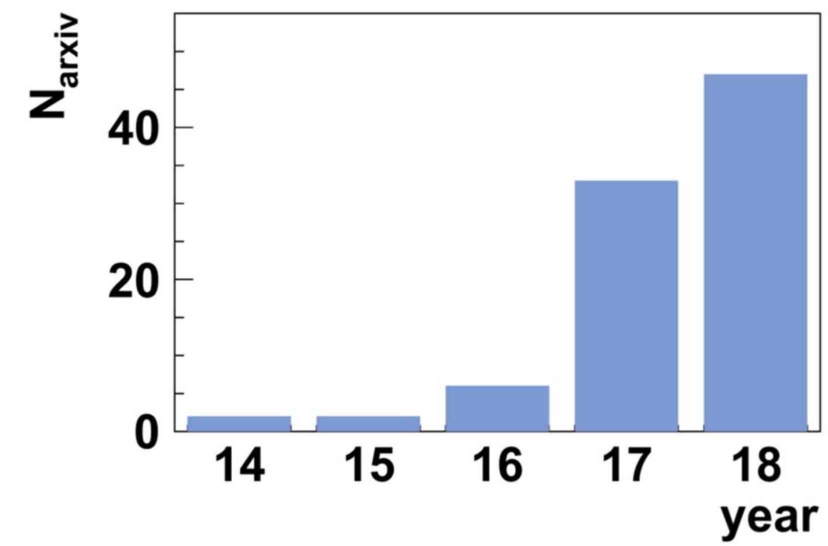
Joint, interdisciplinary activities & developments: Basic structure?

deep learning in particle & astroparticle physics



Big Data Analytics

1. **Rapid development: HEP at new level of Big Data Analysis**
2. **Education: Anchoring machine learning in curriculum**
3. **Funding: Competence networks of various disciplines & funded structure**

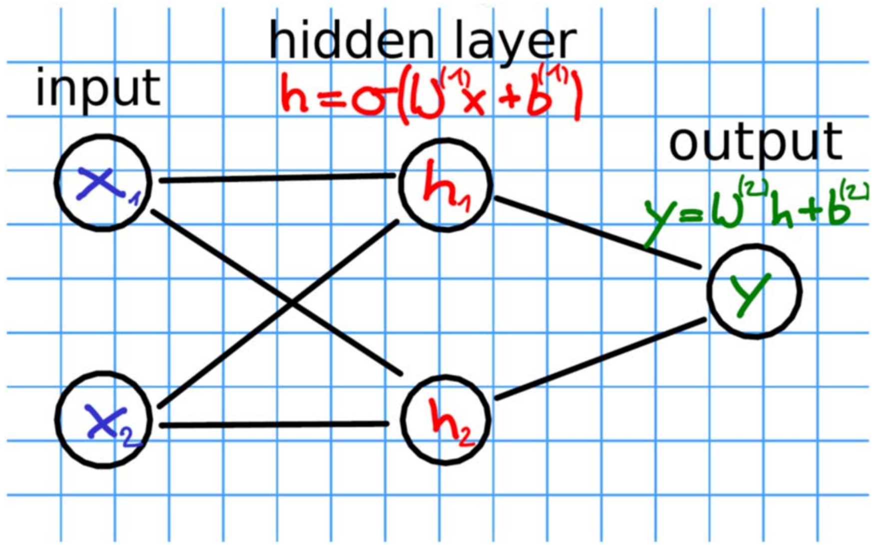
Martin Erdmann, RWTH Aachen University, 27-Sep-2018

1) Rapid development: HEP at new level of Big Data Analysis

Deep Learning, Information Field Theory, Reconstruction Algorithms

Analytics: Deep Learning

Neural networks with ‘many‘ hidden layers



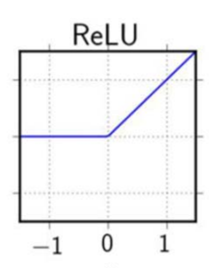
x multi-dimensional input data

W, b to be trained

successively apply 2 operations:

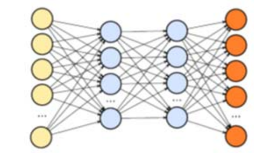
$$\mathbf{y = W\,x + b}$$
$$\mathbf{h = \sigma(y)}$$

departure from linear system

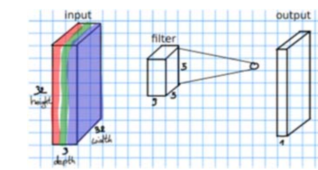


Architectures

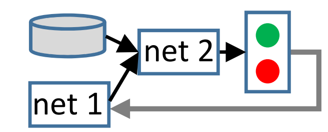
– Fully connected



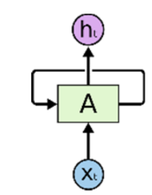
– Convolutional



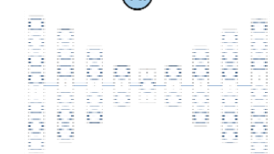
– Adversarial



– Recurrent



– Autoencoder



adapt & combine according to physics requirements

Improved set of tools

Train millions of parameters by:

– Data preprocessing

– Normalization etc

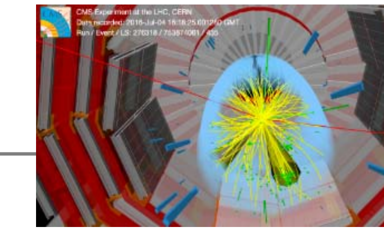
Computing

– Graphics Processing Unit (GPU)

