Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

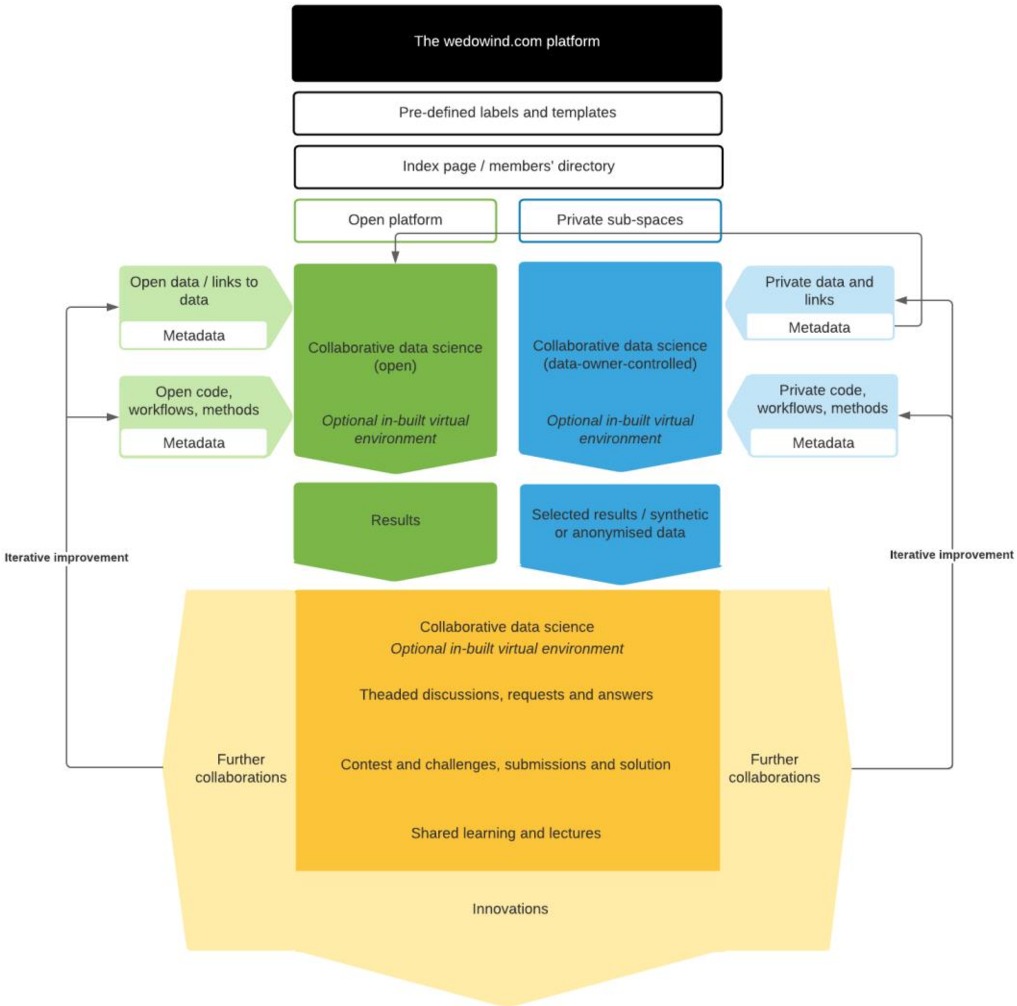
Federal Department of the Environment, Transport, Energy and Communications DETEC

Swiss Federal Office of Energy SFOE
Energy Research and Cleantech Division

Final report dated 31.01.2021

ML-Concept

Concept for The Swiss Wind Energy Machine Learning Platform – improving wind energy project planning and operation



Source: Sarah Barber 2020

Date: 31.01.2021

Location: Bern

Publisher:
Swiss Federal Office of Energy SFOE
Energy Research and Cleantech
CH-3003 Bern
www.bfe.admin.ch

Subsidy recipients:
The Swiss Wind Energy R&D Network
Oberseestrasse 10, 8640 Rapperswil
www.wedowind.ch

Authors:
Sarah Barber, OST, sarah.barber@ost.ch
Imad Abdallah, ETHZ, abdallah@ibk.baug.ethz.ch

SFOE project coordinators:
Katja Maus, katja.maus@bfe.admin.ch
Lionel Perret, lionel.perret@planair.ch

SFOE contract number: SI/502134-01

Summary

In this **report**, the development of a concept for The Swiss Wind Energy Machine Learning Platform is described. We firstly introduce the background and motivation, then define the initial requirements, explain how a suitable partner was identified, introduce the fictional user journeys, describe the implementation of more than 30 user interviews, then describe the development of a mini-demonstrator and finally introduce the final concept. **We have identified a very large interest in and high demand for a self-sustaining platform that provides a central location for wind energy industry and research to share and collaborate on reproducible workflows, data and code whilst respecting confidentiality requirements.** There is a very high potential for this platform and we are committed to continuing its development as quickly as possible, as described in the final chapter of this report. We plan a bottom-up approach, starting with the definition of a Minimum Viable Product and further developing the platform as we introduce more users and better understand their needs, as described in the final chapter.

The **concept** we have developed is **the first central collaboration platform in the wind energy industry that provides real incentives for participation**. It is a Wiki- and cloud-based platform with threaded discussions and searchable labels, allowing users to post requests, challenges and contests on one hand (data-owners), and answers, solutions and submissions on the other hand (algorithm/method-owners) on key wind energy topics, as well as datasets, lectures, information, standards, data standards, and more. The platform is divided into separate "open" and "private" sections, allowing users to share and showcase information on their own terms. It contains an optional docker environment for running and reproducing entire workflows in a standardised manner. The platform does not seek to replace any existing platforms; its power is in providing a central location for solving the most important wind energy industry challenges through sharing and exchanging selected information. For example, the datasets do not have to be stored on the platform (even though they can be), but a link or API to an existing platform can be provided, code can be either uploaded directly or provided via links to existing GitHub repositories, existing videos can be embedded, etc.. As this platform will provide users with a real incentive for participation, it will be able to **sustain itself long-term** by generating revenue from user membership fees, advertising, and fees for particular features. **We have changed the name of the platform to the *wedowind.com* platform, with a slogan "Accelerating innovation through knowledge sharing".**

By providing this central platform where all the information related to particular relevant wind energy challenges and topics can be stored, exchanged and improved in a reproducible manner, **the whole wind energy community benefits**. **Companies** benefit by getting solutions to their key challenges, by improving their workflows, by showcasing their work and by getting access to the best researchers. **Researchers** benefit by getting access to relevant datasets and real industry problems, by improving their solutions, and by getting access to industry partners. The **whole community** benefits as these collaborations lead to more efficient problem solving of the key challenges, ultimately reducing the cost of wind energy. In particular, **Swiss organisations** benefit from the increased opportunities for relevant learning and technology transfer from international and national players as well as collaborations that enable an easier entry in the global wind energy market. The **Swiss Energy Strategy 2050** could benefit from transfer of the ideas and concepts developed here to the entire Swiss energy branch, including tokenomics, incentive schemes, anonymous computing, data sharing and computing without compromising copyrights and authorships. We are already involved in plans to apply the solution to other fields, as well as for the Swiss Wind Atlas. The platform has particularly high potential for applications with a wide range of stakeholders. Other **benefits to Switzerland** in general include the improved visibility and reputation of Switzerland as an innovate country, technology transfer to other fields, including in societal, social and energy applications, and the improved acceptance of wind energy in the general public due to more media coverage and exposure to attractive and innovative projects.

