Structural Biology in the Clouds
Past, Present and Future

Alexandre Bonvin

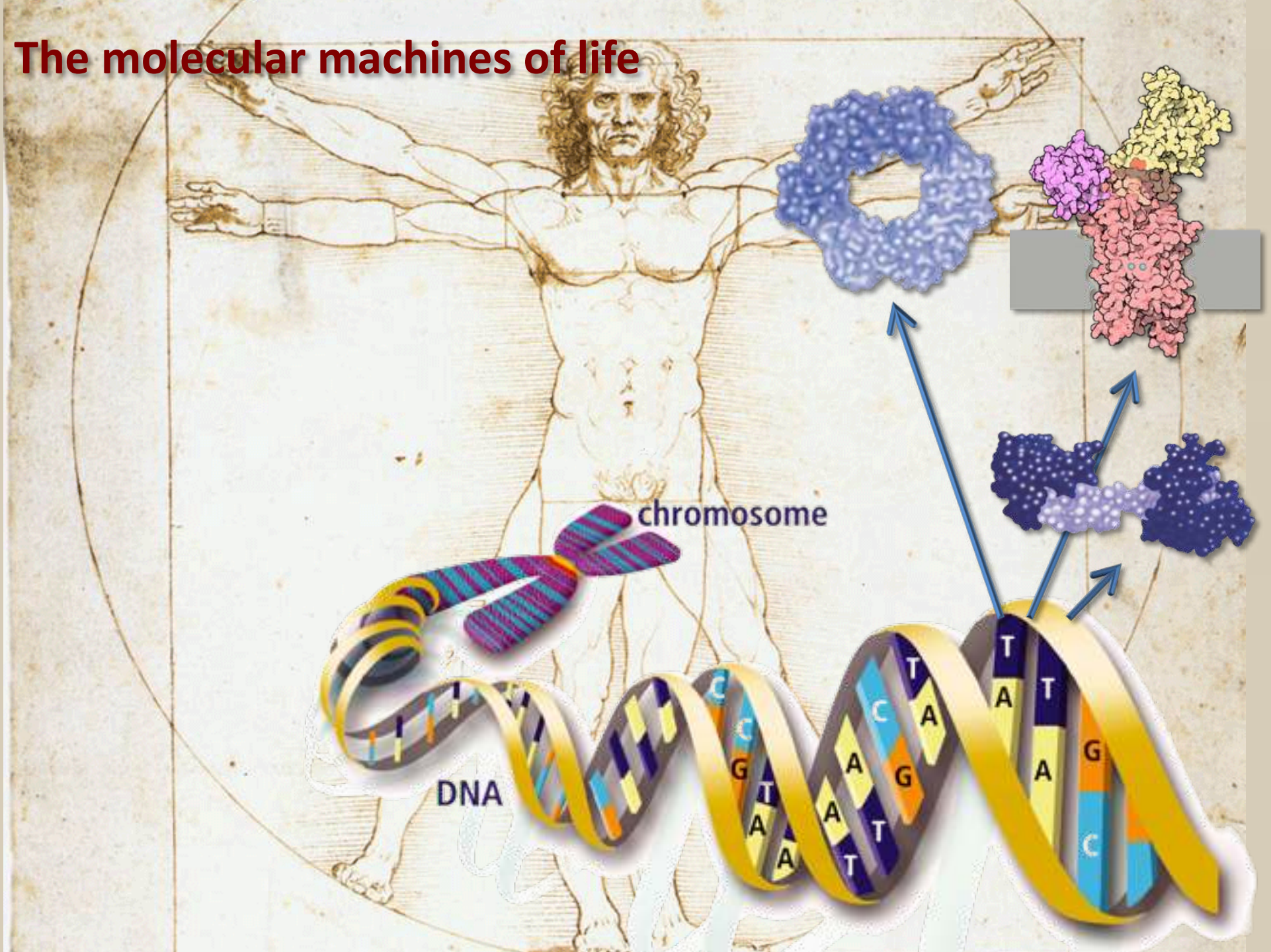
Utrecht University

The Netherlands

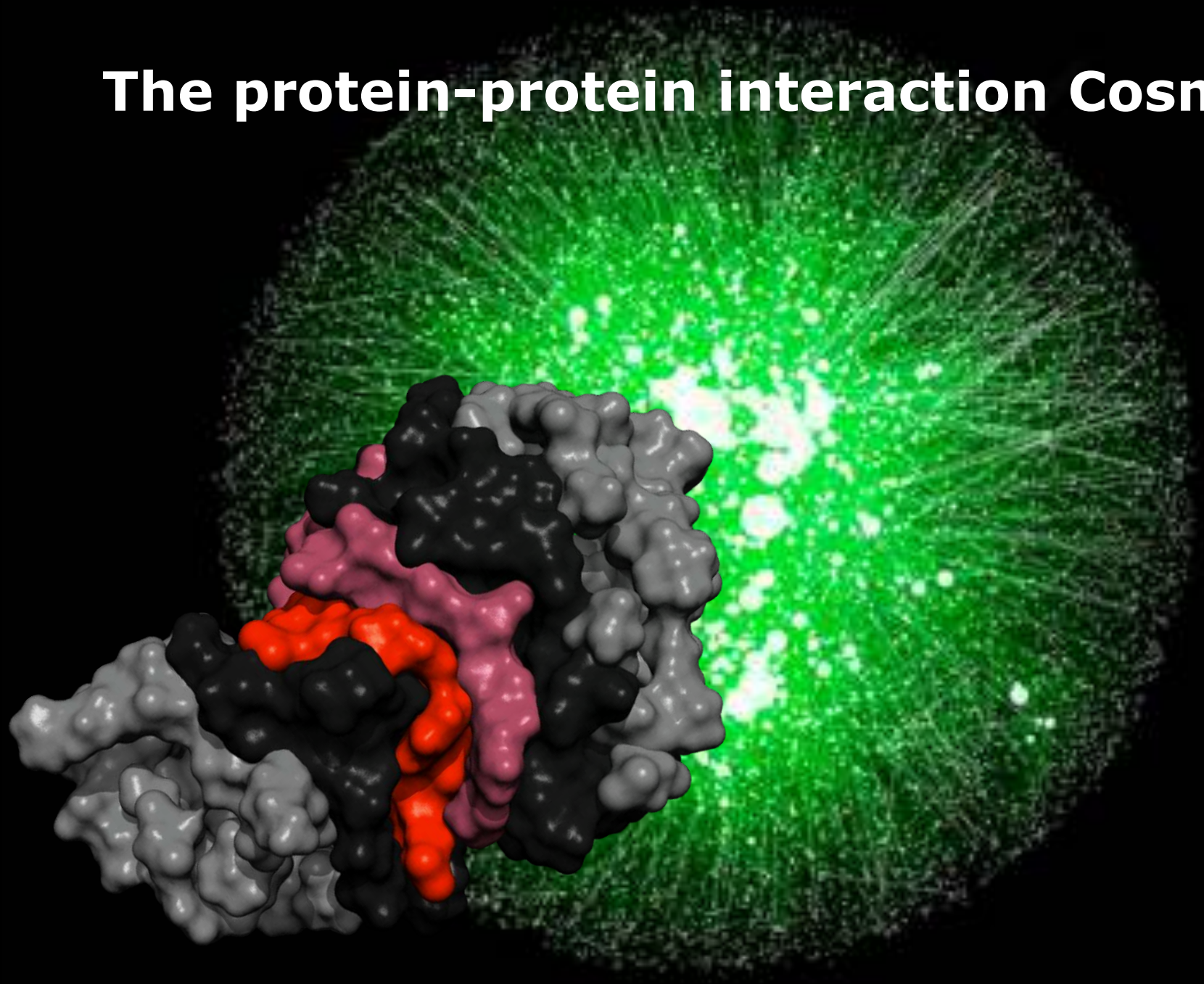
a.m.j.j.bonvin@uu.nl

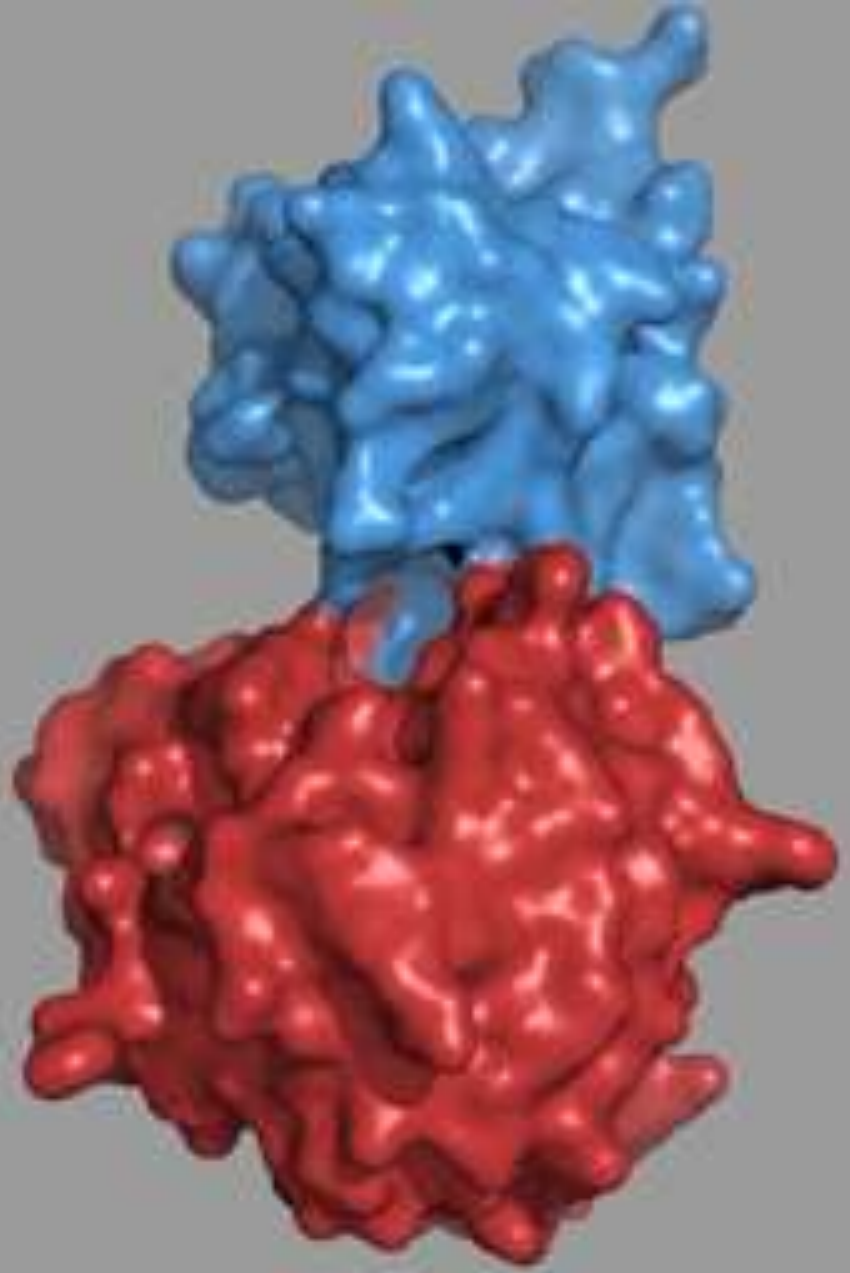
@amjjbonvin

The molecular machines of life

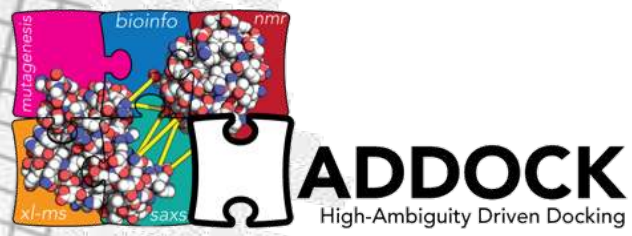


