



Programme: GEN

Project/WP: System Engineering Department

EIROForum Workshop 2020 Big Data - Summary Report

Document Number: ESO-383133

Document Version: 1

Document Type: Correspondence/Memo (COR)

Prepared On: 2020-11-16

Document Classification: Public

Prepared by: Egner, Sebastian Elias

Name



Authors

Name	Affiliation
Sebastian Egner	ESO

Change Record from previous Version

Affected Section(s)	Changes / Reason / Remarks
All	Initial release of the document



Contents

1. Overall Summary	5
2. Programme	6
2.1 Scientific Programme	6
2.2 Social Programme	8
2.3 Detailed Agenda	10
3. Organisation	12
3.1 Scientific Organizing Committee	12
3.2 Local Organizing Committee	12
4. Feedback and Evaluation	13
5. Lessons Learnt and Recommendations	18
5.1 Online Workshops	18
5.2 Preparation	18
5.3 Engagement	18
5.4 Social Interaction	19
5.5 Topics for future workshops	19

EIROFORUM WORKSHOP
BIG DATA
26 - 29 OCTOBER 2020 13:30-17:30 CET

- BIG DATA ✨ CERN EMBL ESA ESO ESRF EUXFEL EUROFUSION ILL
- EXTRACTION / STORAGE / DATA ENGINEERING
- MACHINE LEARNING / NEURAL NETWORK
- HIGH PERFORMANCE COMPUTING / CLOUD

ONLINE WORKSHOP. NO CONFERENCE FEE.
REGISTER BY 18 OCTOBER (SUN) 2020 AT
[HTTPS://INDICO.CERN.CH/EVENT/881752](https://indico.cern.ch/event/881752)

CHRISTIAN JORAM (CERN) / THOMAS SCHNEIDER (EMBL) /
THIBAUT PRODHOMME (ESA) / SEBASTIAN EGNER
(ESO) / JEAN-CLAUDE BIASCI (ESRF) / MARKUS KÜSTER
(EUXFEL) / JOAO FIGUEIREDO (EUROFUSION) / ANDREA
MURARI (EUROFUSION) / PAOLO MUTTI (ILL)

IN PHYSICS BIOLOGY & ASTRONOMY

Logos: CERN, ESA, EMBL, EIRO Forum, ESO, ESRF, EUROfusion, European XFEL, NEUTRONS FOR SCIENCE



1. Overall Summary

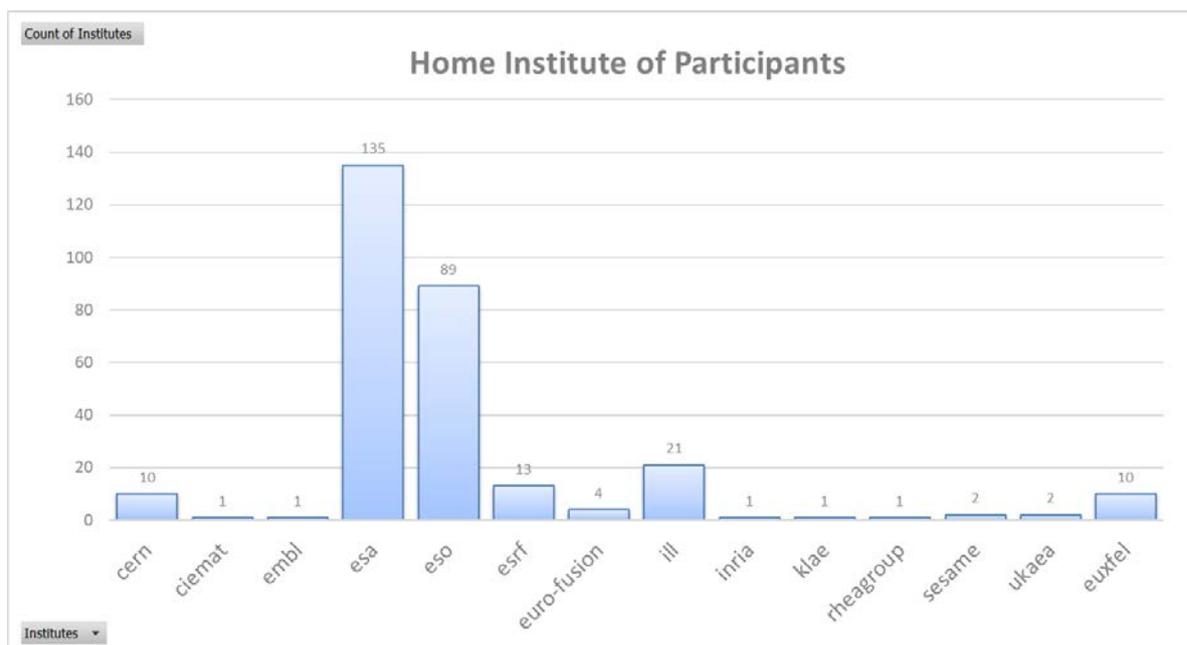
The EIROForum Workshop is a bi-annual event, which is organized by the Instrumentation Working Group (IWG) of the EIROForum (CERN, EMBL, ESA, ESO, EuXFEL, ESRF, EuroFusion, ILL). The workshop covers a different topic every time, which is of common relevance for all the organizations of the EIROForum. The topic of the workshop in 2020 was “Big Data – From acquisition to data mining”.

