

# Netherlands eScience Center

... and the physics community

Daniela Remenska  
Research Engineer

netherlands

eScience center

by SURF & NWO

WLCG GDB Meeting  
Amsterdam, 7th March 2016

# Netherlands eScience Center = digitally enhanced Science



# How we work - overarching

- **eScience research Engineers**
- **calls for project proposals**
  - $\pm$  7 big projects (500K Euro)
  - $\pm$  2 pathfinder projects (50K Euro)
- **Fund projects with universities and research Institutes**
  - capture & transfer knowledge/technology between disciplines
- **Public Private collaborations (PPS)**



# eScience engineers & coordinators

- **Research engineers and projects have a coordinator**
- **One or more research engineers per project, depending on:**
  - **Availability**
  - **Project requirements**
  - **Expertise**



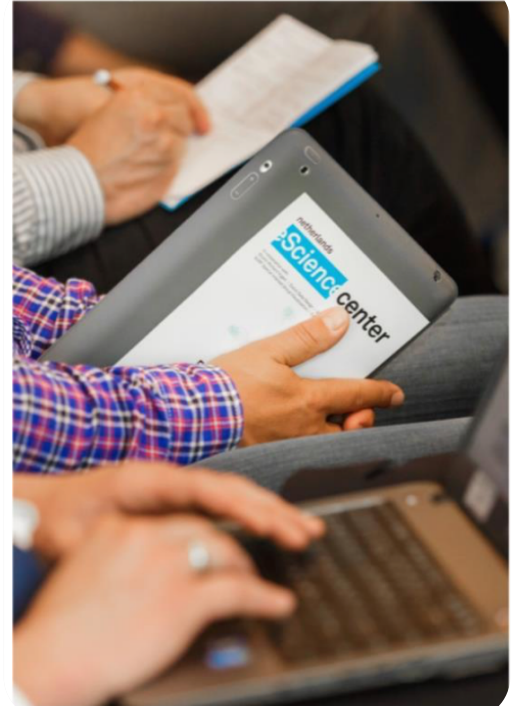
# 'All' of research

- **Environment & Sustainability**
  - climate, ecology, water management, agriculture & food
- **Life Sciences & eHealth**
  - next generation sequencing, biobanking, molecules & man
- **Humanities & Social Sciences**
  - SMART cities, text analysis, eBusiness, creative technologies
- **Physics & Beyond**
  - astronomy, high-energy physics, advanced materials



# Core eScience Technologies

- **Optimized Data Handling**
  - database optimization, sensor networks  
structured & unstructured data
- **Big Data Analytics**
  - statistics, machine learning, text mining
- **Efficient Computing**
  - distributed computing, efficient algorithms



# Implementation in eStep

- eStep = eScience Technology Platform
- Coherent set of technologies to tackle the grand challenges in eScience
- Ensure easy access to data and e-Infrastructures
- Generalized software developed in projects  eStep



# eStep

- **Prevent duplication, fragmentation**
- **Sustainability**
  - Enforce software engineering guidelines and best practices
  - Open source / open access, unless...
  - Community coding