The protein-protein interaction Cosmos





Haddock web portal



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

De Vries *et al.* Nature Prot. 2010

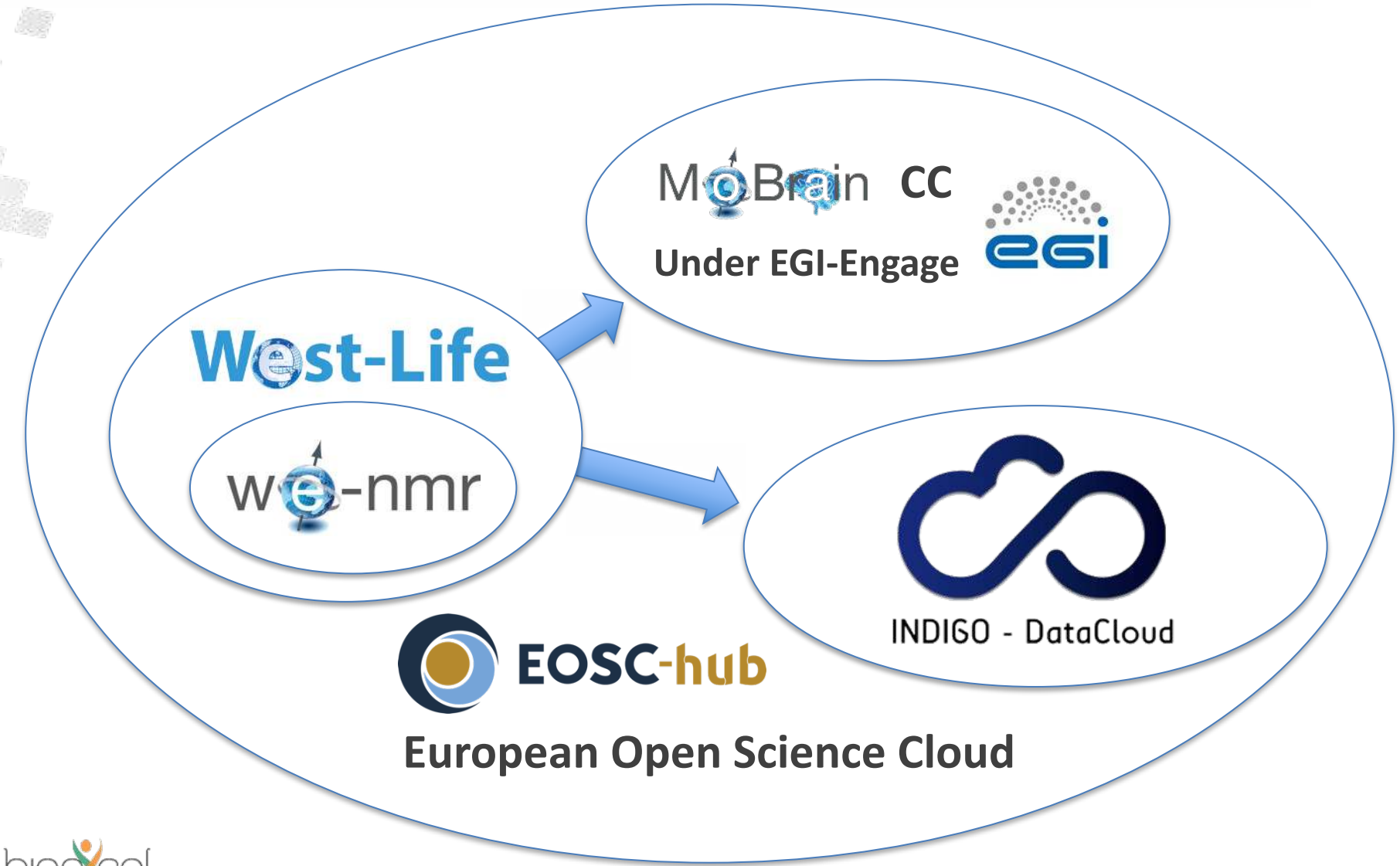
Van Zundert *et al.* J.Mol.Biol. 2016

EOSC-hub

e-infrastructure


>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



EOSC-hub

Virtual Research Community

INDIGO - DataCloud

West-Life Structures for Life

bioexcel Centre of Excellence for Computational Biomolecular Research

e-infrastructure

WENMR | A WORLDWIDE E-INFRASTRUCTURE FOR NMR AND STRUCTURAL BIOLOGY

ABOUT NEWS EOSC-HUB SERVICES SUPPORT TUTORIALS

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.

