The DUTCH e-infrastructure

LIFE SCIENCE COMMUNITY

Establishing life science ICT infrastructure
SURFsara
ELIXIR – NL: big data context

ELIXIR VT, January 16, 2013

SURF SARA

Life Science ICT support – NL

SURF:
- **SURFnet : network, collaboration**
- **SURFmarket : software**
- **SURFshare : innovation**
- **SURFsara : computing, storage**

From January 2013

BiG Grid: computing infrastructure

BiG Grid
the dutch e·science grid

Netherlands Bioinformatics Centre
- **BioAssist and Bioinformatics Research Support**
- **BioRange: science collaboration**

nbic
Ends 2013

Netherlands e-Science Centre: software, interdisciplinary

netherlands
eScience
center

e-BioGrid project

e-BioGrid : BiG Grid e-infrastructure gateway for life science research

When: 1 September 2010 – 31 december 2012
Funding: 2 m€

* BiG Grid computing resources
Enabling life science compute and data intensive research
Selected technology area main projects (2 years), e.g. data a
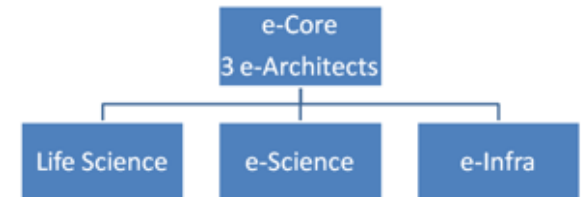Small ad-hoc support projects (2-3 months)

* e-BioGrid human resources (UvA, UL, SARA)
Multidisciplinary e-core coordination team
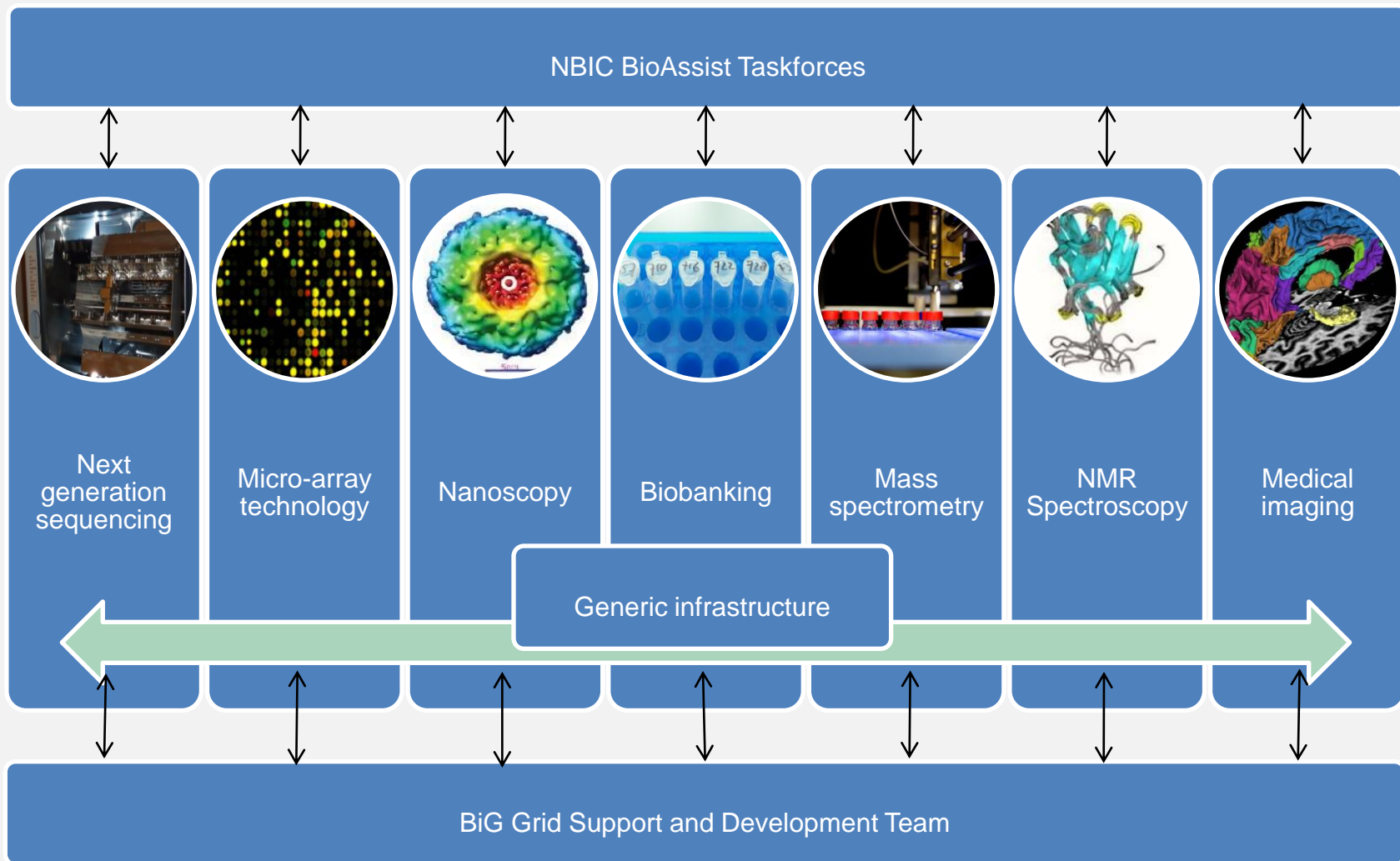e-Science infrastructure support team
Decentralised scientific programmers (located at selected bioscience labs)
An evolving e-bioscience community to build and use the infrastructure

