### IMPROVING DATA DISCOVERY, ANALYSIS, AND VISUALIZATIONS WITH CLOUD-BASED USER SERVICES

*Geoffrey T. Stano1, Navaneeth Selvaraj1, Alan Subedi1, Sara Graves1, Aaron Kaulfus2*

1University of Alabama in Huntsville, Huntsville, Alabama, United States

2NASA Marshall Space Flight Center, Huntsville, Alabama, United States

#### ABSTRACT

The Global Hydrometeorology Resource Center (GHRC) Distributed Active Archive Center (DAAC) is one of 12 DAACs managed by the United States National Aeronautics and Space Administration (NASA) Earth Science Data and Information System (ESDIS) project [1]. GHRC and the other DAACs are designed to process, archive, document, and distribute NASA Earth-observing data, ranging from satellite missions to field campaigns [2]. A major goal of the DAACs is to enable science with these data. Science enabling can be difficult as datasets can be very large, use multiple formats, come from numerous platforms, and require three-dimensional visualization. GHRC is using its expertise with cloud-based technologies to develop open source and open science tools to empower users to explore, coincidentally visualize, and analyze multiple datasets. Being open source, the user community can develop visualizations for their own datasets. This presentation will expand on this objective and highlight the capabilities available to the international community now.

***Index Terms—*** GHRC, DAAC, FCX, user services

**1. INTRODUCTION**

The role of the United States National Aeronautics and Space Administration (NASA) Distributed Active Archive Centers (DAACs) is to serve as public repositories of NASA’s Earth science data. This means that the data available at the DAACs is free of charge and open to users around the globe. Each of NASA’s 12 DAACs have specific areas of expertise and discipline communities. For the Global Hydrometeorology Resource Center (GHRC) DAAC, the emphasis is on storm hazards, lightning, precipitation, convective and tropical systems, and the underlying dynamical and physical processes. Furthermore, GHRC specializes in field campaign observations and cloud-based technologies.

The mission of GHRC and the other DAACs has been expanding. The aim is to serve as more than just curators of data. The DAACs are continuing to explore new ways to enable science to be conducted with the available data. A significant component of this has been open science.

The year 2023 was declared the year of open science by NASA. The principles of open science are, “to make publicly funded scientific research transparent, inclusive, accessible, and reproducible. Advances in technology, including collaborative tools and cloud computing, help enable open-source science” [3]. As a DAAC, GHRC’s core mission aligns with open science, making Earth science data freely available. However, GHRC has leveraged its expertise in data visualization and cloud technologies to improve accessibility and use of GHRC’s holdings.

One of the largest problems for users who wish to utilize DAAC data is easily visualizing data. As an example, GHRC hosts hundreds of datasets from over two dozen field campaigns. Field campaigns provide unique research opportunities as they bring together a wide array of observational systems to investigate specific phenomena. These data include ground-, airborne-, and satellite-based observations. These can range from point observations, gridded model data, to fully three-dimensional datasets. Being able to visualize just one dataset can be extremely time consuming, particularly for users who may be unfamiliar with the dataset or are developing computer programming skills. The issue is further complicated by the large volume of many datasets. Large data volumes can be unfeasible to download the data.

Given these issues, GHRC has been developing open source, cloud-based systems to visualize and analyze datasets. As a cloud-based system, users do not need to download software or data to their local computer. GHRC provides the platform so that users can visualize multiple, three-dimensional datasets simultaneously. Additionally, these tools have been made open source. This enables users to build on GHRC’s capabilities for their own data.

The next sections will describe GHRC’s Field Campaign Explorer (FCX) [4]. This will include a brief technical description. Examples of FCX’s abilities and future efforts will also be described.

**2. FCX TECHNICAL DETAILS**

FCX has been specifically built to be a cloud-based, serverless system. This provides advantages in both the short- and long-term time frames. FCX to takes advantage of GHRC being the first NASA DAAC to have transitioned all of its holdings into the cloud. As a cloud-based system that can access GHRC’s cloud archive, FCX requires no downloads by end users. GHRC uses Amazon Web Service (AWS).

