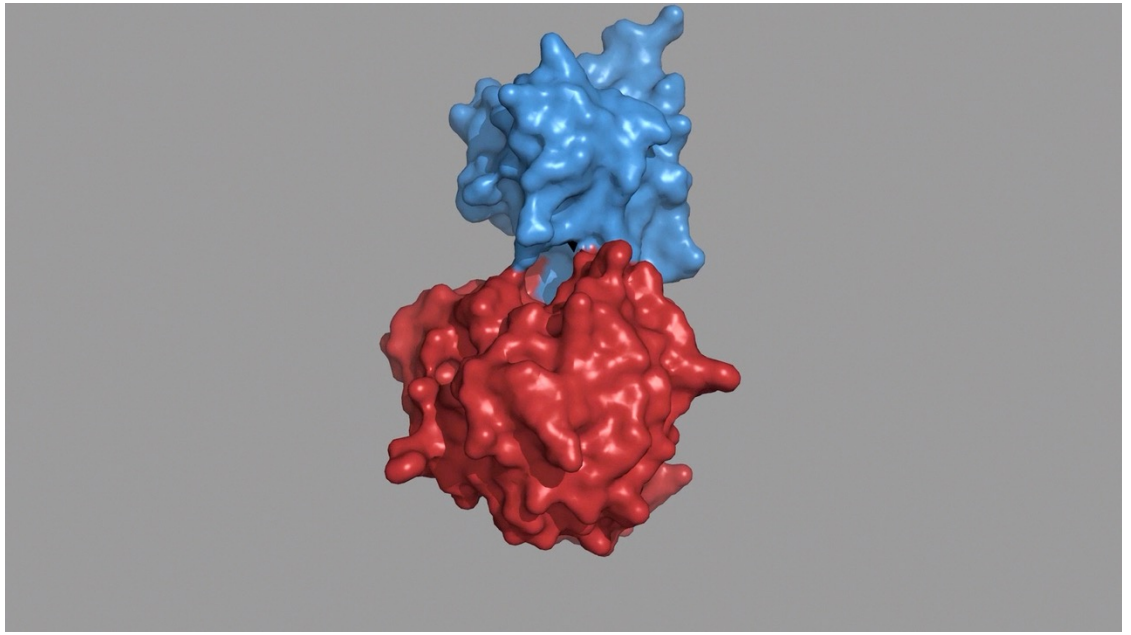
Outline

- Domain and background of the scientific community
- Community and Impact
- Challenges and solutions
- Supporting COVID-19 research
- Future perspectives

Domain: life sciences / structural biology

Understanding life at the molecular level





Main use cases:

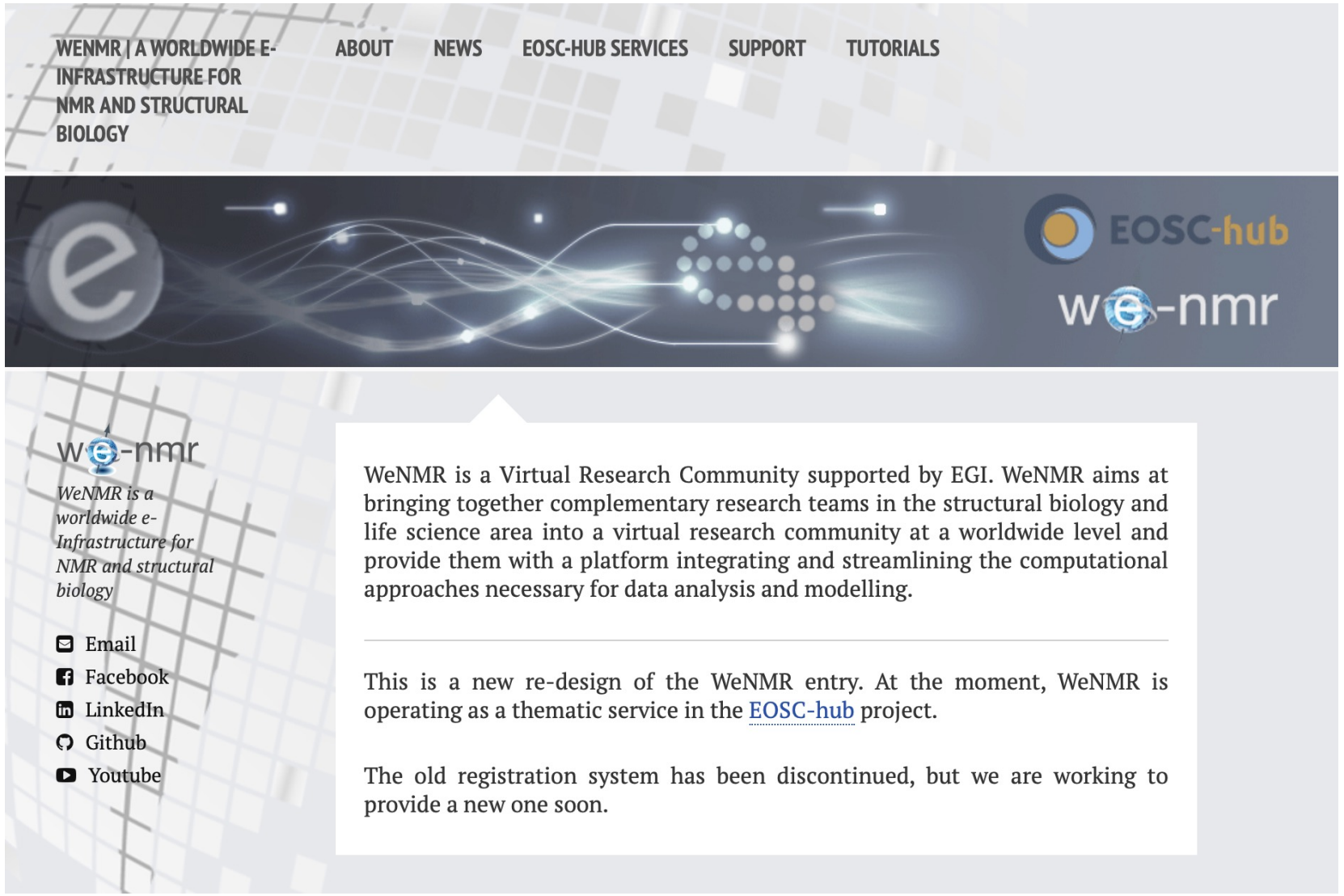
- Disease-causing mutations
- Engineer better molecules for material, health or food applications
- Obtain a starting point for drug design to combat disease

EOSC-hub

Virtual Research Community

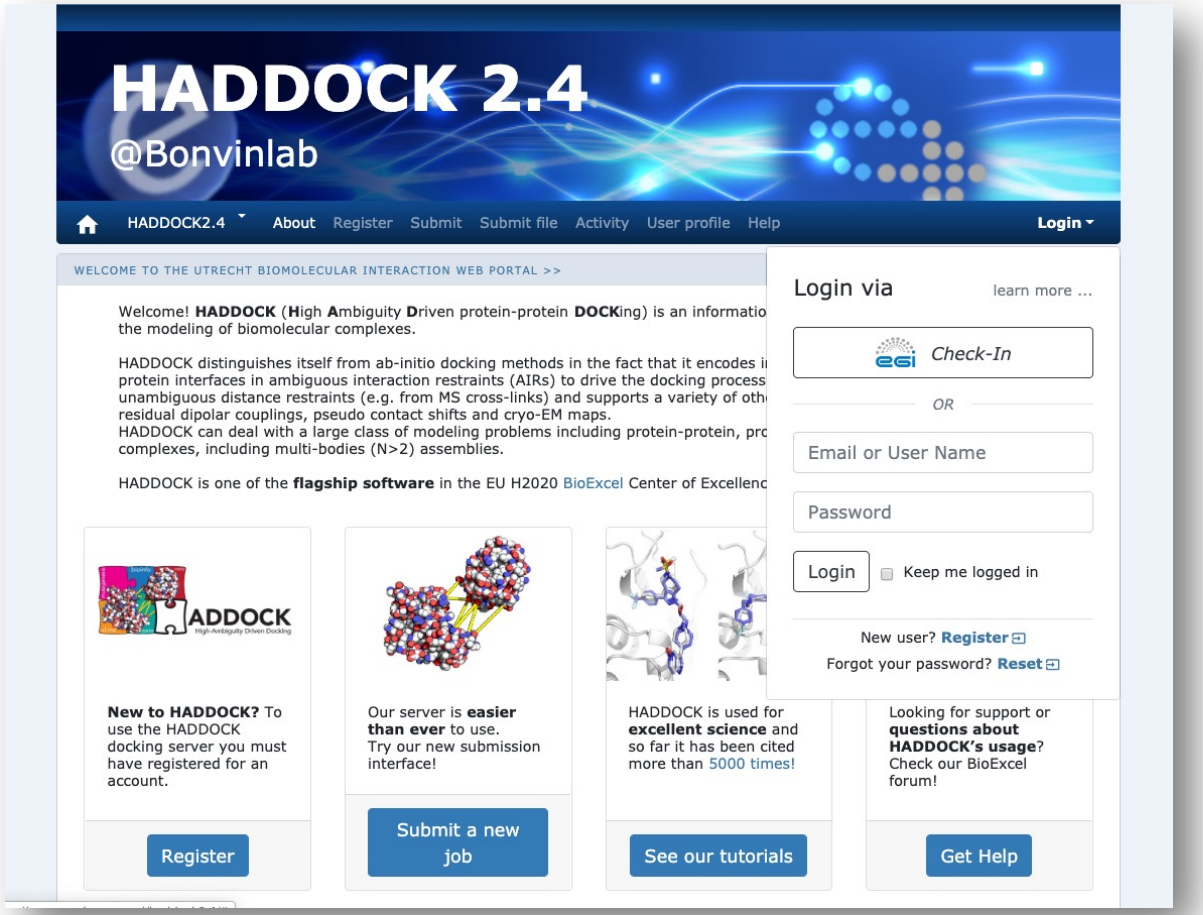
>11 years of serving this research community

Made possible via HTC resources provided by FP7 and H2020 e-Infrastructure projects over the years

