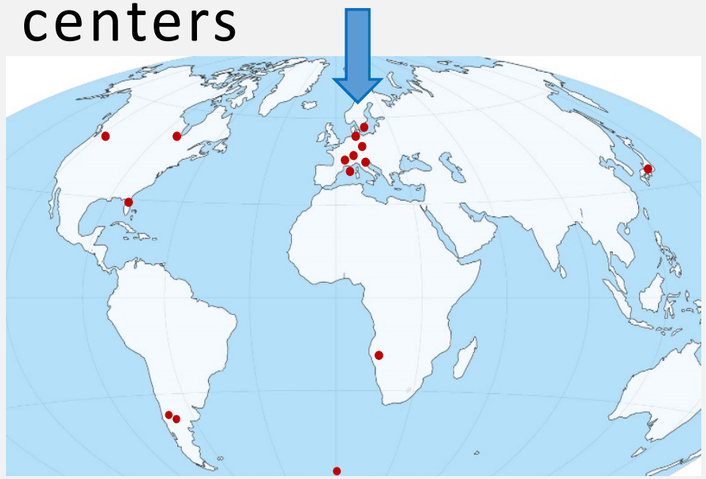
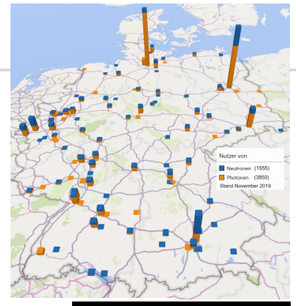
Digital Transformation in the Research on Universe & Matter

1. Organisational matters
2. Deep Learning




Research at largescale (inter)national research centers





Forschung mit Synchrotronstrahlung	2,300
Forschung mit Neutronen	1,600
Rat Deutscher Sternwarten	1,500
Hadronen- und Kernphysik	1,500
Elementarteilchenphysik	1,300
Astroteilchenphysik	500
Beschleunigerphysik	200
Forschung mit nuklearen Sonden und Ionenstrahlen	100
Total number of scientists with doctoral degree	**9,000**

Martin Erdmann, RWTH Aachen University 18-Jan-21

We wish to thank the representatives of the BMBF for their initiative and enormous support for the future of digitization in ErUM research

A short history

BMBF Referat 711

ErUM=Erforschung Universum & Materie

Scientists

9,000 PhD scientists in Germany organized in 8 committees

2018 **Modern digitization? → Workshop**

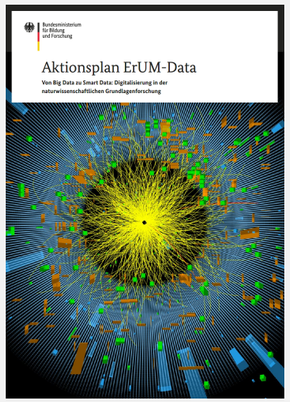
2019

2020

Aktionsplan ErUM-Data

120 M€ for 10 years
Calls for project funding,
Translation, Transfer

& office ErUM-Data-Hub



1. Document of need

2. Umbrella organization (new)

Coordination	**Overview Board (OB)** 8 Committee Chairs, 1 Resource Provider, 1 Representative of the BMBF, 1 Representative of DESY-PT		
	Spokesperson / 5 Co-spokespersons	**Digitization Board (DB)** Spokesperson, 5 Co-spokespersons, 8 Experts from committees, 1 Resource Provider	**Resource Provider Board (RB)** 10 Resource Providers, 8 Experts from committees
	Administrative Office (Hub) 1 Administration coordinator, 5 Assistants, 1 Communicator, 1 Secretary, 1 Technical Support	**Annual Conference** of the ErUM-Data Working Groups	**International Advisory Board (IAB)** ca. 5 from Science, Industry
Topic Boards	**Federated Infrastructures** Board: Coordinators, Experts Compute power Utilization Workflows ...	**Big Data Analytics** Board: Coordinators, Experts Algorithms Autonomization Control & preservation ...	**Research Data** Board: Coordinators, Experts Data models Management Curation ...	**User Interface** Board: Coordinators, Experts Scientists questions Developers work User support ...	**Knowledge distribution** Board: Coordinators, Experts Tenure track programme Workshop, schools ...