The process involves downloading raw field campaign data from Earthdata Search and uploading it to the AWS Simple Storage Service (S3). A Python script generates various cloud-native format files (cloud-optimized GeoTIFF or COG, Zarr, and 3D Tiles) for visualization. These files undergo conversion from raw formats (netCDF, HDF, ASCII) to Cesium visualization format (Point Clouds, Web Map tiles, 3D Tiles), subsequently stored in S3 buckets. A Web Map Tiling server (Terracotta) is deployed to serve COGs, and FCX metadata is stored in S3 buckets (with ongoing conversion to a GraphQL Application Programming Interface or API). The application's frontend, a React-based FCX-WebApp, is deployed in the AWS environment, making the entire system a native cloud-based application. The FCX-WebApp connects to the deployed APIs to visualize the field campaign data.

GHRC has also focused on making FCX open source, to the code for any user to create a separate copy of FCX or for users to create their own visualization APIs. User created APIs is a significant advantage for FCX. Instead of users waiting for GHRC to add a particular dataset, end users can develop their own visualizations and analyses that are derived from the FCX framework. GHRC has primarily focused FCX on field campaign datasets within its archive. The open source nature enables FCX to be extended well beyond this initial capability.

The open source code can be found in several ways. First, GHRC has established an FCX “playground”. This is an isolated software environment where users can experiment with and prototype new code and capabilities without affecting the operational version of the system. This can be found at:

* <https://ghrc.earthdata.nasa.gov/fcx-playground>

Specific code for the FCX front end and back end playground is available at:

* <https://github.com/ghrcdaac/fcx-playground-frontend>
* <https://github.com/ghrcdaac/fcx-playground-backend>

This has also been published to the Python Package Index: <https://pypi.org/user/ghrc/>.

**3. DEMONSTRATION OF CAPABILITIES**

FCX is a three-dimensional data exploration tool. Currently, data visualization is FCX’s primary capability. However, users can access associated documentation for each dataset as well as the parent field campaign. FCX provides the link to the datasets in Earthdata Search as well so that users may download the data if desired. A recently added subsetting tool enhances this as users may specify a specific time range. FCX then packages only the data within this timeframe for users to download.