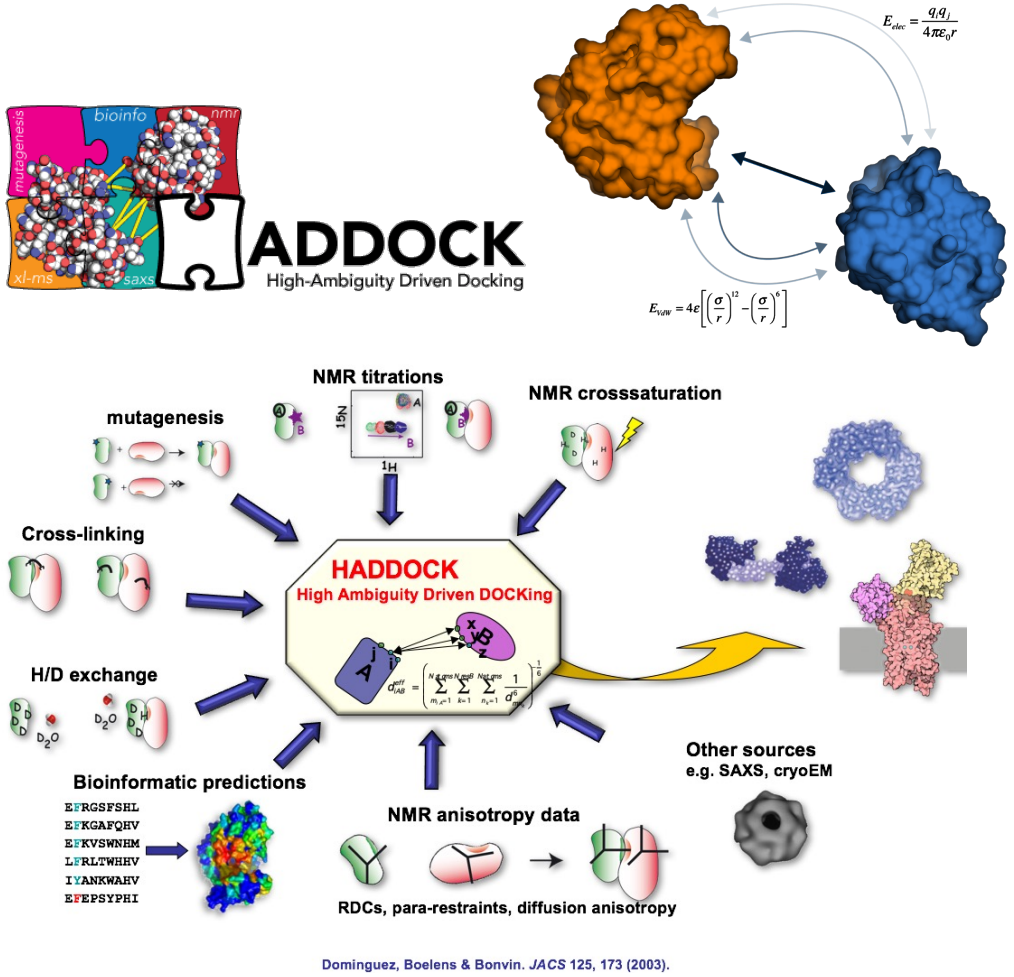

WENMR | A WORLDWIDE E-INFRASTRUCTURE FOR NMR AND STRUCTURAL BIOLOGY
ABOUT
NEWS
EOSC-HUB SERVICES
SUPPORT
TUTORIALS
EOSC-hub
we-nmr
we-nmr
WeNMR is a worldwide e-Infrastructure for NMR and structural biology
Email
Facebook
LinkedIn
Github
Youtube
WeNMR is a Virtual Research Community supported by EGI. WeNMR aims at bringing together complementary research teams in the structural biology and life science area into a virtual research community at a worldwide level and provide them with a platform integrating and streamlining the computational approaches necessary for data analysis and modelling.
This is a new re-design of the WeNMR entry. At the moment, WeNMR is operating as a thematic service in the EOSC-hub project.
The old registration system has been discontinued, but we are working to provide a new one soon.

www.wenmr.eu

HADDOCK - Integrative Modelling Platform



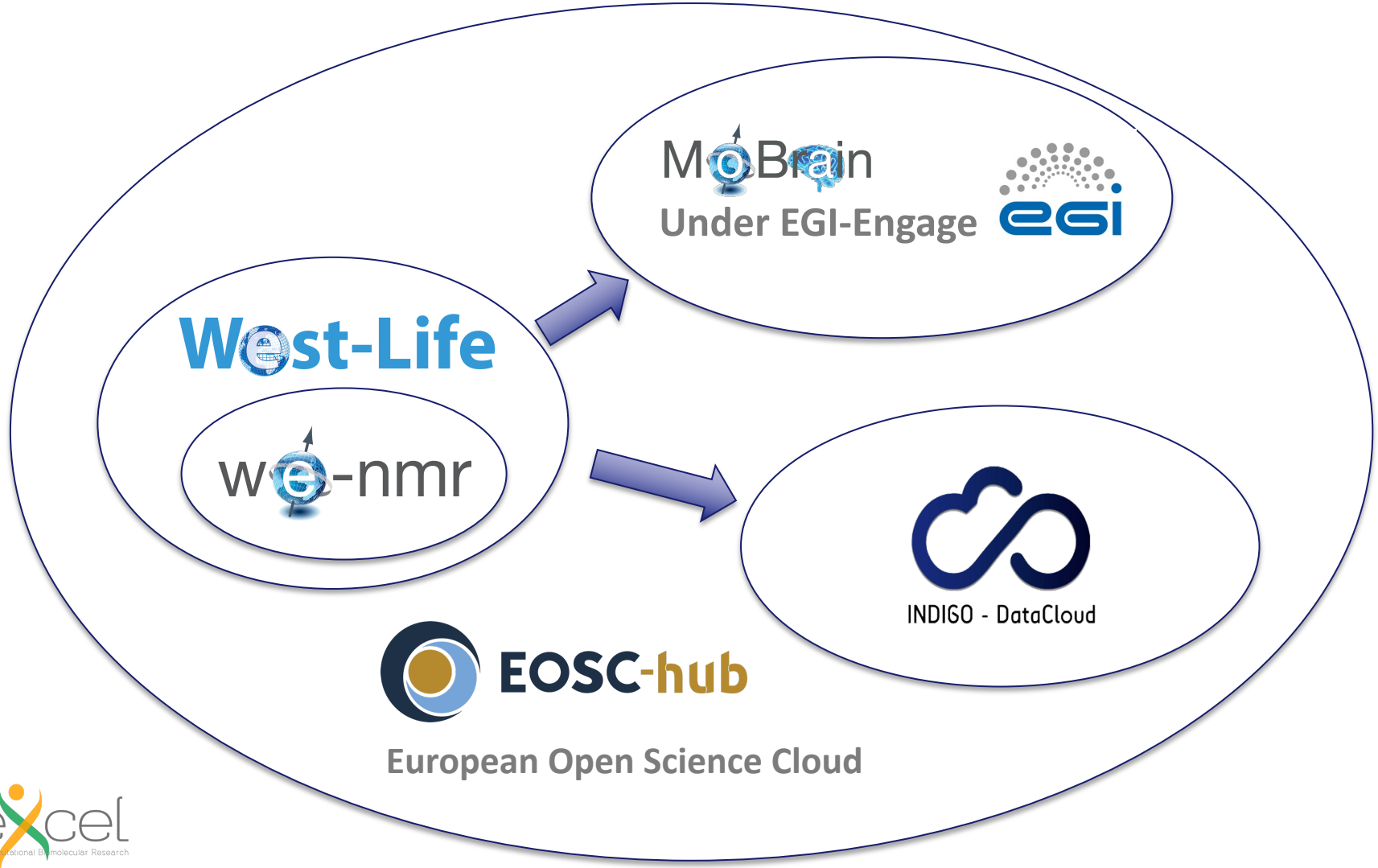
https://wenmr.science.uu.nl/haddock2.4/



De Vries *et al.* Nature Prot. 2010

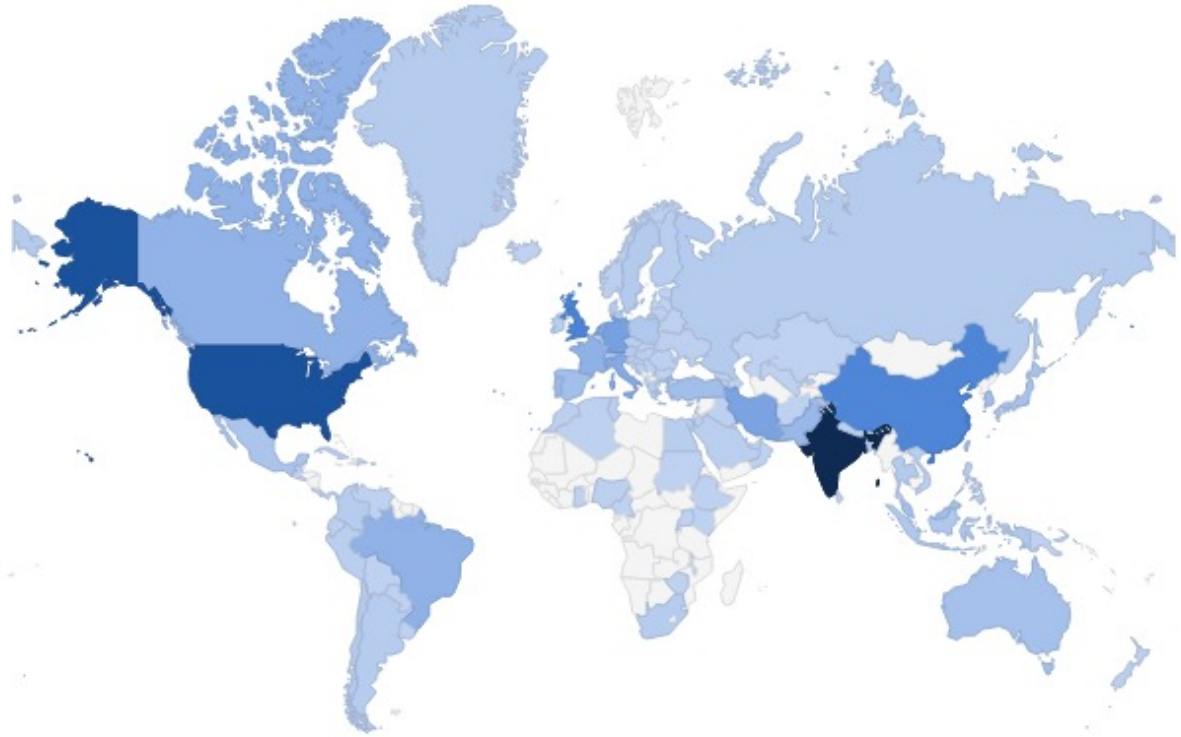
Van Zundert *et al.* J.Mol.Biol. 2016

e-Infrastructure landscape over the years



Scientific community background

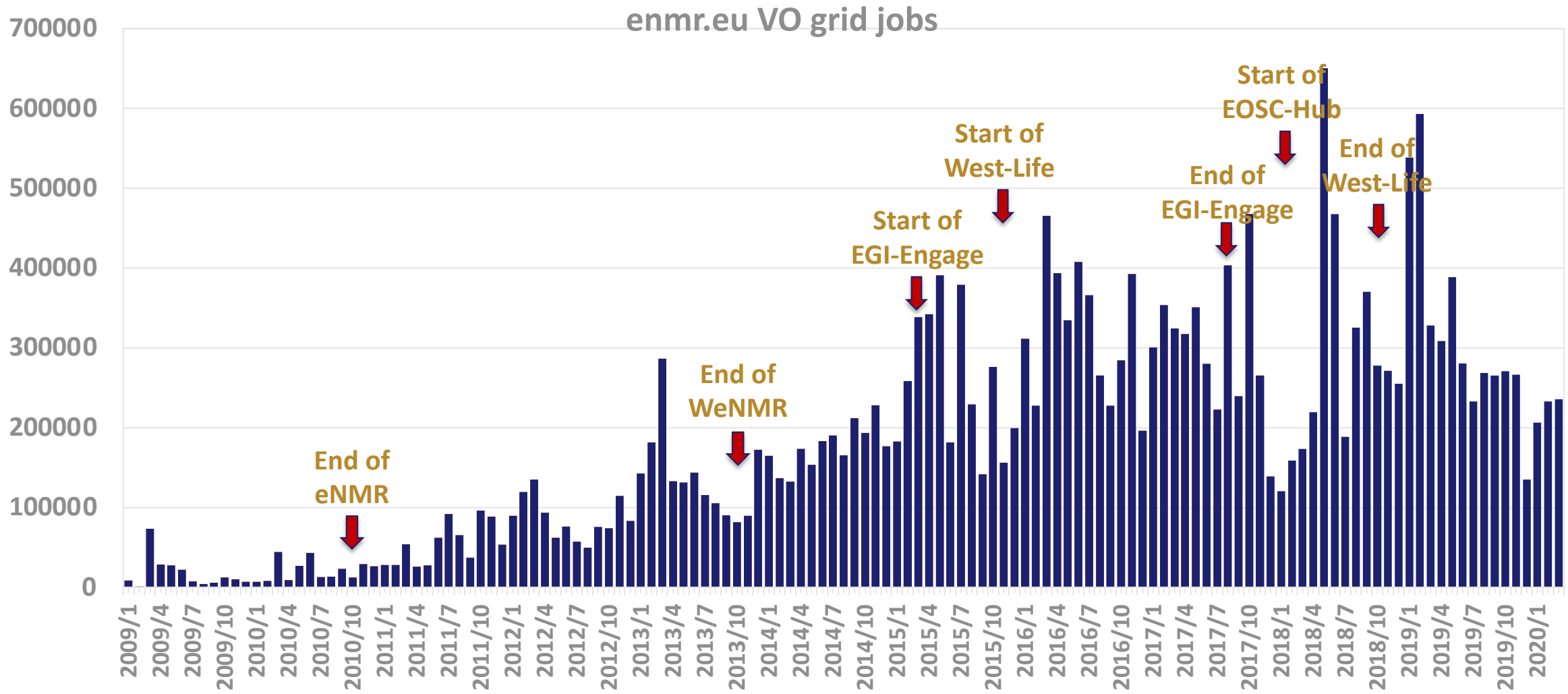
- ~ 22,000 registered users
- > 120 countries reached
- Both researchers and students
- Sustained growth of the community