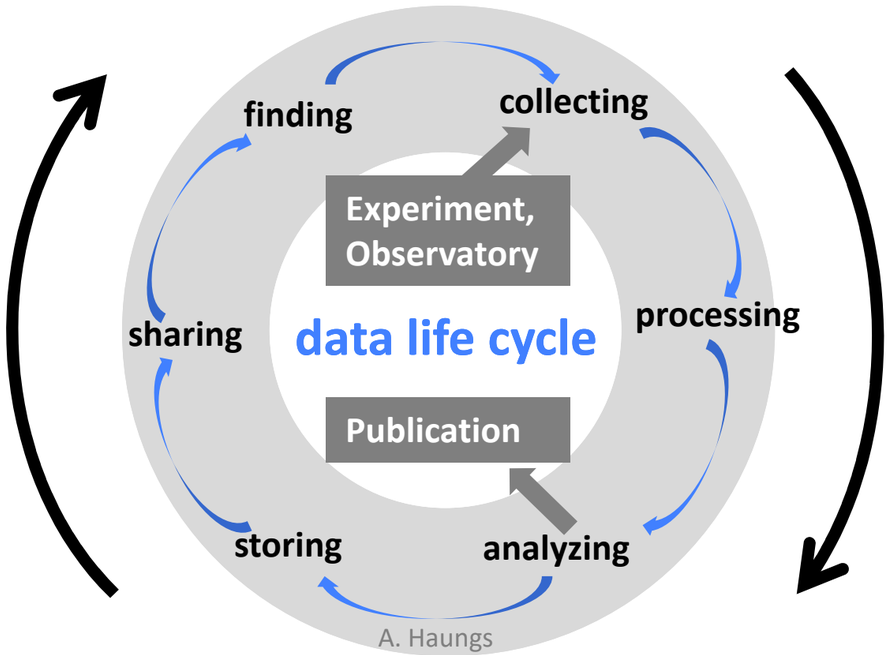
3. Preparing application for office

2021 Funding/location decision on the office

Why digitalization? why now?

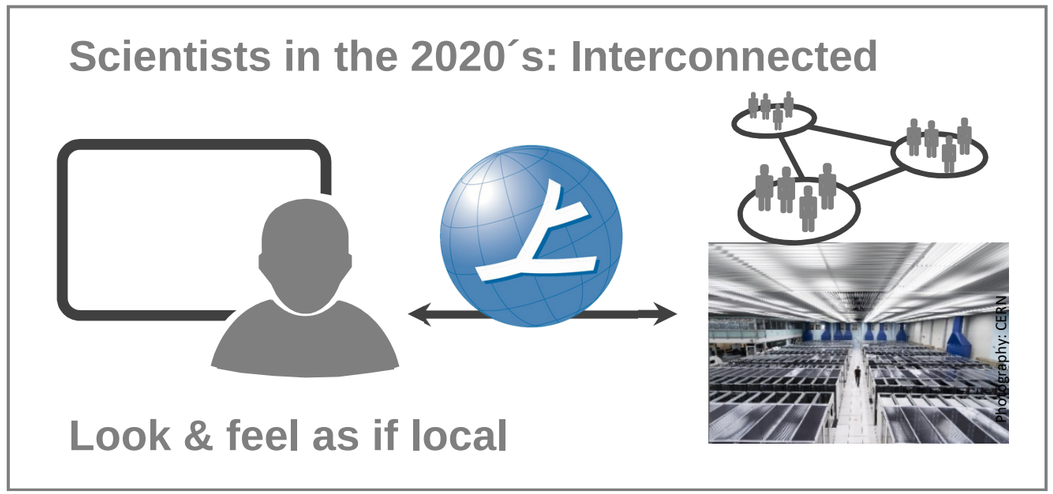
1. Data measurement & simulation (theory)

2. Algorithms to answer scientific questions



A. Haungs

3. Large federated **Digital Infrastructures** their utilization



Multiple challenges that cannot simply be solved by out-of-the-box methods

Blueprint

Challenges and Opportunities of Digital Transformation in Fundamental Research on Universe and Matter

Recommendations of the ErUM Committees

[ErUM - Exploration of the Universe and Matter]

29 April 2019

Editorial board:
Prof. Dr. Martin Erdmann 1 (Lead Editor)
Prof. Dr. Christian Gutt 2e
Dr. Andreas Haungs 3a
Dr. Klaudia Hradil 4d
Prof. Dr. Thomas Kuhr 5b
Dr. habil. Marcel Kunze 6g
Prof. Dr. Anke-Susanne Müller 3c
Prof. Dr. Günter Quast 3b
Prof. Dr. Matthias Steinmetz 7h

https://www.ketweb.de/stellungnahmen/e300611/Strategiepapier_ErUM-Data_Final_2019-04-29.pdf

Recommended measures

strongly endorsed by all committees

Partnership → Structure of new Umbrella Organization

You

Please, feel encouraged to participate: grassroots organization

Overview Board (OB)
8 Committee Chairs, 1 Resource Provider, 1 Representative of the BMBF, 1 Representative of DESY-PT

Coordination

Spokesperson / 5 Co-spokespersons	**Digitization Board (DB)** Spokesperson, 5 Co-spokespersons, 8 Experts from committees, 1 Resource Provider	**Resource Provider Board (RB)** 10 Resource Providers, 8 Experts from committees
Administrative Office (Hub) 1 Administration coordinator, 5 Assistants, 1 Communicator, 1 Secretary, 1 Technical Support	**Annual Conference** of the ErUM-Data Working Groups	**International Advisory Board (IAB)** ca. 5 from Science, Industry

Topic Boards

Federated Infrastructures Board: Coordinators, Experts Compute power Utilization Workflows …	**Big Data Analytics** Board: Coordinators, Experts Algorithms Autonomization Control & preservation …	**Research Data** Board: Coordinators, Experts Data models Management Curation …	**User Interface** Board: Coordinators, Experts Scientists questions Developers work User support …	**Knowledge distribution** Board: Coordinators, Experts Tenure track programme Workshop, schools …