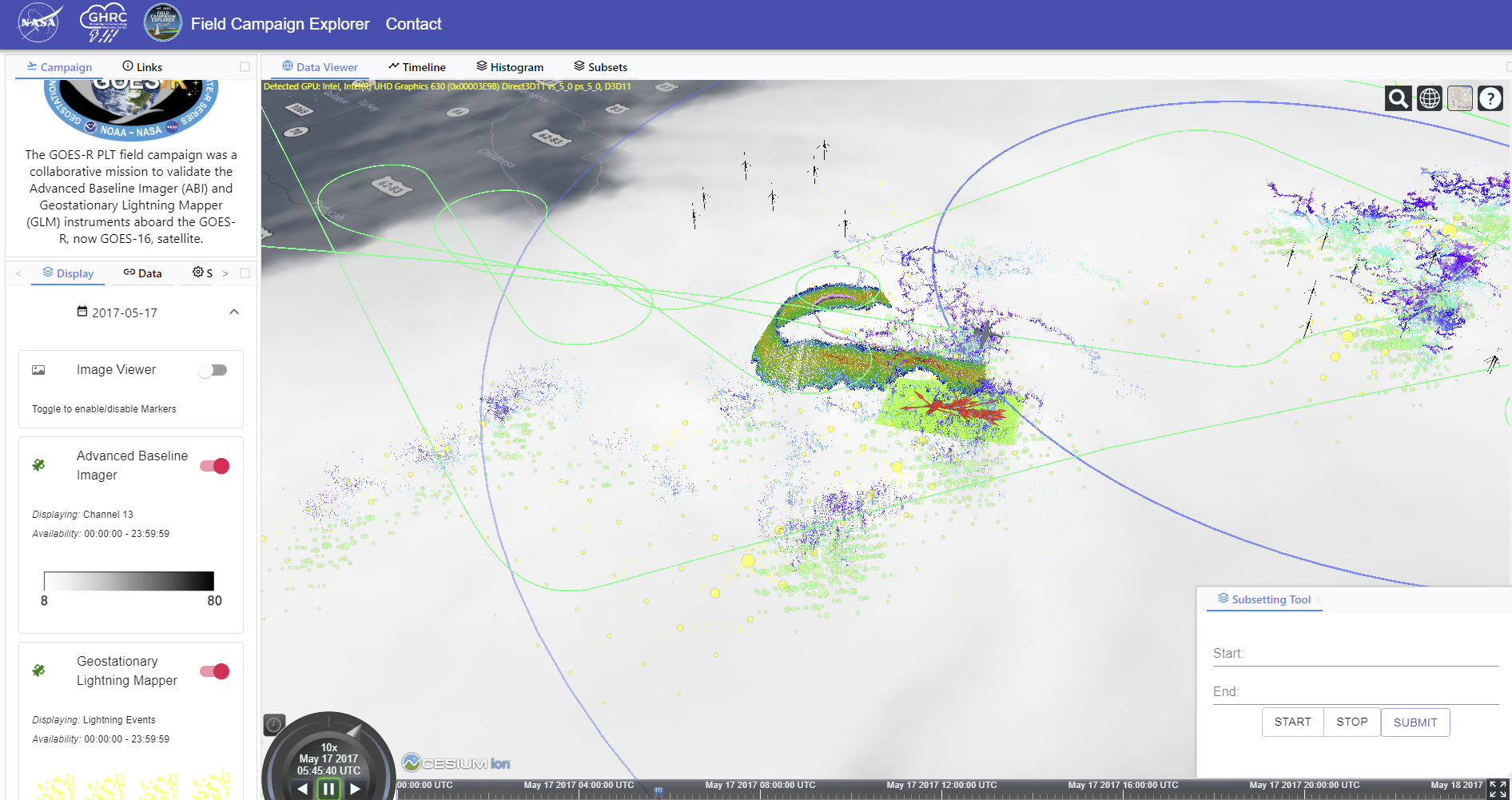
The remainder of this section will demonstrate some of the key capabilities of FCX. The link to access FCX is:

