

Big Data Science  
in Astroparticle Physics:  
**sharing and exploiting  
data & knowledge**

Martin Erdmann & Andreas Haungs

22-Feb-2017

# Workshop: Big data science in astroparticle research

## 20-22 Feb. 2017

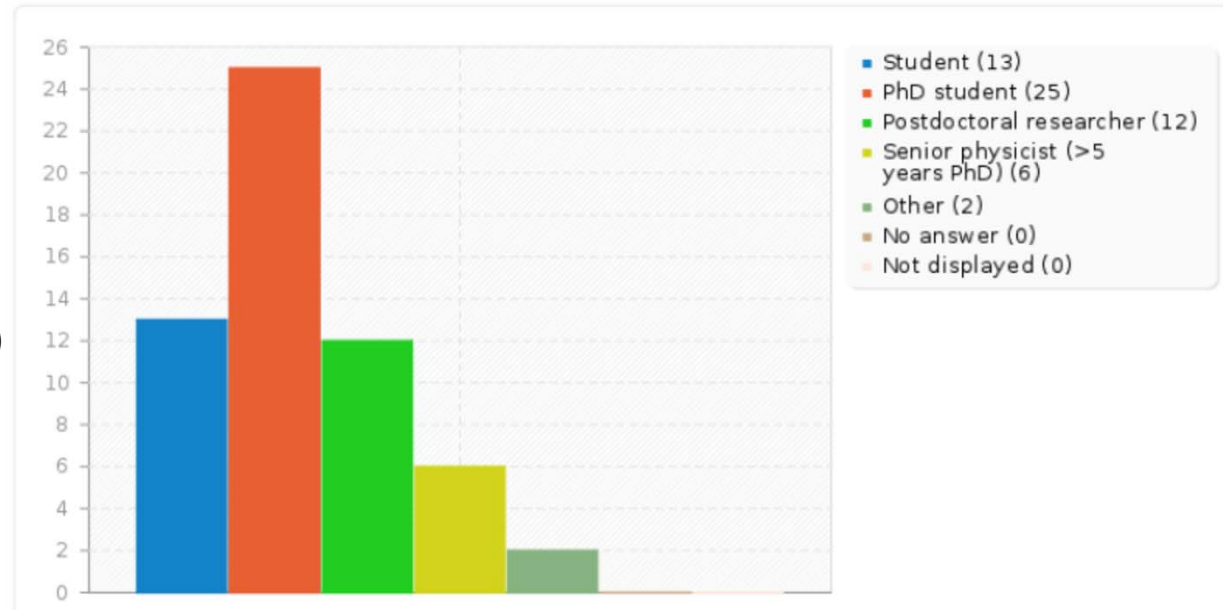
	Monday 14:00	Welcome: Dean of Faculty of Mathematics, Computer Science and Natural Sciences	Prof. SCHAEL, Stefan
	Monday 14:15	Deep Learning Tutorial: basic concepts, fully connected networks	Mr. WALZ, David
	Monday 16:15	Workshop photo	
	Monday 16:15	Deep Learning Tutorial: advanced architectures, convolutional networks	Mr. WALZ, David
Deep Learning	Tuesday 09:00	Organisational Matter	Prof. ERDMANN, Martin
	Tuesday 09:15	Deep Learning for neutrino telescopes	Dr. GEIßELSÖDER, Stefan
	Tuesday 09:40	Deep Learning in Astroparticle Physics exemplified by the Reconstruction of Muon-Neutrino Events in IceCube	Mr. HÜNNEFELD, Mirco
	Tuesday 10:05	Recognizing patterns in the arrival directions of ultra-high energy cosmic rays using deep neural networks	Mr. WIRTZ, Marcus
	Tuesday 10:30	Message of the Vice-Rector for Research and Structure of the RWTH Aachen University	Prof. MATHAR, Rudolf
	Tuesday 11:15	Convolutional Networks in Computer Vision	Prof. LEIBE, Bastian
	Tuesday 12:05	Machine Learning in gamma-ray astronomy: More than just Background-Suppression	Mr. NOETHE, Maximilian
	Tuesday 12:25	Pattern recognition in KM3NeT: A multi-dimensional challenge	Dr. COELHO, Joao
	Tuesday 14:30	Spatiotemporal Integration in Recurrent Deep Neural Networks	Prof. BEHNKE, Sven
	Tuesday 15:15	Exploring deep network architectures to reconstruct cosmic ray induced air showers	Mr. GLOMBITZA, Jonas
Tuesday 15:35	Using neural nets to predict SUSY yields at the LHC	Prof. TATTERSALL, Jamie	
Tuesday 15:55	Event reconstruction and classification in proton-proton collisions using deep neural networks	Mr. RATH, Yannik	
Tuesday 16:00	Evaluation (web)		
Tuesday 16:45	Biological neuronal networks - from structure to activity	Prof. HELIAS, Moritz	
Tuesday 17:25	CREDO: a global cosmic-ray analysis and data processing challenges	Mr. SUSHCHOV, Oleksandr	
Tuesday 17:45	Workflow Management for user analyses in particle physics	Mr. FISCHER, Robert	
Tuesday 18:05	GPU usage of Auger data in the Offline Framework	Dr. RAUTENBERG, Julian	
Tuesday 19:00-22:00	Workshop Dinner Pontgarten		
Open data	Wednesday 09:00	The path to the SKA: Big data challenges in radio astronomy	Dr. BARR, Ewan
	Wednesday 09:50	Data Publishing, Standards, and the Payoff: A Brief Account of the Astronomical Virtual Observatory	Dr. DEMLEITNER, Markus
	Wednesday 11:10	KCDC: data preservation and publication	Dr. HAUNGS, Andreas
	Wednesday 11:35	Accelerating the Digital Transformation in Science	Dr. GAST, Mikael
	Wednesday 12:00	Discussion on Big Data Science in Astroparticle Physics	ERDMANN, HAUNGS