News:

- All committees support guidelines for organization
- M. Erdmann as spokesperson, RWTH Aachen as central office
- RWTH Aachen confirms preparedness for hosting office

Office activities ErUM-Data-Hub

Kontakt- und Transferstelle

- zwischen ErUM-Data Partnern (Universitäten, Helmholtz-Gemeinschaft, Max-Planck Gesellschaft, Leibniz-Gemeinschaft, Fraunhofer-Gesellschaft)
- Informatik, Mathematik
- BMBF-Kompetenzzentren künstliche Intelligenz
- Großforschungsanlagen
- Ressourcen-Betreiber für Computing
- Nationale & europäische Aktivitäten
- Industrie, Wirtschaftsunternehmen, Methoden- und Technologieentwickler

Aus- und Weiterbildung, Workshops

- Koordinierung von Aus-, Weiterbildung, Wissens- und Innovationstransfer

Education/Schools

Workshops

Medien

- Koordinierung der Öffentlichkeitsarbeit
- Maintenance effizienter Datenbanken zu Informationen, Ideen, Kontakten, Workshops, Konferenzen, Journale, Sprachdefinitionen

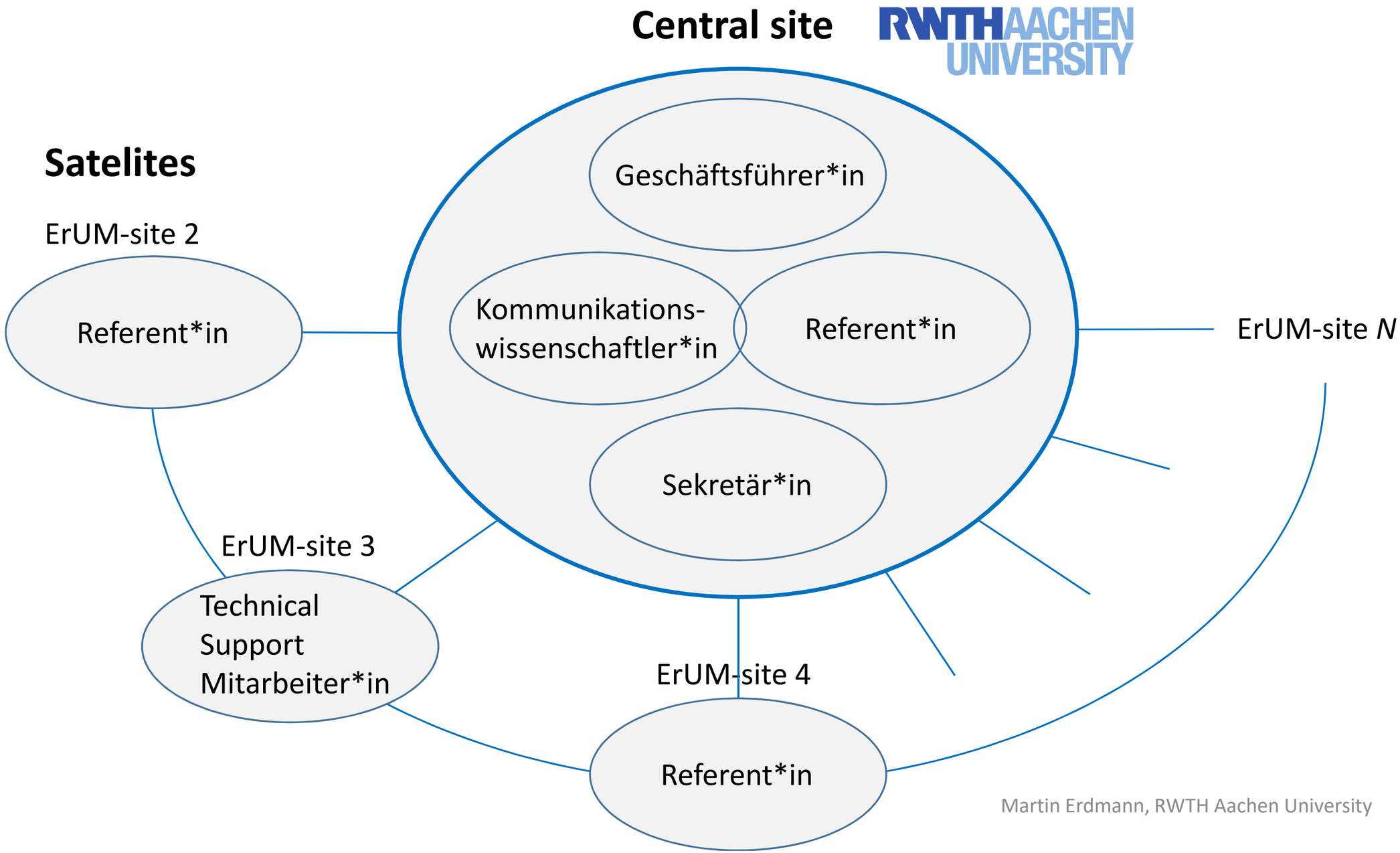
Social media

Beispiele:

ErUM-Arbeitsgruppe: Anfrage spezifisches Problem zu Deep Learning Verfahren: **Referent*in** kennt Akteure und Datenbankeinträge zum Thema, kann an aktive Nutzer*innen und FAQ verweisen.