* <https://ghrc.earthdata.nasa.gov/fcx/index.html>

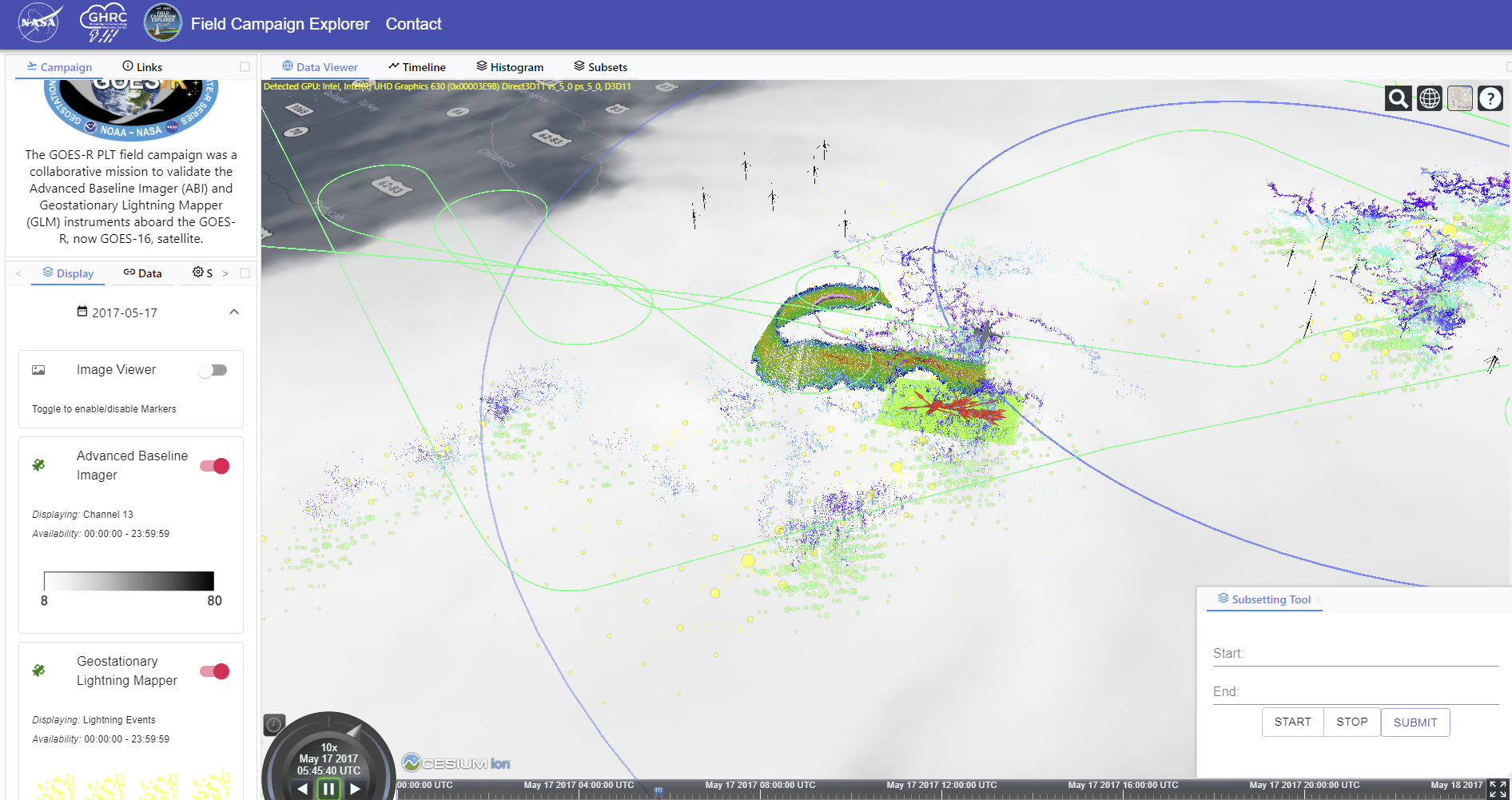
Figure 1 shows a still image within FCX taken from the GOES-R Post Launch Test field campaign that took place from April-May 2017. As shown in the upper left of the image, the campaign was a collaborative mission to validate the Advanced Baseline Imager and Geostationary Lightning Mapper instruments aboard the GOES-R (now GOES-16) satellite. This campaign combined ground- and airborne-based instruments to calibrate and validate the two satellite instruments.

The image shows eight coincident datasets. This includes the Advanced Baseline Imager, Geostationary Lightning Mapper, International Space Station Lightning Imaging Sensor, Cloud Radar System, Cloud Physics LIDAR, Fly’s Eye GLM Event Simulator, Lightning Instrument Package, and the Oklahoma lightning mapping array. These are supported by three supplemental displays; ground station locations and range rings for the Oklahoma lightning mapping array and the flight track of the ER-2 research aircraft.

Figure 2 is the same as Figure 1 but highlights several of the tools available in FCX. The data navigation section (left side in solid black box) provides users options to select a specific data or to toggle datasets on or off. The “Display” and “Data” links allow users to toggle between the visualization shown or to access the links to each individual dataset. The large, dashed box shows the time feature. Users can use a slider to select a specific time. To the left is a display providing the current date and time of the displayed data and gives the playback tool. This can go forwards or backwards, pause, or speed up / slow down. The small, dashed box at top provides options for the Data Viewer (what is shown) as well as the Timeline, Histogram, and Subsets, which will be shown in other figures. Lastly, the dotted black box highlights the subsetting tool that selects the start and stop time of the subset.

