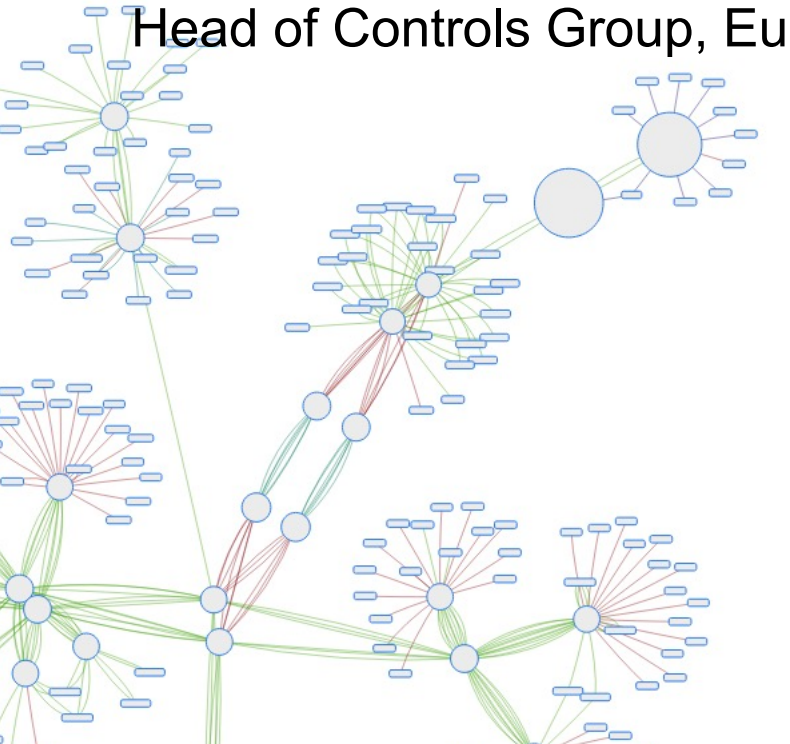
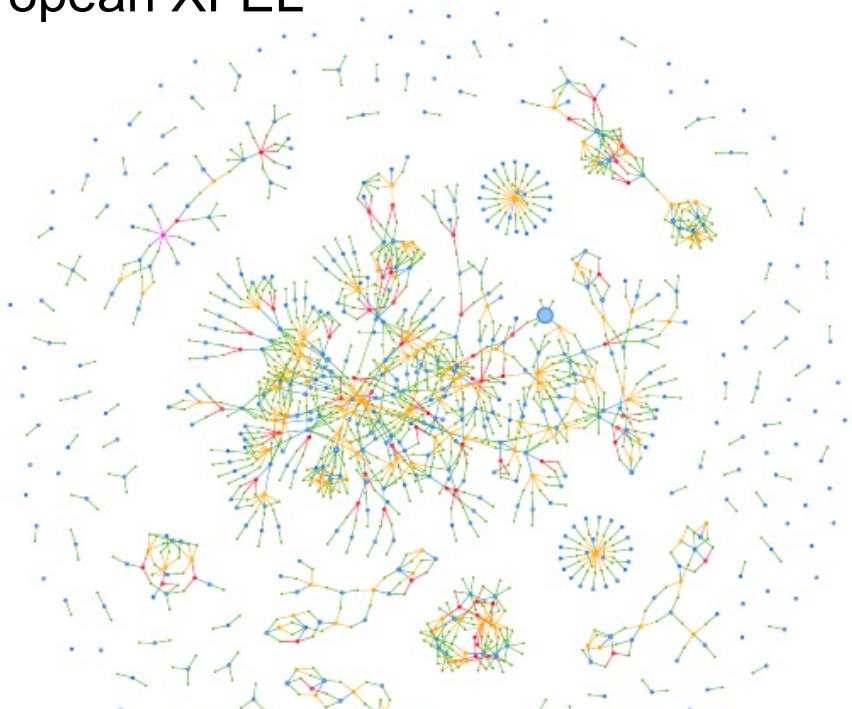
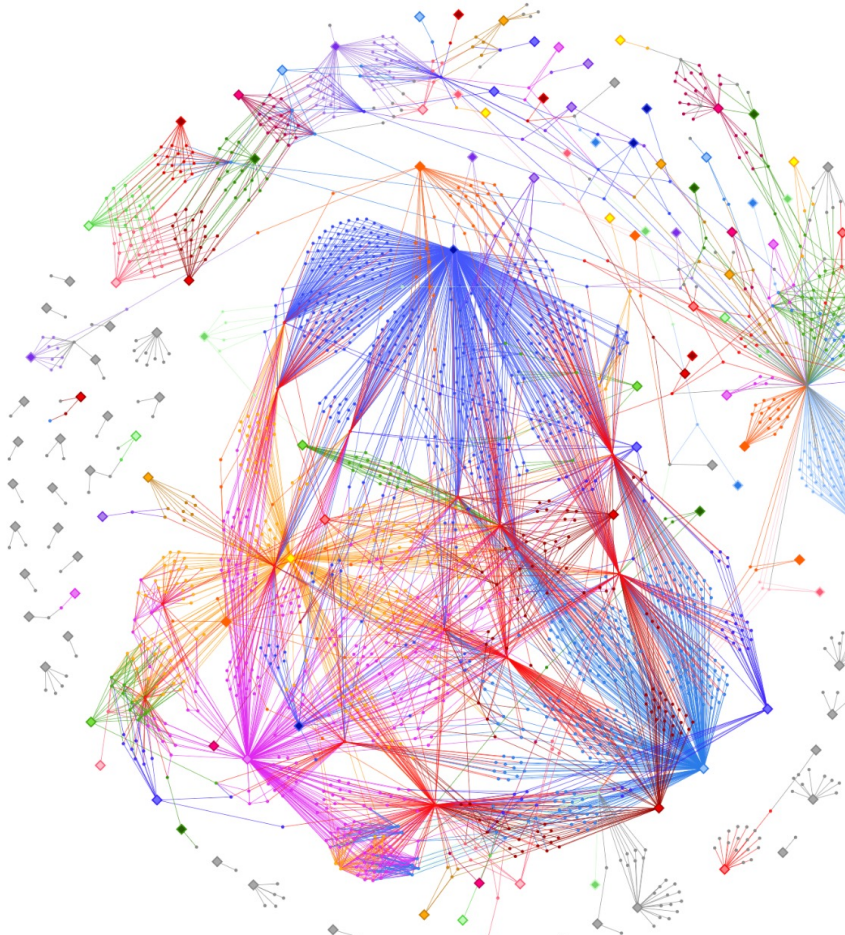
The Data Operation Center: supporting data challenges at the EU.XFEL

Steffen Hauf

Head of Controls Group, European XFEL



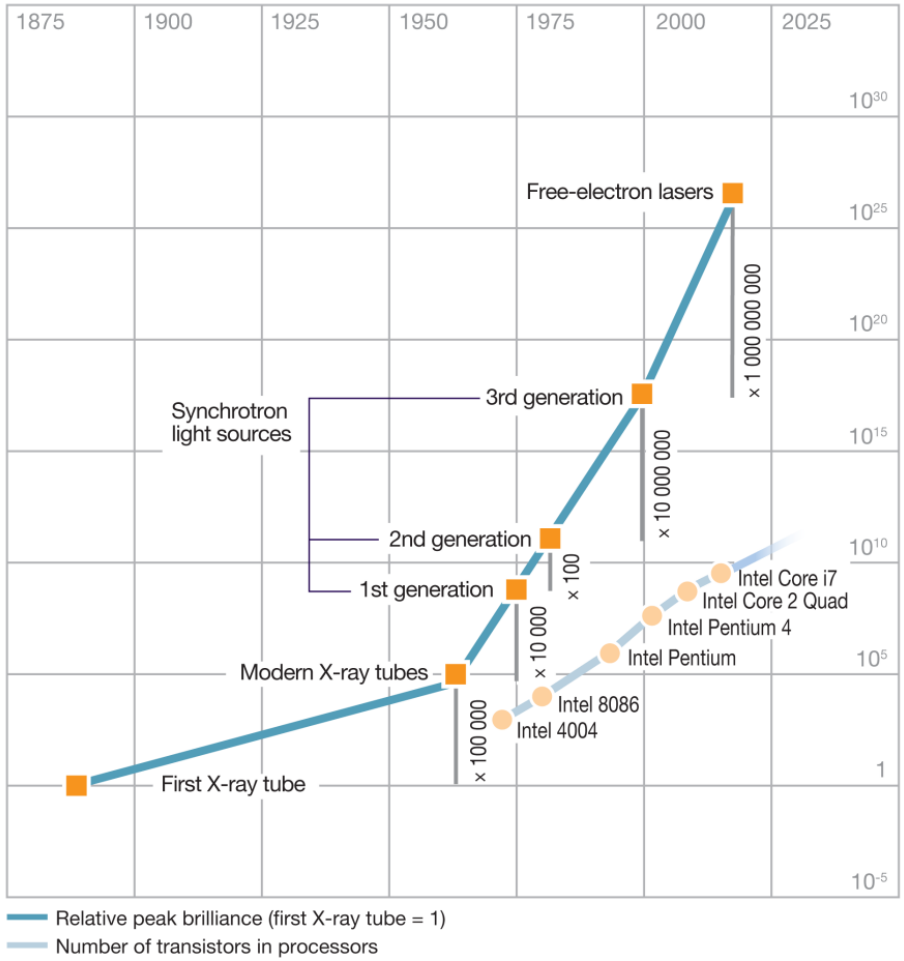




Outline

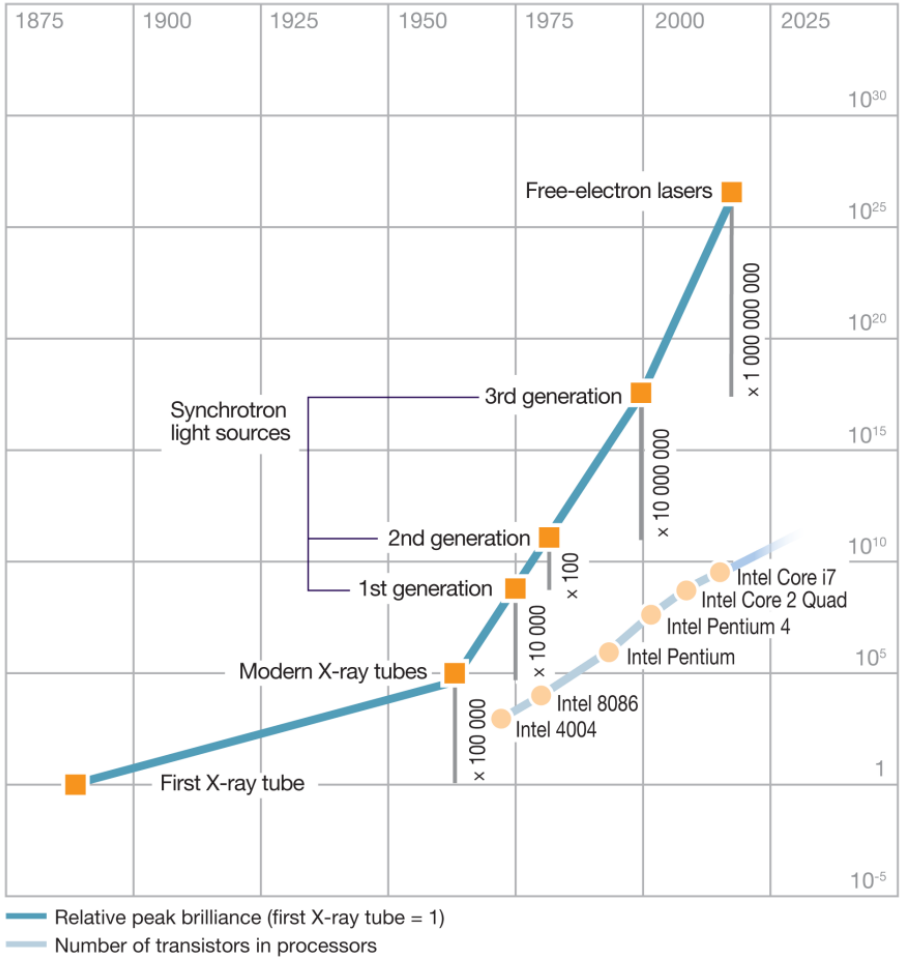
* The European XFEL – a data perspective
* The EU.XFEL Data Operation Center
* Experiences from running the Data Operation Center

The European XFEL – a Data Perspective

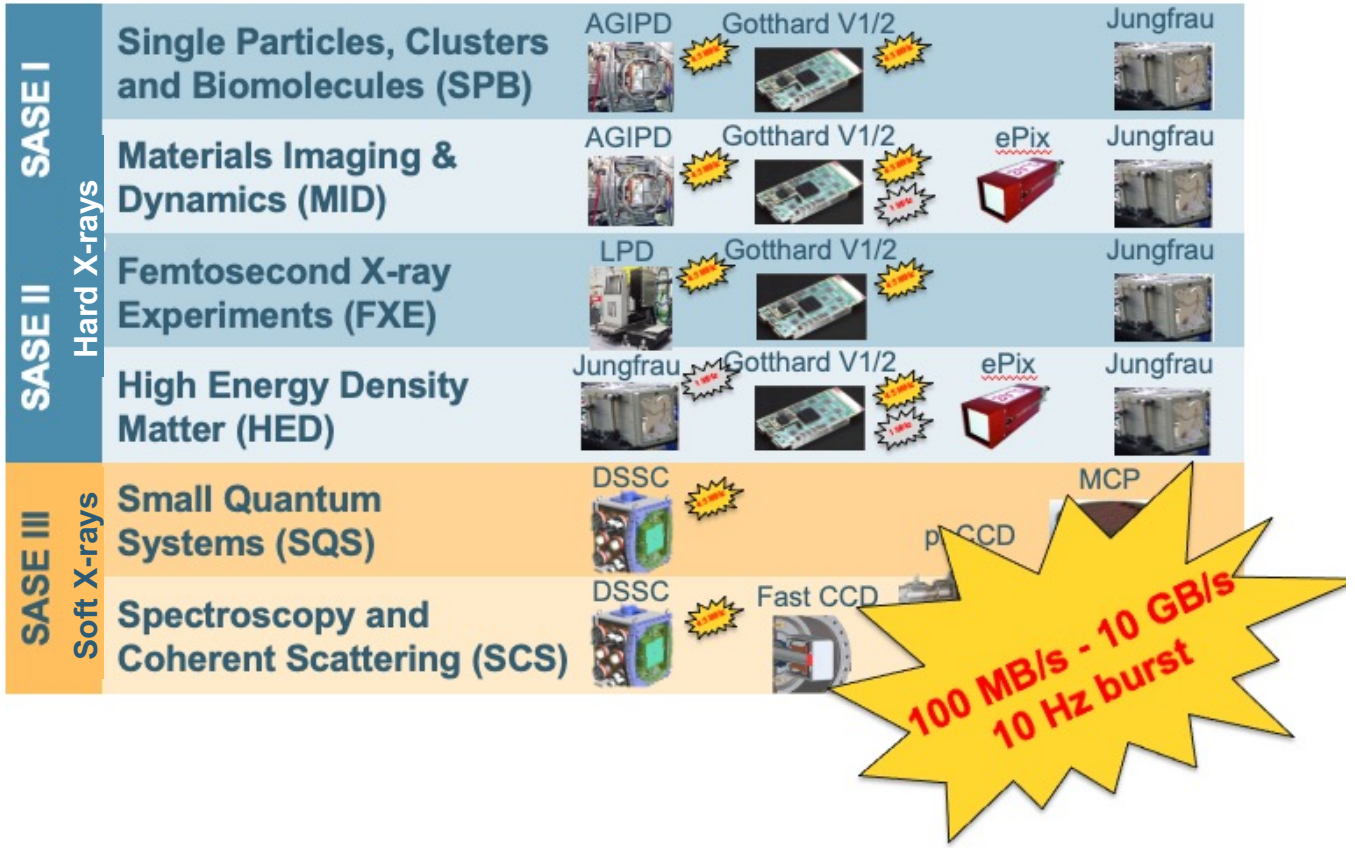


* The development of light source facilities has been faster than the increase in computer processing capacity (i.e., Moore's Law)