Technology exchange

e-BioGrid: results

- **Community building, e.g. regular workshops and user meetings, newsletters**

- **Growing number of life science users (universities, medical centres)**

- **Life scientists trained to use computing resources**

- **Guiding, monitoring and managing life science e-infrastructure needs**

- **Awareness of needs to plan and budget e-infrastructure**

- **Technology specfic infrastructure, e.g.**

  - **Human genome analysis pipeline**

  - **analyse phenotyping and genotyping (SNP) data from large-scale plant breeding experiments**

  - **Integration of computationally intensive software in a metabolomics data processing tool chain**

- **Generic infrastructure, e.g.**

  - **Life Science Grid Portal**
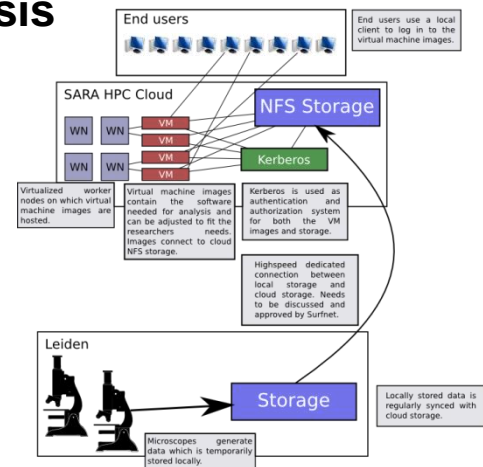
  - **NBIC Galaxy on Cloud**

www.ebiogrid.nl/results

SURF SARA

e-BioGrid: example projects
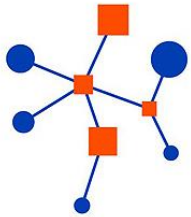
- Analysis of 750 human genomes

  - Genome analysis pipeline running on Grid

  - Using Compute: http://www.molgenis.org/wiki/ComputeStart

- Infrastructure image analysis

  - Authentication

  - Cloud computing

NBIC Galaxy Server on Cloud

e-BioGrid: lessons learned

- **Local infrastructures are sometimes favored**

- **Abundant communication concerning user needs, available expertise and solutions, and progress is essential.**

- **Training is best set up build on use-cases and on-site**

- **Monitoring of resources usage is not straightforward and may be improved**

- **Different user groups require different solutions and support: from pilot to workflow frameworks**

- **Dynamic field of life science research**

  - **requires both structural support for e-infrastructure developments, as well as ad-hoc support for short implementation projects.**

  - **Short-cycle implementation projects often favor cloud over grid computing.**

SURF SARA

e-BioGrid website   info@ebiogrid.nl

www.e-biogrid.nl

SURFsara

Dutch e-infrastructure

- **High-performance computing and storage resources,**

    - The Life Science Grid
    - National supercomputer
    - National compute cluster, GPU's
    - High performance compute Cloud
    - Hadoop
    - Network and light paths
    - Petabytes of disk and tape storage
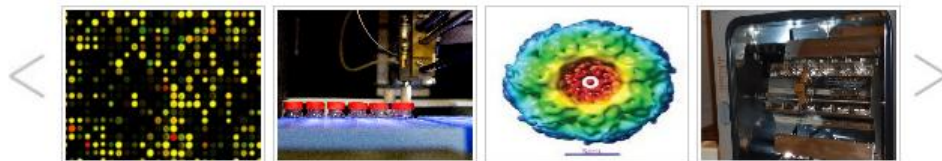
- **Tailored support, training and advice**