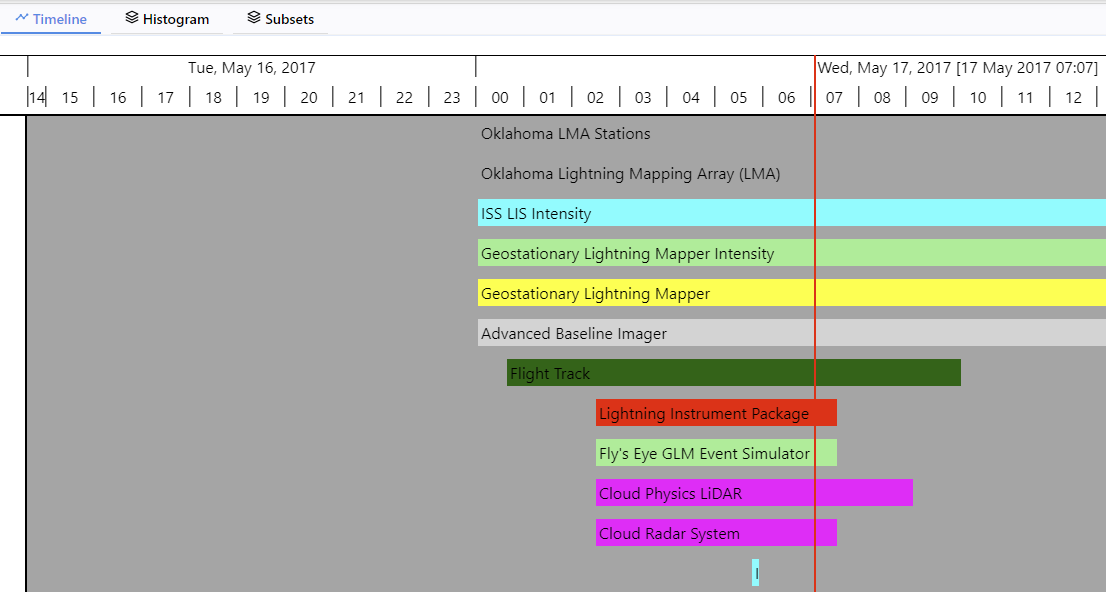


*Figure 1: Still image from the GOES-R Post Launch Test field campaign at 05:45:40 UTC on May 17, 2017. This shows the user interface for FCX as well as eight coincident datasets. The Advanced Baseline Imager (1), Geostationary Lightning Mapper (2), International Space Station Lightning Imaging Sensor (3), Cloud Radar System (4), Cloud Physics LIDAR (5), Fly’s Eye GLM Event Simulator (6), Lightning Instrument Package (7), and the Oklahoma lightning mapping array (8). Also shown are the Oklahoma lightning mapping array sensor location (A) and range rings (B) and the ER-2 flight track (C).*

