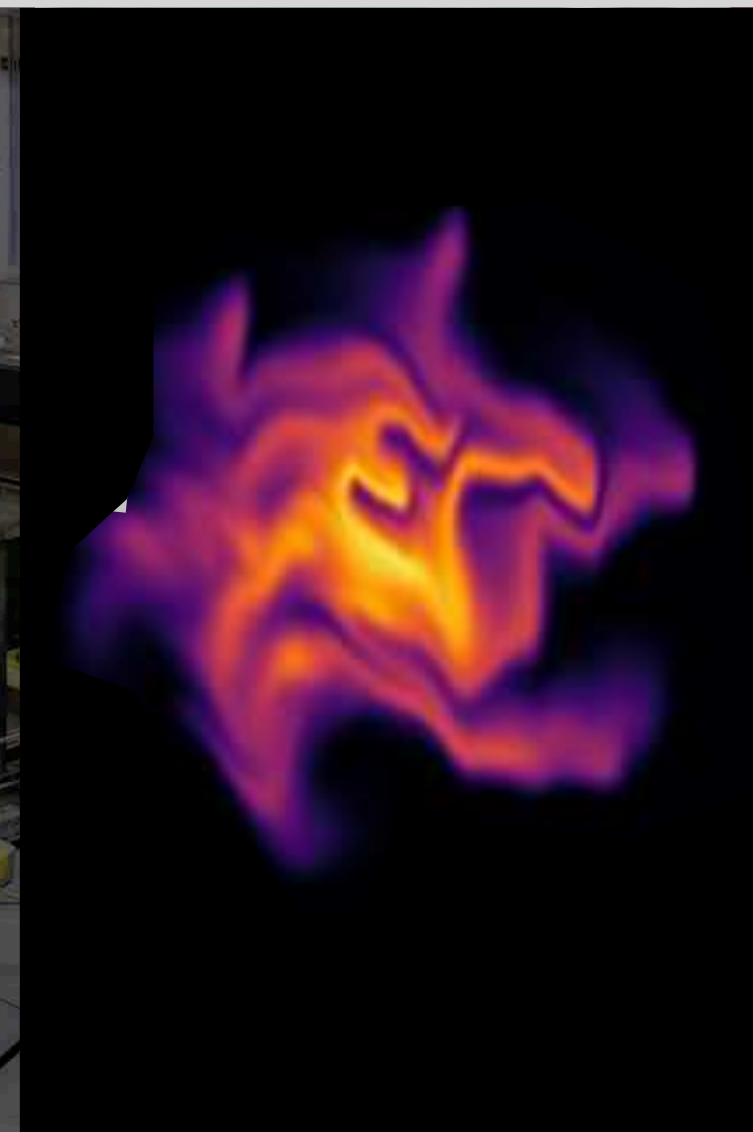