SURFsara

Range of services
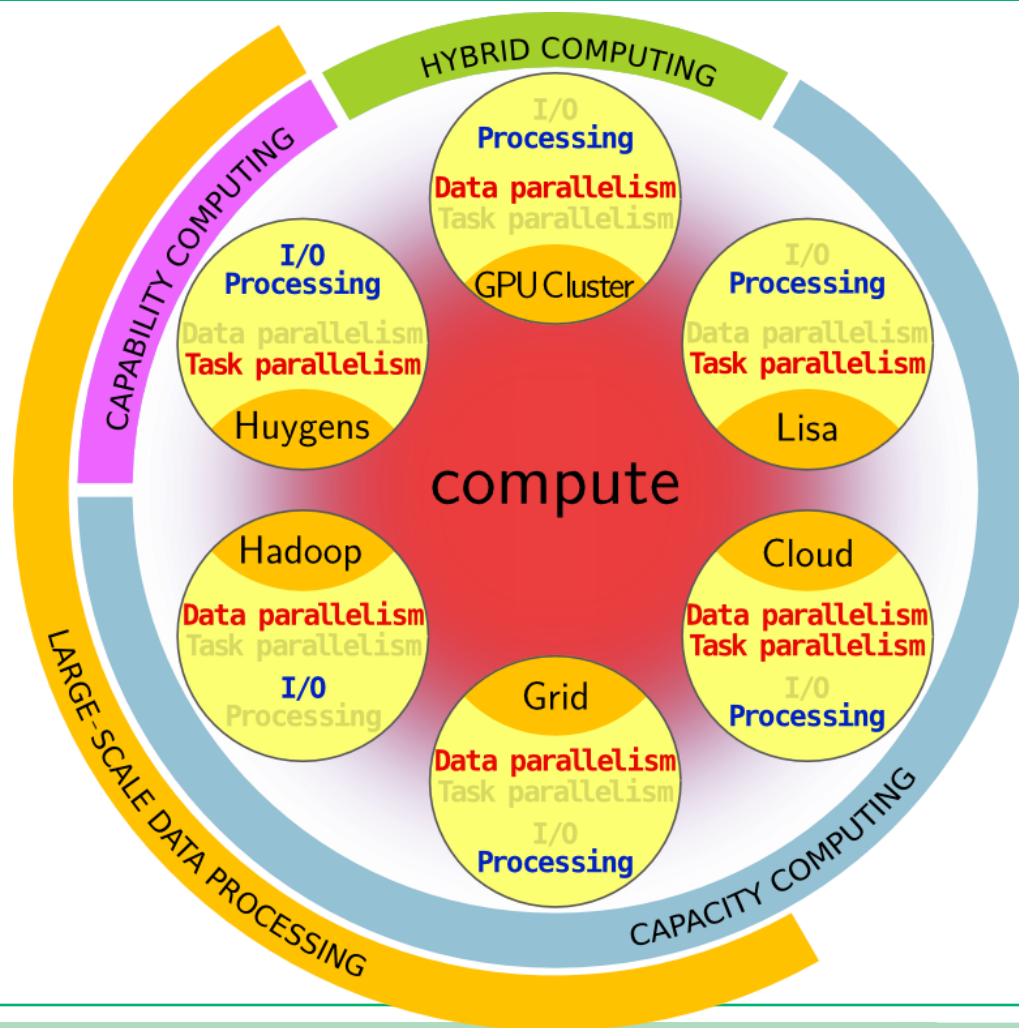
Life Science Grid
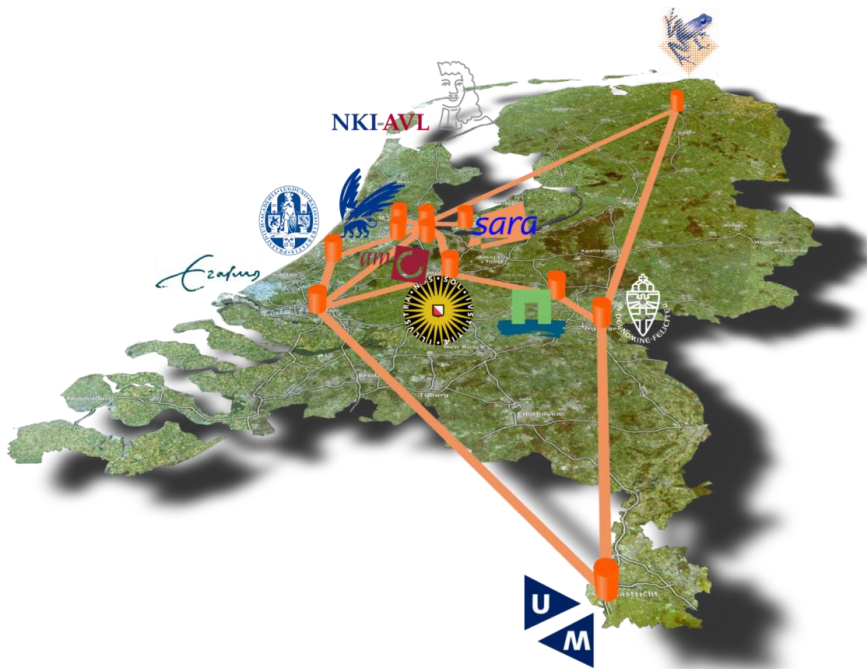
WHAT IT IS