Contents

1 Introduction

In this chapter, we examine the background information and the current situation, including a review of the state-of-the-art in relevant topics followed by a review of Swiss capabilities, and then introduce the project purpose and objectives.

1.1 Background information and current situation

1.1.1 Background

Digitalisation is one of the key topics in the wind energy industry today, due to its extremely high potential for increasing the efficiency of process, reducing costs and increasing revenue. This focus is well-reflected in the focus on the topic in the EERA JP Wind R&I Strategy[1], the SET Wind SET-Plan[2], the WindEurope events and webinars[3] as well as the recent formation of the new IEA Wind Task 43 "Wind energy digitalization"[4]. The IEA Wind Task 43 is focused on developing recommendations for the wind energy community on specific topics related to digitalisation, including data standards and data sharing, machine learning and AI, data analytics and visualisation, open source tools and IoT instrumentation. The entire wind energy community is currently searching for the most optimal way through the "digitalisation jungle".

Additionally, digitalisation is a key topic in Switzerland, especially in the energy branch. For example, the Digital Innovation Office of SFOE has recently developed a concept for a Swiss Hub for Energy Data (SHED)[5], which aims to introduce a national data hub with uniform and fair data access in order to improve the efficiency of billing processes and enable innovative business models beyond the energy sector.

This is therefore the ideal time for Switzerland to develop and introduce innovative tools to improve the digitalisation of the entire wind energy community. The topic of machine learning has a particularly high potential to be exploited in wind energy, for example in site-specific yield prediction, anomaly detection, predictive maintenance, wind atlas improvement and bird detection. Swiss wind energy projects could particularly benefit from machine learning due to the difficulty of analytically describing the effects of mountainous conditions such as complex terrain, turbulence, icing, etc. on wind turbine behaviour. However, the application of machine learning to wind energy is currently a new research field, and there is no coordinated effort to apply it in a targeted manner in the area of wind energy.

In the next section, the current state-of-the-art of the two relevant topics to this project, (1) machine learning and (2) sharing and collaboration platforms, are described, with a focus on the application to wind energy.

1.1.2 State-of-the-art in machine learning for wind energy

As mentioned above, machine learning has a high potential to be exploited in wind energy, especially for Swiss wind energy projects due to the difficulty of analytically describing the effects of mountainous conditions such as complex terrain, turbulence, icing, etc. on wind turbine behaviour. Some examples

[1] https://www.eerajpwind.eu/eera-jp-wind-ri-strategy/
[2] https://setwind.eu/
[3] https://windeurope.org/events/
[4] https://www.ieawindtask43.org/home
[5] https://pubdb.bfe.admin.ch/de/publication/download/10073

of previous applications of machine learning in wind energy include (a) Power prediction, (b) Fault / anomaly detection, (c) Forecasting, (d) Computational Fluid Dynamics (CFD), (e) Bird detection, (f) Virtual sensing on unmeasured systems. These are discussed below:

(a) Power prediction:

The accurate prediction of the expected power production of a wind turbine at a particular site is important in both the wind farm planning and operation phases. For wind farm planning, the power curve provided by the manufacturer (OEM) is used in combination with the expected long-term wind speed frequency distribution to calculate the expected Annual Energy Production (AEP), and therefore the expected Return of Investment on a project. However, the OEM's power curve is not site-specific, which can have a large effect on the accuracy of the results, especially in non-typical conditions such as mountainous terrain. Machine learning has been shown to have high potential for predicting site-specific power curves, specifically using regression trees [1], [2]. The method could be used by wind farm owners to improve wind turbine performance as well as the accuracy of their budget planning. As well as this, power curve modelling has been investigated using an Artificial Neural Network with multi-stage modelling techniques, showing that the method could be successfully used with up to six input parameters [3].

(b) Fault / anomaly detection:

Traditional fault detection systems such as adaptive filtering and acceleration enveloping rely on physical modelling of the wind turbine, which is very challenging. Therefore, data-driven methods based on measurements are highly attractive for the wind energy industry. Basic data-driven methods such as such as principle component analysis (PCA) cannot be applied due to the multiple factors that can induce an anomaly as well as due to the presence of the varying wind, to which the method should not be sensitive. Anomaly detection techniques that have been successfully applied to wind turbines include genetic algorithms [4], regression models and neural networks [5]. In [4], an anomaly identification model was developed using Supervisory Control and Data Acquisition (SCADA) data as input parameter. In order to increase the accuracy of the prediction model a Back Propagation Neural Network algorithm was applied. Gearbox and bearing temperature were used as state parameters. The proposed method was applied on a real 1.5 MW wind turbine, and the analysis shows that the proposed method is effective in wind turbine anomaly identification. In [5], Adaptive Neuro-Fuzzy Interference System models were employed to SCADA data in order to learn the "normal" behaviour in a training phase, where the component condition can be considered healthy. In the application phase the trained models were applied to predict the target signals, e.g. temperatures, pressures, currents, etc. Two examples of real measured faults show the capabilities of the proposed method. The method can be applied both to existing and new built turbines without additional hardware installation.

Other examples include: (a) analysing fault and alarm data from a turbine on the Southern coast of Ireland in order to identify periods of nominal and faulty operation, and then applying classification techniques to detect and diagnose faults by taking into account other SCADA data such as temperature, pitch and rotor data [6], (b) applying Deep Learning to condition monitoring in wind turbines, finding that the application of Deep Learning to SCADA data is impeded by their relatively low dimensionality, and we suggest ways of working with higher-dimensional SCADA data [7], (c) investigating classification techniques using clustering algorithms and principal components analysis for capturing fault signatures, showing that PCA and SOFM are both successful in producing persistent indicators that are well-distinguished from those generated based on no-fault data [8], (d) A based on machine learning and statistical process control tools applied to SCADA data of critical components [9], (e) applying the grey correlation algorithm to extract the eigenvectors of monitoring data and using the genetic algorithm and cross validation method to optimise the parameters of the support vector regression (SVR) model [10], and (f) applying SCADA data analysis methods for assessing the importance of how wind turbines align in patterns to the wind direction [11].

These methods can be applied to diverse components, including main bearing damage and blade leading edge erosion, as planned in this project.

(c) Forecasting

A similar application of machine learning to wind energy is to short-term forecasting, which is used for grid integration and trading purposes. Many studies have been done on this topic, and a review of all of them is beyond the scope of this report. A good review is available in [12]. This compares the main models based on physical methods, statistical methods and hybrid methods over different time-scales. Furthermore, this paper gives emphasis on the accuracy of these models and the source of major errors, thus problems and challenges associated with wind power prediction are presented. On example can be found in [13], in which a deep learning neural network was constructed to predict wind power based on a very high-frequency SCADA database with a sampling rate of one second.

(d) CFD:

Machine learning has recently been applied to wind energy applications including wake modelling and wind modelling in complex terrain, in which complex behaviour that can only be modelled accurately using high-fidelity Computational Fluid Dynamics is "taught" to a surrogate model such as Artificial Neural Networks in order to significantly reduce the required computational power [14]. This then allows optimization problems including wind farm layout design in complex terrain as well as wind farm control to be solved much more efficiently and accurately. In the future, this method could be applied for improving the Swiss wind atlas.