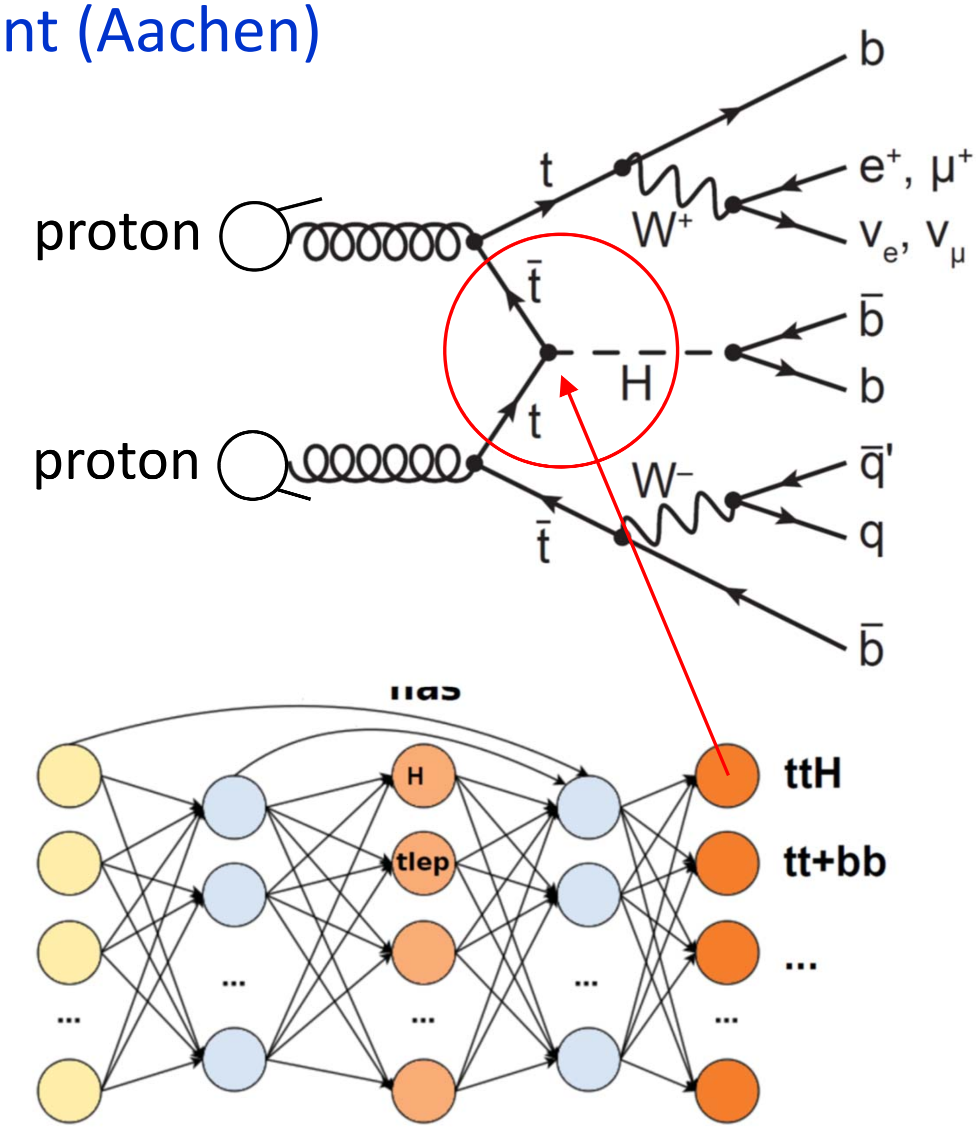
– Software Libraries

CMS Collaboration, Phys. Rev. Lett. 120, 231801 – Published 4 June 2018

LHC: Coupling Top-Quark – Higgs Boson

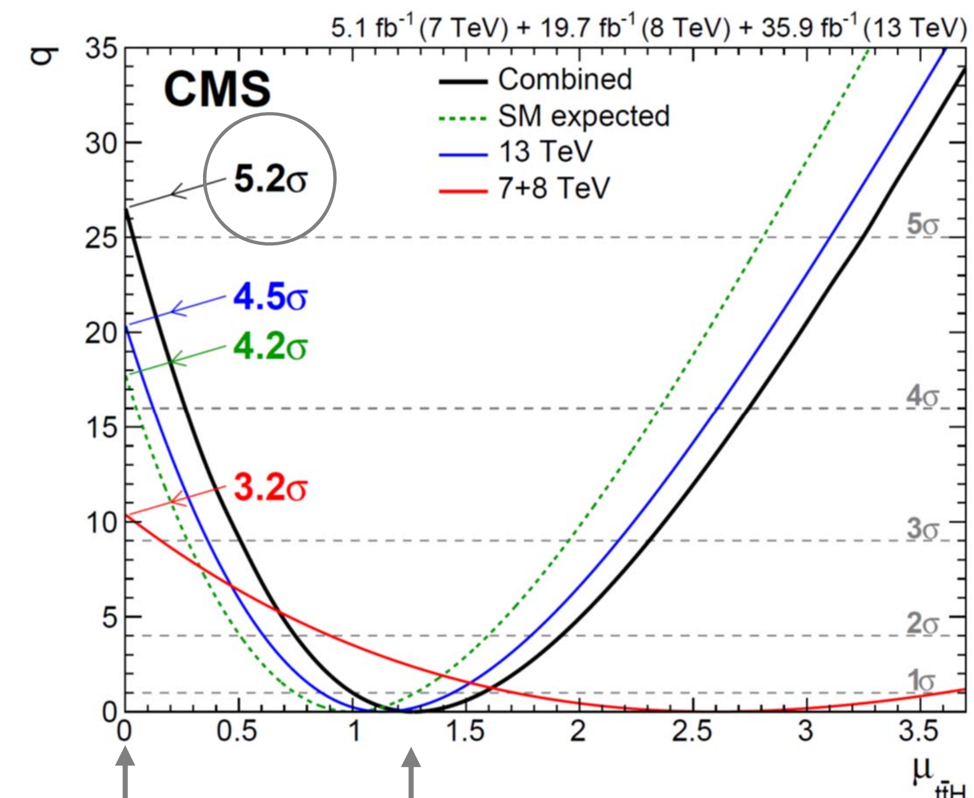


Deep Learning predicts physics process for each event (Aachen)



Observation of ttH production

H→ bb
H→ WW
H→ ZZ
H→ γγ
H→ ττ



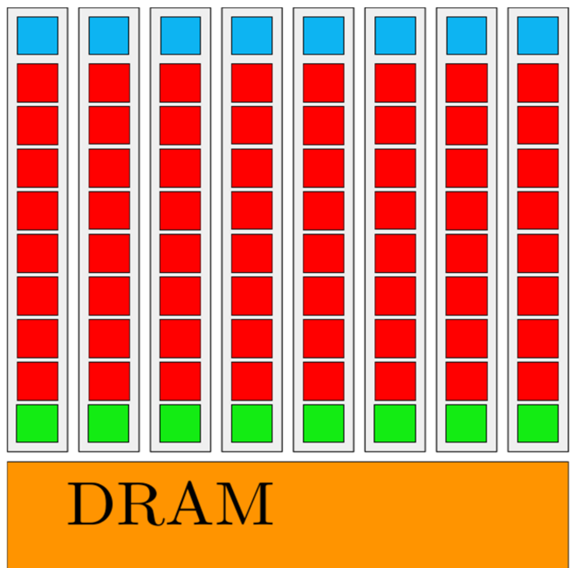
excluded: no coupling

Measured signal close to Standard Model of particle physics (μ=1)

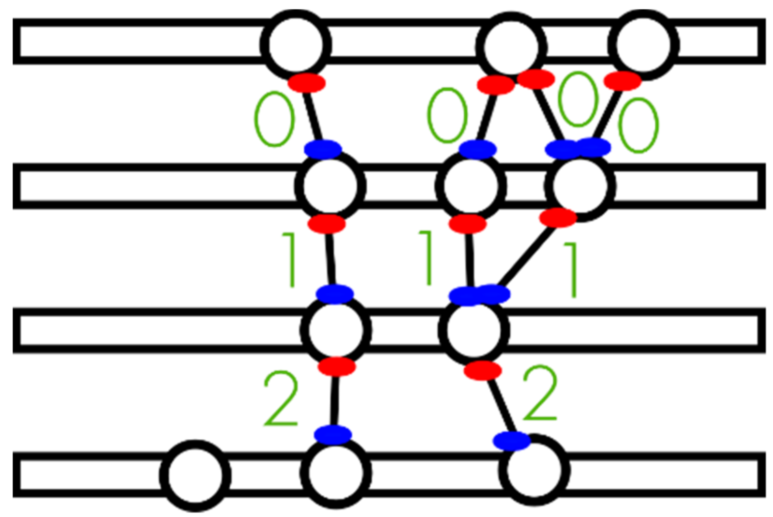
Deep Learning arrived at particle physics publications

Pantaleo, Schmidt, Innocente, Hegner, Pfeiffer, Meyer, CMS-TS-2017-028 ; CERN-THESIS-2017-242

New data processing algorithms: track seeding

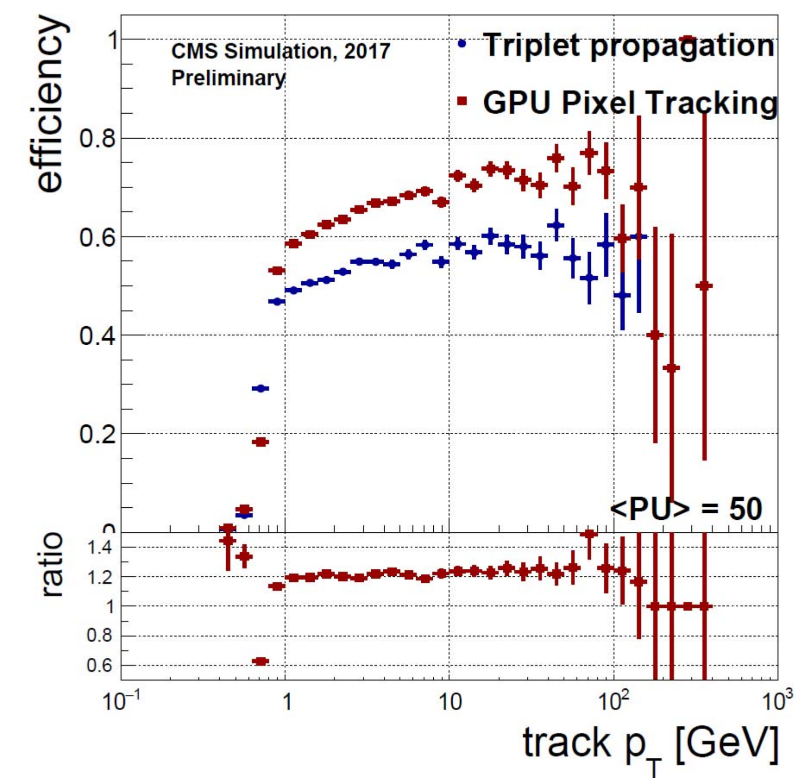


- GPU



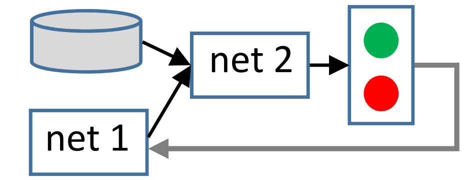
new seeding algorithm based on parallel-friendly algorithmic structure (cellular automaton), computing time grows linear with Pileup

	time per event CPU (ms)	time per event GPU (ms)
Triplet propagation	66.3	
cellular automaton	22	1.6