# Participants

58 answers of 96 participants

I participate in this workshop as

- Student
- PhD student
- Postdoctoral researcher
- Senior physicist (>5 years PhD)
- Other



# „Digital Agenda“

## Reading as astroparticle physicist

1. Accelerating digital change in science
2. Access to knowledge as the basis for innovation
3. Education for the digital knowledge society
4. Exploiting the innovative potential of digitalization
5. Through research, understand digital change
6. Culture and media

- Physicists take advantage of new opportunities arising with deep networks
- Experiment data, analyses, publications
- University: courses, seminars, research
- Improve exploitation of new and old data
- Open data to the public

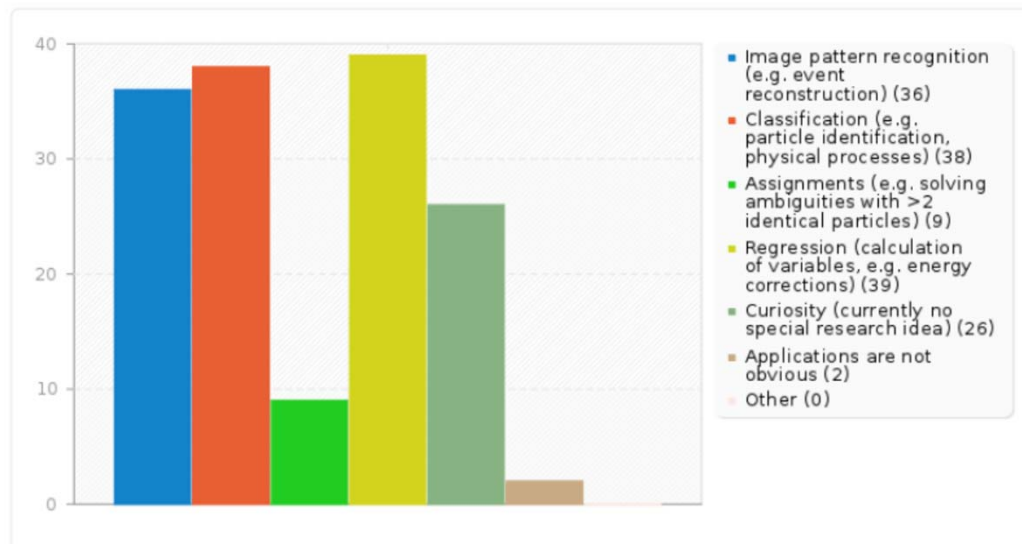
(translated from text of German government concerning Education, Research, Science, Culture, Media, 2014)