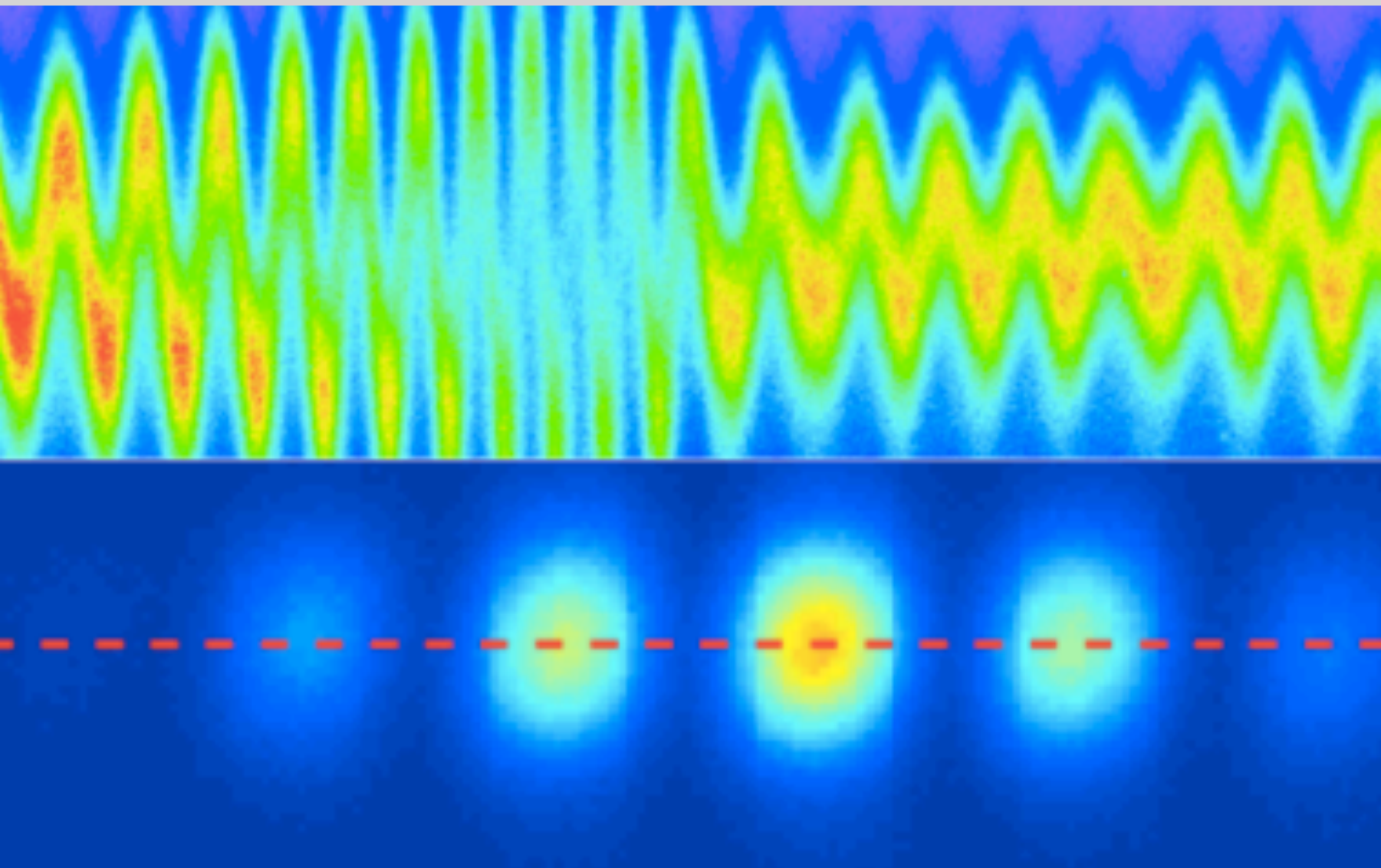