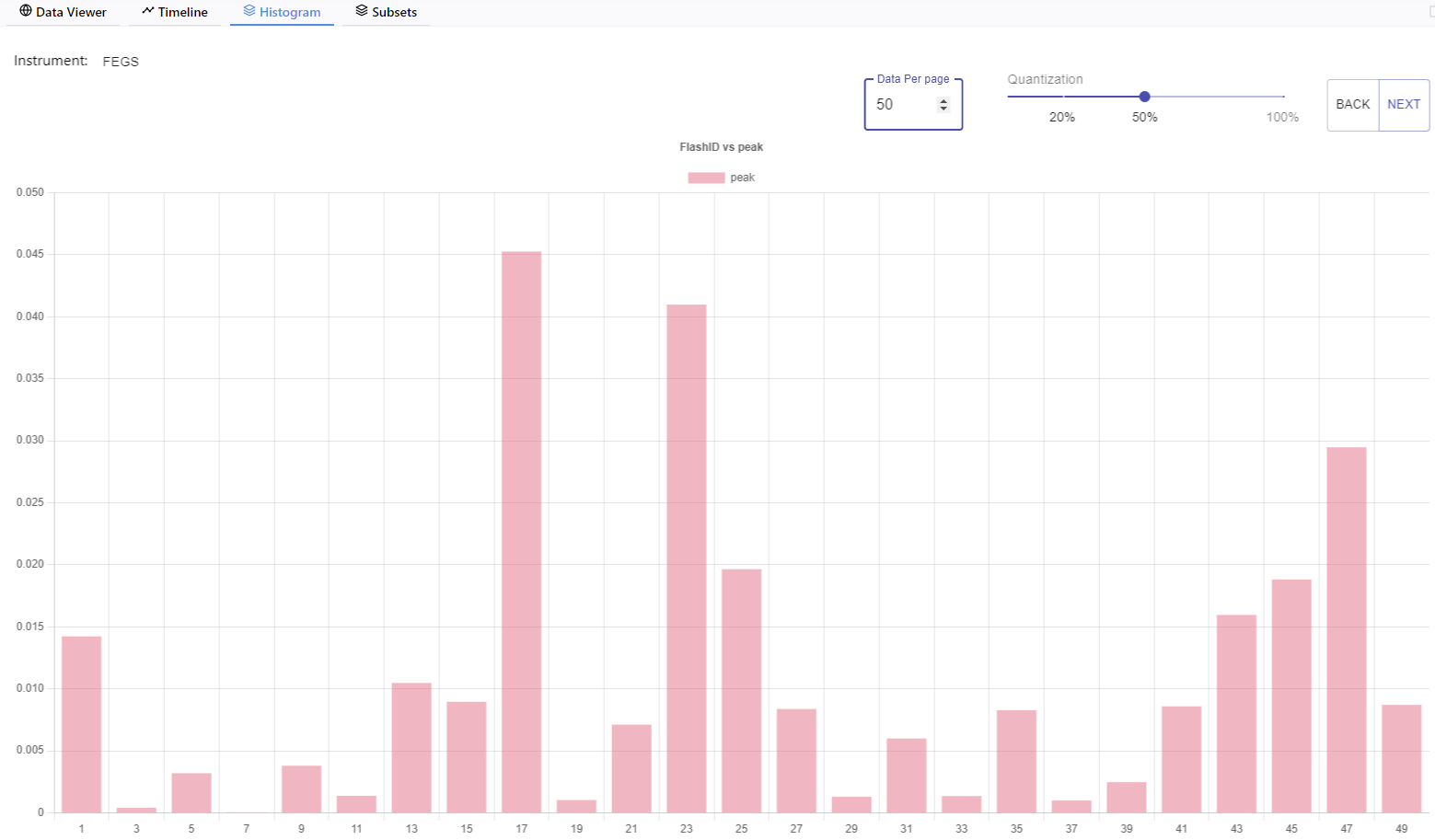


*Figure 2: The same as Figure 1, but this emphasizes the data navigation (solid box, left), the time feature (large, dashed box, bottom), other FCX features (small, dashed box, top), and the subsetting time selection (dotted box, right).*

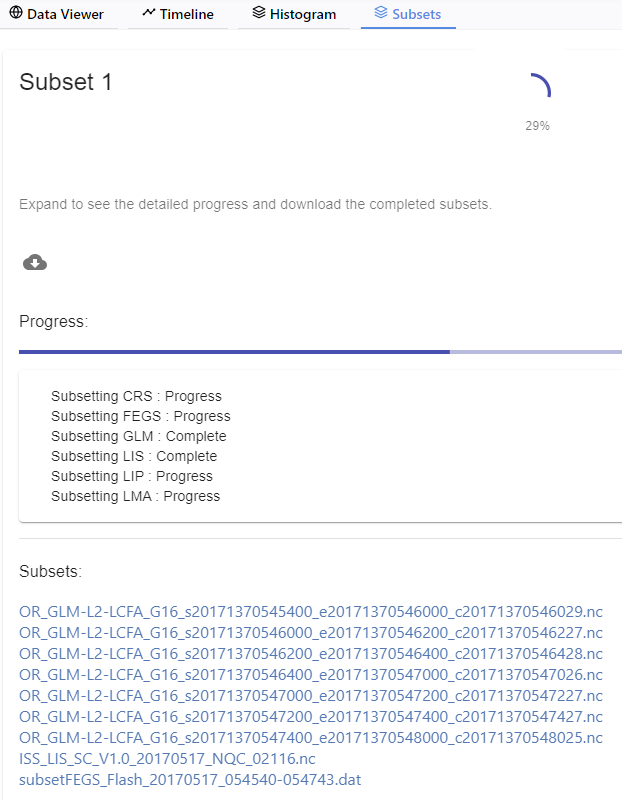
Figure 3 shows the timeline viewer highlighted in the small, dashed box in Figure 2. Viewing data concurrently is a necessary component of FCX. However, airborne- and satellite-based datasets do not provide continuous coverage in a specific location. It is vital to know when data are available. The timeline viewer example in Figure 3 shows when all of the datasets from May 17, 2017, were available. This is very useful to identify when the International Space Station Lightning Imaging Sensor was available as it has the smallest temporal availability.