The workshop was originally scheduled as an in-person meeting in spring 2020, but due to the COVID-19 situation, was postponed to autumn 2020 and held as an online workshop. It took place from 26th to 29th October 2020 and was hosted by ESO. Due to the online nature, the workshop turned out to be very popular, with typically between 85 and 130 attendees to any one talk and over 310 individual registrations in total, from all EIROForum organizations:

The workshop ran over 4 afternoons, with in total 22 talks on Data Acquisition, Data Storage, Data Mining and Machine Learning in the field of Particle Physics, Nuclear Fusion, X-ray Free-electron Laser, Synchrotron Radiation, Earth Observation and Astronomy. On Tuesday evening, a special event was organized, with a virtual tour through the Paranal Observatory of ESO and an introduction to Black Holes, the topic of the Nobel Prize in Physics of 2020.

The feedback by the participants was overall very positive, the topics of the talks were considered interesting, apparently filling a gap in workshops on this topic. The format of 4 afternoons seems to be the right one for this kind of workshop, and technically the employed tools were working very well. On the other hand, the participants pointed out that more thinking needs to go into how to foster the social interactions and networking during an online workshop. In conclusion, more than 90% of the attendees considered the participation to this workshop worthwhile, making the EIROForum workshop 2020 a success, and a reinforcement to continue organizing this kind of workshop.

Further information about the workshop can also be found on the dedicated website at <https://indico.cern.ch/event/881752/>



2. Programme

2.1 Scientific Programme

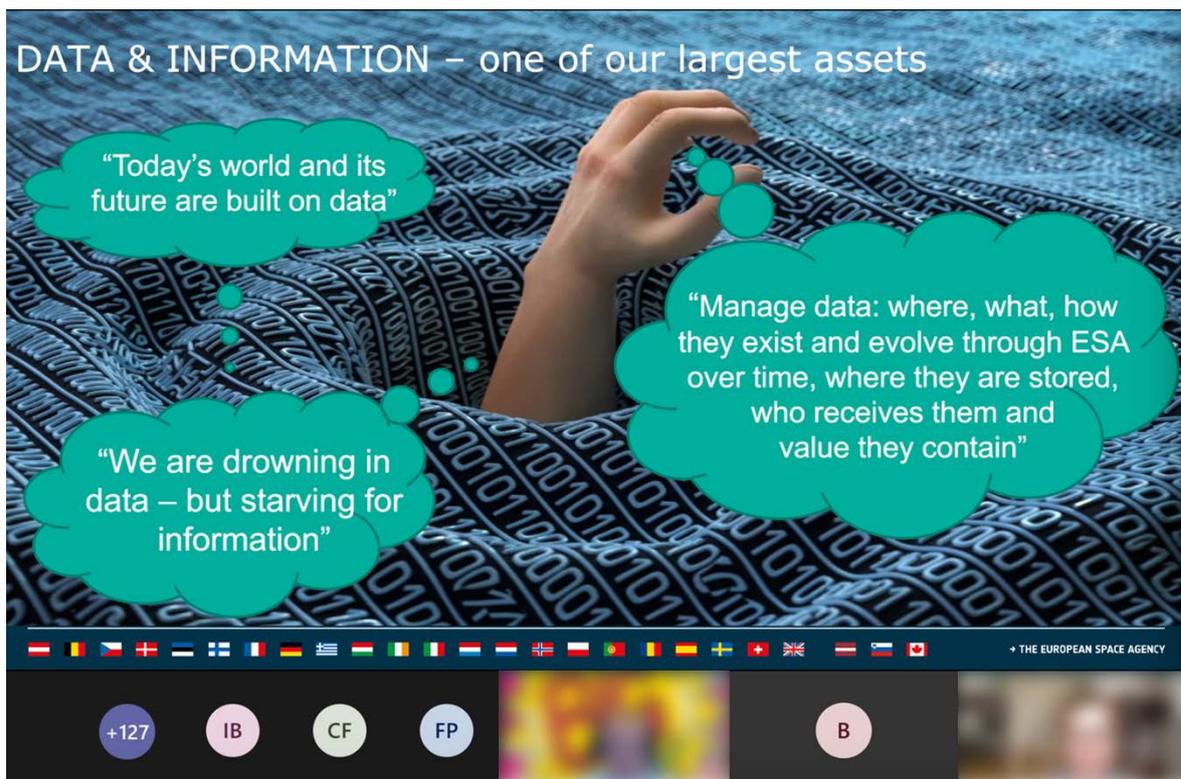
The scientific programme was organized according to the theme of the presentations into 8 sessions, covering the topics

- Data Analysis (2 sessions),
- Data Management (2 sessions),
- Machine Learning (4 sessions).

