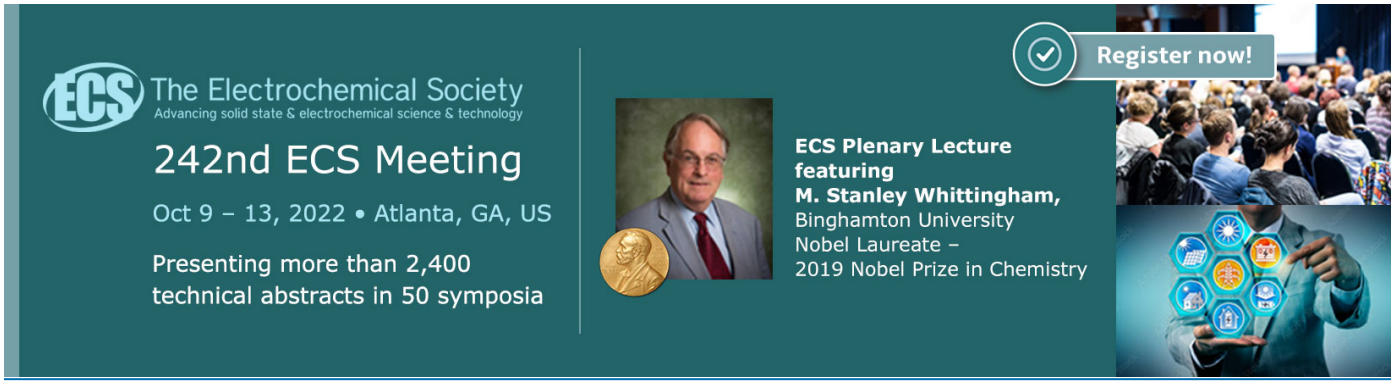
WindsPT e-Science platform for wind measurement campaigns



WindsPT e-Science platform for wind measurement campaigns

D F Gomes[1], J C Lopes[1,2], J M L M Palma[2], F Senra[1], S Dias[1], I L Coimbra[2]

[1] INESC TEC, Porto, Portugal
[2] Departamento de Engenharia Mecânica, Faculdade de Engenharia, Universidade do Porto, Porto, Portugal

Abstract. Experimental field campaigns for collecting wind data, essential for academic research and the wind energy industry, are non-trivial due to the complex equipment and infrastructure required. This paper reports the latest developments of the WindsPT e-Science platform for planning, executing, and disseminating wind measurement campaign data. Existing e-Science platforms have been developed for more generic domains, preventing them from capturing the details and requirements of the field. Additionally, we propose a protocol for transferring large volumes of data from the in-site devices to our platform, ensuring data replication. With an easy-to-use Web interface, WindsPT promotes collaboration between participants, disseminates results among the stakeholders, publishes metadata, uses DOI, and includes metadata that enables machine-to-machine communication. The platform has multiple sections, with maps, images, and documents, where there is information about the location of the stations, positioning of the sensors, operating dates, photos, technical sheets, calibration documents, among others. The WindsPT platform has been used to host the Perdigão 2017 experimental campaign and proved to be a valuable tool during all the phases of this large field experiment. A new version of WindsPT, designed to be FAIR, host multiple campaigns, and include multiple cross-campaign shared features, as full-text search capabilities, is now developed and tested.

1. Introduction

Experimental campaigns to assess wind and atmospheric conditions are critical for collecting valuable data for the academy and wind energy industry. However, these field campaigns are non-trivial to organise due to complex equipment and infrastructure required. Hence, before each campaign execution, careful planning takes place to define the intended data to be collected, the equipment to be used, and their layout. Nonetheless, up to this point, current working methods relied on handcrafted filling of spreadsheets exchanged through e-mail and other communication systems, which makes the planning process even more cumbersome.

Moreover, once the campaign ends, it is necessary to make the collected data available to the interested communities. Over the past few years, there has been an increasing concern about storing data, since the current paradigm is not suitable for easy data integration and reuse by the community after publication. To address this concern, science founders, publishers, and governmental agencies have pushed the development of a set of guidelines to increase management and stewardship for data generated in publicly funded experiments. In response,

