

Structural Biology in the Clouds

Past, Present and Future

Alexandre Bonvin

Utrecht University

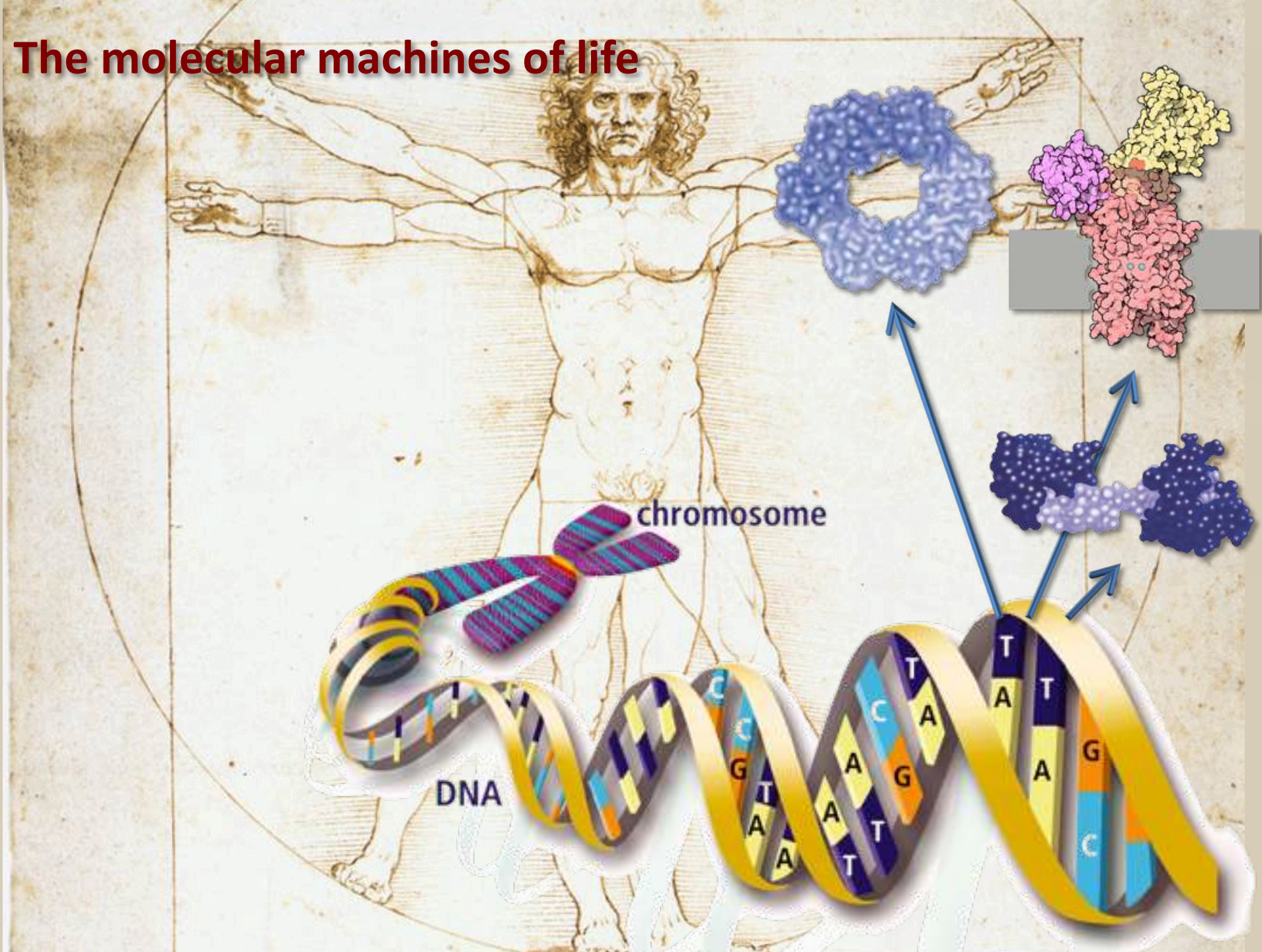
The Netherlands

a.m.j.j.bonvin@uu.nl

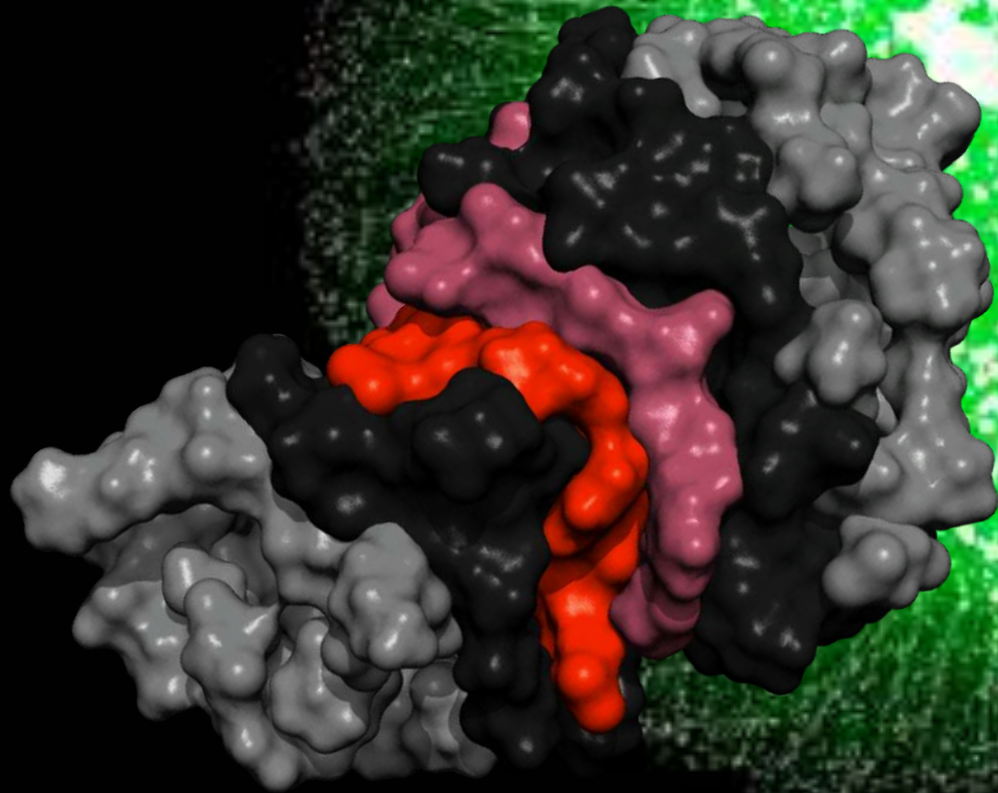


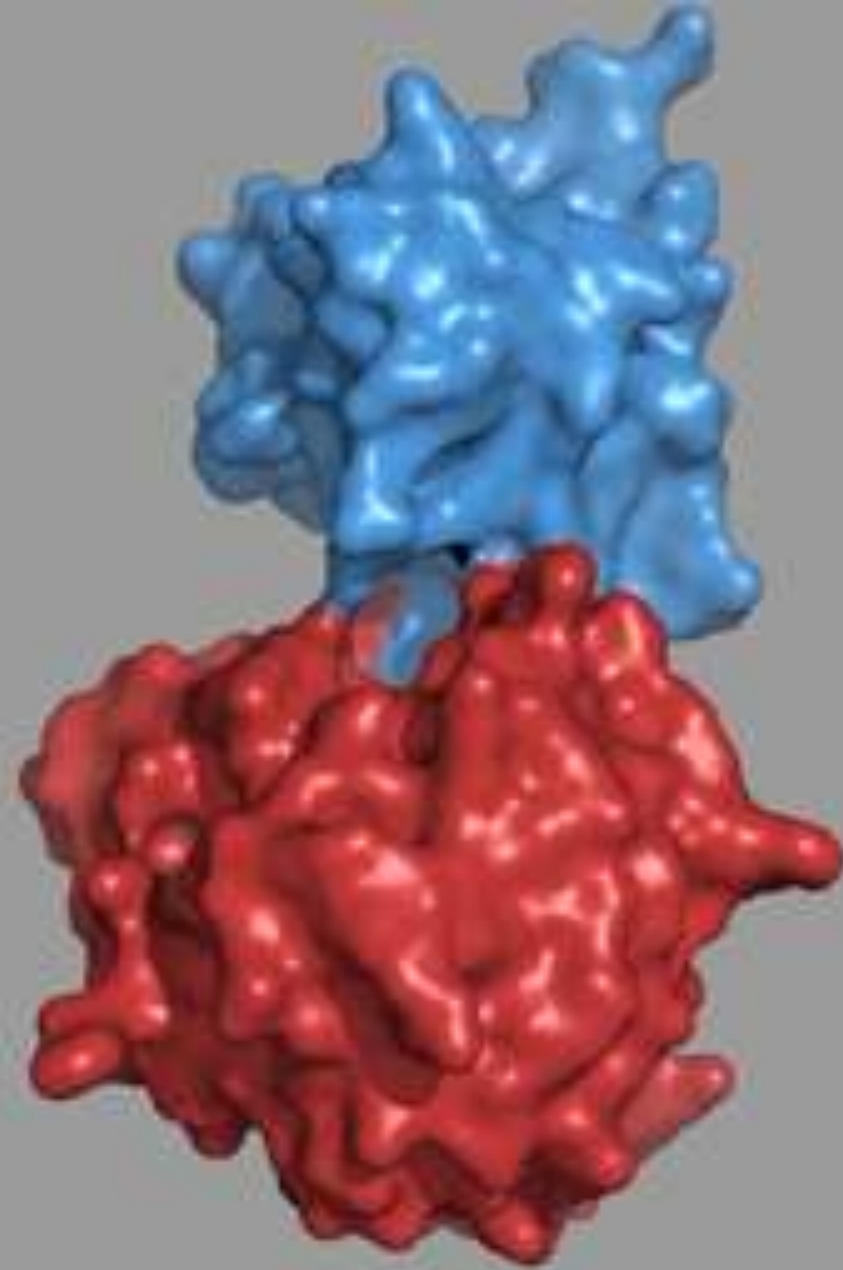
[@amjjbonvin](https://twitter.com/amjjbonvin)

The molecular machines of life

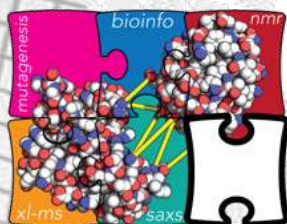


The protein-protein interaction Cosmos





Haddock web portal



HADDOCK
High-Ambiguity Driven Docking

- > 14000 registered users
- > 240000 served runs since June 2008
- > 44% on the GRID

De Vries *et al.* Nature Prot. 2010

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

HADDOCK2.2

WeNMR/West-Life GRID-enabled web portal

WELCOME TO THE WENMR WEB PORTAL >>

HADDOCK (High Ambiguity Driven protein-protein DOCKing) is an information-driven flexible docking approach for the modeling of biomolecular complexes. HADDOCK distinguishes itself from ab-initio docking methods in the fact that it encodes information from identified or predicted protein interfaces in ambiguous interaction restraints (AIRs) to drive the docking process. HADDOCK can deal with a large class of modeling problems including protein-protein, protein-nucleic acids and protein-ligand complexes.

More information about HADDOCK2.2 can be found on the [HADDOCK2.2 website](#)

Read also what an independent review by Moreira *et al.* has to say about our software...

HADDOCK is one of the flagship software in the EU H2020 BioExcel Center of Excellence for Biomolecular Research.

HADDOCK WEBSERVER

REGISTRATION: The use of the HADDOCK WeNMR GRID-enabled docking server is free for academic users. Access to the server is managed through Single Sign On (SSO) authentication using your WeNMR account. Old-style HADDOCK web server accounts are still supported. How to proceed:

1. Become a member of the WeNMR Virtual Research Community at [www.wenmr.eu](#)
2. Once logged in, go to the "My Services" tab in your account profile and subscribe to the HADDOCK web portal. Follow the instructions on screen.
3. Once you are a member of the WeNMR VRC it is easy to subscribe to the many services WeNMR has to offer, some of which will however require a valid X509 personal certificate

SERVICES:

- HADDOCK server: the Easy interface
- HADDOCK server: the Prediction interface
- HADDOCK server: the Expert interface (requires Expert level access)
- HADDOCK server: the Refinement interface (requires Expert level access)
- HADDOCK server: the Guru interface (requires Guru level access)

