



Data Generation & Simulations WG

# The Data Generation & Sims WG

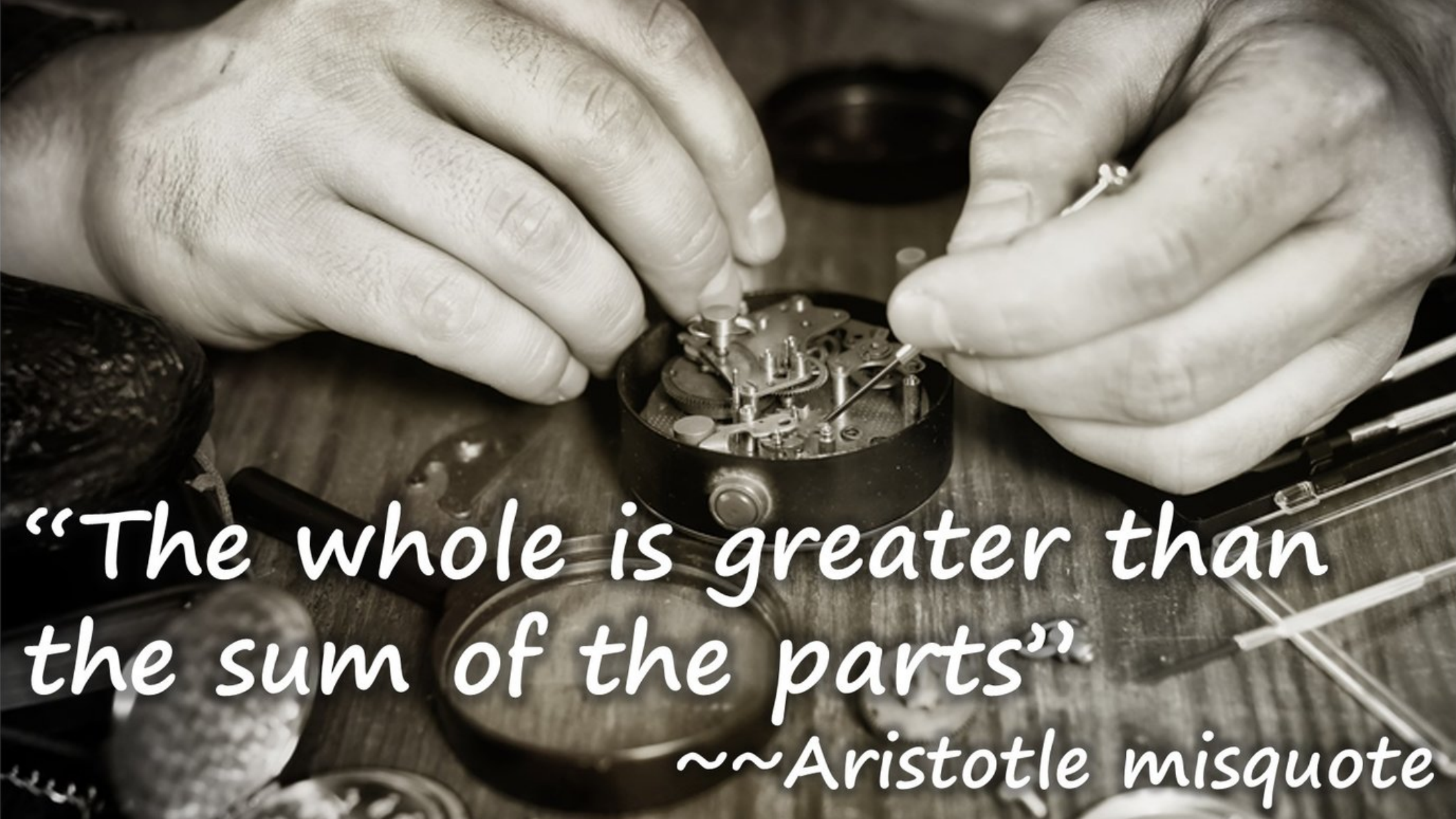
<input type="checkbox"/> A.K.Mistry@GSI.DE Andrew K. Mistry	<input type="checkbox"/> francesco.massimo@UNIVERSITE-PARIS-SACLAY.FR Francesco Massimo
<input type="checkbox"/> a.oeftiger@GSI.DE Adrian Oeftiger	<input type="checkbox"/> francis.osswald@IPHC.CNRS.FR Francis Osswald
<input type="checkbox"/> adnan.ghribi@GANIL.FR Adnan Ghribi	<input type="checkbox"/> gianluca.valentino@UM.EDU.MT Gianluca Valentino
<input type="checkbox"/> amelia.pollard@STFC.AC.UK Amelia Pollard	<input type="checkbox"/> hayg.guler@IJCLAB.IN2P3.FR Hayg Guler
<input type="checkbox"/> andrea.santamaria@KIT.EDU Andrea Santamaria	<input type="checkbox"/> kachelho@CC.IN2P3.FR Thomas Kachelhoffer
<input type="checkbox"/> barbara.dalena@CEA.FR Barbara Dalena	<input type="checkbox"/> kubytsky@IJCLAB.IN2P3.FR Viacheslav Kubytskyi
<input type="checkbox"/> brigitte.cros@UNIVERSITE-PARIS-SACLAY.FR Brigitte Cros	<input type="checkbox"/> lefmann@NBI.KU.DK Kim Lefmann
<input type="checkbox"/> cassou@IJCLAB.IN2P3.FR Kevin Cassou	<input type="checkbox"/> mauro.migliorati@UNIROMA1.IT Mauro Migliorati
<input type="checkbox"/> chenran.xu@KIT.EDU Chenran Xu	<input type="checkbox"/> samuel.marini@CEA.FR Samuel Marini
<input type="checkbox"/> Christine.darve@ESS.EU Christine Darve	<input type="checkbox"/> simon.hirlaender@PLUS.AC.AT Simon Hirländer
<input type="checkbox"/> damien.minenna@CEA.FR Damien Minenna	<input type="checkbox"/> simone.liuzzo@ESRF.FR Simone Liuzzo
<input type="checkbox"/> davide.di.croce@CERN.CH Davide di Croce	<input type="checkbox"/> tatiana.pieloni@EPFL.CH Tatiana Pieloni
<input type="checkbox"/> Elena.Donegani@ESS.EU Elena Donegani	<input type="checkbox"/> Verena.Kain@CERN.CH Verena Kain