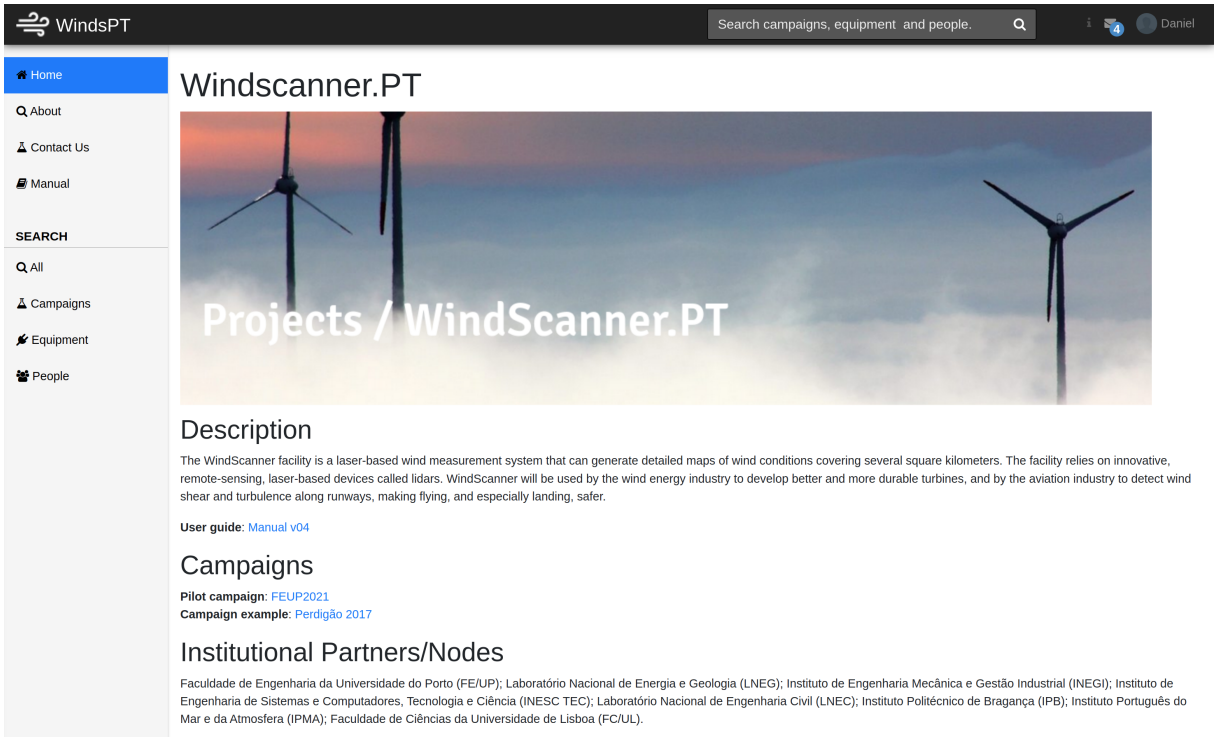


Figure 1. The WindsPT e-Science platform for planning, executing, and disseminating wind measurement campaign data.

a set of principles with the acronym FAIR (Findable, Accessible, Interoperable, and Reusable) have been gaining traction in Europe [13].

The FAIR principles go ahead of previous initiatives by focusing on humans and machines. Since humans increasingly rely on computational support to handle datasets in the expanding world of research, FAIR guiding principles offer a "how-to guide" towards machine-actionability, aiding researchers to find data of interest and reproduce its usage. Aiming to help data producers and publishers in making their data FAIR, e-Science platforms started to comply with these principles. Existing e-Science platforms up to this point have only been developed for more generic domains, or other than wind measurements, preventing them from capturing the details and requirements of this area. Thus, efforts were focused on the development of a platform that would help researchers and companies during all the steps of a wind measurement campaign. The design process of the currently called WindsPT e-Science platform is detailed in this work.

The Winds@UP [6] was the first prototype of the WindsPT e-Science platform, designed and prototyped during the *windscanner.eu*[1] European project [8]. The platform architecture was designed to ensure consistent and efficient research data access, as well as to foster new research opportunities provided by having a collaborative data infrastructure. This last goal was achieved by enabling the usage of the platform by users via a Web interface or by machines using an internal API (Application Programming Interface).

As a task of the Portuguese funded project, Windscanner.PT, the WindScanner e-Science and User Platform, another version of WindsPT, was developed (Figure 1). The platform, including the data server, was aligned with the policies on open access, as set out by the European Commission, to ensure the best possible dissemination of generated results, where access to research data was granted free of charge to users.

[1] https://www.windscanner.eu/

From these early prototypes, the WindsPT platform has been created to, beyond data management, handle personnel and equipment management, making it a handy tool for campaign planning, execution and data storage. The easy-to-use Web platform promotes collaboration between participants, disseminates results among the stakeholders, publishes the metadata, uses DOI, and includes meta-information that enables machine-to-machine communication. The platform was designed in conjunction with the Perdigão field campaign and was used to help the planning and to host the gathered field data (see https://perdigao.fe.up.pt). Recently, the WindsPT platform has been updated to be able to host multiple campaigns and to include cross-campaign shared features, such as full-text search capabilities (see https://windscannerpt.fe.up.pt and Figure 1) [4; 9].

2. Designing an e-Science platform for wind campaigns

To design the platform, the requirements elicited through interviews and study groups between stakeholders (researchers and software engineers) have been gathered in the course of several European projects: *windscanner.eu* (The European WindScanner Facility — FP7-INFRASTRUCTURES, 312372) from 2012 to 2015, NEWA (New European Wind Atlas Joint Programme — ERA-NET Plus, topic FP7-ENERGY.2013.10.1.2, NEWA/0001/2014) from 2015 to 2020, Windscanner.PT (Infraestrutura Nacional WindScanner — PINFRA/22207/2016), from 2017 to 2021.

After elicitation, the requirements were analysed by the Software Engineering team and the system features were described as *User stories*[2] and the description of the concepts of the Domain as *UML Class Diagrams*[3]. The non-functional requirements were also identified and are briefly detailed in the following paragraphs. The identified functionalities were prototyped in the platform and validated by interaction with the stakeholders of the platform.

During requirements specification, it was observed that within each stage of the campaign, requiring actions were needed to be executed by different actors: *campaign managers*, *data producers*, *researches*. Thus, the user's access privileges at WindsPT were constrained accordingly. For example, during planning, only *campaign managers* are able to register the campaign setup, e.g. stations and equipment. During execution, where data is gathered and curated, the datasets are registered in the platform only by *data producers*. Finally, the datasets are made available for open access by all researchers.

Since WindsPT aims to aid wind measurement campaigns, a set of tools related to all measurement campaign stages was included in the platform; these are tools for: planning, execution (registration, data collection, defining catalogues and datasets and their metadata) and open access dissemination (searching, consultation, and download). For the campaign planning, research groups could register their interest in participating in the campaign and contribute to the scientific equipment pool, allowing layout design and estimations of power and data transfer requirements. For the execution, data could be collected and stored on the platform, including dataset and metadata definitions, that later could be openly available to interested parties (open access). The following sections detail the design of the implemented features in e-Science platform.

