

Platform of eScience and Data Research Centers in Europe

e-IRG Workshop Rome 10-11 November 2014

Patrick J.C. Aerts

PLAN-E:

European co-operation of eScience centers

netherlands

eScience center

by SURF & NWO

The world around us

- **Science and society are intimately connected**
 - **Science becomes increasingly problem-driven**
 - **Science increasingly inter-, multi-, trans-disciplinary**



Mission

- **Enabling digitally enhanced research through efficient use of scientific software, data, and e-infrastructure**



Application Domains

**Life Sciences & eHealth, Environment & Sustainability,
Humanities & Social Sciences, Physical World & Beyond**

netherlands eScience center

e-Infrastructure

**Computing, Networking
Storage & Visualization**

Scenery: eScience in The Netherlands

- NLeSC=Netherlands eScience Center
- NLeSC has a national and coordinating role
- University and institutes have their own Data Research Centers
- Platform of escience/data research centers in the Netherlands

netherlands

eScience center

by SURF & NWO



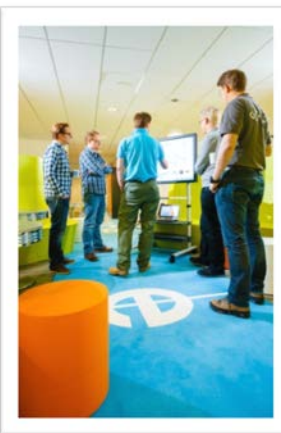
Netherlands eScience Center = *enhanced* Science

To reinforce and accelerate multi-disciplinary and data-intensive research by developing and applying eScience methods and approaches.



enhanced Science is about promoting new scientific breakthroughs and innovation by bridging scientific disciplines via ICT

Undertaking science only possible using advanced ICT



netherlands

eScience

by SURF & NWO



The Collaboratorium

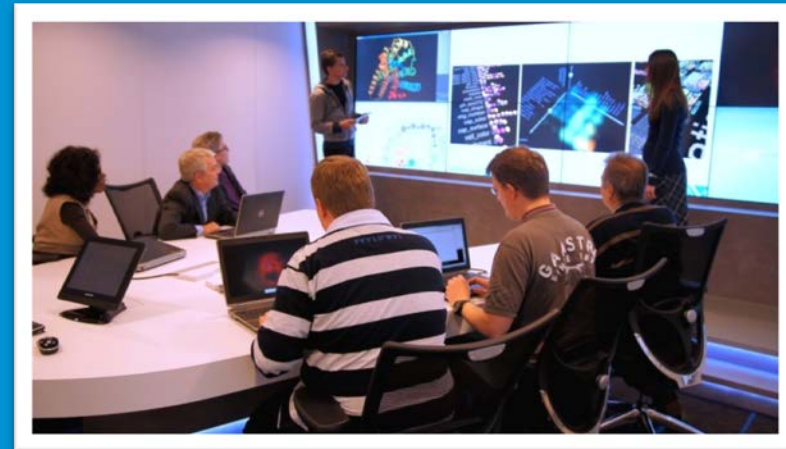
Collaborative environment, supported by High Resolution Data representation capabilities

Stimulate a culture of knowledge sharing & collaboration.

Big Data: challenging data complexity and fragmentation.

Put the scientists together with access to data and tools for its analysis.

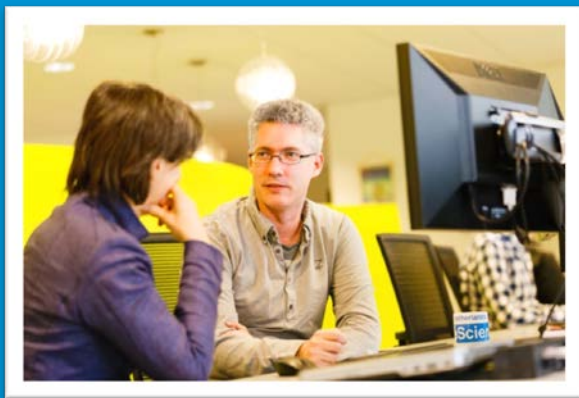
Modern “data-common” to stimulate best practice in scientific project working.





eScience Research Engineers = Digital Scientists

- broad oriented scientists at the interface of research and IT
- collaborating with domain researchers to implement eScience concepts and tools
- mostly PhDs with domain knowledge and IT skills
- Involved in projects, funded in cash and in kind
- As an applicant expect an academic scientist, highly skilled in ICT, use of e-infrastructures, able to design, adapt, code but who speaks “your language” as well



eScience Integrators = Trail blazers

- **eScience integrators:**
 - externally employed network of leading researchers from a variety of disciplines
 - mostly professors or institute leaders
 - broad experience and understanding of the possibilities of eScience.
 - playing a key role, for example by opening up access to and combining data and creating new types of collaboration between different disciplines
 - lead partners for many of NLeSC's collaborative project
 - our voice in the domain world



