WP3/WP7 discussion on Big Data

Eva Sicking (CERN)

Cremlin Workshop on Big data Management – Moscow
February 16, 2017

WP3/WP7 discussion on Big Data: Participants

- Session: "Big Data in particle physics and in research with ion sources"

- WP7 participants:
  - Massimo Lamanna (CERN)
  - Lucie Linssen (CERN)
  - Alexei Anisenkov (BINP)
  - Yuriy Tikhonov (BINP)
  - Eva Sicking (CERN)
- WP3 participants:
  - Jürgen Eschke (FAIR / GSI)
  - Thorsten Kollegger (FAIR / GSI)
  - Vladimir Korenkov (JINR Dubna)
- And many more...

09:00 → 09:10 **The requirements for computing for experiments at SCT** 10m

Speaker: Iouri Tikhonov (Budker Institute of Nuclear Physics (RU))

09:10 → 09:20 **Software & Computing for Linear Collider Experiments** 10m

Speaker: Eva Sicking (CERN)

CLIC.pdf

09:20 → 09:30 **WLCG: Worldwide LHC Computing Grid** 10m

Speaker: Vladimir Korenkov (Joint Institute for Nuclear Research (RU))

09:30 → 09:31 **FAIR computing (already presented in Session I)** 1m

Speakers: Jurgen Eschke (GSI - Helmholtzzentrum fur Schwerionenforschung GmbH (DE)), Thorsten Kollegger (GSI - Helmholtzzentrum fur Schwerionenforschung GmbH (DE))

20170215-Cremlin.pdf

09:31 → 09:32 **Computing for experiments at the LHC (already presented in Session I)** 1m

Speaker: Massimo Lamanna (CERN)

09:32 → 09:33 **CRIC: the evolution of the ATLAS Grid Information System for other collaborations (already presented in Session I)** 1m

Speaker: Alexey Anisenkov (Budker Institute of Nuclear Physics (RU))

CRIC.CREMLIN.Feb....

09:33 → 10:30 **Discussion** 57m

Which big data challenges exist?

How can we share existing information and solutions?

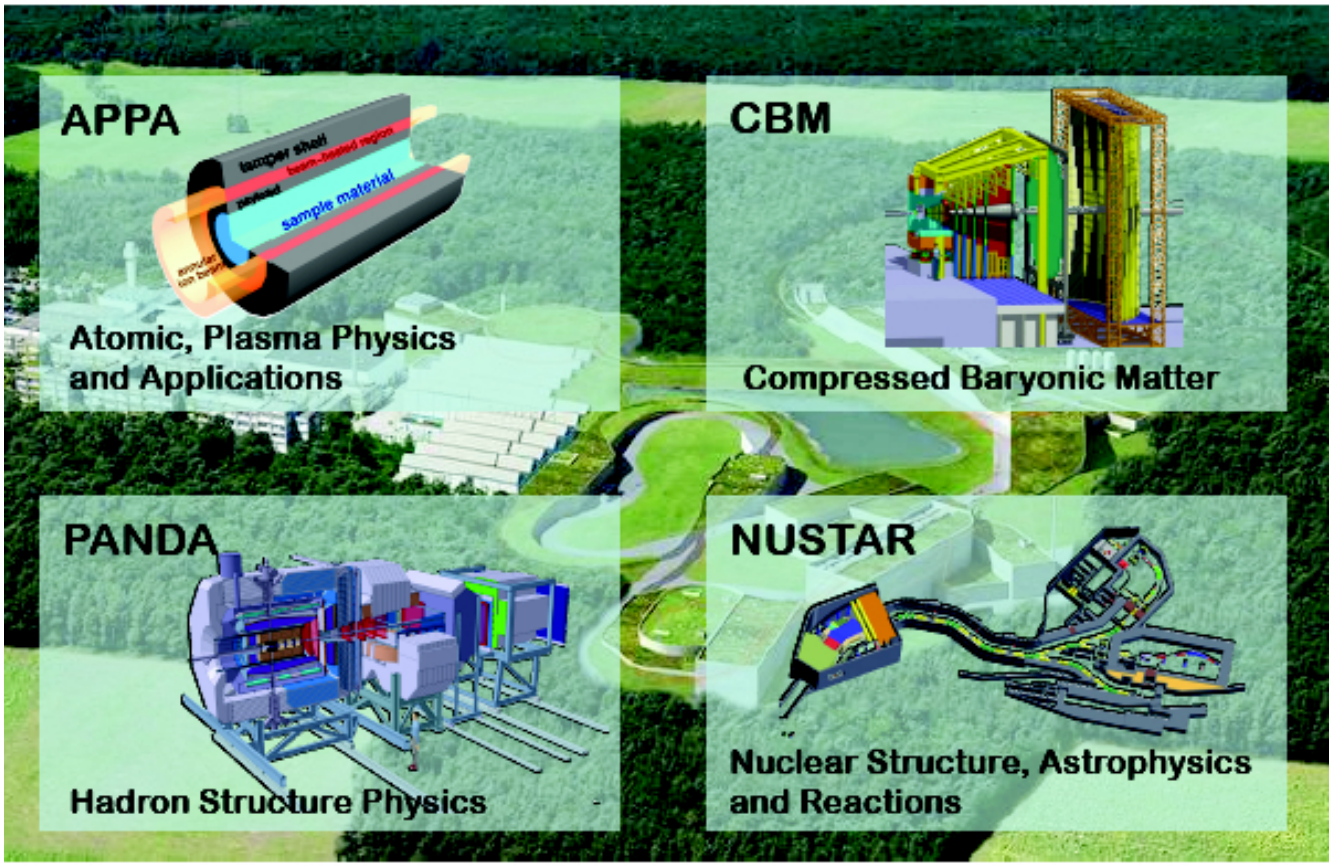
How can this link to Cremlin deliverables? (e.g. task 7.3 "Cremlin data management platform for lepton colliders" <=> intended as a platform for sharing information, software applications, etc).