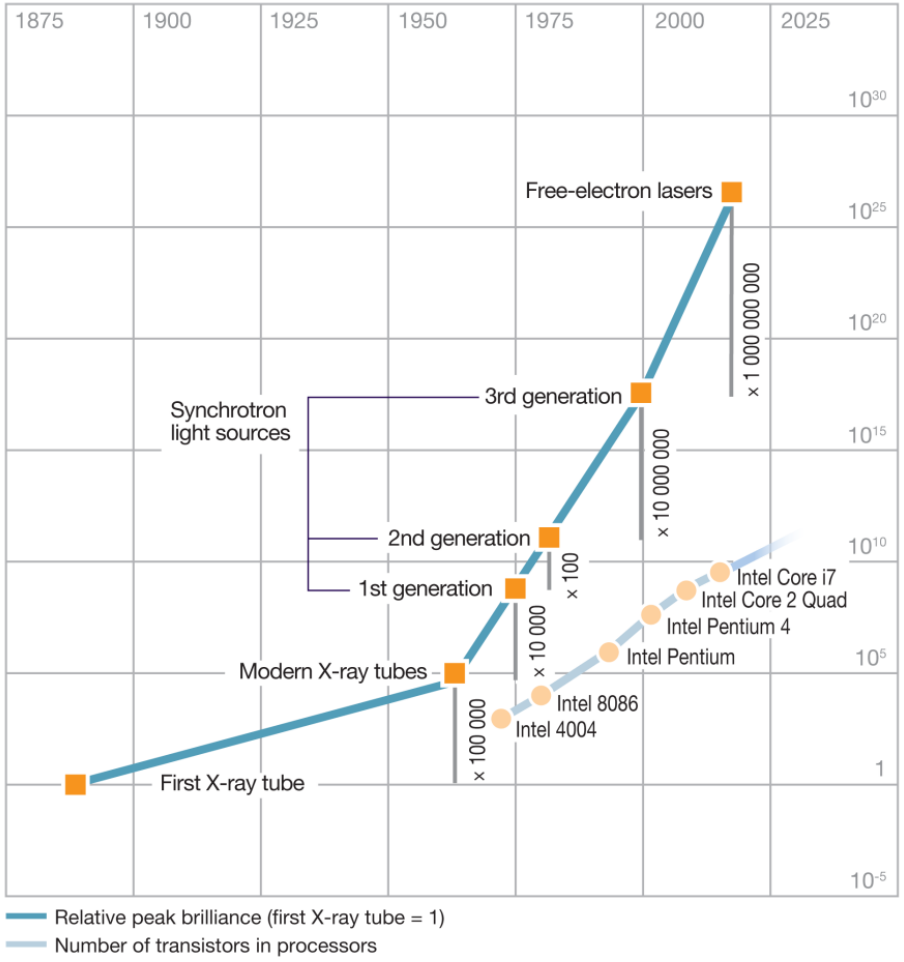
The European XFEL – a Data Perspective



X-Ray Detectors at EU.XFEL Instruments



The European XFEL – a Data Perspective



Custom FPGA-based Data Producers at EU.XFEL

MicroTCA Crates

Large 12 slot 9U and small 6 slot 2U (including MCH, Power Supply and CPU)

X2Timer

XFEL Timing System module for synchronization (clocks and triggers) and pulse parameters from NAT

DAMC2

Required for Clock & Control system for fast 2D detectors, VETO System, Machine Protection System and photon beam loss monitors from DESY

SIS8300

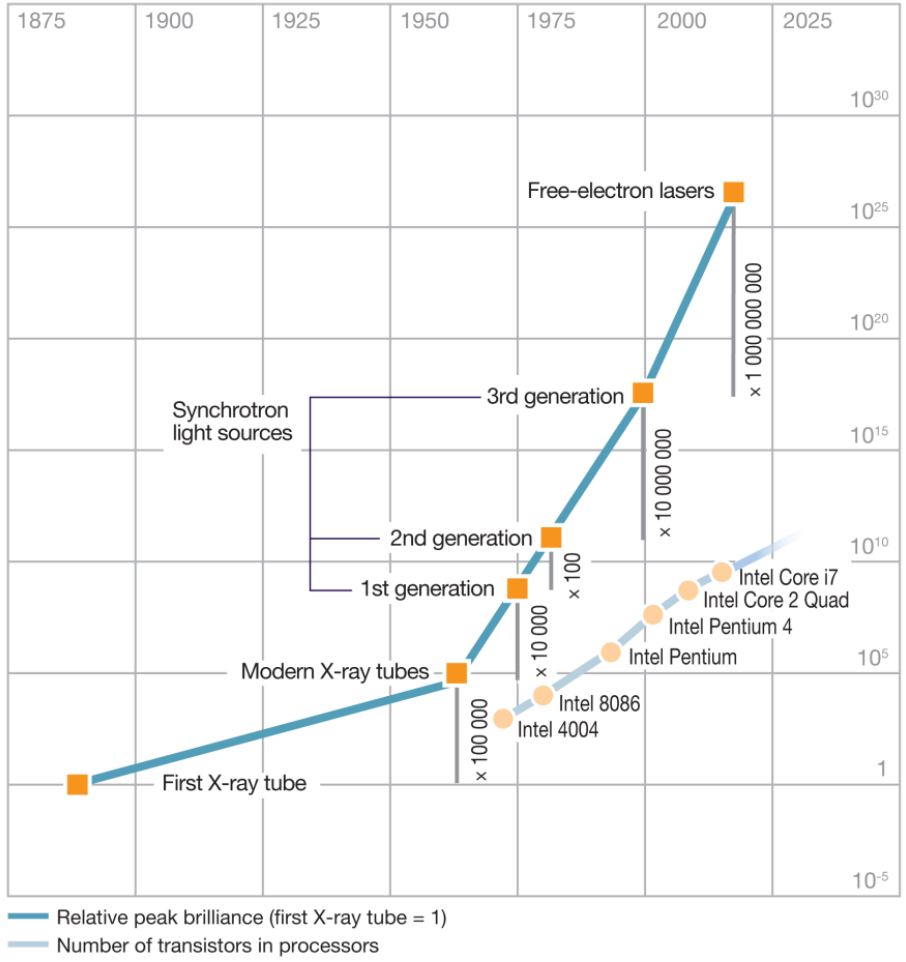
Fast 125MSPS ADC with 10 channels and 16bit resolution for diagnostics and detectors from Struck Innovative Syst e

ADQ412/ADQ14/ADQ7

High-speed digitizers from 1.8GSPS to 10GSPS with 12 to 14 bit resolution from T yne SP Devices

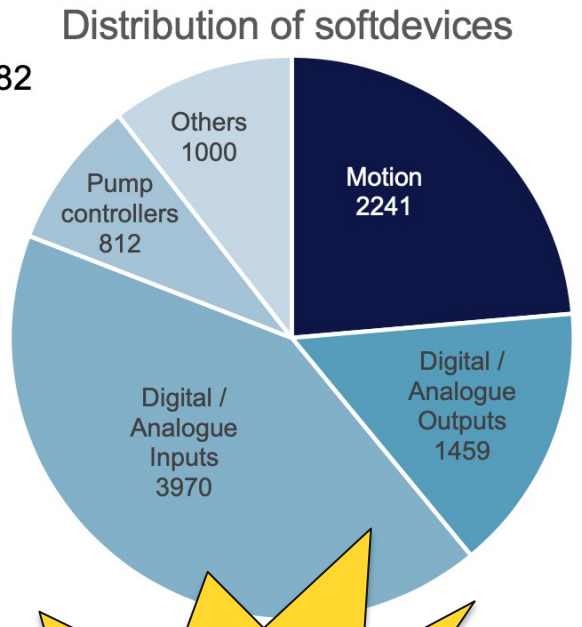
1 MB/s - 1 GB/s
10 Hz burst

The European XFEL – a Data Perspective



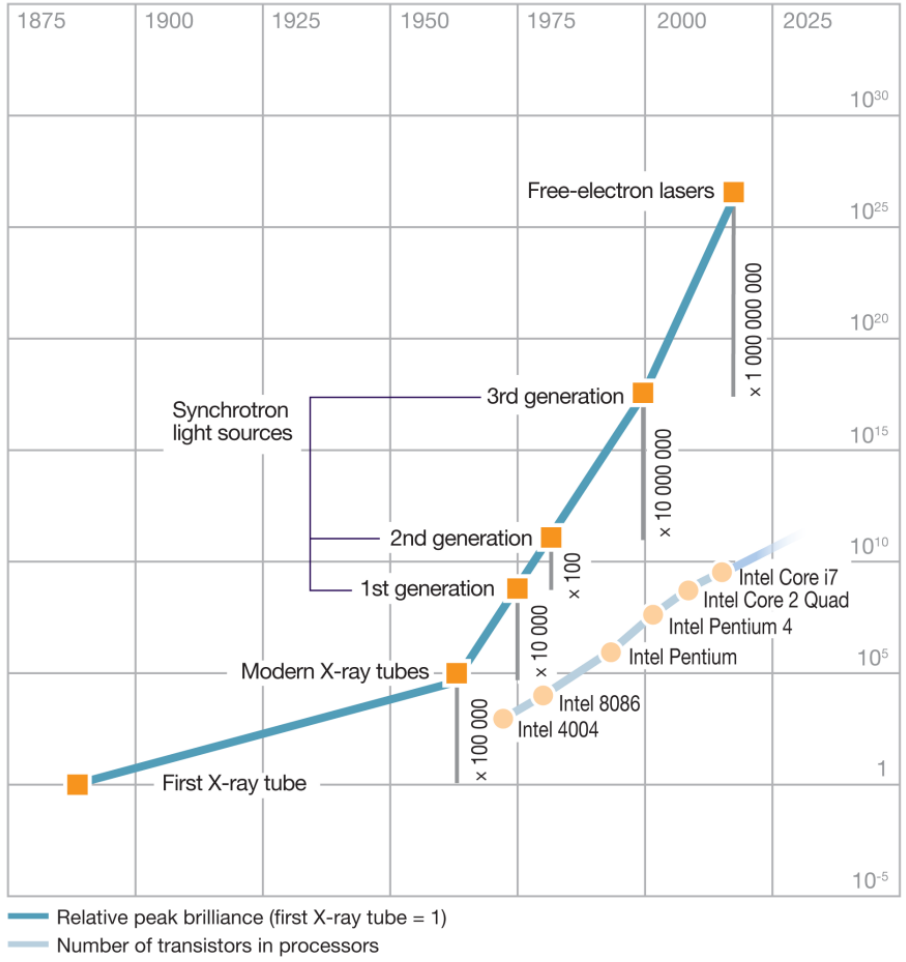
PLC Systems at EU.XFEL

- Softdevices/Equipment: 9.482
- Number of Terminals: 8.452
- PLC Loops / CPUs: 120
- PLC Modules / crates: >500
- MicroTCA Systems: 35
- Digitizer channels: >280

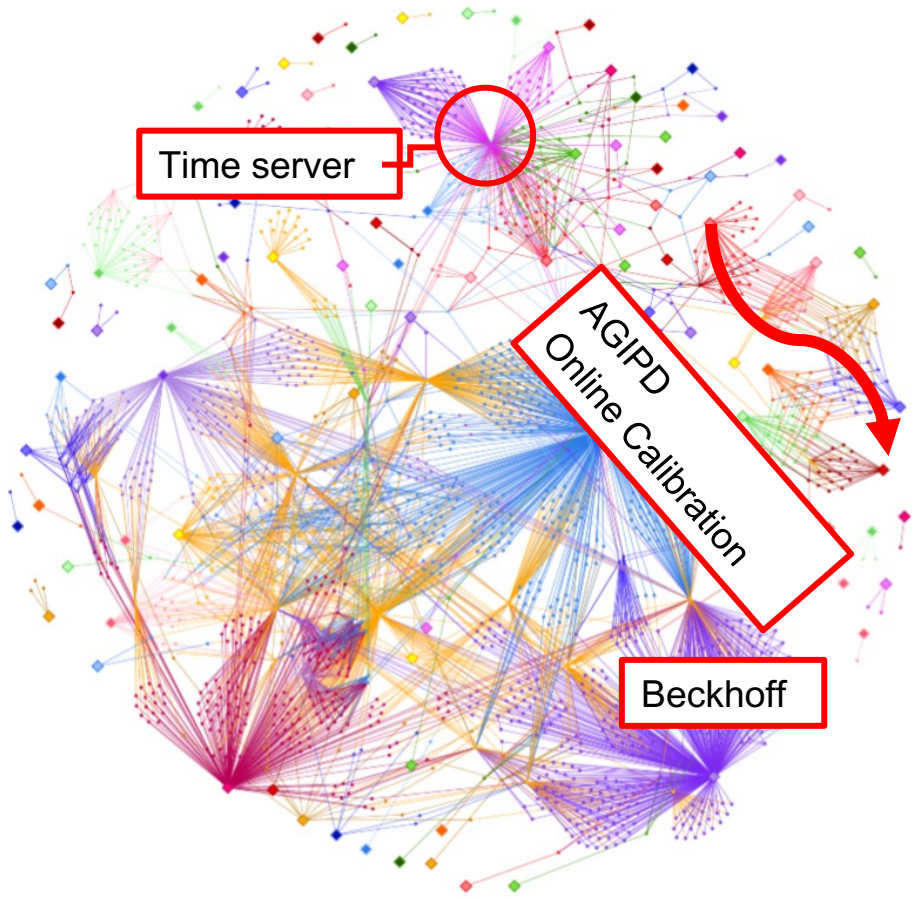


1 B/s - 1 MB/s event driven

The European XFEL – a Data Perspective



The Karabo Control System for the MID instrument

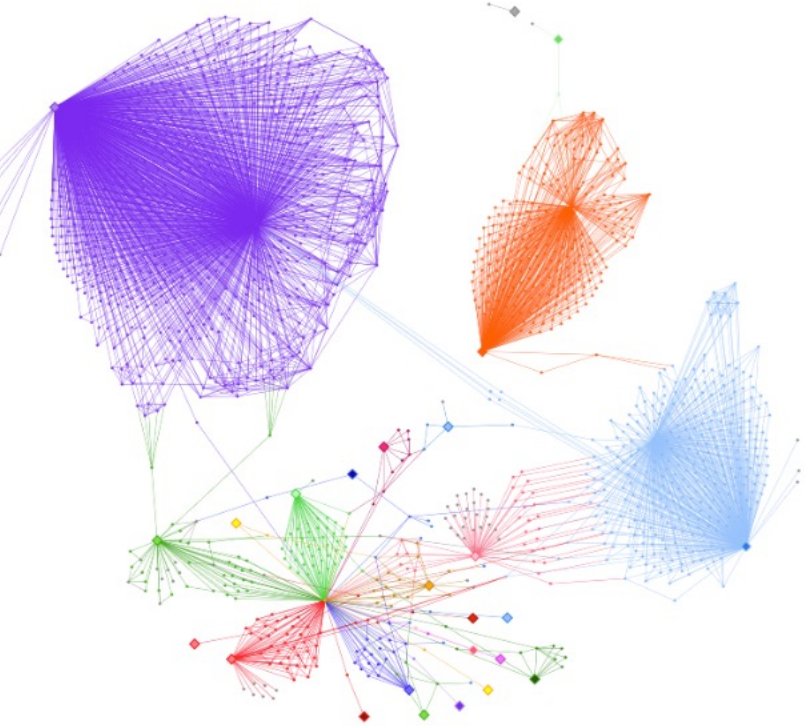


topic	# properties	# devices
LA1	26914.0	257.0
LA2	28103.0	250.0
LA3	32013.0	298.0
SA1	192656.0	1588.0
SA2	274426.0	2260.0
SA3	233216.0	1886.0
FXE	121905.0	905.0
HED	268688.0	1892.0
MID	258094.0	1789.0
SCS	205557.0	1444.0
SPB	267424.0	1872.0
SQS	306817.0	1954.0
Total	2215813.0	16395.0