There were 2 to 3 presentations by members of every EIROForum Organization, with 30 minutes each. One presentation was typically providing an overview and the general framework of Big Data in the organization, with the other one or two presentations going into more detail and covering one specific topic.

This core programme was amended with a Special Event on Tuesday evening, and a Social Event on Wednesday evening.

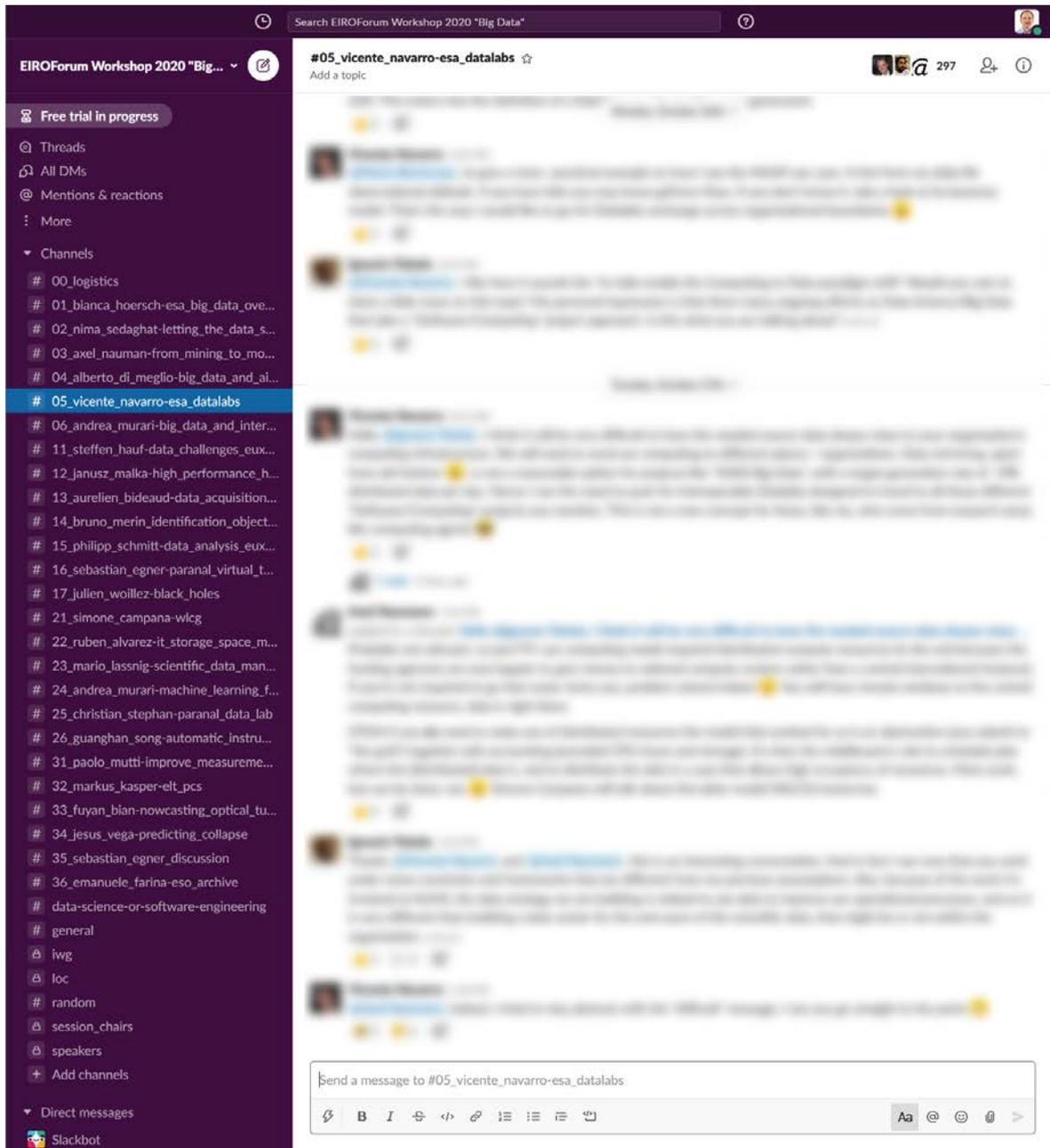
We used Microsoft Teams, hosted by ESO, for streaming the presentations. This tool worked very reliably, and could cope with more than 130 participants without issues:





To transport the interactive nature of such a workshop into the virtual world, several concepts were experimented with during the workshop, with varying success.