## 1. Accelerating digital change in science

# 1. Physicists take advantage of new opportunities arising with deep networks

My research interests concerning deep learning networks are

- **Image pattern recognition** (e.g. event reconstruction)
- **Classification** (e.g. particle identification, physical processes)
- **Assignments** (e.g. solving ambiguities with >2 identical particles)
- **Regression** (calculation of variables, e.g. energy corrections)
- **Curiosity**
- **Applications not obvious**
- **Other applications**



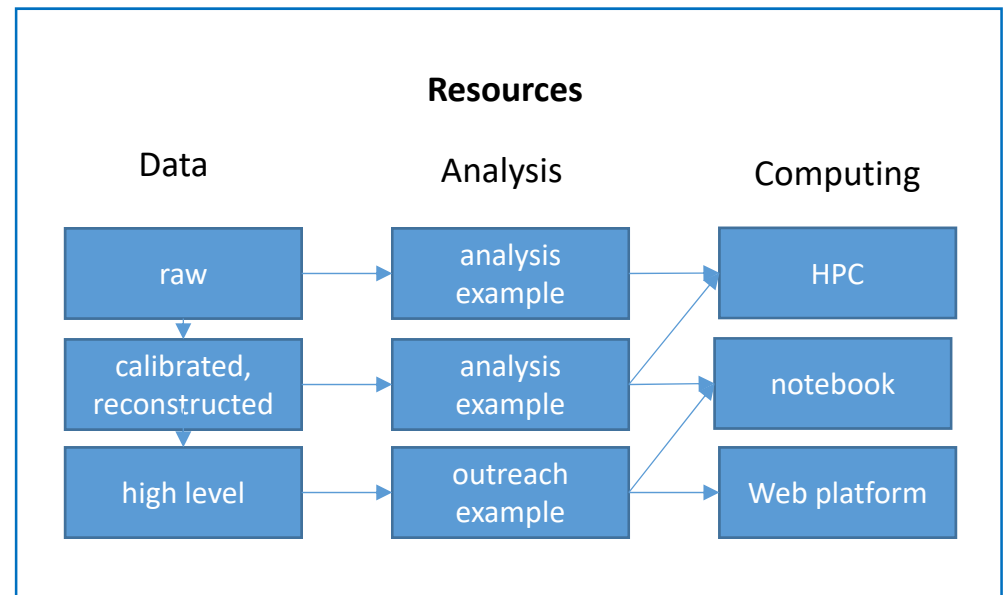
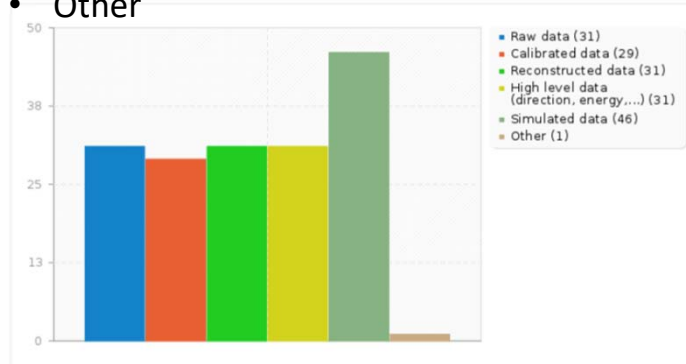
Strong demands, expect exponentially increasing applications

