Satellite workshop on Karabo control and data analysis at European XFEL

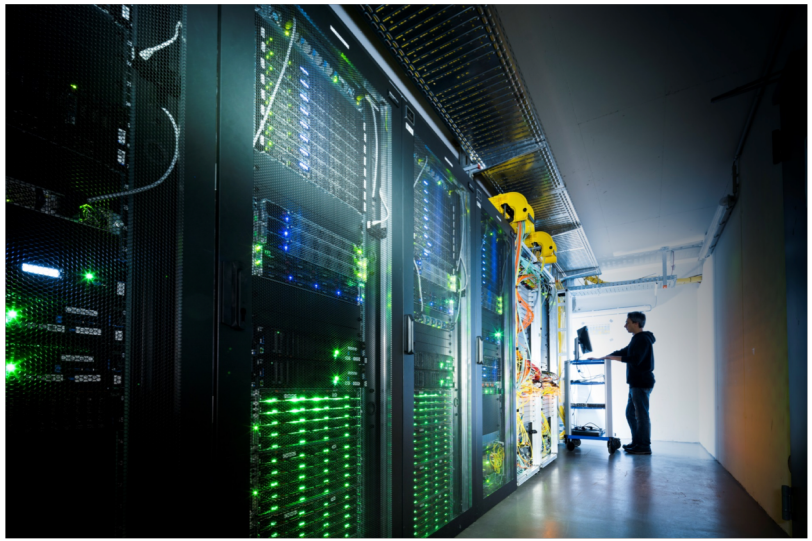
XFEL User Meeting 2019, 24 January 2019

- Agenda
  - 14:00 Welcome (Sandor Brockhauser)
  - 14:00 Overview Karabo Control and Data Analysis (Hans Fangohr)
  - 14:10 Overview Karabo (Gero Flucke)
  - 14:40 Detector Calibration (Steffen Hauf)
  - 15:10 Data Management (Krzysztof Wrona)
  - **15:30 Break (Coffee)**
  - 16:00 Offline Data Analysis at XFEL (Martin Bergemann)
  - 16:25 Online Data Analysis at XFEL (Thomas Michelat)
  - 16:45 Discussion and close

Overview:
Karabo control data analysis at European XFEL

Hans Fangohr
Control and Analysis Software Group
Senior Data Analysis Scientist

University of Southampton
United Kingdom



DESY, Auditorium, 24 January 2019

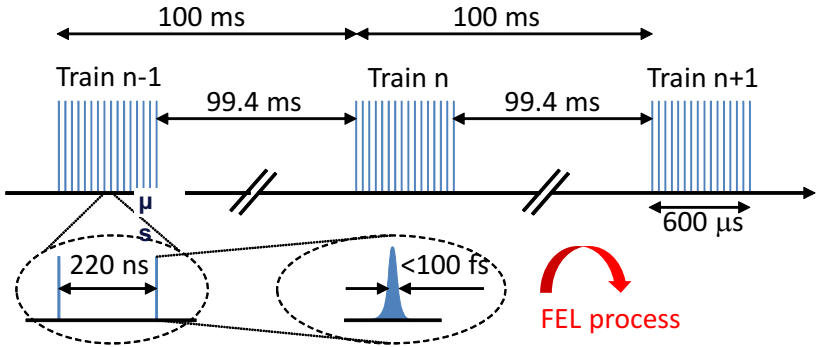
Outline

- Trains and Pulses
- Karabo, devices and pipelines
- Online cluster, Maxwell cluster and data migration
- Near-real time data analysis
- File based data analysis
- Jupyter Notebook
- Summary

European XFEL: Photons come in trains containing short pulses

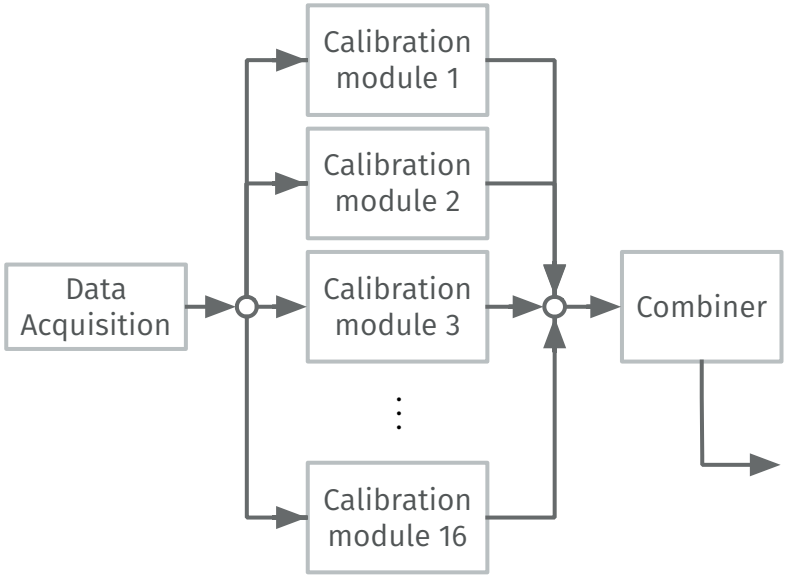
Pattern:

- **Trains** with 10 Hz
  - Trains are numbered (“train ID”)
  - Relevant time unit for data storage
  - and streaming of data
- Up to 2700 **pulses** per train



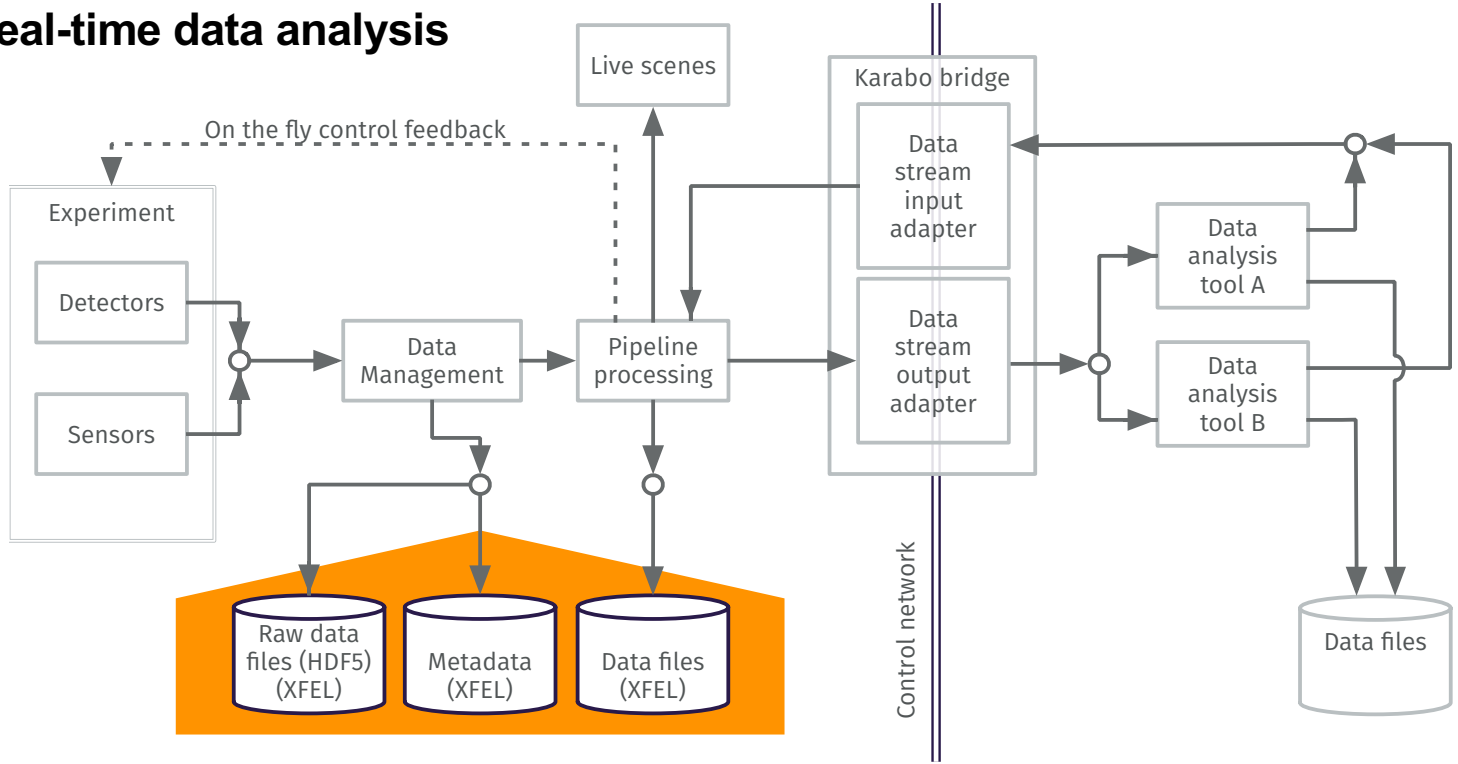
Karabo distributed control system

- Karabo is framework for control and data
  - Processing units called “devices”
  - Data tokens pass through pipeline
  - Devices can be distributed over hardware
  - Simplified example in figure: calibration for detector modules carried out in parallel