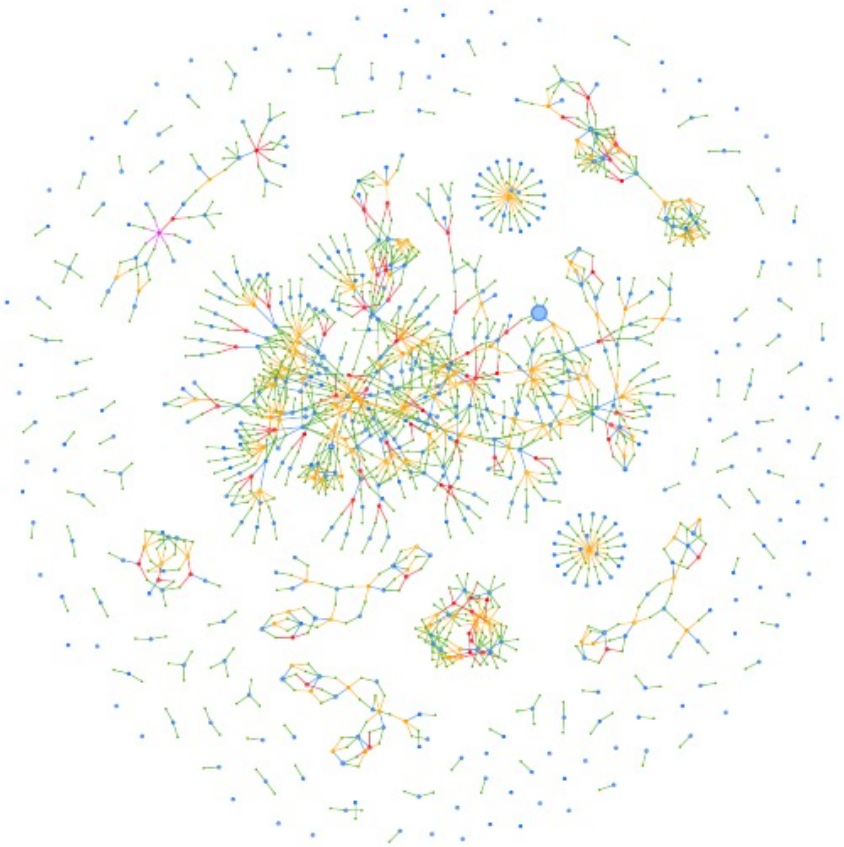
Number of control properties at EU.XFEL

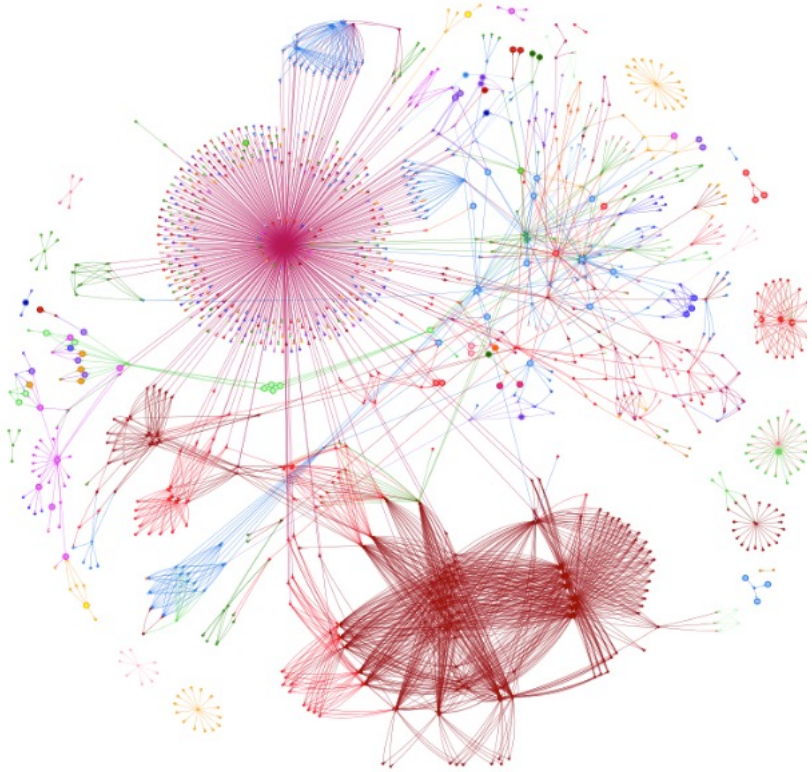
The European XFEL – a Data Perspective

Karabo Host topology of SASE1



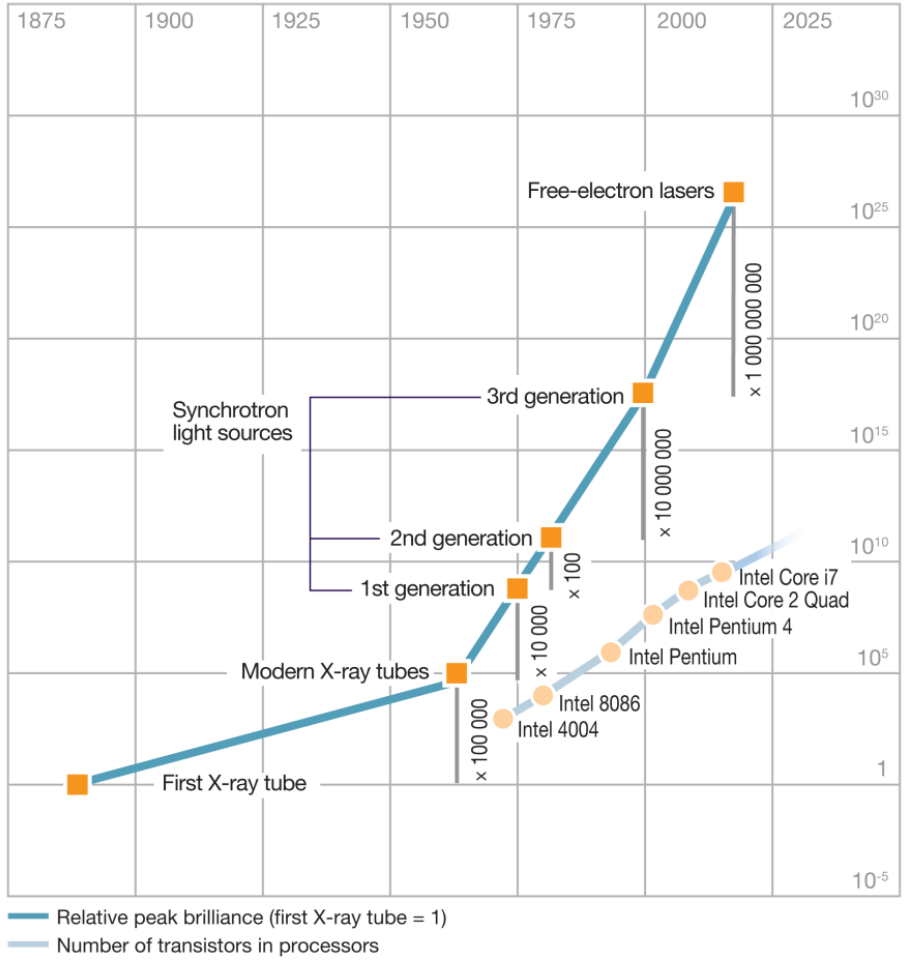
Interlock definitions and references in SASE1 PLCs



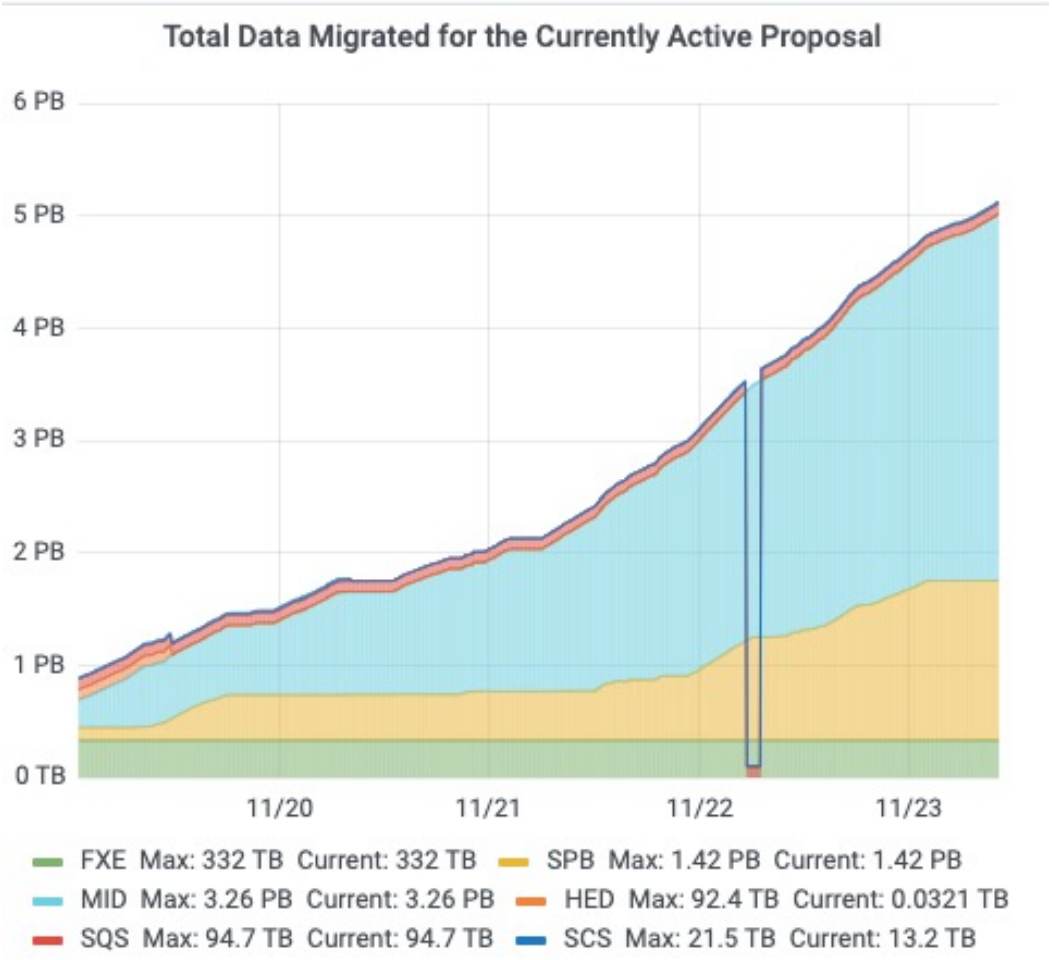


Links between synomptic panels of SASE1 in Karabo

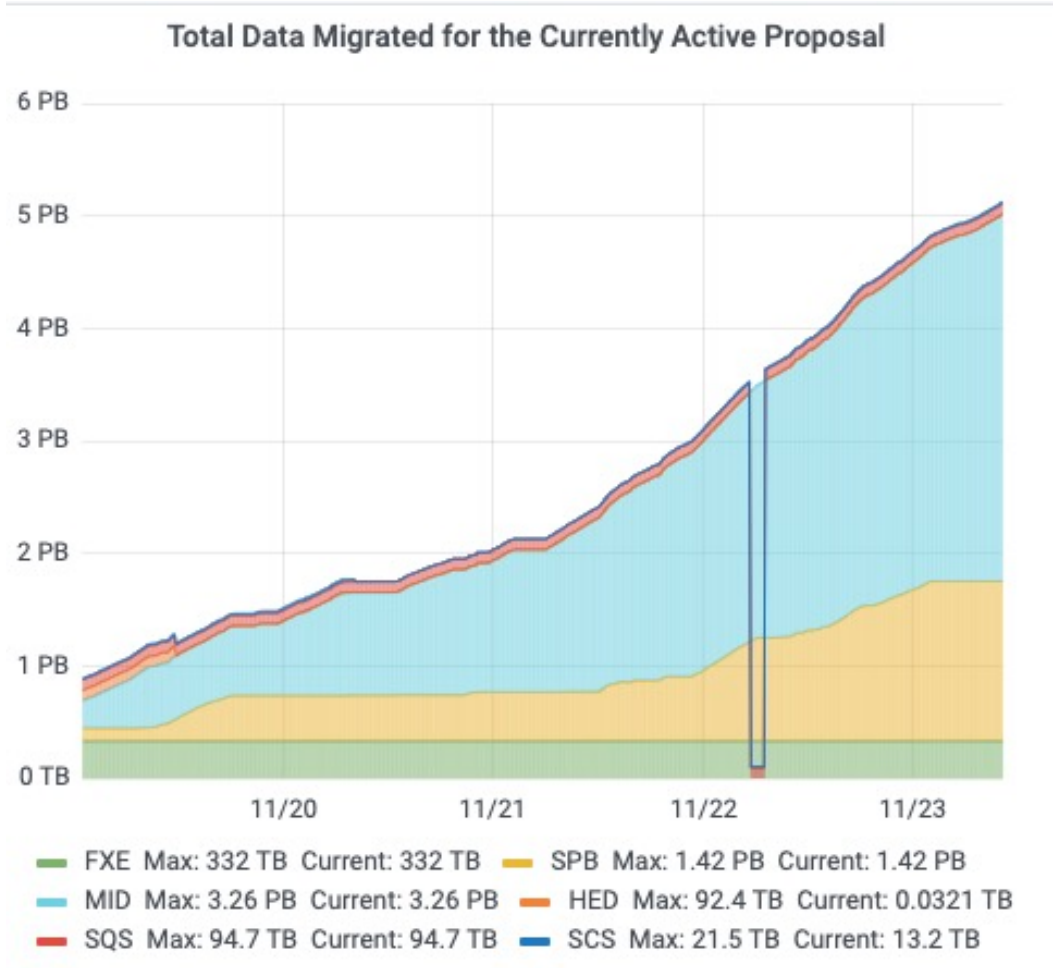
The European XFEL – a Data Perspective



Data Produced in CW46 2021 – All Data Systems Working



The European XFEL – a Data Perspective

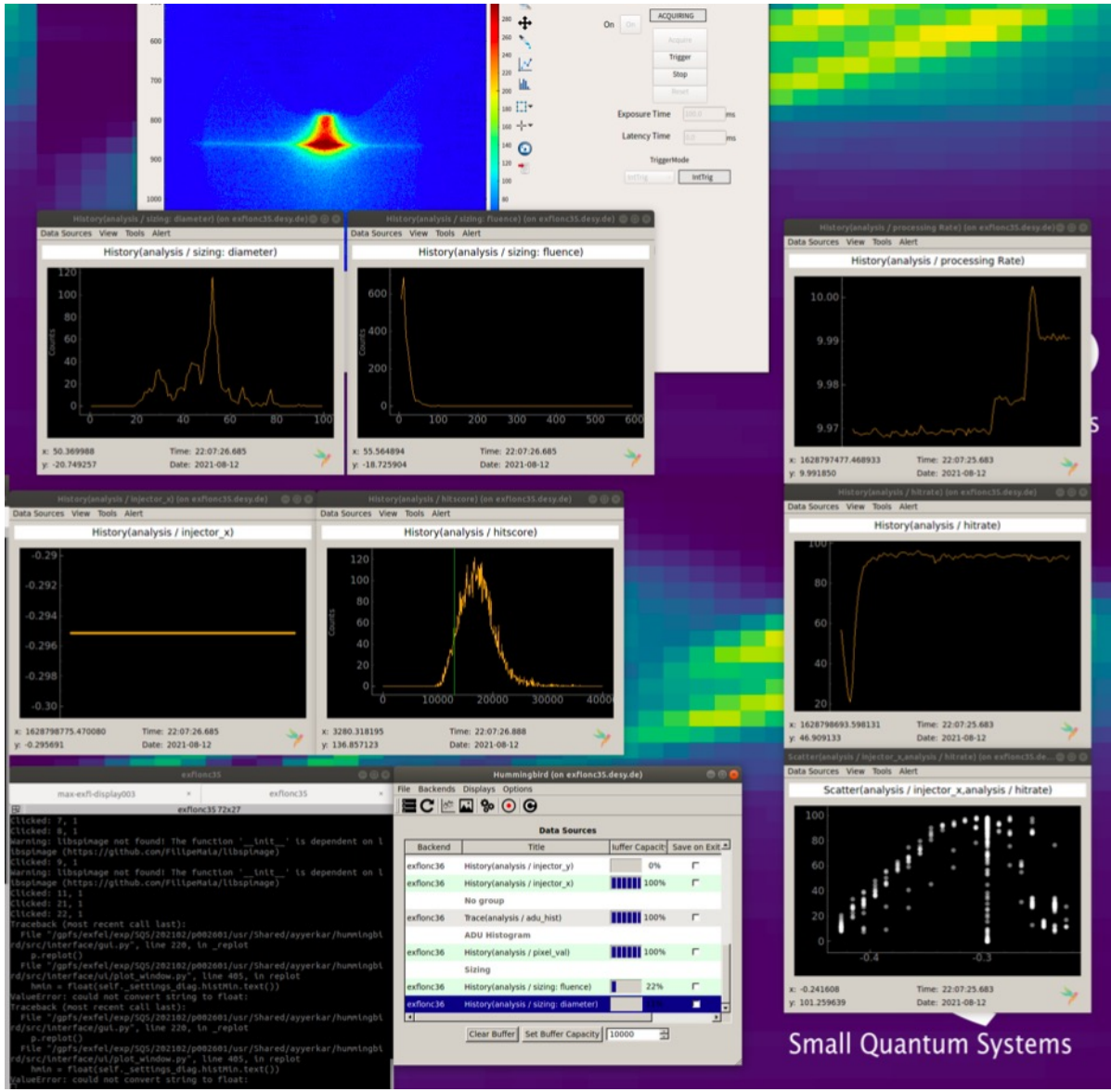


If data services are working, facilities like EU.XFEL easily produce scientific data in the Peta-Byte range in a few days

The European XFEL – a Data Perspective

Online Processing is equally “data-intensive“:

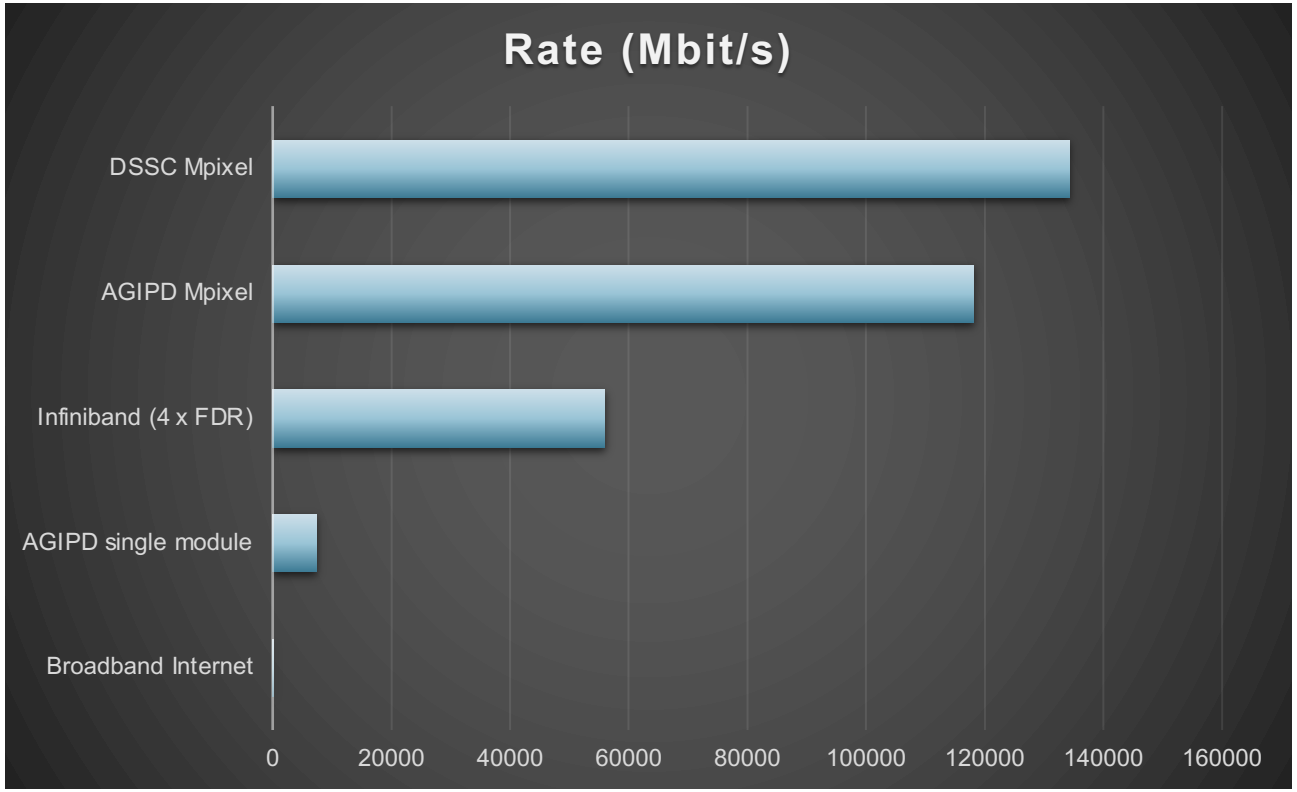
- Near-realtime streaming of up to 4KHz into „user space“
- Critical parameters:
  - Throughput: how may images per second can be calibrated, assembled and passed on to user-supplied code
  - Latency: with which delay does data arrive for user analysis