2.1. Collaborative planning

The planning phase of an experimental campaign aims to answer, amongst others, the following questions: (i) where will the campaign take place?, (ii) what equipment will be necessary, and what electrical and data infrastructure will be necessary?, and (iii) who will participate in such a campaign?. The WindsPT platform offers key features that allow to register and visualise such

[2] https://www.reqview.com/doc/user-stories-template/
[3] https://www.uml.org/

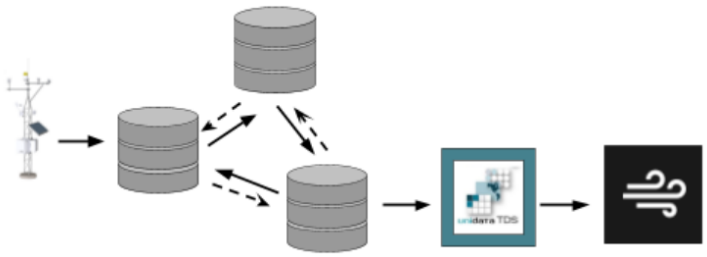


Figure 2. The data flow, from the equipment to the web portal: the data captured by the equipment and uploaded to one of any nodes of a predefined pool is then catalogued via a TDS server. Finally, the WindsPT web portal displays this information, with additional searching and download capabilities.

information to aid the planning and serve as contextual and meta information, which later can be used to better understand the context in which each dataset is collected.

The platform exposes two surveys to be answered by the managers of selected institutions invited to participate in a given campaign. These surveys ask (i) the partner institutions about their interest in participating in such campaigns and (ii) personnel and equipment (quantities and categories) made available to the campaign. The results of these surveys can then be viewed by the *campaign manager*, helping to guide the planning of the campaign layout. The surveys are better suited for more extensive campaigns and are intended to be answered months or years before one. Nonetheless, the information can be inserted or updated in a more compact, single-step form, better suited for smaller and quickly organised campaigns.

2.2. Equipment, station, and infrastructure registrations

A wide variety of equipment, from analogue to digital and from local to remote sensing, can be installed on-site, becoming stations used in a field campaign. Thus, two kinds of stations were created in WindsPT: remote sensors, e.g. Lidars and sodars, and conventional meteorological masts, with sensors connected to a datalogger for digitising and recording signals. In the registration part, the user can fill up information regarding the stations' location, sensor positioning, operation dates, photos, datasheets, calibration documents, etc., which is automatically stored and displayed in the form of maps, images, and sections in the platform, as exemplified in Perdigão use case (Section 4). In the campaign map, for example, the stations' locations can be visualised with different icons, according to the station type (illustrated in Figure 8), and exported to different formats.

2.3. Data collection and archive

Depending on the nature of each experimental campaign, the data collected by each sensor is transferred to a server of a pool of predefined servers (Figure 2). These are periodically synchronised using the Linux *rsync* protocol, ensuring the data is replicated across the pool, also minimising data transfer between nodes. One of such nodes hosts a THREDDS Data Server (TDS) that exposes the data through a basic HTTP (Hypertext Transfer Protocol) API (Application Programming Interface)[4]. In TDS, the gathered files and folders can be reorganised in catalogues and made available through different protocols: OPeNDAP (Open-source Project for a Network Data Access Protocol), OGC WCS (Open Geospatial Consortium Web Coverage Service), and OGC WMS (Open Geospatial Consortium Web Map Service), particularly for netCDF files.

[4] This project took advantage of netCDF software by UCAR/Unidata (http://doi.org/10.5065/D6H70CW6).

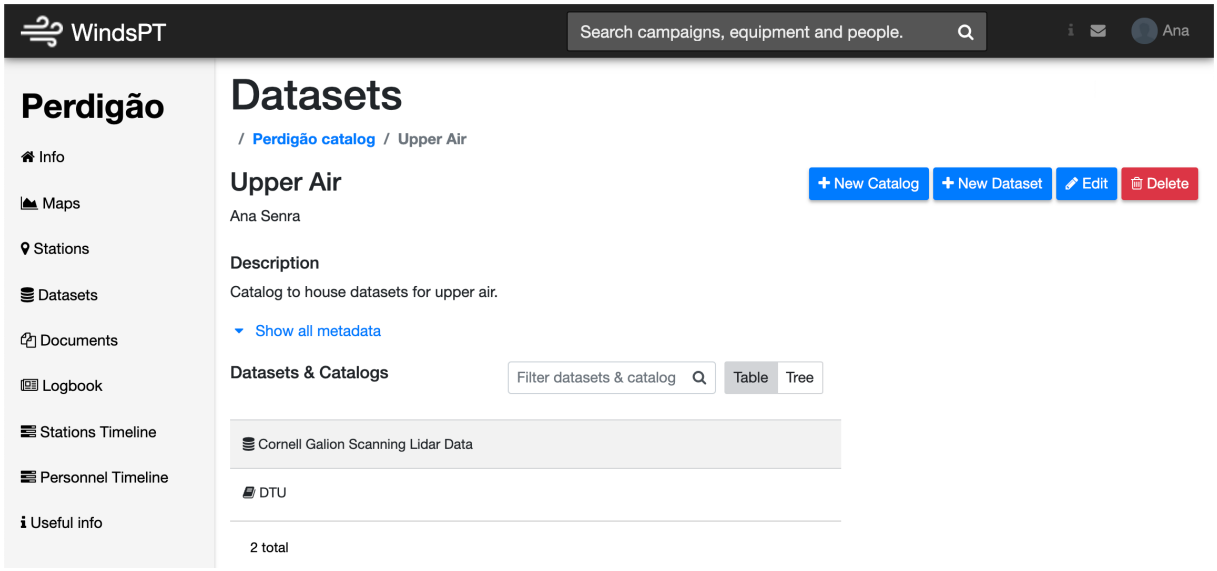


Figure 3. View of a catalogue with a sub-catalogue and a dataset.