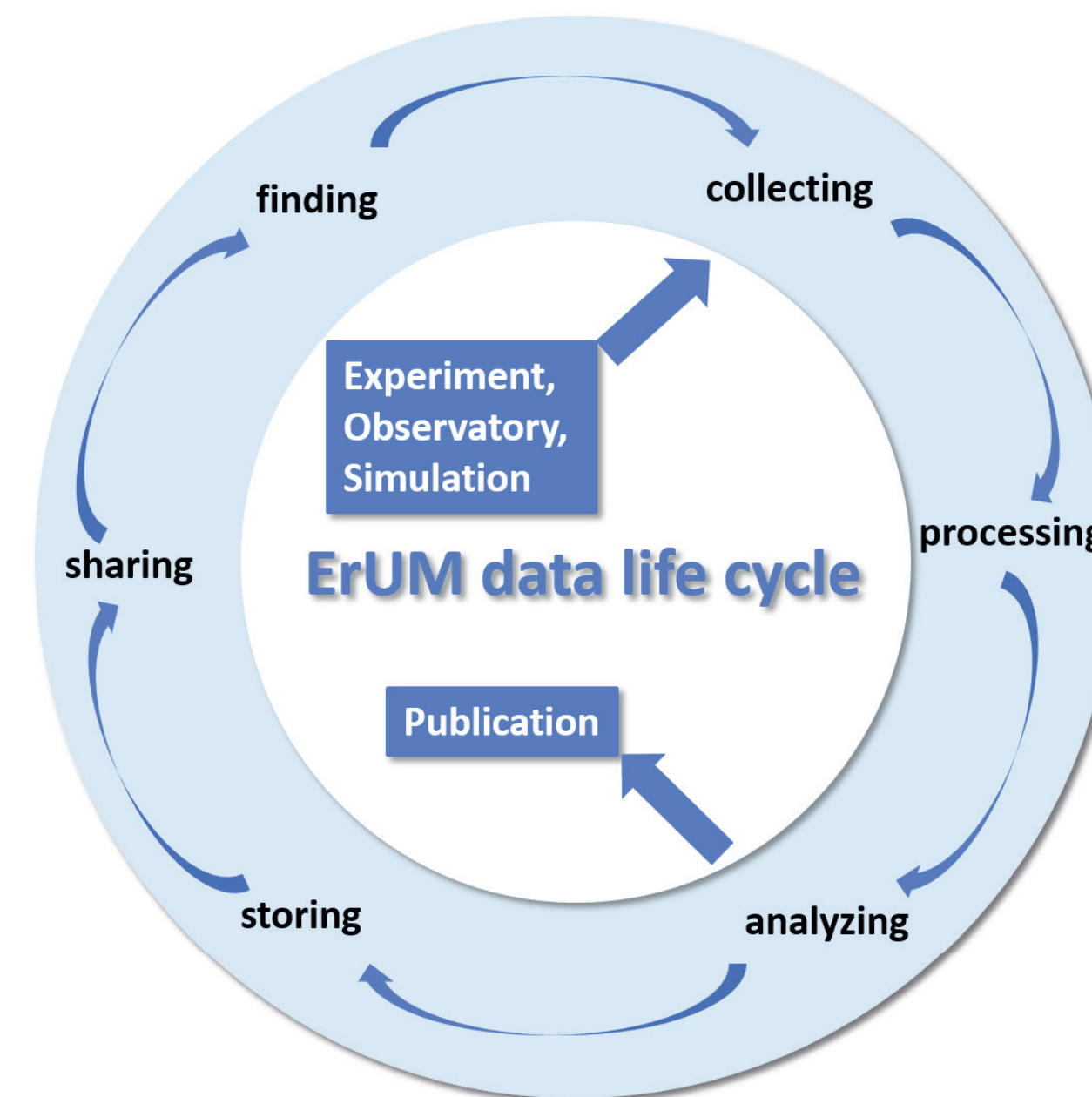
# Digitalisierung

A.-S. Müller



# Digitalisierung in der Grundlagenforschung

- Digitalisierung betrifft alle Wissenschaftszweige, auch Beschleunigerforschung
- BMBF-Workshop am 4./5.10.2019:
  - Erarbeitung Grundlagenpapier
  - 3 Arbeitsgruppen:
    - Förderierte Infrastrukturen
    - Big Data Analytics
    - Forschungsdatenmanagement
- Dachorganisation



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

# Grundlagenpapier

1 The digitization era in fundamental research	4
1.1 Impact of basic research	4
1.2 Prerequisites for new discoveries	4
1.3 Big data science in basic research	4
1.4 Scientists' institutions and working environments	4
1.5 Scientific computing in ErUM	4
1.6 Survey of areas with development needs	4
1.7 Workshop and report of the outcome	4



Figure 1: In the 2020's, the scientists' research environment will look and feel like working in a local environment, allowing access to large-scale data volumes and massive computing resources to process the data, as well as interconnecting with colleagues.

# Grundlagenpapier

---

2 Working Group 1: Federated Infrastructures	7
2.1 Introduction.....	7
2.2 Status and requirements.....	7
2.3 Conclusion.....	8
3 Working Group 2: Big Data Analytics	9
3.1 Introduction.....	9
3.2 Status and requirements.....	10
3.3 Conclusion.....	10
4 Working Group 3: Research Data Management	11
4.1 Introduction.....	11
4.2 Status and requirements.....	11
4.3 Conclusion.....	12

# Grundlagenpapier

---

5 Recommended measures and cost estimates	13
5.1 Federated infrastructures .....	14
5.2 Integration of workflows to exploit infrastructures .....	15
5.3 Comprehensive management of research data .....	16
5.4 Modern Big Data Analytics in fundamental research .....	17
5.5 Scientists' integrated web working environment.....	18
5.6 Tenure-Track programme: knowledge in digitization.....	20
5.7 Partnership for Innovative Digitization .....	21
5.8 Cost estimates.....	23