The European XFEL – a Data Perspective

Online Processing is equally “data-intensive“:

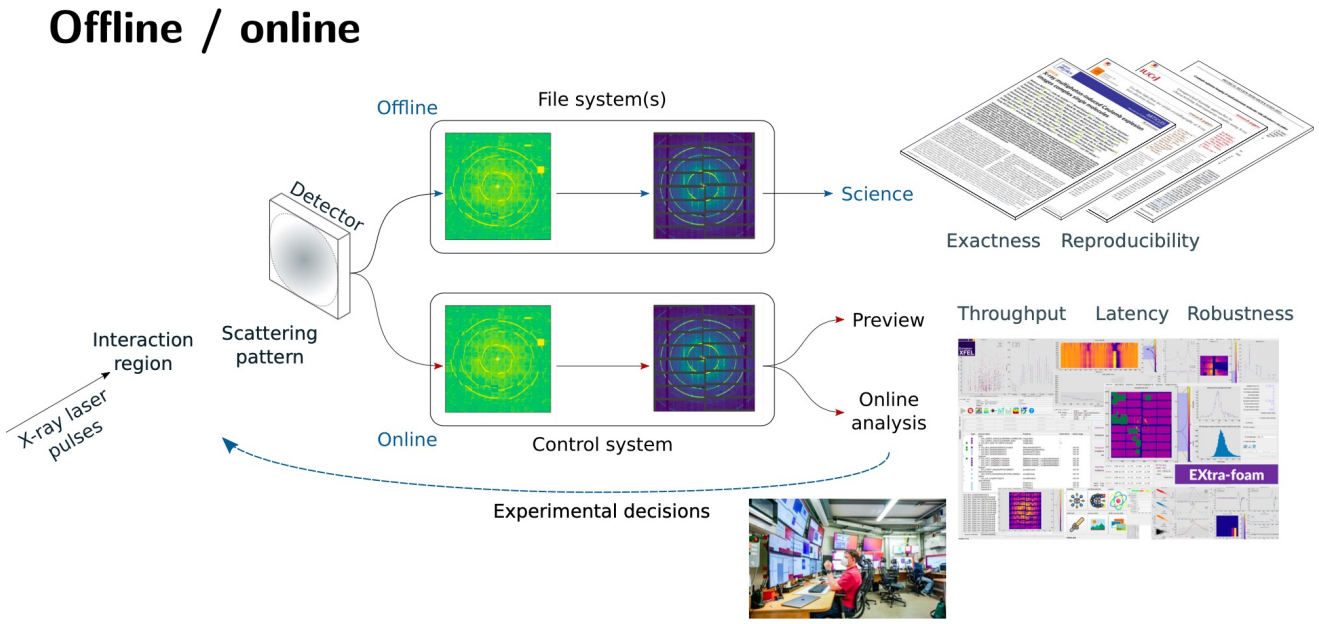
- Near-realtime streaming of up to 4KHz into „user space“
- Critical parameters:
  - Throughput: how may images per second can be calibrated, assembled and passed on to user-supplied code
  - Latency: with which delay does data arrive for user analysis
- **A full Mpix detector exceeds the Infiniband bandwidth of a single online calibration host**



The European XFEL – a Data Perspective

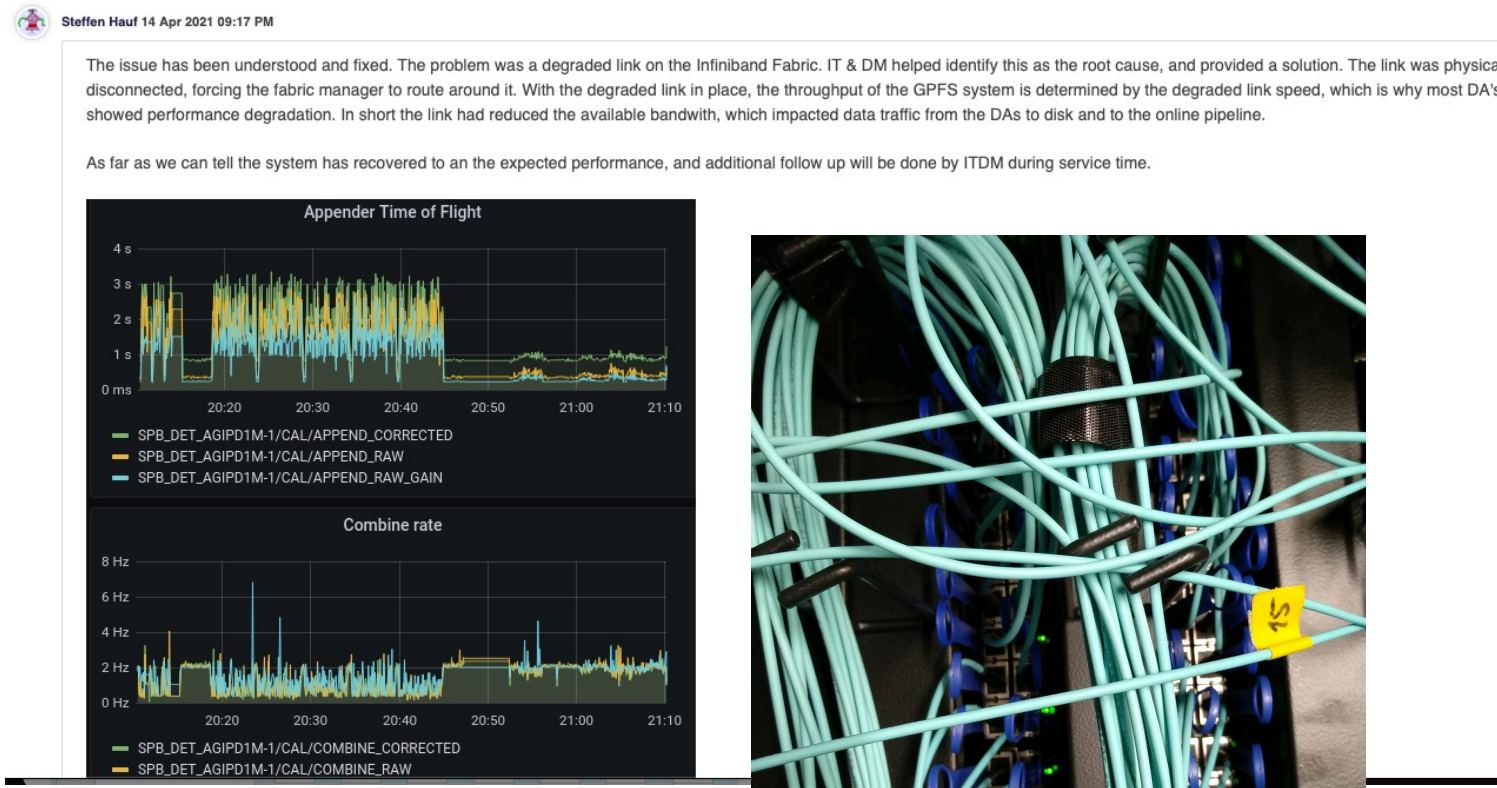
Additional data-related services that are critical to an experiment‘s success:

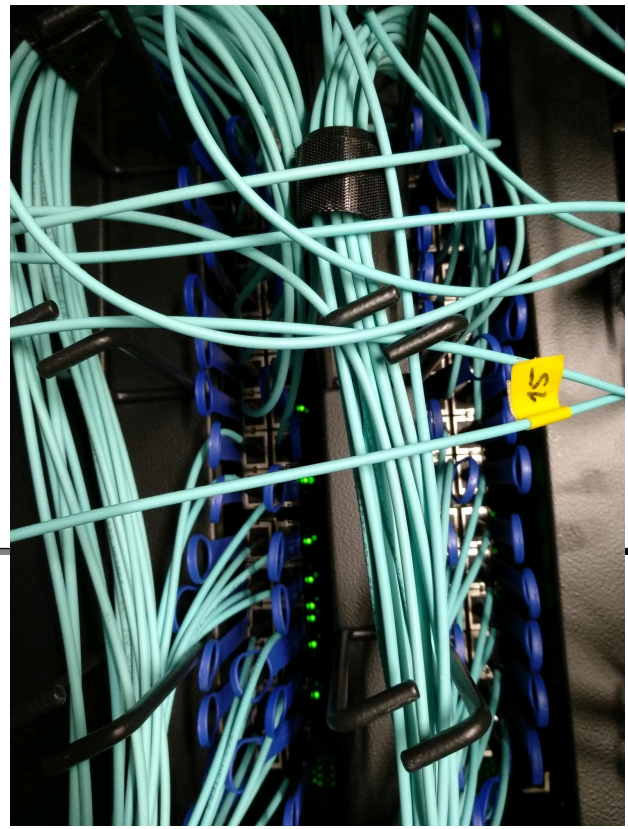
- Detector configurations optimized to the experimental needs
- Online and offline calibration
  - Processed data is calibrated and often the primary data product
  - Calibration constants are curated and selected by the facility
- Standardized data analysis tool-chains
  - Offline
  - Online
- Data access libraries
  - Abstract fragmented data originating from concurrent acquisition



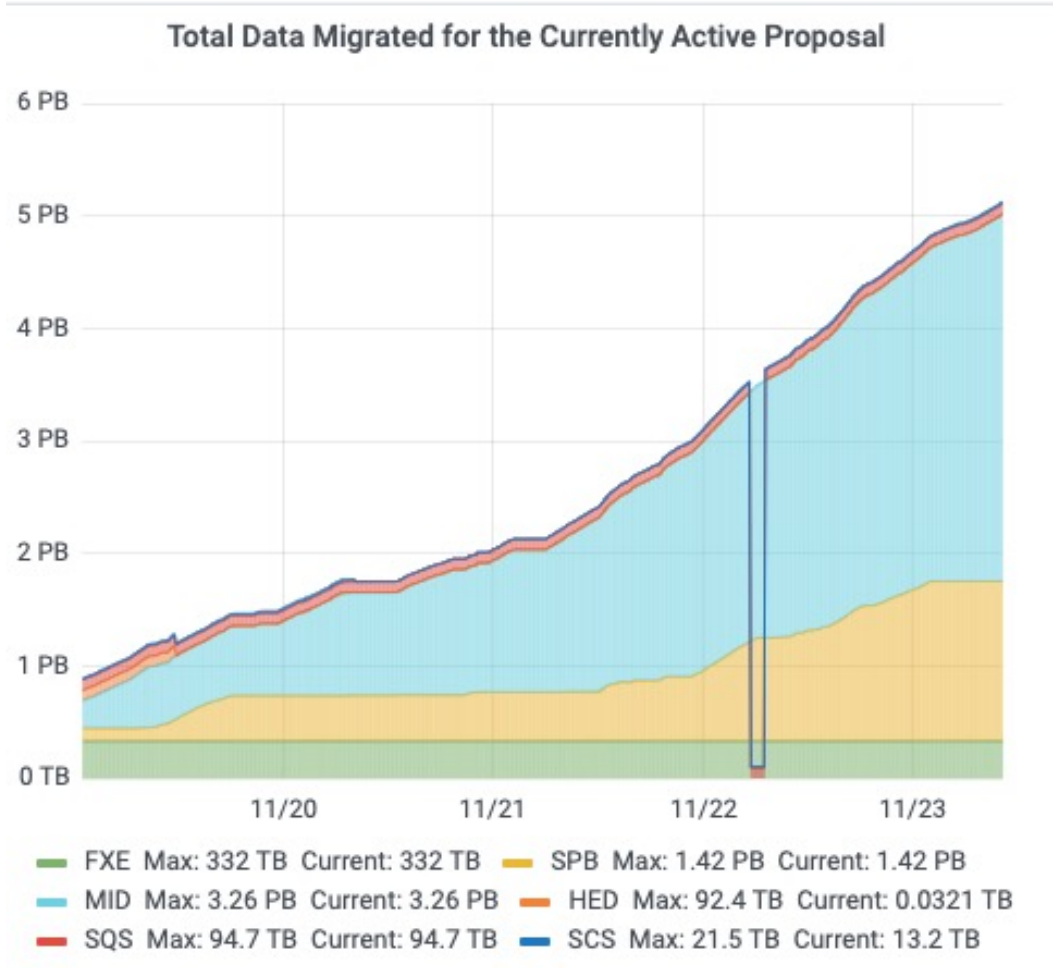
The Mystery of the Broken Fiber - an example of Complexity

* DOC observes strong jitter on AGIPD detector online preview, advises SPB instrument on issue (AGIPD is their main detector)
* SPB and DOC iterate settings, no apparent misconfiguration
* DOC observes correlation with DAQ - at some point DAQ and preview are mutually exclusive –if data is recorded SPB is blind
* Experts from data dept. groups reduce problem set to link-layer
* IT experts track possible bottleneck in network topology
* Second-level IT support unplugs cable
* **Example of a critical issue that could have stopped scientific data acquisition, and required expertise from multiple technical groups to locate and solve**





The European XFEL – a Data Perspective

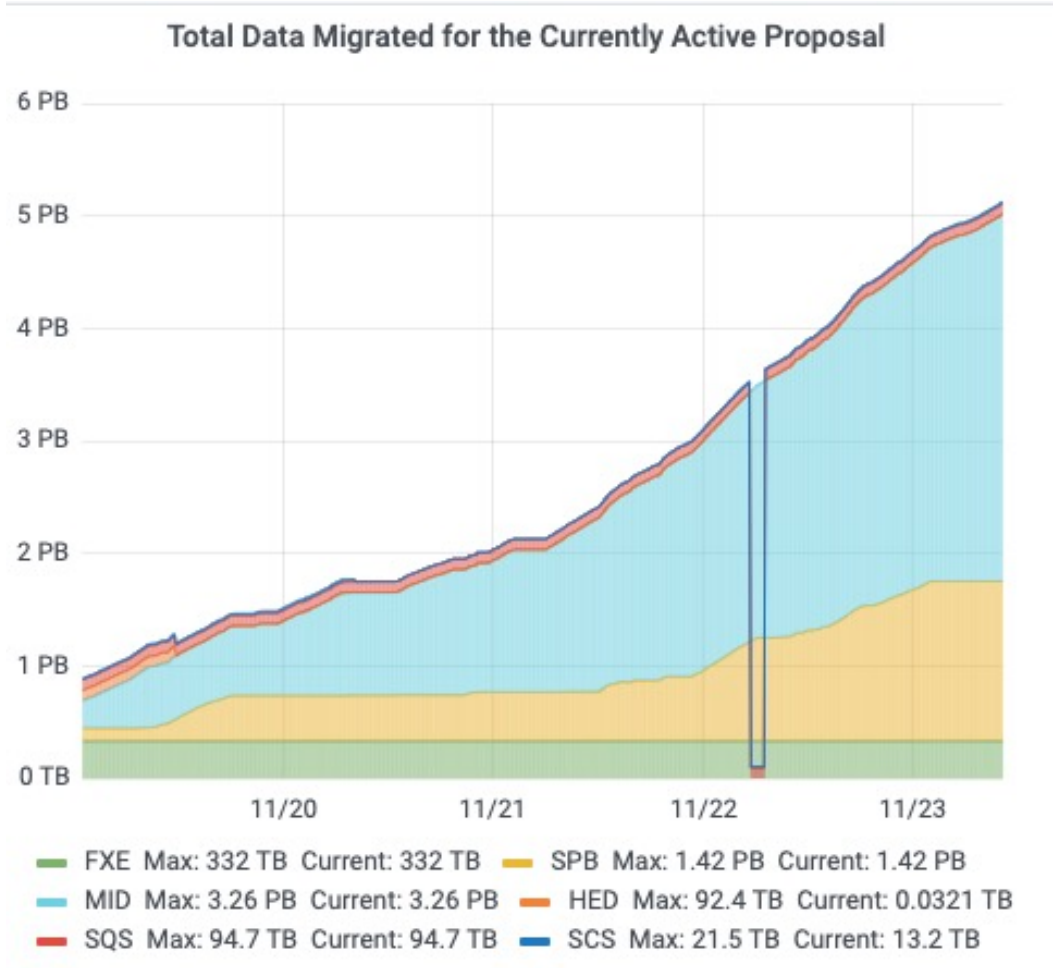


If data services are working, facilities like EU.XFEL easily produce scientific data in the Peta-Byte range in a few days

The complexity of the systems necessary for this significantly exceeds the knowledge and experience of any single person*

* At EU.XFEL data-related services e.g. are provided by the 5 groups of the DATA dept: Control, Electronic and Electrical Eng., IT & Data Management, Detectors, Data Analysis.