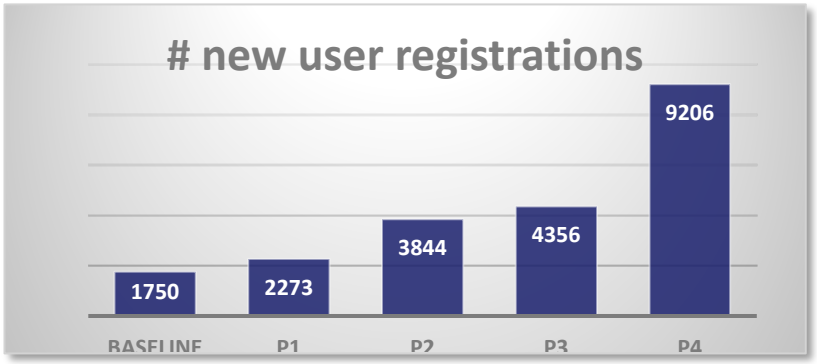
	Country	All_Users	HADDOCK	DISVIS	POWERFIT	SPOTON
1	Total Users	21,924	21,126	3,073	2,445	2,803
2	EU Users	4,589	4,345	652	456	510
3	India	4,549	4,445	605	532	601
4	United States	3,174	3,054	421	298	371
5	United Kingdom	1,327	1,272	148	124	127

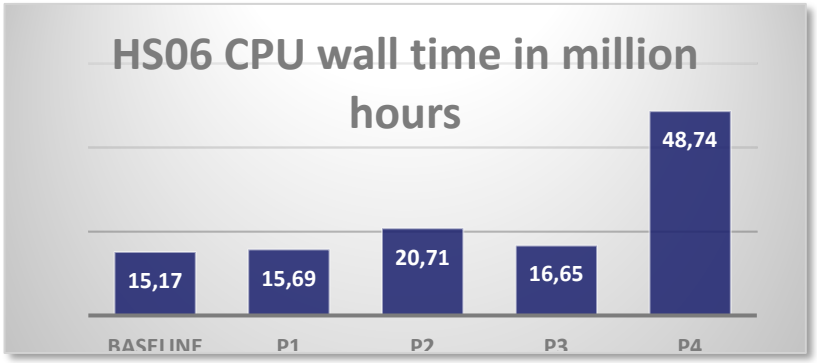
https://wenmr.science.uu.nl/user_map

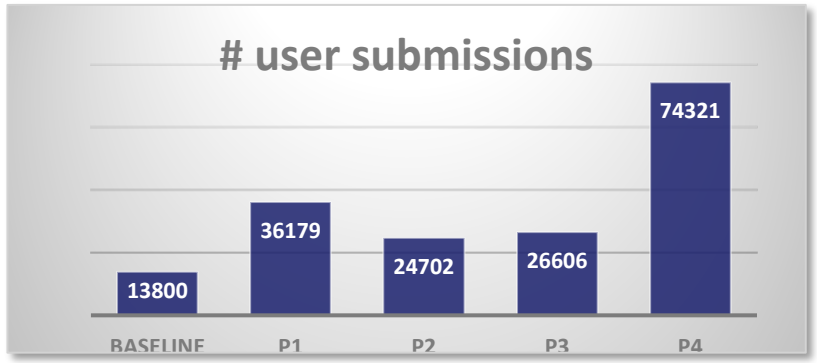
Operational since >10 years



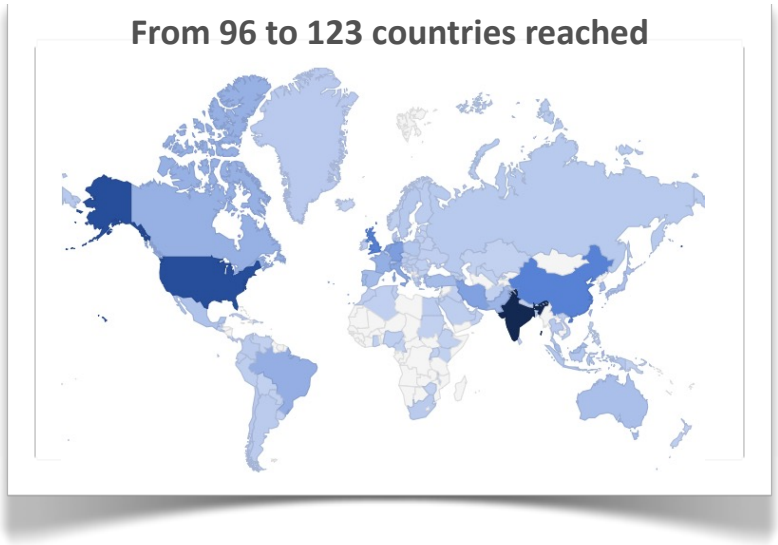
WeNMR – Impact

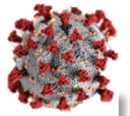


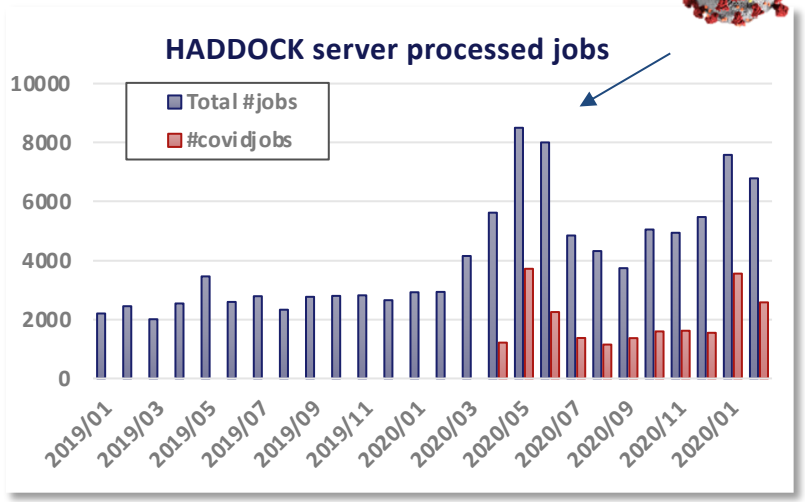


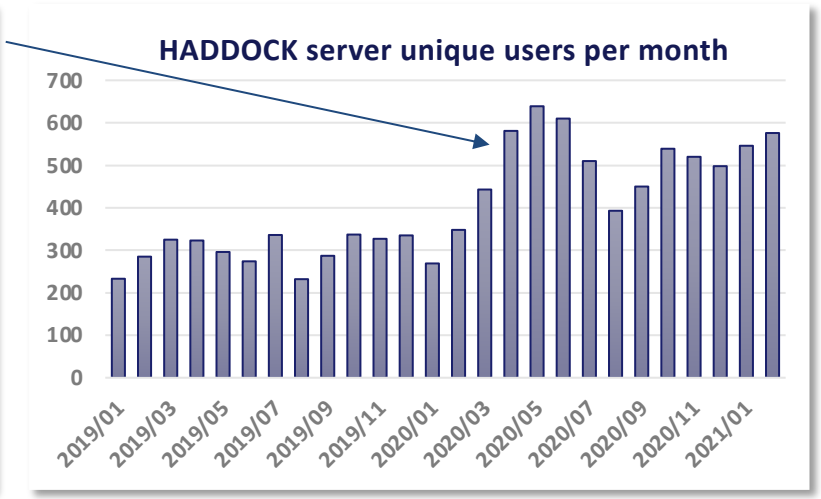


- Steady increase in number of users over the duration of the project
- Worldwide impact
- Relevance of the WeNMR services for COVID-related research made clear by the huge increase in usage over the last period
- ~80% of submissions run on EOSC HTC resources