•Several clusters at academic medical centers and universities.

•Open to all life scientists based in the Netherlands.

•Clusters can be used separately or together by using Grid middleware.

•Currently more than 3700 job slots for concurrent processing.

•Distributive storage capabilities and automatic replication of data.

•Maintained and supported by SARA.

Life Science Grid sites
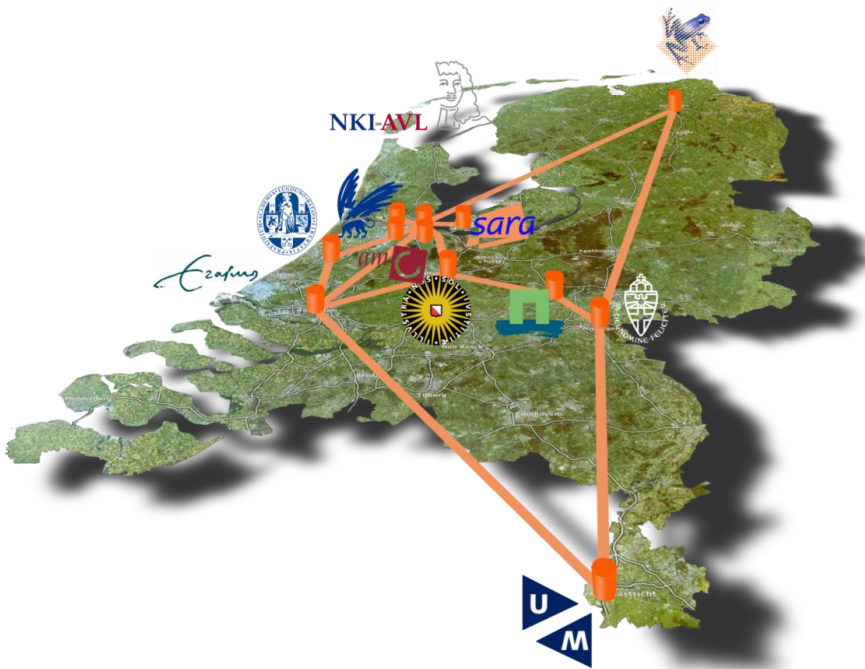
LUMC – Leiden
WUR – Wageningen
UU – Utrecht
NKI – Amsterdam
AMC – Amsterdam
Radboud Universiteit – Nijmegen
ErasmusMC – Rotterdam
Technische Universiteit - Delft
Rijksuniversiteit Groningen
Keygene - Wageningen

Each of these clusters have either 16 or 32 cores and 18 Tb of storage space.