(e) Bird detection:

Machine learning methods can also be applied for bird detection and recognition, which could significantly increase the operation envelope of wind turbines operating in Switzerland. Previous studies have shown that deep learning methods are especially able to achieve very high recognition rates on remote monitoring acoustic data, despite weather noise, low SNRs, wide variation in bird call types, and even with mismatched training data [15]. A separate proposal is being planned on this topic.

(f) Virtual sensing on unmeasured systems (wind turbines components)

The load effects and degradation of structural components of wind turbines (WTs) do not uniformly evolve across a wind farm due to, largely, wake effects and variability in the inflow conditions (and waves for offshore wind farms). The assessment of fatigue damage accumulation on assumption of availability of direct load effects measurements on all main structural components across all wind turbines within a wind farm is not realistic. The assumption of availability of high-fidelity aero-servo-elastic simulators of the investigated wind turbines coupled to site-specific inflow and wake models is convenient, but often not borne out of the actual reality experienced by wind farm owners and operators. At best, the designer/manufacturer of the wind turbine might make a single comprehensive one-off set of output load effects of such wind farm simulations available to the owner/operator of the wind farm. A further and, from the view of this paper, perhaps more important hindrance lies in the precise estimation of structural response signals that are typically not available in the standard data emanating from the SCADA monitoring systems embedded on wind turbines. A number of works therefore attempt condition monitoring on the basis of SCADA availability. However, when structural monitoring information becomes available, then this can be exploited as a more direct proxy to diagnose sudden damage, or to further accurately assess the remaining useful lifetime of a wind turbine in a given wind farm. We therefore posit that considerable improvements to the operation, maintenance and prediction of remaining useful life of a wind turbine can be accomplished by delivering access to transparent, simple, yet powerful and interpretable data-driven machine learning based predictive models. Such models could be trained, tuned and updated via (fairly) easily accessible and cheap structural response observations from a limited number of appropriately selected wind turbines in a given wind farm.

1.1.3 State-of-the-art in sharing and collaboration platforms

The wind energy industry is not particularly well advanced on the topic of sharing and collaboration. For the purpose of initial site prospecting, wind data is available all over the world at a fairly low resolution and accuracy (Global Wind Atlas, New European Wind Atlas, Swiss Wind Atlas). However, beyond this, the wind energy industry is extremely secretive and the involved parties are very reluctant to share data at all. The wind turbine manufacturers (OEMs) are especially sensitive to giving away information that may allow reverse engineering of their wind turbines, and the wind farm operators often do not want to let the wind farm owners take the operation into their own hands, because this can increase the operator's internal costs.

However, due to the recent pressure on reducing wind energy costs and increasing revenue, especially offshore, the wind energy community is becoming particularly interested in benefitting from **digitalisation** by e.g. sharing data, information, standards, etc. Existing solutions and initiatives focused on this topic within and connected to the wind energy industry include:

- **Commercial solutions** offered to owners to optimise their wind farm performance include data platforms such as those from WinJi, Nispera and Breeze, analysis services such as Greenbyte and Sentient Science, as well as measurement devices such as RomoWind and Sereema. These services are strictly confidential, and do not allow sharing of data or algorithms between different organisations. This limits the learning potential immensely. Other industries are further ahead in this respect, including in the aircraft and gas turbine industries, and in the areas of building technology, solar energy, economics, demographics, geography and mobility[6].

- **User meetings** involving sharing experiences in wind farm operation, such as the ones organised by BWE in Germany[7], are a very good start, but tend to have very limited technical content and low transfer potential due to lack of external communication (observed experience).

- Limited amount of **open operational data** are a good start, but are not well coordinated. Also, the work done related to the data cannot be tracked and there is no possibility for easy multi-way communication between data and solution providers / among solution providers (in order to develop a common data cleaning solution, for example):
    - Limited operational data of 52 wind turbines in Brasil (only wind speed and direction, turbine power, rotor speed and nacelle position): https://zenodo.org/record/1475197#.XkJXuiNCfmF.

[6]https://opendata.swiss/de/

[7]https://www.bwe-seminare.de/

- Extensive operational data and maintenance log books as part of a public challenge "Wind Turbine Failure Detection" from the company EDP: https://opendata.edp.com/pages/challenges/description#description.
    - Operational data from the Engie Open Wind Farm: https://opendata-renewables.engie.com/.
    - ORE Catapult POD, the dissemination platform for operational data: https://pod.ore.catapult.org.uk/.
    - Extensive operational data and load measurements from the Alpha Ventus wind farm, available for research purposes after signing a user agreement: https://www.rave-offshore.de/en/data.html.
    - Atmosphere to electron database (The Data Archive and Portal, or DAP, was established by the U.S. Department of Energy Office of Energy Efficiency and Renewable Energy's Wind Energy Technologies Office to provide secure, timely, easy, and open access to all laboratory, field, and benchmark model data and offshore data produced by the Atmosphere to Electrons (A2e) initiative): https://a2e.energy.gov/data.
    - DSWE dataset: the datasets used in the book, Data Science for Wind Energy, are publicly available, a total of 150 MB real wind farm data and 400 MB simulated data are organized into ten datasets. A description of the ten datasets can be found at the website: https://aml.engr.tamu.edu/book-dswe/dswe-datasets/.
    - CREW: Sandia National Laboratories offers a wind energy database through the Continuous Reliability Enhancement for Wind (CREW) project: https://energy.sandia.gov/new-crew-database-receives-first-set-of-data/
    - Orsted Offshore Operational Data (a data portal where a direct data request may be made. "If sufficient information is supplied in the application, the project proposals will be reviewed by internal specialists and the applicant will be informed of the result."): https://orsted.com/en/our-business/offshore-wind/offshore-operational-data.

- Limited amount of **open measurement data** on the sharewind.eu platform, which has recently been developed as part of the EERA JP Wind programme, for example:
    - High-resolution full-scale wind field measurements of the ECN's 2.5 MW aerodynamic research wind turbine using DTU's 3D WindScanner and SpinnerLidar for IRPWind's and EERA's benchmark (https://sharewind.eu/records/aae047f75fb64b9894ad5b0e07fe6cd8).
    - Lidars for Wind Tunnels (L4WT) and Lidar Data for Use (LD4U) - Joint experiment and open data set within the IRPWind project (https://sharewind.eu/records/a1c2ef96a1e446fa9bd7a2a46d2242d4).

- **Benchmarking services** involving submitting data and receiving results in terms of anonymised comparisons and indicators, which is very useful for operators but is limited in scientific value due to the generic indicators:
    - "webs" from the Offshore Renewable Energy Catapult in the UK: https://ore.catapult.org.uk/stories/wind-energy-benchmarking-service/.
    - The DNV GL GEMINI platform: https://ore.catapult.org.uk/stories/wind-energy-benchmarking-service/.
    - The WInD-POOL from Fraunhofer IEE: https://wind-pool.iee.fraunhofer.de/wind_pool_de/WInD-Pool/.
    - SPARTA trend analysis project: https://www.sparta-offshore.com/SpartaHome
    - O2O platform (planned): https://www.esi-africa.com/industry-sectors/renewable-energy/wind-turbine-owners-launch-database-to-improve-performance-standard/.