TDS lacks in user experience and does not allow the connection of the gathered datasets to the remaining campaign information; thus, the WindsPT web portal offers its own dataset repository. Here, datasets can be registered, linked with folders at TDS, and annotated with the necessary metadata to ensure compliance with FAIR standards. Because these datasets are associated with the remainder platform database, WindsPT platform enables a researcher to interactively find the area and station where the dataset was collected, the details of the used equipment, as its reports, and even get in contact with the dataset author, through the internal messaging system.

2.4. Management of dataset catalogues

A subsequent implementation in WindsPT aimed to bring the platform closer to the FAIR principles [13], thus an internal study was conducted [11]. A subset of the identified changes was implemented and tested, where catalogues and datasets could be created or edited, as well as their metadata.

A catalogue can have datasets or other catalogues inside, as illustrated in Figure 3. It should be noted that the catalogues in the platform are independent of the catalogues available at the TDS data server, where derived data and curated data was stored. This separation allows a complete segregation between the structure for data collection and the structure available for data dissemination.

To create or edit a dataset or a catalogue metadata, the user fills a form with the requested data, as illustrated in Figure 4. For this, a controlled vocabulary based on Data Catalog Vocabulary (DCAT) [1] was developed to be used as meta-information for catalogues and datasets [11]. In addition to the metadata being searchable inside the platform, WindsPT provides easy metadata export (see Figure 5), suitable to be ingested in a standard metadata catalogue to be published.

When a dataset is created, it is associated with the station that produces it. The platform assists the user by providing the list of stations in the campaign. In this way, the relationship between the equipment and the data they produce is available for search in the platform.

To add contents to a dataset, the user can choose between all the available folders in the TDS server, as illustrated in Figure 6 with an example dataset [2].

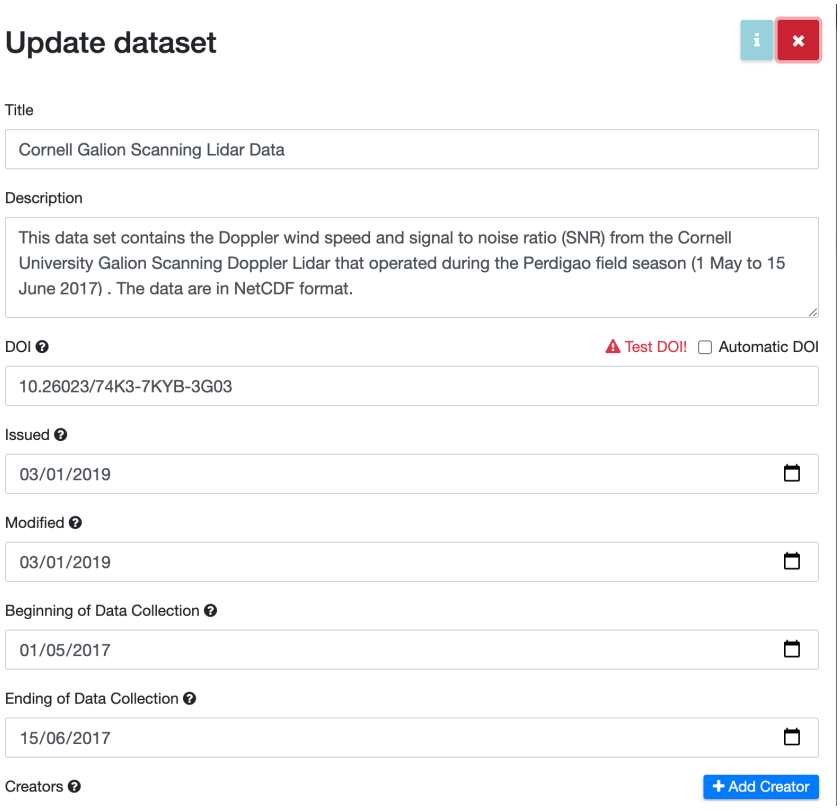


Figure 4. Form to create or edit a dataset.

