FRBSTATS: A WEB-BASED PLATFORM FOR VISUALIZATION OF FAST RADIO BURST PROPERTIES

Apostolos Spanakis-Misirlis\textsuperscript{1,2} and Cameron L. Van Eck\textsuperscript{3}

\textsuperscript{1} Department of Informatics, University of Piraeus, Greece
\textsuperscript{2} Institute of Astrophysics, FORTH, Dept. of Physics, University of Crete, Voutes, University Campus, GR-71003, Heraklion, Greece
\textsuperscript{3} Research School of Astronomy and Astrophysics, The Australian National University, Canberra ACT 2611, Australia

ABSTRACT

The study of fast radio bursts (FRBs) is of great importance, and is a topic that has been extensively researched, particularly in recent years. While the extreme nature of FRBs can serve as a tool for researchers to probe the intergalactic medium and study exotic aspects of the Universe, keeping track of FRB properties is challenged by the frequent detection of new bursts. We introduce the FRBSTATS platform, which provides an up-to-date and user-friendly web interface to an open-access catalogue of published FRBs, along with a statistical overview of the observed events. The platform supports the retrieval of fundamental FRB data either directly through the FRBSTATS API, or in the form of a CSV/JSON-parsed database, while enabling the plotting of parameters and their distributions, for a variety of visualizations. These features allow researchers to conduct population studies and comparisons with astrophysical models, describing the origin and emission mechanism behind these sources. So far, the inferred redshift estimates of 813 bursts have been computed, providing the first public database that includes redshift estimates inferred from dispersion measure entries for nearly all observed FRBs, as well as host redshifts (where available). Lastly, the platform provides a visualization tool that illustrates associations between primary bursts and repeaters, complementing basic repeater information provided by the Transient Name Server. In this work, we present the structure of the platform, the established version control system, as well as the strategy for keeping such an open database up to date. Additionally, we introduce a novel, computationally-efficient, clustering-based approach that enables unsupervised classification of hundreds of bursts into repeaters and non-repeaters, resulting in the discovery of one new FRB repeater. The platform is accessible at: https://www.herta-experiment.org/frbstats

1. INTRODUCTION

Fast radio bursts (FRBs) are bright, short-duration pulses of unknown origin, observed at radio wavelengths. Since the first detection by Lorimer et al. (2007), hundreds of such events have been observed, with varying emission characteristics (see Fig. 1). Despite the 15-year period of FRB observations however, the emission mechanism behind these sources remains unknown (Petroff et al. 2022). Although the exact origin of FRBs is highly debated, nearly all FRBs detected so far are extragalactic, with varying spectral, polarization, and time-domain emission properties. With the exception of the Galactic magnetar SGR 1935+2154 (CHIME/FRB Collaboration et al. 2020), the dispersion measure of FRBs implies that their origin is extragalactic. It is worth noting, that some FRBs exhibit a repeating behaviour (Spitler et al. 2016), which has ruled out, altered, and even introduced a variety of new theoretical models describing the radiation process (Platts et al. 2019). An unusual subclass of repeating FRBs appears to show a periodic activity (Chime/Frb Collaboration et al. 2020; Rajwade et al. 2020), which may further narrow down the list of candidate models.

Because of their enigmatic nature and the rapid rate of discoveries, gaining a better and more complete understanding of these astrophysical objects is an essential prerequisite to uncovering a significant piece of the puzzle behind the mysterious presence of this class of sources. Due to the unpredictability of FRBs however, the consistent monitoring of individual sources is a highly time-consuming and challenging task. While each observation comes with its own derived properties, carrying out literature reviews for each individual burst becomes impractical, especially as the number of detected bursts increases. Without the most significant properties of each detected burst indexed in an accessible database, population studies and extensive analysis of FRBs (taking every observed event into consideration to build statistical models) become laborious.

While previous efforts relevant to cataloging FRBs, such as the Transient Name Server (TNS; Yaron et al. 2020) and FRBCAT (Petroff et al. 2016) have been able to collect a meaningful amount of data from FRB publications, such platforms have struggled to remain up to date with new bursts (occasionally with several values being reported inaccurately), and are arguably limited in terms of their functionality and ease of use.