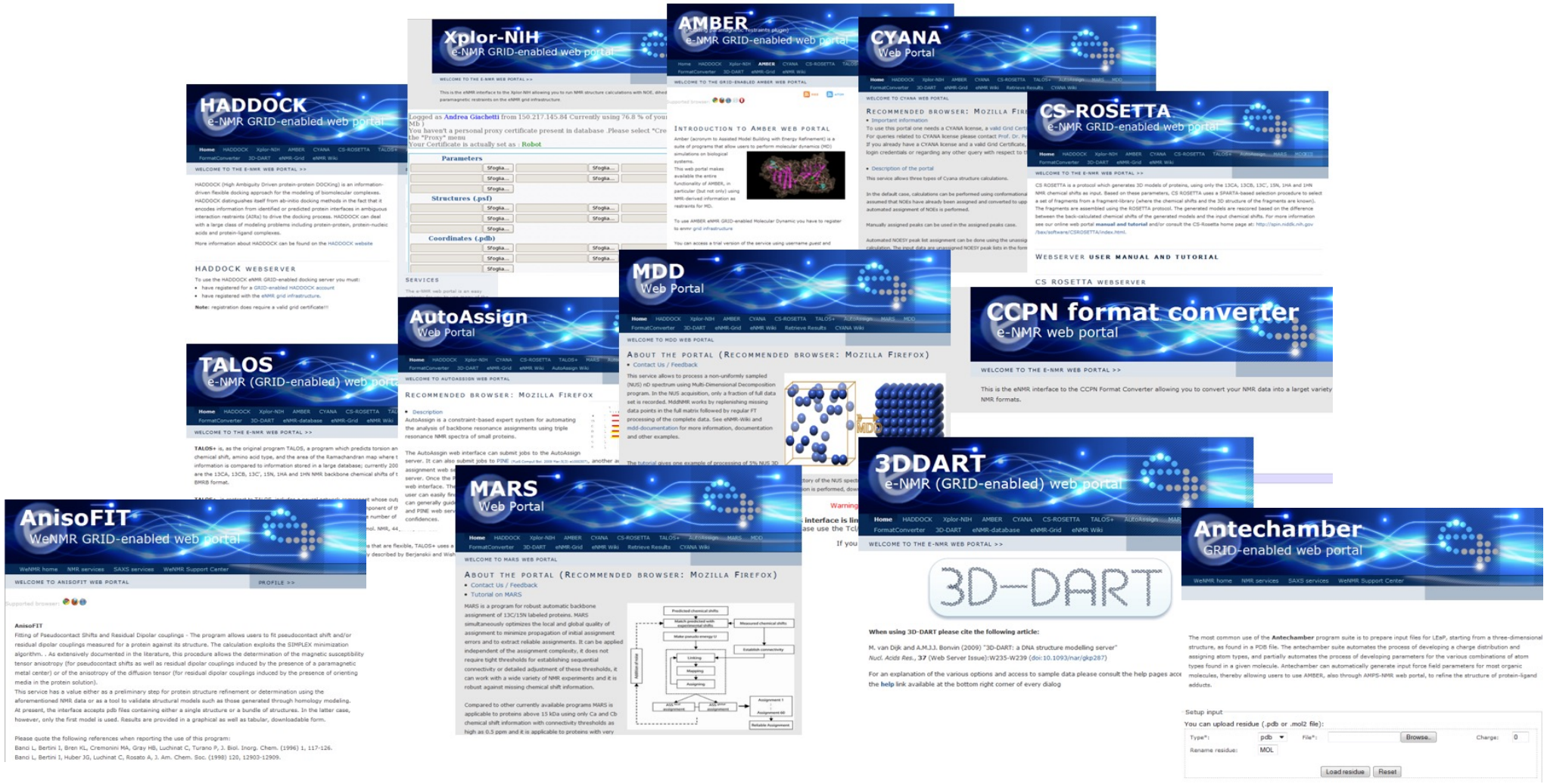



Challenges and solutions

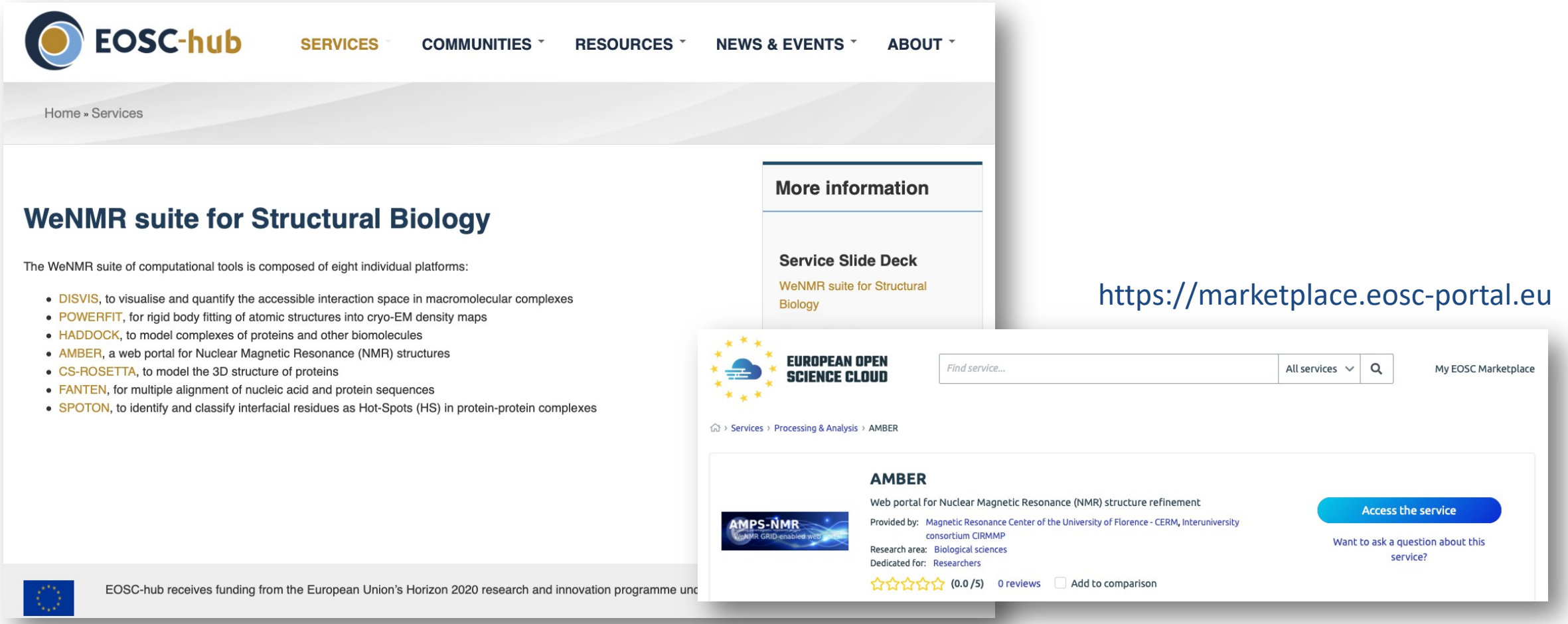
Challenges & e-Solutions

- **Attract users!**
  - **Offer them top of the line eScience solutions for their research ... which means top of the line software**

The WeNMR services portfolio over the years



The WeNMR services @ EOSC



https://marketplace.eosc-portal.eu

https://www.eosc-hub.eu

Challenges & e-Solutions

- **Attract users!**
  - **Offer them top of the line eScience solutions for their research ... which means top of the line software**
  - **Provide them training, tutorials and support**

The WeNMR virtual research community

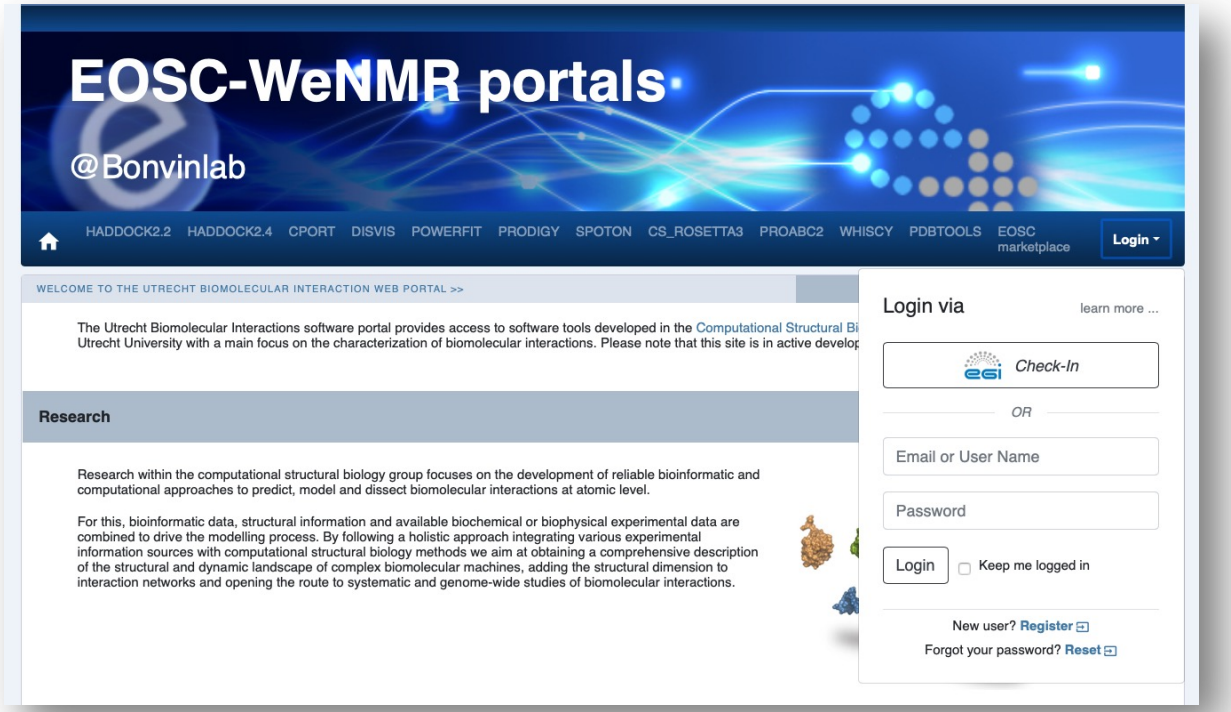
- Support center
- Tutorials
- YouTube channel
- Many workshops …

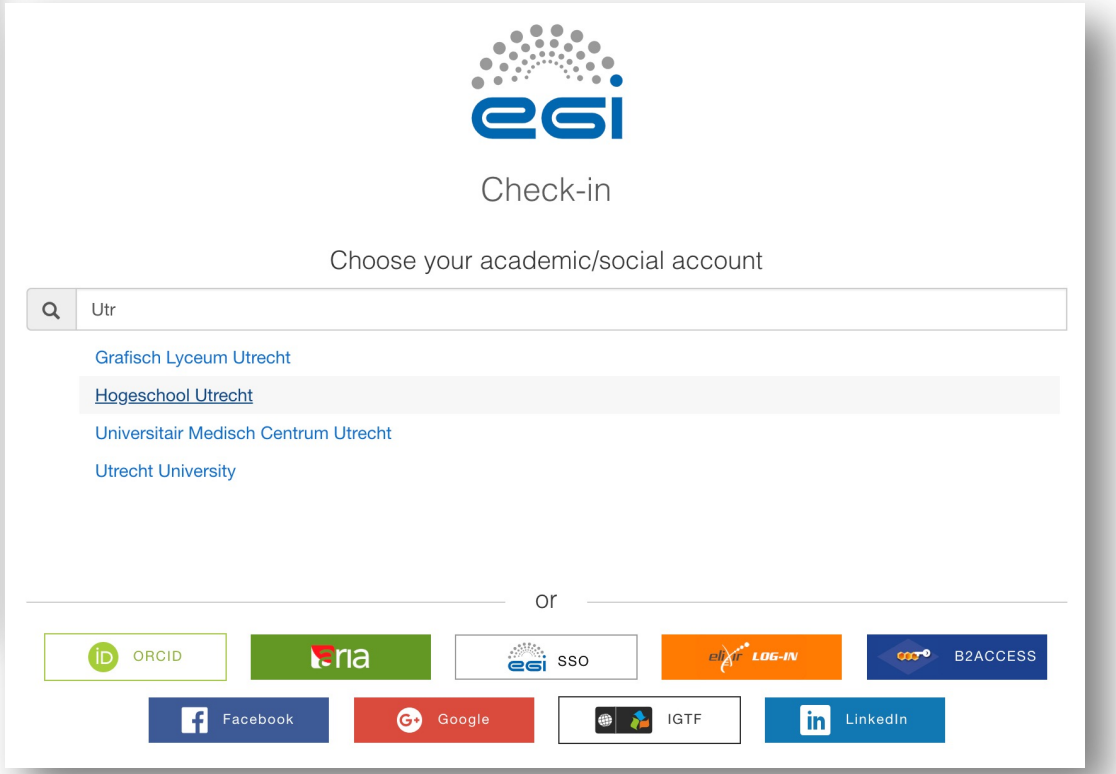
Challenges & e-Solutions

- **Attract users!**
  - **Offer them top of the line eScience solutions for their research ... which means top of the line software**
  - **Provide them training, tutorials and support**
  - **Make their life easier**

Single Sign On

- **SSO Integration with European Open Science Cloud EGI CheckIn**





User Satisfaction

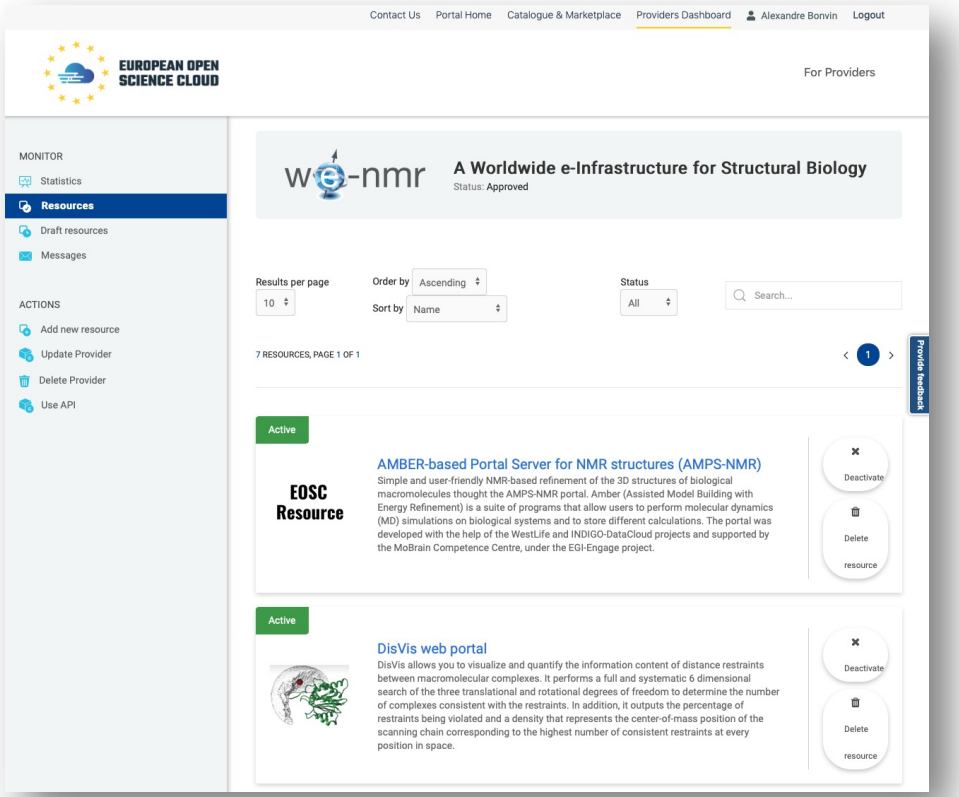
- Keep monitoring user satisfaction and get input from users