SURFsara HPC Cloud
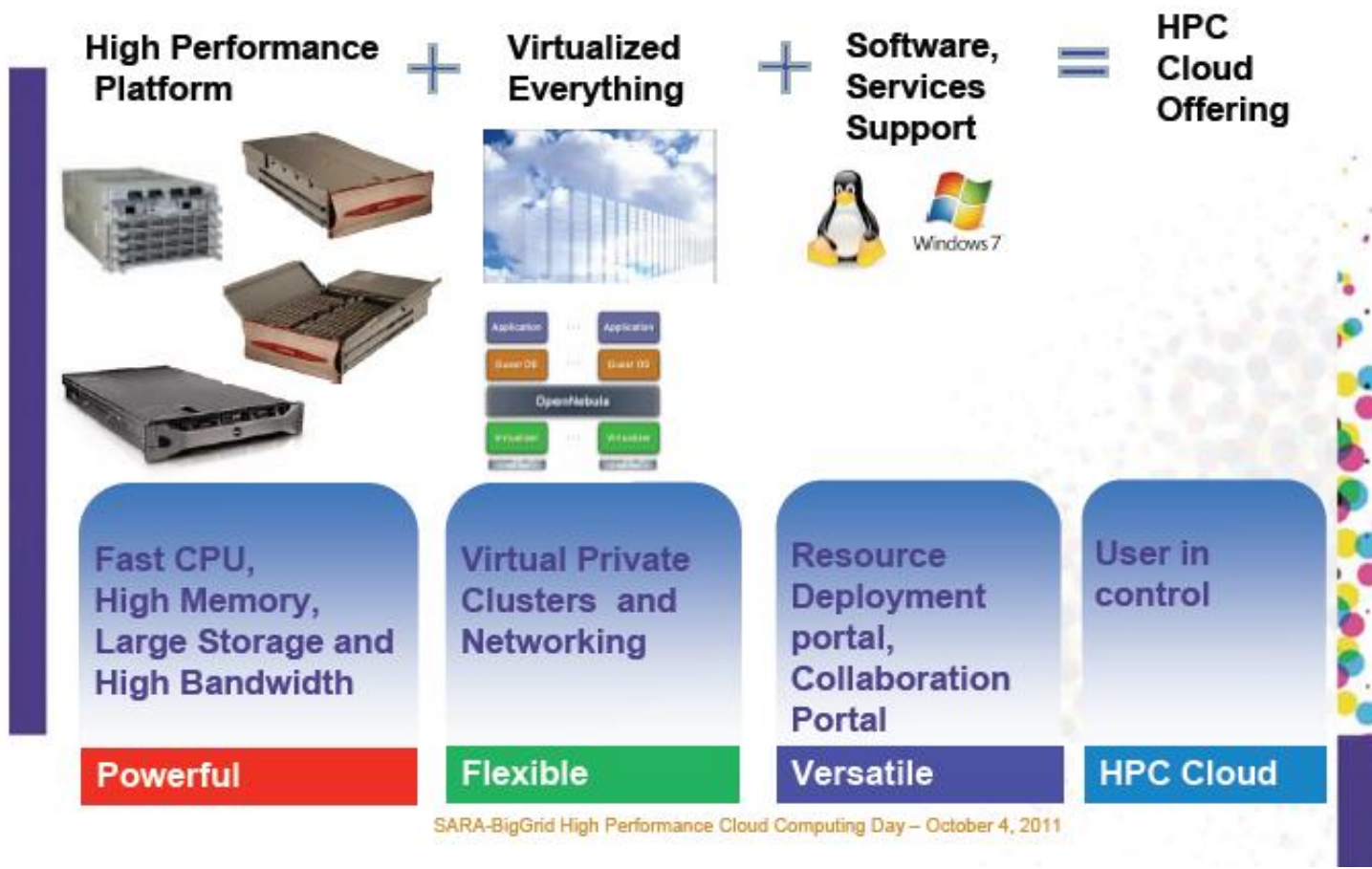
SURFsara

Continues
- **Life science research e-infrastructure support (e-BioGrid)**
- **Life science community building**
- **Development of e-infrastructure based on life science research requirements**

Ongoing:
- **WS-Pgrade install (with AMC)**

We would like to hear from users how they experience WS-Pgrade

- **Upgrade of Life Science Grid**
- **trainings**
- **technology exchange, e.g. Workshop with CSC Finland**

SURFnet : network and lightpaths

ELIXIR - NL

**Compute
Training
Tools
Standards**

DTL: Dutch techcentre for life sciences
DISC: Data Integration and Stewardship Centre