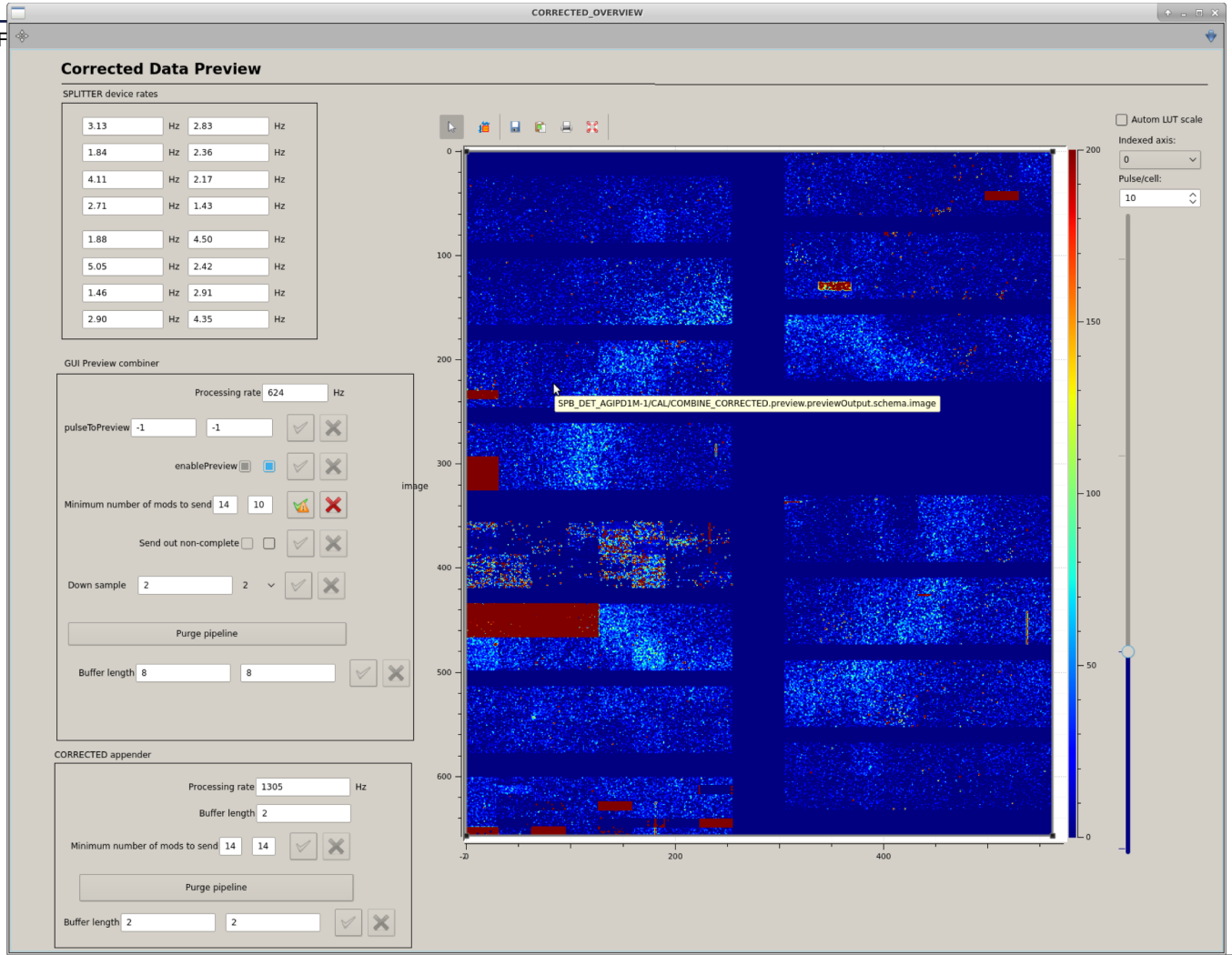


- More details:
  - 14:10 Gero Flucke: “Karabo overview”
  - 14:40 Steffen Hauf: “Calibration”