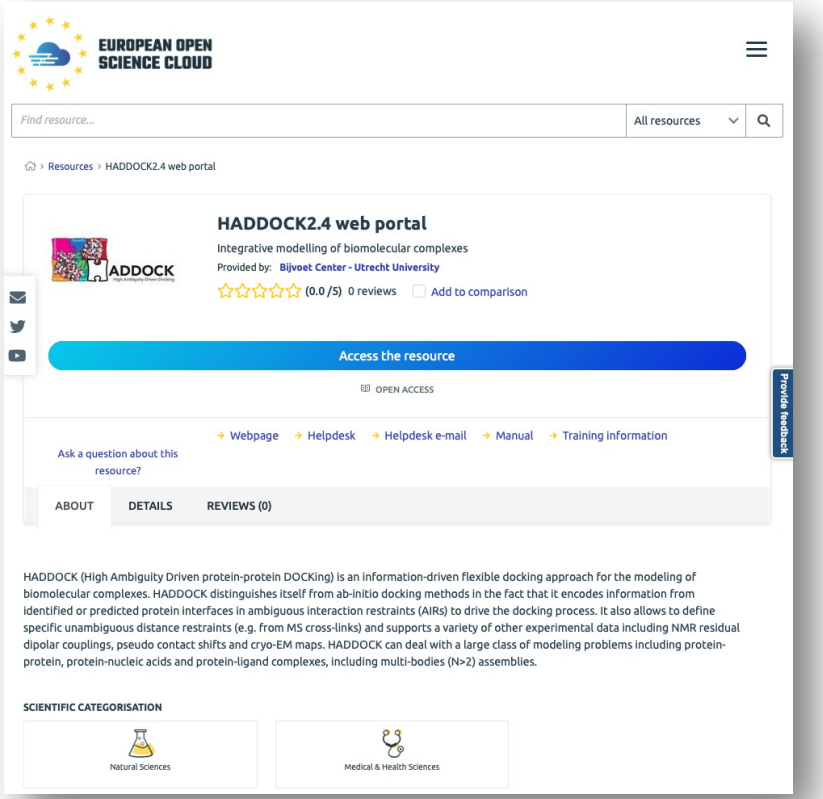
- User satisfaction ratings 

  - CS-ROSETTA3: 5.0 (from 5 respondent)
  - DISVIS: 4.8 (from 38 respondents)
  - HADDOCK: 4.9 (from 3263 respondents)
  - HADDOCK2.4: 4.9 (from 103 respondents)
  - POWERFIT: 4.8 (from 14 respondents)
  - PRODIGY: 4.7 (from 792 respondents)
  - SPOTON: 4.7 (from 74 respondents)

EOSC marketplace

Integrated in EOSC marketplace and in full operation from day 1





Challenges & e-Solutions

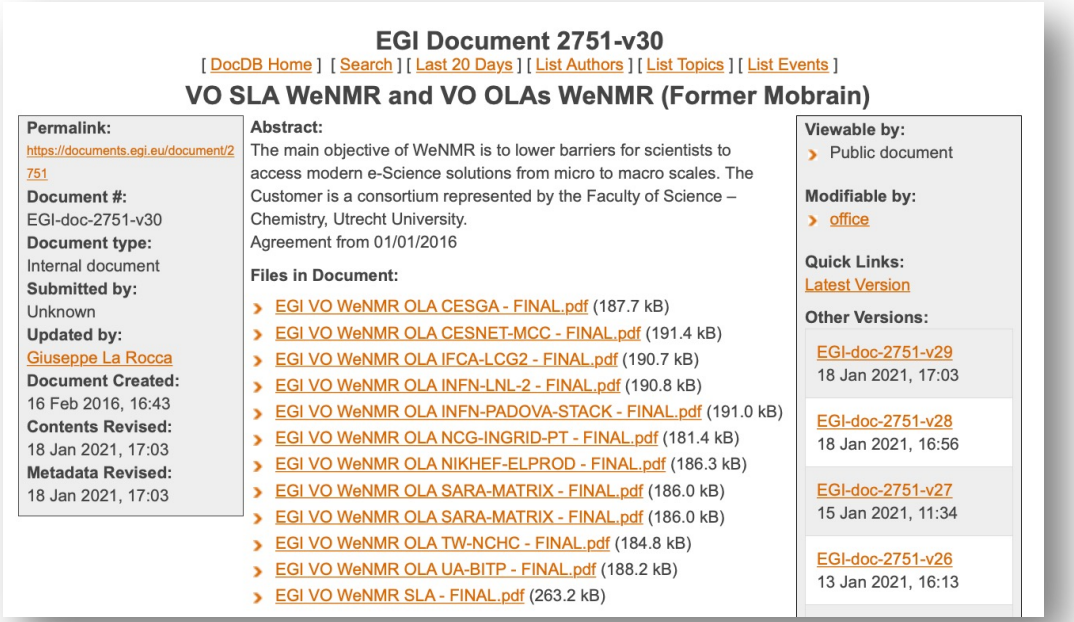
- **Attract users!**
- **Access to e-Infrastructure**

HTC and cloud resources

- The WeNMR services have been in production since >11 years under various projects (eNMR, WeNMR, EGI-Engage, West-Life, EOSC-Hub)

- Access to resources formalized through a SLA agreement valid until 6/2023
  - **53 million CPU hours** (opportunistic access)
  - **412 cloud CPU cores**
  - **59 TB storage**

"The FP7 WeNMR (project# 261572), H2020 West-Life (project# 675858), the EOSC-hub (project# 777536) and the EGI-ACE (project# 101017567) European e-Infrastructure projects are acknowledged for the use of their web portals, which make use of the EGI infrastructure with the dedicated support of CESNET-MCC, INFN-PADOVA-STACK, INFN-LNL-2, NCG-INGRID-PT, TW-NCHC, CESGA, IFCA-LCG2, UA-BITP, SURFsara and NIKHEF, and the additional support of the national GRID Initiatives of Belgium, France, Italy, Germany, the Netherlands, Poland, Portugal, Spain, UK, Taiwan and the US Open Science Grid."



EGI Document 2751-v30

[DocDB Home] [Search] [Last 20 Days] [List Authors] [List Topics] [List Events]

VO SLA WeNMR and VO OLAs WeNMR (Former Mobrain)

Permalink: https://documents.egi.eu/document/2751
Document #: EGI-doc-2751-v30
Document type: Internal document
Submitted by: Unknown
Updated by: Giuseppe La Rocca
Document Created: 16 Feb 2016, 16:43
Contents Revised: 18 Jan 2021, 17:03
Metadata Revised: 18 Jan 2021, 17:03

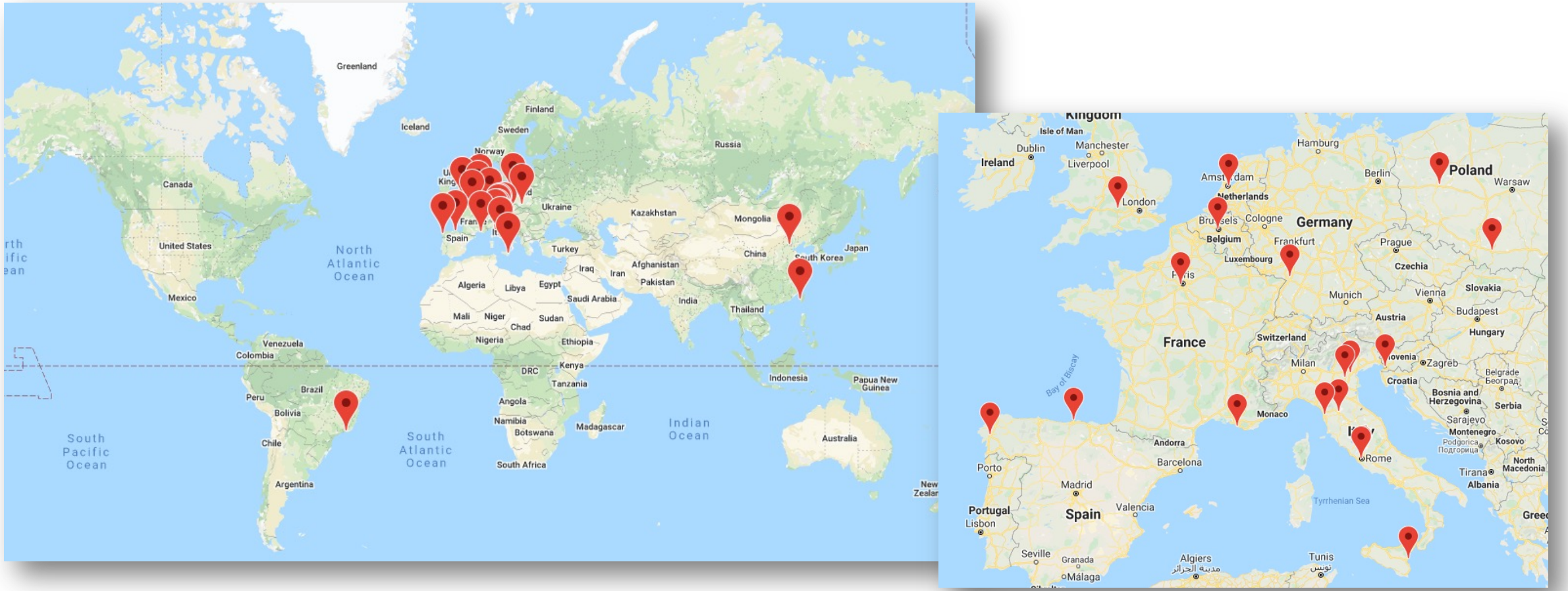
Abstract:
The main objective of WeNMR is to lower barriers for scientists to access modern e-Science solutions from micro to macro scales. The Customer is a consortium represented by the Faculty of Science – Chemistry, Utrecht University.
Agreement from 01/01/2016

Files in Document:
- EGI VO WeNMR OLA CESGA - FINAL.pdf (187.7 kB)
- EGI VO WeNMR OLA CESNET-MCC - FINAL.pdf (191.4 kB)
- EGI VO WeNMR OLA IFCA-LCG2 - FINAL.pdf (190.7 kB)
- EGI VO WeNMR OLA INFN-LNL-2 - FINAL.pdf (190.8 kB)
- EGI VO WeNMR OLA INFN-PADOVA-STACK - FINAL.pdf (191.0 kB)
- EGI VO WeNMR OLA NCG-INGRID-PT - FINAL.pdf (181.4 kB)
- EGI VO WeNMR OLA NIKHEF-ELPROD - FINAL.pdf (186.3 kB)
- EGI VO WeNMR OLA SARA-MATRIX - FINAL.pdf (186.0 kB)
- EGI VO WeNMR OLA SARA-MATRIX - FINAL.pdf (186.0 kB)
- EGI VO WeNMR OLA TW-NCHC - FINAL.pdf (184.8 kB)
- EGI VO WeNMR OLA UA-BITP - FINAL.pdf (188.2 kB)
- EGI VO WeNMR SLA - FINAL.pdf (263.2 kB)

