



ELIXIR DK

Danish National Supercomputer for
Life Sciences



European Life Sciences Infrastructure for Biological Information

www.elixir-europe.org

National need for life science supercomputer

Approx. 2% of Denmark's GDP is related to Lifescience

- Large pharmaceutical sector (foundation based)
- Historic tradition for PPP's in research
- National strategy for HPC
- Large user community

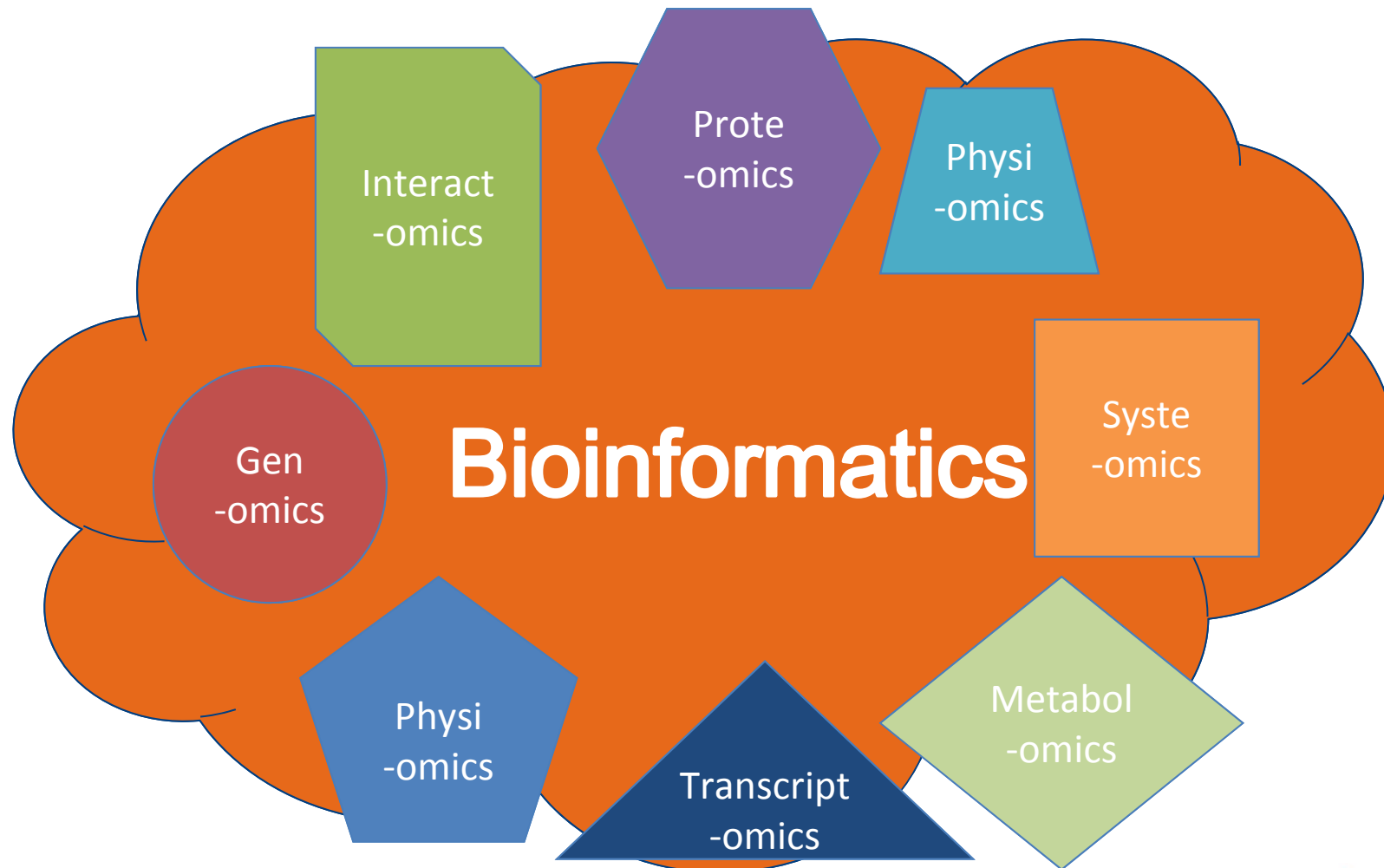


Life science data deluge

- **Massive unstructured data** from several areas DNA, patient journals, imaging, Omics, **Lifestyle** ...
- Impacts Industry, Environment, Health
- Societal grand challenges
- Cheap sequencing technologies results in **explosion of DNA data**