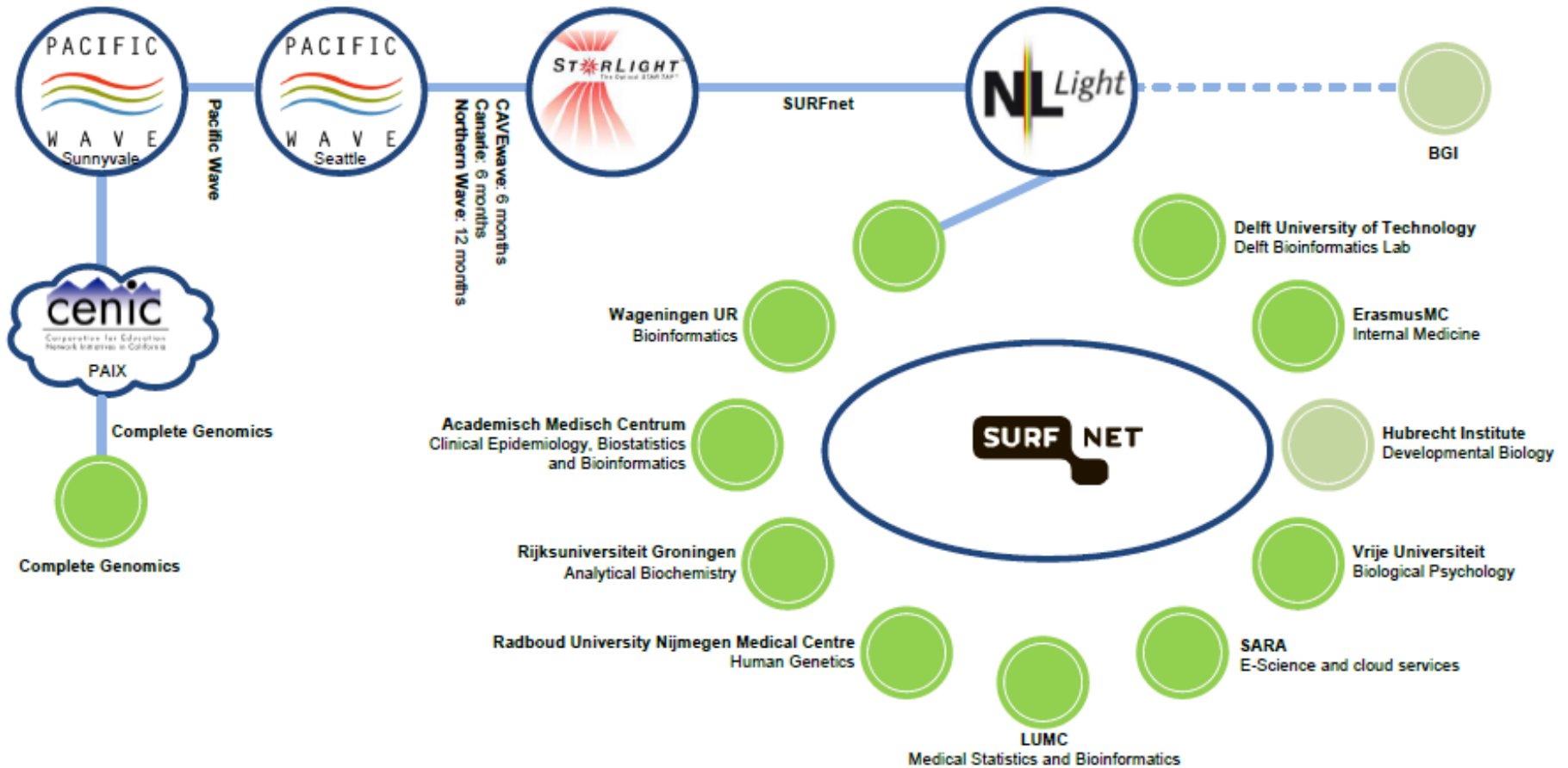Services: SURF, NLeSC, and others

Netherlands Proteomics Centre (NPC)
Netherlands Metabolomics Center (NMC)
Center for Translational Molecular Medicine (CTMM)
Wageningen University & Research Centre
Radboud University Medical Centre
UMC Utrecht
Hubrecht Institute
TNO Quality of Life
UMC Groningen
Netherlands Center of Systems Biology (NCSB)
Leiden University Medical Centre
VU University Amsterdam

..

LIFE SCIENCE IS THE NEXT BIG SCIENCE

HIGH TECH - 'BIG' SCIENCE

discovery & innovation depend on advanced technologies and on cross tech collaboration

easy and cost-effective access to key technologies is essential

HIGH VOLUME - 'BIG' DATA

data volumes double every 9 months

data integration needed to understand complex biological systems

progress needs data sharing & open access of public data

strategy needed for an integrated technology & data infrastructure

DISC Services

Multi & Cross Technologies

e.g.