For asking questions to the speakers, discussing technical details or following-up topics from the presentations, a slack workspace was created. This turned out to be quite popular, with several lengthy follow-up discussions on various topics:



2.2 Social Programme

For the Virtual Coffee Breaks, we offered a virtual room in wonder.me, where colleagues could walk around their avatar and approach other colleagues. Once you are within a certain distance, the video and audio stream would start, allowing you to have a video-chat with your colleague. The concept seems to be nice and technically worked well:



For the Dinner Get-Together on Wednesday evening, we used gather.town, which is a similar app, but offers more features:



Unfortunately, the popularity of the two tools for the Virtual Coffee Breaks / Dinner Get-Together among the participants was not so great. In the feedback, the participants pointed out that there were constraints of being in the office / at home, which did not allow them to participate to the social events (e.g. family / work duties). Furthermore, the usage of a separate tool for the social events did not help to make to it a smooth experience, and it is

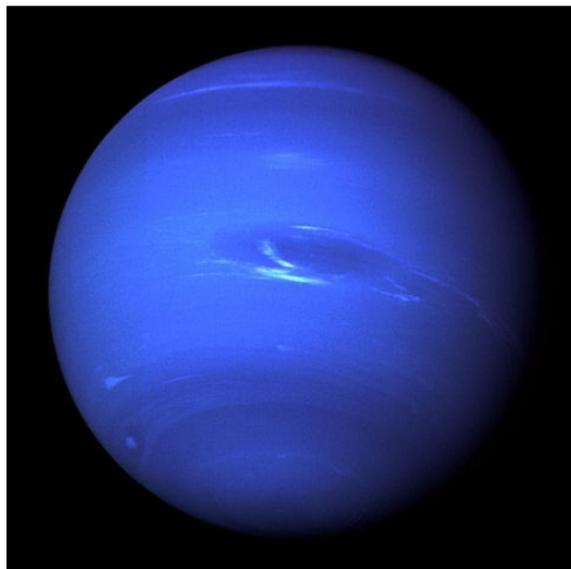
recommended to have a look at different technical solution. Several colleagues also pointed out that offering incentives to join the virtual rooms for the social events (e.g. run the quizzes in the virtual rooms / teamwork assignment) might help.

To further increase the interactivity of the workshop, Quizzes were organized during the Coffee Breaks on Monday, Tuesday and Wednesday with trivia questions related to Data Science, Internet, Computing and ESO. During each day, four attendees could win an ESO Calendar for 2021 as a prize. Typically, around half of the audience participated to the quizzes, with overall positive feedback. An example of a quiz questions is given below:

Big Data Quiz - Monday

Neptune

In 1846, Urban Le Verrier used the observations of the planet Uranus and its deviations from the expected positions to predict the presence of another planet beyond Uranus - Neptune. Le Verrier sent his prediction of the location of Neptune to Johann Gottfried Galle at the Berlin Observatory. Galle indeed discovered Neptune during the next night at only 1 degree away from the predicted position.



What amount of data did Le Verrier analyze to predict the position of Neptune (assuming the data consists of 4-Byte floats)?

- 250 Bytes
- 5 kByte
- 100 kByte
- 2.5 MByte
- 50 MByte



2.3 Detailed Agenda

The detailed agenda of the workshop is given below.

The slides of all the presentations are available online at the website of the workshop at: <https://indico.cern.ch/event/881752/timetable/#all>

Monday, 26th October

Session #1: Data Analysis

Bianca Hoersch (ESA)	ESA Big Data Overview Talk
Nima Sedaghat (ESO)	Letting the data speak for itself
Axel Naumann (CERN)	From Mining to Modeling: Data Analytics in High Energy Physics

Session #2: Data Analysis

Alberto di Meglio (CERN)	Big data and AI applications for data analysis
Vicente Navarro (ESA)	ESA Datalabs for space science and GNSS
Andrea Murari (EuroFusion)	Big data and the inference problem in thermonuclear fusion

Tuesday, 27th October

Session #3: Data Management

Steffen Hauf (EuXFEL)	Data challenges at EuXFEL
Janusz Malka (EuXFEL)	High performance and high capacity storage systems
Aurelien Bideaud (ESRF), Samuel Debionne	Development of an advanced data acquisition framework for 2D X-Ray Detectors

Session #4: Machine Learning

Bruno Merin (ESA)	Automatic identification of objects in HST images with machine learning
Philipp Schmitt (EuXFEL)	Data analysis approaches at EuXFEL

Special Event

Sebastian Egner	Virtual Tour Paranal Observatory
Julien Woillez (ESO)	Black Holes: Nobel Prize in Physics 2020



Wednesday, 28th October

Session #5: Data Management

Simone Campana (CERN)	WLCG: an infrastructure for HEP computing
Ruben Alvarez (ESA)	IT storage solutions for big data space missions
Mario Lassnig (CERN)	Scientific data management for exascale experiments with Rucio

Session #6: Machine Learning

Andrea Murari (EuroFusion)	Machine learning for the Sciences: from Causality Detection to Data Driven Theory
Christian Stephan (ESO)	The Paranal DataLab: first steps towards advanced system monitoring and maintenance schemes
Guanghan Song (ILL)	Automatic instrument configuration using deep learning methods