- **Existing energy data portals** and initiatives are very useful but appear to be too general for much of wind energy community to be interested in participating in:
    - Open Energy Data Initiative (OEDI) from the US government (providing access to open data from DOE-funded projects): https://openei.org/wiki/Open_Energy_Data_Initiative_(OEDI)
    - Open energy data on the European Data Portal: https://www.europeandataportal.eu/data/datasets?categories=ener&page=1&locale=en&minScoring=0
    - European Open Science Cloud (an environment for hosting and processing research data to support EU science): https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/european-open-science-cloud-eosc_en

- **General** sharing and collaboration platforms can be very useful but are difficult to use for collaborations due to the very broad topic scope and the functional focus (e,g, only data storage or only code sharing or only contests, etc.):
    - Zenodo (allows for the storage and sharing of large data sets, data sets are associated with a unique digital object identifier): https://zenodo.org/.
    - GitHub/GitLab (version control platforms allow programmers to share code, public commit history allows for a permanent record of a code through unique commit hashtags, large data files are discouraged): https://about.gitlab.com/. Examples of useful openly available GitHub contributions include:
        - The "windpowerlib" is a library to model the output of wind turbines and farms: https://github.com/wind-python/windpowerlib.
        - "Energy Wind Farm Estimation" is a tool to estimate energy production of a wind farm. Furthermore, the option of wind data analysis is possible: https://github.com/mpizosdim/EnergyWindFarmEstimation.
        - The "Wind Energy Toolbox" is a collection of Python scripts that facilitate working with (potentially a lot) of HAWC2, HAWCStab2, FAST or other text input based simulation tools: https://github.com/DTUWindEnergy/WindEnergyToolbox.
    - Kaggle (data science contests with prizes): https://www.kaggle.com/.
    - Papers With Code (sharing code with the associated science manuscript): https://paperswithcode.com.
    - Collaborative data science platform for reproducible science (by the Swiss Data Science Center): https://renkulab.io/
    - MatrixDS provides a state-of-the-art collaborative platform. This provides real-time access to tools and infrastructure for building data-focused solutions effectively and efficiently: https://matrixds.com/.
    - Collaborate on data science workflows and publish the work: http://www.melda.io/.

- Data **standards** and **semantics** are a key part of data sharing and collaboration, providing a common language for sharing data and saving enormous amounts of time having to clean and understand data. However, no common definition yet exists and there is no common place to disseminate and share the work on this topic:
    - Metadata and related wind energy semantic created during IRPWind (2017) https://data.windenergy.dtu.dk/ontologies/view/wtax/en/.
    - Data catalogue of IEA Wind Task 41 on Distributed Winds built up by customizing to the IRPWind taxonomy: https://www.vindenergi.dtu.dk/english/research/research-projects/iea-wind-tcp-task-41.
    - Wind ontology developed in IEA Wind Task 37: https://data.windenergy.dtu.dk/ontologies/view/en/.

    - Universal data standard for Wind Resource Assessment developed as part of IEA Wind Task 43: https://www.ieawindtask43.org/work-package-4-digital-wra.

- International collaborations via **IEA Wind Tasks** have the aim of aligning research activities between international groups, and differ greatly in their effectiveness. Some very fruitful collaborations results from these tasks; however the main difficulties are: (1) the slow progress due to lack of funding (many people attend but not a lot gets done because the work is not funded specifically), (2) the lack of knowledge transfer and transferability, (3) the low participation of companies, (4) the lack of focus on real industry challenges (because most participants are from academia.

These rapidly growing number of initiatives and solutions indicate the increasing interest in trying to improve information sharing and collaboration. Although we can learn from the existing initiatives, many do not provide a large enough incentive for participation, and there is no coordination or consolidation of all of them. Additionally, none of the existing initiatives allow a real collaboration over the entire workflow (data, code, application, analysis, understanding, comparison, further development) specifically for the wind energy community. This strongly limits the learning potential. Our concept therefore focuses on bringing existing solutions together, rather than replacing them, and providing real incentives for collaboration.

1.1.4 Swiss capabilities

There are many Swiss organisations with relevant experience in machine learning and data sharing. This includes:

- **Utilities owning wind farms** (e.g. BKW, IWB, EKZ, ewz, Axpo, etc): as far as we have been able to establish, all these companies are greatly involved in the digitalisation of their workflows, including the usage of data platforms to consolidate all their data as well as machine learning algorithms for data analytics. The amount of internal vs. external development varies greatly between companies, but many employ external services such as WinJi and Nispera to do this for them. These companies are highly interested in our project (see Section 2.4.).
- **Data platform providers for the wind energy industry** (e.g. WinJi, Nispera, Sqooba (Open Systems)): these companies provide platforms for wind farm owners to track all their wind turbines and wind farms in one place, and provide some additional data analytics services. These companies work with the Swiss utilities but also with international partners. They are all very interested in our project (see Section 2.4).
- **Data platform providers for the other industries** (e.g. Misurio): as well as the companies mentioned above, who don't just work in the wind energy industry, there are many other companies offering similar services to other industries, such as the photovoltaic industry. We haven't yet spoken to these companies but we expect their interest in an expansion to other applications to be high.
- **Data analytics / big data / digital providers** (e.g. Cognite, Akselos, Frontwerks, Statworx, BSG Group, Frox, Siemens, IBM, Yokogawa, SOREAL, etc.): these companies have general experience providing data analytics and big data services to many different industries. We haven't yet spoken to these companies but we expect their interest in an expansion to other applications to be high.
- **Machine learning research groups at all Swiss universities:** there are too many research groups involved in machine learning at Swiss universities to list here, but examples include the wind energy group at OST (Sarah Barber) and the Chair of Structural Mechanics and Monitoring at ETHZ (Imad Abdallah), as well as the Swiss Data Science Center at EPFL. All Swiss universities are increasing their activities in this area rapidly. For example, OST has just launched a machine learning competence centre.

- **Other organisations:** Opendata.ch is a non-profit association committed to making data public and easy to access and use, and are very interested in working with us. The Swiss Data Science Center aims is to accelerate the use of data science and machine learning techniques within academic disciplines of the ETH Domain.

As well as this, SFOE has recently published a concept for a Swiss Hub for Energy Data (SHED)[8], which aims to introduce a national data hub with uniform and fair data access in order to improve the efficiency of billing processes and enable innovative business models beyond the energy sector. We see a large potential for connecting our planned platform to the SHED, and generalizing it for the entire energy market. Additionally, many parties could benefit from working together on key topics such as standardised and fair data. Our concept is not competing with this idea, but complementing it by focusing on collaboration and interaction between people and real industry challenges.

To summarise, it can be seen that there is a large amount of experience in the area of digitalisation, data sharing and machine learning in Switzerland. The Swiss Wind Energy R&D Network is an ideal position to work with all of these organisations and create a central platform for collaboration related to machine learning and wind energy.

1.2 Purpose of the project

The purpose of this project is to develop a concept for The Swiss Wind Energy Machine Learning Platform for coordinating activities on the application of machine learning for improving wind energy project planning and operation. As will be explained in this report, the idea has since then developed into a web-based data sharing and collaboration platform, which we have re-named to the *wedowind.com* platform. "We do wind" is the slogan of The Swiss Wind Energy R&D Network and at the same time reflects the collaborative nature of the platform. The main focus is still on machine learning solutions, but is not limited to this topic.

1.3 Objectives

In order to develop the concept, the project consists of the following objectives:

- Definition of the initial requirements.
- Identification of a suitable partner.
- Definition of user journeys
- Implementation of user interviews
- Development of a mini-demonstrator.
- Development of final concept.
- Definition of a plan for future development.

[8] https://pubdb.bfe.admin.ch/de/publication/download/10073

2 Procedures and methodology

In this chapter, we firstly describe the definition of the initial requirements, then the identification of a suitable partner, the definition of user journeys, the implementation of user interviews, then the development of a mini-demonstrator. The final concept is presented in the Chapter 3, and the future development plan in Chapter 5 (Chapter 4 presents the conclusions).