Background: omics data



Background: Heterogeneous data types

One size doesn't fit all.

Each “-omics” needs **at least two** of:

- Very high single core performance
- Very high node and multi-node scalability
- Very fast disk read/write
- Large amounts of storage per project (30TB+)
- Large amounts of RAM (1TB+)

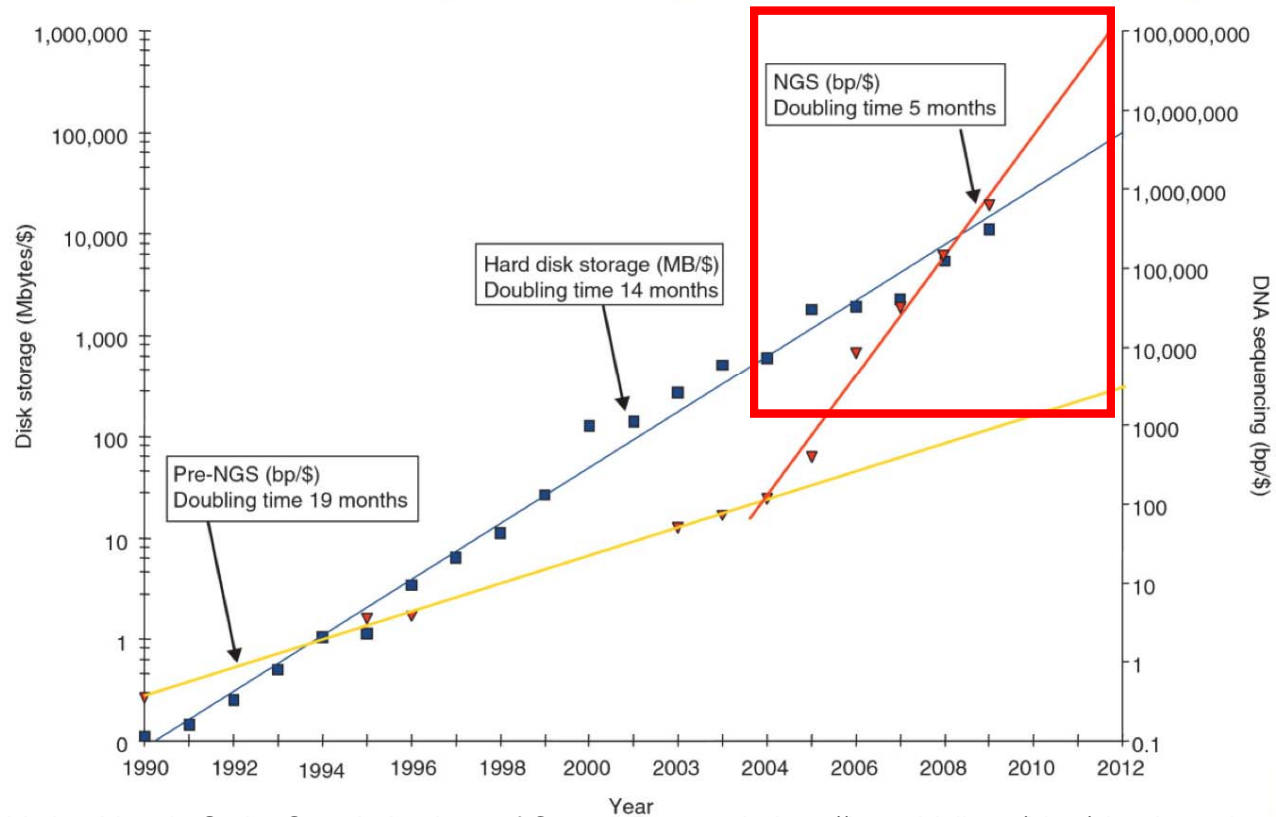
It is referred to as “*omics onslaught*”