'omics'

Biobanking

imaging

Data Design

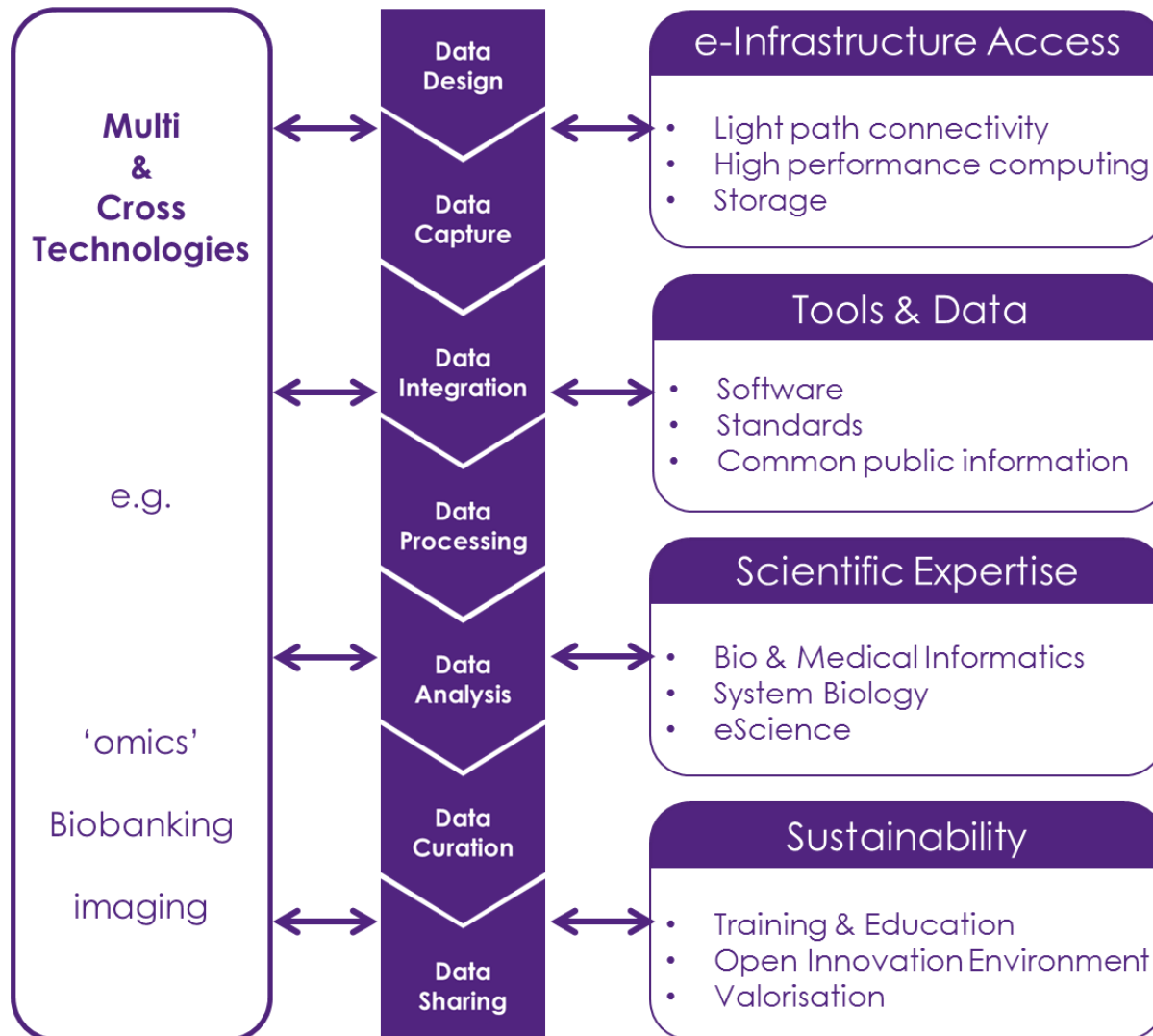**Data Capture**

Data Integration

Data Processing

Data Analysis

Data Curation

Data Sharing

e-Infrastructure Access

- Light path connectivity
- High performance computing
- Storage

Tools & Data

- Software
- Standards
- Common public information

Scientific Expertise

- Bio & Medical Informatics
- System Biology
- eScience

Sustainability

- Training & Education
- Open Innovation Environment
- Valorisation

SET UP PPP CONSORTIUM BASED UPON EXISTING NETWORK