*Figure 3: The timeline view showing the time span with which each dataset from May 17, 2017, was available.*



*Figure 4: An example of the histogram tool showing the intensity of the first 50 lightning flashes detected from the ER-2 versus the Geostationary Lightning Mapper.*



*Figure 5: Example of the subsetting tool getting data from May 17, 2017, at 05:45:40 UTC through 05:47:43 UTC.*

Figure 4 is an example of the first analysis tool incorporated into FCX. This is currently in demonstration mode. In this example, the first 50 flashes detected by the airborne Fly’s Eye GLM Event Simulator have their peak intensity compared against the corresponding Geostationary Lightning Mapper event intensities.

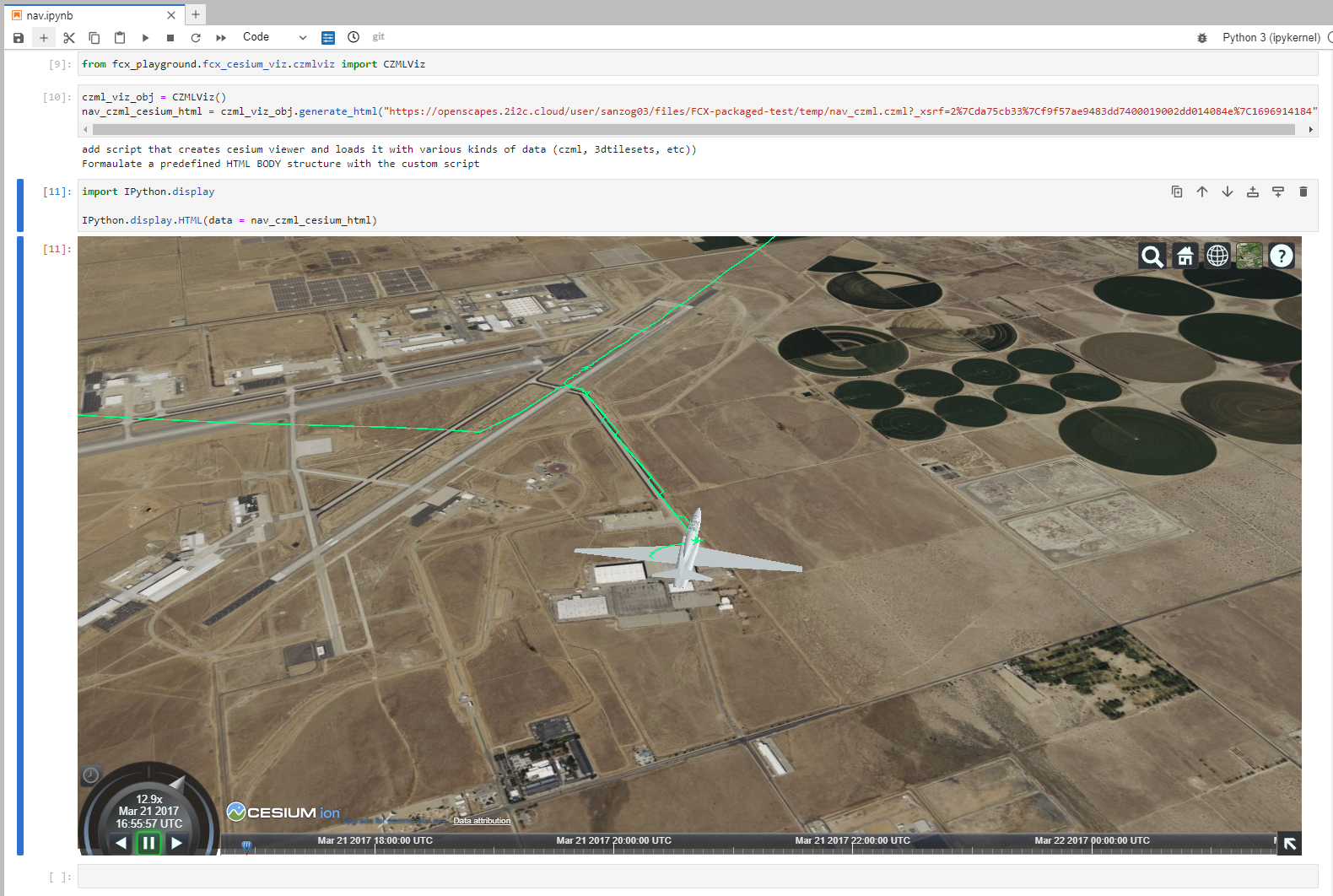
Lastly, Figure 5 provides an example of the subsetting tool. In the example, data from 05:45:40 UTC to 05:47:43 UTC on May 17, 2017, were selected. The image shows the status display FCX provides as well as a percentage complete in the upper right corner. At the bottom, the subset files are provided. Any files that required temporal subsetting have “subset” appended to the front of the filename.

