Generated using the official AMS LaTeX template v6.1

# Events Data Base for Data Retrieval from the CASA DFW X-band Radar Network

Aleksei Rutkovskii,[a] , Kathleen McLaughlin,[a] David Westbrook, [a] Brenda Philips,[a] Michael Zink,[a]

[a] *University of Massachusetts Amherst*

*Corresponding author*: Aleksei Rutkovskii, arutkovskii@umass.edu

ABSTRACT: Accessing and utilizing high-resolution atmospheric data is critical for advancing research in mesoscale meteorology, numerical weather prediction, and machine learning applications. However, challenges such as complex data formats, high computational demands, and fragmented access methods hinder efficient data retrieval. This work introduces the CASA Data Portal, a web-based platform designed to enhance the accessibility and usability of X-band radar data from the CASA DFW radar network. Inspired by SEVIR and Dynamical.org, the portal integrates event-based data retrieval, cloud-optimized storage, and efficient data extraction with Xarray and Zarr, along with an intuitive web interface to streamline data access. Researchers can query and extract data guided by NOAA's Storm Events Database records and/or CASA's internal archives, facilitating efficient dataset selection for scientific analysis and machine learning workflows. By bridging the gap between complex meteorological datasets and practical research needs, the CASA Data Portal enhances rapid, scalable, and efficient data-driven workflows.

## 1. Introduction

In atmospheric sciences, data availability and usability are essential for advancing mesoscale, synoptic, and climate research and applications. Atmospheric data is gathered through various methods, including satellite observations, ground-based stations, radar systems, and emerging technologies like drones and IoT sensors. Each method contributes unique data; however, accessing various atmospheric science datasets presents many challenges. Atmospheric data is typically archived by time and location rather than by specific meteorological phenomena, making it difficult for researchers to retrieve relevant data without prior knowledge of the event (Oakley and Daudert 2016). Additionally, the size and diversity of atmospheric data, collected at varying temporal and spatial resolutions, require familiarity with data formats such as NetCDF, which demand specialized software and expertise to interpret, as well as substantial computational resources for storage, processing, and analysis, often necessitating sophisticated data management techniques (Veillette et al. 2020). For example, working with ERA5, a comprehensive reanalysis dataset used to develop and train machine learning models, involves managing vast amounts of data across multiple variables like temperature, humidity, and wind speed. This requires large computational resources for downloading, storing, and preprocessing. This workflow is complex and challenging for beginners. It is often slow due to the need for specialized knowledge and tools to handle large datasets and perform the necessary data transformations. As machine learning research expands, an increasing number of researchers encounter these challenges. This complexity presents a significant barrier to broad data accessibility and usability, slowing down research progress and innovation in atmospheric science.

This research focuses on improving the accessibility of CASA's (Engineering Research Center for Collaborative Adaptive Sensing of the Atmosphere) X-band Radar data for atmospheric research. The CASA DFW Network is a network of six X-band radars in North Texas that have been generating high spatiotemporal resolution single radar moments and multi-radar products since 2016 (Philips et al. 2017; Chandrasekar et al. 2018). We draw inspiration from concept introduced by SEVIR — a curated, annotated dataset of over 10,000 weather events — and from approach utilized by Dynamical.org, a public-benefit corporation that provides long-term, open-access weather and climate data with a focus on sustainability and usability. SEVIR is specifically designed for deep learning applications in meteorology, providing temporally and spatially aligned image

sequences depicting weather events captured over the contiguous US (CONUS) by the GOES-16 satellite and the mosaic of NexRad radars (Veillette et al. 2020). The structured nature of SEVIR reinforced the importance of this research, highlighting the value of well-curated datasets for scientific study. SEVIR's user-friendly approach allows scientists to focus more on model development and analysis rather than on data preparation, further underscoring the motivation to enhance access to CASA data. As for Dynamical.org, it utilizes technologies that enable the provision of scalable, cloud-optimized access to weather and climate data. By leveraging cloud-native storage solutions such as Zarr and advanced data management frameworks like Xarray, Dynamical.org enables efficient, on-demand retrieval of large-scale atmospheric datasets. Its approach reduces the computational overhead associated with traditional data access methods and enhances reproducibility in meteorological research. Inspired by these methodologies, our work incorporates cloud-optimized storage, efficient indexing, and event-driven data access, ensuring that high-resolution CASA radar data is not only accessible but also efficiently retrievable for a diverse range of applications.