Viewable by: Public document
Modifiable by: office
Quick Links: Latest Version
Other Versions:
- EGI-doc-2751-v29 18 Jan 2021, 17:03
- EGI-doc-2751-v28 18 Jan 2021, 16:56
- EGI-doc-2751-v27 15 Jan 2021, 11:34
- EGI-doc-2751-v26 13 Jan 2021, 16:13

https://documents.egi.eu/document/2751

HTC resources



World-wide (excl. OSG) April 2020:
> 500’000 CPU cores from 22 sites

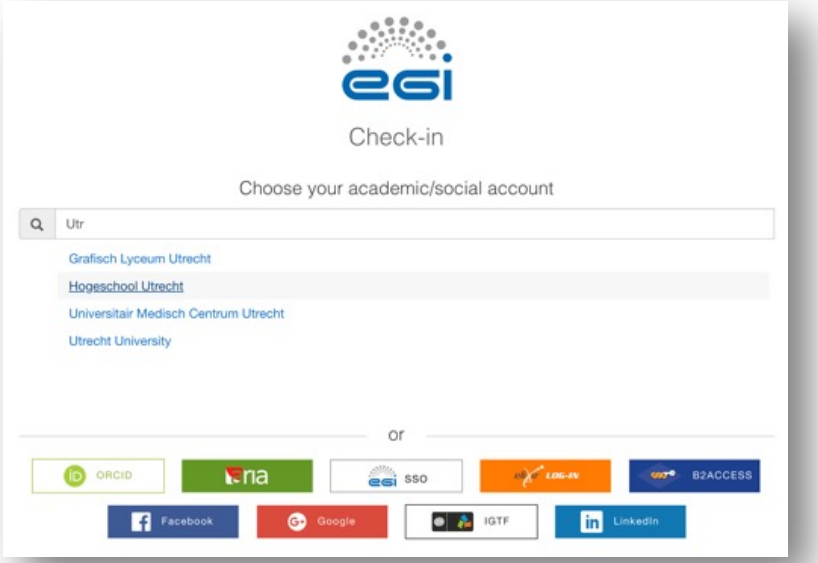
Challenges & e-Solutions

- Attract users!
- Access to e-Infrastructure
- **Develop software and maintain and operate a complex infrastructure**

High-level architecture overview

The **WeNMR services** are based on:

- A web page front end to the service exposed to the user
- Back end consisting of a variety of software and scripts
- Make use of the **DIRAC4EGI/EGI workload** service for distributing compute jobs and of some **gLite** components in some cases
- Make use of the **HTC EGI resources (grid**, cloud, GPGPUs**)** of EOSC-Hub to distribute the computations
- User registration and authentication mechanisms connected to **EGI Check-in**
- Some portals use the **INDIGO-Datacloud udocker** solution



Architecture behind the portals

User DB

User not found

Input error

Validation

Pre-processing
+
Input files packaging

Submission to local nodes

CPU-calculation

OR

Submission to grid node

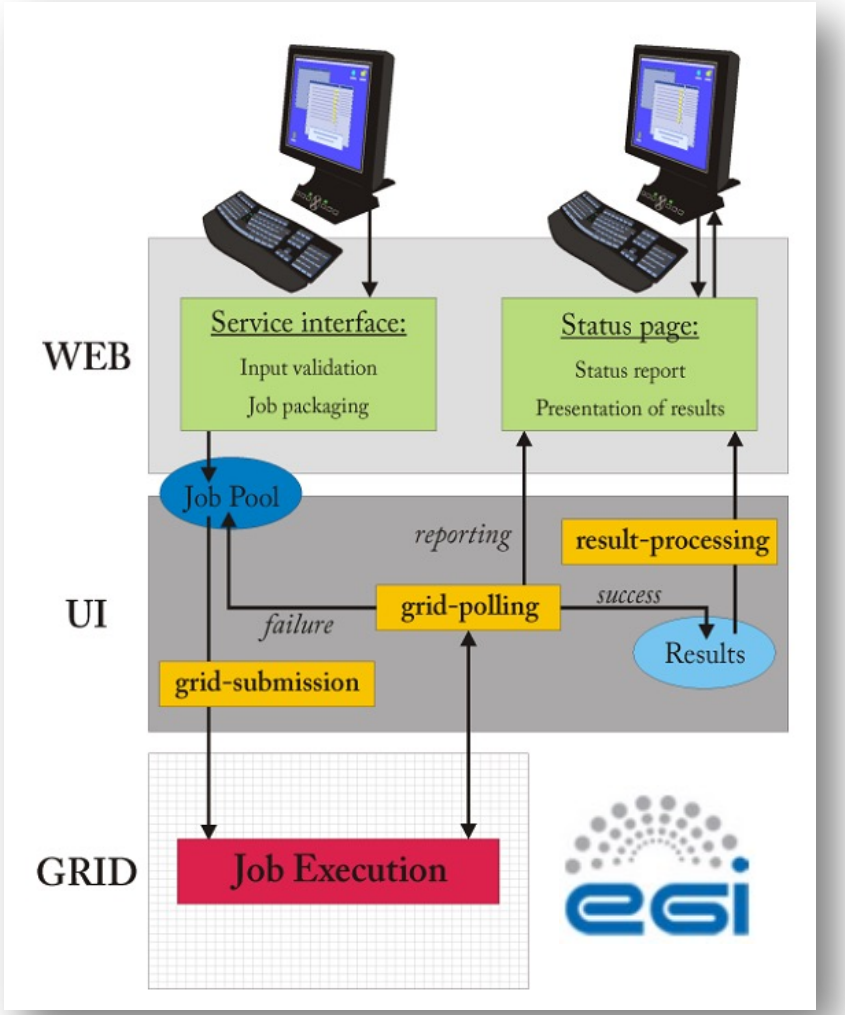
Output files packaging
+
submission of image generation

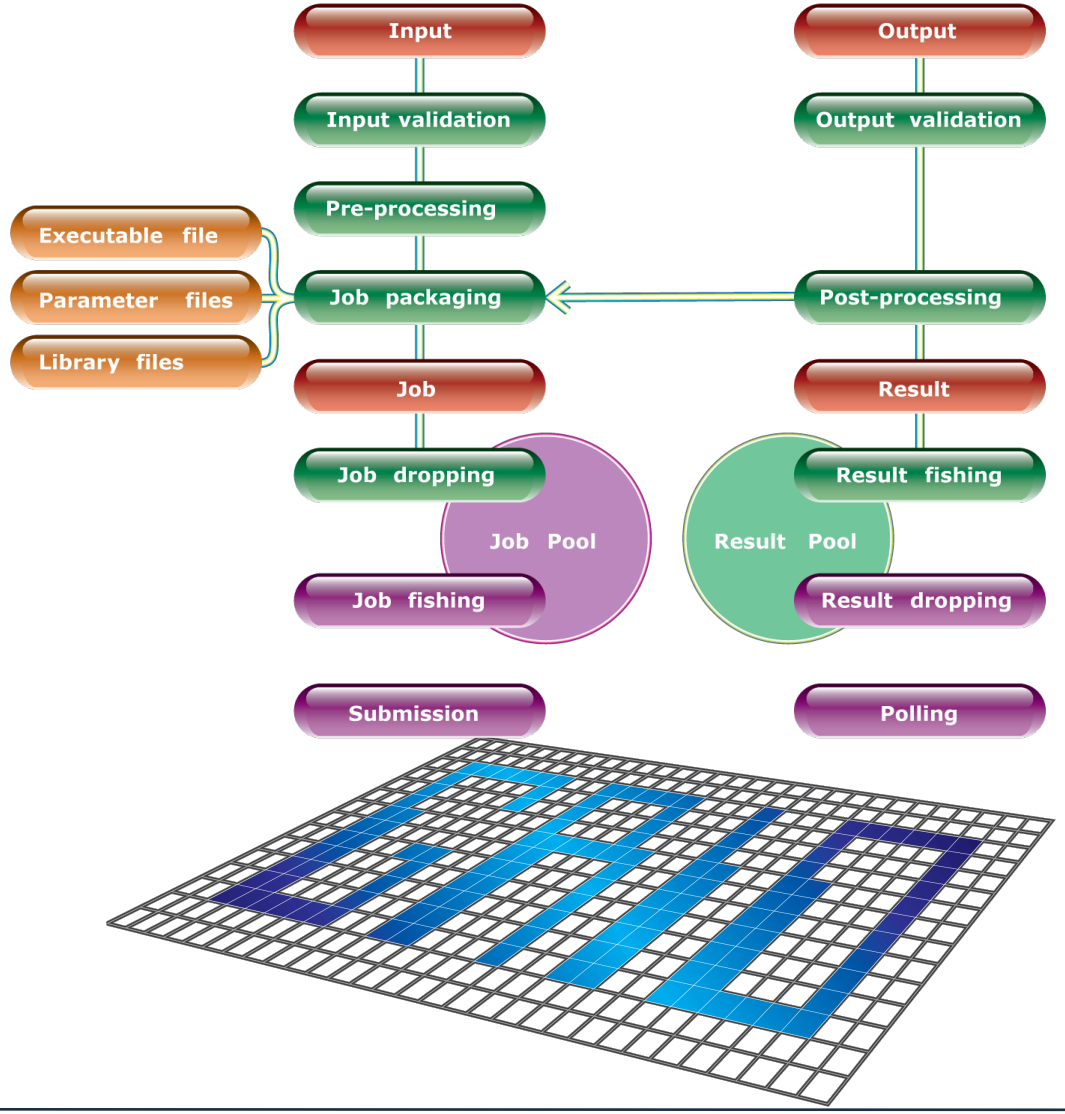
Chimera image generation

Post-processing
+
Results formatting

WEB CLIENT | WEB SERVER | MASTER NODE | WORKING NODE

The daemons behind the portal





Efficient HTC jobs submission - DIRAC4EGI

- **DIRAC4EGI allows for simplified and efficient job submission and monitoring**
  - No root access required
  - Only DIRAC software installation required
  - You can already benefit from it to transform your laptop into an HTC server