PARTNERS >>

we-nmr
West-Life
bioexcel
e-infrastructure

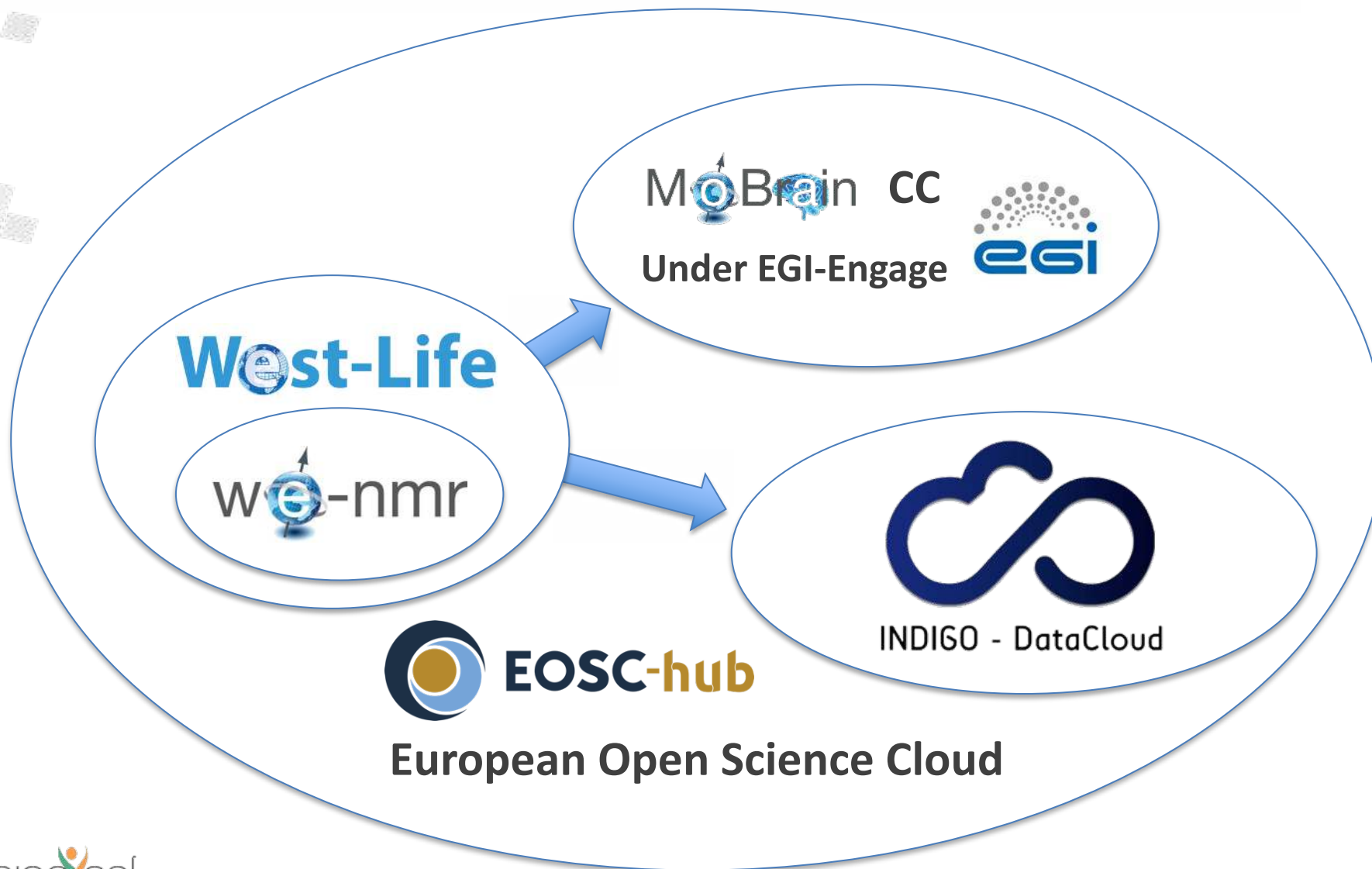
EOOSC-hub



**>10 years of serving the research
community**

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

The eInfrastructure landscape over the years





Virtual Research Community



WENMR

WeNMR is a
worldwide e-
infrastructure for
NMR and structural
biology

-  Email
-  Facebook
-  LinkedIn
-  Github
-  Youtube

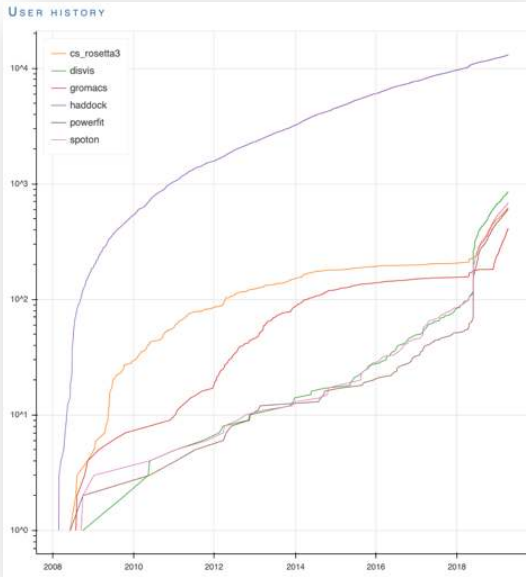
WeNMR is a Virtual Research Community supported by EGI. WeNMR aims at bringing together complementary research teams in the structural biology and life science area into a virtual research community at a worldwide level and provide them with a platform integrating and streamlining the computational approaches 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

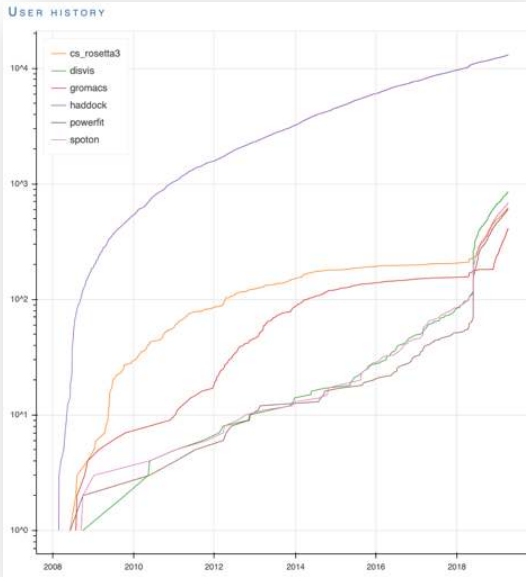
Sustained growth of our user base



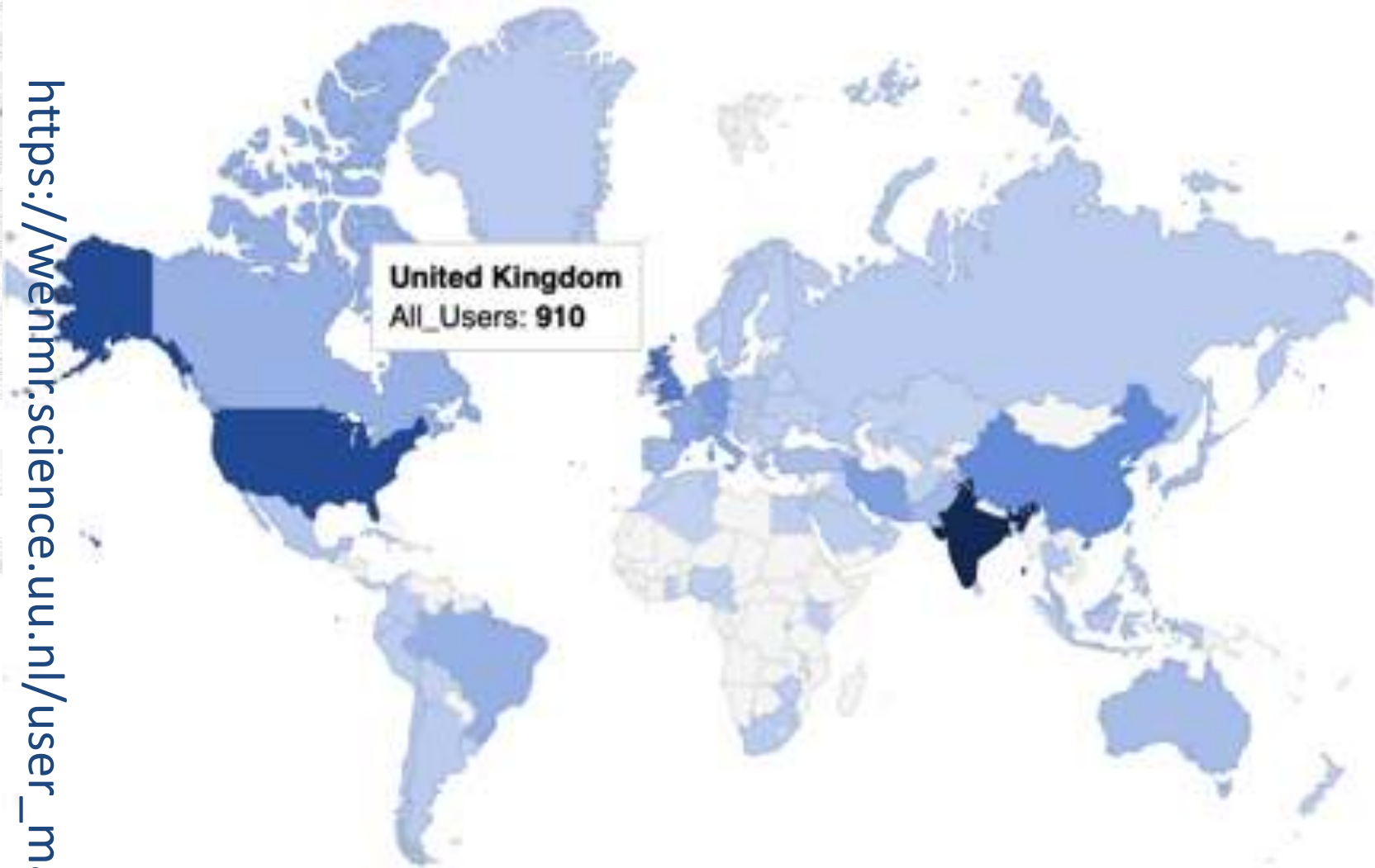
https://www.mr.science.uu.nl/user_map

	Country	All_Users	HADDOCK	DISVIS	POWERFIT	SPOTON	CS_ROSETTA3	GROMACS
1	Total Users	14,867	14,253	1,229	901	1,033	856	695
2	EU Users	3,321	3,098	310	181	201	192	131
3	India	3,185	3,128	199	172	211	153	181
4	United States	2,321	2,227	190	116	150	116	76
5	United Kingdom	910	863	64	55	41	57	34