## 2. Access to knowledge as the basis for innovation

# 2a. Experiment data, analyses, publications

For performing my research I need access to

- Raw data
- Calibrated data
- Reconstructed data
- High level data
- Simulated data
- Other



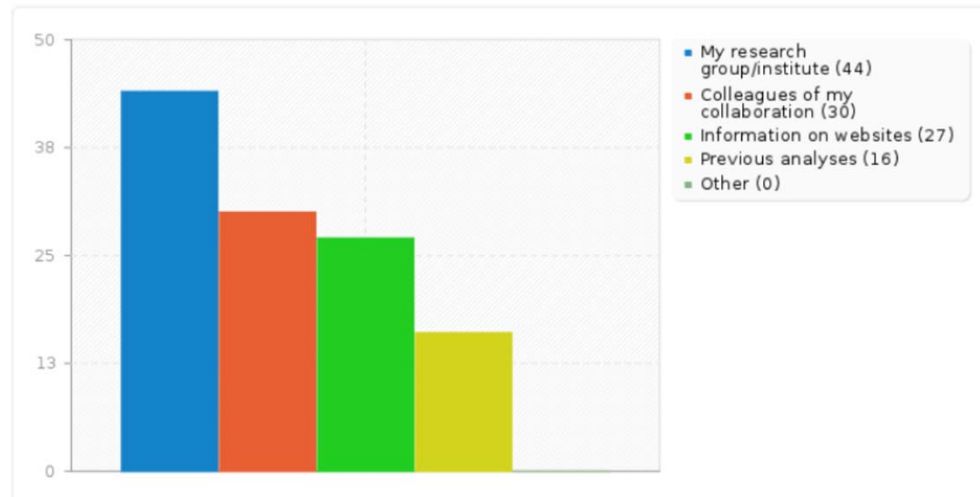
Resource intensive, aim to maximally exploit old & new data

## 2. Access to knowledge as the basis for innovation

# 2b. Experiment data, analyses, publications

I get the most important analysis support from

- My research group/institute
- Colleagues of my collaboration
- Information on websites
- Previous analyses
- Other



Support is an obvious prerequisite for successful exploitation of data

### 3. Education for the digital knowledge society

## 3. University: courses, seminars, research

I have used machine learning techniques before

- Deep learning networks
  - Neural networks (shallow)
  - Boosted decision trees
  - None
  - Don't know
  - Other
- Random Forests, Support Vector  
Machines, Naive Bayes ...