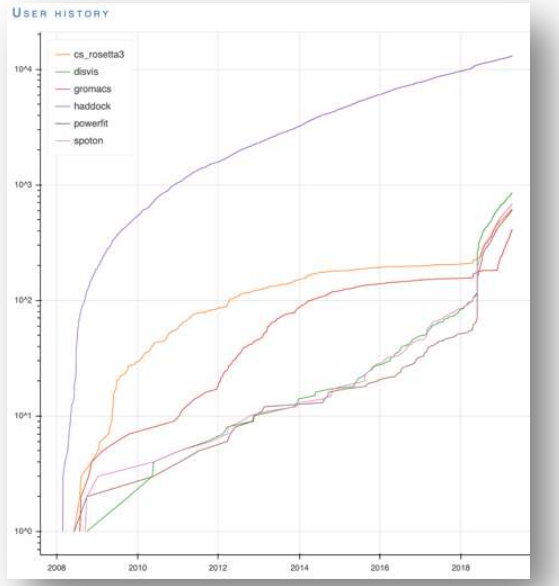
EOSC-hub

www.wenmr.eu

e-infrastructure

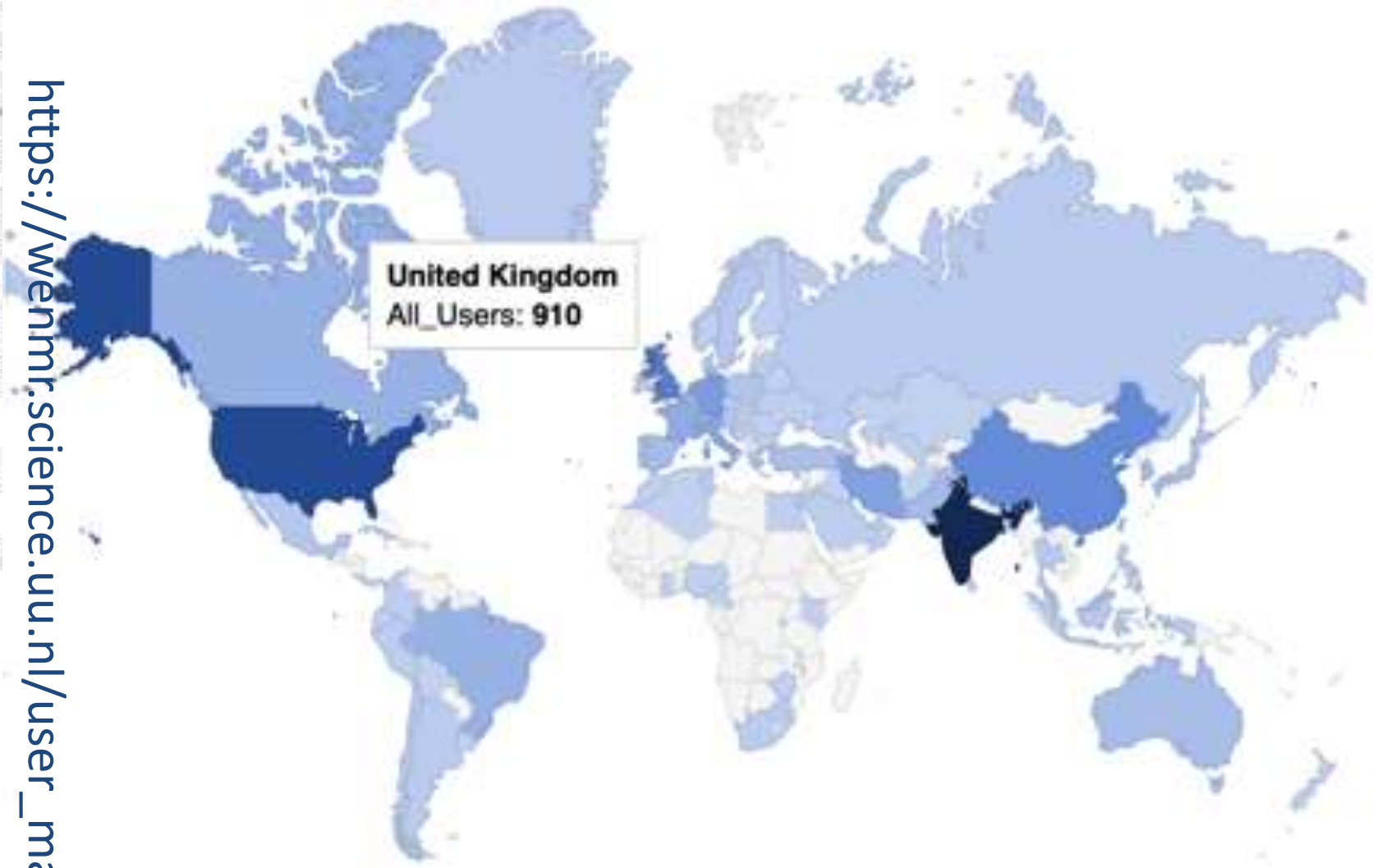
https://wenmr.science.uu.nl/user_map

Sustained growth of our user base

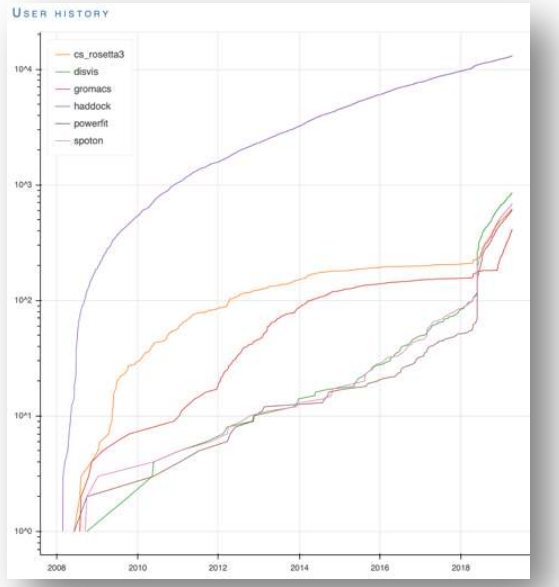


	Country	All_Users	HADDOCK	DISVIS	POWERFIT	SPOTON	CS_ROSETTA3	GROMACS
1	Total Users	14,867	14,253	1,229	901	1,033	856	695
2	EU Users	3,321	3,098	310	181	201	192	131
3	India	3,185	3,128	199	172	211	153	181
4	United States	2,321	2,227	190	116	150	116	76
5	United Kingdom	910	863	64	55	41	57	34