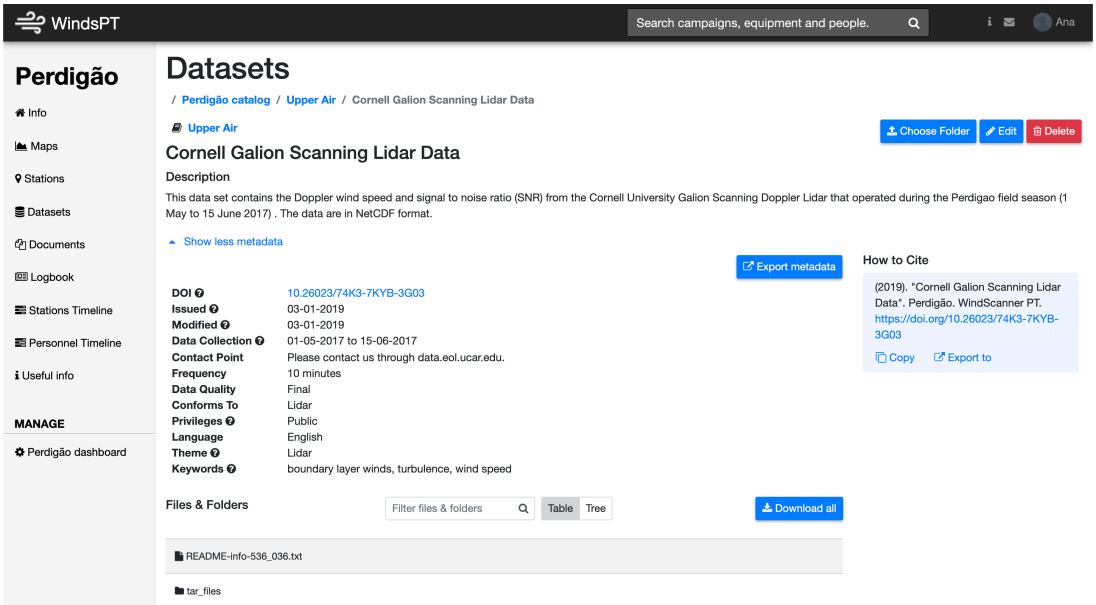


Figure 5. View of a Dataset with its metadata.

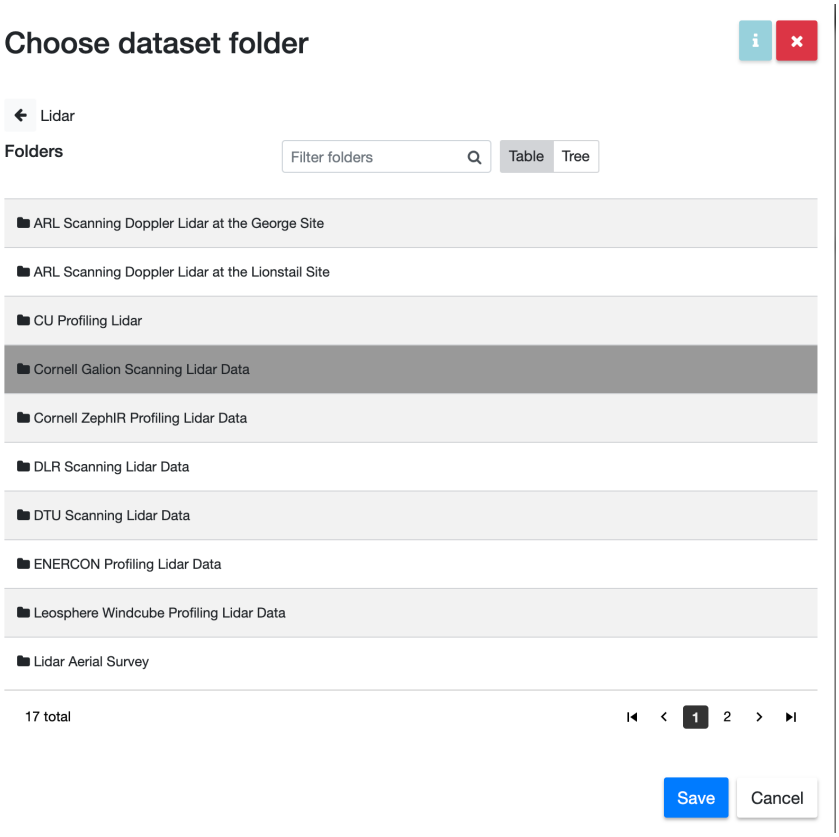


Figure 6. Form to choose the TDS folder of a Dataset.

We chose to keep this flow instead of allowing the user to directly upload files through the WindsPT platform due to the large upload times required for the significant files associated with a dataset. By having the files on a TDS server, the user can set up a folder in the personal computer that uploads the files in the background and later quickly create a dataset and add its metadata in the WindsPT platform.

If the user chooses to, the platform can publish a DOI for a dataset and manage the metadata associated with it automatically. We used the DataCite REST[5] API to publish and manage the DOI that the platform is responsible for.

To comply with the *Accessible* principle of FAIR, the metadata needs to be always accessible even if the contents are not. Thus, the metadata of a dataset or catalogue page in WindsPT is always accessible, even if the dataset or catalogue is embargoed, or in the case that authentication is required to acces the dataset. In this case, the contents are not visible by external users (only by managers).

2.5. Cross-campaign shared support features

While a campaign possesses individual goals, people and equipment can be common to multiple campaigns. In WindsPT, equipment and people databases are stored in their respective galleries, allowing the reuse of such information.

Besides data management, the WindsPT platform hosts many documents, datasheets, manuals, logbook posts, etc. Quickly finding a document on the platform from a set of