Firma BRUKER (Hersteller von Röntgenanlagen): Anfrage zur Beteiligung an Erstellung und Kuration von Metadaten und Laborbüchern für ihre kommerziellen Anlagen: **Referent*in** kennt die Förderinstrumente für Industriebeteiligungen und erstellt Diskussions- und Entscheidungsvorlagen.

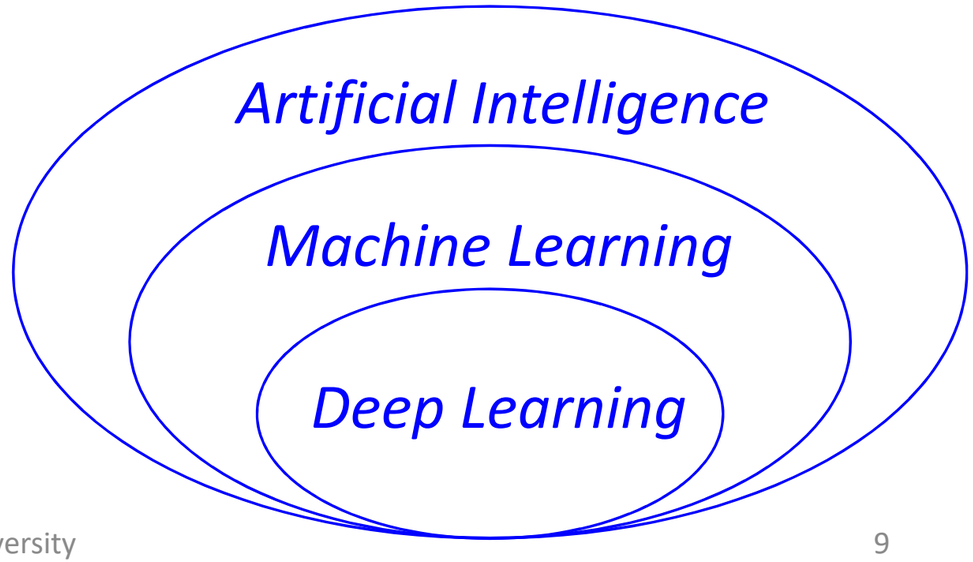
ErUM-Data-Hub Distributed Office



Software & Algorithms

- *machine learning (deep learning)*
- *optimized algorithms for experiment software (parallelization, vectorization) and simulation/theory*
- *adaptation to heterogeneous computing architectures (i.e. GPUs, FPGAs)*
- *intermediate/long term future: quantum computing*

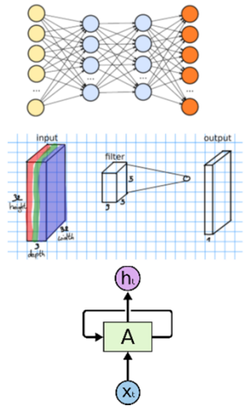
Data-driven knowledge gain by deep learning

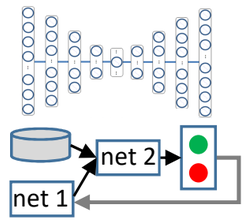


Deep Learning is based on Neural Networks

Concepts (symmetries in data)

- Fully connected
- Convolutional
- Graph
- Recurrent
- Lorentz Boost Network
- Autoencoder
- Adversarial
- Reinforcement
- Invertible





Improved set of tools

Train millions of parameters by:

- Data preprocessing
- Normalization
- Regularization
- Short cuts
- Gradient descent
- Momentum
- …

Computing

- Graphics Processing Unit (GPU)
- Software Libraries
  - TensorFlow
  - keras…

sufficent: Python language

McCulloch, W.S., Pitts, W.: Bulletin of Mathematical Biophysics (1943) 5: 115.
Frank Rosenblatt, Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. Spartan Books, Washington DC, 1961

Neural Network Operations

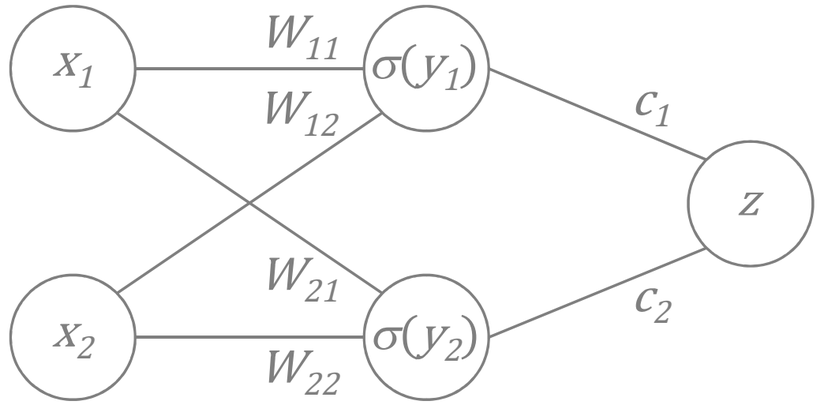
x multi-dimensional input data

W, b to be trained

successively apply 2 operations:

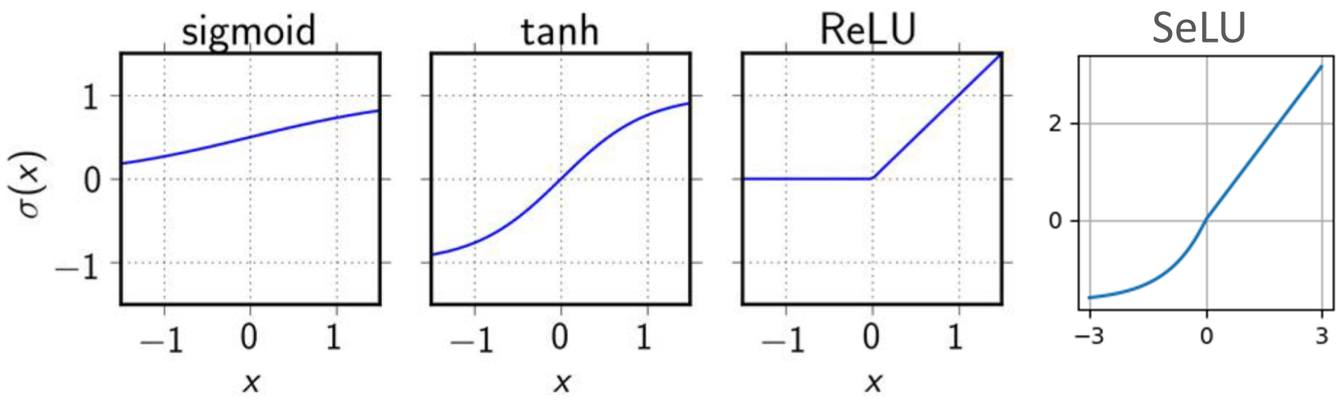
$$y = Wx + b$$

$$h = \sigma(y)$$

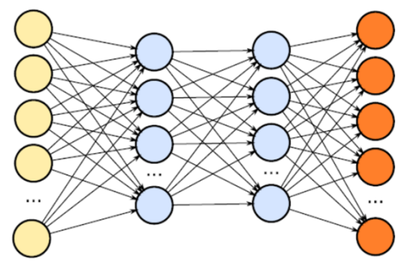


system differentiable

activation function: departure from linear system

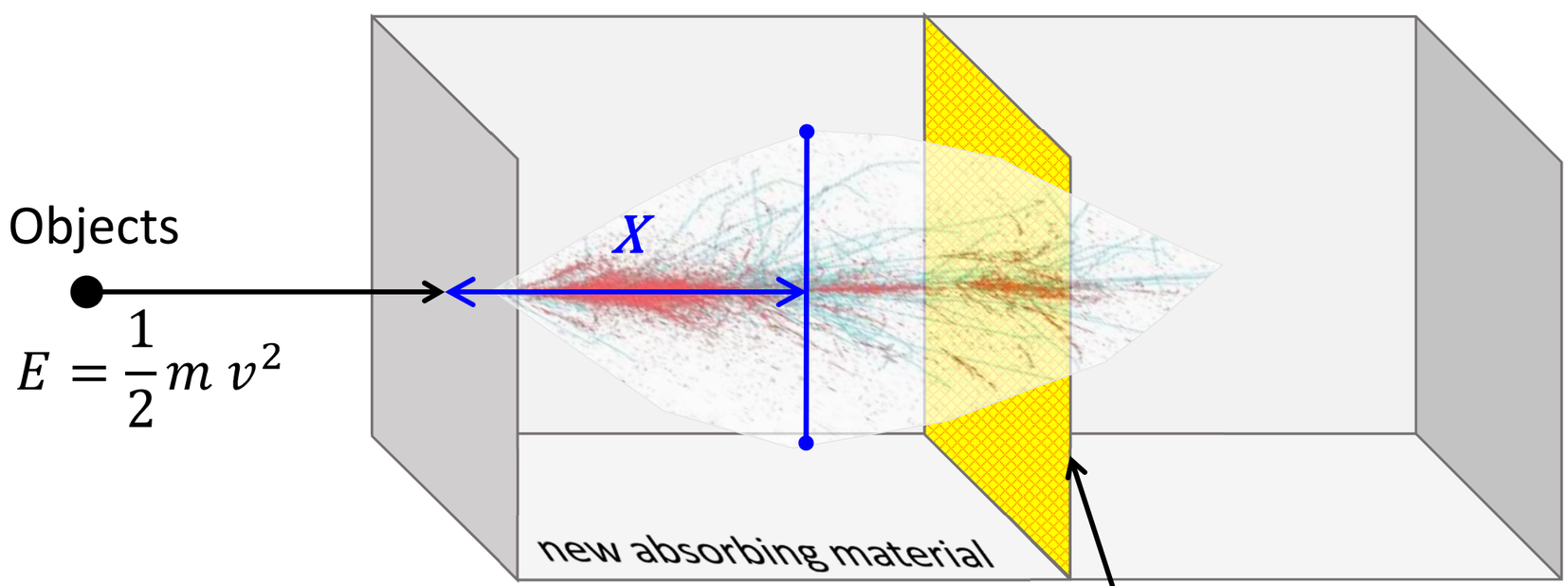


up-scale



Thought experiment new absorbing material

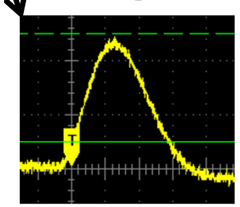
Wanted: X = path to maximum transverse extent



1. Photography

X from image analysis by convolutional network

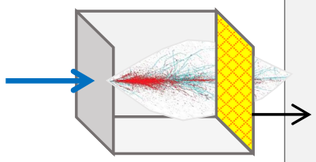
2. Signals of what `remained´ on *single* plane



X after scan of signals by recurrent network

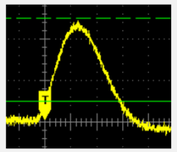
Deep neural networks add scientific value to experiments, examples in backup

Simulation mismatch can be repaired

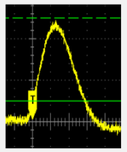


$$E = \frac{1}{2} m\, v^2$$

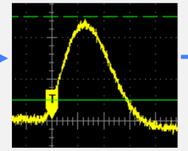
data traces



badly simulated traces

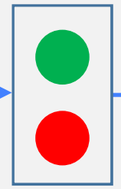


Neural network 1 improves simulated trace



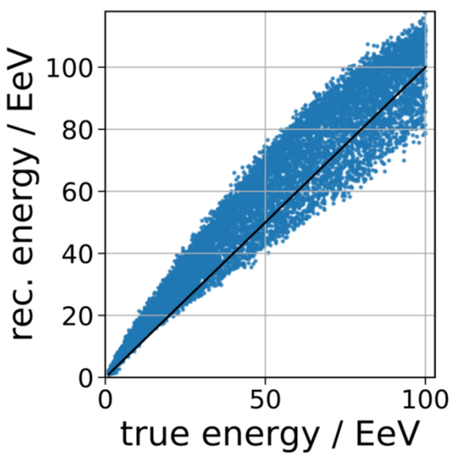
Neural network 2 evaluates similarity data↔simulation