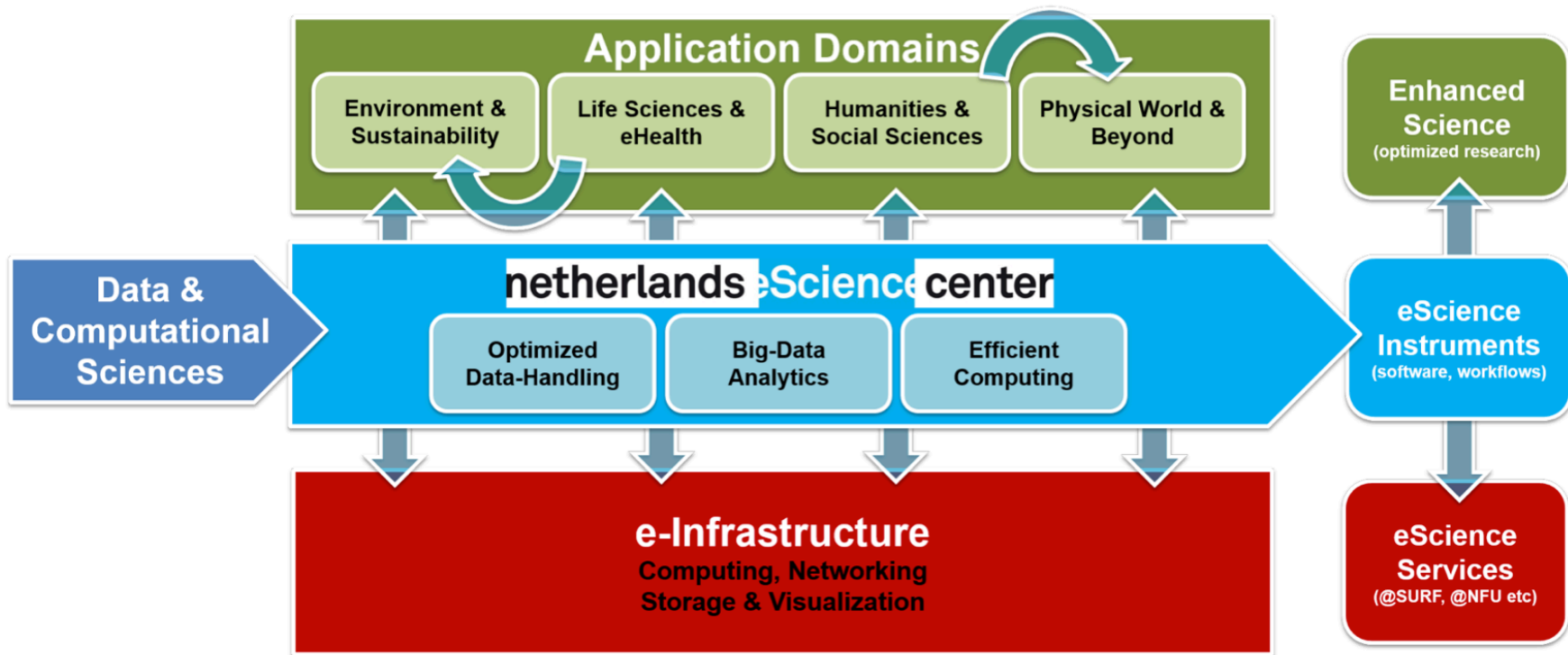




# Working with NLeSC

- **NLeSC = project partner**
- **Engineer(s) typically spend 40%-60% of their time “on location”**
- **Research engineer normally co-author of project papers**
- **Project partners co-author of eScience papers**
- **Similarly for PR, disseminations, outreach, etc.**



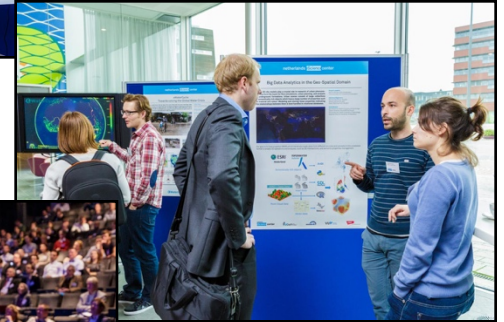


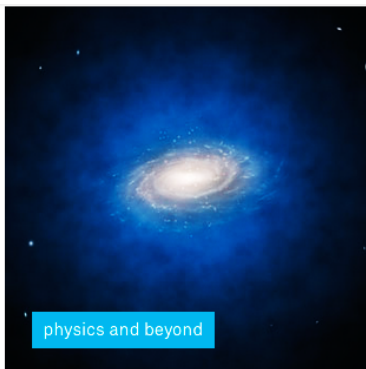
State-of-the-art e-science → impact in discipline → excellence in science



# Communication

- NLeSC web site
- eStep
- Movies
- Yearly symposium
- Magazines
- Papers
- Trainings
- ...