Sustained growth of our user base

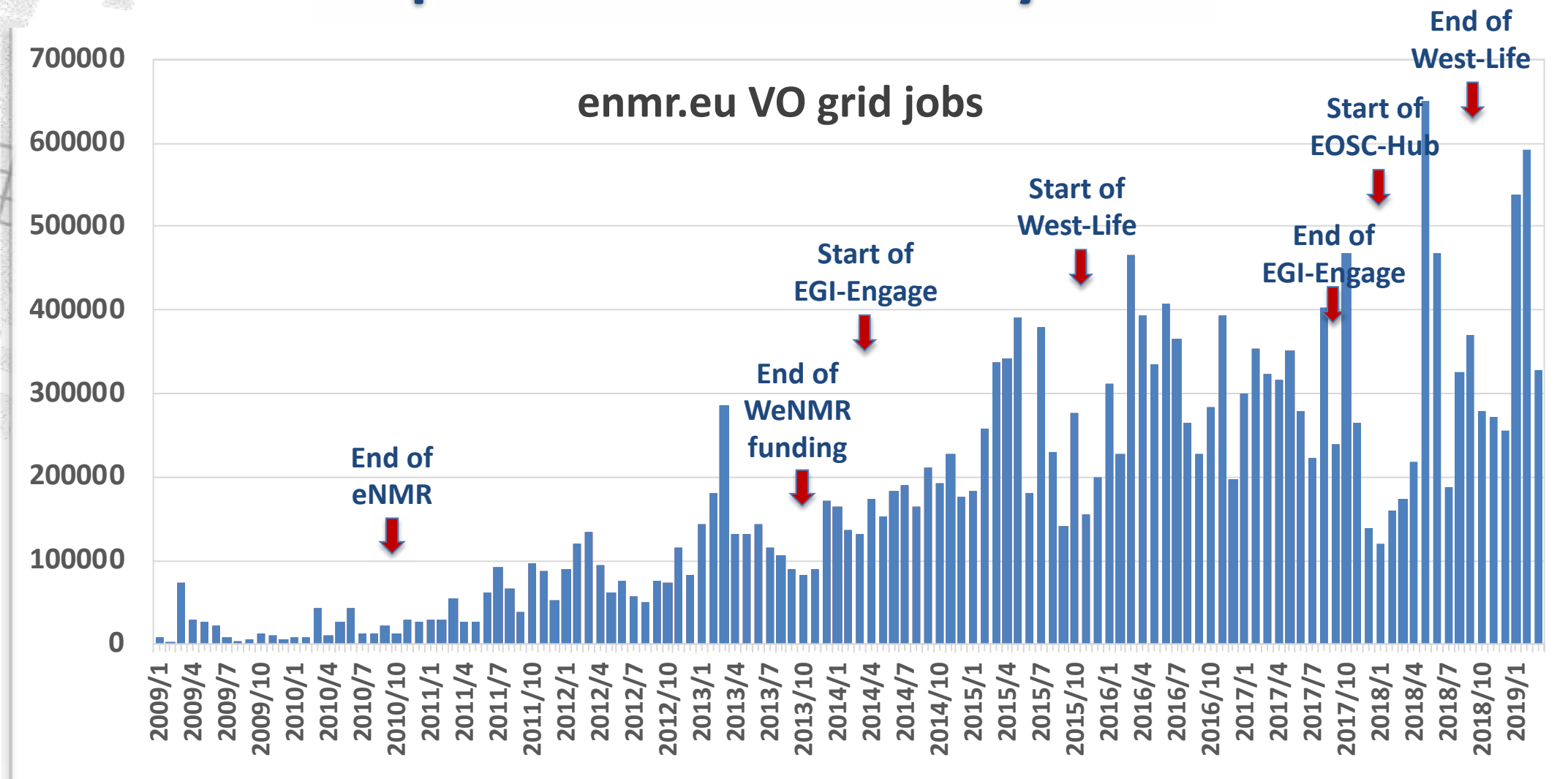


https://wenmr.science.uu.nl/user_map



	Country	All_Users	HADDOCK	DISVIS	POWERFIT	SPOTON	CS_ROSETTA3	GROMACS
1	Total Users	14,867	14,253	1,229	901	1,033	856	695
2	EU Users	3,321	3,098	310	181	201	192	131
3	India	3,185	3,128	199	172	211	153	181
4	United States	2,321	2,227	190	116	150	116	76
5	United Kingdom	910	863	64	55	41	57	34