Adversarial training



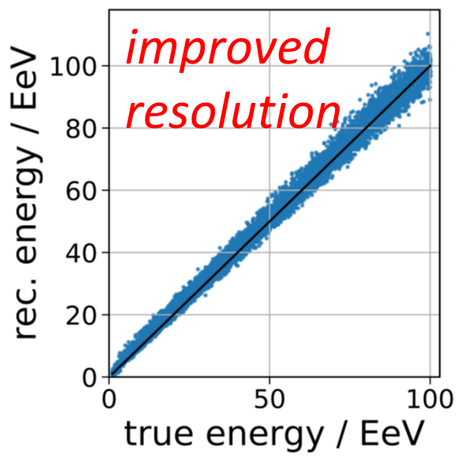
Neural network 3 predicts energy E from trace

trained on bad simulation, evaluated on data

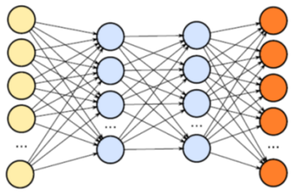


Neural network 3 predicts energy E from trace

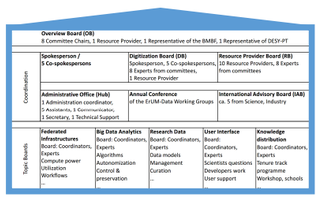
trained on data-like simulation, evaluated on data



Conclusions



- **Open doors to anyone interested in gaining knowledge by data-driven methods**
- New focus is controlled autonomization of algorithms (previously: automation)
- Added scientific value to experiments (also economizes financial resources)



- **Umbrella organization: together shaping digital methods for your research**
- Grassroots organization: plenty of room for your contribution
- Upcoming office will support, but scientific momentum needed from you
- Name for organization? refer to backup

THIS MEETING

- We need to understand what exactly are your questions to the data
- Common research questions → find your partners for upcoming funding application

backup

Your suggestion for the name of the umbrella organization ?