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

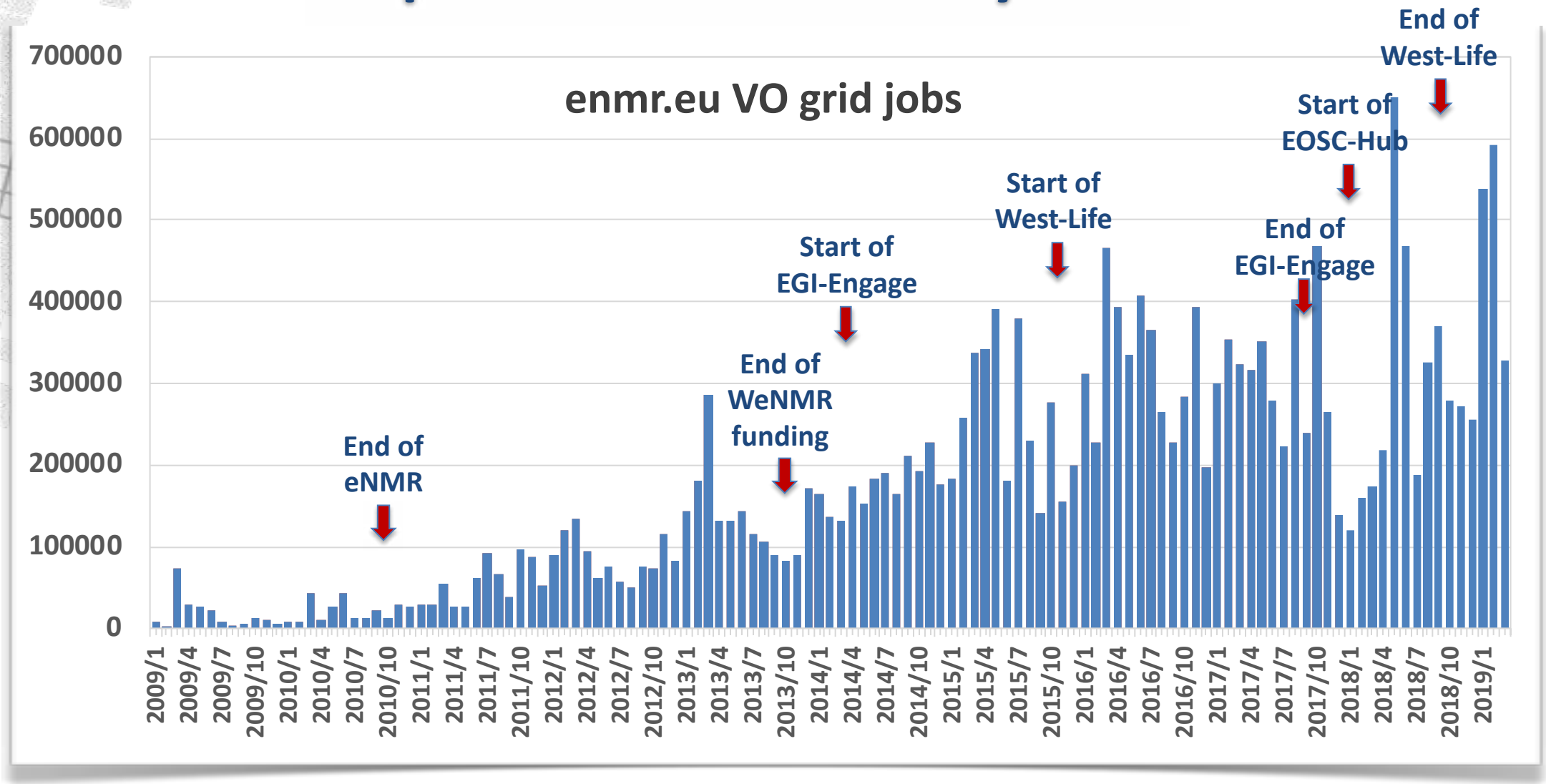


Sustained growth of our user base



	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

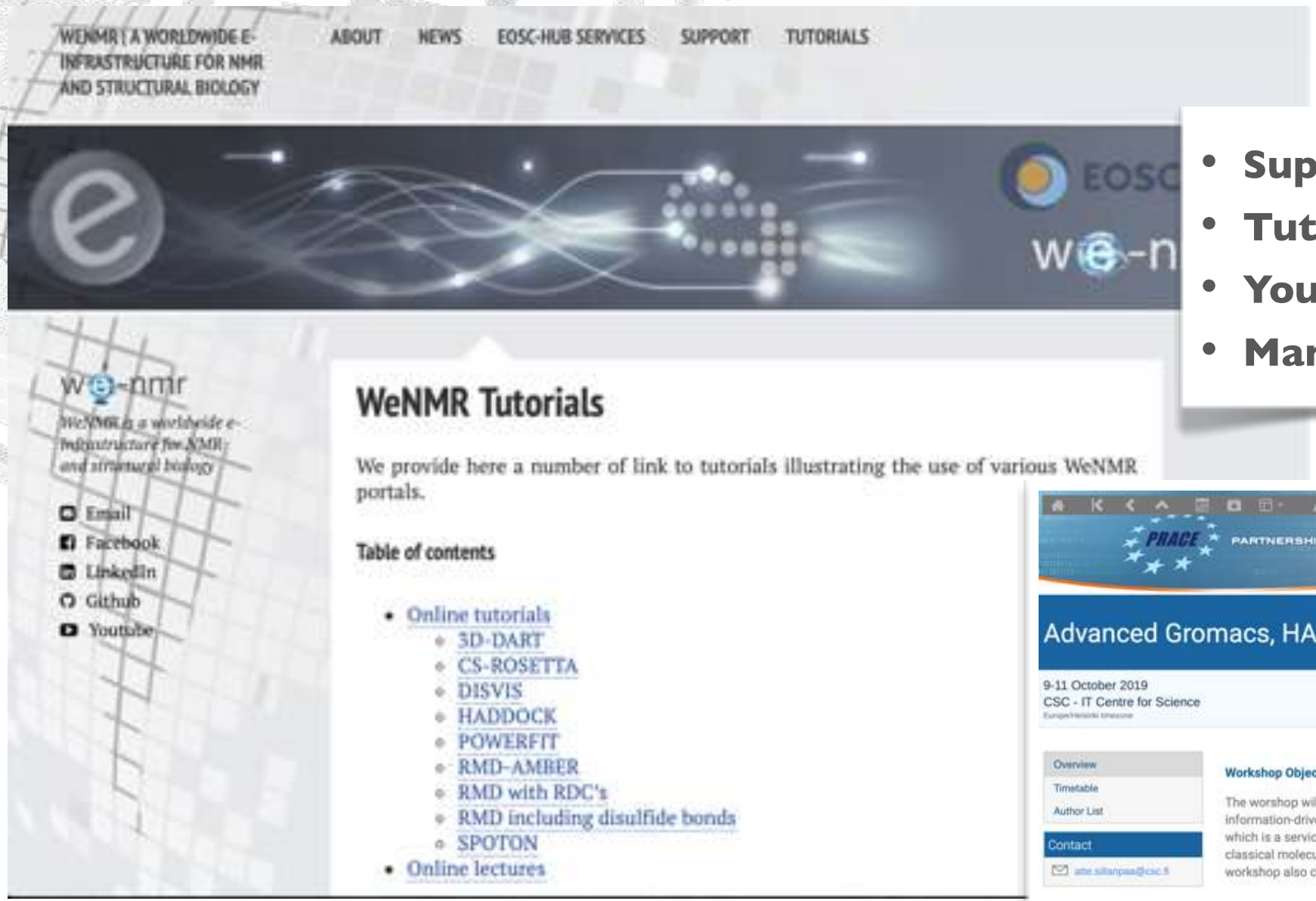
- **Attract users!**
  - **Offer them top of the line eScience solutions for their research ... which means top of the line software**