prerequisite to process the High Luminosity - LHC data

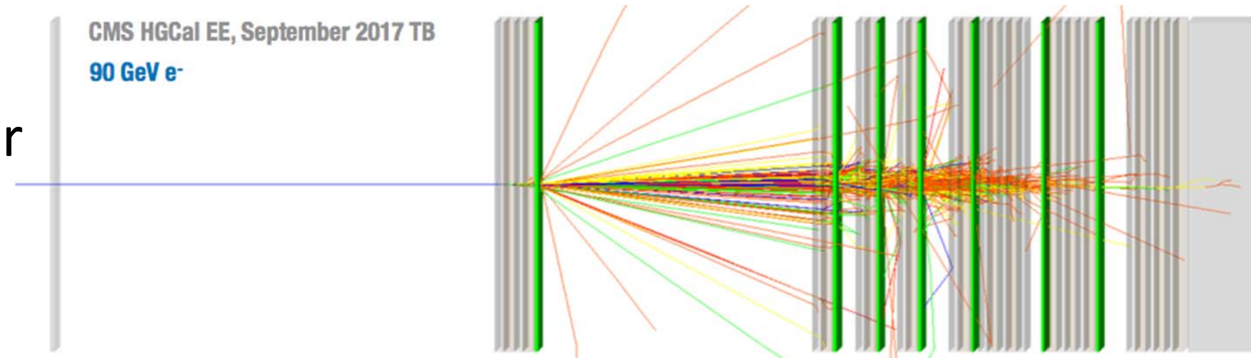
Ultra-fast simulations: WGAN

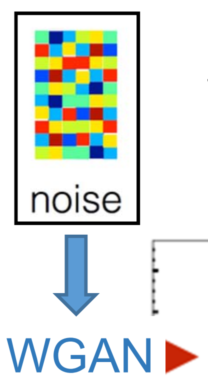


M.E., J. Glombitza, T. Quast

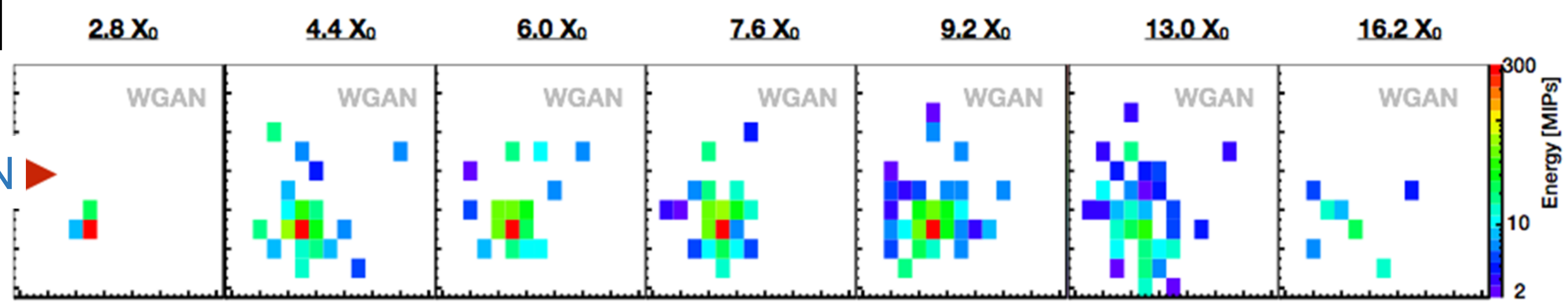
arxiv 1807.01954

Electron calorimeter in CERN test beam





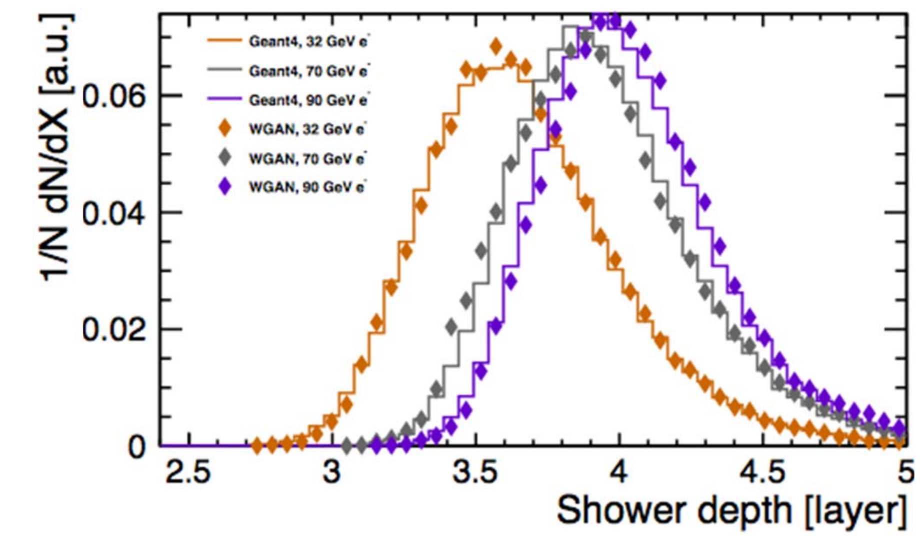
Wasserstein-based **G**enerative **A**dversarial **N**etwork



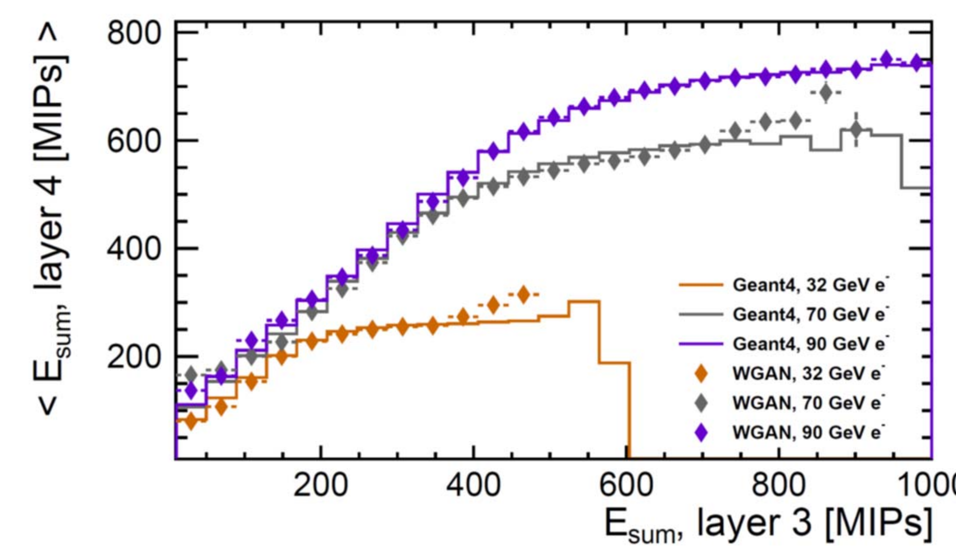
Generation Method	Hardware	milliseconds/ shower
GEANT4	CPU	**2000**
WGAN	CPU	52
	GPU	**0.3**