Building upon the event-based structure of SEVIR and the scalability of Dynamical.org, we introduce the **CASA Data Portal**, a platform designed to bridge the gap between atmospheric data complexity and usability. This platform includes a web interface that allows users to navigate and selectively extract atmospheric data of interest. Researchers can use the NOAA Storm Events database and CASA's own events archive to easily discover significant weather events at a scale relevant to CASA data. A key strength of our platform is its dual access approach: a web interface enables users to visually explore and select specific data subsets, while Zarr format access is streamlined through Kerchunk-generated (https://fsspec.github.io/kerchunk/) reference files for efficient data retrieval. This dual access method combines the simplicity of web-based data exploration with the power of scalable, programmatic data processing, accommodating a wide range of research needs. To support this enhanced accessibility and usability, the data is stored in a cloud S3 bucket hosted on the Open Storage Network (https://www.openstoragenetwork.org/), ensuring high availability and reliability. The backend and frontend of the platform are hosted on NERC (https://nerc.mghpcc.org/) and/or on popular web providers such as Amazon Web Services or Google Cloud Platform. This state-of-the-art solution improves both data availability and usability, accelerating advancements in mesoscale meteorology research in mesoscale meteorology,

numerical weather prediction (NWP), machine learning (ML), and climate science. Our platform enables researchers to develop predictive models and applications more rapidly and effectively, leveraging both the accessibility of the web interface and the efficiency of Zarr-based data access for high-performance computing environments.

## 2. Related Work

Historically, large atmospheric data platforms have been designed primarily for data storage and distribution rather than for efficient access and usability. Systems such as NOAA's data services, ERA5, and the CEDA Archive provide extensive datasets crucial for meteorological and climate research. However, as the volume and complexity of atmospheric data has grown, these platforms have become overwhelming and often inefficient for researchers who require quick, streamlined access to specific subsets of data. Navigating these systems frequently requires specialized expertise, and working with the data demands substantial computational resources, posing challenges for a broad range of users (NOAA 2025b; Service 2025; Science and Council 2025).

In response to these inefficiencies, modern datasets and platforms such as SEVIR and Dynamical.org have adopted curated approaches and cloud-optimized formats such as Zarr, along with libraries like Xarray, to enhance usability and efficiency. These initiatives focus on scalable, analysis-ready data solutions that facilitate programmatic access while reducing the need for extensive preprocessing. By structuring their datasets to align with machine learning applications and enabling on-demand retrieval, they have made significant strides in optimizing data accessibility for researchers.

To contextualize our approach, we review key existing platforms and datasets, highlighting their strengths and limitations in relation to the vision of the CASA Data Portal. Our platform aims to bridge the gap between accessibility and high-performance data processing by combining an intuitive web-based interface with efficient, programmatic access using Xarray and Zarr, enabling researchers to discover and interact with CASA data in a more streamlined and effective manner.

### *a. NOAA Services*

NOAA's ecosystem offers an impressive breadth of atmospheric data through multiple discovery tools and access methods. The OneStop portal, for instance, provides extensive filtering options by scientific discipline, data format, temporal range, and geographic location. This capability is particularly useful for users seeking specific datasets—from historical climate records to real-time radar products—making it a valuable resource for research. Additionally, NOAA's Data Access Application and Climate Data Online APIs facilitate programmatic downloads and automated

data retrieval, which can be integrated into machine learning workflows (NOAA 2025b,a; for Environmental Information NCEI).

However, these same features also introduce significant challenges. The complexity of NOAA's interfaces—combined with the need for specialized knowledge in data formats such as NetCDF—can make initial data exploration and extraction a steep learning curve, particularly for newcomers. Moreover, the multi-step ordering process and high volume of data sometimes result in delays and require considerable computational resources, which further underscores the need for more streamlined and user-friendly platforms like our CASA Data Portal.

*b. CEDA Archive and UK Composite Rainfall Data*

The CEDA Archive, hosting the UK Composite Rainfall Data from the Met Office's NIMROD system, provides high-resolution precipitation information that is critical for short-term forecasting and machine learning applications. With detailed documentation, dynamic file organization, and both bulk and individual download options, the platform is well-suited for researchers who need to access extensive temporal records and high-resolution spatial data. The rich metadata and well-structured archives enable efficient exploration and integration into modeling frameworks (Science and Council 2025).