Social Event

Thursday, 29th October

Session #7: Machine Learning

Paolo Mutti (ILL)	Improve measurement strategy with AI: the ILL perspective
Markus Kasper (ESO)	ELT PCS R&D – AO predictive control and calibration using ML techniques
Fuyan Bian (ESO)	Nowcasting the Optical Turbulence at Paranal Observatory

Session #8: Machine Learning

Jesus Vega (EuroFusion)	Predicting collapse: adaptive and transfer learning
Emanuele Paolo Farina (ESO)	Mining the Southern Sky via the ESO Archive: Data-streams, cross matching, and forward look

Discussion & Closing remarks



3. Organisation

The main organizer and chair of workshop, Sebastian Egner, was supported by:

- A Scientific Organizing Committee (SOC) to elaborate the scientific programme of the workshop and identify the speakers,
- A Local Organizing Committee (LOC) to organize the workshop logistics.

It is worth noting here that the SOC received also support from various members of the EIROforum IT working group in identifying the workshop speakers.

3.1 Scientific Organizing Committee

The Scientific Organizing Committee (SOC) consisted of all members of the Instrumentation Working Group (IWG) of the EIROForum, particularly of:

Sebastian Egner (ESO, chair)

Christian Joram (CERN)

Thomas Schneider (EMBL)

Thibaut Prod'homme (ESA)

Jean-Claude Biasci (ESRF)

Markus Kuster (European XFEL)

Andrea Murari and Joao Figueiredo (EUROfusion)

Paolo Mutti (ILL)

3.2 Local Organizing Committee

The Local Organizing Committee (LOC) consisted of:

Sebastian Egner (ESO, chair)

Iris Bronnert (ESO)

Mylène Francois (ESO)

Martine Peltzer (ESO)



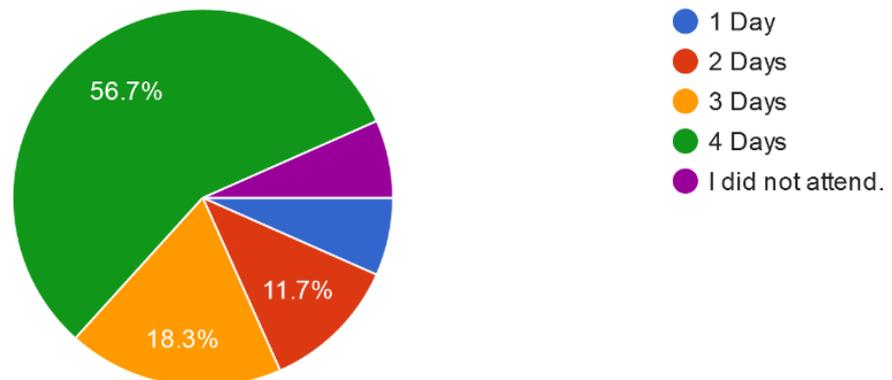
4. Feedback and Evaluation

After the end of the workshop, a questionnaire was distributed to all participants, to collect their feedback. Especially since this was the first time that we organized such an online workshop within the EIROForum and we are still experimenting with this kind of format, we considered it useful to get some insight from the audience. In total we received 60 filled responses.

Around 80% of the participants attended the workshop on 3 or 4 days. This confirms that the presentations and topics were relevant for them and the participants did not leave disappointed after the first day.

How many days did you attend the workshop?

60 responses



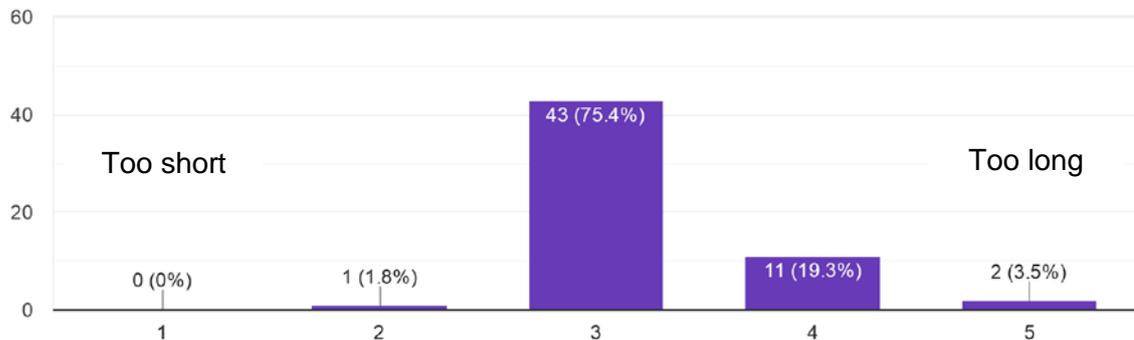
To participate to the workshop, one needed to register on the workshop website. We received around 300 registrations, from all EIROForum organizations. But not everybody, who registered, also joined the presentations. We estimate that around half of the colleagues, who registered, actually participated to the workshop.



The format of the workshop of 4 afternoons seems to be the right one for this kind of workshop, according to the vast majority of the participants. Only 20% would recommend shortening the workshop to 3 afternoons:

How did you like the format of the workshop (4 half-days, 2 sessions / day)?