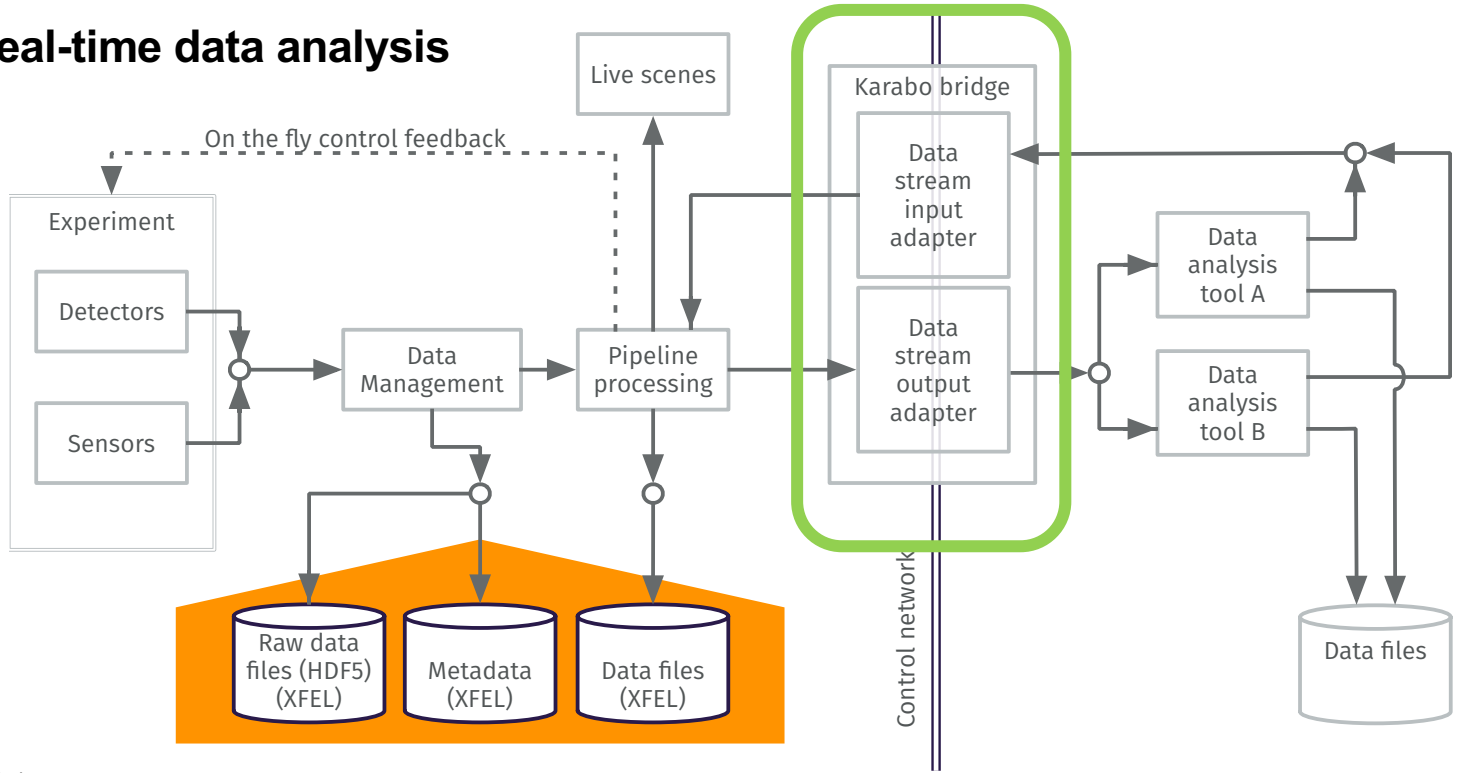
Near real-time data analysis



Online data analysis: Rapid feedback through GUI



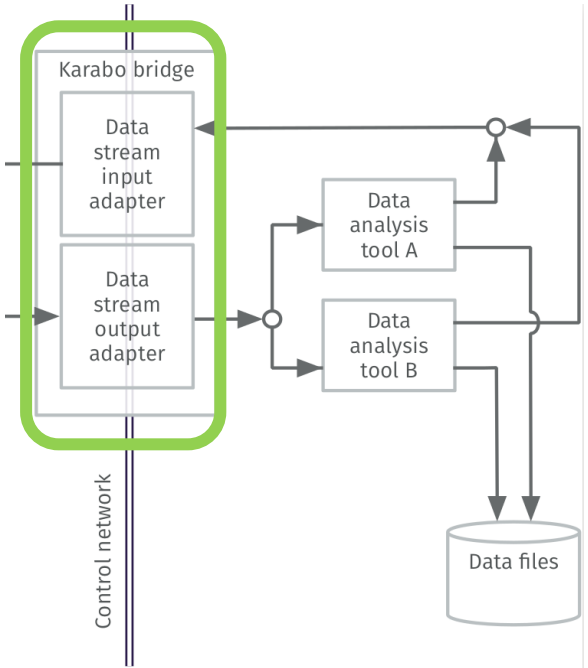
Near real-time data analysis



Online data analysis

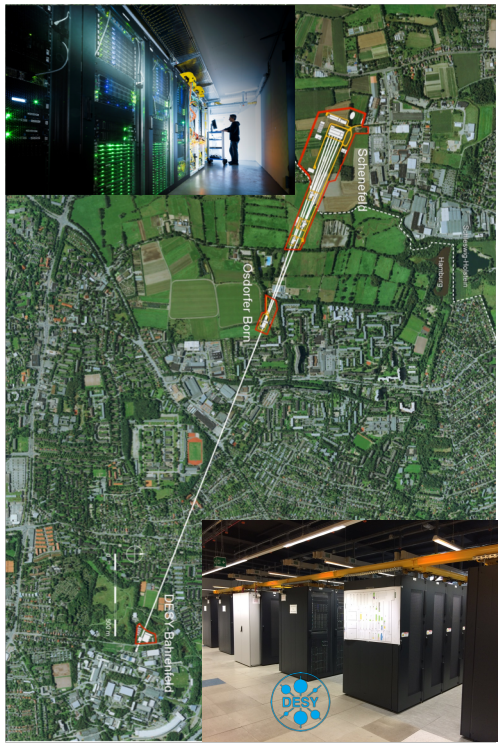
Karabo Bridge – export data stream

- interface to listen to Karabo pipelines
  - Integrate existing (complex) user tools
  - Quick (dirty) specific scripts to use during an experiment
- Development in collaboration with CFEL Chapman Group (S. Aplin, A. Barty, M. Kuhn, V. Mariani)
- 16:25 Online Data analysis (Thomas Michelat)