Goal: use measured data to re-train WGAN

- Shower depth



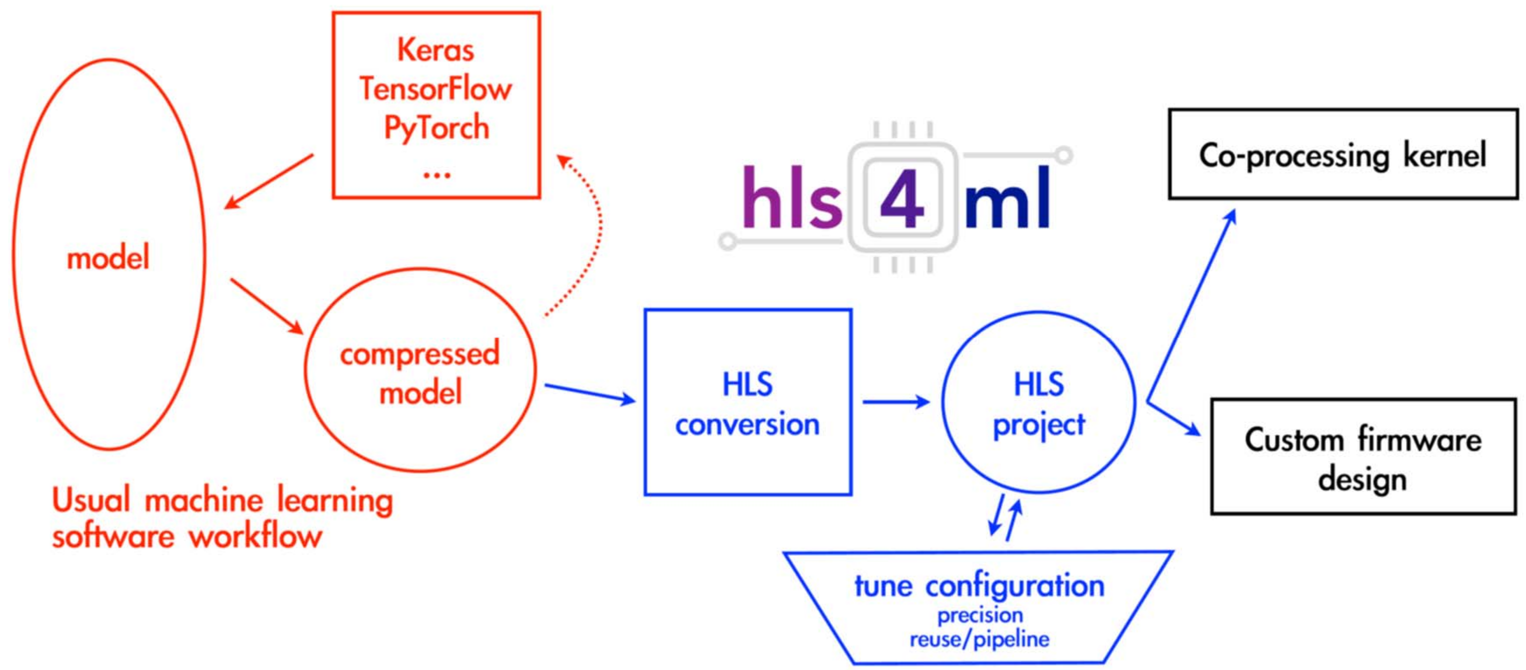
- Correlations between layers



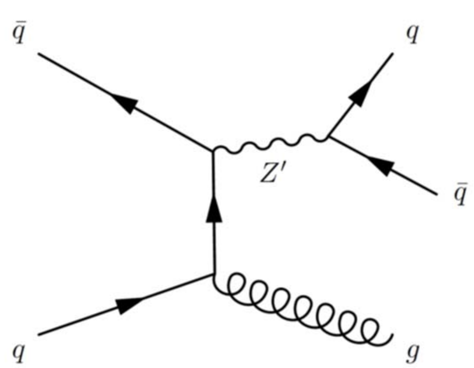
- Challenge: low-energy depositions

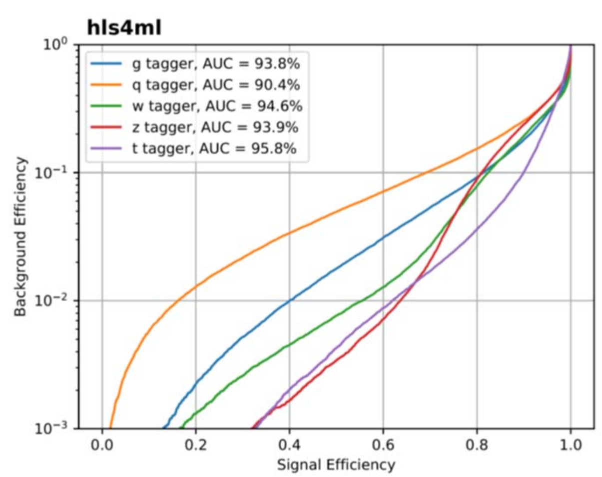
Javier Duarte et al., arXiv:1804.06913

Deep neural networks in FPGAs



Distinguish jets from quarks, gluons, `fat‘ jets





Remarkable

- Fully connected neural network to identify jets with 4389 parameters
- Implemented in FPGA using network compression & reduced precision
- Latency of inference 75–150 ns with clock frequency 200 MHz → LHC Trigger

Information Field Theory

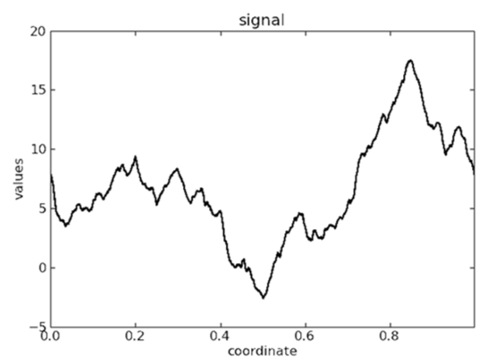
Torsten Enßlin, arXiv:1804.03350
http://www.mpa-garching.mpg.de/ift/

Bayesian method to fuse multiple information sources, learning from a single data set

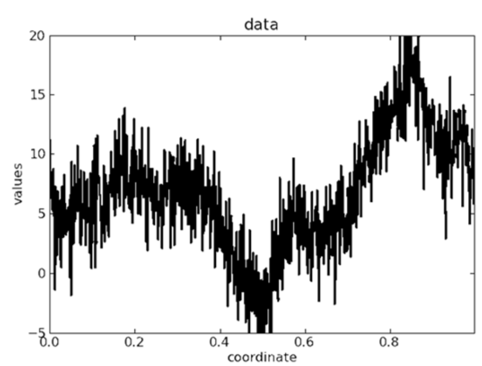
$$\mathcal{P}(s|d) = \frac{\mathcal{P}(d|s)\,\mathcal{P}(s)}{\mathcal{P}(d)} = \frac{1}{Z(d)}\, e^{-H(d,s)}$$

$$s = \text{signal}$$
$$d = \text{data}$$

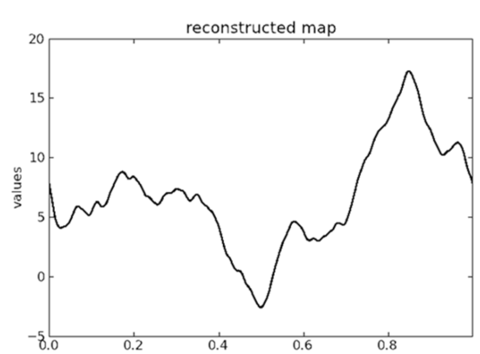
Simulated signal



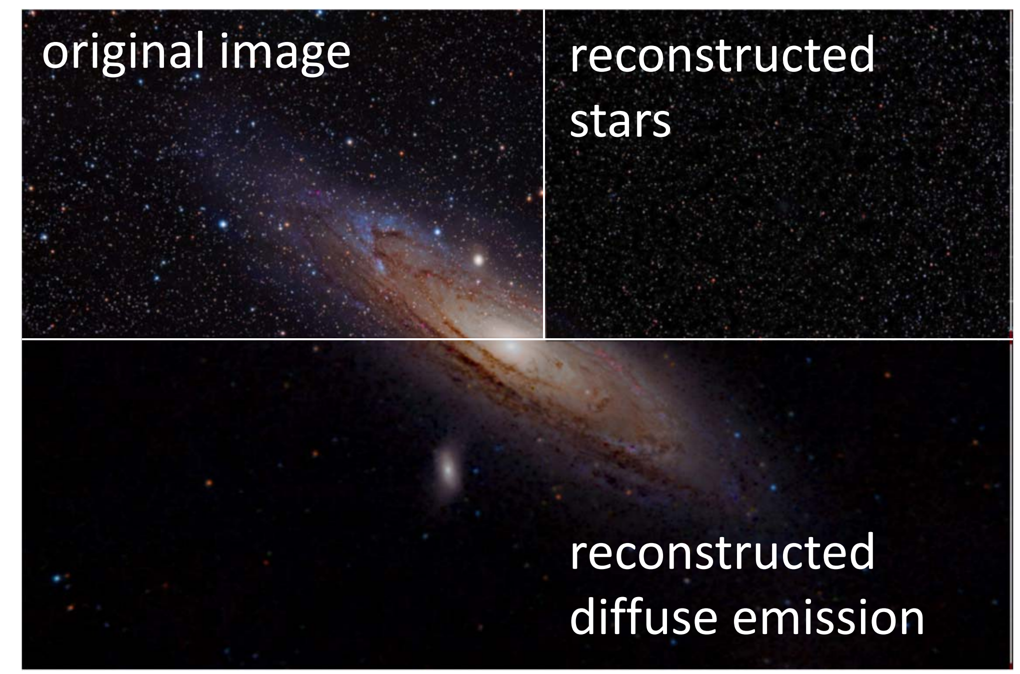
Simulated data



Reconstructed signal



Separate contribution to Hubble's Andromeda galaxy



2) Education: Anchoring machine learning in curriculum

„Broschüre“

KAT + AKPIK: M.E., U. Katz, T. Enßlin, A. Hamm,
K. Mannheim, V. Markl, K. Morik, …

„Broschüre“

„We want to achieve that physics students receive an educational offer in the application of machine learning in physics in the standard curriculum of their studies“ (Uli Katz)

Contents

1. Training Concept: Machine Learning for Physics Research
2. Examples applications of deep learning concepts in physics
3. Recommendations
   - Broadening of the course portfolio...
   - Installation and maintenance of graphics processors at university computer centres...
   - Establishment of a training platform for the exchange of course materials ...

→ Konferenz der Fachbereiche Physik

Draft Version 1.0,
15-Juli-2018

Arbeitskreis Physik, moderne Informationstechnologie und Künstliche Intelligenz der DPG

Ausbildung für Physikstudierende im Erkenntnisgewinn durch moderne datengetriebene Methoden: Vertiefungsrichtung Deep Learning

Autoren: Martin Erdmann für das AKPIK Vorstandsgremium

Die Entwicklung neuer Technologien ist entscheidend für Erkenntnisgewinne in Grundlagen- und angewandten Wissenschaften. Fortschritten bei modernen Datenanalyse-verfahren kommt dabei besondere Bedeutung zu: Sie sind der Schlüssel, um neue Erkenntnisse aus Messdaten zu schöpfen und können im besten Fall entscheidenden Mehrwert aus Messapparaturen erwirken. Zahlreiche Neu- und Weiterentwicklungen im Bereich von Analyseverfahren finden kontinuierlich statt, so dass hier eine Momentaufnahme erfasst wird.