Despite these strengths, the platform presents several hurdles. User registration is required, and the use of non-standard file formats (such as the NIMROD .dat files) often necessitates additional preprocessing steps. In some cases, the absence of automated handling for missing data means that researchers must directly contact data providers to resolve gaps, thereby delaying analysis. These challenges point to the broader issue of accessibility in atmospheric datasets and motivate efforts to simplify data retrieval and processing, as seen in our work.

*c. ERA5 Reanalysis*

ERA5 is widely recognized for its comprehensive global atmospheric reanalysis, offering high spatial and temporal resolution data that underpin many climate and weather prediction models. The dataset covers a broad range of atmospheric variables—from temperature and humidity to wind and precipitation — and is available via the Climate Data Store (CDS) interface and its supporting API. These features facilitate tailored data selection and efficient programmatic access,

making ERA5 a cornerstone for both long-term climate studies and machine learning applications (Hersbach et al. 2020; Service 2025).

Nevertheless, ERA5 is not without its drawbacks. The sheer volume of data and its intricate structure require substantial computational resources, and users must often navigate a steep learning curve to fully exploit its capabilities. Moreover, known biases—especially in the precipitation records—necessitate careful quality control and preprocessing. This combination of strengths and challenges illustrates the persistent need for platforms that can make high-quality atmospheric data more accessible and easier to work with, providing clear motivation for our CASA Data Portal initiative.

### *d. German Weather Service (DWD) Products*

The German Weather Service (DWD) provides a suite of high-resolution weather products, including RADOLAN, RADKLIM, and composites such as RV and WN. These platforms excel at delivering quantitative precipitation estimates by integrating radar observations with rain gauge measurements, thus enhancing the accuracy of real-time and historical data. The high spatial resolution (often at the 1 km scale) and frequent updates (every 5 minutes) make these datasets invaluable for both operational forecasting and long-term climate studies of Germany (Kreklow et al. 2019).

However, the wealth of data provided by DWD comes with its own set of challenges. The extensive archives and multiple data formats (such as GRIB and binary files) can make it difficult for non-specialists to identify and extract the specific datasets they need. The lack of an integrated user interface and the necessity for scripting or programming knowledge to automate downloads further limit immediate usability. These hurdles emphasize the ongoing need for platforms that streamline access and simplify data processing.

### *e. TAASRAD19*

TAASRAD19 by Franch et al. (2020) offers a high-resolution radar reflectivity dataset that is particularly useful for precipitation nowcasting in complex, mountainous regions. The dataset is collected by the Civil Protection weather radar of the Trentino South Tyrol Region, in the Italian Alps. The dataset is characterized by its high temporal (5-minute intervals) and spatial (sub-

kilometer resolution) granularity, along with detailed annotations and noise mitigation strategies. These features make it an excellent resource for developing machine learning models that require precise and reliable inputs, especially in challenging terrains where radar signals are often affected by topographical variations. Yet, TAASRAD19 also faces limitations regarding data integration and standardization. The use of multiple data formats (e.g., ASCII, HDF5, NetCDF) and the need for specialized techniques to correct for beam blockage at lower levels and other artifacts add layers of complexity to its use.

### *f. SEVIR*

The SEVIR dataset by Veillette et al. (2020) was explicitly developed to address the challenges of aligning multimodal weather data for deep learning applications. By integrating temporally and spatially aligned radar and satellite imagery, SEVIR offers a curated collection of over 10,000 weather events with comprehensive metadata. Its standardized file formats, along with the availability of supporting code and baseline models, simplify data ingestion and model training, allowing researchers to bypass many of the typical preprocessing hurdles. Additionally, SEVIR is hosted on AWS S3, leveraging cloud-based storage to facilitate scalable access to its dataset. This curated approach significantly enhances accessibility for end users, providing a structured and efficient foundation for machine learning applications in atmospheric science.

### *g. Dynamical.org: Sustainable, Cloud-Optimized Weather Data Access*

Dynamical.org is a public benefit corporation providing long-term, open-access weather and climate data with a focus on sustainability and usability. It facilitates efficient access to datasets and model predictions from organizations committed to cloud-optimized data sharing. The platform adopts a cloud-native approach using the Zarr format, enabling scalable, analysis-ready datasets with efficient, on-demand access. It currently offers an hourly NOAA GFS analysis dataset (2015–present) and plans to expand to GEFS, HRRR, and other high-resolution forecasting models.