Biographynet



Ontologies for Food text mining



Visualization



eWater Cycle

Biochemical Network Reconstruction



Bio marker Boosting

eScience methodology



Twitter analysis



Dr. WATSON goes medical



Beyond the

Massive Point Clouds



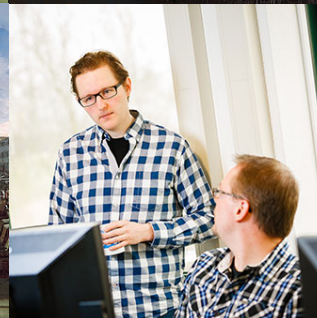
Virtual lab Plant breeding



Searching public discourse



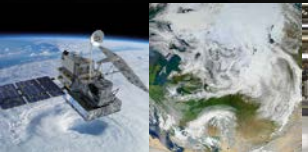
HADRIANVS



Via Appia in 3D



eEcology



SIM CITY

netherlands

eScience center

Co-ordination

- NLeSC is also concerned with coordinating activities in its domain
- Initiatives in data stewardship, software sustainability
- Initiatives to bring people together and mobilize their common strength, by combining forces

NL Platform for eScience

- A platform is formed by the Netherlands eScience/Data Research Centers at universities and research institutes
- The Platform:
 - Shares a common definition of escience
 - Defines common grounds for co-operations
 - Turns individual escience/data research groups and organizations into a strong collective and federated movement
 - Has an extensive to-do list

But first: anyone remembers ARCADE?

- ARCADE= Advanced Research Computing Academic Discussion group Europe
- It was the first (1995) in its kind of a community building effort for HPC, from an organizational, point of view
- Funding and policy agencies, HPC computing centers, etc.
- Many members from across Europe, sharing similar goals, with a sense of urgency
- Website, documents, database, “Overview of Academic Supercomputing in Europe”, annual report
- Database and annual report carried over to the e-IRG and forms now the heart of <http://knowledgebase.e-irg.eu/>
- Activities resolved into the e-IRG +Support Program

Implementation of PLAN-E

- Why a new entity in Europe?
- What and how?

Why?

- There are many reasons, but let's pick a few
- e-Infrastructures remain extremely important, but their complexity goes way beyond what scientists in application domains can really (or should be supposed to) oversee
 - Grids, a variety of clouds, HPC in many flavors, Clusters, GPU's
 - Mobile
 - Visualization
- ICT tooling from the domain of informatics and computer science continuously yield many new products
 - “Jungle computing”, data base technology for real time data,
 - analytics, pattern recognition
- Data come from many new resources and in abundance:
 - Twitter resources, other social media
 - Sensors in a variety of domains, fixed and mobile
 - Scientific research equipment
 - Previously undisclosed or inaccessible resources

More why

- **Attract attention:**
 - for the new dimension in science: escience, largely, but not exclusively driven by data
 - for innovative ways to do science where domains cross, using whatever is available in ICT tooling and software and in (national, European, Global) e-infrastructure
- To get support at large for the development of centers that develop the new skills required for bridging the gap between ICT/e-infrastructures and science at an academic level
- To form a community of entities that can bi- or multilaterally co-operate in future H2020 and other calls

More why

- High level support requires an academic skills level (PhD) by people that:
 - Understand the scientist's questions
 - Can help to articulate those questions
 - Can help to translate questions into the efficient use of available tools and hardware
 - Can help overarching disciplines (re-use of methods and tools)
- eScience/Data research Scientists should become *recognized* as academically skilled persons and their output recognized as important prove for further continued academic careers

What?

- PLAN-E is a new European Platform (a network) of organizations responsible for/strongly involved in eScience/Data Research
- It 's goals are
 - To strengthen the position of data research and escience as a community of practice
 - To bring forward the importance of re-use and the re-usability of software, tools and methodologies across disciplines
 - To focus attention on the importance of layer between ICT and e-Infrastructures on the one hand and the science domain on the other, closely together with the scientists
 - To aid discovery in science that without innovative use of software and tools would not have been possible

How?

