

---

# DA newsletter

Mar 02, 2021



<b>1</b>	<b>July 2020</b>	<b>3</b>
1.1	EXtra-data 1.2 . . . . .	3
1.2	Oscovida . . . . .	4
1.3	<i>Did you know...</i> the data analysis documentation? . . . . .	5
<b>2</b>	<b>August 2020</b>	<b>7</b>
2.1	EXtra-foam 1.0 . . . . .	7
2.2	<i>Did you know...</i> that JupyterHub can be used on the online cluster? . . . . .	8
2.3	<i>Did you know...</i> that there are control networks PCs outside the hutches? . . . . .	8



Released on a monthly basis, the data analysis newsletter reports on a mixture of topics surrounding analyzing scientific data at European XFEL. These range from recent release highlights of both in-house and community developed software, experience reports from users and collaboration projects as well as general tips and tricks when working with our infrastructure.

For further information about many of the covered topics or if you have any questions, please consider looking at the documentation for [data analysis at European XFEL](#) or email us directly at [da-support@xfel.eu](mailto:da-support@xfel.eu).



## 1.1 EXtra-data 1.2

EXtra-data release 1.2 is out and comes with many new tools, features and improvements! We present a few of them in this letter. The full list can be seen [here](#).

### 1.1.1 lsxfel

lsxfel has a new `--detail` option to show more detail on selected sources.

Example, checking if instrument sources have missing data:

```
% lsxfel --detail "*" /gpfs/xfel/exp/SQS/201901/p002388/raw/r0138 | grep % -B 2
- SA3_XTD10_XGM/XGM/DOOCS:output
  - data:
    data for 6474 trains (100.00%), up to 1 entries per train
--
- SQS_AQS_LAS/SPEC/800:output
  - data:
    data for 5645 trains (87.19%), up to 1 entries per train
--
- SQS_DAQ_SCAN/MDL/KARABACON:output
  - data:
    data for 21 trains (0.32%), up to 1 entries per train
--
- SQS_DIAG3_PAM/CAL/SPEC_REF:daqOutput
  - data:
    data for 0 trains (0.00%), up to 0 entries per train
```

### 1.1.2 extra-data-validate

extra-data-validate is significantly faster, and validating a run with several terabytes of data now takes only a few

seconds.

Fig. 1: Validate run data integrity

### 1.1.3 extra-data-make-virtual-cxi

`extra-data-make-virtual-cxi` has a new `--fill-value` option to set the default value for missing data.

### 1.1.4 extra-data-locality

`extra-data-locality` is a new command line tool to check how the files are stored.

Reading files may hang for a long time if files are unavailable or require staging in dCache from the tape. The program helps finding problematic files. If it finds problems with the data locality, the program will produce a list of files located on tape, lost or at unknown locality.

```
% extra-data-locality /gpfs/xfel/exp/XMPL/201750/p700000/raw/r0002
Checking 120 files in /gpfs/xfel/exp/XMPL/201750/p700000/raw/r0002
120 on disk, 0 only on tape, 0 unavailable
```

### 1.1.5 karabo-bridge-serve-files

`karabo-bridge-serve-files` has a new `--append-detector-modules` option to combine data from multiple detector modules. This makes streaming large detector data more similar to the live data streams.

---

**Note:** Documentation of all EXtra-data command line tools is available [here](#)

---

### 1.1.6 write\_frames

EXtra-data detector components have a new `write_frames` method to write selected detector frames to a new EuXFEL HDF5 file.

## 1.2 Oscovida

The Data Analysis team and members of the PaNOSC project have been working on a COVID19 tracking website as part of efforts to improve the public awareness and understanding of the current coronavirus situation.

The Open Science COVID19 Analysis (OSCODIA) package is used to programatically create a website that provides COVID19 tracking information, which can be used by scientists and the wider public to understand spread of the global pandemic, and in some regions, where the first wave has passed, to detect, monitor and study localised or wider renewed outbreaks. The availability of the reproduction number R and its behaviour as a function of time helps to put the discussions in politics and media into context, the change in daily deaths and cases can help motivate measures such as social distancing or easing of restrictions. You can view the tracking page (updated twice a day) here: <https://oscovida.github.io>

For example, this is the [Hamburg tracking page](#), each tracking page has the same set of plots and the meaning of the plots is explained in our 'standard plots' educational page. As the analysis pages themselves are generated from



Jupyter Notebooks, we provide a link on each page to easily open the notebook in an interactive MyBinder instance which allows users to explore the data themselves however they want.

Currently tracking pages are provided for: all countries, German Landkreise, American states, and Hungarian counties. If you would like more detailed regions for your country to be included on the website please [create an issue on the Oscovida GitHub repository](#).