Background: Growth

Only one thing in common: GROWTH

NextGen Sequencing a Game-Changer



Graphic by: Lincoln Stein, Ontario Institute of Cancer Research, <http://ivory.idyll.org/blog/cloud-not-the-solution.html>



HPC Challenge: National Life sciences Center

Number of cross-mounted file systems: 16+

Number of actively used software tools: 180+

Number of jobs in the queue: 20.000+

Diverse user profiles: 200+

Turnaround time for life sciences applications rapidly going down

Policy challenges:

- Data needs to be backed up in multiple copies.
- Disaster recovery time
- Ensuring software compatibility
- Data access, security and sensitive data



Lessons Learned

1. Storage – it's not only about size:
 - It's about throughput, authentication and manageability

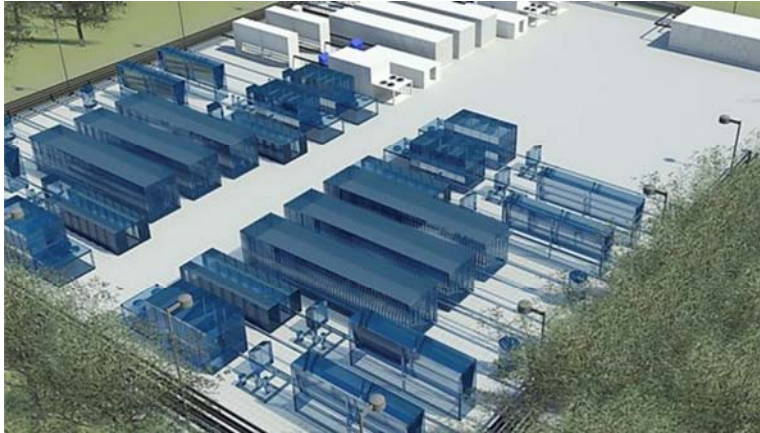
2. Compute – it's not about the “GHz”
 - It's about Scalability and I/O

3. Access control – it's not about a user
 - It's about groups and ease of sharing
 - It's about user-defined environments

4. Integration – it's not about building an island
 - It's about integration with other systems
 - Industry \rightleftharpoons Research \rightleftharpoons Healthcare
 - Relevant = accessible



Computerome #121 Top500: Summary



Computing Power

- 525 x HP XL230
 - 2 x 14 core Intel Haswell
 - 128 GB DDR4
- HP Apollo 6000
 - 36 x HP DL560
 - 4 x 8 core Intel Ivy Bridge
 - 1TBDDR3