## Heterogene Community:

- Strahlungsquellen (FELs, Synchrotrons, ...) an Großforschungszentren
  - Collider
  - Beschleunigertestanlagen (SRF, RF, Plasma, Dielektrika)
- heterogene Anforderungen

# Föderierte Infrastrukturen



## Problemklasse a) Bsp. feedback loops, full start-to-end simulations, machine optimization

- Computing Modell: lokal – high throughput + high performance, high data rate (TB/s)
- Archivierung, Datenreduktion für User sind Big Data Challenges
- Spezialarchitekturen werden intensiv verwendet: GPUs + FPGAs in Kontrollsystemen

## Problemklasse b) Bsp. Simulation von Plasmabeschleunigern

- Computing Modell: föderiert (z.B. JEWELS) und lokal, high performance computing, multi-million core hour Einzelsimulationen mit reduzierten Modellen
- Verwendung von Spezialarchitekturen: GPUs, KNL, ...

# Föderierte Infrastrukturen

---

## Zukünftiger Bedarf im Vergleich zu heute

- Sowohl HTC Echtzeitanwendungen (machine feedback + optimization) als auch HPC Stapelverarbeitung (Start-to-end Simulationen, Plasma Wakefield)
- In Zukunft werden beide Anwendungsfälle kombiniert werden müssen
- Erfordert signifikantes Wachstum bei Rechenleistung, Datenvolumen und Netzwerkbandbreite (Zeitraum ~10 Jahre)

## Nutzung von Cloud Technologien?

- HPC Science Cloud nützlich für Simulationen (auch in Verbindung mit Container), bisher nicht implementiert, bedarf F&E Support
- HTC Probleme scheinen ungeeignet
- Expertensupport für die Lösung zukünftiger Computing Probleme im Feld der Beschleunigerforschung muss gestärkt werden.
- Heterogene Problemstellungen bei Optimierung von Algorithmen, Methoden, Analysewerkzeugen etc. auf Multicore und Spezialarchitekturen

*J. Osterhoff*





# Big Data Analytics

- Questions and Topics

- Data analysis
- Simulations and modeling

Offline



Precise/accurate modeling for better control

- Tuning, optimization and control
- Prognostics, alarm handling, anomaly-breakout detection