In der Physik haben Datenanalysen schon immer zentrale Bedeutung. Im Vergleich mit bisherigen Analyseverfahren schöpfen moderne Verfahren allerdings die Daten weitgehender aus, was Erkenntnisprozesse zumindest beschleunigt oder sogar erst ermöglicht.

Die Grundlagen für den Erkenntnisgewinn durch aktuelle datengetriebene Methoden sollen daher im Physikstudium stärker verankert und in entsprechenden Kursen gelehrt werden.

Aktuell werden in großer Anwendungsbreite Verfahren der Deep Learning Technologie erprobt, die vielfach herausragende und neue Möglichkeiten dieser Technologie für wissenschaftliche Anwendungen zeigen (siehe unten: Forschungsbeispiele). Ein verstärkender Faktor für den Einsatz in der Physik sind die Entwicklung vergleichsweise komfortabler Softwarebibliotheken durch Großkonzerne, die als Open Source Programme zur Verfügung gestellt werden.

Für die kommenden Jahre wird daher erwartet, dass die Verfügbarkeit, Anwenderfreundlichkeit sowie die Anwendungsbreite der Analyseverfahren mit Deep Learning eine gute Grundlage für die vertiefte Ausbildung von Physikstudierenden im Bereich der Erkenntnisgewinne durch moderne datengetriebene Methoden darstellt (siehe unten Kursbeispiel). Voraussetzung für entsprechende Kurse sind selbstverständlich solide Grundkenntnisse in statistischen Verfahren zur Datenanalyse, die im Bachelorstudium der Physik gelehrt werden.

Fertigkeiten und Kenntnisse in modernen Methoden der Datenanalyse haben über die wissenschaftliche Forschung hinaus Bedeutung. Die Mehrzahl der Physik-Studierenden wird eine Stelle in der freien Wirtschaft annehmen, so dass auch hier aktuelle Datenanalysemethoden zum Einsatz kommen werden, um neu aufkommende Fragestellungen in der wirtschaftlich und gesellschaftlich fortschreitenden Digitalisierung professionell zu beantworten.

1

Education: Deep Learning in Physics Research

12 lectures and exercises

1. Fundamentals of deep learning
2. Regularization & generalization
3. Optimization & hyperparameter tuning
4. Convolutional neural networks
5. Classification of magnetic phases
6. Advanced computer vision methods
7. Application in astroparticle physics
8. Autoencoders & application in solid state physics
9. Generative adversarial networks
10. Restricted Boltzmann machines
11. Recurrent networks
12. Summary & a bit more on recurrent networks

3) Funding: Competence networks of various disciplines

"Digitalisierung in ErUM": BMBF-Workshop 4.-5-Oct-2018

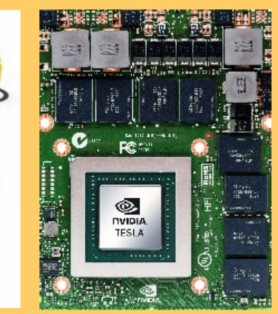
KAT, KET, KfB, KFN, KFS, KHuK, RDS

Federated Infrastructures	Big Data Analytics	Research Data Management

The Digital Basic Delivery for Scientists

1980th
local

2020th: interconnected



Interactive & batch access via web browser to all scientific data, the required computing resources, exchange of information with colleagues

→ Should look & feel as if local
→ New own ideas through better technologies

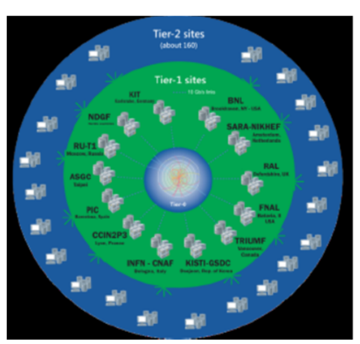
Numerous great activities KET, KAT, …

Gefördertes Projekt: Förderung von ausgewählten Schwerpunkten der Erforschung von Universum und Materie auf dem Gebiet „Physik der kleinsten Teilchen“
Innovative Digitale Technologien für die Erforschung von Universum und Materie

Proposal: Entwicklung und Erprobung von Kurationskriterien und Qualitätsstandards von Forschungsdaten im Zuge des digitalen Wandels im deutschen Wissenschaftssystem.
FAIRe Forschungsdaten aus der Hochenergie Astroteilchenphysik: Teilchenschauer

Federated Infrastructures	Big Data Analytics	Research Data Management

New computing model following The Worldwide LHC Computing Grid



in 2017:
~750k CPU cores
~1 EB of storage
10-100 Gb links
>2 million jobs/day

VISPA internet platform: development environment for data analysis in web browser

Jupyterhub: Multi-user web server for Jupyter Notebooks (Belle II, SWAN@CERN)

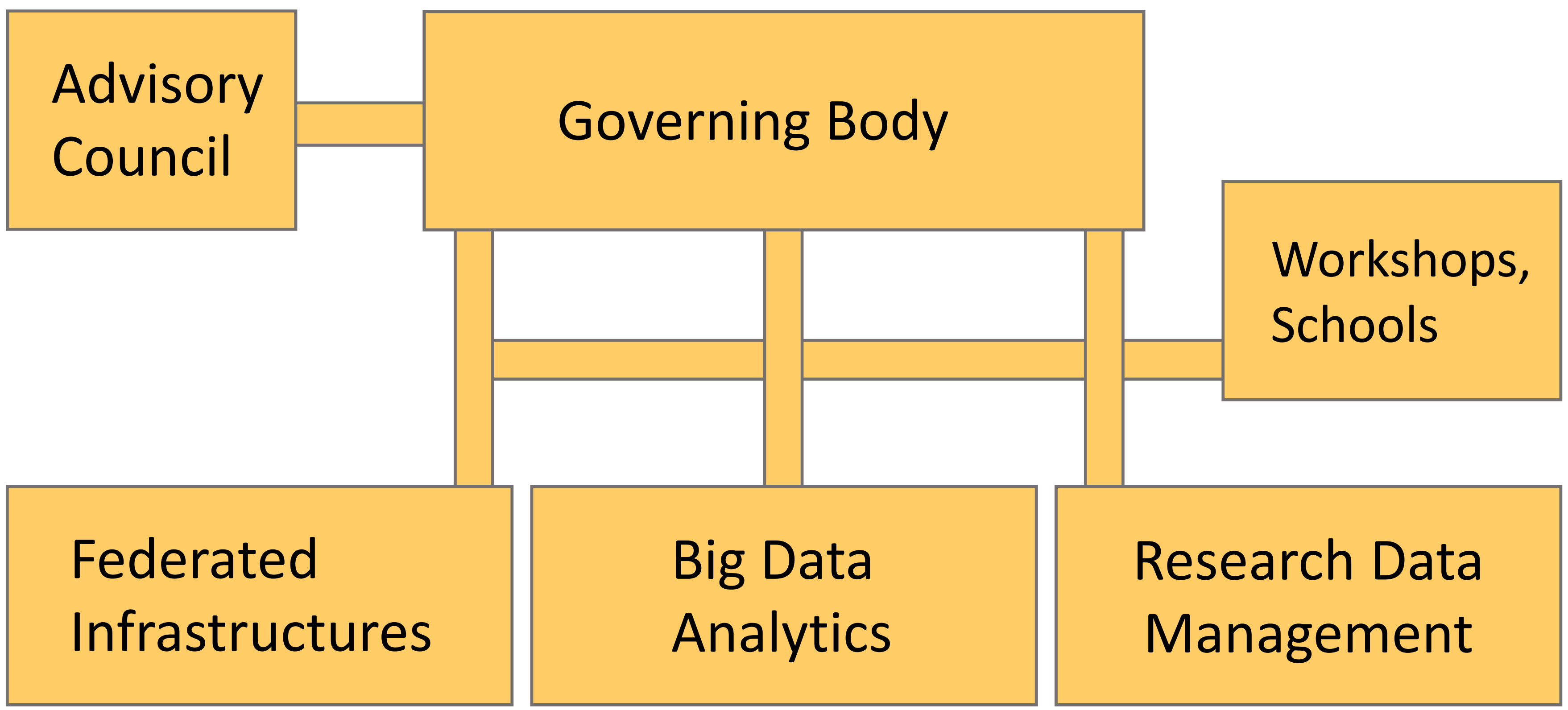
Workshops: Big Data Science in Astroparticle Research in Aachen 2017/2018/2019, CERN machine learning 2017/2018, Dortmund 2019,..