The European XFEL – a Data Perspective



If data services are working, facilities like EU.XFEL easily produce scientific data in the Peta-Byte range in a few days

The complexity of the systems necessary for this significantly exceeds the knowledge and experience of any single person

„Traditional“ support e.g. mainly through beamline staff and on-call/on-email experts reaches it‘s limits.

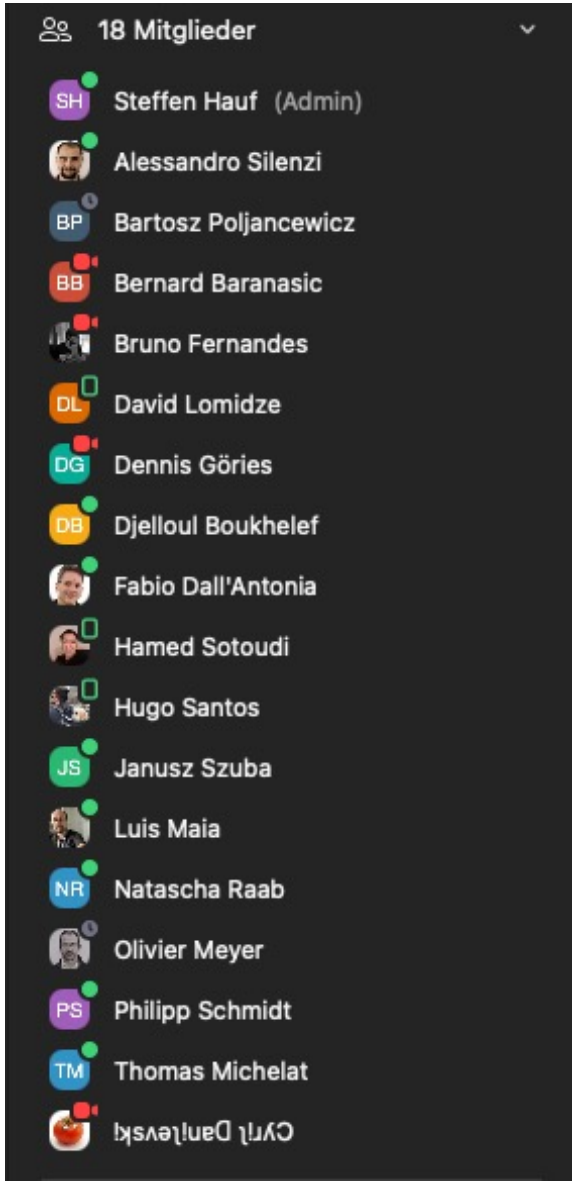
The Data Operation Center (DOC) – The Vision

* **Single point of entry** into DATA support during X-ray operation
* **Proactive**: monitors DATA provided services essential during X-ray operation
* **Cross-functional and inter-group**
* Not static: will enhance its capabilities over time

Apollo 11 had the computational power of a calculator, Mission control that of a few iPhones

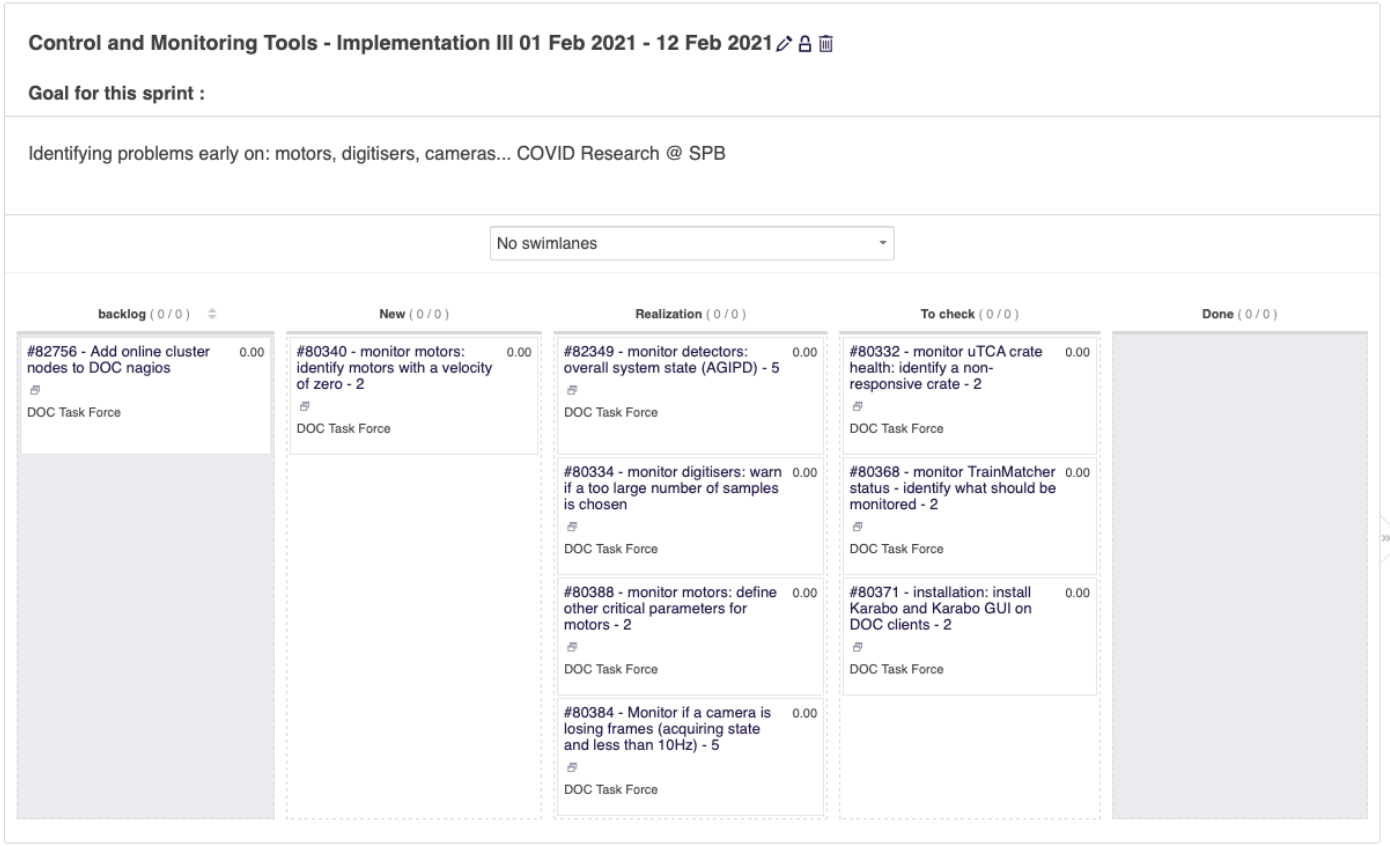
Creating the DOC

* A DOC Task Force was launched Dec 18th 2020
* DOC to reach initial operational capability in early 2021
* 18 TF members from across DATA dept., all groups represented
* Emphasis on extensible but always usable capability set from day-1 on
  → an agile approach lends itself to this
  * 3 Task force teams: hardware, documentation, monitoring tools
  * 3 concurrent two week sprints
  * TF members contributed between 20-40% of their time
  * DATA management was stakeholder in the effort

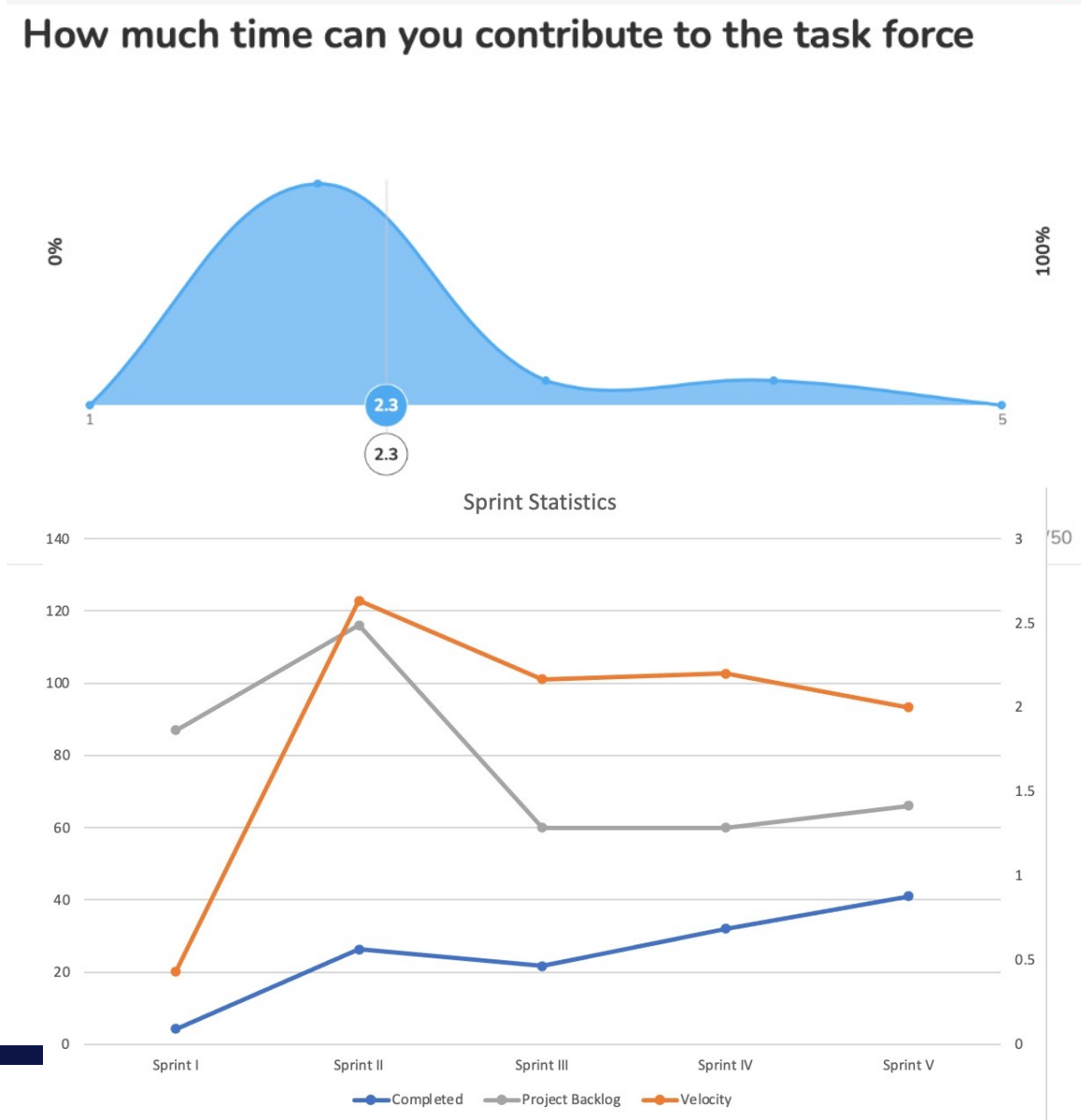


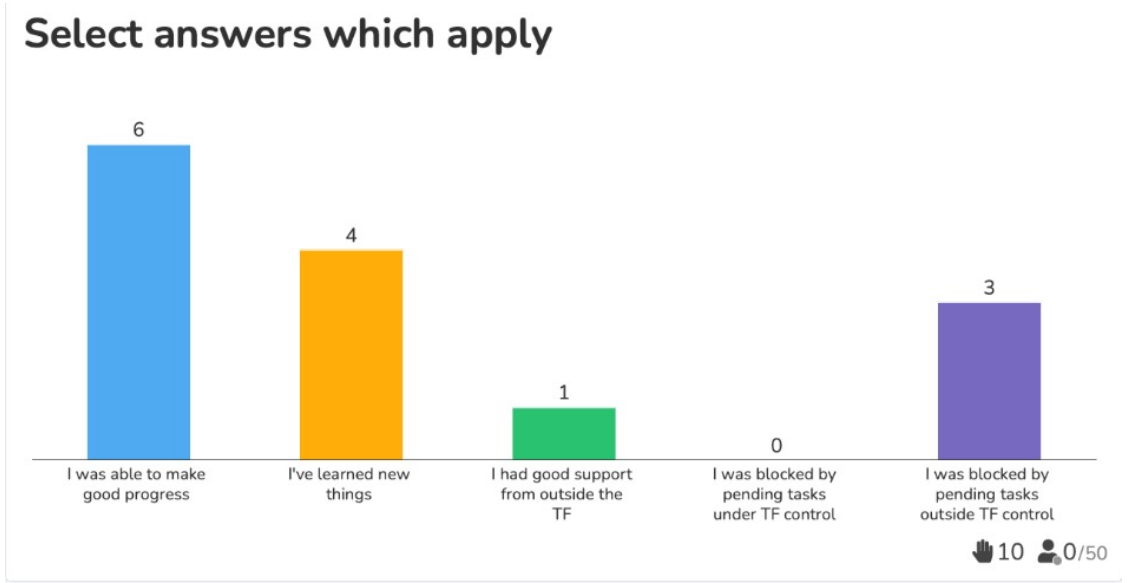
Task Force Work Organization

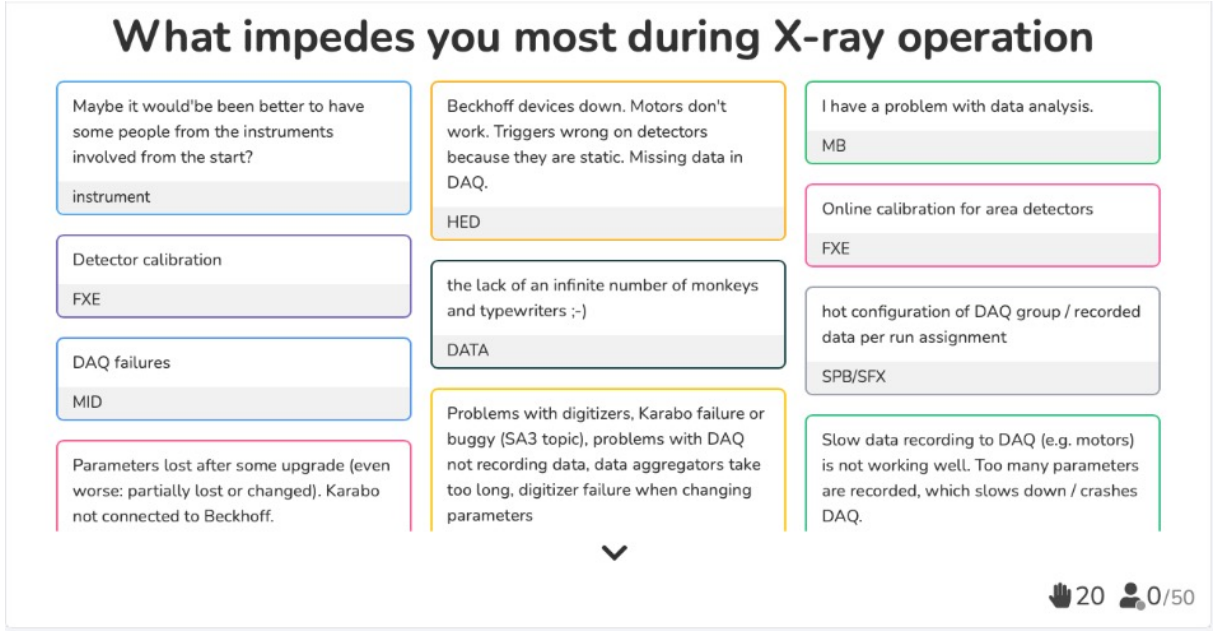
* SCRUM and task board
* Product owner defines overall priority in coordination with stakeholders
* Team defines effort for tasks
* PO and team agree on *Definition of Done*
* Anybody in the team can take on a task
* A task may be further specified in a sprint but should not change in scope