Operational since >10 years



~2670 normalized CPU years over 2018

Challenges & e-Solutions

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

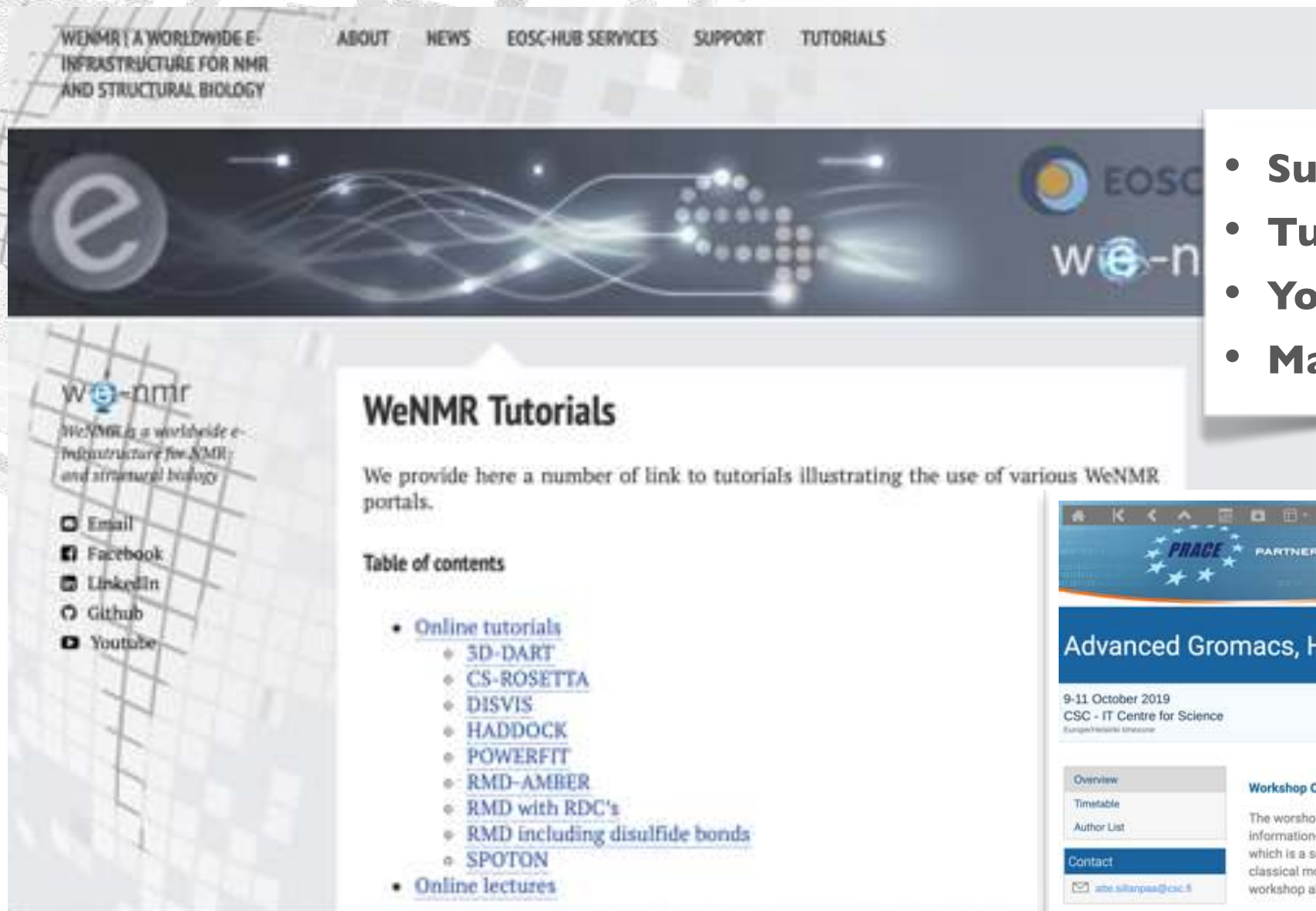
The WeNMR services portfolio



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



The screenshot shows the WeNMR website homepage. At the top, a navigation bar includes links for ABOUT, NEWS, EOSC-HUB SERVICES, SUPPORT, and TUTORIALS. Below this is a large banner with a stylized 'e' logo and the text 'WeNMR a worldwide e-INFRASTRUCTURE FOR NMR AND STRUCTURAL BIOLOGY'. The main content area is titled 'WeNMR Tutorials' and contains a list of links to various online tutorials and lectures, including 3D-DART, CS-ROSETTA, DISVIS, HADDOCK, POWERFIT, RMD-AMBER, RMD with RDC's, RMD including disulfide bonds, SPOTON, and Online lectures. A sidebar on the left lists social media links for Email, Facebook, LinkedIn, Github, and Youtube.

WE-NMR A WORLDWIDE E-INFRASTRUCTURE FOR NMR AND STRUCTURAL BIOLOGY

ABOUT NEWS EOSC-HUB SERVICES SUPPORT TUTORIALS

WeNMR Tutorials

We provide here a number of link to tutorials illustrating the use of various WeNMR portals.

Table of contents

- [Online tutorials](#)
 - [3D-DART](#)
 - [CS-ROSETTA](#)
 - [DISVIS](#)
 - [HADDOCK](#)
 - [POWERFIT](#)
 - [RMD-AMBER](#)
 - [RMD with RDC's](#)
 - [RMD including disulfide bonds](#)
 - [SPOTON](#)
- [Online lectures](#)

Left sidebar:

- Email
- Facebook
- LinkedIn
- Github
- Youtube

- **Support center**
- **Tutorials**
- **YouTube channel**
- **Many workshops ...**



The screenshot shows the PRACE Events Portal. The header includes the PRACE logo and the text 'PARTNERSHIP FOR ADVANCED COMPUTING IN EUROPE'. The main content area is titled 'Advanced Gromacs, HADDOCK + PMX Workshop @ CSC, Finland' and includes the dates '9-11 October 2019' and the location 'CSC - IT Centre for Science'. The bioexcel logo is also present. A sidebar on the left lists 'Overview', 'Timetable', 'Author List', and 'Contact'. The main text describes the workshop objectives, mentioning the use of HADDOCK, PMX, and Gromacs.

PRACE PARTNERSHIP FOR ADVANCED COMPUTING IN EUROPE

EVENTS PORTAL

Advanced Gromacs, HADDOCK + PMX Workshop @ CSC, Finland

9-11 October 2019
CSC - IT Centre for Science
European Research Infrastructure

bioexcel
Centre of Excellence for Computational Biomolecular Research

Workshop Objectives

The workshop will introduce **HADDOCK** (High Ambiguity Driven protein-protein DOCKing) which is an information-driven flexible docking approach for the modeling of biomolecular complexes, and **PMX** which is a service for users who need to do free energy calculations. PMX utilizes the **Gromacs** classical molecular dynamics simulation engine to perform calculations at the background. The workshop also covers advanced usage of Gromacs itself in particular on HPC environment.

Left sidebar:

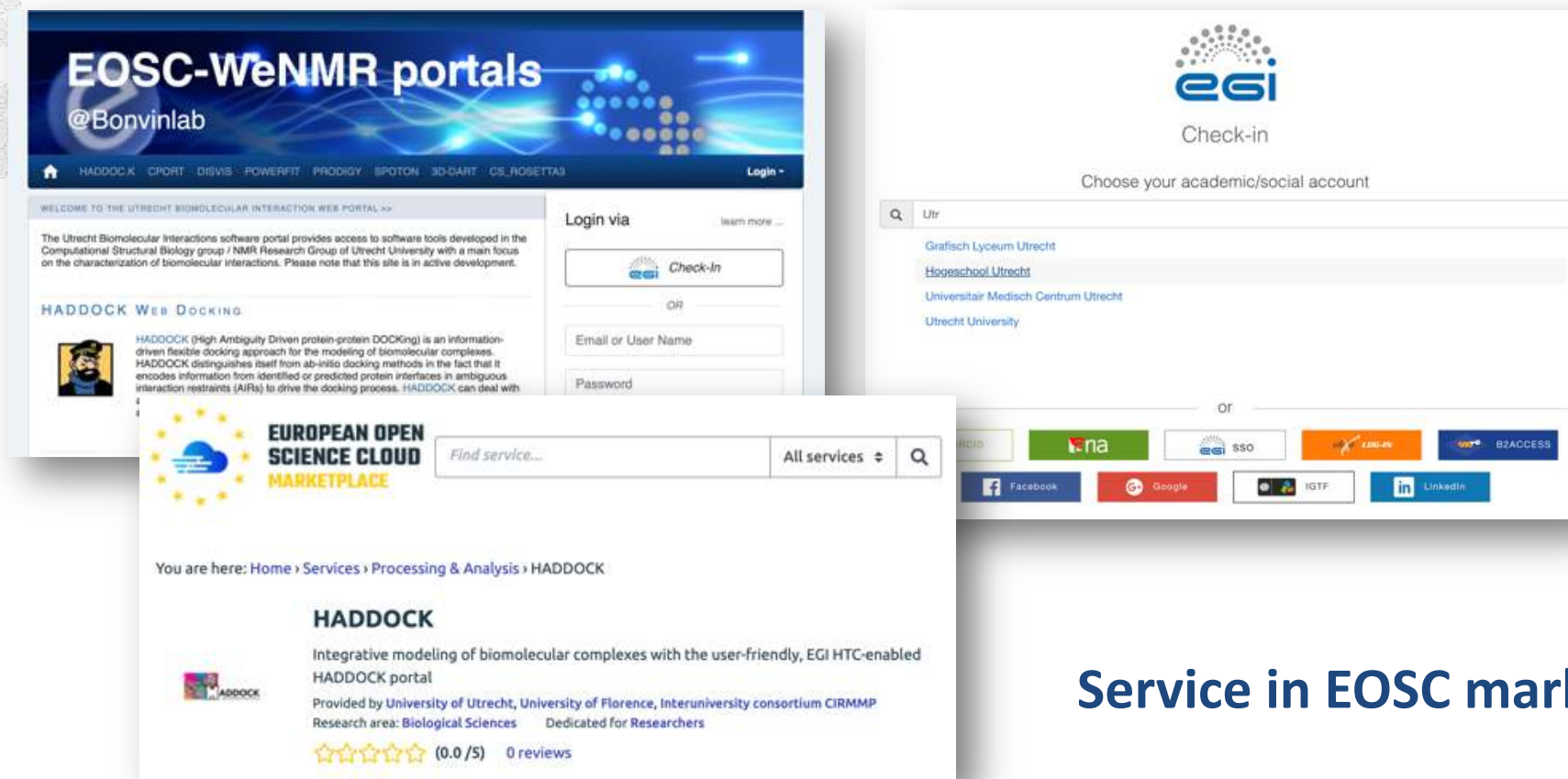
- Overview
- Timetable
- Author List
- Contact