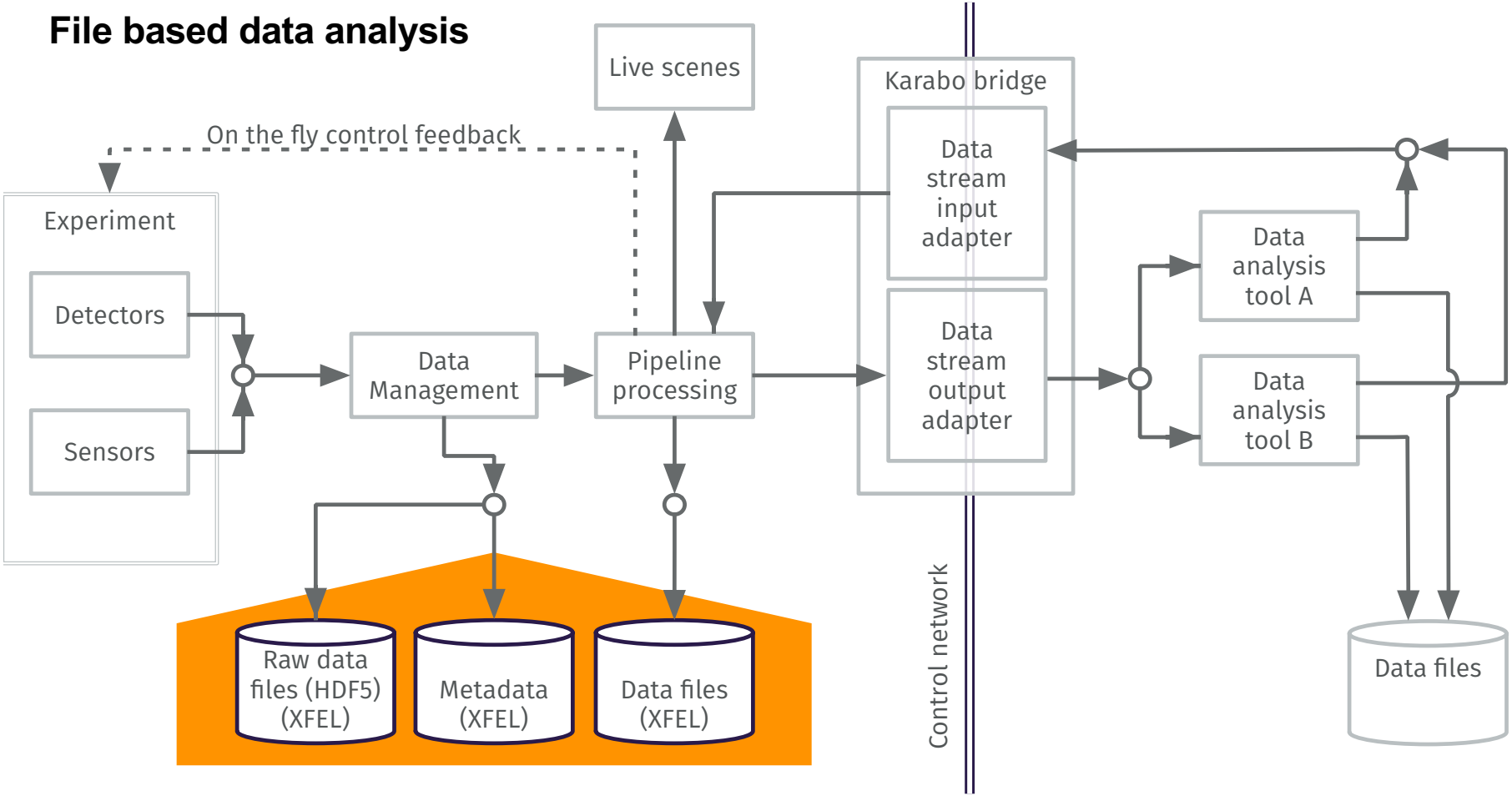


Data management: hutch → Maxwell

- During measurement (run)
  - Streaming data available in hutch (GUI, Karabo-bridge)
- Data migration after each run
  - After each run, data manager decides on quality of the data: “good”, “unclear”, “not interesting”
- Analyse files on Maxwell cluster
  - 180 nodes with 40 cores and ~512GB each
- 15:10 Data management (Krzysztof Wrona)