Schools on Computing & Machine Learning: GridKa school 2017/2018, DESY school 2018, Dortmund 2018…

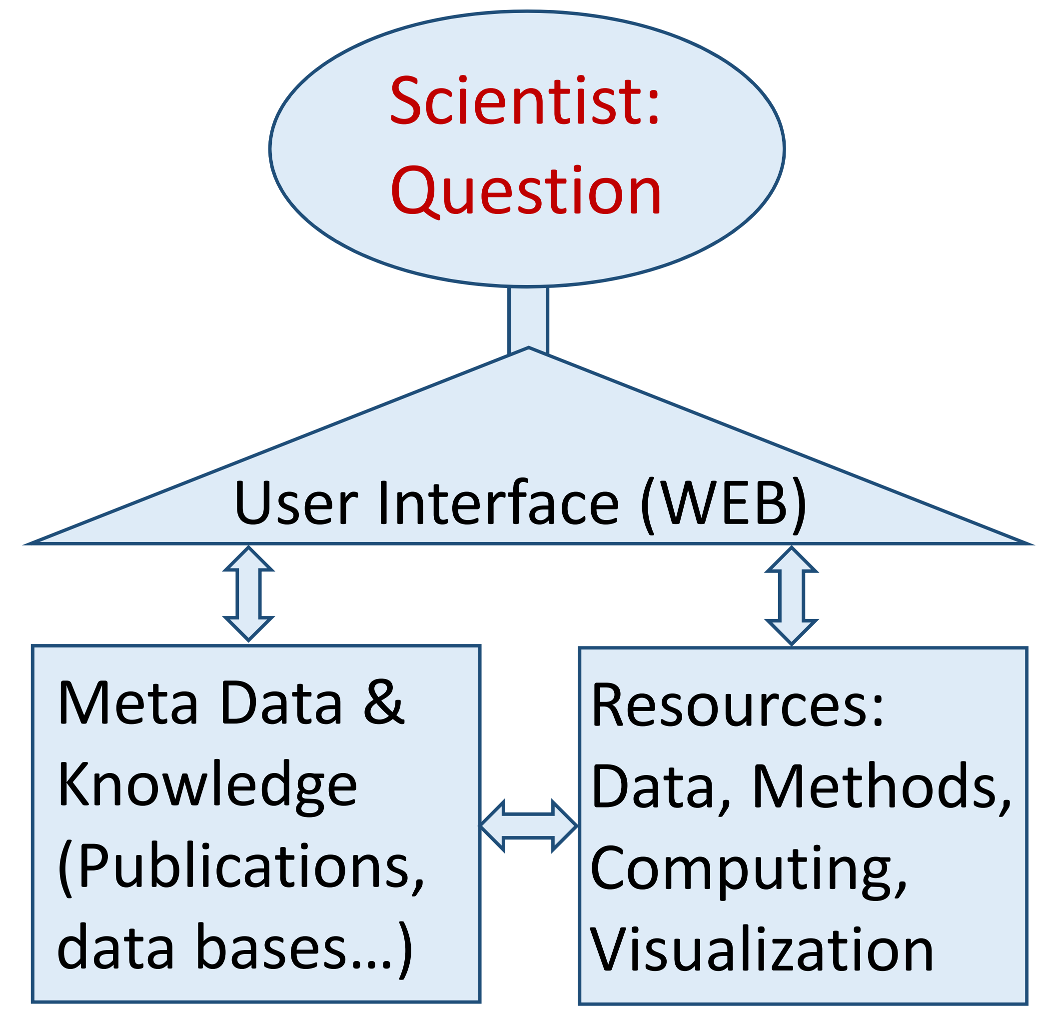
→***We need a home*** *for all these activities to stay focused on the scientist & her/his questions !*

asking for ErUM-Alliance

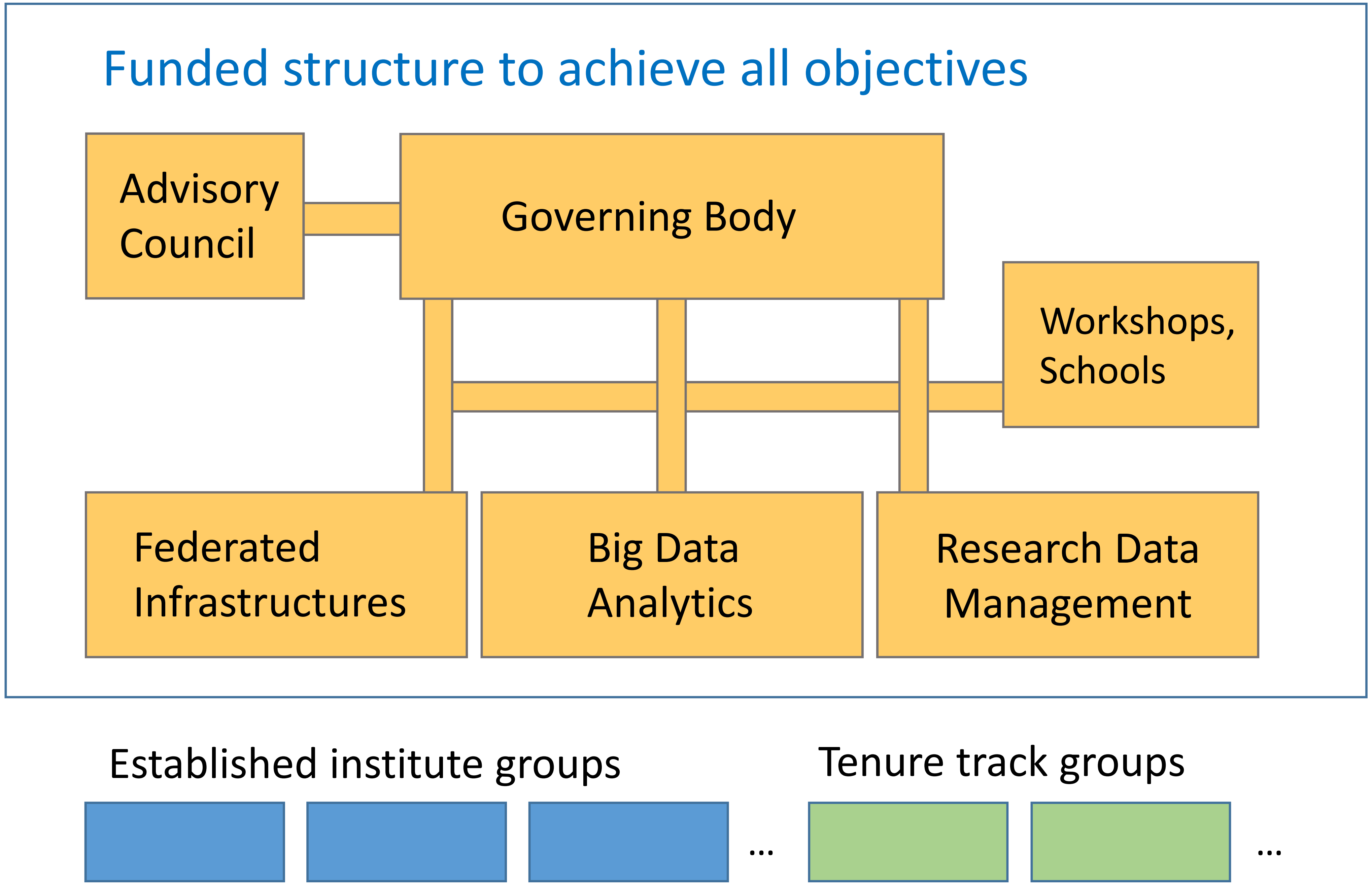
Funded structure to achieve all objectives