Broadcast new possibilities arising from deep learning to physics institutes



### 3. Education for the digital knowledge society

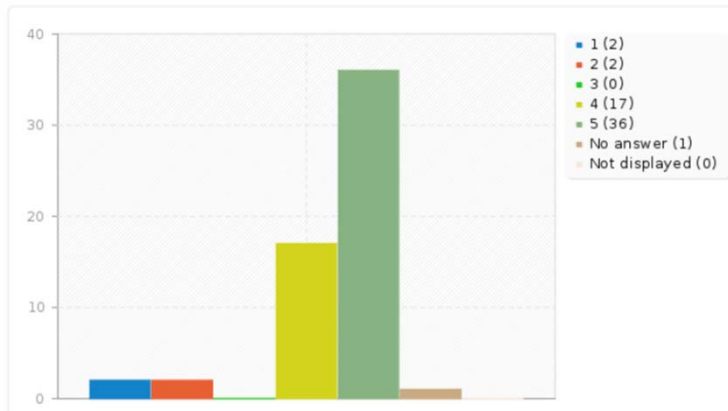
## Workshop tutorial

*New master modul @RWTH: Deep Learning in Physics Research , summer term 2017*

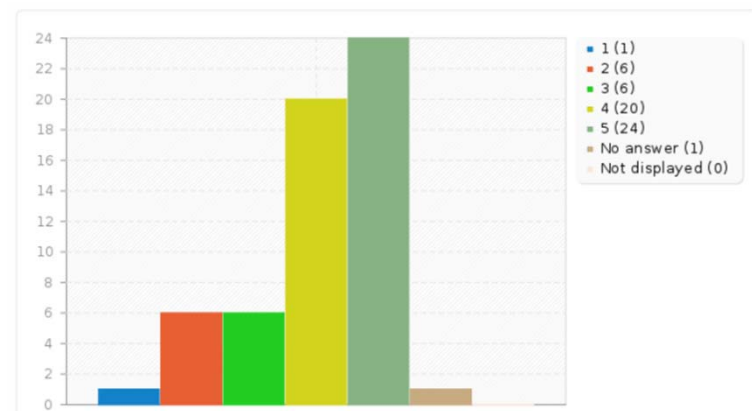
[D. Walz, S. Schipmann, M. Erdmann, U. Klemradt, Lectures & exercises, 2h/week]

- Introduction to machine learning with deep neural networks.
- Applications to image recognition and to specific problems in solid state, particle and astroparticle physics.

The deep learning introductory tutorial was helpful



The example applications deepened my understanding of deep networks



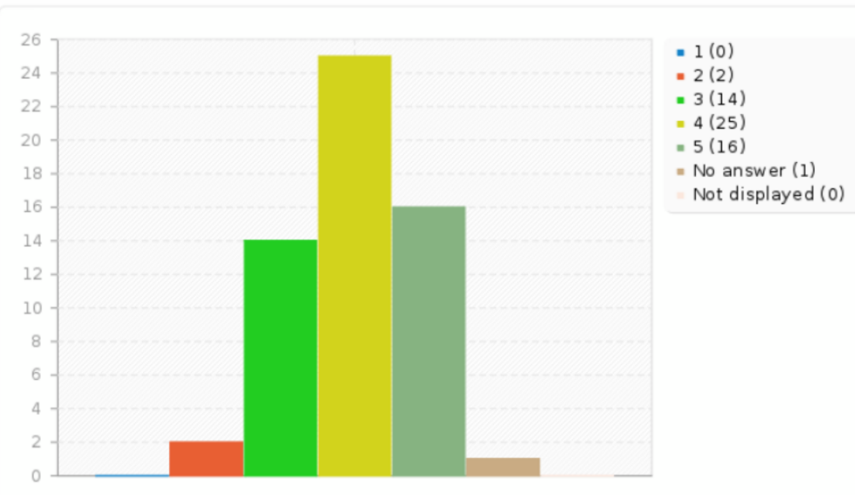
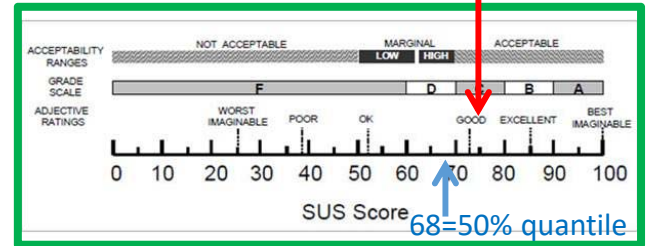
1=strongly disagree, 2=disagree, 3=undecided, 4=agree, 5=strongly agree