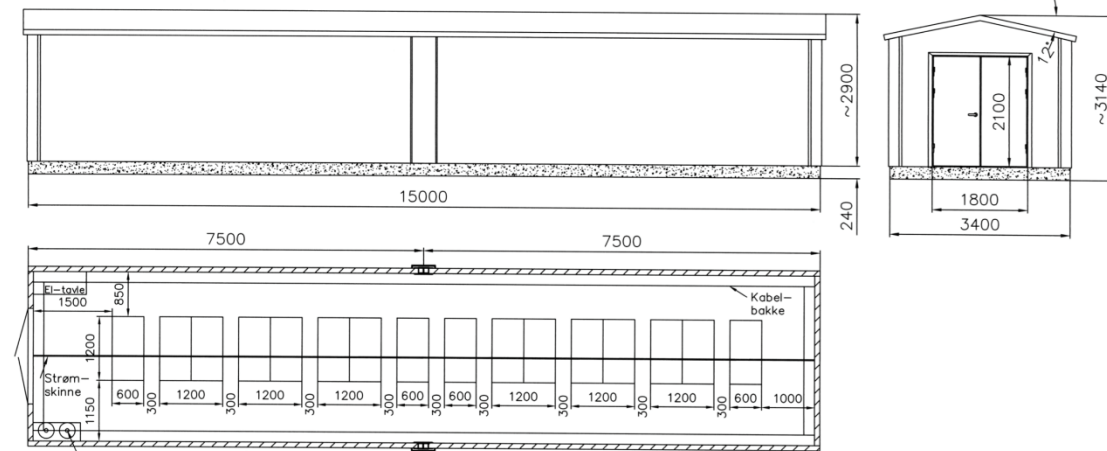
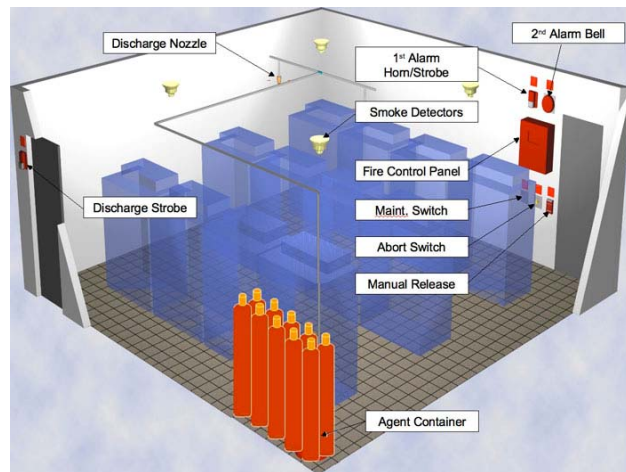
Storage

- 3 PB of highly available storage
 - +75GB/s aggregated bandwidth
 - SSD's for fast caching
 - fully backed up at two physical locations
 - Data in 4 copies



Green field facility

Made-to-fit facility – 7 Mega Watts



- Supporting system growth **10x** of the new system
- **Redundancy** in every single support system
- Heat **re-used** for heating the Town of Roskilde







Cloudification of Life Science applications and data

On the Danish national life Sciences
supercomputer.