Discussion topics

- What are our big data challenges?
- How can we share existing information and solutions?
- How can this link to Cremlin deliverables?
  - Example: Task 7.3 "Cremlin data management platform for lepton colliders"

Big Data challenges at FAIR

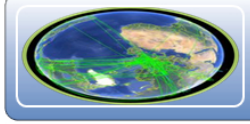
- Facility for Antiproton and Ion Research: FAIR
- Currently under construction
- Start in 2025



- 1 TByte/s into online farms
- 35 PByte/year disk
- 30 PByte/year tape
- 300.000 cores (majority on-site in common compute center)

- Green Cube (data center for FAIR and GSI)
- Common software framework for all FAIR experiments and beyond

Big Data challenges at JINR: WLCG & NICA



JINR grid sites of WLCG/EGI: Tier-1 for CMS
Tier-2 for ALICE, ATLAS, CMS, STAR, LHCb,
BES, biomed, fermilab

Cloud infrastructure

Heterogeneous(CPU + GPU) computing cluster HybriLIT

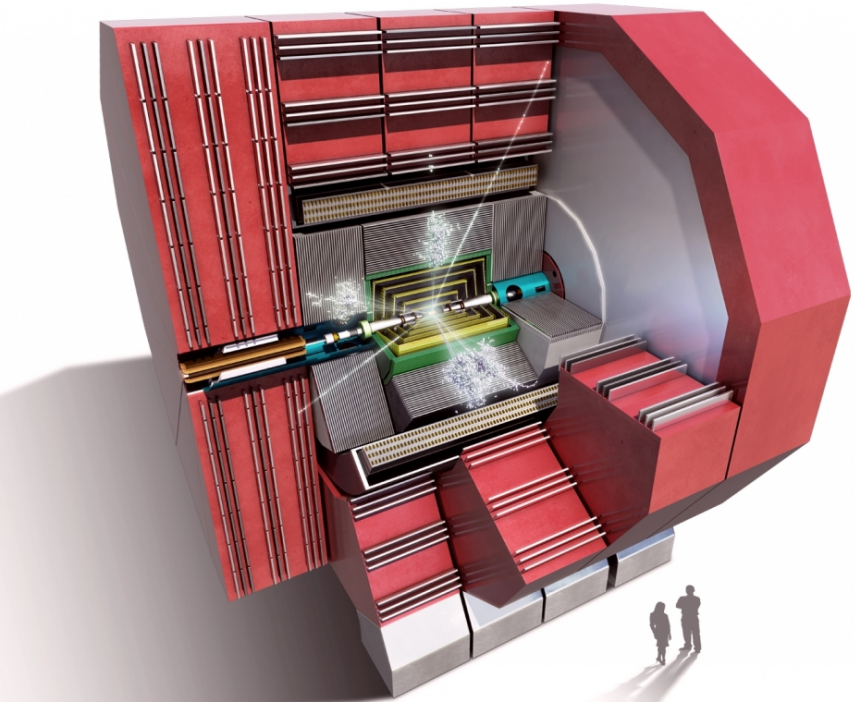
Off-line cluster and storage system for BM@N, MPD,
SPD Storage and computing facilities for local users