### 3. Education for the digital knowledge society

# VISPA platform

My overall score for the VISPA platform

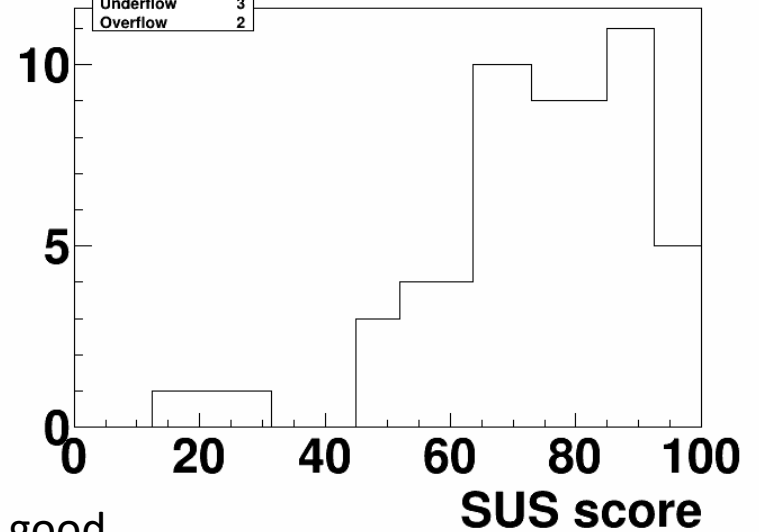
Non-linear quality scale



SUS score

**N**

h	
Entries	58
Mean	74.6
RMS	15.9
Underflow	3
Overflow	2



1=insufficient, 2=sufficient, 3=ok, 4=good, 5=very good

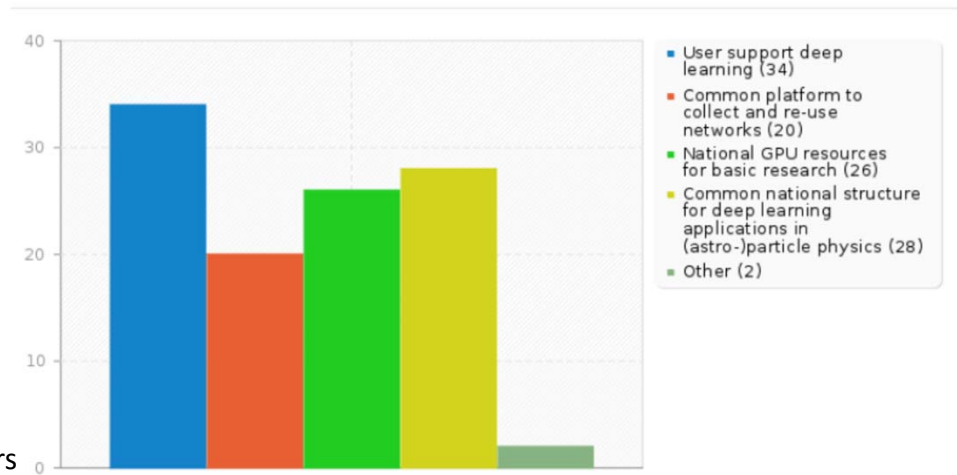
## 4. Exploiting the innovative potential of digitalization

# 4. Improve exploitation of new and old data

The following **measures** are important for achieving progress in my research

- User support deep learning
- Common platform to collect & re-use networks
- National GPU resources for basic research
- Common national structure for deep learning applications in (astro-)particle physics
- others

Common *international* structures for deep learning in physics, info exchange between the community members



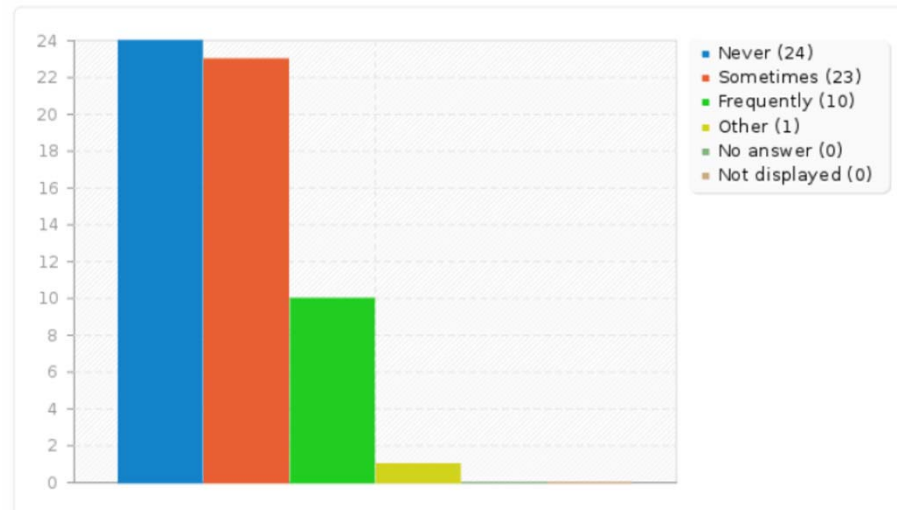
Need sustainable structures for accelerating research & development