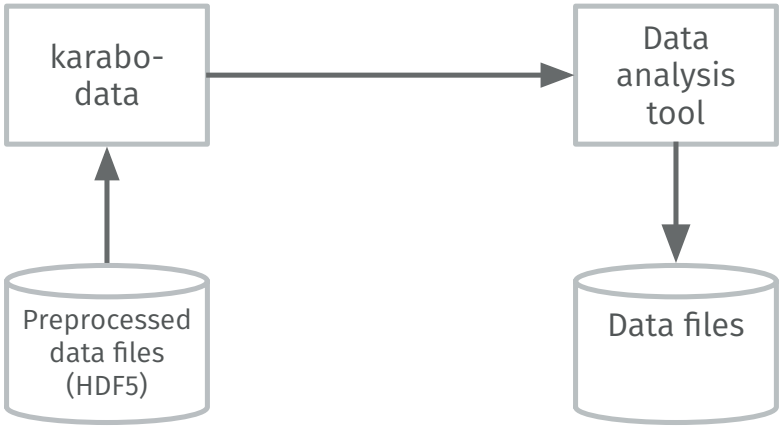
File based data analysis



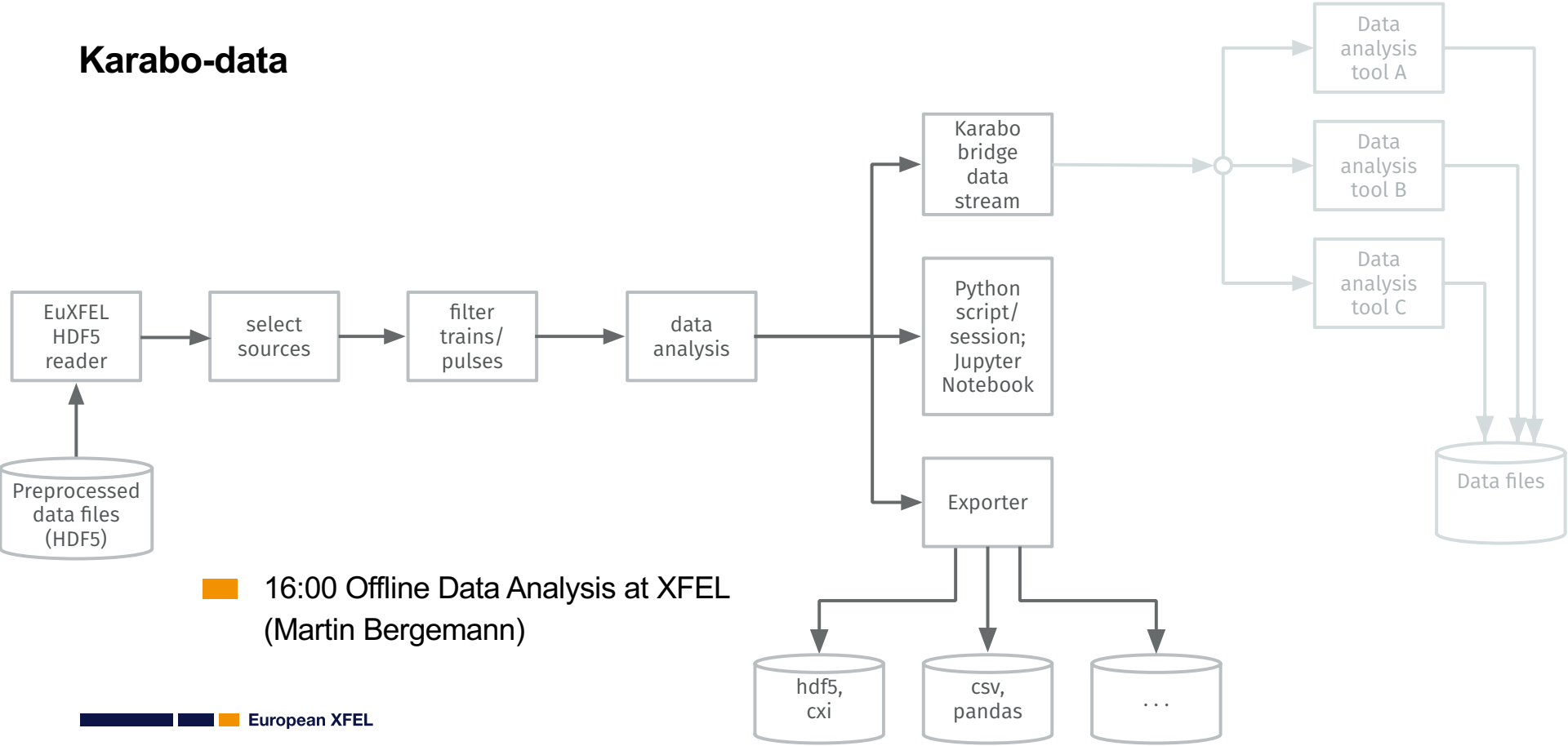
```
[fangohr@max-exfl014]/gpfs/exfel/exp/SPB/201701/p002012/raw/r0359% ls -lh
total 92G
-rw-r--r-- 1 xdata xdata 6.1G Jul  6 11:03 RAW-R0359-AGIPD00-S00000.h5
-rw-r--r-- 1 xdata xdata 6.1G Jul  6 11:03 RAW-R0359-AGIPD01-S00000.h5
-rw-r--r-- 1 xdata xdata 6.1G Jul  6 11:03 RAW-R0359-AGIPD02-S00000.h5
-rw-r--r-- 1 xdata xdata 241K Jul  6 11:03 RAW-R0359-AGIPD03-S00000.h5
-rw-r--r-- 1 xdata xdata 6.1G Jul  6 11:03 RAW-R0359-AGIPD04-S00000.h5
-rw-r--r-- 1 xdata xdata 6.1G Jul  6 11:03 RAW-R0359-AGIPD05-S00000.h5
-rw-r--r-- 1 xdata xdata 6.1G Jul  6 11:03 RAW-R0359-AGIPD06-S00000.h5
-rw-r--r-- 1 xdata xdata 6.1G Jul  6 11:03 RAW-R0359-AGIPD07-S00000.h5