In this work, we present a platform that hosts an accurate and complete catalogue of FRBs, with the primary goal of enabling researchers to easily derive fundamental properties of bursts, with minimal effort. In Sect. 2, we begin by describing the high-level ar-
Fig. 1.— Total number of published bursts (from 622 unique FRB sources) observed as a function of their detection time. The notable increase of observed bursts in July 2019 came with the commissioning of the Canadian Hydrogen Intensity Mapping Experiment Fast Radio Burst (CHIME/FRB) Project. FRB 20180725A marks the very first detection by CHIME, representing the first ever FRB observed below 700 MHz.

architecture of the system, as well as the strategy for maintaining and keeping the catalogue up to date, as autonomously as possible. Accessible via a user-friendly web interface, FRBSTATS includes an open-access catalogue of all FRBs published to date (Sect. 3), along with a statistical overview of their properties. Furthermore, the platform hosts a page that illustrates the association between primary bursts and repeaters (Subsect. 3.1), complementing basic repeater information provided by the TNS. In addition to the ability of plotting of parameter distributions for a variety of visualizations (Subsect. 3.2), the platform supports the retrieval of burst data, directly through the FRBSTATS API (Subsect. 3.3). We finally conclude by summarizing our work and discussing future enhancements for the platform (Sect. 4).

2. PLATFORM ARCHITECTURE

In order to ensure all data and code supporting the platform is findable, accessible, interoperable, and reusable, the source of the website is hosted on a Git repository.

Based on a non-relational document model, MongodB offers a suitable NoSQL database, which serves as the catalogue infrastructure for the platform. By automatically parsing the CSV catalogue (obtained from the spreadsheet) as a JSON-formatted file, the catalogue page permits quick search queries, based on either the designation of the burst of interest, or its parameters.

3. CATALOGUE

The catalogue consists of the observed properties for each burst: the topocentric Coordinated Universal Time (UTC) and Modified Julian Date (MJD) of the observed event, the corresponding telescope that detected the burst, and any new burst reported on the TNS is automatically queried and merged to FRBSTATS, on a daily basis.

1 https://www.wis-tns.org
2 https://github.com/HeRTA/FRBSTATS
3 https://github.com/mongodb/mongo
the burst, and its sky location in equatorial (right ascension α and declination δ, along with their associated errors), as well as Galactic coordinates (longitude λ and latitude b). Additionally, the database includes the frequency of each burst, its dispersion measure and error, host and redshift estimate inferred from the dispersion measure (computed with the help of the fruitbat package: Batten 2019), peak flux density, burst width and fluence, as well as the signal-to-noise ratio of the detection. For the redshift inference, the fruitbat package calculates the excess dispersion measure by subtracting the contribution of the Galactic foreground, before using a relation between DM_{IGM} and z (Ioka 2003; Inoue 2004; Zhang 2018). The catalogue also contains a reference for each burst to direct users to the original data source.

Event data are obtained from the TNS, FRBCAT, the First CHIME/FRB Fast Radio Burst Catalog. The Astronomer’s Telegram (Rutledge 1998), and the corresponding publications cited on the FRBSTATS catalogue. We note that since Virtual Observatory Events (VOEvents) from CHIME have not been verified and characterized, they have not been incorporated into the catalogue. Since the TNS remains highly relevant for the submission of new FRB observations, appropriate CI/CD pipelines have been set up and scheduled to monitor the database for new entries on an hourly basis, updating the FRBSTATS catalogue automatically. For some bursts, parameters obtained and merged from other catalogues have been reviewed and corrected based on the publications associated with each burst.

### 3.1. Repeaters

As mentioned in Sect. 1, a fraction of FRB sources appear to exhibit a repeating activity; a single FRB source is capable of producing more than one burst. This clue becomes especially significant if the observed properties appear very similar to the previously-emitted bursts.

The discovery of the first repeater (FRB 20121102) by Spieler et al. (2016) demonstrated that such a class of FRBs could not originate from cataclysmic events involving the collapse of supramassive neutron stars (Fulinke & Rezzolla 2014) or neutron star mergers (Hansen & Lyutikov 2001). Although it is not yet clear whether a single emission mechanism is responsible for all FRBs (repeaters and non-repeaters), we find that this approach can be applied to FRBs using merely three parameters: the candidate’s coordinates (α, δ), as well as its dispersion measure (DM).