[5] https://en.wikipedia.org/wiki/Representational_state_transfer

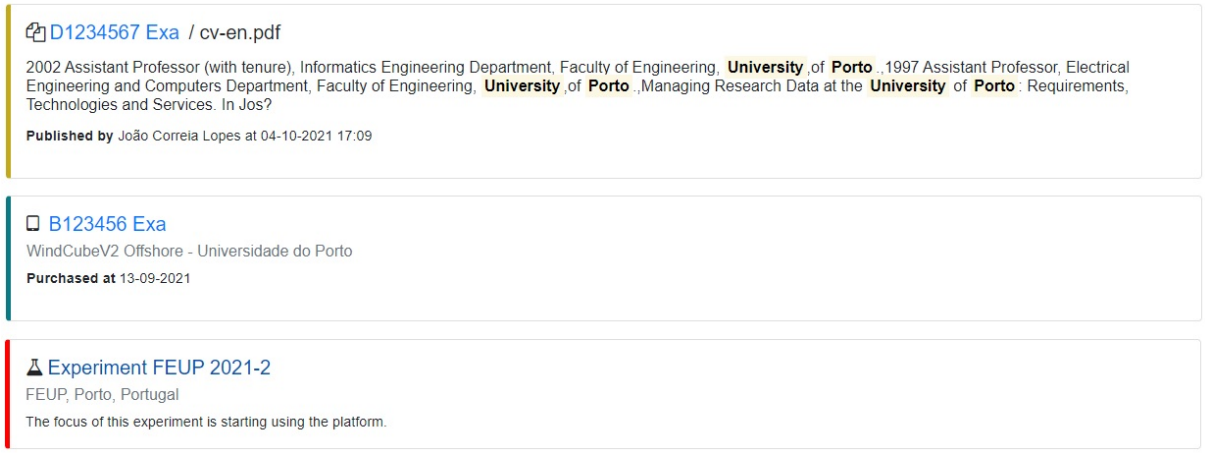


Figure 7. Excerpt from the search results returned for the query "University of Porto".

keywords is of utter importance. Thus, a full-text search (FTS) solution, based on Elasticsearch[6] and Lucene[7], is implemented to allow global searching and the retrieval of the most relevant documents related to a given information need. The platform's full-text search capabilities help find relevant information across the whole platform, such as in documents and logbooks. In a full-text search, a search engine examines all of the words in every stored document as it tries to match the search criteria (for example, the text specified by a user) and gives back a set of documents ordered by relevance. In addition, an Elasticsearch database and an FTS component were implemented for users to find relevant documents for both data and metadata.

A study [3] was conducted to select the set of attributes of the documents to be automatically stored in the Elasticsearch database. Also, an FTS component was implemented and can now be used to search for relevant documents for both data and metadata, as illustrated in the Figure 7.

3. Implementation

The e-Science platform for planning, execution, and dissemination of wind measurement campaign data, WindsPT, includes:

- surveys for gathering the interest of partner institutions, of particular interest for more extensive campaigns co-organized by multiple partners;
- the geo-registration of measurement stations and their equipment, of high value for a better understanding of the campaign itself and the gathered datasets; and
- data repositories for hosting the collected data and relevant documentation.

After the identification of the functional and technical requirements and the design of the corresponding solutions, the development took place spanning several sprints[8] to achieve the following results:

(i) design and development of the set of Web services (API), with Express[9];
(ii) development of the server-side logic to use the data server API, with Node.js[10];

[6] https://www.elastic.co/what-is/elasticsearch
[7] https://lucene.apache.org/
[8] https://en.wikipedia.org/wiki/Scrum_(software_development)
[9] http://expressjs.com/
[10] https://nodejs.org/

(iii) development of the web access Front-end to the API to manage the infrastructure and the campaigns, with Angular.JS[11];
(iv) development of the web accesses Font-end to the API related to the researchers's work, with AngularJS;
(v) enable the e-infrastructure to manage equipment, traceability and usages; and
(vi) design and support an *Application Profile*, based on DCAT[12].

The technical details of the methodologies and technologies used on the Computing Science domain, outside the scope of this paper, are available as project reports [3; 5; 11]

4. Perdigão use case

To demonstrate the platform, a use case is herein detailed. The platform was used in the Perdigão-2017 measurement campaign [4] and its full range of implemented features can be viewed at `https://perdigao.fe.up.pt`. This project was intended to provide the wind energy sector with more detailed resource mapping capabilities in a new digital EU wind atlas. Its primary goal was to quantify errors of wind resource models against a benchmark dataset collected in complex terrain. One other goal was to identify the physical and numerical weaknesses of models and discover methods to overcome such deficiencies [4].

4.1. How was the platform used in Perdigão and how was its performance?

The platform was used to gather collaboration intentions of the multiple partner institutions and build the catalogue of used equipment, collaboratively. Since multiple institutions joined the campaign during the planning phase, the platform needed to be adjusted to allow opening and closing surveys for each institution. In total, 24 institutions answered the surveys, committing to contribute with 10 towers, 328 tower-based sensors and 14 lidars.

Because different institutions utilised different vocabulary regarding the categories of equipment, following the surveys, discussions were held and the categories and subcategories gathered were pruned, resulting in the structure shown in Figure 8.

From the equipment information, the campaign managers decided on the station's layout and infrastructure necessary to assist such stations. The measurement cluster (Figure 8) consisted of 108 stations with 50 tower stations and 58 remote ground stations, e.g. LiDARs, microphones, etc.

For gathering the measured data, a network of data servers was established between the DTU[13], UCAR[14], and U.Porto[15]. The latter being the server of the WindsPT Perdigao platform and TDS server (see Section 2.3). The protocol described in Section 2.3 has been followed and the data made available on the web platform. During the *in-situ* phase, researchers collected their datasets using their tools and finally uploaded their data to one of the nodes, at their own pace. Datasets were periodically synchronised, every four hours, between the three servers of the network. In addition, through the dissemination of Perdigão datasets, several articles have already been published with results from this measurement campaign [7; 10; 12].

The details of each tower station are depicted in the 2D sketch diagram illustration, which contains the sensors deployed on the mast, as illustrated in Figure 9. The illustration can be accessed from the map with all stations, or from the datasets produced.

Different researchers also maintained their personal logbooks with news about the campaign. These have been periodically archived onto the platform logbook manually, by a *data curator*.