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



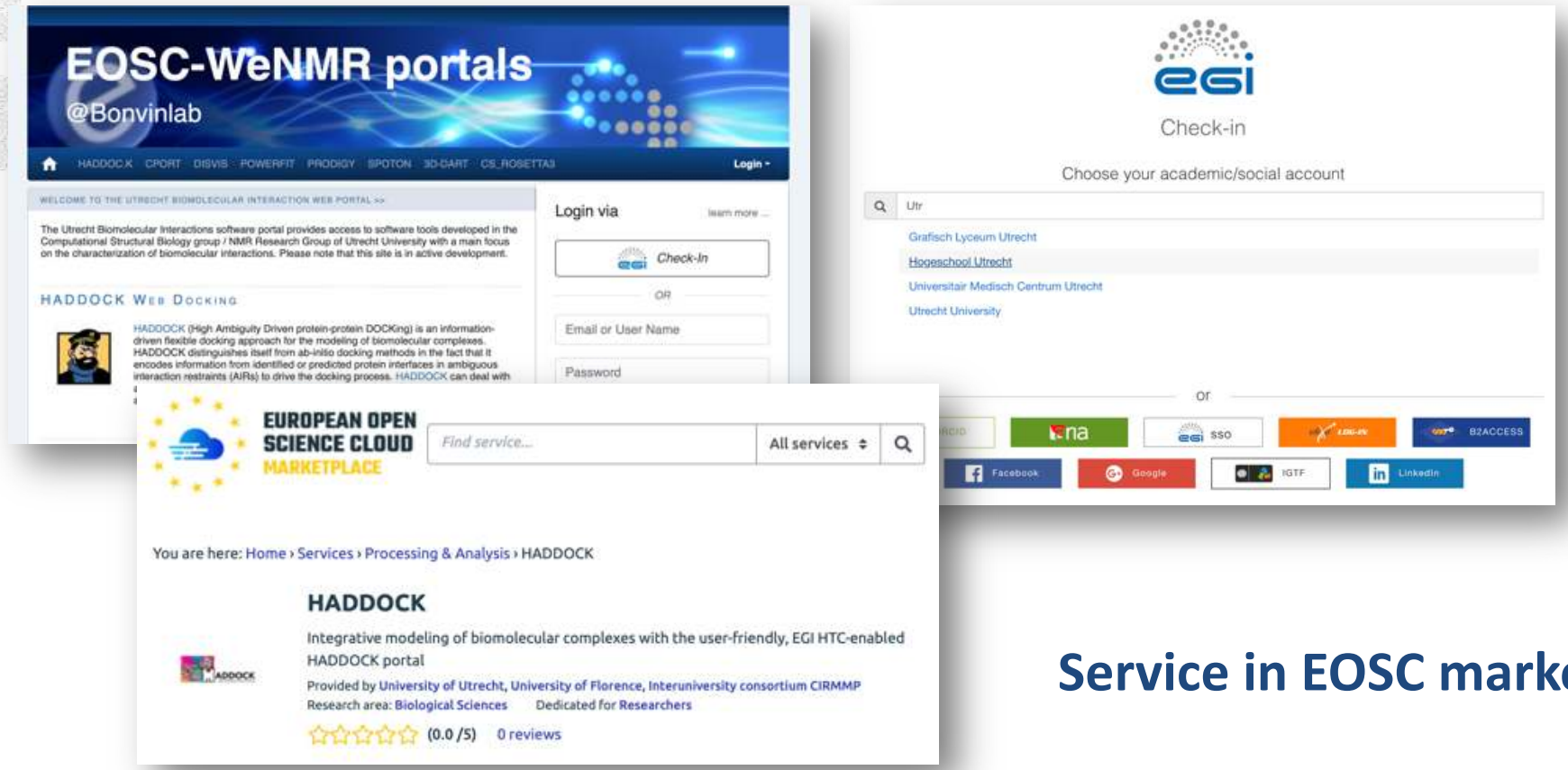
- Support center
- Tutorials
- YouTube channel
- Many workshops …

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**



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	1,730,491	2,323,139	9.81%
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%			