57 responses

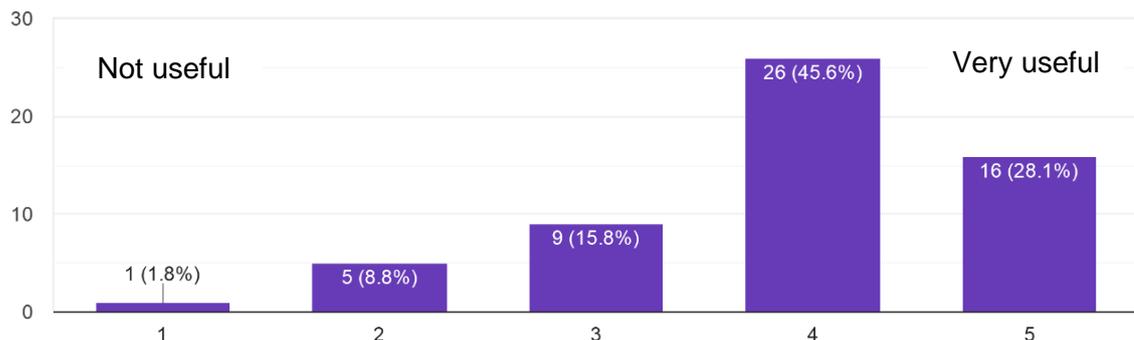


Several responses appreciated that half-days are the preferred format for such a workshop. This allows one to work in the mornings and then be able to concentrate on the workshop in the afternoons. Full-day online workshops usually do not work so well, as this collides with the other work-related duties.

Almost three-quarters (73.2%) found the workshop useful or very useful:

How useful have you found the workshop?

57 responses



Participants appreciated the variety of the covered topics and applications of Machine Learning, as well as the combination of overview talks and detailed technical talks. This gave the audience an overview of what is happening in the area of Big Data for scientific applications, outside of the direct area of work.

Several responses praised the high quality of the speakers and their presentations. The presentation by Andrea Murari was mentioned as especially interesting, as it provided a general and more abstract introduction into the topic of Machine Learning. More talks in this direction, with an introduction to the topic for newbies would have been appreciated by the participants.



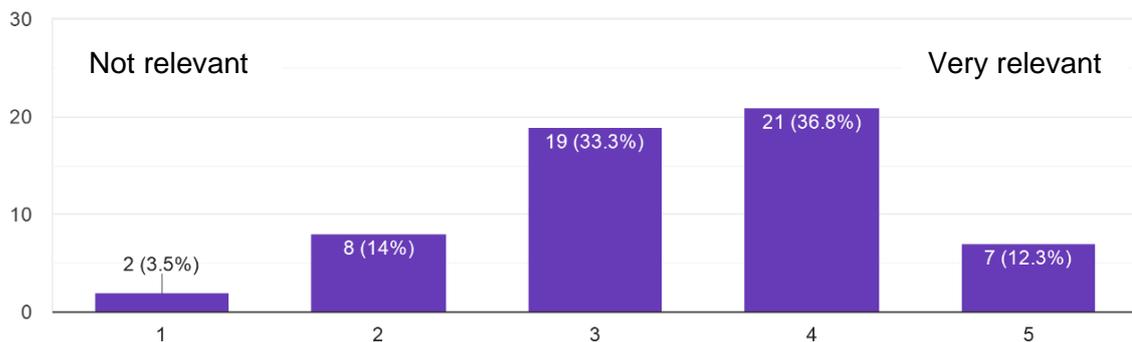
Some participants recommend adding one general keynote presentation about Big Data at the beginning of the workshop. Furthermore, having the slides available immediately after the sessions would have been appreciated, as well as the availability of the recorded talks.

Despite the usefulness of the workshop, the presentations provided more a general overview of the topic and the concepts of Big Data, than to have a direct relevance for the day-to-day work of the participants. This was also highlighted by several comments, saying that it was useful to get an overview, and an excellent opportunity to learn new things, but due to the very different topics that the various organizations are working on, the overlap in the actual topics is rather limited.

To improve on this aspect, some of the participants would have preferred to have more technical presentation on machine learning and more details on the actual applications. Others suggested to have a more interactive session, like a small hands-on project, or a special session with shorter presentations by several colleagues. Also a more varied range of topics, such as Research of new AI/ML algorithm, Applications of existing AI/ML libraries or algorithms, Big data in Operations, and Big data in Science Research, might have helped on that aspect.

How relevant were the presentations for your area of work?