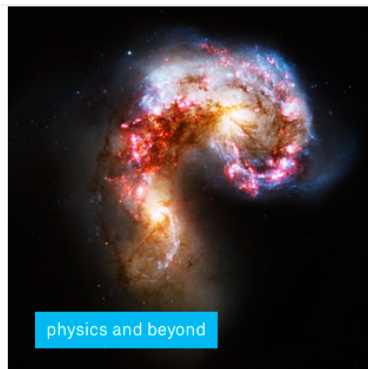




physics and beyond

**iDark**

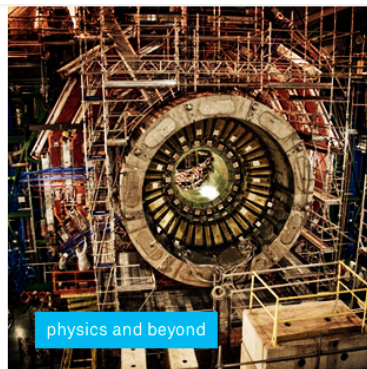
The intelligent Dark Matter Survey



physics and beyond

**AA-ALERT**

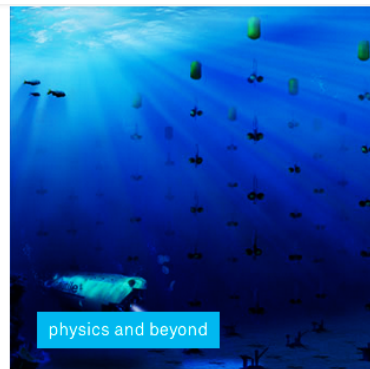
Access and Acceleration of the Apertif  
Legacy Exploration of the Radio  
Transient Sky



physics and beyond

**Automated Parallel Calculation of  
Collaborative Statistical Models**

Large scale statistical data analysis in  
particle physics



physics and beyond

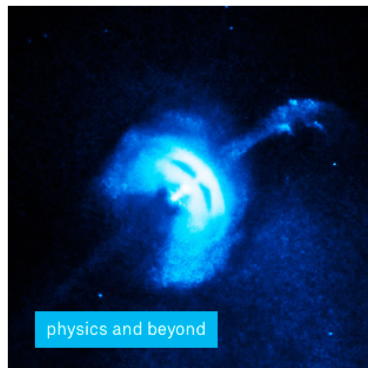
**Real-time detection of neutrinos  
from the distant Universe**

Observing processes that are  
inaccessible to optical telescopes



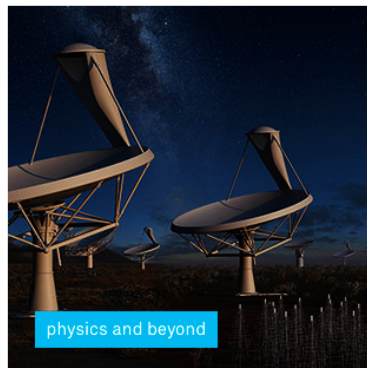
physics and beyond

**Giving Pandas a ROOT to Chew on**  
Modern Big Data front and backends  
in the hunt for Dark Matter



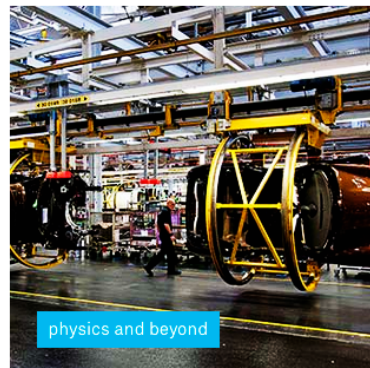
physics and beyond

**Compressing the sky into a large  
collection of statistical models**  
Optimized data handling for  
observations in astronomy