DIRAC THE INTERWARE

http://diracgrid.org

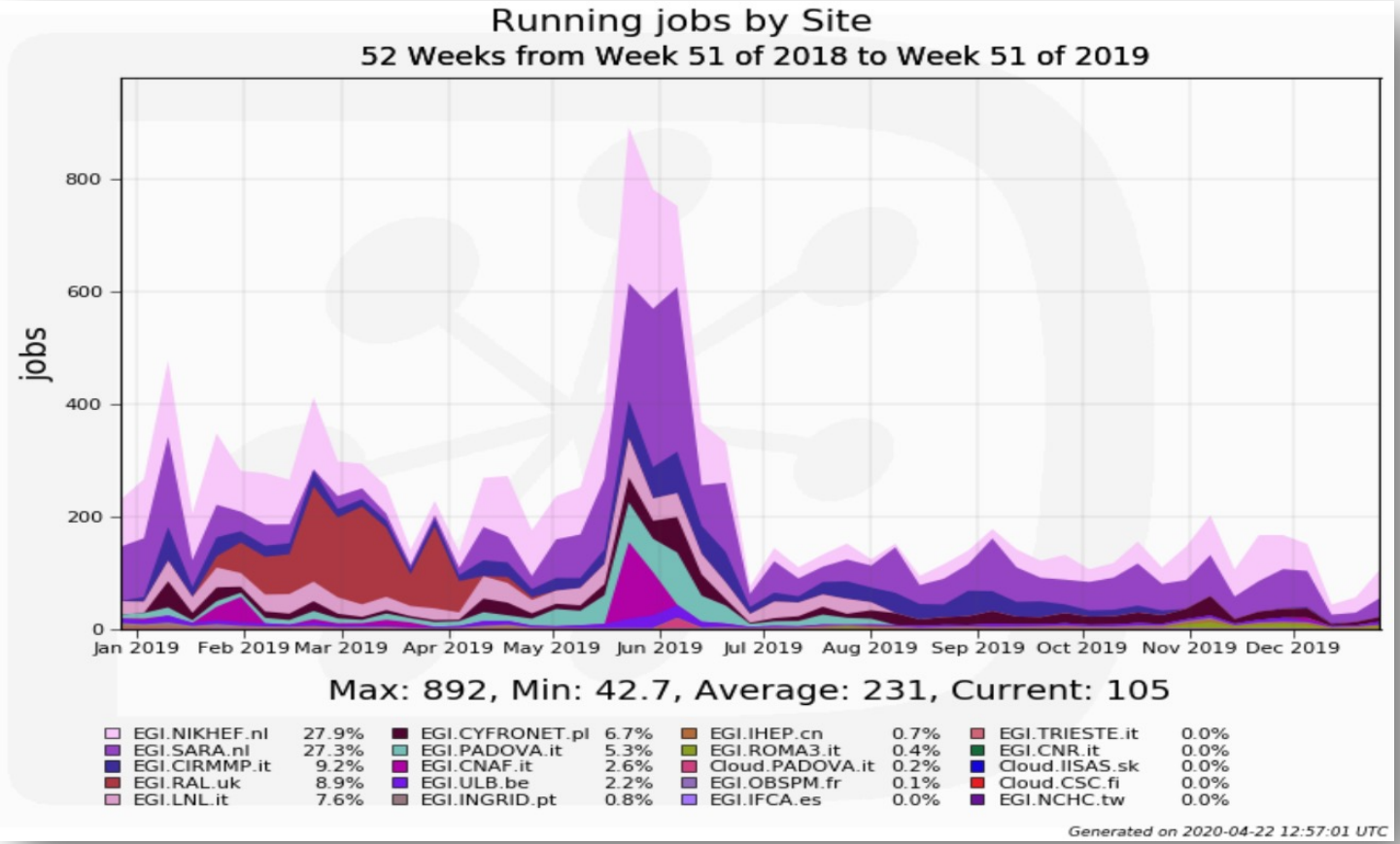
Andrei Tsaregorodtsev

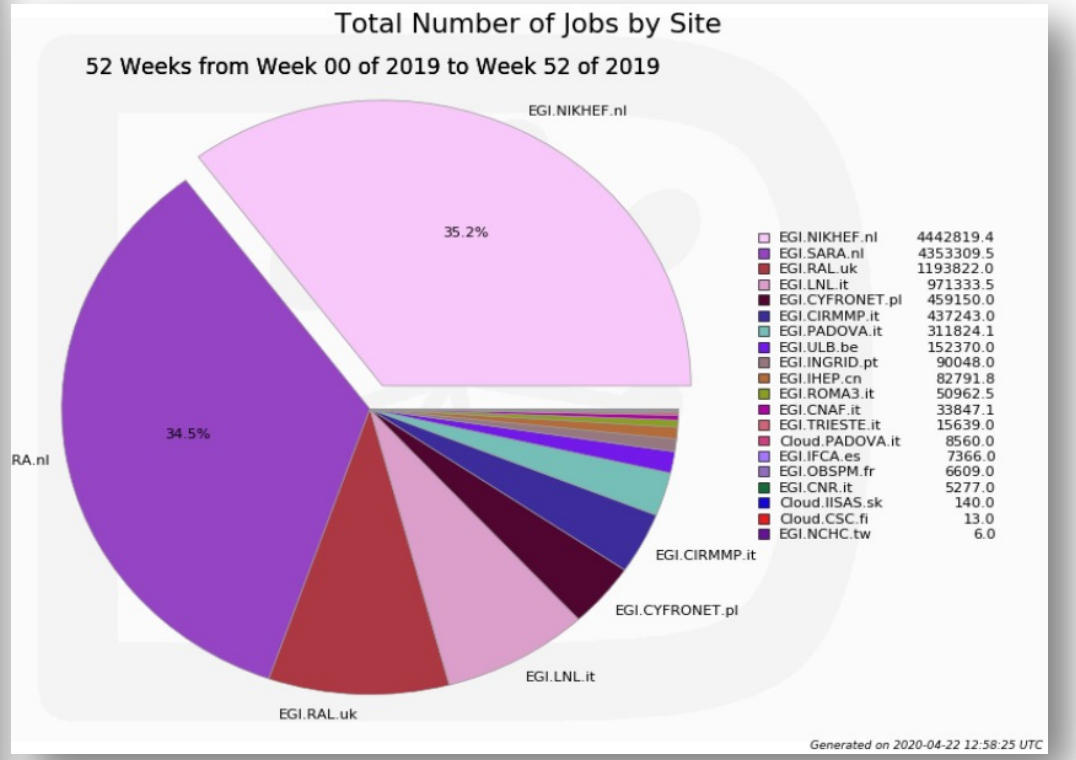
Ricardo Graciani

```
JobName = "dirac-xxx";
CPUTime = 100000;
Executable = "dirac-xxx.sh";
StdOutput = "dirac-xxx.out";
StdError = "dirac-xxx.err";
InputSandbox = {"dirac-xxx.sh","dirac-xxx.tar.gz"};
OutputSandbox = {"dirac-xxx.out", "dirac-xxx.err","dirac-xxx-result.tar.gz"};
```

- **Very efficient submission, high job throughput** (used for our servers since 2015)

Some DIRAC4EGI statistics

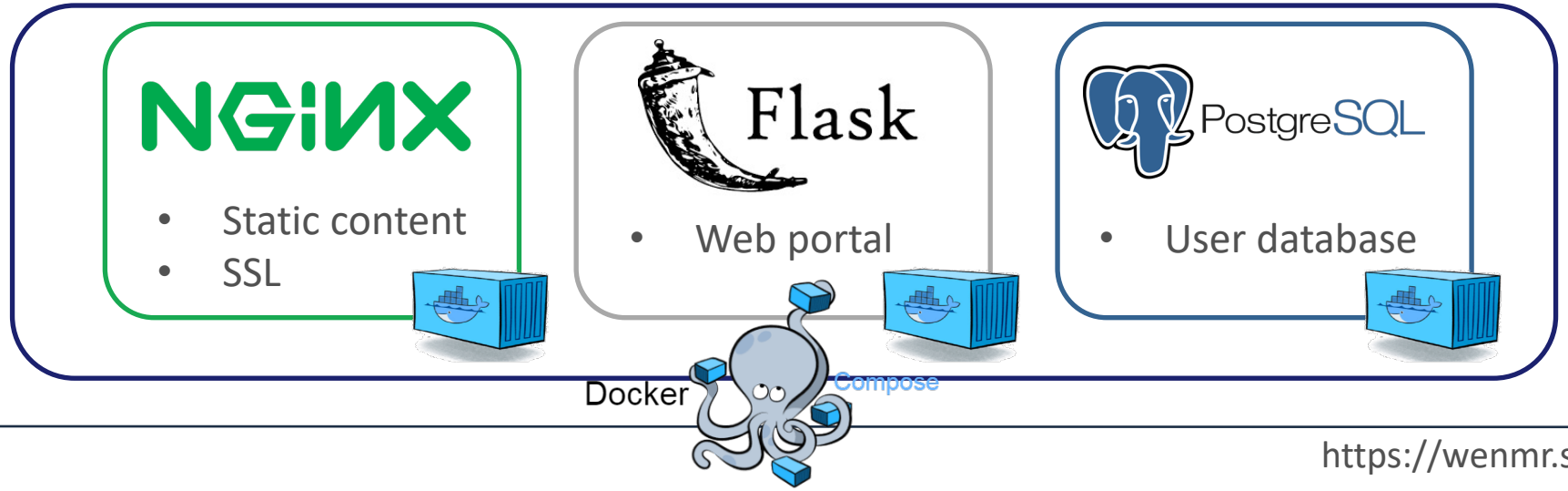




Automated deployment of the HADDOCK server

Migration from local provisioning to provisioning via docker-compose

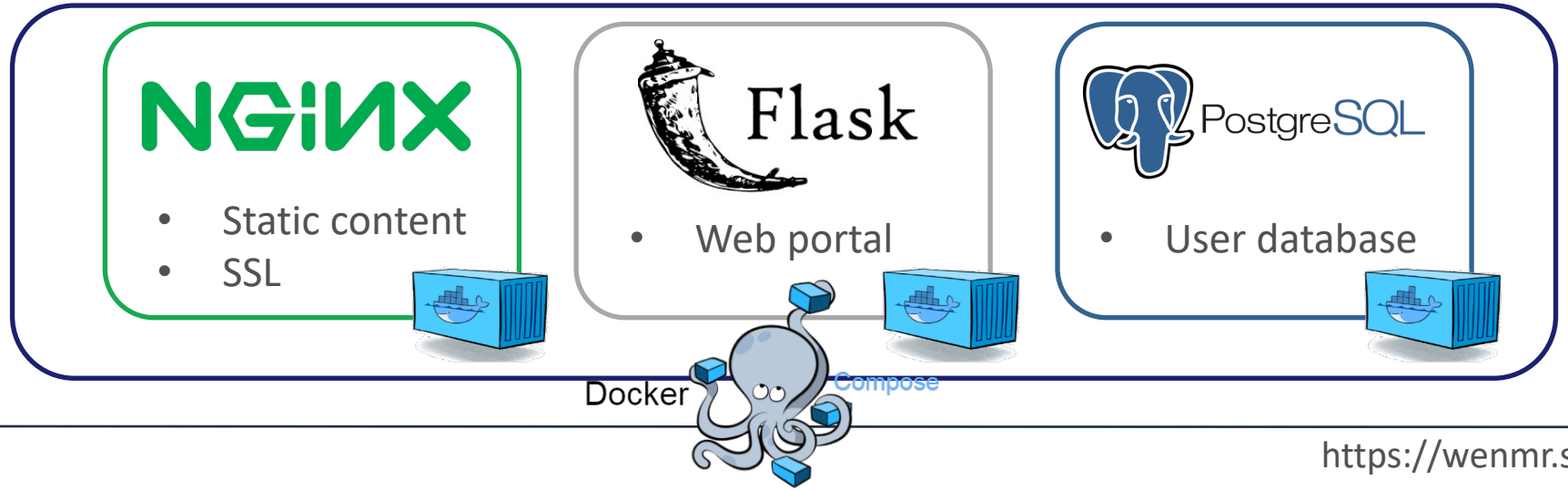
- Improves portability and development
- Separation of the different components
- Important configuration stored in docker compose file
- Easy to switch between development/production configuration
- Easy scaling by deploying to a docker swarm
- Used in HelixNebula pilot project



Automated deployment of the HADDOCK server

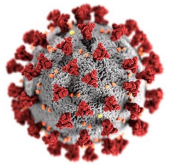
Migration from local provisioning to provisioning via docker-compose

- Improves portability and development
- Separation of the different components
- Important configuration stored in docker compose file
- Easy to switch between development/production configuration
- Easy scaling by deploying to a docker swarm
- Used in HelixNebula pilot project

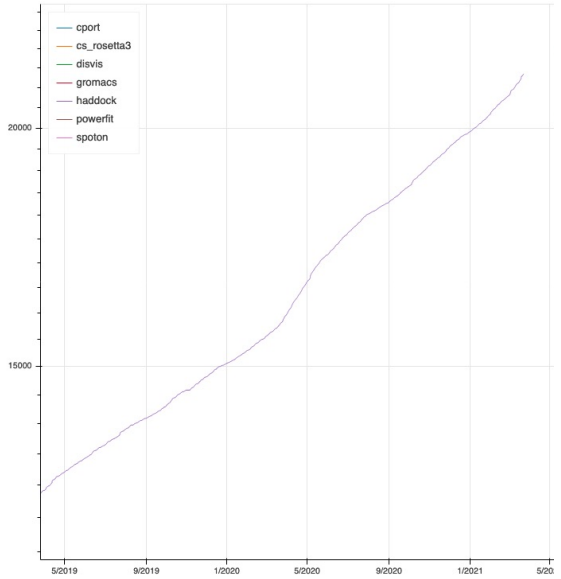


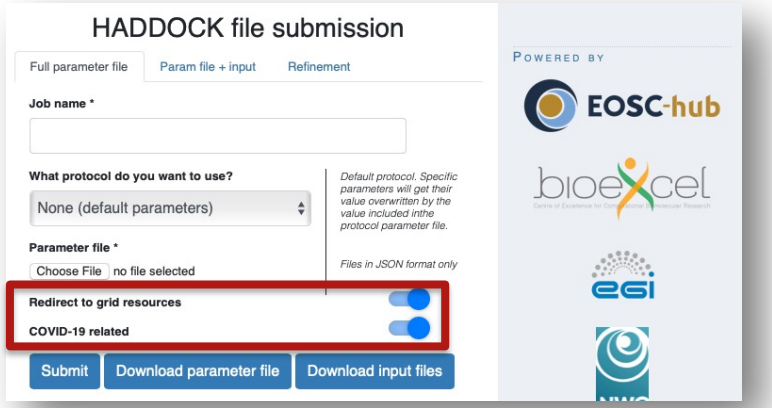

Supporting COVID-19 related research

Meeting the demand

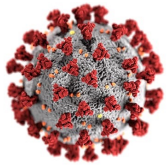


- HADDOCK can be used to model interactions between virus and human proteins and for drug screening
- Increased number of registrations since several weeks (months)
- We have doubled our processing capability by modifying the backend machinery managing the HADDOCK workflow
  - From ~95 docking runs per day to ~183 per day
- Users can now tag their submissions as COVID-related