57 responses

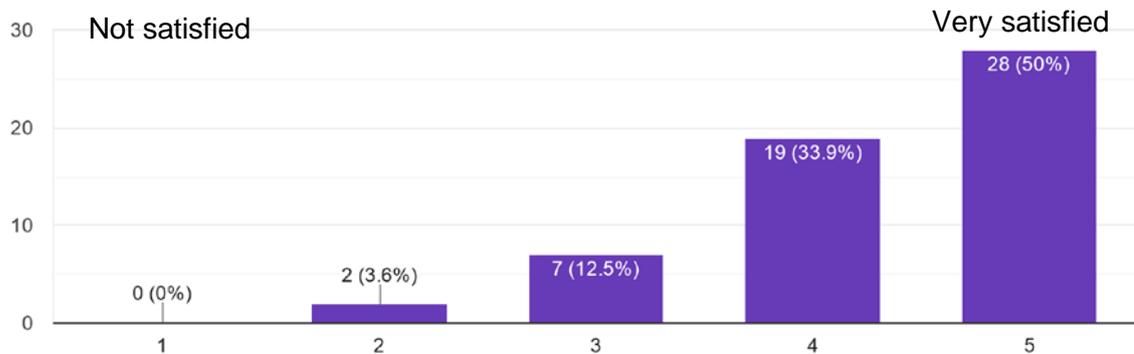




On the technical side, it seems that Microsoft Teams worked very well, more than 80% of the participants were satisfied with the quality of the video-feed:

How satisfied were you with the quality of the streaming of the presentations via MS Teams?

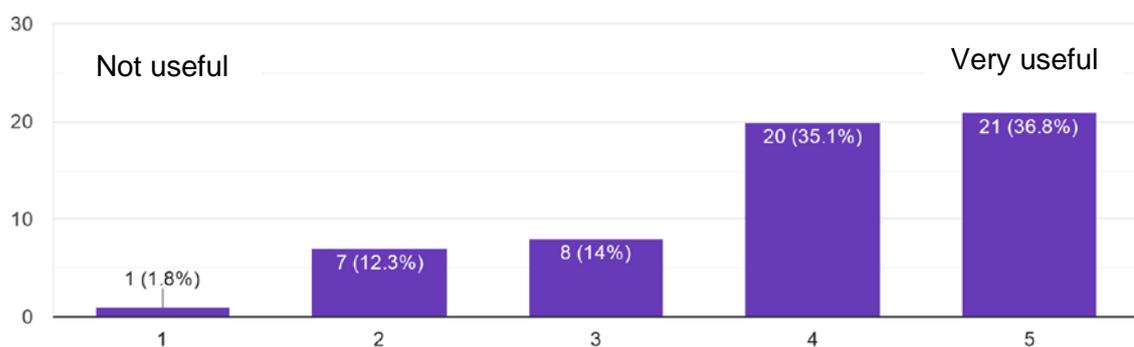
56 responses



Around 70% of the participants found the usage of slack a good idea for asking questions to the speaker and to continue the discussion of the topics afterwards. However, there were also several comments, that a more direct interaction with the speakers would have been useful in order to iterate on the questions. Furthermore, a more interactive approach (e.g. to ask the questions not at the end of the presentation, but on specific slides during the presentation) would have been something worthwhile to try with an online workshop.

How useful do you think is slack for asking questions / interacting with the speakers?

57 responses

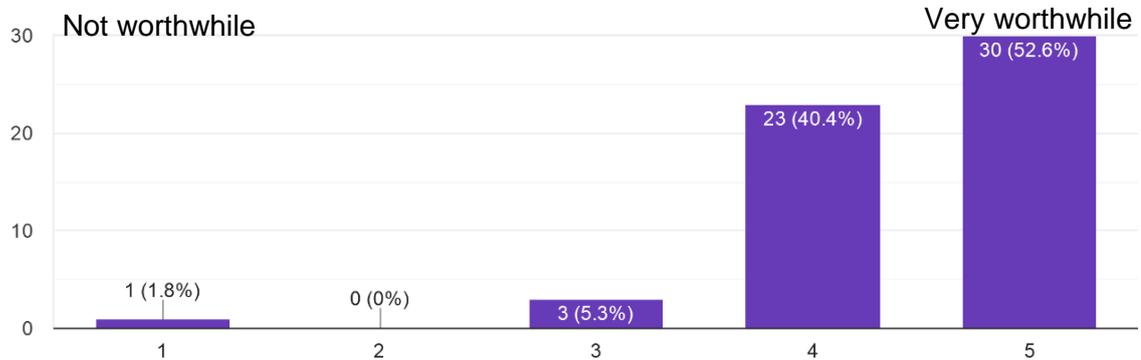




Overall, more than 90% of the participants found the workshop worthwhile to attend:

Overall, was it worthwhile to participate to this workshop?