## Topics:

- Accelerator physics
- Accelerator technology
- Linacs / Rings
- Hadrons / Light sources
- Plasma WF accelerators
- Neutron scattering

# Goals for today

- **Exchange** on individual goals & experience
- Develop the **concept**:  
**Identify** joint objectives & deliverables
- **Structure** our WG: outline WPs & Tasks
- ✓ Gain enough momentum & ideas for next phase  
(formalising consortium & preparing proposal!)



*“The whole is greater than  
the sum of the parts”*  
~~Aristotle misquote

# White Paper to Shape Ideas

- data: bottom-up strategy for WG
- 12 study cases (+1)
- Compare data management solutions, develop strategy
- Investigate methods for active learning

WHITE PAPER

---

A Route toward Sustainable  
Data Generation in Accelerator Science

---

**Contributors:**

Adnan GHRIBI, CEA/IRFU/CNRS/IN2P3/GANIL, France

Adrian OEFTIGER, GSI, Germany

Gianluca VALENTINO, University of Malta, Malta

Francis OSSWALD, IPHC, CNRS, France

Andrea Santamaria Garcia, KIT, Germany

Chenran Xu, KIT, Germany

Kevin Cassou, IJClab, CNRS/IN2P3, France

# Pool of Study Cases

Exploring Resonance  
Diagrams  
(Adrian Oeftiger et al)

Enhanced Emittance  
Evaluation  
(Marcel Heine,  
Francis Osswald)

Surrogate modeling of  
beam losses in the LHC  
collimation hierarchy  
(Gianluca Valentino)

Surrogate Modelling of a  
Low-energy Linac  
(Chenran Xu, Andrea  
Santamaria Garcia)

DA and collective effects  
(Adnan Ghribi, Barbara  
Dalena, Quentin Bruant)

Design of a multi-stage laser  
wakefield accelerator at  
LPGP (Francesco Massimo)

DA and loss rate predictions and  
FCC optimisation  
(Davide Di Croce, Tatiana Pieloni,  
Ekaterina Krymova, et al)

Machine learning based control  
for temporal and spatial laser  
shaping at CLARA  
(Amelia Pollard)

PIC simulations dataset for laser-  
plasma injector optimisation  
(Kevin Cassou)

Surrogate modeling at  
IJCLAB (Hayg Guler)

Automated data collection for  
laser-plasma injector beam  
quality optimisation  
(Kevin Cassou)

EARLI, Design of a laser  
wakefield accelerator for AWAKE  
(Damien Minenna)

# Key aspects from first discussions

- Develop metadata standard
- Open science (FAIR principles) → publish data to enable new cross-machine/cross-institutional models/training
- Data management aspects: large data sets?
- Active learning to improve on grid parameter scans
- Enable new studies/technology (large-scale models, federated training) vs sustainability
- Investigate active learning methods jointly on study cases
- Economic cluster usage (more energy-efficient computation)
- Hackathon-style events to jointly push progress on study cases
- ...

# Speed Dating

- Split up into study case presenters & visitors
- 5min round of discussion: use slide
- Cycle through
- Switch presenters & visitors after full cycle



# Name / Study Case:

Goals of Study Case

(what method do you have in mind?)

Ideas for Collaboration

(what would ARTIFACT enable you to do that you can only do in the collaboration?)

Connection to Industrial Partners

Experience on Topic

(what skills do you want to offer to the collaboration?)

Experience in Open Science

(Open-access Publications?  
Open-source Software / Analysis?  
Open Data?)

Barriers to Study Case

(from “unclear how exactly to profit from ML”  
→ to “stuck on implementational details”?)

# Ideas for Objectives

- Data management: exploit EOSC (European Open Science Cloud) and national infrastructure
  - ESCAPE (OSSR, DIOS), PUNCH4NFDI
- Develop tool to integrate parameter space sampling (+launch on cluster?) and sharing data
  - Active learning algorithms
  - Data-lake solution(s) for sharing
  - Publication of data via DOI
  - Try tool on study cases, develop along

# Work Packages & Tasks

- Keywords from call should appear in WP structure, stay generic, e.g.:
  - Tools, methods and simulation
  - Data and computing
- WPs then comprise Tasks, Deliverables and Milestones