World-wide: Dec

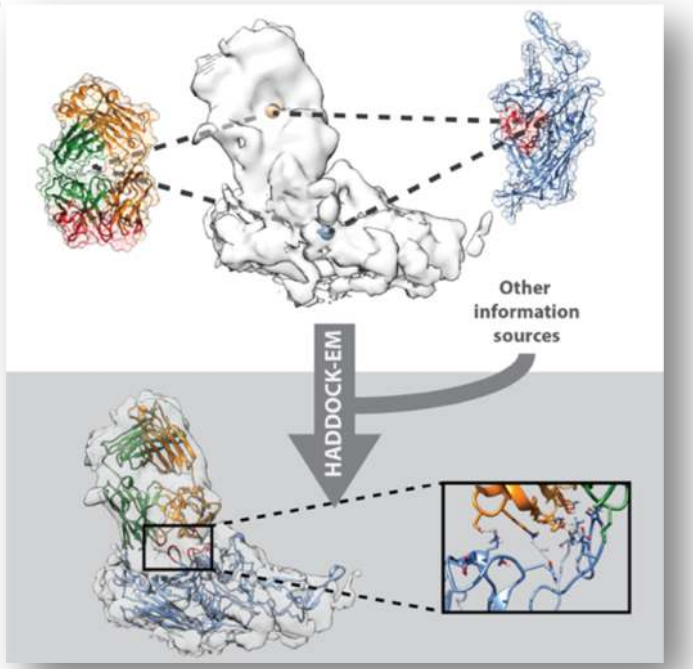
Challenges & e-Solutions

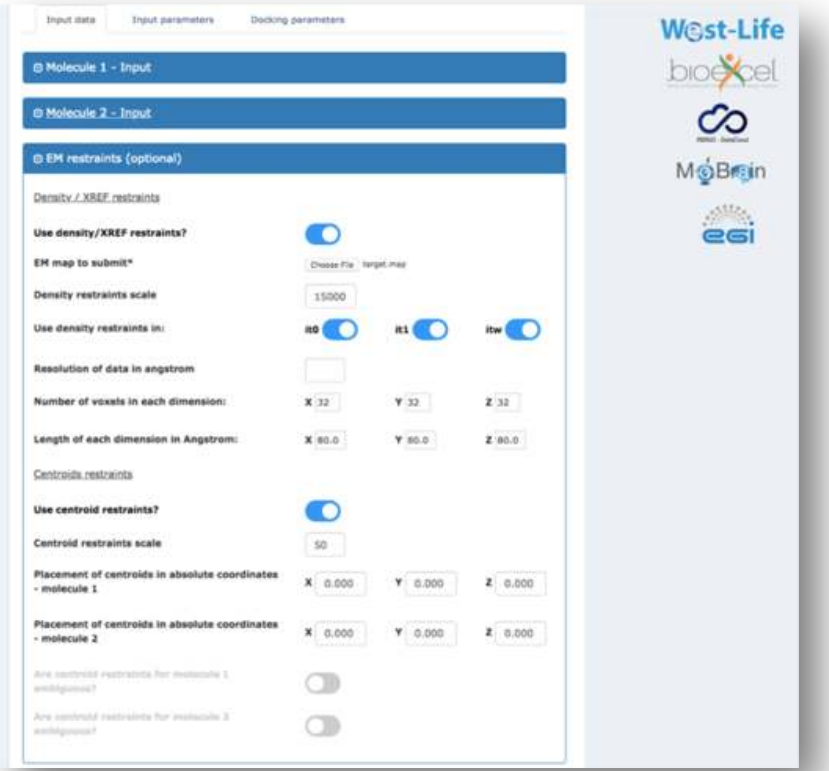
- Attract users!
- Access to e-Infrastructure
- Develop software and maintain and operate a complex infrastructure


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 data[1], cyclic peptides, etc.



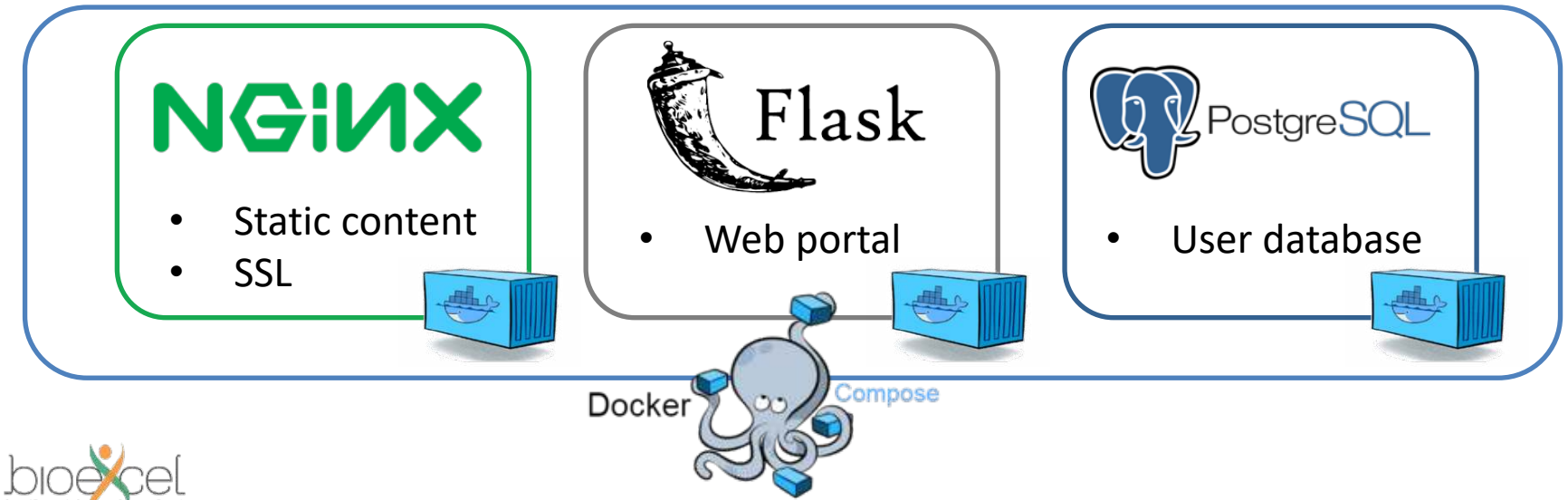


Trellet et al., *Protein-protein modelling using cryo-EM restraints,* 2018, submitted

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




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

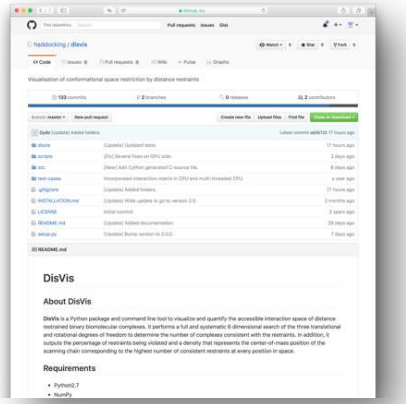
Because of complex software dependencies we use docker containers

- Python2.7
- NumPy 1.8+
- SciPy
- FFTW3
- pyFFTW 0.10+
- OpenCL1.1+
- pyopencl
- clFFT
- gpyfft