- Basically by bundling all forces across Europe that share these goals;
- By forming a platform at the organizational level
- For exchanging knowledge and expertise in the field in order to strengthen the European position in the escience domain;
- With a solid, yet light weight organization
- Based on Terms of Reference
 - Based on voluntarity
 - With a kernel group of active/leading members prepared to engage in the activities
 - With goals, targets and clear communicatio
 - Work with annual plans and targets
 - Open to European organizations that share the goals

Next step: a European Platform

- **Agree on Terms of Reference**
 - Defining escience/data research domain
 - Defining the scope of activities, based on shared common interests
 - Form a kernel group of active/leading members prepared to engage in the activities
 - Work with annual plans and targets

Terms of Reference

- Address the new dimension in science: escience, largely, but not exclusively driven by data
- Address innovative ways to do science where domains cross, using whatever is available in ICT tooling and software and in (national, European, Global) e-infrastructures

Before we dive into the terms:

What PLAN-E is not:

- PLAN-E does not service/host e-infrastructures
- PLAN-E adds value to the existing scene but is not led, governed or dominated by any specific e-infrastructure (provider)
- PLAN-E is not concerned with computer science (informatics), ICT development as such
- PLAN-E is not concerned with services close to the e-infrastructure
- PLAN-E is not even closely similar to the e-IRG

eScience is a community of practice characterized -but not exclusively or limitedly- by the following: (1)

- It is concerned with innovative ways in which ICT can *be applied* to complex scientific or industrial problems;
- It is concerned with the support of multi-disciplinary research, for example through but not limited to cross-type data integration, the managing of structured and unstructured data sets, data-driven research (“Big Data” research) and data analysis;
- It is the application of computer technology to the undertaking of modern scientific investigation, including the preparation, experimentation, data collection, results dissemination, and long-term storage and accessibility of all materials generated through the scientific process;
- It applies computer algorithms and tools for the interactive specification and maintenance of models and their analysis, visualization and simulation, in order to support scientific *in silico* experiments;
- Bohle, S. "What is E-science and How Should it Be Managed?" Nature.com, Spektrum der Wissenschaft (Scientific American), http://www.scilogs.com/scientific_and_medical_libraries/what-is-e-science-and-how-should-it-be-managed/.

eScience is a community of practice, characterized -but not exclusively or limitedly- by the following: (2)

- It is concerned with the optimal use and/or optimization of the use of larger parts of an e-infrastructure for scientific applications for complex scientific problems and/or
- It is concerned with the optimal use and/or optimization of the use of (high-end) computers in scientific applications for complex scientific problems and/or
- It is concerned with addressing scientific usage of computers and/or e-infrastructures in cases where the problems may for example be one of the following or a combination of these
 - Compute bound
 - Data size bound
 - Data streaming bound
 - Data complexity bound
 - Latency bound
 -
- In general it is concerned with the *application*, *re-use* and *re-usability* rather than the (from scratch) *development* of ICT methods, methodologies and tools to support solving complex scientific and/or industrial problems;
- It encompasses and advocates strongly advanced visualization and pattern recognition in support of its goals.

Goals and Action lines (1)

- Forms a forum for exchanging knowledge and expertise in the field in order to strengthen the European position in the escience domain;
- Discusses common approaches to escience;
- Communicates about escience and the way it is showing results in all disciplines;
- Represents the European escience scene as embodied by the PLAN-E community externally and internationally in addition to the individual representations from the participating members where applicable. In particular towards the EC in relation to future funding schemes;
- Proposes evaluation criteria for the quality, impact and benefits of escience activities;

Goals and Action lines (2)

- Supports actions towards data stewardship and software availability and sustainability
- Will take endeavours to stimulate quality and quality ranking of escience publishing means;
- Facilitates the interaction between its members;
- Will encourage and provide escience requirements towards improved e-infrastructure provisioning and usage;
- Will communicate best escience practices regarding the use of e-infrastructures and ICT tools;
- Will strive for the improvement of the skills-level of students and researchers in escience techniques and stimulate the upgrading of the status of escience technologists and engineers.

Organizationally

- The PLAN-E platform is based on voluntary participation of organizations that share the goals of the platform. In a later stage, if the organization has proven robust, it might be considered to base the organization on formal Letters of Intent. The PLAN-E will consider working with a chosen board with a secretary responsible for the communication. The board will consider drafting a white paper on escience and its present and future impact on science and the way it is conducted.
- In principle the documents produced by PLAN-E are public.

Cooperationally

- PLAN-E is to be a neutral and e-infra independent supporter and advisor to:
 - PRACE
 - EGI
 - EUdat
 - RDA
 - OpenAir
 - e-IRG
 - ESFRI/Projects
 -

Schedule

- PLAN-E is now in place
- Established September 29-30, 2014
- Annual Plan being compiled today
- Broadcasted next months to stakeholders
- Participating countries:
 - (Sweden, Norway, Finland), United Kingdom, Bulgaria, Germany, Hungary, Spain, Italy, The Netherlands, Romania, Serbia, Poland, Belgium, Denmark, Greece, France
 - Observers: EC, EGI, JISK