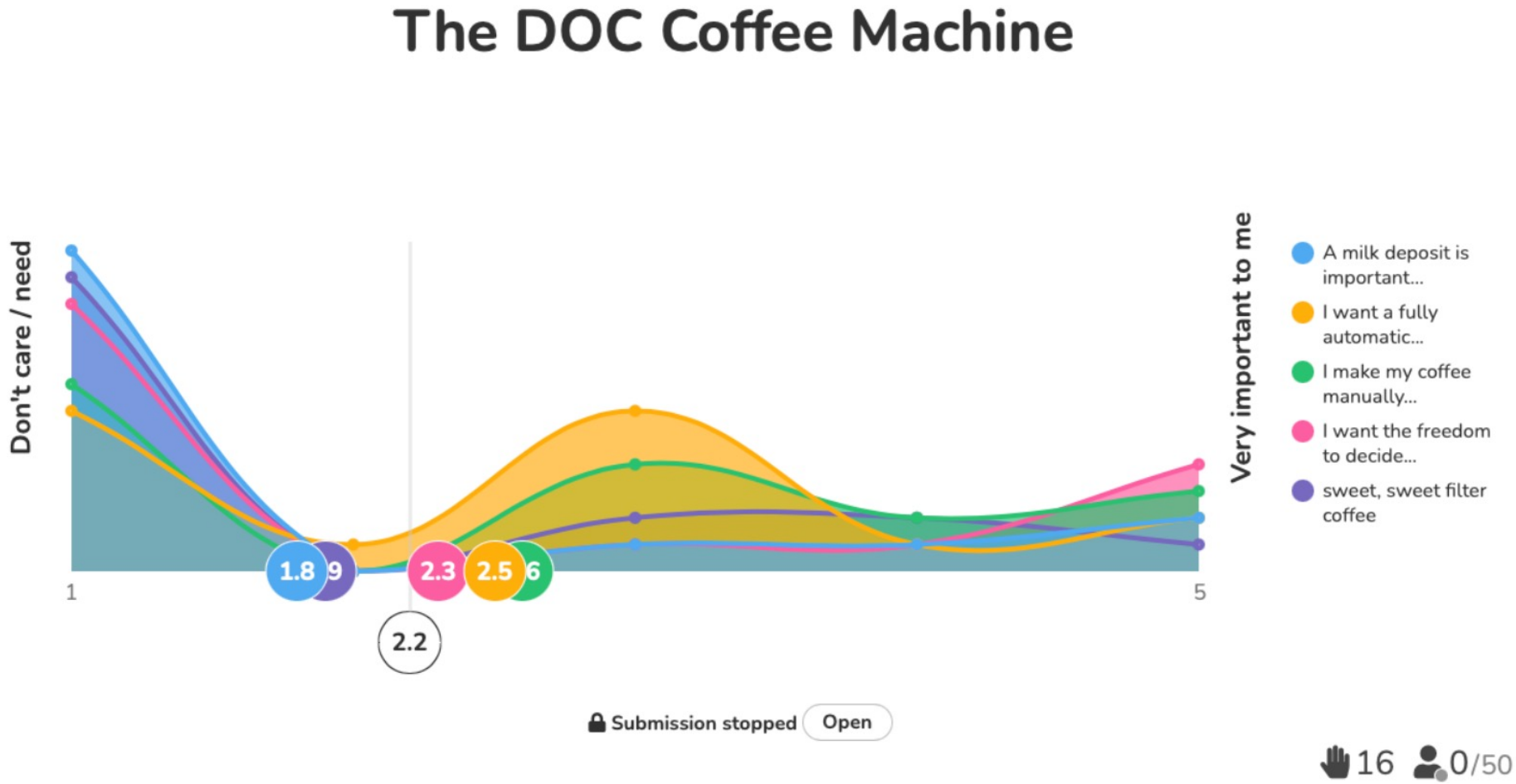
DOC Taskforce – Measuring Progress







DOC Taskforce – Other Interesting Surveys



The Data Operation Center (DOC) – The Vision

* **Single point of entry** into DATA support during X-ray operation
* **Proactive**: monitors DATA provided services essential during X-ray operation
* **Cross-functional and inter-group**
* Not static: will enhance its capabilities over time
* Two people on shift and a Data Run Coordinator (DRC)
  * DRC is the interface of the DOC to facility coordination
  * DRC is an escalation path

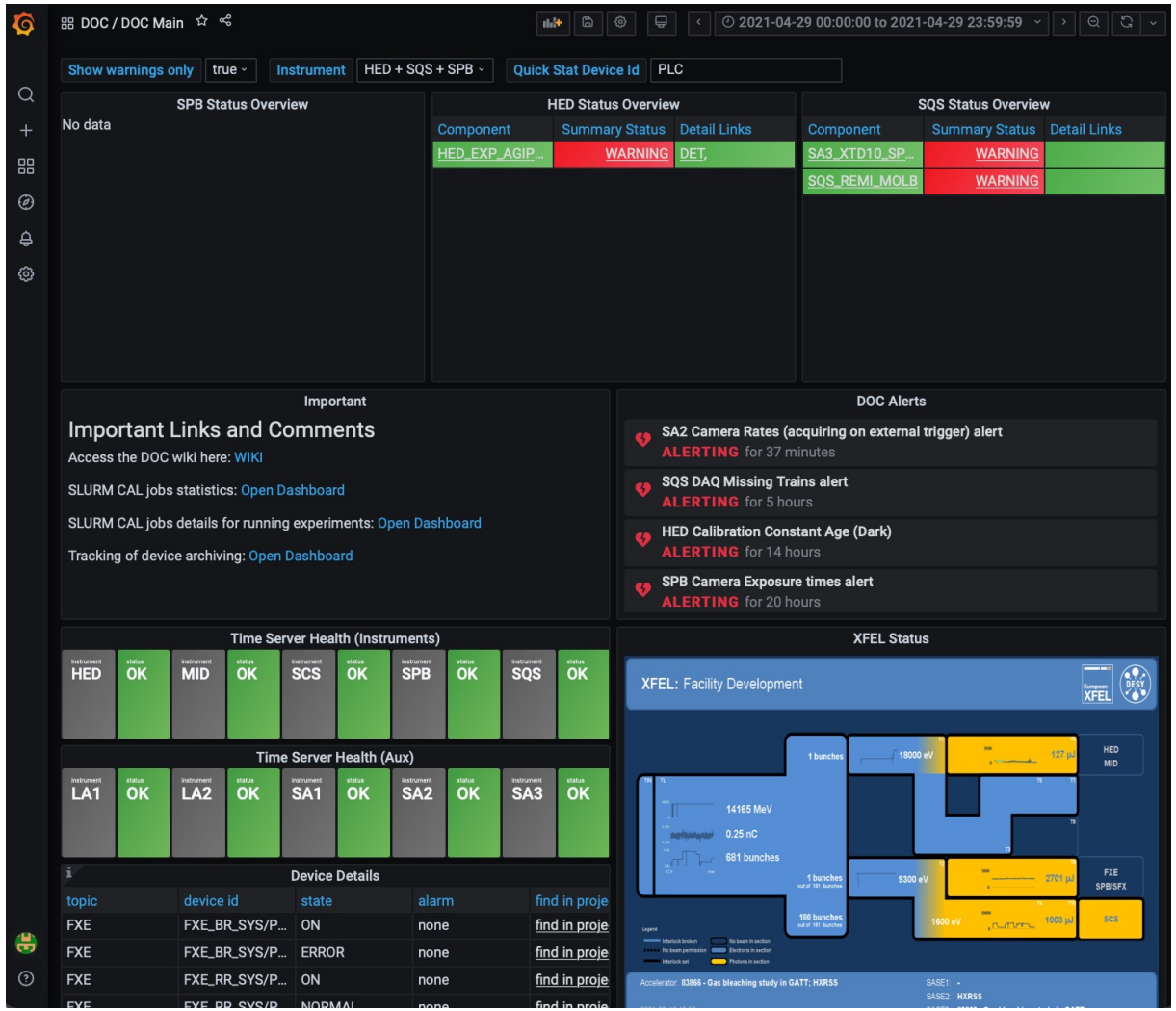
XFEL DOC has access to a state-of-the art timeseries DB, monitor 100+ servers, 10000+ devices and jobs on an HPC cluster. We do have less screens than NASA, but probably a similar amount of pixels available

DOC Technology Choices

* Three months to setup: early decision - do not reinvent the wheel

* Identify existing systems to use
  - ► for communication (Zoom) - now testing Zulip
  - ► documentation (Redmine Wiki)
  - ► ticketing (Redmine)
  - ► logs (PSI elog)

* But... for monitoring
  * ~~Karabo~~: somewhat obvious, but large knowledge differences. Programming it, is an „expert“ skill
  * **Grafana + Influx: was new to almost everyone, we learned together and supported each other**

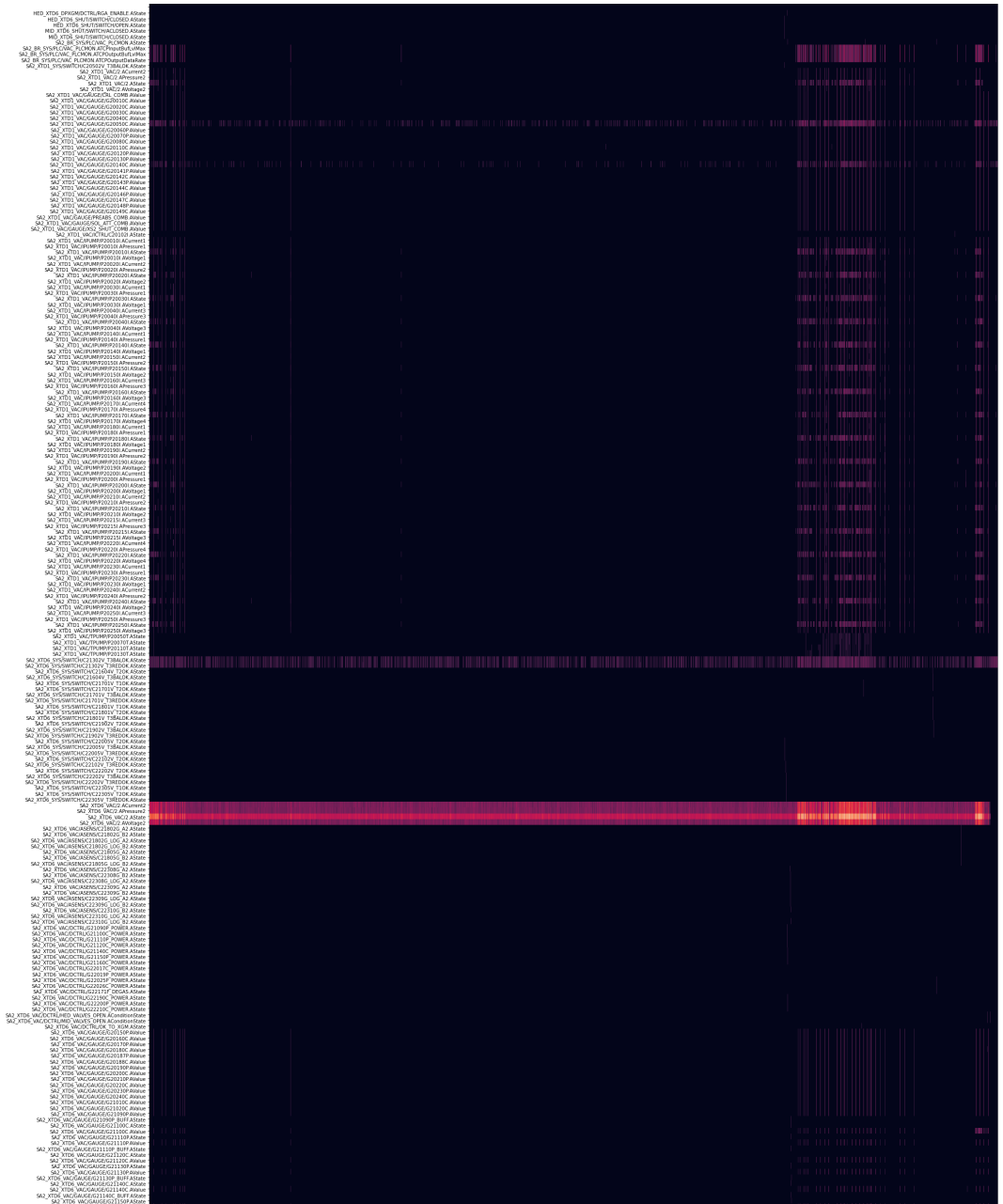


DOC Technology Choices

* Three months to setup: early decision - do not reinvent the wheel

* Identify existing systems to use
  - ► for communication (Zoom) - now testing Zulip
  - ► documentation (Redmine Wiki)
  - ► ticketing (Redmine)
  - ► logs (PSI elog)

* But... for monitoring
  * ~~Karabo~~: somewhat obvious, but large knowledge differences. Programming it, is an „expert“ skill
  * **Grafana + Influx: was new to almost everyone, we learned together and supported each other**



Data Operation Center Roles

Data Run Coordinator (DRC)

- Responsible for DATA operation during a given week
- Information Path
- Escalation path
- On-Site or remote

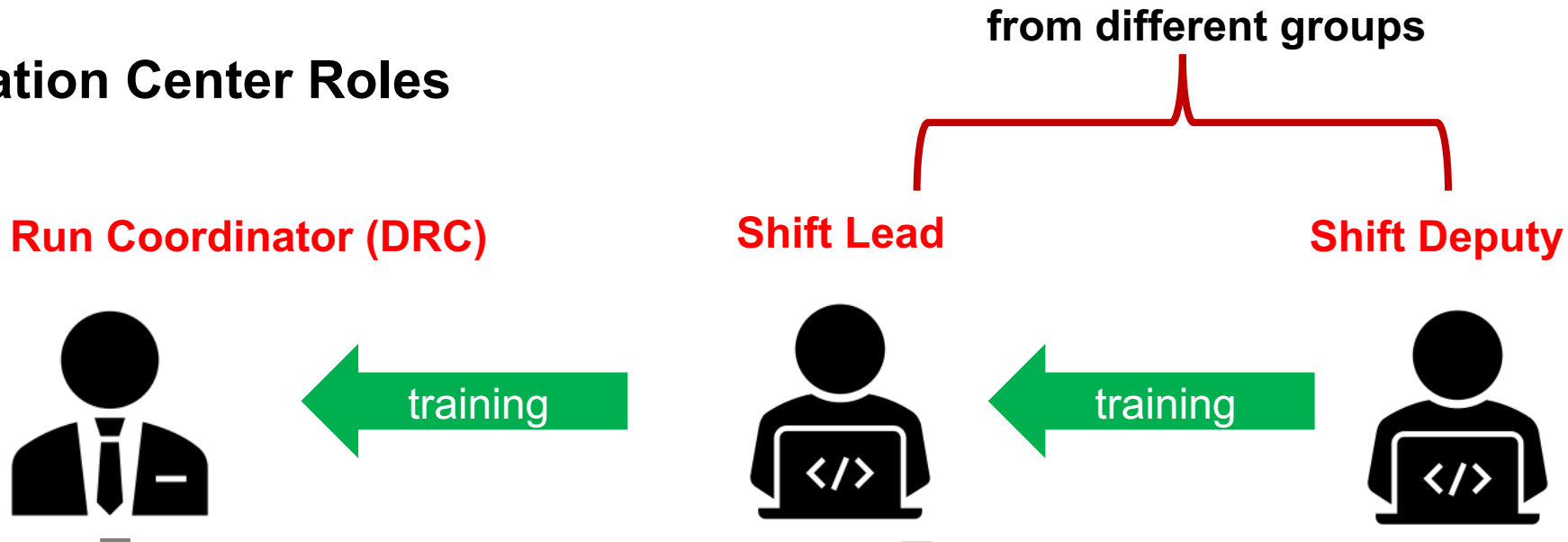
Shift Lead

- Responsible for a given shift
- Experienced enough to categorize most problems
- Proactive
- Reactive
- On-Site

Shift Deputy

- Supports DATA services at the facility
- Proactive
- Reactive
- On-Site

Data Operation Center Roles

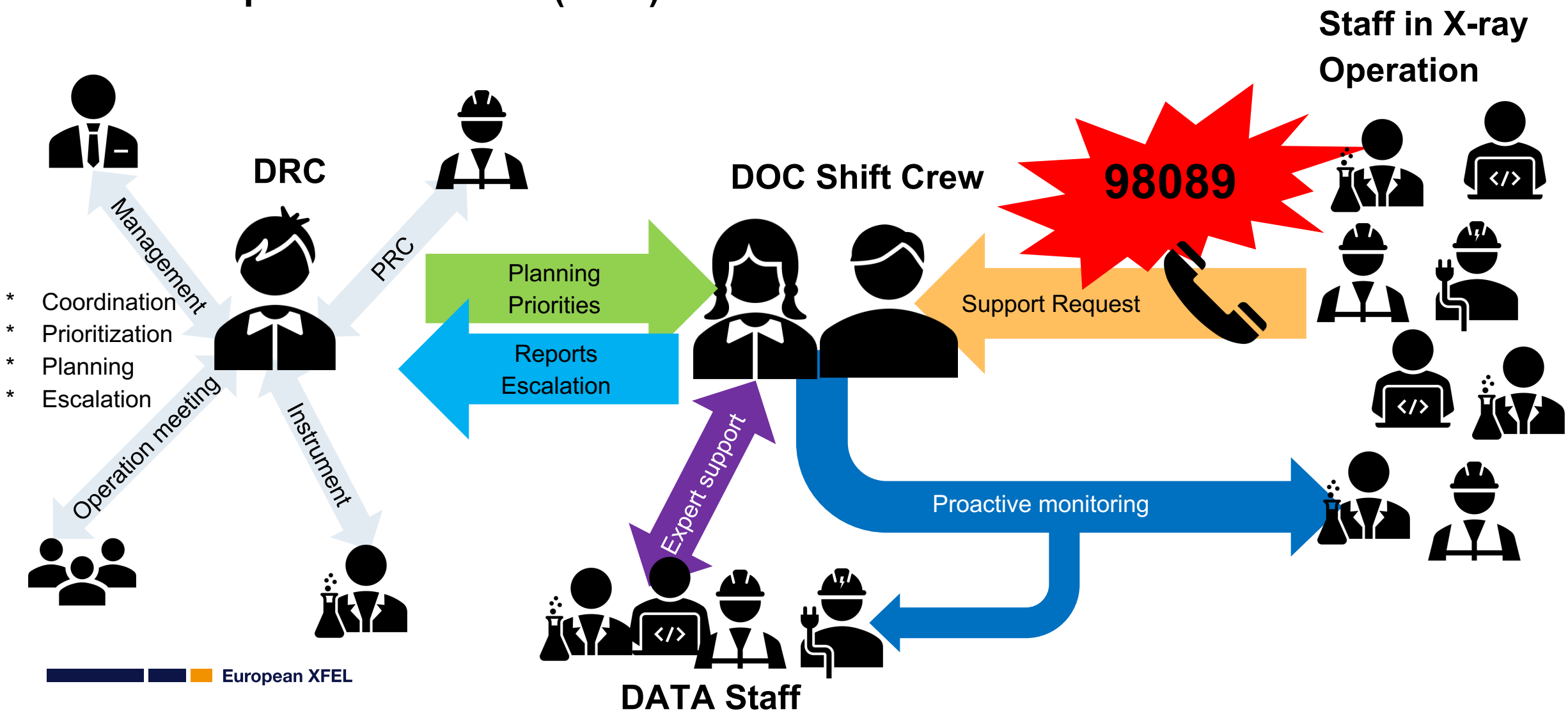


- Responsible for data operation during a given week
- Information Path
- Escalation path
- On-Site or remote