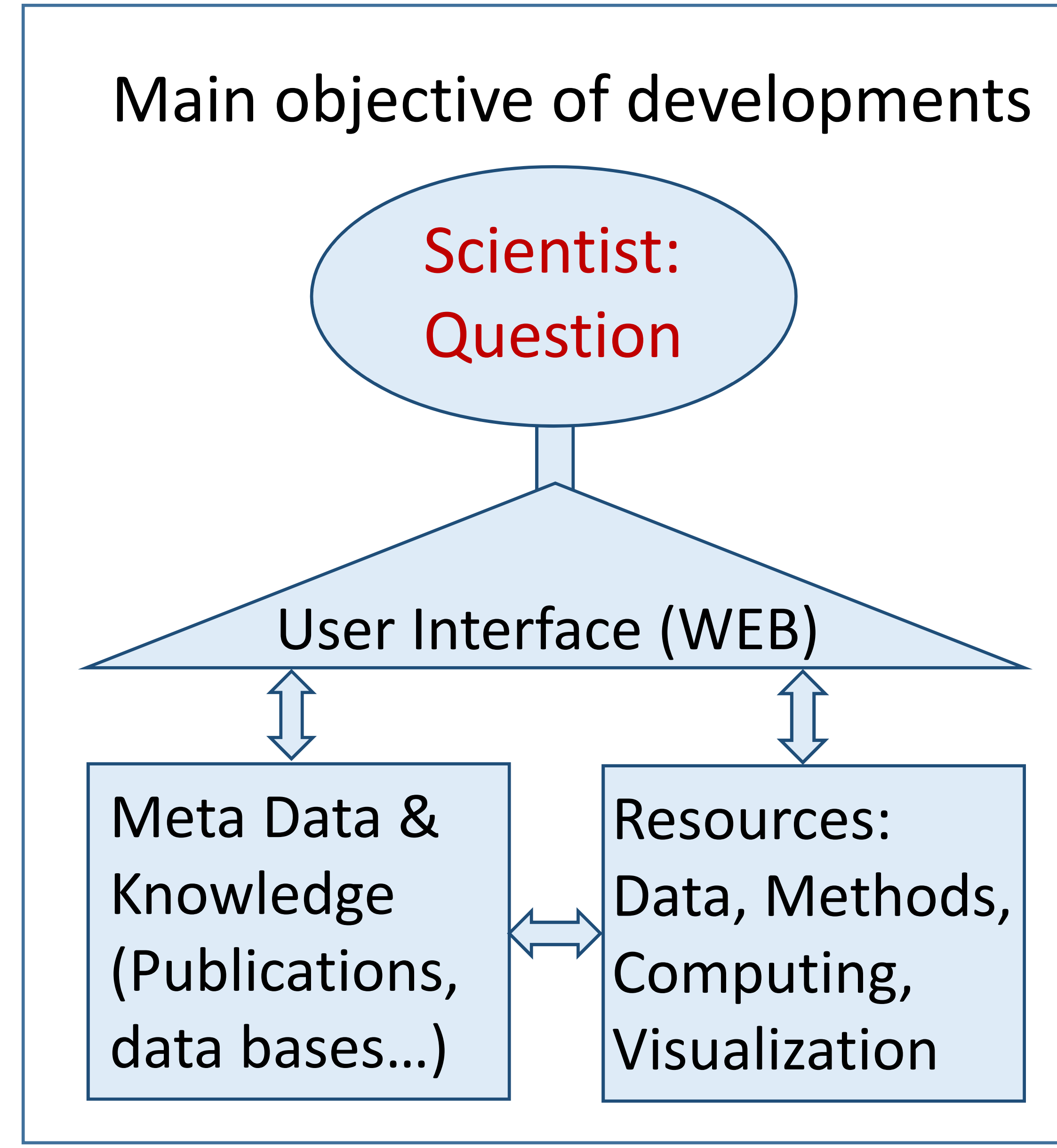


Main objective of developments



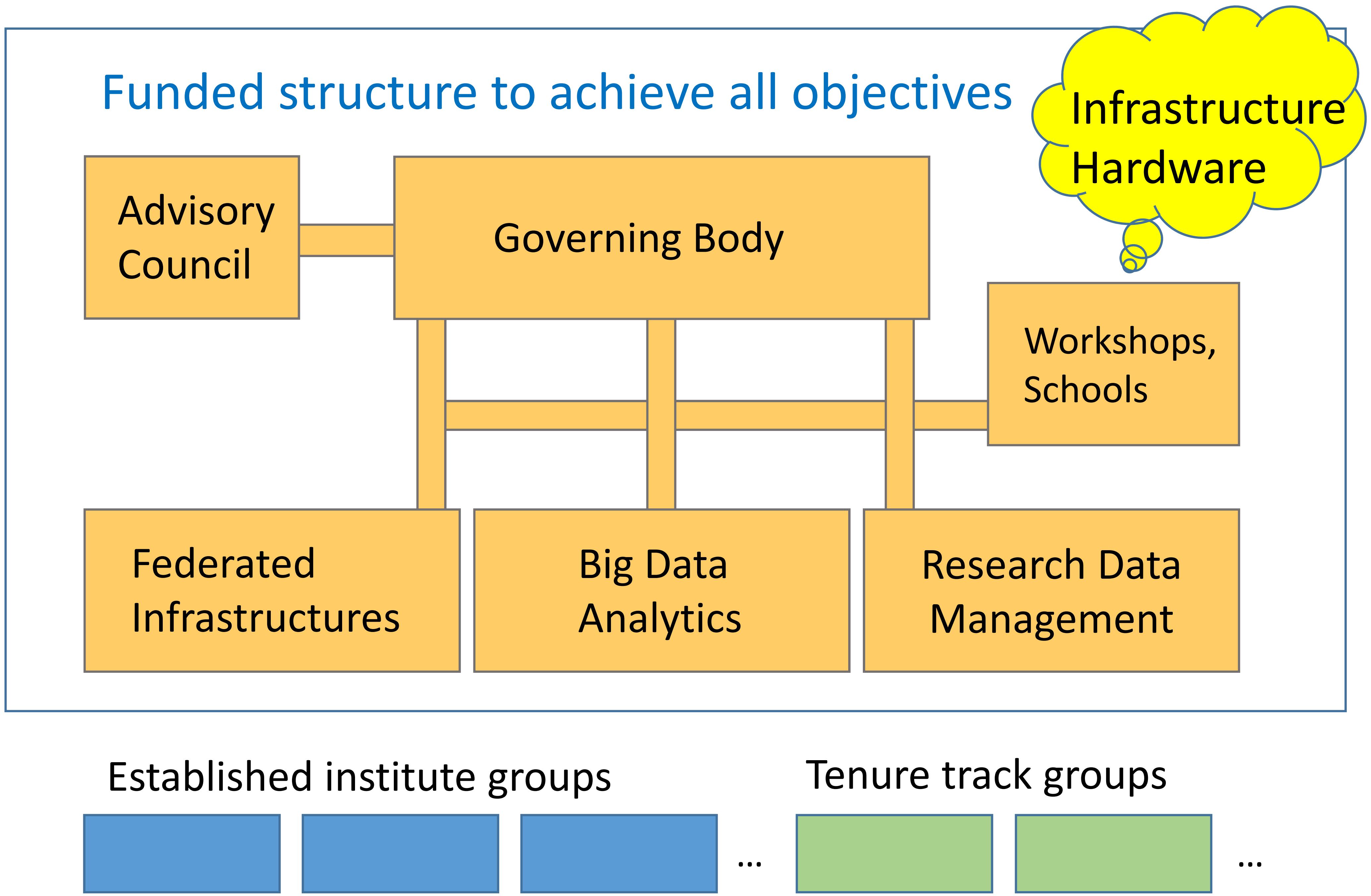
asking for ErUM-Alliance

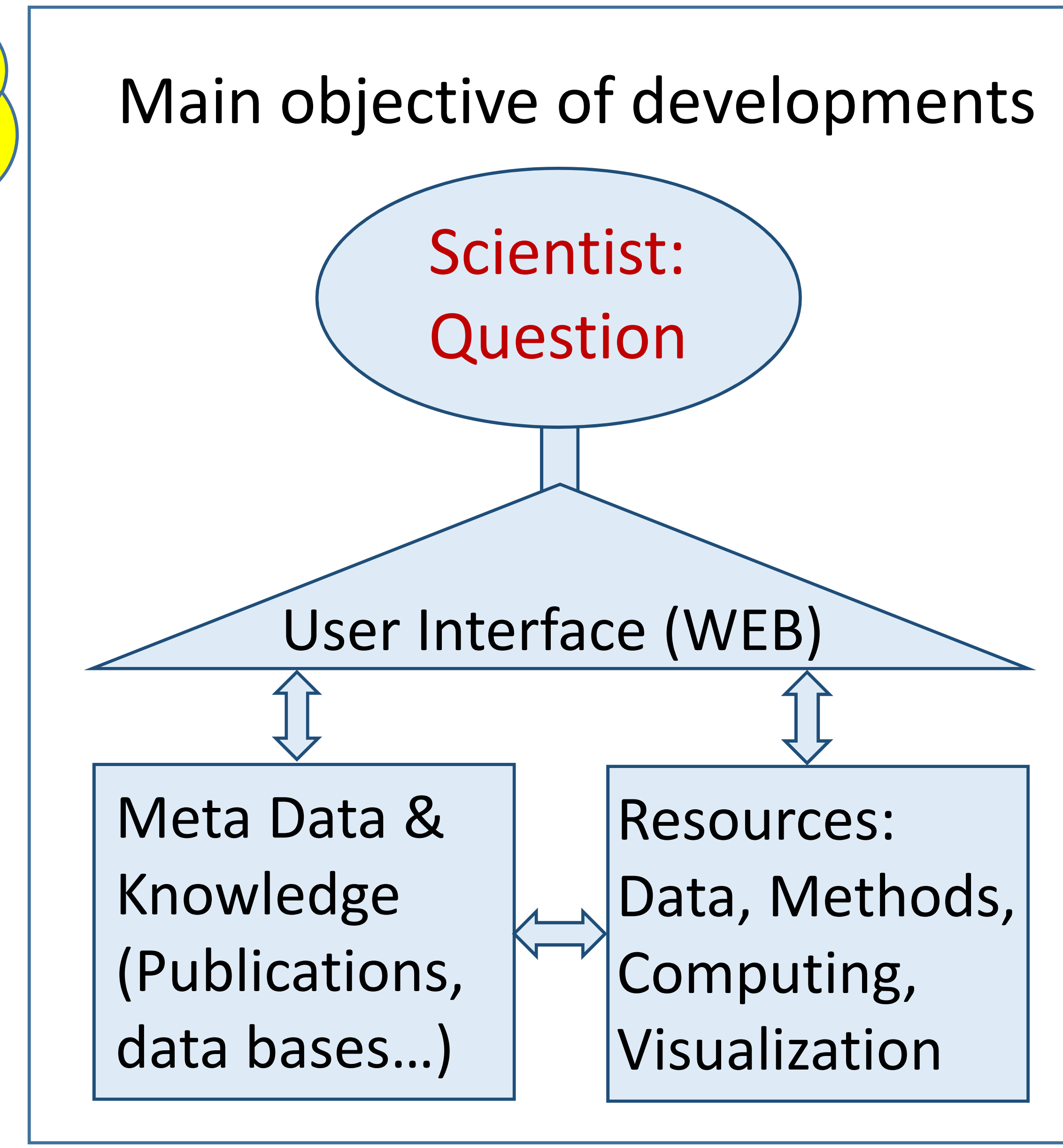




Newly establish strong physicists groups on Big Data Infrastructures & Management & Analytics

asking for ErUM-Alliance





Newly establish strong physicists groups on Big Data Infrastructures & Management & Analytics

BMBF working group: Big Data Analytics

Big Data Analytics

BMBF specifications for working group topics Big Data Analytics:
The considerably higher resolution of detectors and experiments is accompanied by a rapid increase in the resulting measurement data...

Translation into concrete discussion topics

- Trigger: "on-the-fly" data analysis & data selection
- Simulation: new approaches e.g. generative models
- Algorithms: optimal match with computer technology, scalable, parallel, "green
- Signal/background (noise): machine learning (deep learning, information field theory)
- Systematic effects: new approaches e.g. adversarial decrease
- Precision: Combination of big data methods for more accurate results
- Suitability: Big Data methods also for complex, unstructured data sets
- Results, Meta-Results: New Conclusions from Data Using Big Data Methods

Artificial intelligence?

Big Data Analytics

Consideration of other aspects

Disciplines (particles, matter and universe) with the same challenge:

- *Access for all*: new methods for efficient use of measurement data
- *Method development*: develop new, specifically designed methods.

Aim of this working group: "common challenges & concrete approaches":

- *Enable solutions as well as cooperation* between different areas.
- *Mathematics & Informatics*: Opportunities for interdisciplinary cooperation & possible approaches.

Additions:

- *Young talent qualification*
- *Interdisciplinarity*
- *internationalisation*
- *Cooperation with industry*

Questions to KET: approximation of language comprehension & needs of community

Big Data Analytics

Questions and expectations for Big Data Analytics
- Which questions should be solved?
- What expectations are associated with Big Data Analytics?
- How much experience is there already?
- Evaluation of past experiences?