2.1 Initial requirements

The initial requirements were defined in order to initiate a dialogue about the platform. They are expected to be developed or changed based on the results of the user journeys and user interviews. The platform is based on the fact that data owners need machine learning methods and method owners need data, resulting in a win-win situation for both parties, as shown in Figure 1.

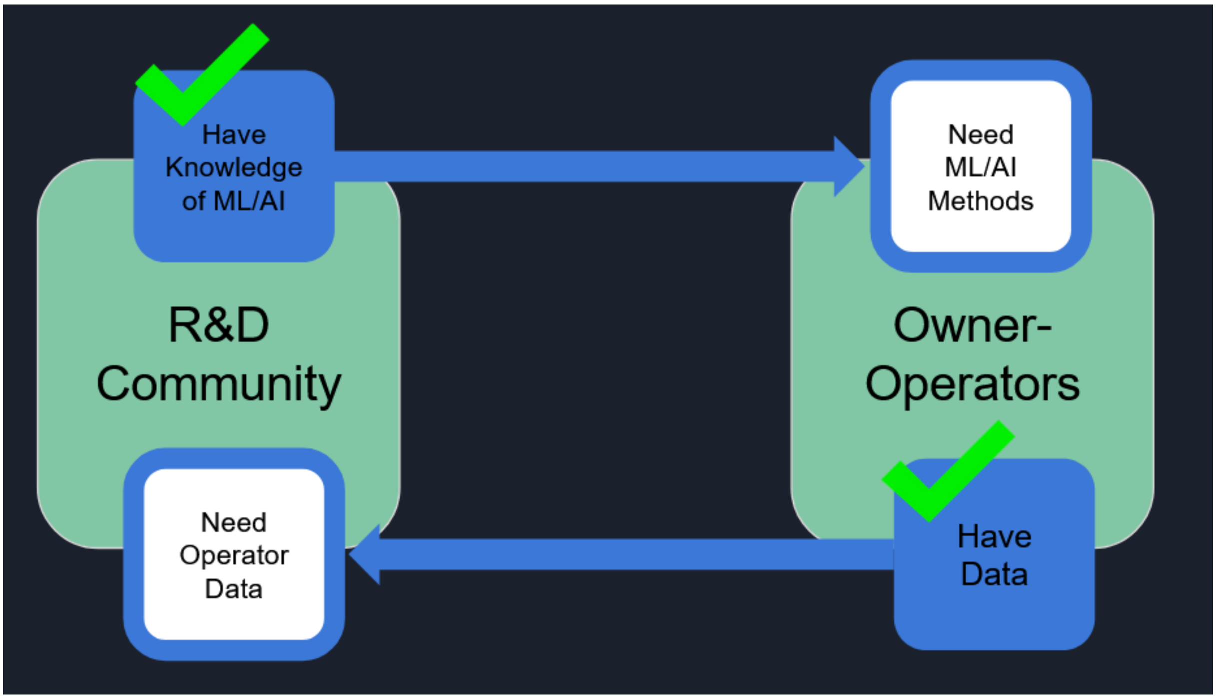


Figure 1. Win-win situation for both parties.

The initial idea of the platform that was used in order to visualise these requirements is shown in Figure 2. This shows data owners on the left and workflow owners on the right. The data and workflows have been split into "private", "selected access" and "open access" categories. The "private" information does not actually get stored on the platform, only its metadata (i.e. a description of it). The data and workflows get combined to create new results and innovations, which can be published on the platform either fully publically, partially publically or privately, depending on the requirements of the users.

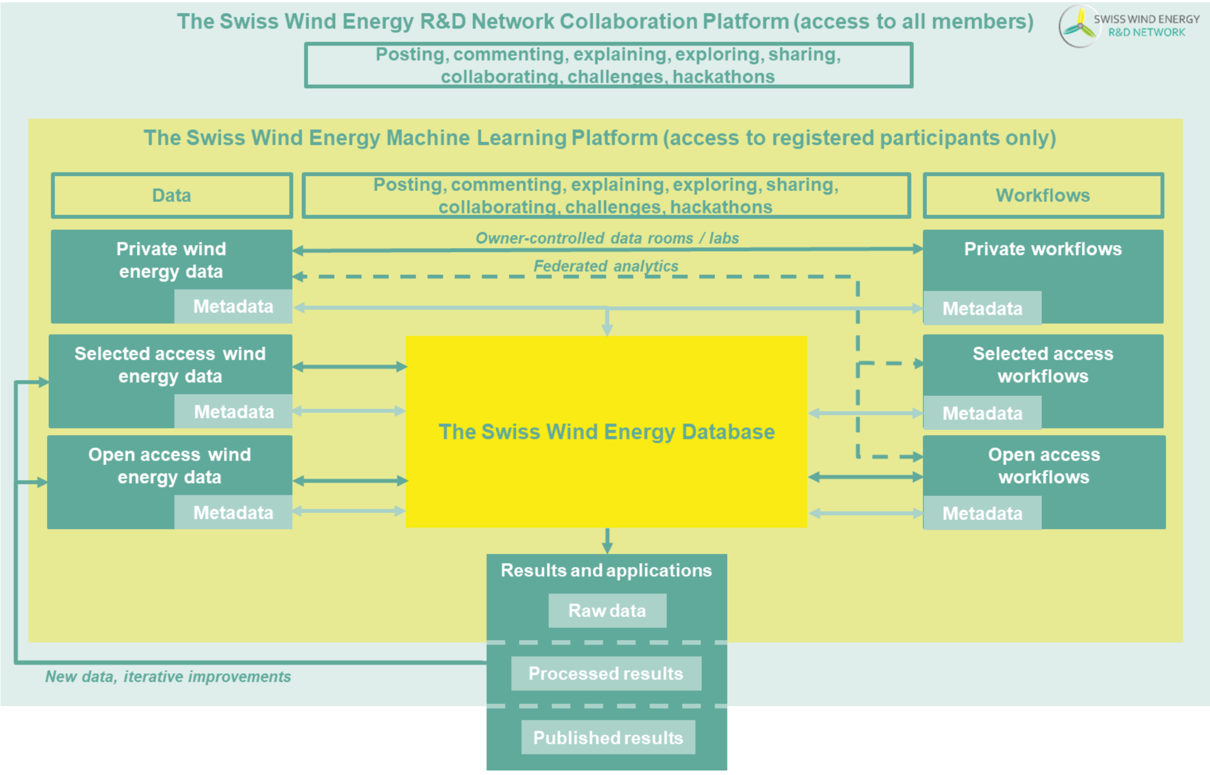


Figure 2. Overview of initial concept.

The initial requirements defined were:

- Provide a central location for wind energy industry and research to share reproducible workflows, data and code in order to boost innovation whilst respecting confidentiality requirements of all parties (the focus is not on hosting large amounts of raw data, but facilitating remotely hosted data on client servers via APIs).
- Embedded in a wiki-based collaboration platform, to which all registered participants have access, which allows posting, commenting, explaining, exploring, sharing, collaborating, posting challenges, hosting hackathons etc. (the name may change in the future).
- Allow data and workflows to be connected to the platform either as "private", "selected access" or "open access":
    - Private data and workflows are not in the database, but connected to each other via owner-controlled data rooms and/or federated and featured analytics. The metadata is searchable and requests for access can be sent. Synthetic data sets can be generated out of these datasets and made accessible to the other users.
    - Selected access data and workflows can only be accessed by pre-defined groups. The data can be uploaded or links can be set up to existing databases. The metadata is searchable and requests for access can be sent.
    - Open access data and workflows can be accessed by registered users. The data can be uploaded or links can be set up to existing databases. The metadata is searchable.
- Provide access to different types of results (= combinations of data and methods) depending on access settings, allowing reproducibility.