The DM describes the integrated free-electron column density between the observer and the FRB of interest, and based on the model of Macquart et al. (2020), the observed DM can be split into four contributing factors,

$$DM_{FRB} = DM_{disk} + DM_{halo} + DM_{cosmic} + DM_{host},$$

where the contribution of the Galactic disk and halo, the extragalactic distribution of ionized gas and the host galaxy of the FRB source are all taken into account.

Despite the uncertainty of DM_{host} described by James et al. (2022), Yang & Zhang (2017) show that variations in the dispersion measure of repeating FRBs are small. The stability of this parameter makes the DM a great alternative to using source distance as a clustering parameter, which, due to the limitation of free-electron distribution models being confined to Galactic sources (Cordes 2004; Schnitzeler 2012), remains unknown.

Assuming the three parameters (α, δ, DM) of repeaters used for clustering possess negligible variation over time/remain constant, we can employ a custom metric to reliably evaluate the distance between two given points (FRBs), using the three-dimensional Euclidean distance that separates them.

Let \(x_1, y_1, z_1\) and \(x_2, y_2, z_2\) be the Cartesian coordinates of FRB_1 and FRB_2, respectively. The position of each burst expressed in terms of α, δ and DM is then given by

\[
\begin{align*}
    x_i &= DM_i \cos \delta_i \sin \alpha_i \\
    y_i &= DM_i \cos \delta_i \cos \alpha_i \\
    z_i &= DM_i \sin \delta_i,
\end{align*}
\]

where \(i \in \{1, 2\}\) refers to the FRB of interest. We can then compute the Euclidean distance \(d\) between FRB_1 and FRB_2 using

\[
d(FRB_1, FRB_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}.
\]
However, considering the error values of $\alpha$, $\delta$ and DM are commonly reported, it is important to incorporate uncertainties into the clustering analysis so that the resulting clusters are reliably determined. This is especially useful when the plain Euclidean distance is insufficient for bursts whose reported uncertainties are relatively large.

The objective is to determine the ratio

$$\frac{d}{\sigma_d}, \quad (5)$$

where $d$ is the Euclidean distance and $\sigma_d$ is its standard uncertainty, derived using the variance formula for error propagation (Ku 1966). This ratio is used as the distance metric for DBSCAN, taking measurement uncertainties into account.

In order to derive $\sigma_d$, we begin by computing the partial derivatives of the Cartesian coordinates of the two bursts $(x_i, y_i, z_i)$, with respect to $\alpha$, $\delta$ and DM. By applying the variance formula, the standard uncertainties of the coordinates become:

$$\sigma_{x_i} = \sqrt{\left(\frac{\partial x_i}{\partial \alpha_i}\right)^2 \sigma_{\alpha_i}^2 + \left(\frac{\partial x_i}{\partial \delta_i}\right)^2 \sigma_{\delta_i}^2 + \left(\frac{\partial x_i}{\partial DM_i}\right)^2 \sigma_{DM_i}^2},$$

$$\sigma_{y_i} = \sqrt{\left(\frac{\partial y_i}{\partial \alpha_i}\right)^2 \sigma_{\alpha_i}^2 + \left(\frac{\partial y_i}{\partial \delta_i}\right)^2 \sigma_{\delta_i}^2 + \left(\frac{\partial y_i}{\partial DM_i}\right)^2 \sigma_{DM_i}^2},$$

$$\sigma_{z_i} = \sqrt{\left(\frac{\partial z_i}{\partial \alpha_i}\right)^2 \sigma_{\alpha_i}^2 + \left(\frac{\partial z_i}{\partial \delta_i}\right)^2 \sigma_{\delta_i}^2 + \left(\frac{\partial z_i}{\partial DM_i}\right)^2 \sigma_{DM_i}^2},$$

where $\sigma_{\alpha_i}$, $\sigma_{\delta_i}$ and $\sigma_{DM_i}$ are the reported errors of the right ascension, declination and DM, respectively.