-rw-r--r-- 1 xdata xdata 6.1G Jul  6 11:03 RAW-R0359-AGIPD08-S00000.h5
-rw-r--r-- 1 xdata xdata 6.1G Jul  6 11:03 RAW-R0359-AGIPD09-S00000.h5
-rw-r--r-- 1 xdata xdata 6.1G Jul  6 11:03 RAW-R0359-AGIPD10-S00000.h5
-rw-r--r-- 1 xdata xdata 6.1G Jul  6 11:03 RAW-R0359-AGIPD11-S00000.h5
-rw-r--r-- 1 xdata xdata 6.1G Jul  6 11:03 RAW-R0359-AGIPD12-S00000.h5
-rw-r--r-- 1 xdata xdata 6.1G Jul  6 11:03 RAW-R0359-AGIPD13-S00000.h5
-rw-r--r-- 1 xdata xdata 6.1G Jul  6 11:03 RAW-R0359-AGIPD14-S00000.h5
-rw-r--r-- 1 xdata xdata 6.1G Jul  6 11:03 RAW-R0359-AGIPD15-S00000.h5
-rw-r--r-- 1 xdata xdata 788M Jul  6 11:03 RAW-R0359-DA01-S00000.h5
-rw-r--r-- 1 xdata xdata  38M Jul  6 11:03 RAW-R0359-DA02-S00000.h5
```