Increased HTC capacity



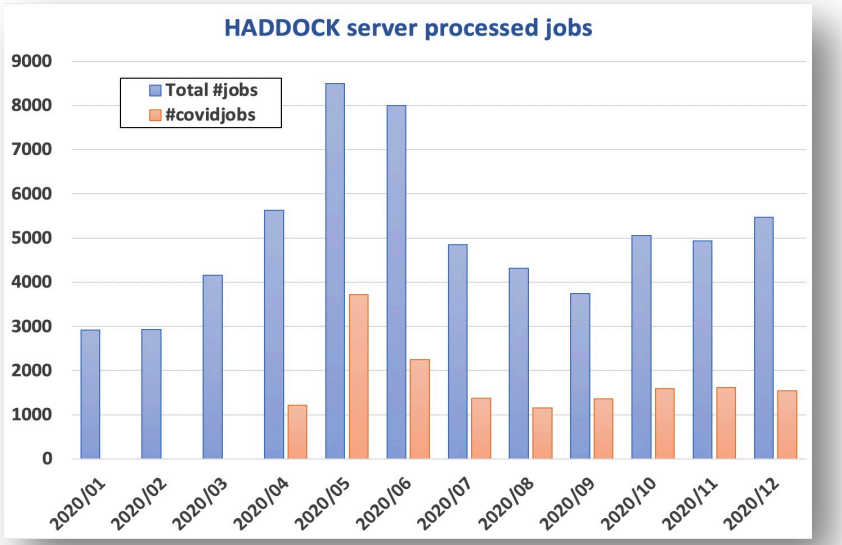
- **Via contacts between EGI and the US Open Science Grid, WeNMR has now access (via DIRAC4EGI) to the US HTC resources**

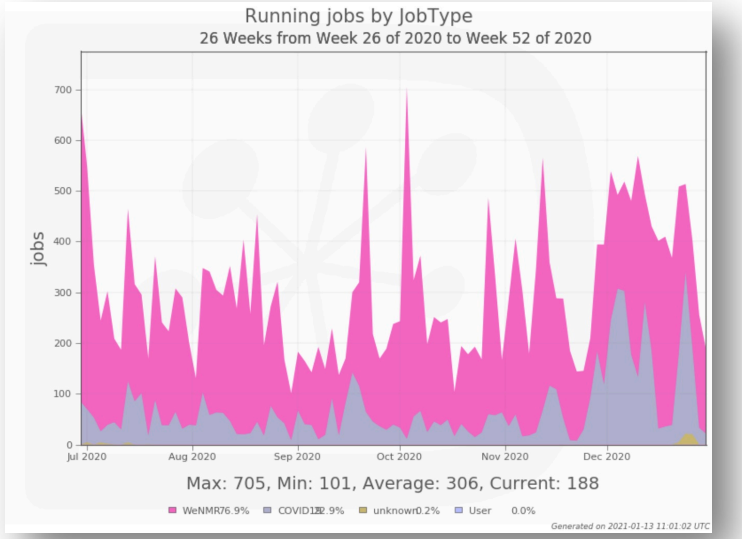
- **Via contacts with high energy physics (WLCG** - Worldwide Large hadron collider Computing Grid**), additional sites are now supporting WeNMR COVID-related jobs**
  - **Centre de Physique des Particules de Marseille (CPPM)**
  - **Karlsruhe Institute of Technology**
  - **Spanish LHCb Tier2 (USC-LCG2) site**

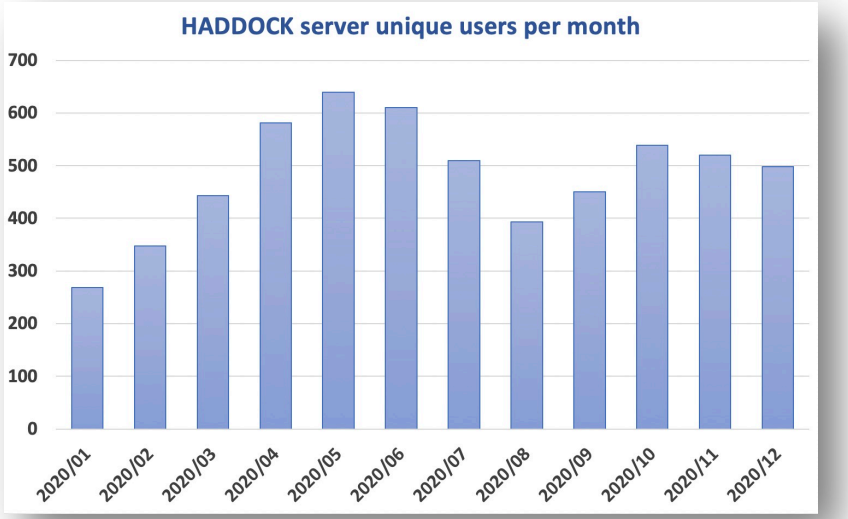
- **Mechanism in place in DIRAC4EGI to tag submissions as COVID and direct those to sites especially supporting this research** (thanks to Andrei Tsaregorodtsev)

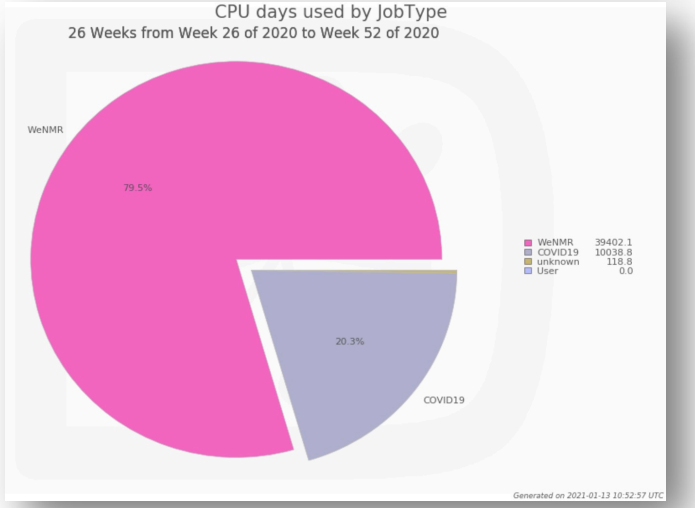
```
Tag = {"COVID19"};
JobType = "COVID19";
```

Increased Usage







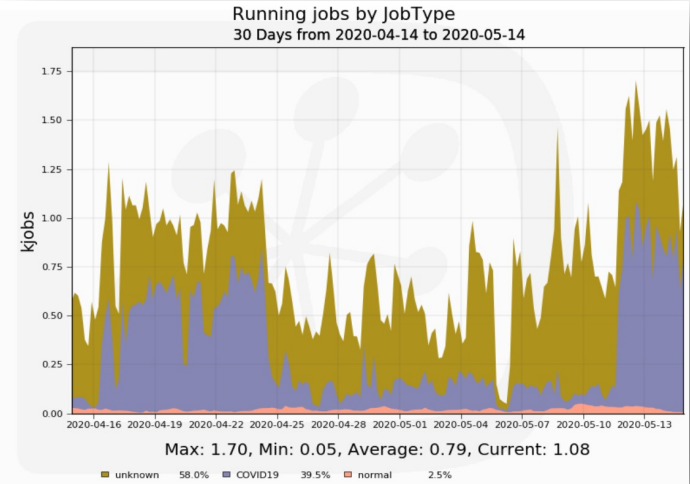


COVID efforts

Modelling of various Sars-Cov2 – human protein interactions

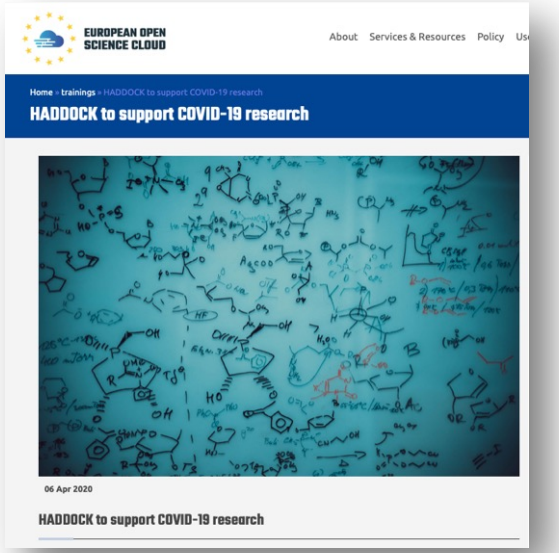
Screening of approved drugs against the protease with HADDOCK

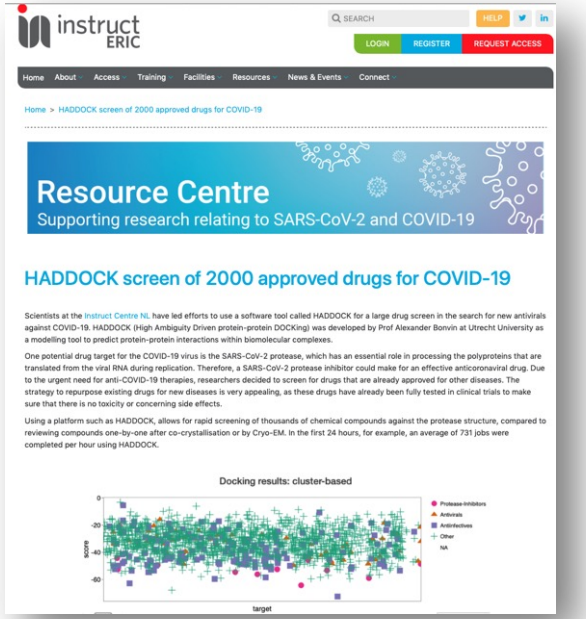
- Docking of ~2000 approved drugs run on EGI/EOSC/OSG HTC resources in ~ 3 1/2 days



See: **bonvinlab.org/covid**

https://instruct-eric.eu/haddock-screen-of-2000-approved-drugs-for-covid19





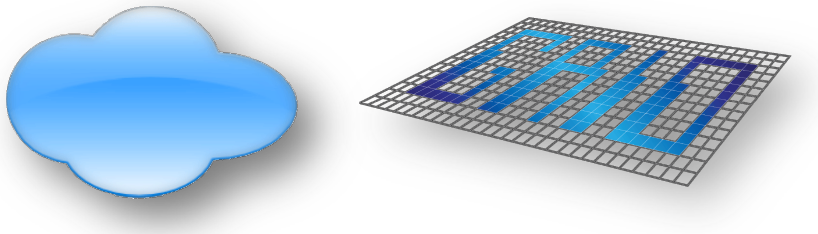
https://www.eosc-portal.eu/news/haddock-support-covid-19-research


Future perspectives

Ambition and timeline

- Keep operating state-of-the-art portals
- Keep ensuring a smooth operation
- Provide user support and training
- Adopt (when needed / makes sense) new solutions
- Keep monitoring user satisfaction 

WeNMR services



- Harvest transparently both
- DIRAC4EGI can handle both without the additional burden of managing cloud VMs
- We still have much more grid than cloud resources
- Efforts within BioExcel to move HADDOCK toward Exascale (e.g. large-scale simulations on HPC resources, using singularity containers)
- Deployment of the HADDOCK portal in the FedCloud for COVID-19 specific projects

Acknowledgements

EOSC-hub

instruct ERIC
EUROPEAN RESEARCH INFRASTRUCTURE CONSORTIUM

NWO

EGI

EGI-ACE

EUROPEAN OPEN SCIENCE CLOUD

bioexcel
Centre of Excellence for Computational Biomolecular Research

INDIGO - DataCloud

DIRAC
THE INTERWARE

SURF SARA

West-Life

we-nmr

MoBrain