- Workflows can include algorithms, notebooks and manual steps. Data cleaning, preparation and standardisation is dealt with through sharing workflows themselves.
- Be financially self-sustaining.

2.2 Identification of suitable partner

This concept is being developed by The Swiss Wind Energy R&D Network, consisting of wind energy experts and researchers, not of computer scientists. It was therefore decided that an external partner is required to develop the details of the platform itself.

We discussed a collaboration with the following partners:

- The OST Digital Lab put collaborates with the international company Cognizant. During the summer, we developed a concept with them, but the main focus was on the database itself and the proposed costs were very high.
- We had a few meetings with the company Sqooba in Switzerland, who already work together with IWB on a data platform for their power stations, including wind farms. However, they did not have enough resources to pursue the idea with us.
- We had some meetings with the company Octue in the UK, who already work together with OST on the development of a Digital Twin for a different project. However, they did not have enough resources to pursue the idea with us.
- The most fruitful initial conversations were with the company Stacker Group in the USA, who supports the clean energy sector by constructing data systems required to connect and analyse data at fleet scale. They build modern data platforms in a fraction of the time and cost required by traditional approaches, by using low-code techniques that minimize the amount of code written. For this reason, they have led the buildout of numerous data systems for leaders in the renewables space, generating efficiencies while saving money. They have vast experience on the development of networking and data sharing platforms, amongst others in the wind energy industry.

It quickly became clear that Stacker Group were the ideal partner for this project, despite the fact that they are not a Swiss company. The whole development of this concept presented in this report has been done in close collaboration with them. We have met them on average once a week since September and they have participated in a large proportion of the user interviews described in Section 2.4. They have already developed a tool called Relight, which they apply for collaboration purposes in a number of different US cities, and which enables effective collaboration based around people, data and workflows. They are committed to creating inclusive spaces for fostering innovation. The concept developed in this project is based on their Relight platform, and involves developing it further for our needs. They fully understand our needs in terms of bringing people together with data and workflows, and their inputs have led to the transformation of the initial "data sharing platform" idea into an innovative central collaboration platform. No other potential partners got close to understanding our needs beyond the data storage technology part.

The two roles of Stacker Group in this project are:

- **Software developers:** Up until now they have contributed "in kind", but in the future development stage we would need to pay them for their development time, as well as platform licensing costs.
- **Business adviser:** The CEO of Stacker Group, Charles Henderson, has advised us on the business development side a great deal. In the future development of this solution, we have

both agreed that we would like him to continue in this role, as his insights and contacts are vital for the project).

Even though Stacker Group has been a great support to us in this project, we recognise that it may be preferable and more convenient to find a suitable Swiss partner. We plan to start discussions in the next few weeks with companies such as Adranis in order to evaluate their suitability, now it is more clear to us exactly what we need.

2.3 Definition of user journeys

Before carrying out user interviews, we developed a set of fictional user journeys in order to help us further develop the concept. For this, we defined the types of user shown in Table 1. The journeys of each user type are shown in the next figures, followed by a summary in Table 2. The characters in these journeys have been freely invented by us and do not represent real people!

Table 1. Summary of fictional users for user journey.

User group	Data contribution	Workflow contribution	Access type
1a. Small wind farm owner/operator	SCADA	No	Private / selected access
	CMS	No	Private / selected access
1b. Large wind farm owner/operator	SCADA	Yes	Private / selected access
	Advanced measurements	Yes	Private / selected access
2a. Technology developer / data scientist (inspections)	SCADA	Yes	Private / selected access
	Specific measurement data	Yes	Private / selected access
2b. Technology developer / non-data scientist (data platform)	SCADA	Yes	Private / selected access
	Specific measurement data	Yes	Private / selected access
3. Researcher	Simulation results	Yes	Open access
	Experimental data	Yes	Open access
4. Small WTG manufacturer	Measurement data	No	Private / selected access

2.3.1 User journey 1a: Small wind farm owner/operator (Figure 3)

- **Lea** works at a small wind farm owner (utility) and is responsible for data analysis of running and planned wind farms.
- She has access to a large amount of SCADA data from the company's 20 operating wind farms, and the company recently implemented an internal database that collects all the data in the same format and carries out a basic analysis of the monthly performance in order for her to present the results in an understandable way to her manager.
- She has limited resources and spends much of her time compiling and comparing the monthly performance figures. She is an engineer but not a programmer, and prefers to carry out any further analyses that are required in Excel.
- She is under pressure from her manager to create added value for the company from her data analysis, and has always wondered how machine learning could help her work more efficiently.
- She is already a member of The Swiss Wind Energy R&D Network and recently signed up for the collaboration platform in order to follow posts related to wind farm performance.

2.3.2 User journey 1b: Large wind farm owner/operator (Figure 4)

- **Melanie** works at a large wind farm owner (utility) and is part of a large centralized data science team.
- She has access to a large amount of SCADA, CSM and other data from the company's 1'000s of operating wind turbines, and the company recently spent a lot of money on implementing an internal database that collects and stores all the data in the same format. She and her colleagues work on several different projects related to different wind farms in the portfolio.
- Melanie is a talented data scientist and has a Masters' degree in wind energy engineering. She feels like she spends about 80% of her time sorting through and understanding the data relevant to the current project. The rest of the time she gets to actually develop code and analyse the wind farm performance, developing recommendations for person responsible for the wind farm.
- She is under pressure from her manager to create added value for the company from her data analysis, and gets frustrated at the amount of time required to sort and clean data.
- She is quite proud of her own abilities and is not sure whether she can really learn anything from a bunch of researchers who usually focus too much on the theoretical side rather than take into account the needs of the industry. However, she signed up for the platform because she believes that collaborations and sharing improve innovation.

2.3.3 User journey 2a: Technology developer / data scientist (inspection) (Figure 5)

- **Leonie** works at a technology developer in Switzerland, who offer blade inspection services with drones. She is a data scientist and is responsible for applying machine learning algorithms and writing code that interprets the images taken by the drone.
- The company has just started operations, and have successfully managed to acquire a number of projects around the world. However, their workflows are still fairly basic and Leonie is struggling to keep up with the amount of data she is being delivered with by the drones.
- Leonie is under pressure to show her manager that surface damage can be detected successfully and economically by their system. Her workflows have to deliver results, and she

often has to reduce the quality of her workflows and their degree of automation in order to save time and invest most of her efforts in the data analysis itself.

- She feels like she spends 80-90% of his time cleaning and correcting data. Even though most of her data comes from the company's own drones, the drone version is different for every project and the additional data (e.g. SCADA data and log books) that is required varies in format and quality immensely between projects.
- She and her company are very unsure about the benefits the platform can bring them. They feel like they may lose their competitive advantage if they share too much information (data and algorithms). However, Leonie signed up to become a member of The Swiss Wind Energy R&D Network as well as for the new data platform in order to at least follow posts related to wind farm performance in the hope of snapping up some new information.