57 responses

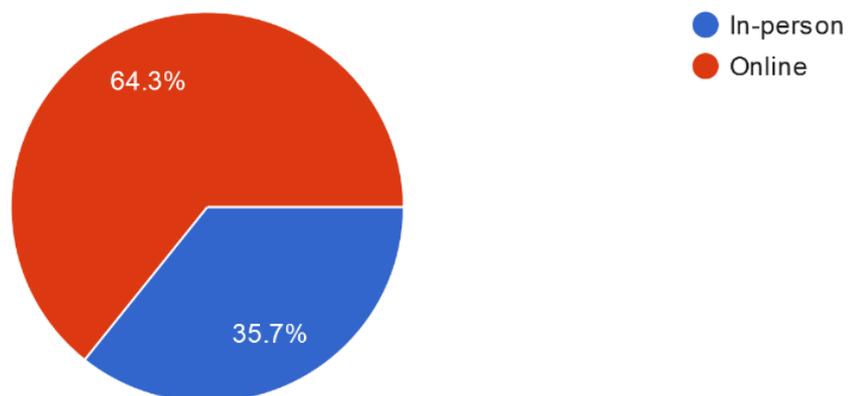


Probably this is well summarized by one feedback by an attendee: *“This is the first virtual workshop that went very well for me. The organization was super and the idea of a quiz was great. Please, continue in this way!”*

The final interesting discovery is that almost 2/3 of the participants would prefer to have the next workshop again as an online event:

Would you prefer to have the next workshop in-person? Or again online?

56 responses





5. Lessons Learnt and Recommendations

5.1 Online Workshops

It is almost certain that this format of online workshops will remain in the future. The good news is that it seems to work very well in practice. The tools are now all available and working reliably. It takes a bit of time to get used to the format, get to know the functionalities and limitations of the various tools and the differences to in-person events. Also, some fine-tuning and optimization of some aspects will certainly still be needed. But generally, the experience of this first online workshop was overly positive and the feedback by the attendees suggest that it was worthwhile to organize it.

We have some concerns about hybrid workshops though, where part of the audience will participate in-person, while the other part will join remotely. Firstly, this will require almost double the organization effort by the host. Secondly, the experience for the two groups of attendees will be very much different, it is technically difficult to have a good audio / video signal to both groups, and interaction with the remote participants is very difficult. We tried a hybrid workshop in the recent past at ESO and it did not really work well, the remote participants mentioned the limited user experience dropped out one after the other.

5.2 Preparation

A good preparation proved to be essential for a successful online workshop, even more than for an in-person event:

- Test the various tools in the same configuration as later during the workshop.
- Organize short test-sessions with all the speakers, to ensure that they know the tool and can share their presentations. We discovered several limitations by the various operating systems and web-browsers used by the speakers. With the support of IT, we could then resolve these issues before the actual workshop.
- Get support by IT, as tools have settings and configurations, which not accessible to the normal user. Furthermore, the colleagues from IT usually know the quirks of the tools better.
- Inform the audience with the practicalities of the various tools (e.g. what is the purpose of slack and how to use it, how to join the workshop, what are the social activities, etc.).
- Start early (>1 month) with playing with the tools. We discovered that some originally planned tools would not work, or that feedback by the vendors can be slow, or certain vendors do not comply with the procurement rules. In either case, you will need time to look for alternatives.

5.3 Engagement

It turned out, that the audience appreciated a lot to have the opportunity for small engagements and interactions with other participants. This re-created a bit the atmosphere of an in-person workshop and it felt less like watching a recorded video. Options include:



- A slack workspace with the possibility to ask questions to the speakers and have follow-up discussions. Another option to investigate is to use slido, which allows a more direct interaction with the speaker, also during the presentation. The current limitations are that a feedback on the answer to the question is hardly possible.
- The trivia quiz during the coffee breaks was quite popular. Other ideas could be to have short surveys during the presentations, or even some hands-on activity during the workshop.
- A social event with a more general presentation, relevant for the technical audience, but not down to the details was well appreciated. Similarly, the slideshow of the virtual tour was very popular. It created a relaxing atmosphere and raised some topics to discuss about during the following coffee breaks.
- The discussion at the end of the workshop was also appreciated by many attendees. However, it might be needed to split this into several topical discussions, to allow more colleagues to actively participate.
- The attendees would have liked to applaud to the speaker at the end of their presentations. As well, the speakers found it a bit awkward, to have no feedback from the audience at the end of the presentation.

5.4 Social Interaction

An important aspect of such a workshop are the social interactions between the attendees. One purpose of organizing such workshops is to get to know each other and establish a network across the organizations with colleagues, who are working on similar topics. As mentioned above, this still needs some refinement. Proposal for improvement could be:

- A more integrated tool (presentations and virtual coffee breaks in the same tool).
- Offer incentives to join the virtual coffee breaks (e.g. quiz, hand-on, dedicated discussions, etc.).

But in the end, it will be difficult to merge social interaction among the workshop participants with the duties of work or family.

5.5 Topics for future workshops

The following is a list of topics, which are getting increasingly important for instrumentation projects, but are typically not addressed properly. They could thus be good subjects for future workshops:

- Scientific software development by researchers and not by software engineers.
- Knowledge management within an organization.
- Quality, configuration, and documentation management.
- Efficient remote collaboration.
- Remote control of experiments.
- Popularizing instrumentation: speaking and writing about instrumentation to the outside world.



- Modelling experiments: instrument performance verification, scientific results validation etc.

———— **End of document** ————