ELIXIR - NL
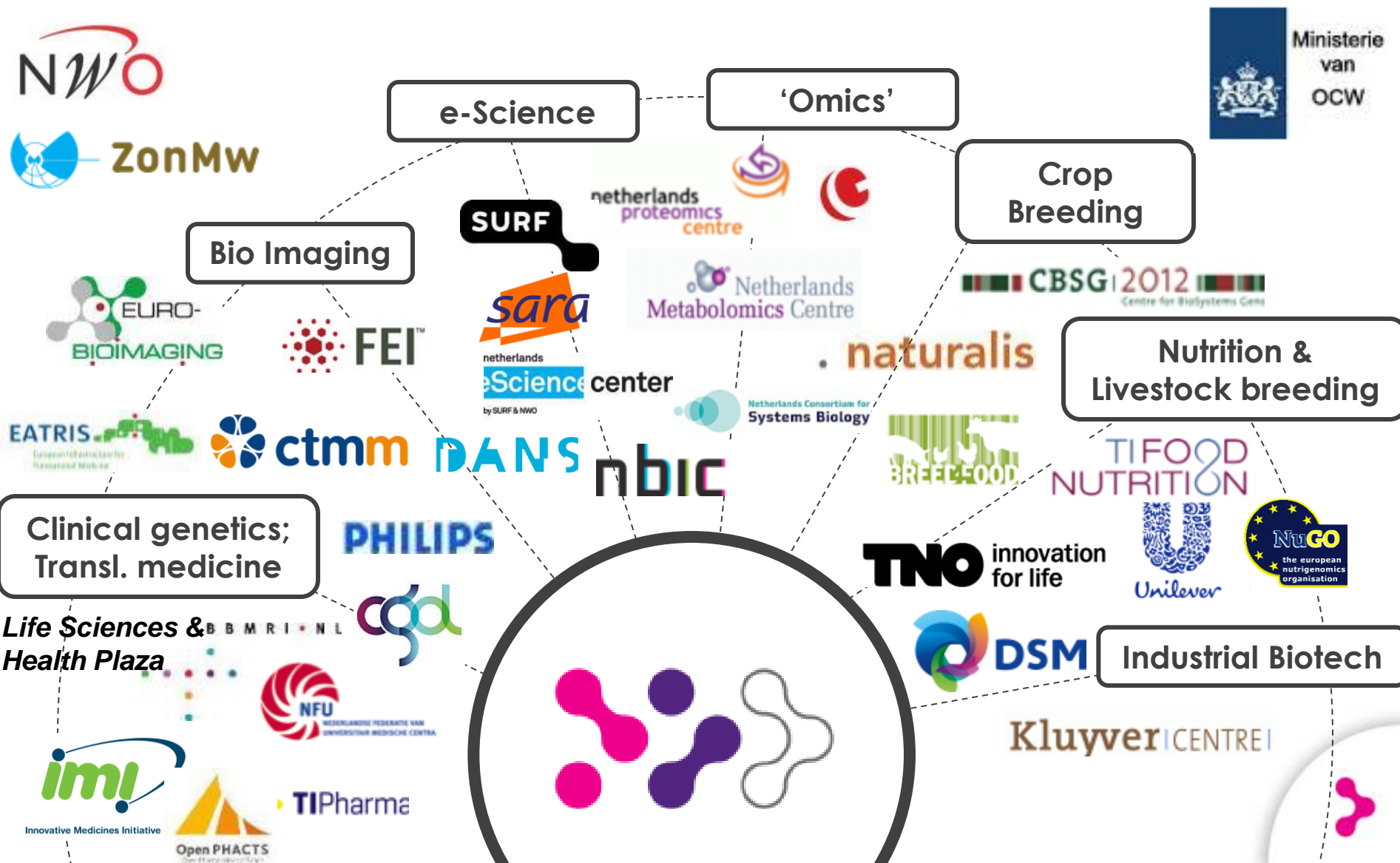
- **Continuing life science ICT infrastructure support: SURF, NLeSC, and others**

- **Establising DTL-DISC**

- **Providing affiliated, and possibly commissioned services**

- **Head node workshop, March 21-22, Amsterdam**

- **April 2013, final recommendation of the ELIXIR Interim Board on Node Application.**

Contact:
irene.nooren@surfsara.nl
(i.m.a.nooren@uva.nl)

www.surfsara.nl

See you at EGI community forum, April 8-12?

WHAT SURF CAN DO