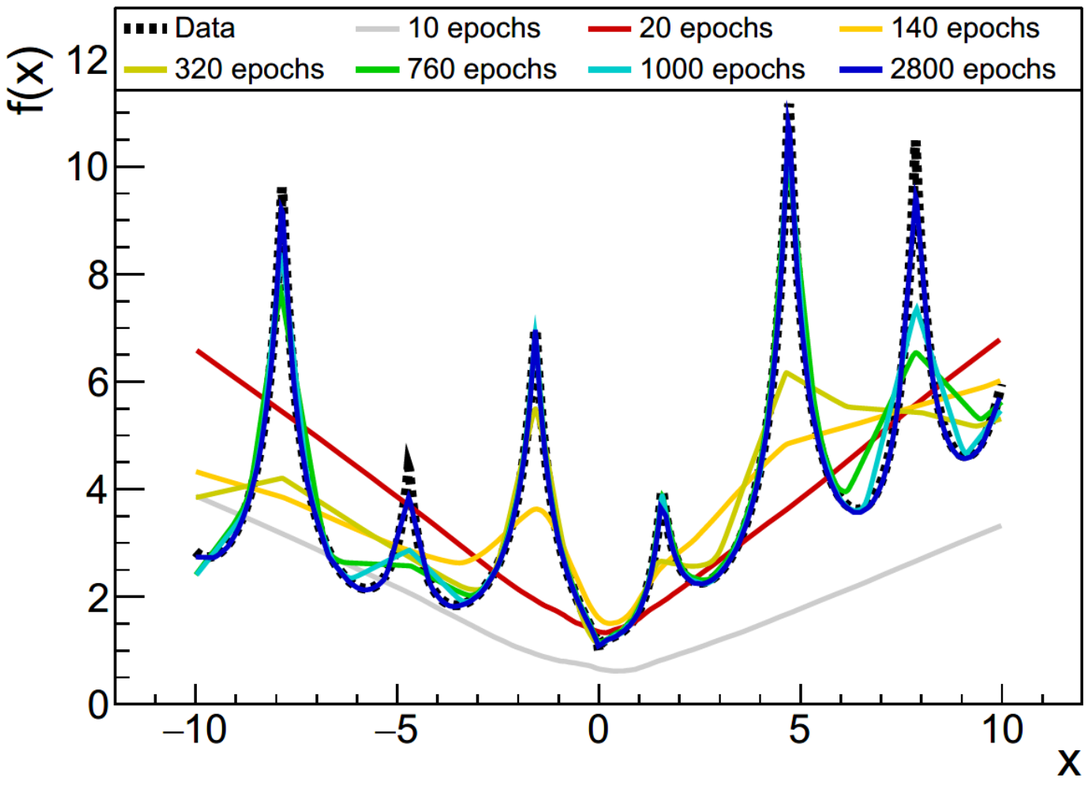
Digital transformation in the research on universe and matter

- DIGUM
- DIGErUM
- DigitaliUM
- DARUM
- DiRUM (Digitalization of the Research in Universe and Matter) Haungs
- DRUM
- DigUM
- DiTRUM (Digital Transformation in the Research on Universe and Matter)
- DUniMat
- DRUniMat
- Digitraum
- Digrum
- Dagum
- Derum

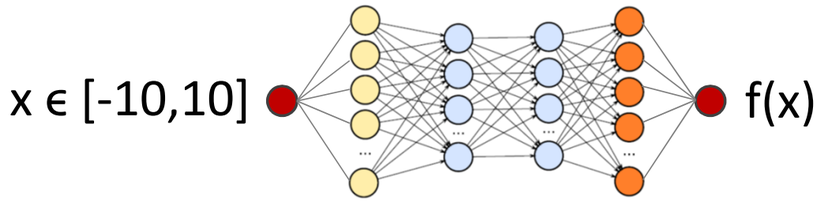
Accepted suggestion:
Aachener Printen

Example: Parameterization of arbitrary function

T. Quast CERN/RWTH Aachen



original function (black symbols):
fair description after 2800 training steps



- Neural network represents ultra-flexible model
- Reality: working in multi-dimensions

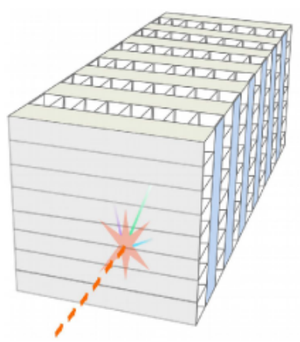
$$\vec{x} \in \mathbb{R}^n \rightarrow \vec{z} \in \mathbb{R}^m$$

- Adjusting network: likelihood fit

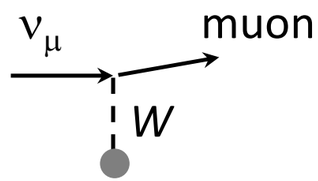
Electron neutrino identification

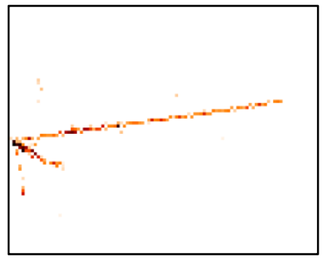
A. Aurisano et al., JINST 11 (2016) P09001