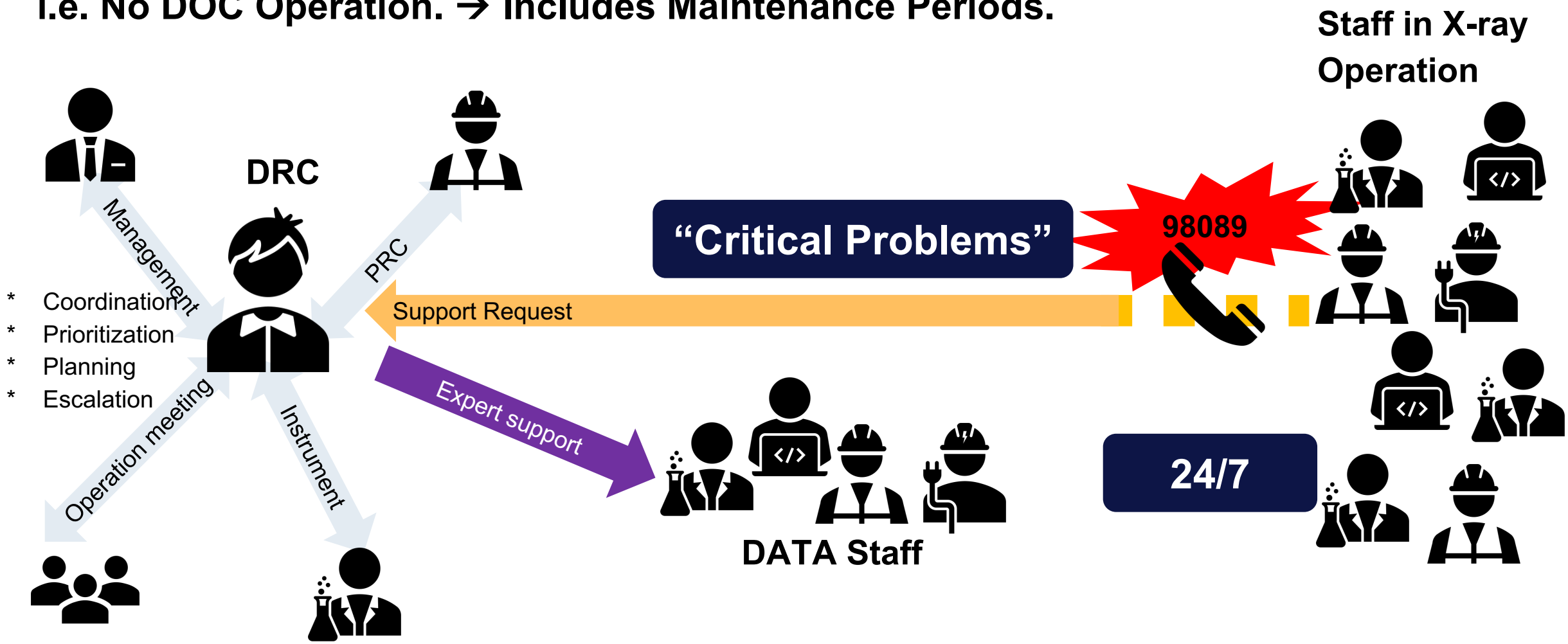
- Experienced enough to categorize most problems
- Proactive
- Reactive
- On-Site

- Supports data services at the facility
- Proactive
- Reactive
- On-Site

The Data Operations Center (DOC) – 7/11 hours



On-Call DATA support outside of priority X-Ray delivery i.e. No DOC Operation. → Includes Maintenance Periods.



What the Data Operation Center Supports

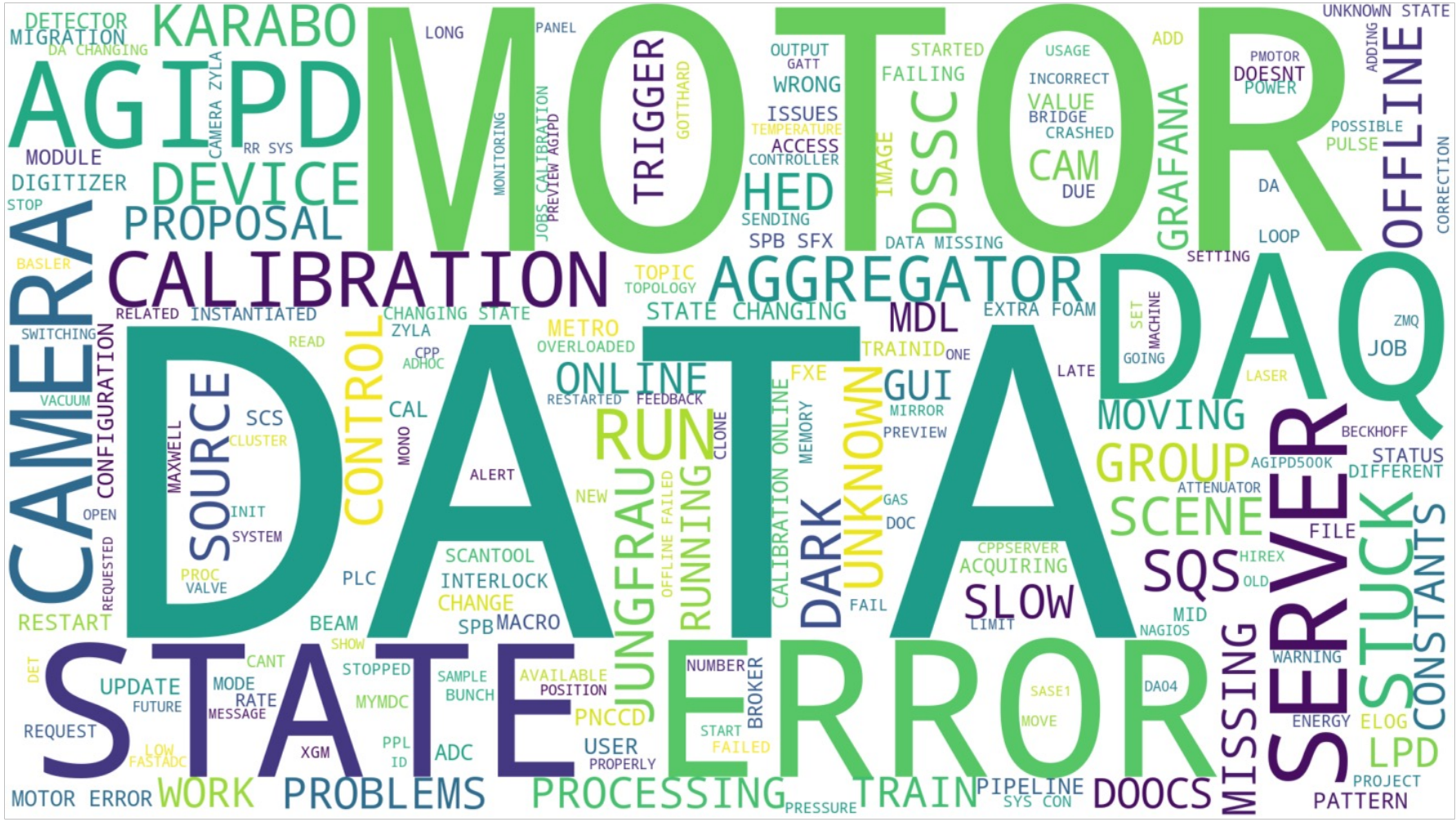
Evaluation of word frequency from ticket subjects:
the larger the more frequent

1420

incidents handled (1.3.21 -25.7.22)

384 follow-ups created

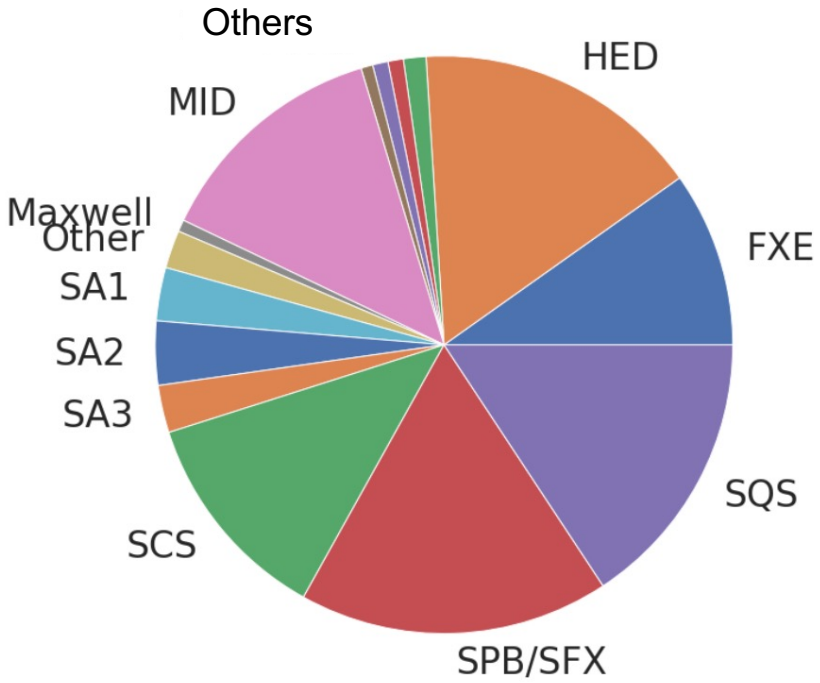
226 closed

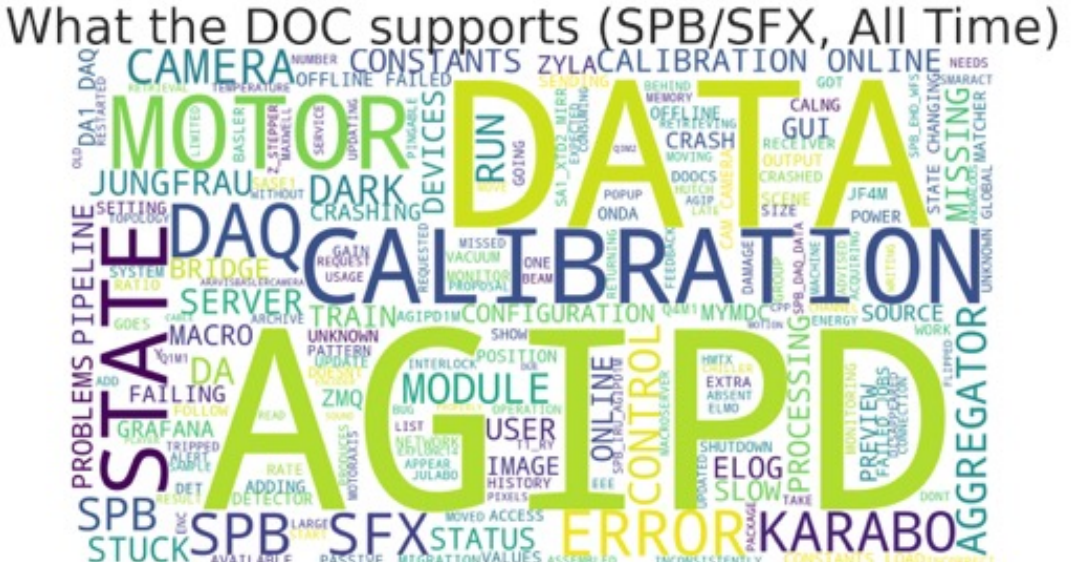


Who the Data Operation Center Supports

- Number of support requests by different instruments are comparable
- Support needs vary significantly, depending on equipment and experimental methods used

Support by Installation

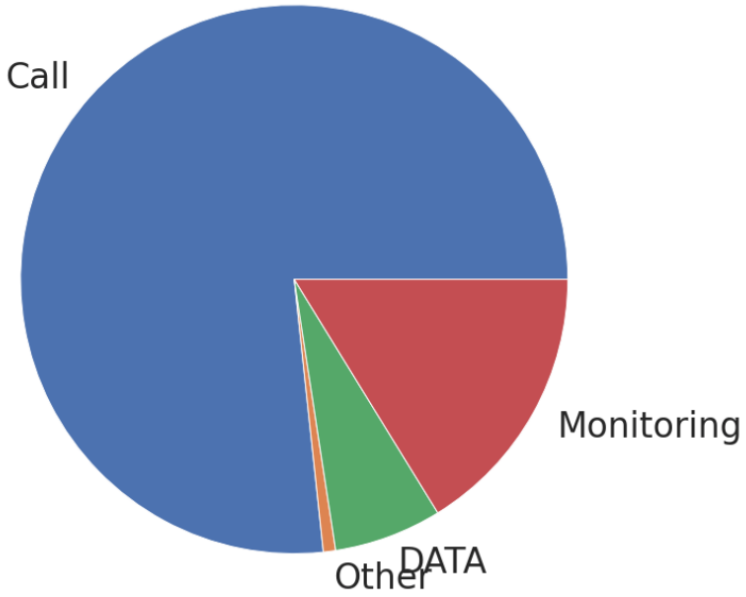




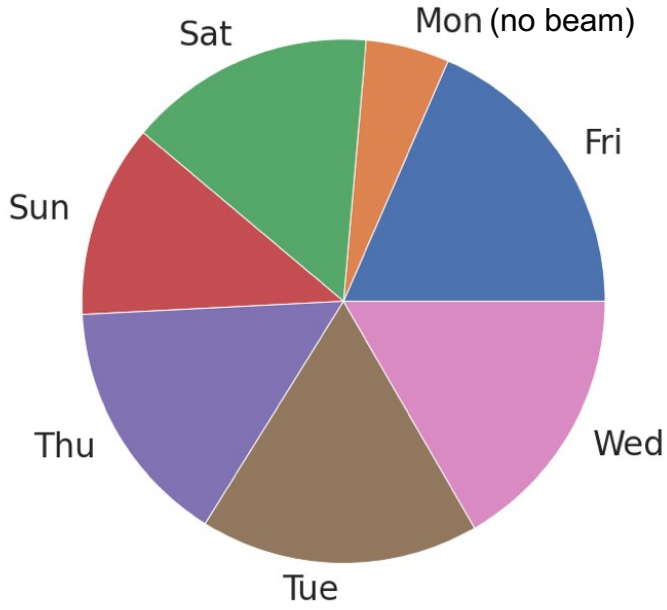
When the Data Operation Center Supports

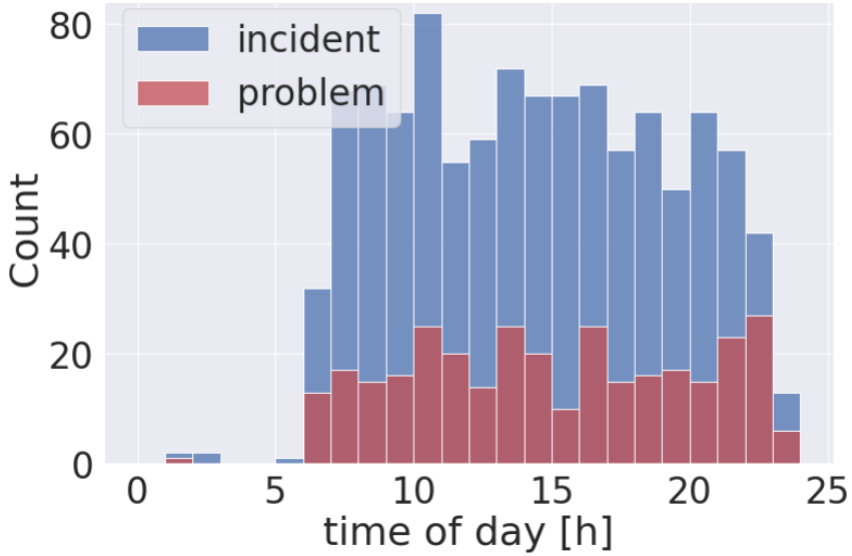
- Support needs are roughly uniform in time
- ~25% of support is initiated through proactive monitoring of systems

How Support is triggered?