By leveraging tools like Xarray and Dask, users can efficiently retrieve specific data subsets in parallel without downloading entire archives, making it ideal for large-scale modeling and AI-driven forecasting. While transitioning from NetCDF or GRIB to Zarr requires adaptation, this approach

enhances accessibility, reproducibility, and efficiency in atmospheric sciences (dynamical.org 2025).

### *h. Summary*

The review of existing atmospheric data platforms highlights significant advancements in data accessibility, cloud optimization, and machine learning integration. However, challenges such as steep learning curves, data fragmentation, and computational overhead remain persistent barriers for researchers. Our CASA Data Portal aims to bridge these gaps by providing an intuitive, high-performance solution that combines user-friendly web access with efficient programmatic interfaces using Xarray and Zarr. By addressing limitations in existing platforms, our work enhances accessibility and usability for a broad range of atmospheric science applications.

## 3. CASA Testbed & Data

### *a. Overview of the CASA DFW Radar Network*

The CASA DFW radar network monitors the lower atmosphere in the Dallas–Fort Worth–Arlington (DFW) region, an area spanning roughly 194 x 201 km and serving approximately 6.5 million people. This network of six advanced X-band, dual-polarization radars is specifically designed for the detection and mitigation of severe weather phenomena—such as hail, tornadoes, and floods — by providing high-resolution observations of the lower atmosphere. Unlike traditional systems, CASA's dense coverage is optimized to capture rapidly evolving localized weather events, providing information that is critical to urban safety. Figure 1 illustrates the spatial layout and coverage areas of both the CASA X-band radars and the NexRad S-band radar (KFWS) over the DFW Metroplex, highlighting the dense network's capability to monitor urban weather phenomena with high spatial and temporal resolution.

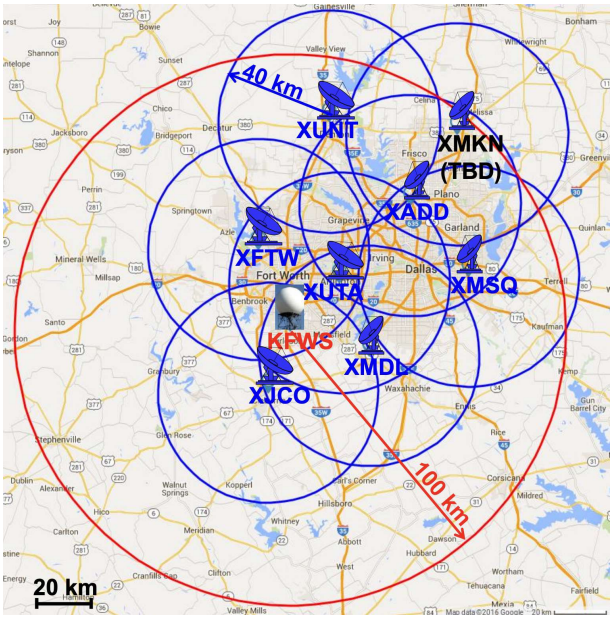



FIG. 1. Location and coverage of CASA and NexRad Radars over DFW region. The layout of S-band KFWS WSR-88DP radar (100 km range ring in red) and the DFW dual-polarization X-band radars (40 km range rings in blue). The letter symbols, such as "XMDL", correspond to the name of various radars. Figure from Chandrasekar et al. (2018).

Furthermore, the DFW network serves as a testbed for integrating advanced technologies, such as attenuation correction and clutter suppression, to enhance the reliability of urban radar systems. The dual-polarization capability, coupled with collaborative multi-radar approaches, provides robust, fault-tolerant data to mitigate the impact of equipment outages.

The article *Principles of High-Resolution Radar Network for Hazard Mitigation and Disaster Management in an Urban Environment* by Chandrasekar et al. (2018) describes in detail CASA X-Band Radar Network, compares WSR-88DP S-band Radar (NexRad) with DFW X-band Radar (CASA), provides locations of CASA radars, and explains principles behind the radar network.

### *b. Data and Capabilities*

Since March 2016, the CASA network in DFW has produced a range of high-resolution weather products, including: Quantitative Precipitation Estimates (QPE) at a resolution of 554 m $\times$ 549 m; specialized products for identifying hail events (hail detection); both single-radar and composite reflectivity measurements; single-radar velocity products capturing storm dynamics.