Network infrastructure

Engineering infrastructure

- Large existing infrastructure
- Already numerous existing collaboration with many experiments
- Tier 1 for CMS
- Tier 2 for all LHC experiements
- Strong cooperation between FAIR and NICA

Big Data challenges at CLIC

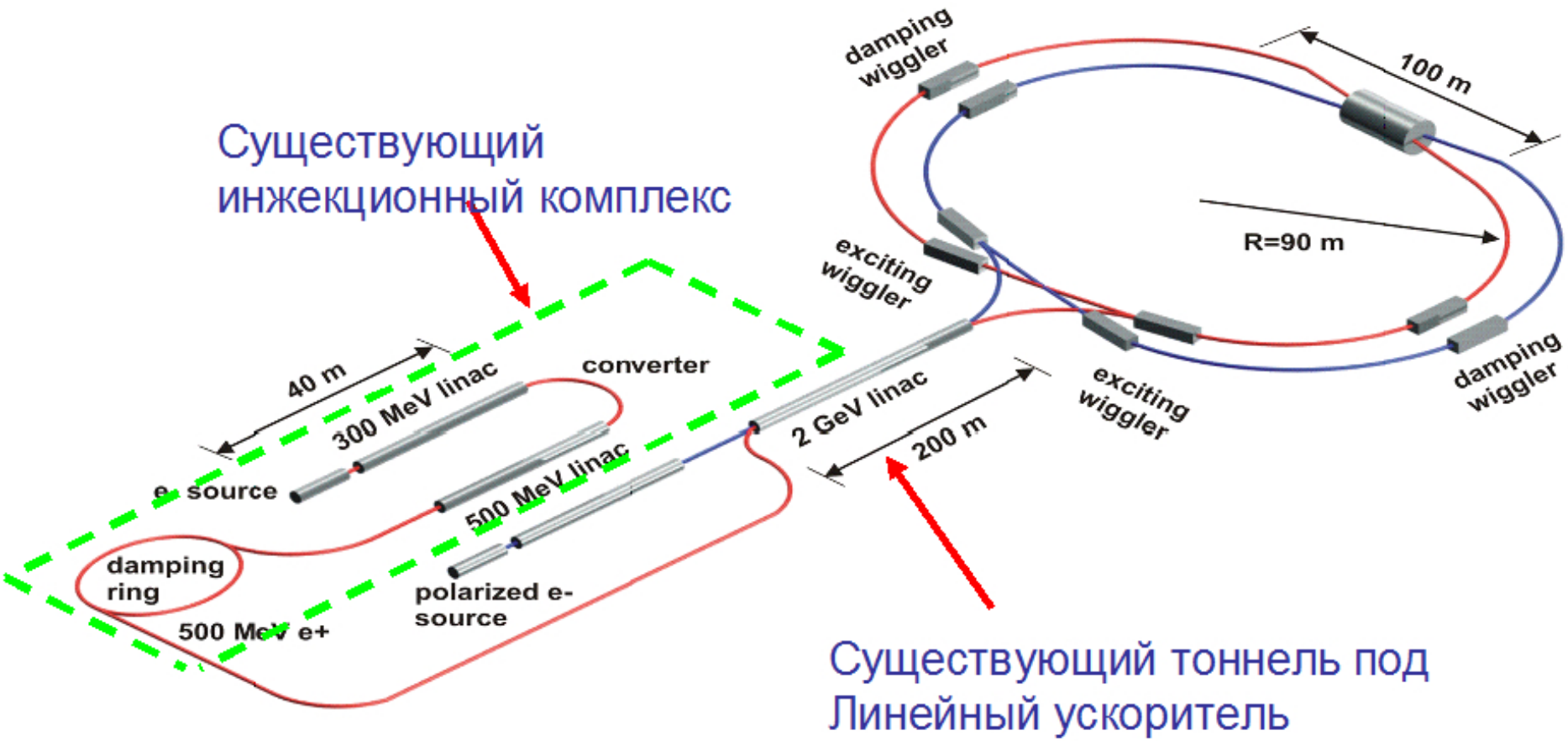


- Current simulation studies:
  - Grid solution iLCDirac, using WLCG, OSG
  - 5.5 PB files
  - 41 grid sites
- Data during CLIC operation:
  - Bunch trains, trigger-less readout
  - 10 GByte/s

- CDR in 2012, now prepare for next European Strategy meeting
- Possible start date: 2035
- Simulation studies to identify physics potential, detector R&D
- Example: Higgs physics paper in 2016

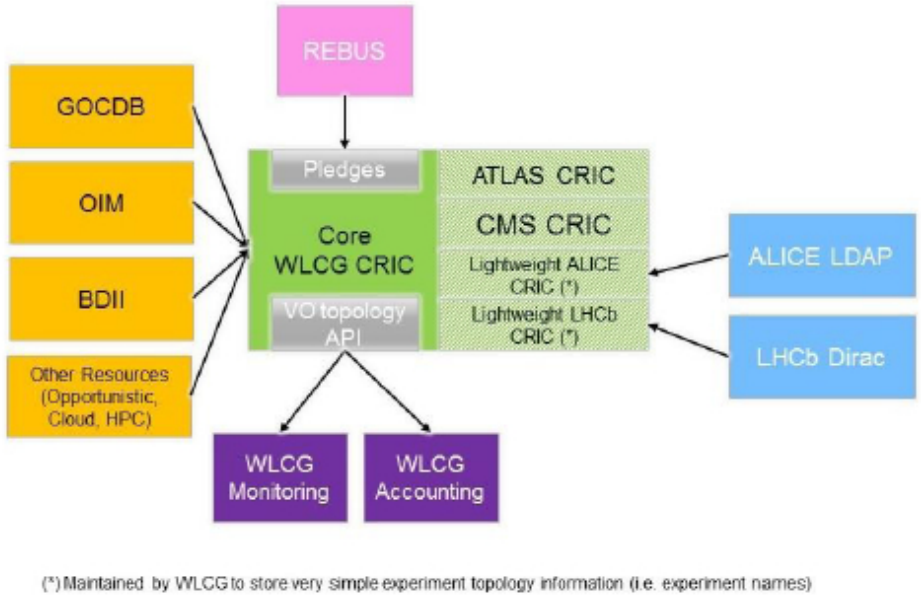
- Small linear collider community: Cooperation with other projects
- Common software solutions shared between CLIC, ILC, CALICE, FCC, neutrino experiments, LHC
- Use existing infrastructure: Grid, EOS

Big Data challenges at SCT



- 8 GByte/s data rate
- ∼50 PB/year

CRIC: BINP& CERN collaboration



- Computing Resource Information Catalogue (CRIC) → replaces AGIS
  - Single entry point for WLCG topology and service configuration
  - Currently used by ATLAS
  - Proposed to be used for other LHC experiments, too
  - Combine CRIC and Dirac (LHCb)
  - Option: Integrated CRIC into CLIC/ILC Grid submission system (iLCDirac)

Challenges → Expand Cooperation

BINP SCT:

- Increase data link to Novosibirsk
  0.5 GBit/s link currently → 10 GBit/s link currently
- Increase size of local data center by factor 100

Link to WP7 Cremlin deliverables

- Ongoing WP7 actions:
  - Setup of web-based platform for sharing exsiting information and software tools between lepton collider projects
- Potential option for further collaboration:
  - Integrated CRIC system developed at BINP into CLIC grid submission system