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



The image displays two overlapping screenshots. The background screenshot shows the 'EOSC-WeNMR portals @Bonvinlab' website. It features a navigation bar with links to various tools like HADDOCK, CPORT, DISVIS, POWERFIT, PRODIGY, SPOTON, 3D-GART, CS-ROSETTA3, and a 'Login' button. The main content area includes a 'Login via' section with an 'EGI Check-In' button, an 'OR' separator, and fields for 'Email or User Name' and 'Password'. Below this, there's a section for 'HADDOCK WEB DOCKING' with a brief description of the tool.

The foreground screenshot shows the 'EUROPEAN OPEN SCIENCE CLOUD MARKETPLACE'. It has a search bar with the text 'Find service...' and a dropdown menu showing 'All services'. Below the search bar, there's a breadcrumb trail: 'You are here: Home > Services > Processing & Analysis > HADDOCK'. The main content area displays the 'HADDOCK' service, described as 'Integrative modeling of biomolecular complexes with the user-friendly, EGI HTC-enabled HADDOCK portal'. It lists the providers as 'University of Utrecht, University of Florence, Interuniversity consortium CIRMMP' and the research area as 'Biological Sciences'. The service has a rating of 0.0/5 and 0 reviews.

Service in EOSC marketplace

User satisfaction

– User satisfaction ratings



- CS-ROSETTA2: 5.0 (from 5 respondent)
- DISVIS: 4.8 (from 28 respondents)
- HADDOCK: 4.9 (from 2522 respondents)
- POWERFIT: 4.7 (from 11 respondents)
- PRODIGY: 4.7 (from 410 respondents)
- SPOTON: 4.9 (from 56 respondents)

Challenges & e-Solutions

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

Resources

- **The WeNMR services have been in production since ~10 years under various projects (eNMR, WeNMR, West-Life)**
- **Access to resources formalized through a SLA agreement valid until 12/2020**
 - 60 million CPU hours (opportunistic access)
 - 160 cloud CPU cores
 - 250 TB storage

HTC resources

VO Admin — Normalized CPU time (hours) by Country and Half-year

Country	May 2018 — Oct 2018	Nov 2018 — Apr 2019	Total	Percent *
Netherlands	7,037,159	8,465,237	15,502,396	65.43%
United Kingdom	1,932,729	2,001,129	3,933,858	16.6%
Italy	592,649	1,326,401	2,222,139	9.21%
Poland	510,883			
Belgium	157,989			
Portugal	145,309			
China	77,506			
Taiwan	45,650			
France	2,310			
Brazil	0			
Total	10,502,185			
Percent	44.33%			



COMMUNITY • INFRASTRUCTURE • SERVICES • SOLUTIONS • HOW TO? • CASE STUDIES • NEWS

Newsfeed

Newsletters

Presentations

Publications

Press

Videos

HOME > NEWS & MEDIA > NEWSFEED > EGI – MOBRAIN COLLABORATION: AN SLA FOR BETTER RESEARCH

EGI – MoBrain collaboration: an SLA for better research

The EGI Foundation is pleased to announce an agreement between a group of EGI resource centres and the MoBrain research collaboration.

Thanks to the agreement, MoBrain will be able to access the High-Throughput Computing and File Storage services they need for their research.

World-wide: Dec

Challenges & e-Solutions

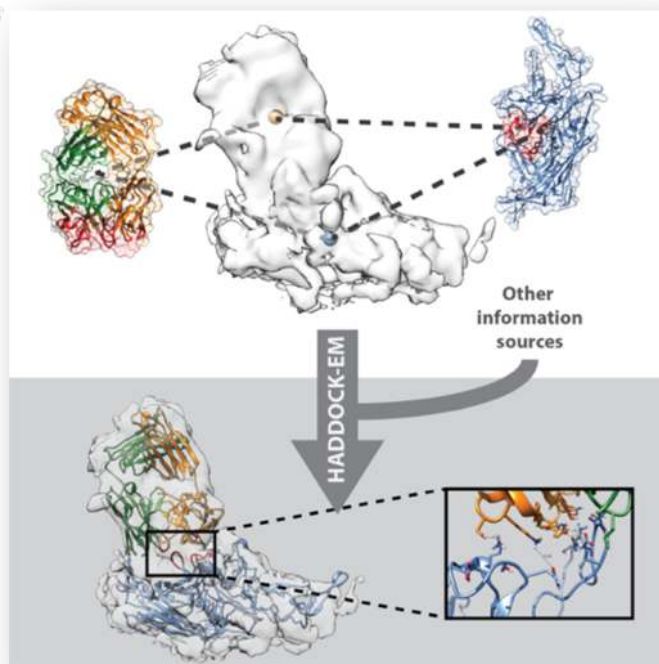
- **Attract users!**
- **Access to e-Infrastructure**
- **Develop software and maintain and operate a complex infrastructure**

The background of the slide is a high-contrast, artistic photograph of a sky. It features dark, heavy, and swirling clouds in shades of deep blue and black. A bright light source, likely the sun, is positioned in the upper center, creating a strong lens flare that radiates across the frame. The light illuminates the edges of the clouds, giving them a white and light blue appearance where they catch the light. The overall mood is dramatic and ethereal.

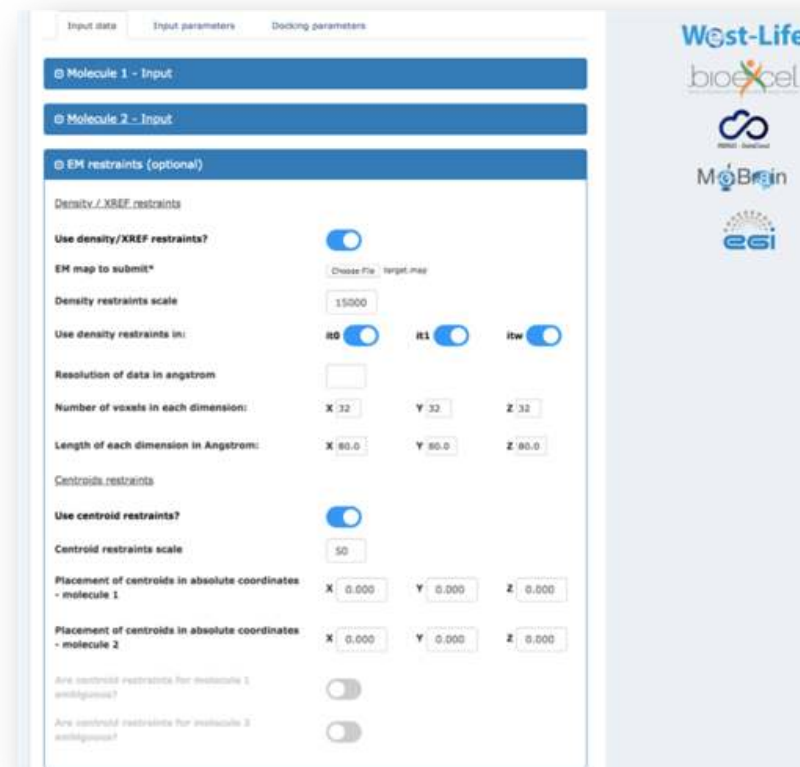
Containerization for running services

HADDOCK development's highlights

- New beta version of the server
<https://haddock.science.uu.nl/services/HADDOCK2.4/>
- Support for cryo-EM data¹, cyclic peptides, etc.