European Life Sciences Infrastructure for Biological Information

www.elixir-europe.org

National Life-sciences supercomputer

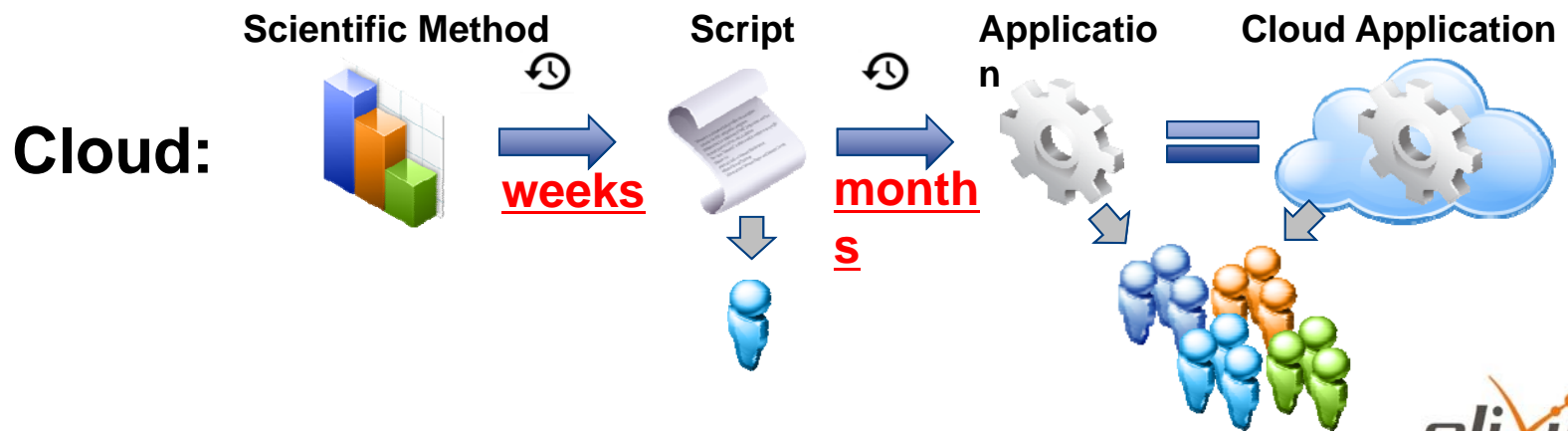
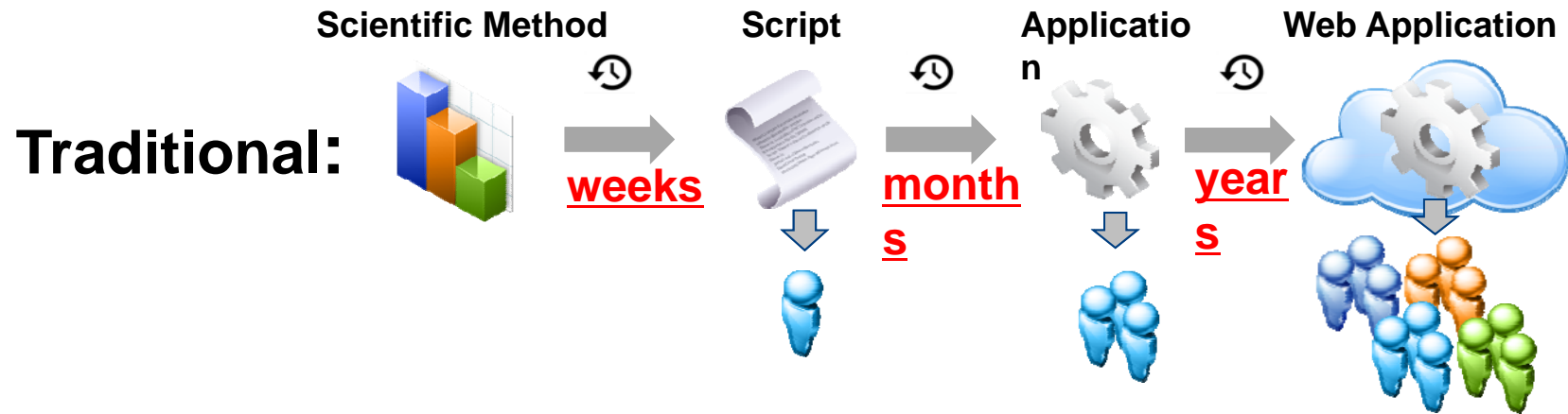
Elixir National Life science Cloud

Enterprise-grade service allowing for secure provisioning of Hybrid HPC environments, secure & reliable execution of life sciences workflows and access to curated data sets.