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

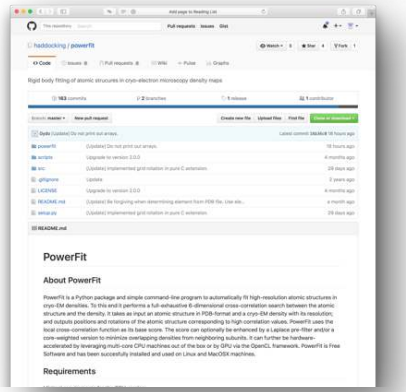
GitHub

INDIGO - DataCloud



docker

indigodatacloudapps/**disvis**



docker

indigodatacloudapps/**powerfit**

SUMMARY SUCCESS FACTORS

SOFTWARE/SCIENCE

MANWARE

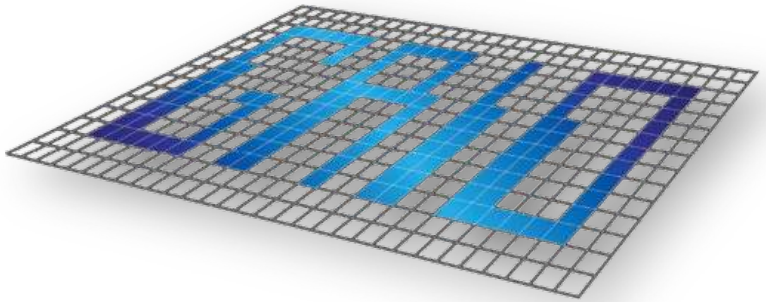
HARDWARE

NETWORKING

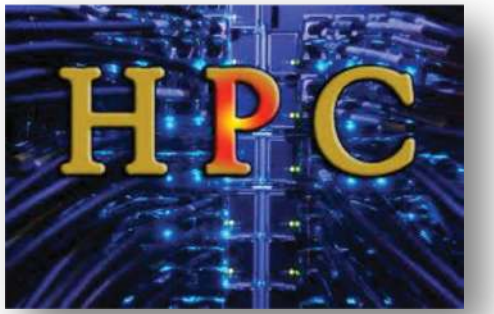
SUPPORT

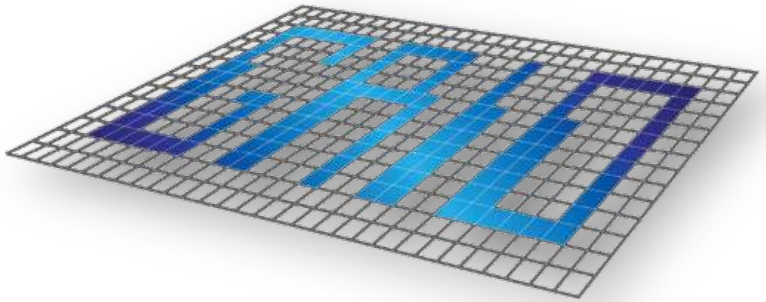
$$$$


Where are we going?

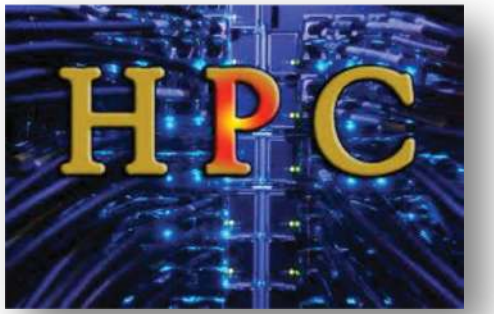


Which solution?

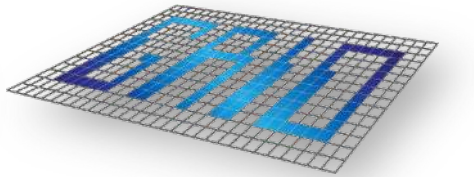




A bit of everything…

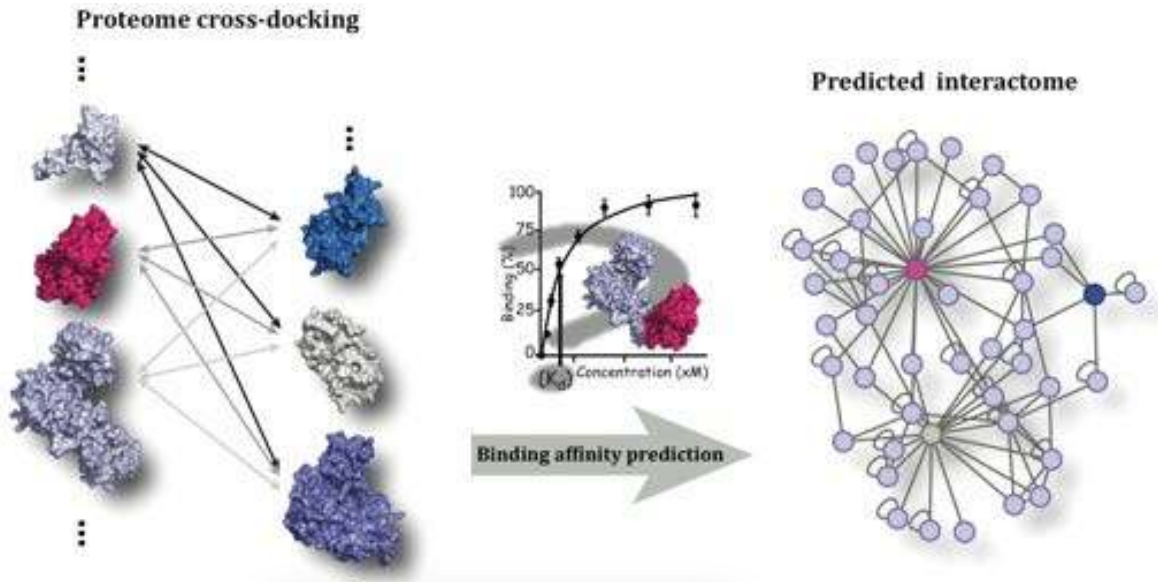


Thematic services under EOSC-Hub

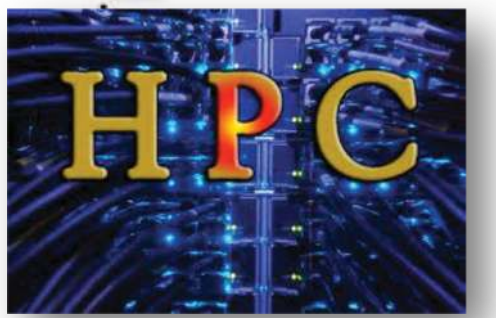
- Harvest transparently both  
- DIRAC4EGI can handle both without the additional burden of managing the cloud VMs
- We still have much more grid than cloud resources
- HADDOCK portal as use case in Helix Nebula Science Cloud

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
Center of Excellence for Computational Biomolecular Research



bioexcel.eu

Acknowledgments:

the CSB group@UU

€€

NWO

VICI
TOP-PUNT

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