physics and beyond

**Beyond the Data Explosion**  
An eScience infrastructure for huge  
interferometric datasets

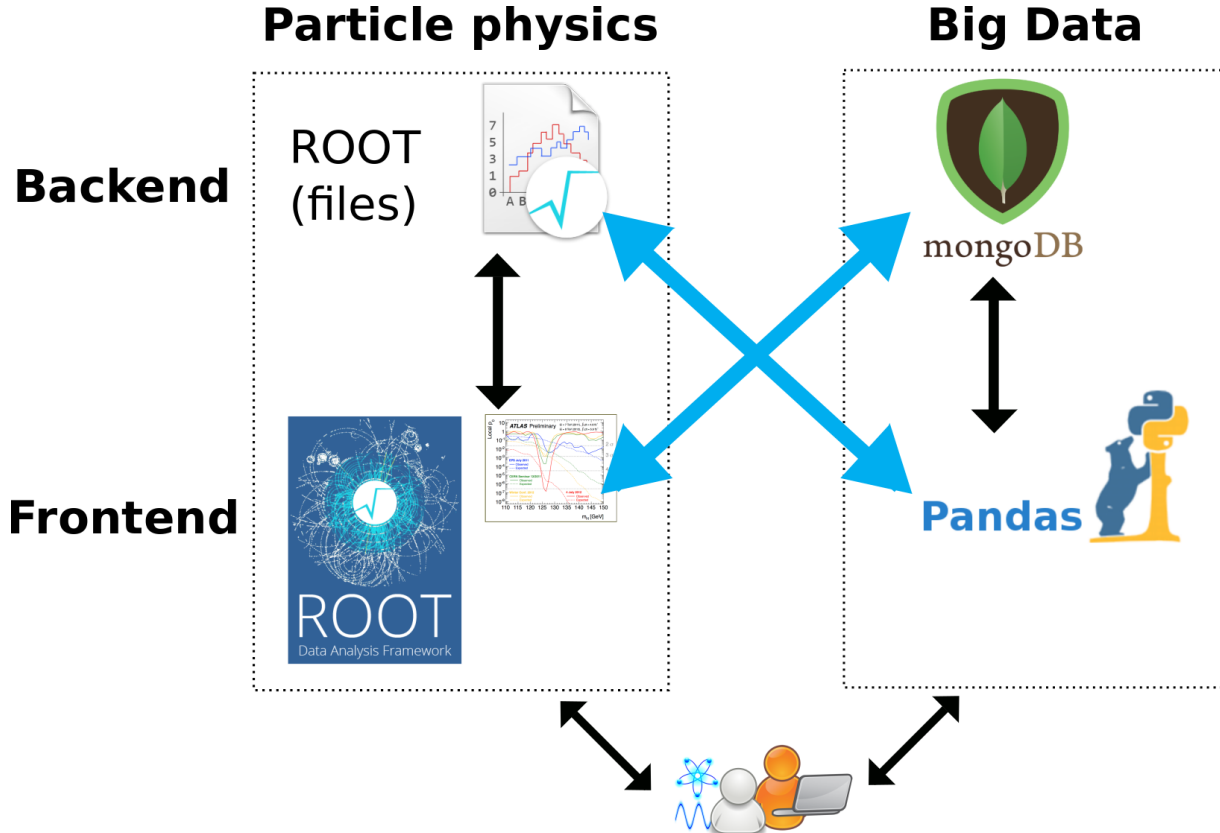


physics and beyond

**PROMIMOOC**  
Process mining for multi-objective  
online control

# Giving Pandas ROOT to Chew On

Modern Big Data front and backends in the hunt for Dark Matter



# Giving Pandas ROOT to Chew On

Modern Big Data front and backends in the hunt for Dark Matter

- ROOT binaries across platforms  
(avoid compiling ROOT on target machine)
- Layers not fully functional /  
lacking Python 3 support;
- Incompatibilities between ROOT  
5/6 versions (also memory leaks)
- Synergize data formats  
(pandas,HDF5,MongoDB,Hadoop)