2.3.4 User journey 2b: Technology developer / data scientist (data platform) (Figure 6)

- **Daniela** works at a technology developer who offer a data platform helping wind farm owners and operators to access all their data in one place. Recently, they have also been developing basic data analysis methods for their customers, too. Daniela is a data scientist without a background in wind energy. She is mainly responsible for developing the data analysis methods.
- The company has recently started operations, and have successfully managed to acquire a number of projects around the world. However, their workflows are still fairly basic and Daniela barely has time to keep up with the day-to-day operations, never mind to develop new methods.
- Daniela is under pressure to show her manager that the data analysis methods provide an added value to the customers. Her workflows have to deliver results, and she often has to reduce the quality of her workflows and their degree of automation in order to save time and invest most of her efforts in the data analysis itself.
- She spends a lot of time compiling and trying to understand the data, which is sometimes difficult due to the lack of information she is provided with and her lack of experience in the wind energy industry.
- She and her company are very unsure about the benefits the platform can bring them. They feel like they may lose their competitive advantage if they share too much information (data and algorithms). However, Daniela signed up to become a member of The Swiss Wind Energy R&D Network as well as for the new data platform in order to at least follow posts related to wind farm performance in the hope of snapping up some new information.

2.3.5 User journey 3: researcher (Figure 7)

- **Maria** works as a postdoc at a university in Switzerland, and has a background in wind energy aerodynamics. She is involved in an exciting project developing a novel aerodynamic measurement system for wind turbine blades. Her work involves experiments and simulations related to the pressure distribution on wind turbine blades. She also develops workflows for analysing and utilising the measurement data in the best way.
- The project has just started, and Maria is a fan of open innovation. She is happy to share her data and workflows and is convinced that this will be beneficial to the project. However, her manager doesn't entirely agree with her, because the funding body for this project requires a business plan for a new product to be developed.

- She has recently conducted a new set of wind tunnel tests and simulations of a 2D wind turbine blades. She is developing a workflow for identifying blade damage, including leading edge erosion and surface cracks.
- Maria is under pressure to deliver scientific content as well as recommendations for the measurement system design. Despite the fact that there are some companies on the Advisory Board of the project, she feels like she is not in touch enough with the "real world" and is not sure about the applicability of her work in the industry.
- She spends a lot of time compiling and trying to understand the measurement and simulation data, as well as trying to work out how to turn the results into valuable scientific content.

She and her university are very excited about this new platform and think they can benefit immensely from exchanges with companies as well as with other research groups. Maria signed up to become a member of The Swiss Wind Energy R&D Network as well as for the new data platform and is very keen to get actively involved.

2.3.6 User journey 4: small wind turbine (WTG) manufacturer (Figure 8)

- **Laura** works at a small WTG manufacturer technology manufacturer who sell wind turbines for urban spaces. They also offer simple Wind Resource Assessments and data analysis services to their customers. Laura is a highly motivated, recently-graduated engineer with decent skills in data science and wind energy. She is part of the engineering team in the company. The other team members are CAD and design specialists.
- The company is a fairly new SME, and have recently installed a number of their new wind turbine for some local customers. Their workflows are still fairly basic and they do not yet have a lot of experience with the operating wind turbine. Laura barely has time to keep up with the day-to-day operations, never mind to develop new methods.
- Laura is under pressure from her manager to analyse the data from the operating wind turbines in order to understand how well the wind turbine operates in real conditions, with the goal of recommending design changes to the rest of the team. She has developed some basic analysis algorithms, but thinks there is a high potential to improve them and is not sure if the parameters she is examining are the best one.
- She and her company are very positive about the new collaboration platform. They hope to learn about the most effective data analysis techniques and are happy to deliver data in return for this.

2.3.7 Summary of user journeys

The user journeys allowed us to start to shape the platform directly towards the needs of the potential user type. A summary of the findings, the consequences for the platform development, and the actions required are shown in Table 2.

Table 2. Summary of findings from user journeys.

Finding	Consequence for platform	Actions
Users need to find documents/entries related to particular topics.	The platform should have a way of labelling documents / entries.	Including labelling,
	Topics should be searchable.	Include a search function.
Entries need to be consistent.	Moderate the platform and provide templates.	Develop templates.
Private messaging should be possible.	Allow users to create private spaces.	Create private spaces functionality.
It should be possible to send private files.	Allow users to create private spaces.	Create private spaces functionality.
Submitted data should be labelled correctly.	Have a consistent way of describing the files.	Create a meta-data template.
Users need to be motivated to "contribute" as well as "take".	Rule imposing contribution quotas based on usage?	Make benefit of contributing content more clear.
		Define a business model.
Users would like to post challenges.	Challenge announcement capability.	Label entries with "challenge" or "solution"
	Challenges should be easy to post.	Standard contracts / procedures?
Users would like to post collaboration requests.	Collaboration request capability.	Label entries with "request" or "offer"
Requests or challenges should actually result in a benefit (= results).	Include clear goals with deadlines.	Provide template for challenges and requests.
Some organisations may only contribute anonymised data.	It should be easy for contributors to anonymise data.	Automatic anonymisation function
Users may like guidelines for project funding once their ideas have been tested.	Consider including funding guidelines or links to them	Consider including funding guidelines or links to them