In comparison with the NexRad S-Band radars, CASA radars exhibit notable differences. CASA radars provide coverage below 1 km, whereas NexRad only covers about 30% of its data below 1 km. In addition, CASA offers updates every 1 minute, in contrast to NexRad's 5-minute update cycle. These differences yield detailed observations of the lower atmosphere with CASA, while NexRad offers broader-scale atmospheric coverage.

## 4. CASA Data Portal Implementation

This section describes the proposed CASA Data Portal, presenting its objectives, functional requirements, system design, and completed work to date. By integrating event-based data retrieval, efficient storage solutions, and user-friendly accessibility, the portal aims to support researchers and practitioners in event analysis and machine learning workflows.

### *a. Goals*

The CASA Data Portal is envisioned to fulfill the following primary objectives:

1. Provide access to event-related atmospheric data, leveraging records from the NOAA Storm Events Database and CASA Alerts for the DFW region (for Environmental Information NCEI; CASA 2025).

2. Host and serve the CASA DFW Dataset, comprising eight years of precipitation data from CASA X-band radars and NexRad "KFWS" S-band radar (June 1, 2016 – May 31, 2024) used in deep generative modeling applications.

3. Provide access to all CASA radar data products from March 2016 (as listed in Section b) to support broader machine learning workflows, contingent on available resources.

### *b. Requirements*

To achieve its intended functionality, the CASA Data Portal must meet the following technical and operational requirements:

- **Accessibility**: The platform must be hosted online to allow remote user access.
- **User Interface**: Provide an intuitive interface for selecting weather events based on user-defined parameters such as date range, location, and event severity.
- **Data Storage and Integration**: Store radar maps in NetCDF format while generating reference files and mappings using Kerchunk for optimized data access.
- **Efficient Data Access**: Enable users to interact with and download data in Zarr format, supported by reference files for rapid retrieval.

- **Command-line Support**: Facilitate dataset downloads via command-line tools, ensuring ease of integration into automated workflows.

*c. Limitations*

Despite its robust design, the CASA Data Portal will operate under certain constraints:

- **Cost Considerations**: To minimize expenses, the system will leverage NERC (https://nerc.mghpcc.org/) servers for backend operations, Vercel (https://vercel.com/) for frontend deployment, and Open Storage Network (https://www.openstoragenetwork.org/) for data storage.
- **Storage Capacity**: The OSN bucket currently provides a 30-terabyte allocation, extendable up to 50 terabytes upon request.
- **Update Frequency**: Event records will be updated according to the NOAA Storm Events Database's monthly update schedule, potentially introducing delays in data availability to the latest month of data.

*d. System Design*

1) Component 1 – Frontend

The frontend will be implemented using Next.js and hosted on Vercel. Next.js, a React-based framework, enables seamless integration with Vercel and simplifies frontend development by abstracting infrastructure management.

The frontend will consist of the following pages:

- **Home:** Explains the platform's purpose and guides users on how-to use the platform.
- **Atmospheric Events:** Provides an interface for users to view files associated with selected events, explore the CASA DFW Dataset, or create custom datasets for ML workflows. While files cannot be downloaded directly from the interface, users can provide their email to receive reference files for the selected events. For downloading the CASA DFW Dataset, users can either receive a reference file via email or obtain a command-line instruction from the interface to download the data directly from the S3 bucket.