[11] `https://angularjs.org/`
[12] `https://www.w3.org/TR/vocab-dcat-2/`
[13] `https://www.neweuropeanwindatlas.eu/`
[14] `https://data.eol.ucar.edu/master_lists/generated/perdigao/`
[15] `https://perdigao.fe.up.pt/`

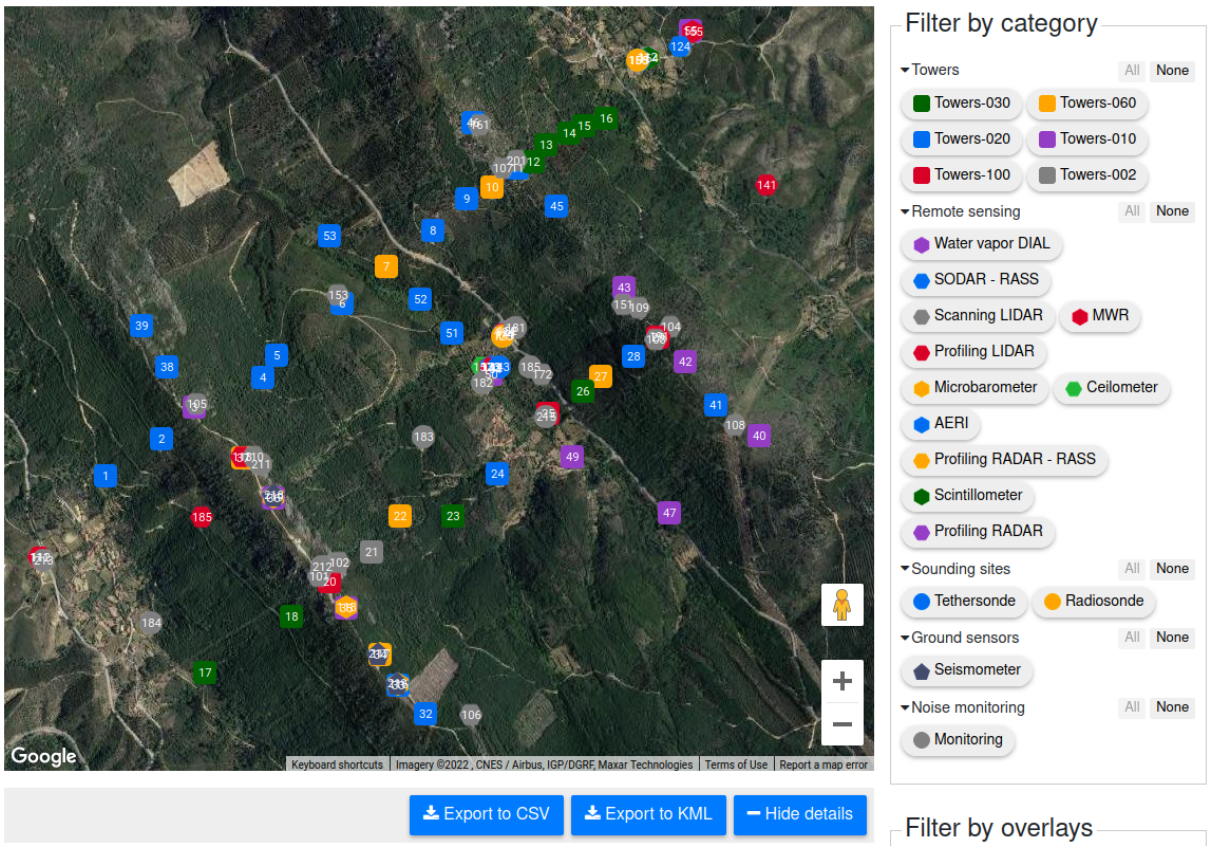


Figure 8. Map with the stations of the Perdigão's campaign. On the left, the planned stations can be seen on the map of the Perdigão's valley. Each station is marked with the icon that corresponds to its category.

One particular and unexpected set of events during the campaign were the fires at Pedrógão Grande. The platform logbook, together with e-mails served as a mechanism to broadcast timely and necessary information (see it with a search of "fires" or "18 june" in the logbook at https://perdigao.fe.up.pt/logbook). Today, and for the future, the full history of the Perdigão campaign can be followed by reading through the logbook.

5. Conclusions

The platform proved to be a valuable tool during all the phases of Perdigão field experiment, from planning to execution and further open-access dissemination of the collected data.

In the current WindsPT multi-campaign installation, for Windscanner.PT, some cross-campaign shared features were included. Further to the implementation of the described full-text search capabilities and the platform's authentication system, the internal messaging system enables that different campaign participants registered on the platform to talk to each other, fostering data and research collaboration.

Future work can also be done to add more tools and facilities in the e-Science platform, such as:

(i) discussion groups — to allow online discussion of the results, possible collaborations, ideas for projects, and a place where early-stage researchers could ask questions and get answers from more experienced researchers;