File based data analysis

- EuXFEL creates a set of files per run
  - Multiple files per detector
  - Sequence files
- Processing EuXFEL HDF5 files
  - Recommend “karabo-data” Python library and tool (https://github.com/European-XFEL/Karabo_data)
  - Hides the multiple file complexity
  - Easy to install
  - Often combined with Jupyter Notebook



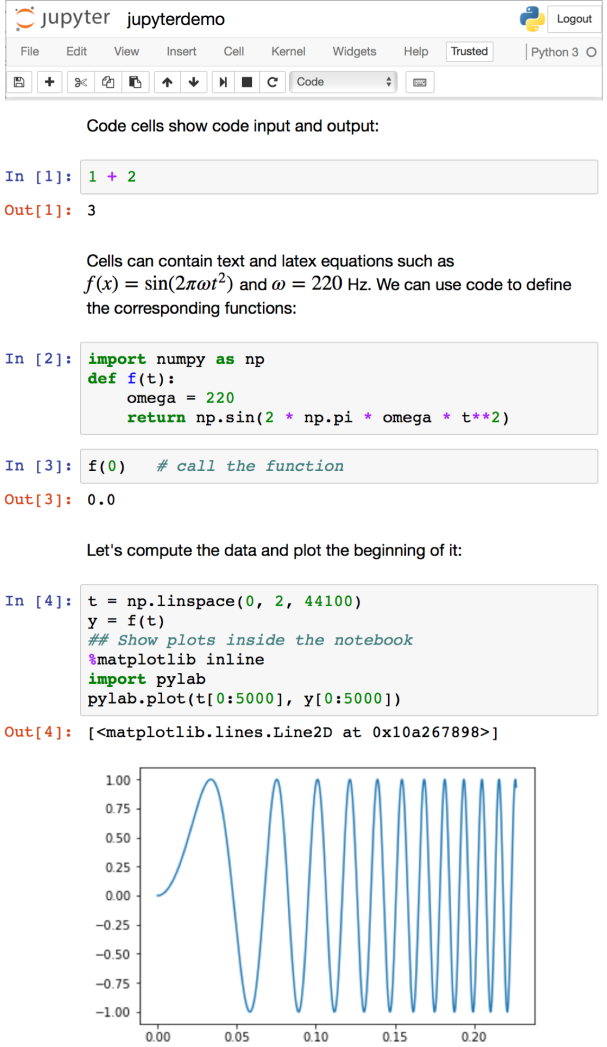
Karabo-data



- 16:00 Offline Data Analysis at XFEL (Martin Bergemann)

Jupyter Notebook

- Jupyter Notebook
  - Executable document
  - Code (typically Python), output, interpretation
  - Remote access through https://max-jhub.desy.de
- XFEL tool Karabo-data integrates in Notebook



Summary

- Introduced concepts and outline of meeting
- Data analysis user support
  - Documentation starting point https://in.xfel.eu/readthedocs/docs/data-analysis-user-documentation/en/latest/
  - Support available (cas-support@xfel.eu)
  - Collaboration with users and other facilities desired
- Slides from this workshop available on http://github.com/european-XFEL/events
- Agenda
  - 14:10 Overview Karabo (Gero Flucke)
  - 14:40 Detector Calibration (Steffen Hauf)
  - 15:10 Data Management (Krzysztof Wrona)
  - **15:30 Break (Coffee)**
  - 16:00 Offline Data Analysis at XFEL (Martin Bergemann)
  - 16:25 Online Data Analysis at XFEL (Thomas Michelat)
  - 16:45 Discussion and close
- Contact
  - hans.fangohr@xfel.eu
  - http://fangohr.github.io
  - @ProfCompMod