Fermilab (Chicago)
→NOvA experiment 810km

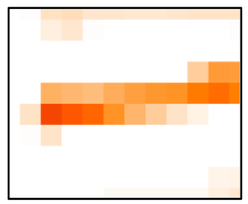


Convolutional neural network
neutrino event classifier

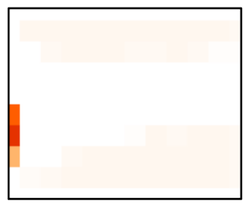




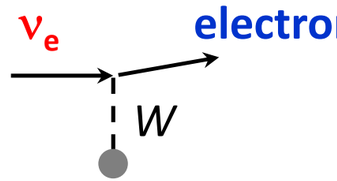
Feature maps
track

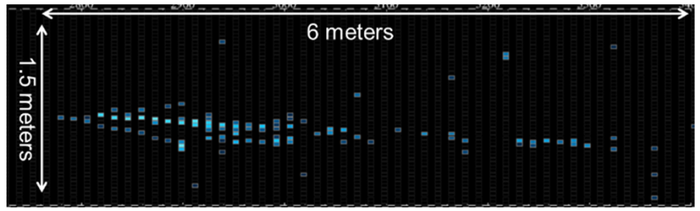


hadronic



Challenge: electron-neutrinos



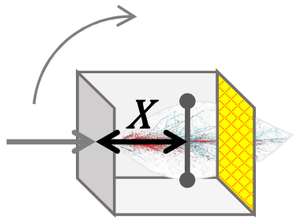


Method	**ν_e efficiency** (same purity)
Physicists algorithm	**35%**
Deep learning neural network	**49%**

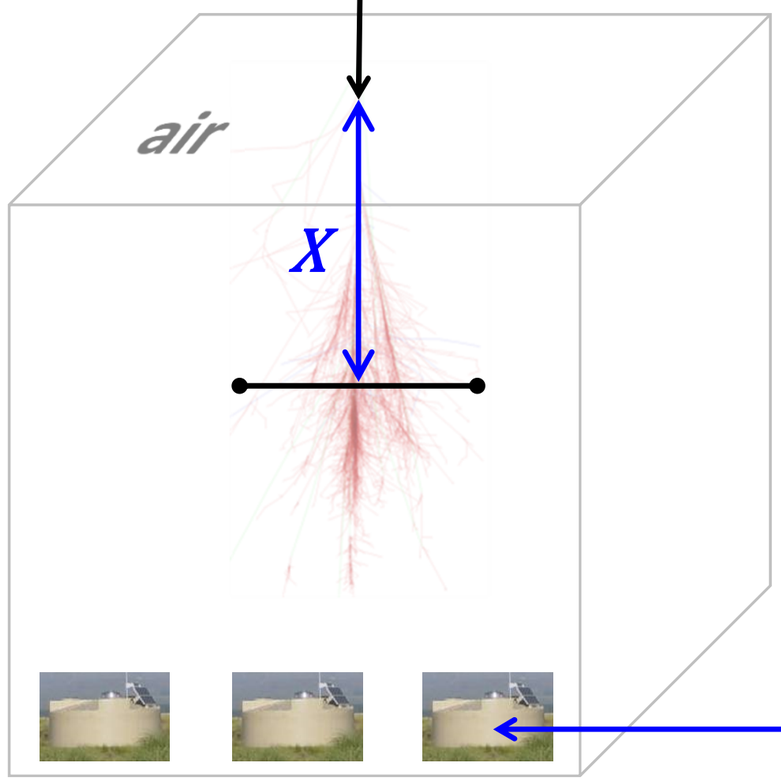
Collaboration wins 40% statistics

https://arxiv.org/abs/2101.02946

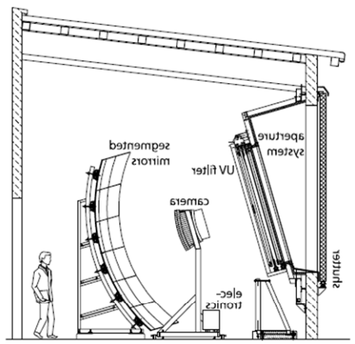
Atmospheric collisions of cosmic rays



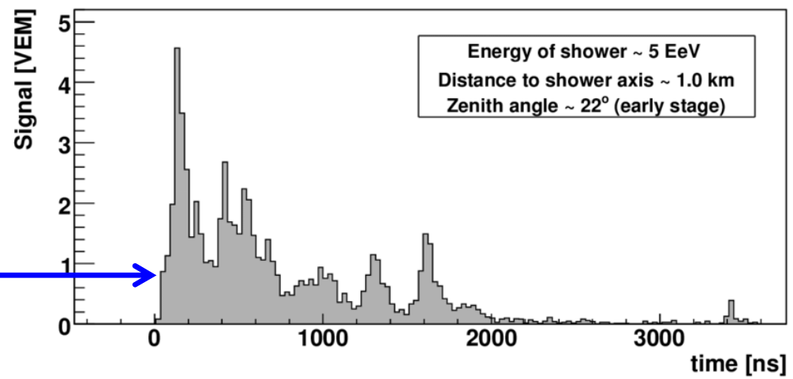
Cosmic rays air showers
Pierre Auger Observatory



1. Photography
at dark *<u>night</u>* → X



2. Scan signal traces *<u>day&night</u>*
deep learning (recurrent LSTM)
predicts → X



Collaboration wins factor 7 statistics of high quality measurements