Trellet et al., *Protein-protein modelling using cryo-EM restraints*, 2018, submitted

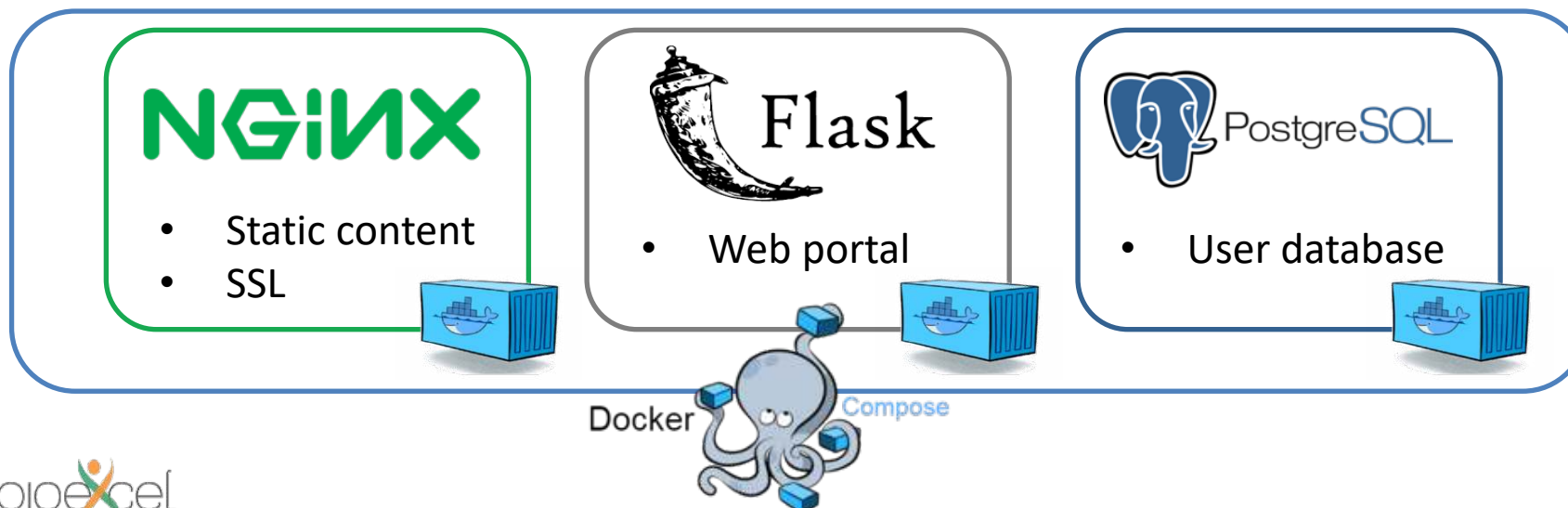




Automated deployment of the HADDOCK server

Migration from local provisioning to provisioning via docker-compose

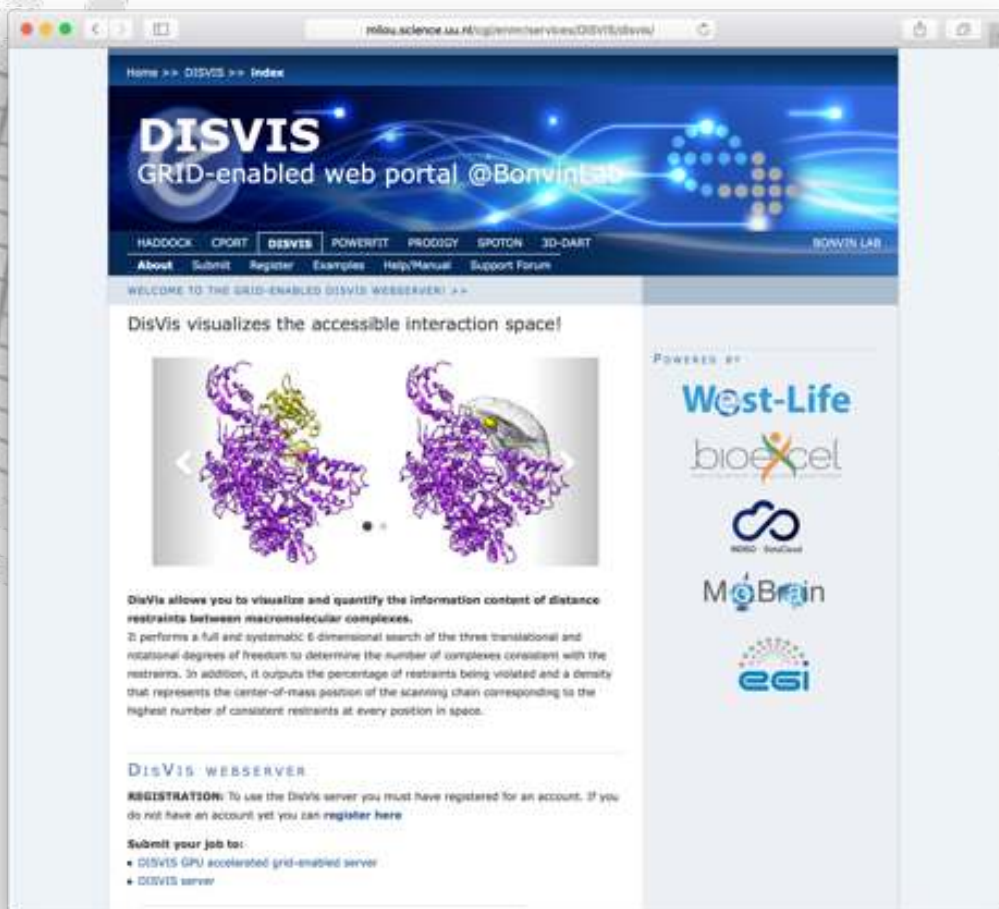
- Improves portability and development
- Separation of the different components
- Important configuration stored in docker compose file
- Easy to switch between development/production configuration
- Easy scaling by deploying to a docker swarm
- Used in HelixNebula pilot project



The background of the slide is a high-contrast, artistic photograph of a sky. A bright sunburst is visible in the upper center, with rays of light breaking through a dense layer of dark, swirling clouds. The clouds are rendered in shades of deep blue, grey, and white, creating a sense of depth and movement. The overall mood is dramatic and hopeful, suggesting a breakthrough or a new dawn.