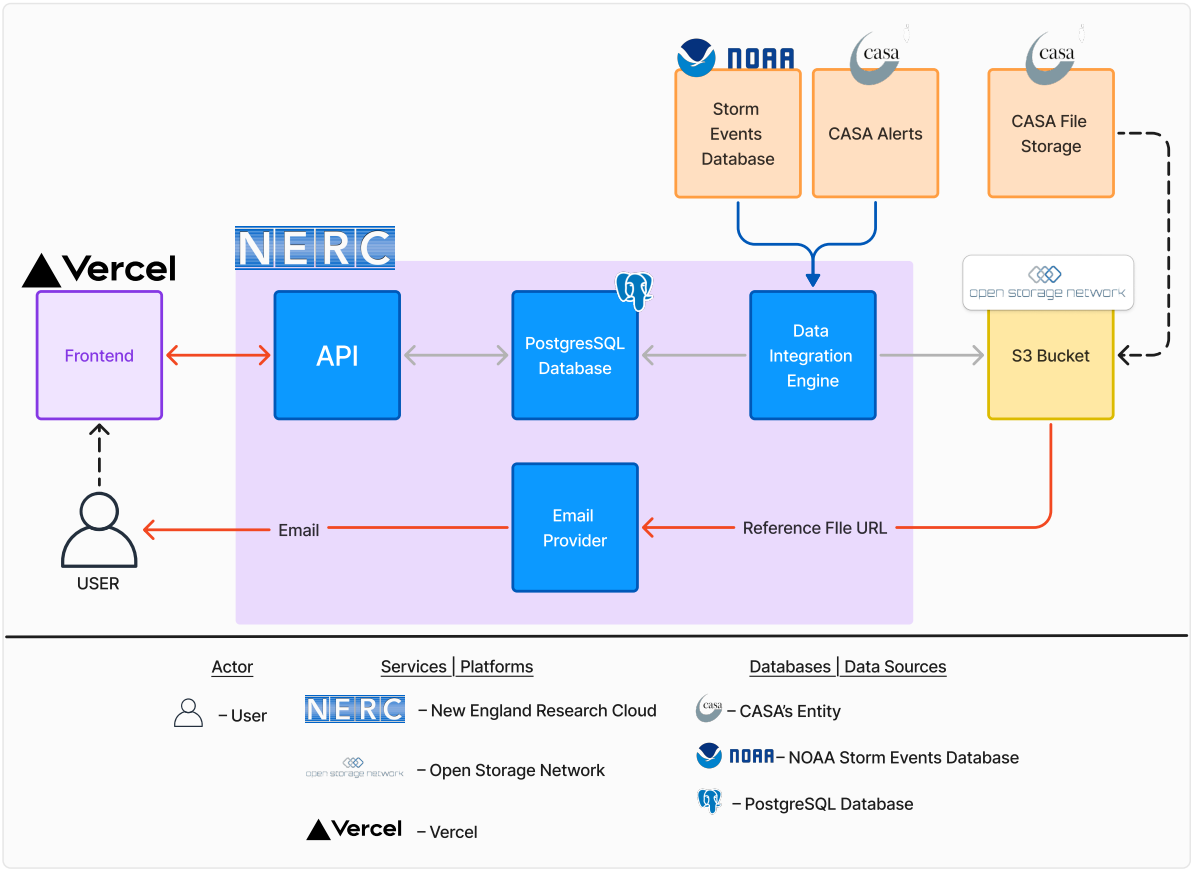



FIG. 2. CASA Data Portal System Design.

- **Documentation:** Provides a concise guide for downloading atmospheric data using reference files, a tutorial on navigating the Atmospheric Events interface, and additional background information about CASA's data.
- **About:** Provides detailed information about CASA, the platform, its goals, and the contributors.
- **Feedback Form:** Enables users to report issues, suggest improvements, or submit requests.

### 2) COMPONENT 2 – BACKEND

The backend will be implemented as a Dockerized FastAPI service. This API will:

- Process queries from the frontend.
- Interact with the PostgreSQL database to fetch relevant data.

- Email reference files and data-download instructions to users.
- Submit content of the Feedback Form.

Docker ensures consistent deployment, while FastAPI offers a high-performance web framework for API development.

### 3) Component 3 - Data Integration Engine

This engine will perform the following functions:

1. Periodically extract new events from the NOAA Storm Events Database and CASA Alerts.
2. Map events to corresponding NetCDF files and update PostgreSQL database.
3. Upload NetCDF files to OSN and trigger Kerchunk to generate reference files.
4. Add reference to NetCDF mappings to the database to enable efficient data retrieval.

### 4) Component 4 - Open Storage Network (OSN)

OSN will store NetCDF files and corresponding JSON reference files generated by Kerchunk. These reference files enable efficient data access using Zarr libraries. This S3-compatible storage solution supports Zarr operations, allowing users to stream or download data efficiently in parallel using the reference files.

### 5) Component 5 - CASA Server

The CASA server, which houses atmospheric data, will serve as the initial source for data uploads to OSN.

### 6) Component 6 - CASA Alerts

CASA Alerts, a service supported by a MySQL database, contains records of atmospheric events reported by public of DFW Metroplex. These records will supplement the NOAA Storm Events Database records.

### 7) Component 7 - NOAA Storm Events Database

The NOAA Storm Events Database, the primary source for event selection, provides annual records with metadata such as event types, start and end times, and descriptions. This data will drive the event selection functionality.