Because repeaters have been observed to exhibit DM variations up to an order of 3% (Dai et al. 2022), $\sigma_{DM_i}$ is determined by the maximum value between the measurement uncertainty and the largest expected DM variation. Thus, the final standard uncertainty becomes:

$$\sigma_d = \sqrt{\sum_{u \in \{x,y,z\}} \left(\frac{\partial d}{\partial u_1}\right)^2 \sigma_{u_1}^2 + \left(\frac{\partial d}{\partial u_2}\right)^2 \sigma_{u_2}^2}. \quad (6)$$

Assuming the threshold for the number of neighbours, minPts, is equal to 2 (definition of FRB repeater; i.e., present at least 1 detection following the primary burst), we find that, by using Equation (5) as a metric for the distance function, reliable results are obtained when the

![One-Off Events](image1.png)

![Repeater Clusters](image2.png)

**Fig. 3.—**: Output scatter plots resulting from the DBSCAN algorithm for $\epsilon = 1.15$. The one-off events are counted as noisy samples/stray data points (left), while the detected 23 clusters are grouped and represented with a corresponding colour associated with an individual repeater (right).

**TABLE 1**: Fundamental burst properties available for direct plotting through the platform, along with a brief description for each parameter.

| Parameter | Source Position | Description |
|-----------|----------------|-------------|
| RA        | Right ascension (HMS) |
| RA Error  | Right ascension error (') |
| Dec.      | Declination (DMS) |
| Dec. Error| Declination error (') |
| Gal. Long.| Galactic longitude (l) |
| Gal. Lat. | Galactic latitude (b) |

**Burst Properties**

| Frequency | Reported burst frequency (MHz) |
| DM        | Reported dispersion measure (pc cm$^{-3}$) |
| DM Error  | Dispersion measure error (pc cm$^{-3}$) |
| Flux Density | Peak flux density (Jy) |
| Burst Width | Derived burst width (ms)$^a$ |
| Fluence   | Measured burst fluence (Jy ms) |

$^a$ Commonly reported as the fitted full width at half maximum (FWHM) of the burst, generally referring to the measured width (prior to scattering correction), depending on the submitted entry.
employed by DBSCAN to evaluate the association between two bursts, the radius value was optimized (fine-tuned) by identifying the \( \epsilon \) threshold that minimizes the number of false positives (FP).

In order to evaluate the confusion matrix and benchmark the robustness of the clustering algorithm, the TNS database was used as a reference for the set of actual (true) repeater parents (22 FRBs).

While the optimal value of \( \epsilon \approx 1.15 \) was identified using brute-force for simplicity, we note that the global minimum is easily retrievable in a matter of minutes (instead of hours), using conventional optimization algorithms. This is because the problem is one-dimensional and the objective function does not present many instabilities.

By using the cluster.DBSCAN class provided by the scikit-learn library (Pedregosa et al. 2011), we identify 23 repeating FRB sources. All repeaters reported on the TNS have been successfully identified, yielding a perfect false negative rate:

\[
FNR = \frac{FN}{FN + TP} = 0, \quad (8)
\]

where FN and TP are the false negatives and true positives, respectively.

The clustering algorithm identified one additional candidate repeater source: FRB 20190425B with one additional sub-burst association (FRB 20190614A). Upon visual inspection, we find that this candidate parent/daughter association is located near the north celestial pole, and has a similar DM with a variation of \( \sim 3\% \), consistent with what has previously been observed for FRB 20190520B (Dai et al. 2022) and FRB 121102 (Spitler et al. 2016; Law et al. 2017). We conclude that based on the available information, this source (highlighted in red on the FRBSTATS platform) could be a true repeater that requires further investigation.

Given the efficiency of the algorithm, repeater clustering of 818 bursts is carried out in \( \sim 60 \) seconds by the FRBSTATS server (1 Intel vCPU; 1 GB of memory). This includes the total data read time of the catalogue, the clustering itself, as well as the plotting of the results. The output of the clustering algorithm for 818 events is provided in Fig. 3.

### 3.2. Plots

Since the analysis of burst properties is a crucial step required to derive parameter distributions of FRBs and conduct population studies, the platform also offers an easy-to-use plotting tool for quick analysis. Similar to the Australia Telescope National Facility (ATNF) Pulsar Catalogue (Manchester et al. 2005), FRBSTATS enables direct parameter visualizations, without the need to download the catalogue and manually run scripts locally, in an offline manner. The burst parameters available for plotting are listed in Table 1.

An example plot obtained using FRBSTATS is shown in Fig. 4, highlighting the distribution of the flux, width, and dispersion measure of bursts. These plots are maintained up to date automatically, and are updated as soon as a new database entry/modification is submitted to the system. This ensures that newly detected events are displayed, and simple comparisons with the remaining FRB population can be made fairly easily.