When is support given

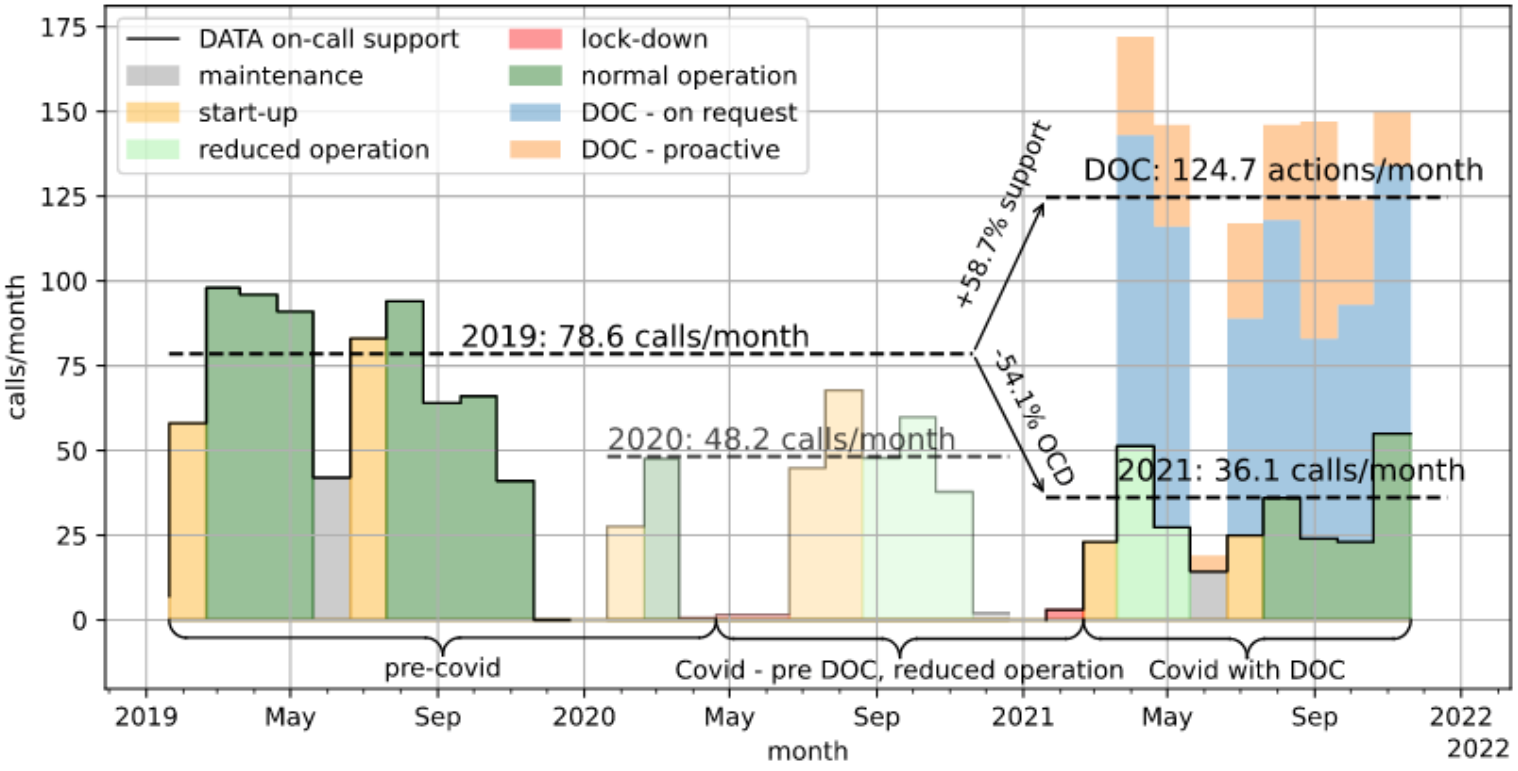




Overall DOC Impact on Support

- More support is given through the DOC than before
- Fewer ad-hoc, unplanned support activities, i.e. less disruption to routine activities of staff members
- Proactive support catches problems before they can impact user operation

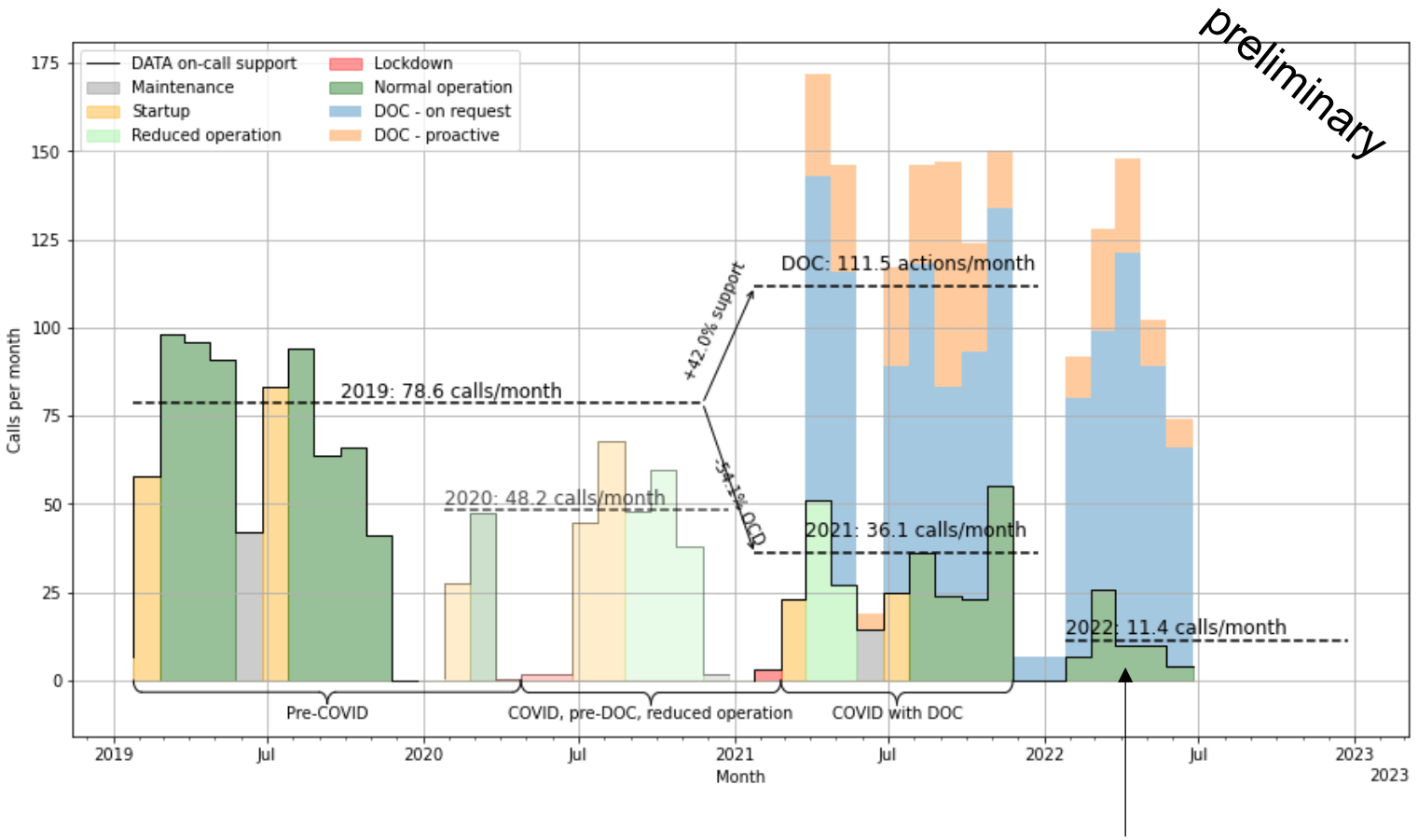
→ **Increase in scientific "efficiency" (more, higher-quality beam time)**
→ **Increase in staff efficiency**



Overall DOC Impact on Support

- More support is given through the DOC than before
- Fewer ad-hoc, unplanned support activities, i.e. less disruption to routine activities of staff members
- Proactive support catches problems before they can impact user operation

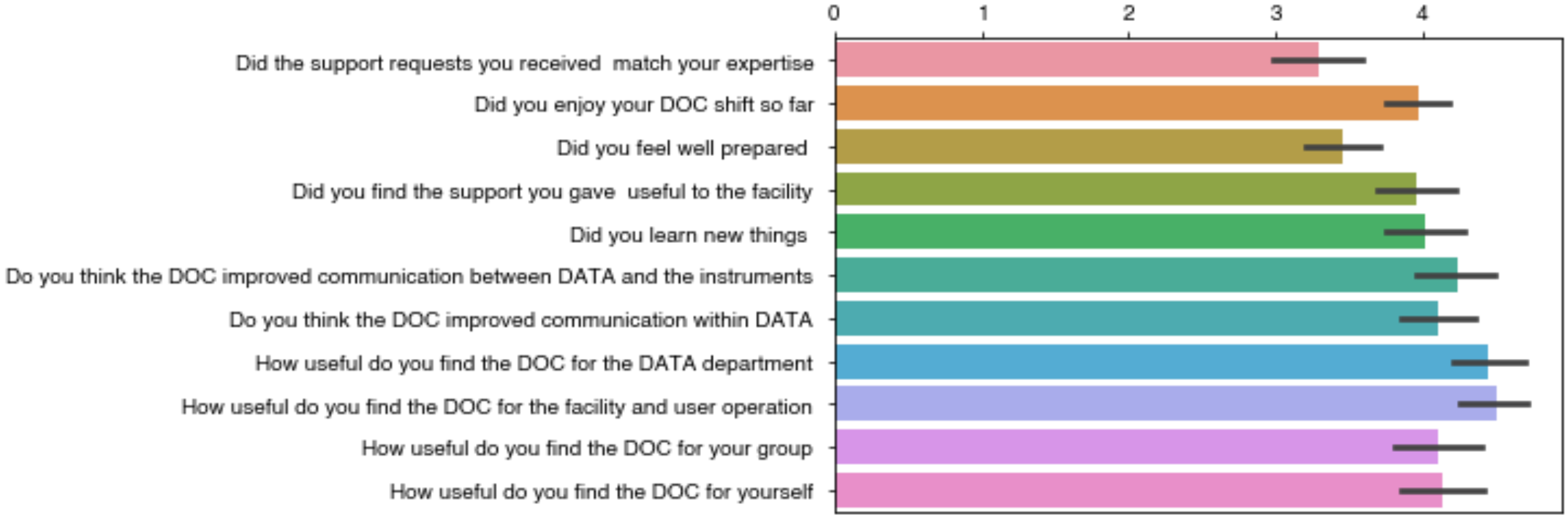
→ **Increase in scientific "efficiency"**
(more, higher-quality beam time)
→ **Increase in staff efficiency**



CTRL group only,
DATA likely at about same level as before

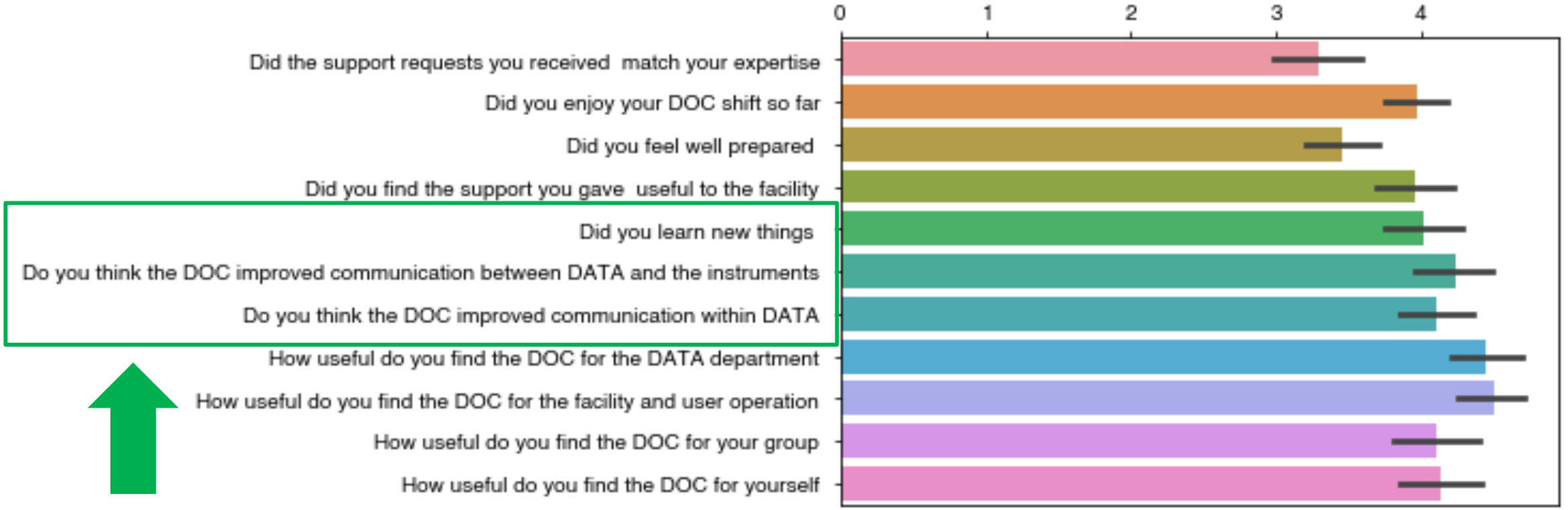
So it‘s a Call Center?

Staff Views on Data Operation Center



Selected results from the DOC shift crew survey. Answers were on a scale from 1 (very negative) to 5 (very positive). The average feedback was generally very positive, with a small spread in answers.

Staff Views on Data Operation Center



Staff Development

Selected results from the DOC shift crew survey. Answers were on a scale from 1 (very negative) to 5 (very positive). The average feedback was generally very positive, with a small spread in answers.

Addendum: The Bring Up Days

* Since 2022 the DATA department conducts a structured hand-over of data services to instruments at the end of maintenance periods (2x annually).
  * Check list of essential data services that are verified together with instrument staff over the course of two sessions:
    ► Essential controls (PLCs, timing, Control System)
    ► Data Acquisition and Processing (full stack, as much as is possible without beam)

* Has lead to a noticeable decrease in number of support calls to the DOC coming out of maintenance periods.
  * Especially, “small” issues like misconfigurations are caught early on and quickly resolved

Summary

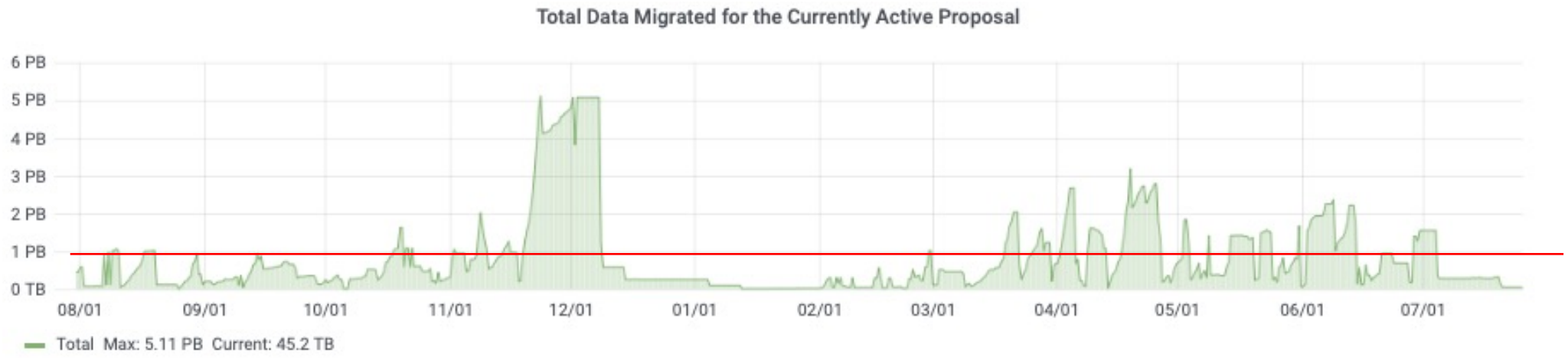
* The Data Operation Center is an innovative way to handle operational support for data services at the European XFEL
  * traditional on-call/mail-in support reaches it‘s limits at complex facilities like EU.XFEL
  * bundles expertise and simplifies the interface for beamline staff
  * proactive, with many issues identified before they impact operation

  → significantly increases operational efficiency in a complex infrastructure and service environment

Summary

* The Data Operation Center is an innovative way to handle operational support for data services at the European XFEL
  * traditional on-call/mail-in support reaches it‘s limits at complex facilities like EU.XFEL
  * bundles expertise and simplifies the interface for beamline staff
  * proactive, with many issues identified before they impact operation

  → significantly increases operational efficiency in a complex infrastructure and service environment



Summary

* The Data Operation Center is an innovative way to handle operational support for data services at the European XFEL
  * traditional on-call/mail-in support reaches it‘s limits at complex facilities like EU.XFEL
  * bundles expertise and simplifies the interface for beamline staff
  * proactive, with many issues identified before they impact operation

  → significantly increases operational efficiency in a complex infrastructure and service environment

* The Data Operation Center fosters Staff Development
  * Cross-training between shift crews
  * Communication and awareness of achievements and problems outside staff‘s usual work portfolio
  * 80+ members of the ~100 people in the DATA department have done shifts so far.

* The Data Operation Center identifies avenues of improvement: we have the statistics and logs to see what‘s going wrong and why. These are used e.g. also in planning the bring up days.