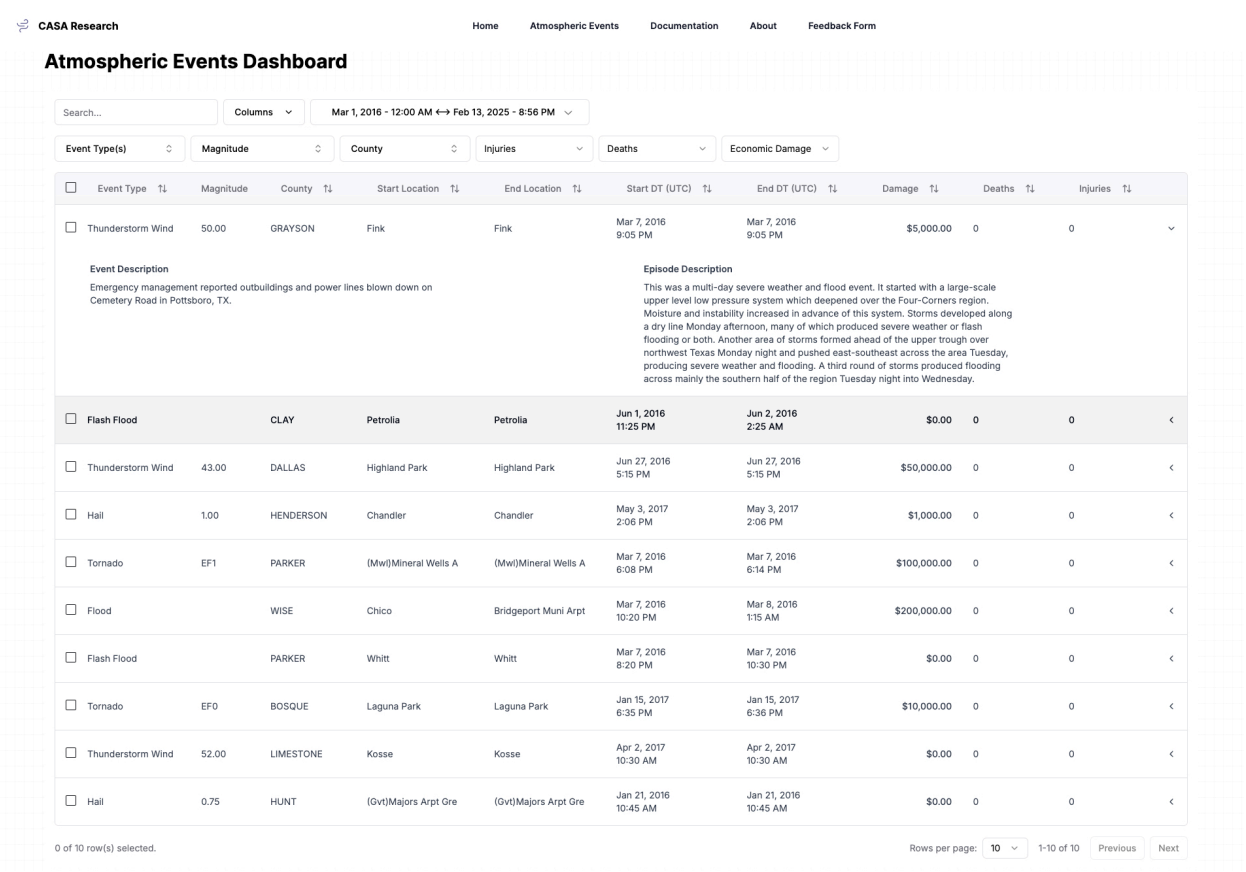



Fig. 3. CASA Data Portal – Atmospheric Events Page

### *e. Completed Work*

At this stage of development, several foundational steps have been achieved:

- **Prototyping and Architecture Setup**: A minimal FastAPI backend has been tested locally, demonstrating the feasibility of using Docker for consistent deployment.

- **Initial Data Ingestion**: A subset of NetCDF files has been uploaded to OSN with corresponding reference files, and their mappings have been store in PostgreSQL Database.
- **Database Integration**: A PostgreSQL schema has been initialized, storing metadata from NOAA Storm Events and CASA Alerts for retrieval.
- **Frontend Development**: Basic Next.js frontend have been prototyped, validating integration with Vercel.