### 3.3. API

An application programming interface (API) enables a convenient integration layer between a server and its clients. In the case of this particular platform, the FRBSTATS API provides researchers and developers access to FRB parameters in a lightweight, rapid, and easy manner, supporting numerous automation applications. All entries present in the database can be obtained with the help of the API, whose functionality is documented on the web platform, along with some examples. The client may submit a GET request to the provided endpoint URL, and query a burst of interest. If the entry is found in the database, the JSON-encoded API response will consist of the parameters of the queried burst, as well as the reference linked to the corresponding catalogue/publication. Thus, each parameter value can be reliably traced back to its original publication source for further verification. In the case of queries that are not as-

---

9 For the exact definition of \( \epsilon \), see: https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN
associated with any burst, the API returns an empty array (i.e., \[
\]
).

It is worth noting that the PHP-based utility automatically neglects the ‘FRB’ prefix, underscores and spaces, in order to improve the parsing of the query and handle different designations (corresponding to the same event) in a compatible fashion. Finally, unlike the catalogue-querying system in place, the server ensures that the search queries submitted to the API are directed to solely match the name of the burst of interest, and does not return irrelevant entries that could happen to match the values of certain parameters.

4. CONCLUSIONS

We have built and presented a versatile web platform consisting of an accurate open-access catalogue of fast radio bursts, a parameter visualization utility, as well as an application programming interface available for querying burst properties, such as inferred/host redshifts. Furthermore, we expanded on the method of automatically classifying FRBs into repeaters and non-repeaters using the DBSCAN algorithm with a custom distance metric. We hope that the platform provides researchers with access to an up-to-date database of FRBs to work with and conduct better population studies, with as little maintenance work as possible.

In future work, we plan on further expanding the catalogue by providing additional information regarding each burst, such as the Galactic, intergalactic, and local contributions to the DM, as well as polarization properties. Additionally, we hope to make dynamic spectra of each individual burst available and accessible, directly through the online catalogue.

ACKNOWLEDGEMENTS

We wish to acknowledge the help provided by Dr. John Antoniadis for the valuable feedback collected and provided to improve the utility of the platform, and Yannick De Clercq for assisting with the deployment of the system.

REFERENCES

| Author(s) | Year | Title/Details |
|-----------|------|---------------|
| Batten A. | 2019 | The Journal of Open Source Software, 4, 1399 |
| Bhattacharyya S. | 2017 | arXiv e-prints, p. arXiv:1711.09083 |
| CHIME/FRB Collaboration et al. | 2020 | Nature, 587, 54 |
| Chime/FRB Collaboration et al. | 2020 | Nature, 582, 351 |
| Cordes J. M. | 2004 | in Clemens D., Shah R., Brainerd T., eds, Astronomical Society of the Pacific Conference Series Vol. 317, Milky Way Surveys: The Structure and Evolution of our Galaxy, p. 211 |
| Dai S., et al. | 2022 | arXiv e-prints, p. arXiv:2203.08151 |
| Ester M., Kriegel H.-P., Sander J., Xu X. | 1996 | in Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. KDD’96. AAAI Press, p. 226–231 |
| Falcke H., Rezzolla L. | 2014 | A&A, 562, A137 |
| Hansen B. M. S., Lyutikov M. | 2001 | MNDRAS, 322, 695 |
| Imoue S. | 2004 | MNDRAS, 348, 999 |
| Ioka K. | 2003 | ApJ, 598, L79 |
| James C. W., Prochaska J. X., Macquart J. P., North-Hickey F. O., Bannister K. W., Dunning A. | 2022 | MNRAS, 509, 4775 |
| Kashiyama K., Ioka K., Mészáros P. | 2013 | ApJ, 776, L39 |
| Ku H. | 1966 | Journal of Research of the National Bureau of Standards, Section C: Engineering and Instrumentation, 70C, 263 |
| Law C. J., et al. | 2017 | ApJ, 850, 76 |
| Liu X. | 2018 | ApSS, 363, 242 |
| Lorimer D. R., Bailes M., McLaughlin M. A., Narkevic D. J., Crawford F. | 2007 | Science, 318, 777 |

This paper was built using the Open Journal of Astrophysics \LaTeX template. The OJA is a journal which provides fast and easy peer review for new papers in the astro-ph section of the arXiv, making the reviewing process simpler for authors and referees alike. Learn more at http://astro.theoj.org.