Harvesting GPGPU resources through containerization

GPGPU, GRID-enabled web portals



<http://milou.science.uu.nl/enmr/services/DISVIS/>



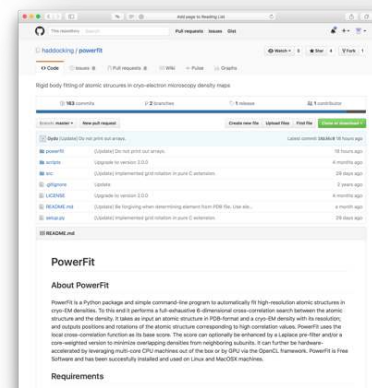
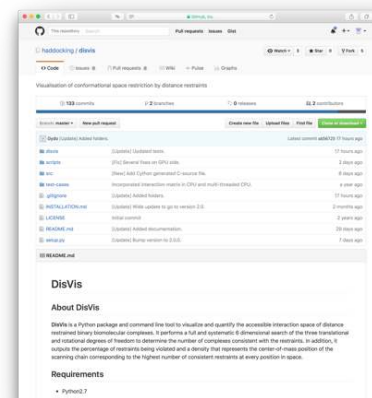
<http://milou.science.uu.nl/enmr/services/POWERFIT/>

Software Provisioning

Because of complex software dependencies we use docker containers

- Python2.7
- NumPy 1.8+
- SciPy
- FFTW3
- pyFFTW 0.10+
- OpenCL1.1+
- pyopengl
- cIFFT
- gpyfft

And to avoid security issues on the grid side, **udocker** from INDIGO



SUMMARY SUCCESS FACTORS

SOFTWARE/SCIENCE

MANWARE

HARDWARE

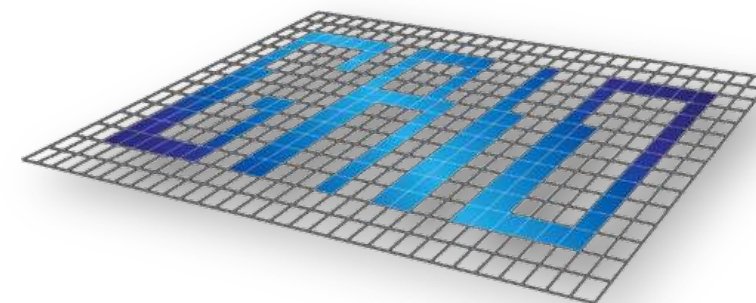
NETWORKING

SUPPORT

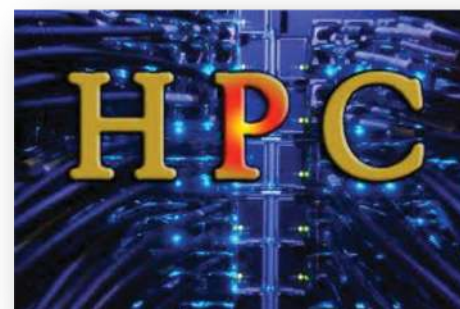
\$\$\$\$\$

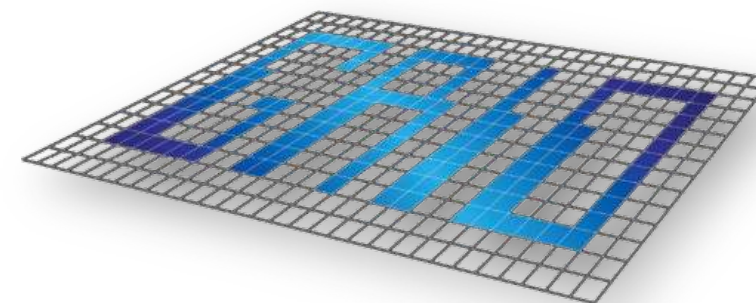
The image shows a vast, turbulent sky filled with dark, swirling clouds. A bright light source, likely the sun, is positioned in the upper center, casting a powerful glow that illuminates the surrounding clouds and creates a rainbow-like spectrum of colors. The overall mood is one of mystery and awe, suggesting a journey or a significant event.

Where are we going?

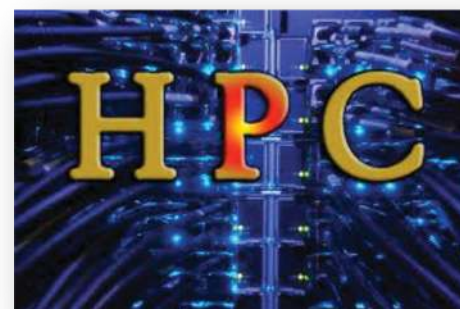


Which solution?


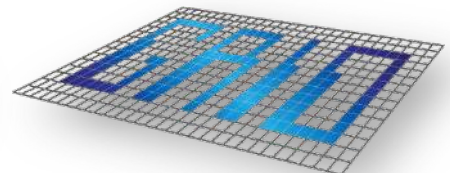




A bit of everything...

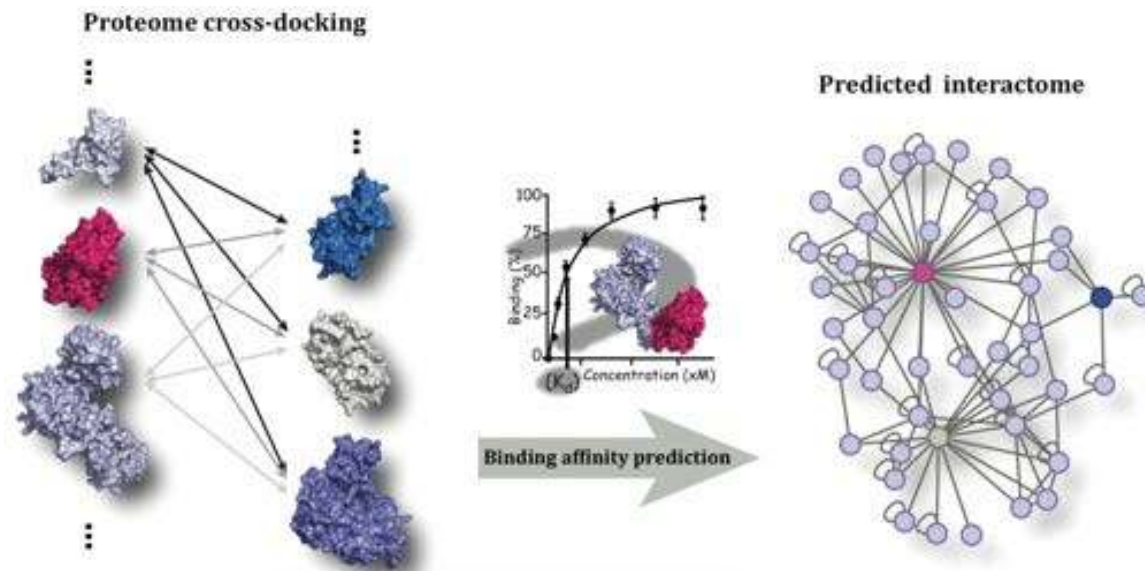


Thematic services under EOSC-Hub

- Harvest transparently both  
- DIRAC4EGI can handle both without the additional burden of managing the cloud VMs
- We still have much more grid than cloud resources
- HADDOCK portal as use case in Helix Nebula Science Cloud

The exascale challenge

- ~20'000 human proteins
- Hundreds of thousands of interactions
- Billions CPU hours and exabytes of data
- Need to make our software ready for it!



bioexcel.eu





Acknowledgments:

the CSB group@UU



VICI
TOP-PUNT



WeNMR
West-Life
EGI-Engage
INDIGO-
Datacloud
BioExcel CoE
EOSC-Hub
SURFSara



