## PUNCH4NFDI - Training, Education, Outreach

Frank Bertoldi, Kevin Kröninger, Arnulf Quadt, Michael Kramer

## Challenges

* Research: Imbalanced educational background of professionals: on-the-job training, learning by doing; Expertise grown on demand, does not transmit throughout community
* Career: Expert personnel (brain) drain due to lack of career options, low pay, gender gap
* University: Physics university curricula lack data science, big data management, advanced programming
* Public Outreach: depends on sporadic resource allocation and lack coordination
* Public relation to science suffers from passivity, lack of context, mismatch to educational level
* Physics in Schools: Data science not represented in school curricula; decline of school physics education, strong gender gap.

## Mission

* Identify and address short- and long-term needs in formal, professional, and public education
* Establish hub & clearinghouse for information & activities; Provide and coordinate open access to data and training resources
* Train the professional physics community on data science methods and access;
* Guide implementation of best practices, support career advancement
* Educate and engage the society at large in data science, especially on Big Data aspects; Coordinate and initiate education, outreach, citizen science programs within existing and new frameworks
* Focus on structural challenges related to gender equality and access to computing and information resources.

## Specific Measures

#### Training Experts

* Training in the application of the tools and practices developed within PUNCH4NFDI
* Coordination and organization of specific training measures and events on NDFI topics for physicists.
* Promotion of measures for career advancement in the data sciences.
* Emphasis on the promotion of women.
* Critical review of the measures developed and development of a feedback system to guide education & training
* Documentation and long-term archiving of training material, the service documentation and tool descriptions in coordination with the TIB Hannover

#### Educating Students

* Market survey of available teaching concepts and material
* Development of standardized curriculum
* Compilation and development of teaching material for courses and independent learning.
* Aggregation of data resources and access to computing infrastructure
* Coordination and initiation of educational events, e.g. visiting seminars.

#### Outreach and Schools

* Website, blog, twitter, communication of PUNCH products
* Networking, PO training, study on effectiveness of outreach, development of evaluation criteria
* Central access to data and software for education and entertainment
* Provide tutorials and material for teachers and students, teaching and learning material for masterclasses
* Study and test pedagogical approaches to promote data science for children
* Pilot extracurricular activities (AG/Projektkurs) in selected schools, promote changes in school curricula
* Organize/support hackathons, rent a scientist, masterclasses (Netzwerk-Teilchenwelt), VO-days, etc.

#### Citizen Science

* Map out potential research applications in Citizen Science
* Prepare data sets and providing soft- and hardware infrastructure
* Run pilot projects for 6-12 months by engaging schools and universities and evaluate results
* Prepare and launch of further projects in collaboration with the physics community

The training of our professional community and of students is a central concern of PUNCH4NFDI.

We need to provide colleagues and students with access to expertise or know-how, to methods, tools, data, or any relevant training resources related to data science and data management in the context of physics research. We believe it is also important to inform, educate and actively engage society at large in data science aspects of our research, and for this also to provide some access to data resources. For the planned measures we are guided by what we perceive as the most important needs or deficits.

1. Physicists are typically well-trained in mathematical methods that relate to the physics problems they work on. However, often they are less prepared in data and computer science, in statistics, in the extraction of reliable information from complex, noisy data sets.

Physicists also know how to program and process data, on a moderate scale, but many are not really equipped for the world of Big Data, with its particular challenges in data storage and data access, or in the analysis using high-performance computers. In our physicist's culture of learning-by-doing, we are also not so familiar with modern software-development and documentation techniques, or how to implement the FAIR principles.

As a specific measure, we will establish a **PUNCH Young Academy** that offers career-development support to scientists with non-permanent contracts, with a particular focus on young female scientists. This academy can have a number of aims and tasks, such as

* collect **material for general career development**,
* collect and update **information** for the introduction of Academy fellows to the German academic system and opportunities for funding,
* organize **best practice and exchange fairs** including representatives from German funding agencies or foundations and institutions supporting science
* identify potential **mentors** among the established senior scientists in our community,
* establish an **alumni platform** and organize alumni days where present fellows can meet former members of the PUNCH community who have moved on to industry and now look for new, young talents.
* put a strong focus on career development for women, or minorities, on their challenges, but also their special opportunities and chances. **Role models** in the national and international community will be invited, special aspects of mentoring will be offered and additional, specialised career development courses will be established.

2. Ideally, university curricula for physics should provide the skills, methods, and tools needed to cope with data. Many of the students later on need to handle large and complex data sets. This is mostly not the case, since courses on modern statistics, applied mathematics, or computer science are usually not mandatory. In fact, many curricula lack basic training on software design principles,

on data structures, sampling methodology, statistics, algorithmic design and optimisation, standards and procedures. Even on the master or doctoral level, there is yet little systematic training in data science, on big data management, on modern statistics, or on machine learning. This deficit is aggregated by the rapid development of the field, which makes any existing course material outdated quickly.

3. Modern science relies more and more on big data, on artificial intelligence or machine learning.

These are actually disruptive technological advancements that in the general public meet fear, suspicion, but also interest especially with young people who love to explore new IT technologies and services. Communicating science to a wide public is therefore important, whether they are interested or not, because it does affect them. However, outreach has never been near the centre of scientific practice. It's usually added on, without dedicated resources. This is a problematic deficit because fundamental science depends not only on public financial support, but on an educated and well-informed public that actually understands and appreciates the science approach, the scientific method. There is wide public interest in the fundamental questions of astronomy and physics. It is therefore quite easy to meet public interest, but it is a bit more difficult to actively and critically engage the public, which is needed to foster a deeper understanding and some kind of empowerment on aspects of Big Data and Science.

4. Because astrophysics and data science are usually not part of the school curriculum, most children have no connection to the rapid developments in these areas. The decreasing popularity of physics in schools does not help, and neither does the dramatic shortage of qualified physics teachers.

There are many initiatives and institutions in our community we can connect to, e.g.

* CERN School of Computing,
* GridKa School,
* the education programs of the Helmholtz Alliance "Physics at the Terascale",
* the International Virtual Observatory Alliance,
* Physikwerkstatt Rheinland
* Schülerlabor initiatives
* Netzwerk Teilchenwelt
* citizen science projects such as Zooniverse, Muon Hunters, CREDO, Einstein@home, (Folding@home Corona)

5. The strong fluctuation of expertise in data science in our communities, because that expertise is grown on demand and disappears again when the experts find better-paid jobs in the private sector.

6. There is a strong gender gap in physics and computer science and it seems very difficult to close that.