## 6. Culture and media

# 6. Experiment data, analyses, publications

I have accessed open (*public*) data for my research purposes before

- Never
- Sometimes
- Frequently

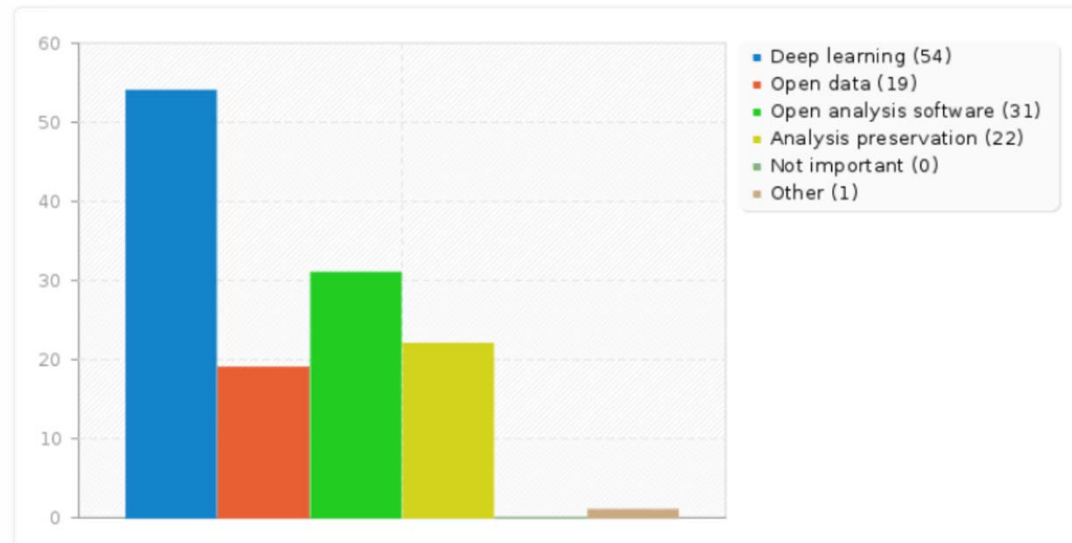


Requires scientific experience to judge what to extract from which data

# Workshops perspectives

For the future, I consider continuation workshops important on topic(s)

- Deep learning
- Open data
- Open analysis software
- Analysis preservation
- Not important
- Other



Strong wish to benefit from each other's progress,  
reconvene in 1 year

# Ongoing discussions: structures & support

Discussions for possibilities on supporting these efforts are ongoing

## Funding

BMBF

DFG

Helmholtz

EU

## Big data science in astroparticle research

Committee representing User Communities

Deep learning methods

Network exchange (pre-trained)

GPU computing resources

User support

Open data / management

Analysis software & preservation

HPC Computing resources

User support

# Thanks very much!



Workshop administration:

[Sabine Bucher](#) (KIT), [Melanie Roder](#) (RWTH)

Local RWTH workshop team:

[Daniel Kuempel](#), [Gero Müller](#), [Yannik Rath](#), [David Walz](#), [Marcus Wirtz](#)

and the VISPA team:

[Benjamin Fischer](#), [Robert Fischer](#), [Erik Geiser](#), [Christian Glaser](#), [Gero Müller](#), [Thorben Quast](#), [Marcel Rieger](#), [Martin Urban](#), [Florian von Cube](#), [David Walz](#), [Christoph Welling](#)

# Backup



# GPU resources

I have access to GPU resources for network training at my research institute