Online

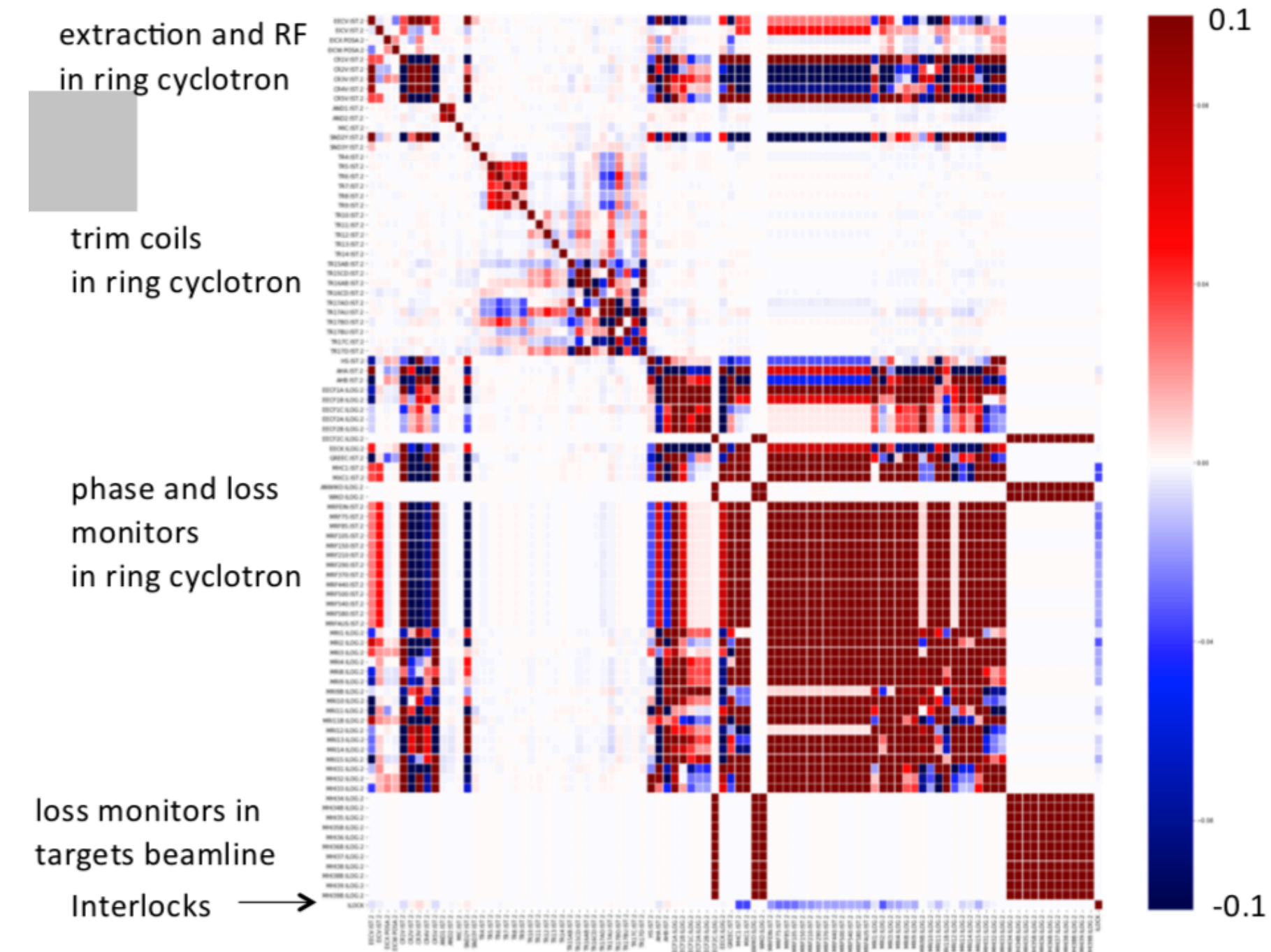
- Challenges

- Rare events (eg UFOs in LHC)
- Very short reaction time (< ms)
- Large variability and high dimensionality of data

- Data mining

- on full set of operating data
- as well data from components

## Beispiel: ML @ p cyclotron



S. Appelt 

# Forschungsdatenmanagement

---

## Metadaten: Auffindbarkeit (Basis DOI/DataCite Kriterien)

- keine übergreifenden Definitionen in der Community -> KfB-Themenworkshop?
- aus dem JACOW-Prozess lernen? -> internationale Bemühungen

## Gemeinsames FDM

- Allgemein anerkannte Workflows?
  - Es gibt Normen, z.B. für strahlenschutzrelevante Daten.
  - Zunehmende Relevanz durch Verwendung von standardisierten Sensornetzwerken, "mutual interest data" in Simulationsdatensätzen
  - internationale Zusammenarbeit ?!

## Datenreduktion

- (Noch) keine gemeinsame implementierte / akzeptierte Workflows zur Datenreduktion
  - jedoch: Fortschritte in Simulation & Sensorik -> lernen aus Best-Practice-Beispielen

# Community-übergreifende Vernetzung

## Committee for

Forschung mit Synchrotronstrahlung  
 Rat Deutscher Sternwarten  
 Hadronen- und Kernphysik  
 Elementarteilchenphysik  
 Forschung mit Neutronen  
 Astroteilchenphysik  
 Beschleunigerphysik  
 Forschung mit nuklearen Sonden und Ionenstrahlen

## Total number of scientists with doctoral degree

## Scientists with doctoral degree

KFS	2,300
RDS	1,500
KHuK	1,500
KET	1,300
KFN	1,000
KAT	500
KfB	200
KFSI	100
	<b>8,400</b>

## Partnership Innovative Digitization in ErUM

### Federated Infrastructures

Storage  
 Fast Network  
 Compute Power

### Enabling Technologies

Big Data Analytic Tools  
 Infrastructure Services  
 Research Data Management

### Sharing Knowledge

Workshops  
 Schools  
 Competence Multipliers

**Common Governance Structure  
 with Connections to Other Communities and International Partners**