### 1.3 *Did you know...* the data analysis documentation?

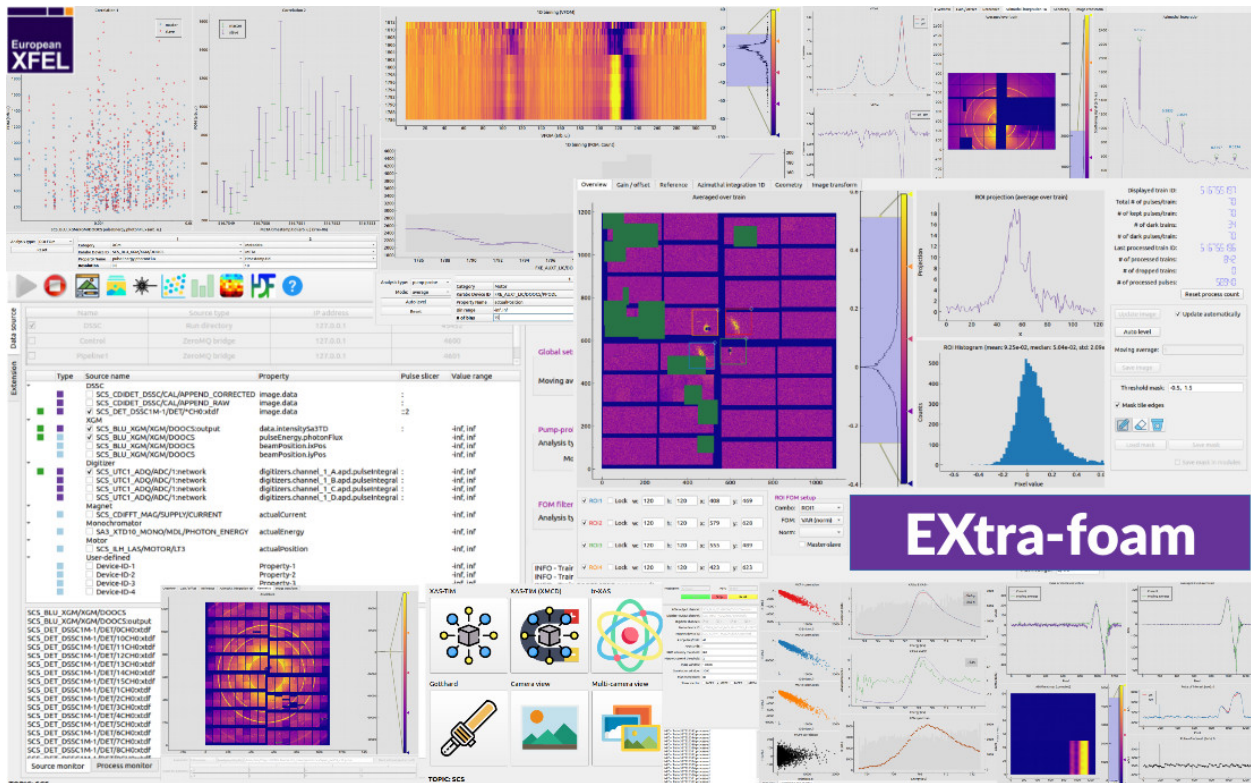
Did you know that the data analysis group maintains an extensive documentation surrounding many aspects of analyzing data at European XFEL? The latest version can always be found here: [Data analysis at European XFEL](#)

It covers a wide array of topics from the internal data structure of files to the tools and services readily available in our infrastructure down to specific deployments provided at each instrument. It is continuously updated and extended, and if you notice anything missing or have further questions, please don't hesitate to email us at [da-support@xfel.eu](mailto:da-support@xfel.eu). You can always find the link on the front page of this newsletter, too.



## 2.1 EXtra-foam 1.0

EXtra-foam release 1.0 is out and comes with many new features and improvements! The full list can be seen [here](#).



Should you have any questions, please feel free to send an email to [da-support@xfel.eu](mailto:da-support@xfel.eu) or open an issue on the [EXtra-foam github repository](#).

## 2.2 *Did you know...* that JupyterHub can be used on the online cluster?

Did you know that every node reserved for a particular instrument (e.g., *sal-onc-fxe* for *FXE*, see [all available nodes](#)) runs a [Jupyter Hub](#) instance? They can be used during experiments to run local scripts and/or to check the data quality, reducing the amount of data that has to be transferred and stored offline. Remember that the offline cluster (i.e. Maxwell) still provides many more resources for complex analyses, including the exclusive node reservations available for each proposal during the experiments. You can find the details, as well as further available options, here: [JupyterHub on the EuXFEL online cluster](#)

## 2.3 *Did you know...* that there are control networks PCs outside the hutches?

Did you you know that in the so-called “green-corners” of XHQ there are several computers that have access to the online clusters? You can find more information and the exact location of those workstations here: [Access to the online cluster](#)