**root\_pandas:**  
**ROOT I/O for Pandas**

*Pandas should  
read ROOT files*



**rootpy: Pythonic ROOT**

*Truly "Pythonic"  
ROOT interface*

**PyROOT:**  
**A Python -- ROOT Bridge**

*Python bindings  
for ROOT*



**C++**

```
conda install root=6 python=3
```





**Tim Head**

@betatim



Follow

OH: TIL installing ROOT is already as easy as  
`conda install -c nlesc root` via @ibabusch

1:30 PM - 18 Sep 2015



kreczko commented on Nov 25, 2015

@remenska : nice work. Following the conversation in the HSF packaging forum  
(<https://groups.google.com/forum/#topic/hep-sf-packaging-wg/h4HWHnVkBA8>)

Will advertise this to our groups.



cdeil commented on Sep 3, 2015

the rootpy project member

@remenska – That's awesome, thank you!



@ndawe – Should these binaries be used for rootpy testing on travis-ci (either exclusively or in addition to what's there now)?



ANACONDA CLOUD

Search Anaconda Cloud



Docs

Contact



NLeSC

NLeSC / Packages / **root-numpy** 4.4.0



0

Conda

Files

Labels

Badges

Builds

Settings

Apache v2

3258 total downloads



# Real-time detection of neutrinos from the distant Universe

- **Trigger based purely on L0 hits**

Challenging: amount of data+combinatorics...

Can we compare each hit with all other hits?



- **Online reconstruction (= determining the neutrino direction)**

Apply more sophisticated reconstruction algorithms to pre-selected set of hits

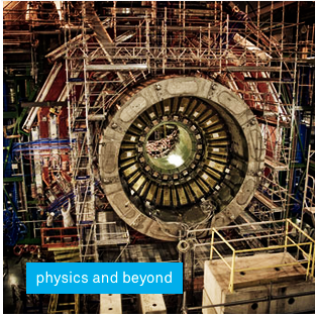
Lowering the threshold/including L0s gives you more chance coincidences

- **This allows real-time alerts to other users**

Also outside of collaboration



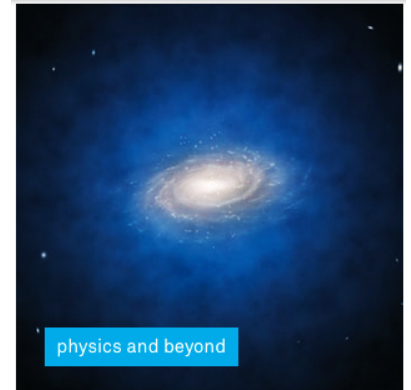
# ... and more HEP NLeSC projects



Automated Parallel Calculation of Collaborative Statistical Models

Large scale statistical data analysis in particle physics

- **RooFit: statistical models of measurements performed by independent teams combined a posteriori without loss of detail**
- **scaling issues**
- **parallel algorithms / new data structures needed**
  
- **Combine the worldwide data within the most general models of Dark Matter**
- **algorithms to find (tiny, fragmented) solution areas in large multidimensional parameter spaces**
- **make a (web-accessible) largely automated “DM model” database**



iDark

The intelligent Dark Matter Survey



# Summary

- **“Research software at the heart of scientific discovery”**
- **NLeSC parallels the HEP Software Foundation in a way**
  - Similar objectives: share expertise, promote commonality & collaboration, raise awareness of existing software solutions, sustainability
  - but we have our own funding and FTEs
- **NLeSC “physics and beyond” projects get in the spotlight in the coming year(s)**
  - Efficient computing and optimized data handling are prevailing themes

See [www.esciencecenter.nl](http://www.esciencecenter.nl) for more