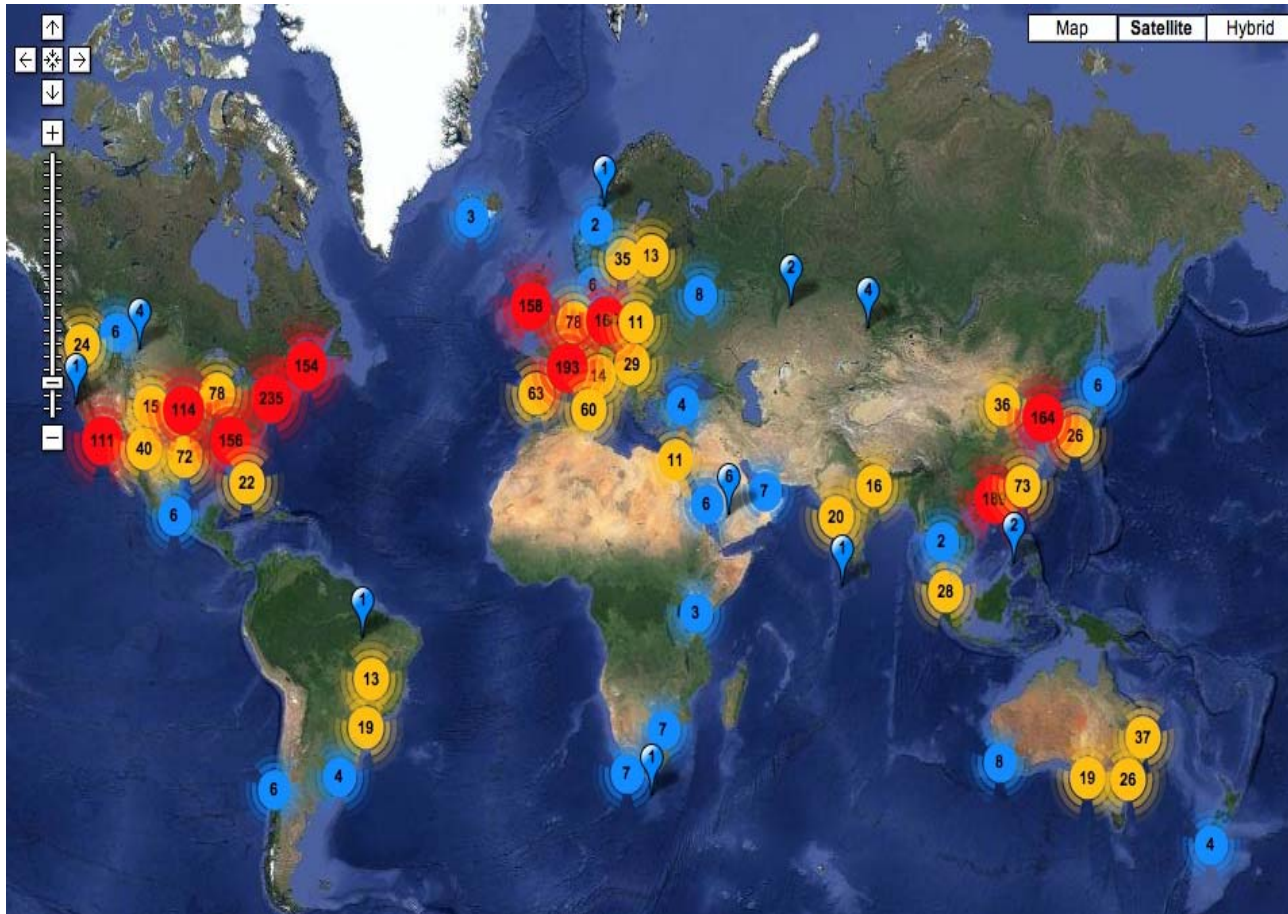


National Life-sciences Cloud

New workflows



Distributed data production



World wide >900 centers

>90 PB/ year

Data transfer and storage becomes an issue

GenomeDK (KU, DTU, AU)
We ship hard drives...

<http://omicsmaps.com>



National Life sciences Cloud

National Life Sciences hybrid cloud provides a collaborative infrastructure

- Cloud for Bioinformatics, Healthcare and Industry partners
- Provide setup for collaborative research on sensitive data
- KU, DTU, BGI, Rigshospitalet, Industry partners
- Bioinformatics support



New Life Sciences Services model

To unleash the power of data in the Cloud

Bioinformatics as a Service	Enabling Healthcare transformation
Analysis as a Service	Marketplace of high value consumable Bioinformatics services
Platform as a Service	Standardized and integrated application development platform and execution
Infrastructure as a Service	Enterprise class, optimized infrastructure

Public– Private - Hybrid



National Life-sciences Cloud – select use cases

- Tryggve
 - An international platform for collaboration on sensitive data and to utilize and connect existing capacities and services at the Nordic countries.
- BIO for Industry
 - The Embassy Cloud is a service that provides secure, flexible infrastructure to tenant organizations such as industry, close to the national life science data & resources.
- IBM Watson Explorer
 - To identify dependencies between drugs, patient symptoms and side effects based on the patient data and global medical journals in different languages impossible to mine by one single user



Watson video

- https://www.youtube.com/watch?v=Y_cqBPo8yuA