## 5. Conclusion and Future Work

Building on the usability principles demonstrated by SEVIR and the cloud-optimized approach adopted by Dynamical.org, our CASA Data Portal is being developed to provide structured, accessible, and efficient access to high-resolution atmospheric data. By integrating event-based retrieval from the NOAA Storm Events Database and utilizing scalable data formats such as Zarr, the platform will enhance both usability and computational efficiency. The completed work — including initial data ingestion, backend prototyping, and frontend development — lays a strong foundation for achieving the objectives set forth in Section a. Given the progress made, we anticipate that the platform will fully satisfy all outlined goals upon completion.

While the core architecture and essential functionality are established, further refinements and integrations remain to be completed. Future work includes:

- **Integrating CASA Alerts** – Incorporating real-time atmospheric event reports from CASA's alerting system to enrich event metadata and provide more comprehensive event discovery.
- **Refining the User Interface and Conducting Testing** – Improving the frontend for seamless user experience and performing extensive user testing to ensure accessibility, efficiency, and robustness.
- **Automating the Data Integration Engine** – Implementing an automated pipeline for continuous data ingestion and processing, incorporating message queues such as RabbitMQ to facilitate communication between backend microservices. RabbitMQ will enhance fault tolerance by ensuring message persistence and retries in case of failures and will improve reliability by providing guaranteed message delivery, load balancing, and efficient inter-service communication.
- **Deploying the System** – Finalizing platform deployment across Vercel (for the frontend) and the New England Research Cloud (NERC) (for backend services and data processing), ensuring stability, scalability, and accessibility.

With these remaining steps, the CASA Data Portal will serve as a state-of-the-art solution for atmospheric research, enabling streamlined access to high-resolution radar data and facilitating advanced machine learning applications. By bridging the gap between data complexity and acces-

sibility, this work contributes to enhancing atmospheric research, numerical weather prediction, and climate science.

*Acknowledgments.* Funding for this work was provided by Jerome M. and Linda L. Paros whose generosity established the Paros Center for Atmospheric Research.

## References

CASA, U., 2025: Umass casa urban testbed applications — casa alerts. URL http://www.casa.umass.edu/main/research/urbantestbed/apps/.

Chandrasekar, V., H. Chen, and B. Philips, 2018: Principles of high-resolution radar network for hazard mitigation and disaster management in an urban environment. *Journal of the Meteorological Society of Japan. Ser. II*, **96A**, 119–139, https://doi.org/10.2151/jmsj.2018-015.

dynamical.org, 2025: dynamical.org. URL https://dynamical.org/.

for Environmental Information (NCEI), N. N. C., 2025a: Noaa national centers for environmental information (ncei) data access. URL https://www.ncei.noaa.gov/cdo-web/.

for Environmental Information (NCEI), N. N. C., 2025b: Noaa storm events database. URL https://www.ncdc.noaa.gov/stormevents/.

Franch, G., V. Maggio, L. Coviello, M. Pendesini, G. Jurman, and C. Furlanello, 2020: Taasrad19, a high-resolution weather radar reflectivity dataset for precipitation nowcasting. *Scientific Data*, **7 (1)**, 234.

Hersbach, H., and Coauthors, 2020: The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **146 (730)**, 1999–2049, https://doi.org/https://doi.org/10.1002/qj.3803, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3803.

Kreklow, J., B. Tetzlaff, G. Kuhnt, and B. Burkhard, 2019: A rainfall data intercomparison dataset of radklim, radolan, and rain gauge data for germany. *Data*, **4 (3)**, 118.

NOAA, 2025a: Noaa coast data viewer. URL https://coast.noaa.gov/dataviewer/.

NOAA, 2025b: Noaa onestop data search platform. URL https://data.noaa.gov/onestop/.

Oakley, N. S., and B. Daudert, 2016: Establishing best practices to improve usefulness and usability of web interfaces providing atmospheric data. *Bulletin of the American Meteorological Society*, **97 (2)**, 263–274.

Philips, B., T. Ryan, V. Chandrasekar, E. Lyons, T. Bradshaw, M. Fox, F. Junyent, and A. Bajaj, 2017: Tracking tornados down streets: Using casa radars in real time severe weather warning operations in north central texas. *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, IEEE, 5973–5976.

Science, and T. F. Council, 2025: Ceda archive. URL https://archive.ceda.ac.uk/.

Service, C. C. C., 2025: Copernicus climate data store. URL https://cds.climate.copernicus.eu/.

Veillette, M., S. Samsi, and C. Mattioli, 2020: Sevir: A storm event imagery dataset for deep learning applications in radar and satellite meteorology. *Advances in Neural Information Processing Systems*, **33**, 22 009–22 019.