**4. FUTURE WORK**

GHRC will continue to enhance the Field Campaign Explorer (FCX) as the DAAC continues to develop capabilities to support science enabling within our user community. As GHRC moves forward with this, the goal will be to bring more of GHRC’s field campaign data into FCX. This is also extended to future missions that GHRC will support in the coming years, such as the Investigation of Convective Updrafts mission [5]. Discussions with our user community and user working group will identify datasets to prioritize as well as new analysis tools.

GHRC also aims to encourage dataset visualization APIs developed by the user community. The first step in this has been to make FCX open source. However, users require more than just the code. In the year ahead, GHRC plans to develop a basic training module to show users how to interact with the FCX playground and how to integrate a new API.

Lastly, FCX offers the ability to create additional tools. A significant amount of effort was spent to create FCX’s visualizations. The code can be extracted and incorporated into jupyter notebooks. These cloud-based notebooks offer users the ability to re-create FCX’s visualizations for individual datasets to create still images and animations. Furthermore, GHRC is incorporating NASA’s Openscapes [6] that is providing a community python library. Users will not have to download or install these libraries. GHRC will collaborate with its users over the next year to select high-value datasets to be supported with these jupyter notebooks.



*Figure 6: Screen capture of the ER-2 navigation animation from OLYMPEX using the jupyter notebook data recipe.*

**5. REFERENCES**

[1] Earth Science Data and Information System (ESDIS) Project, (2020). <https://earthdata.nasa.gov/esdis>.

[2] EOSDIS Distributed Active Archive Centers (DAACs), (2021). <https://earthdata.nasa.gov/eosdis/daacs>.

[3] Open Source Science Initiative, (2023). <https://science.nasa.gov/researchers/open-science/>.

[4] Stano, G. T., Y. Wu, N. R. Selvaraj, M. Maskey and A. Kulkarni, "The Field Campaign Explorer," 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 2021, pp. 3257-3260, doi: 10.1109/IGARSS47720.2021.9553677.

[5] van den Heever, S., Haddad, Z., Tanelli, S., Stephens, G., Posselt, D., Kim, Y., Brown, S., Braun, S., Grant, L., Kollias, P., Luo, Z. J., Mace, G., Marinescu, P., Padmanabhan, S., Partain, P., Petersent, W., Prasanth, S., Rasmussen, K., Reising, S., and Schumacher, C. and the INCUS Mission team: The INCUS Mission, EGU General Assembly 2022, Vienna, Austria, 23–27 May 2022, EGU22-9021, <https://doi.org/10.5194/egusphere-egu22-9021>, 2022.

[7] Robinson, Erin, and Julia S. Stewart Lowndes. "The Openscapes Flywheel: A framework for managers to facilitate and scale inclusive Open science practices." (2022). <https://doi.org/10.31223/X5CQ02>.