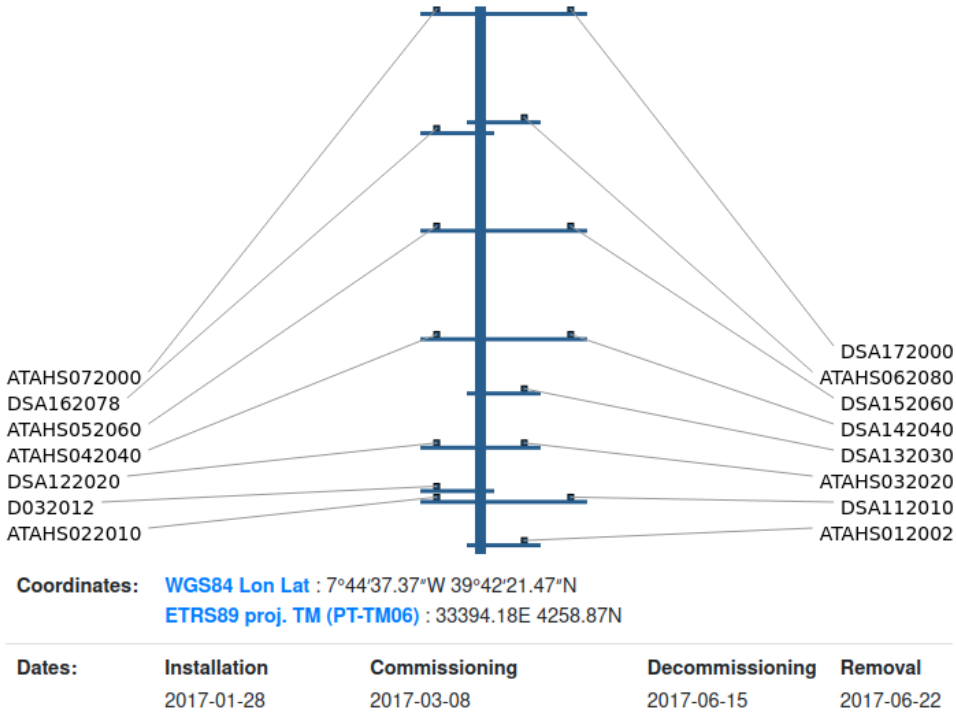


Figure 9. An example of a mast station with its equipment as displayed in WindsPT.

(ii) collaborative writing tool — to facilitate a joint work of a group of authors on publications;

(iii) collaborative data analysis — to allow multiple people to work together on the data analysis and, as it is connected to a well-designed measurement database, to simplify data interpretation;

(iv) code repository — to serve as an online resource of the codes of all the existing tools;

(v) an e-Learning tool with online courses, 'how to' examples, and exercises — to help initialise potential users to comprehend measurement technologies, for example.

Despite the improvements that can be introduced, the platform contains the necessary features to plan, execute and disseminate the results of wind campaigns. WindsPT is designed to be FAIR, host multiple campaigns and include multiple cross-campaign shared features, such as full-text search capabilities, and is now developed and tested.

Acknowledgments

This work was financed by National Funds through the Portuguese funding agency, FCT – Fundação para a Ciência e a Tecnologia within the project "Infraestrutura Nacional WindScanner" — NORTE-01-0145-FEDER-022207 (PINFRA/22207/2016). J M L M Palma and I L Coimbra are part of the LIdar Knowledge Europe (LIKE) project, H2020-MSCA-ITN-2019, funded by the European Union, Grant no. 858358.

References

[1] Riccardo Albertoni, David Browning, Simon Cox, Alejandra Gonzalez Beltran, Andrea Perego, and Peter Winstanley. Data Catalog Vocabulary (DCAT) — Version 2. Technical report, W3C Recommendation 04 February 2020, 2020.

[2] R. Barthelmie and Sara Pryor. Automated wind turbine wake characterization in complex terrain. *Atmos. Meas. Tech.*, 12:3463–3484, 2019.

[3] Sérgio Dias. Using the WindsPT e-Science platform in searching. Technical report, Task#3, *Windscanner.PT*, INESC TEC, 2021.

[4] H J S Fernando and et al. The Perdigão: peering into microscale details of mountain winds. *Bull. Am. Meteorol. Soc.*, 100(5):799–819, 2019.

[5] Daniel Gomes. Development of software architectures and tools for the *Windscanner.PT* infrastructure — Grant Activities. Technical report, Task#3, *Windscanner.PT*, INESC TEC, 2018.

[6] Filipe Gomes, J. Correia Lopes, José Laginha Palma, and Luís Frölen Ribeiro. WindS@UP: The e-Science Platform for WindScanner.eu. *J. Phys. Conf. Ser.*, 524(The Science of Making Torque from Wind 2014 (TORQUE 2014) 18–20 June 2014, Copenhagen, Denmark), 2014.

[7] R. Menke, N. Vasiljević, J. Wagner, S. P. Oncley, and J. Mann. Multi-lidar wind resource mapping in complex terrain. *Wind Energy Sci.*, 5(3):1059–1073, 2020.

[8] Torben Mikkelsen, Søren Siggaard Knudsen, Mikael Sjøholm, Nikolas Angelou, and Anders Tegtmeier Pedersen. WindScanner.eu — a new remote sensing research infrastructure for on- and offshore wind energy. In *Proceedings of the International Conference on Wind Energy: Materials, Engineering, and Policies (WEMEP-2012)*, 2012.

[9] J. L. Palma, M. Belo-Pereira, L. S. Leo, J. Fernando, N. Wildmann, T. Gerz, C. V. Rodrigues, A. S. Lopes, and J. Correia Lopes. Design of a data catalogue for Perdigão-2017 field experiment: establishing the relevant parameters, post-processing techniques and users access. In *AGU Fall Meeting Abstracts*, 2017.

[10] J. M. L. M. Palma, C. A. M. Silva, V. C. Gomes, A. S. Lopes, T. Simões, P. Costa, and V. T. P. Batista. The digital terrain model in the computational modelling of the flow over the perdigão site: the appropriate grid size. *Wind Energy Sci.*, 5(4):1469–1485, 2020.

[11] Filipa Senra. Making the WindsPT e-Science platform FAIR. Technical report, Task#3, *Windscanner.PT*, INESC TEC, 2021.

[12] J. Wagner, T. Gerz, N. Wildmann, and K. Gramitzky. Long-term simulation of the boundary layer flow over the double-ridge site during the perdigão 2017 field campaign. *Atmospheric Chem. Phys.*, 19(2):1129–1146, 2019.

[13] Mark D. Wilkinson and et al. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3:1–9, 2016.