- Which challenges need to be solved?

Data level for Big Data Analytics
- Online "at the measuring sensor"?
- Offline from files?
- On single events (e.g. collision events, X-ray images, ...)?
- On the totality of data sets (e.g. sky maps, ...)?
- Which data structures and formats are used?
- Who stores the data where?

Data volume for "Big Data" analyses
- Typical data volumes per analysis
- Analyses performed per year
- Bytes (or number of variables) per record element

Analysis methods for Big Data Analytics
- Adaptations with very large numbers of parameters (e.g. fits with >10,000 parameters)?
- Methods with high demands on training data (deep networks)?
- Methods with numerical inference methods (e.g. Bayes: information field theory)?

Existing and expected cooperations Big Data Analytics
- Within your own community area (e.g. only particles)
- Neighboring communities (particle astroparticles, synchrotron neutrons, ...)
- International Partners
- Mathematics, Computer Science
- Economy

Need your input now

Messages

- **Lively exchange on current developments in "digitization" in particle physics:**
  - Machines exploit physics data more deeply
  - New data processing algorithms: prerequisite to High Luminosity - LHC data
  - Improved digital basic infrastructure: new own ideas through better technologies
  - Education: training the next generation of scientists in new developments

- **Strong desire in KAT-KET-KHuK Community for joint & interdisciplinary developments:**
  - Needs structure for the "*home*" of the activities (similar to Allianz, FSP,...) for coordinated efforts

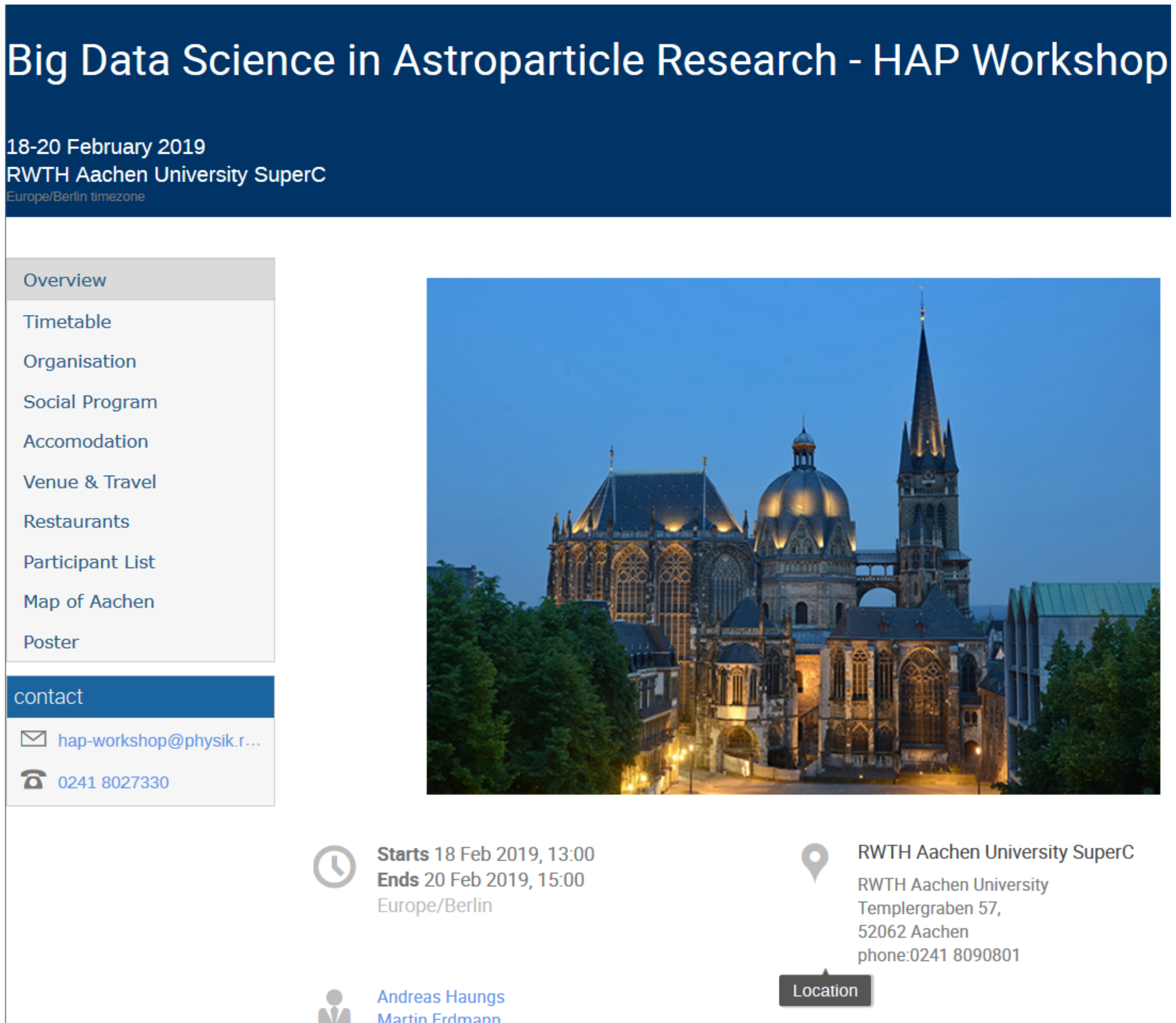
backup

DPG Arbeitskreis „Physik, moderne Informations-technologie und Künstliche Intelligenz“

1. BIG DATA
2. IT
3. KI & ROBOTIK
4. HOCHSCHULE
5. INDUSTRIE und GESELLSCHAFT

http://www.dpg-physik.de/dpg/gliederung/ak/akpik/organisation.html

Big Data Science in Astroparticle Research (III.)



Aachen, 18.-20. Feb-2019

We invitation your contributions

- Progress in deep learning applications
- Data management and data centers
- Software
- Analysis preservation
- Platforms for algorithms & network architectures
- Education material

Keynote speakers

Hands-on tutorial deep networks