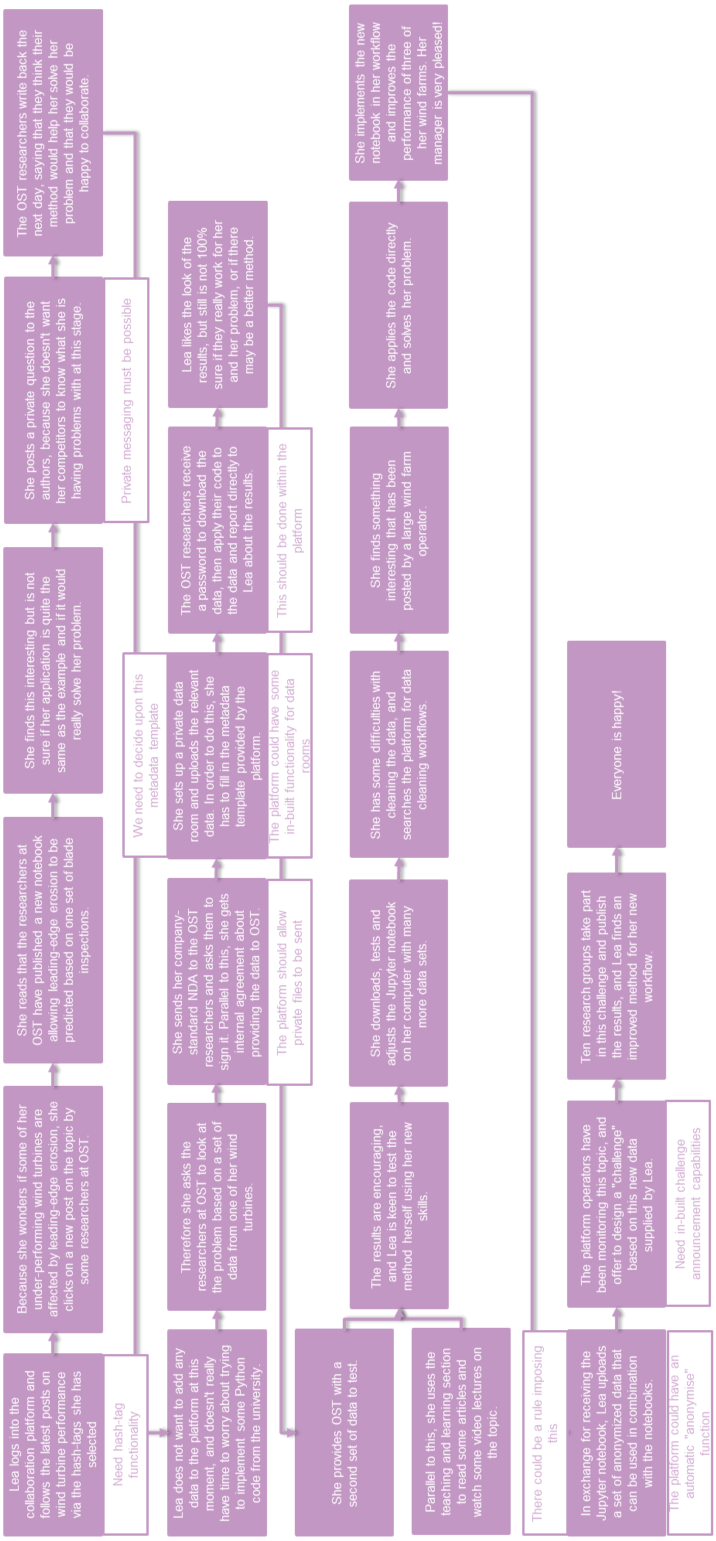


Figure 3. User journey 1a: Small wind farm owner/operator.

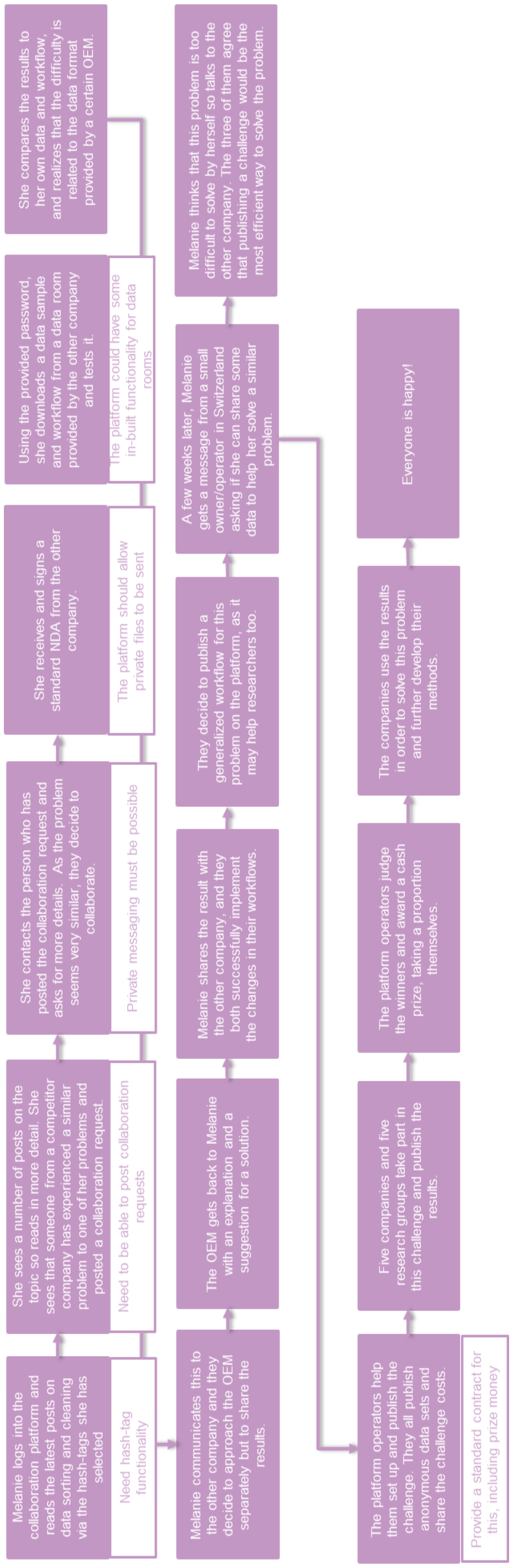


Figure 4. User journey 1b: Large wind farm owner/operator

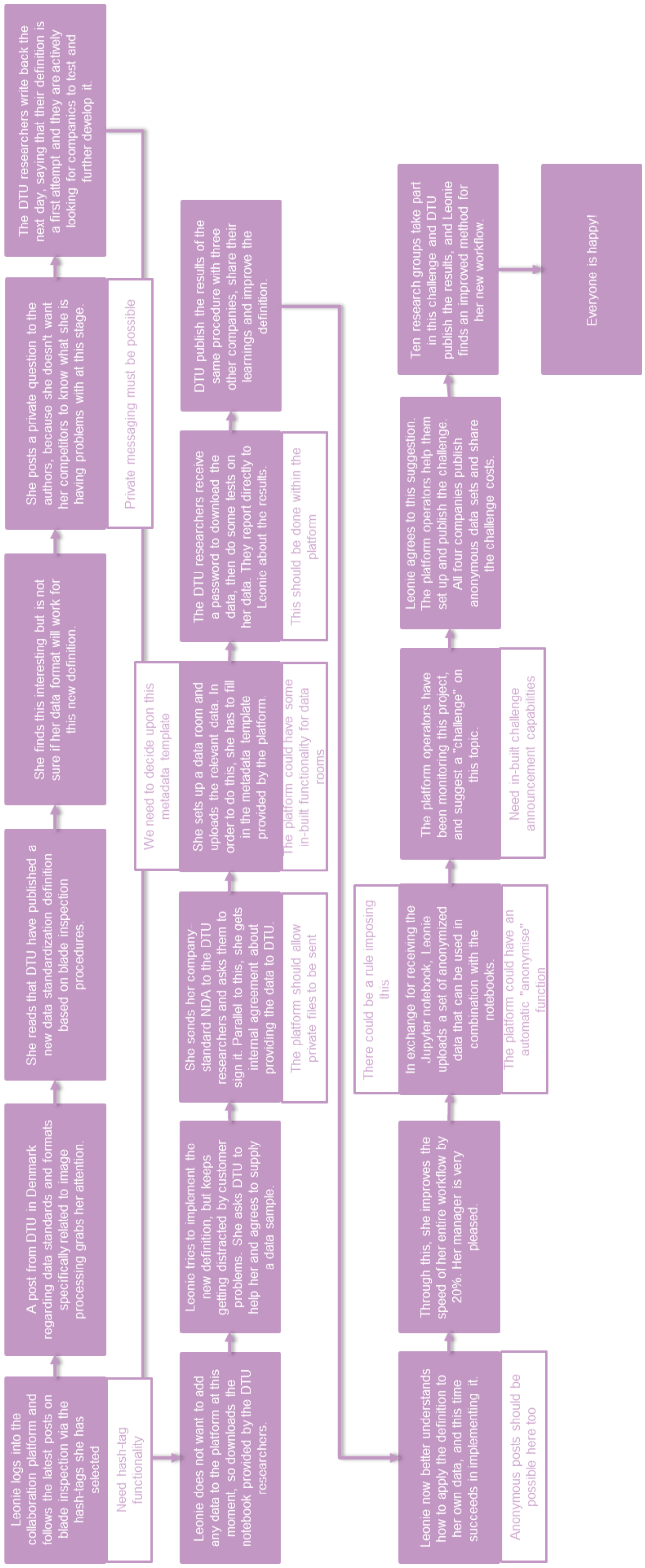


Figure 5. User journey 2a: Technology developer / data scientist (inspection).

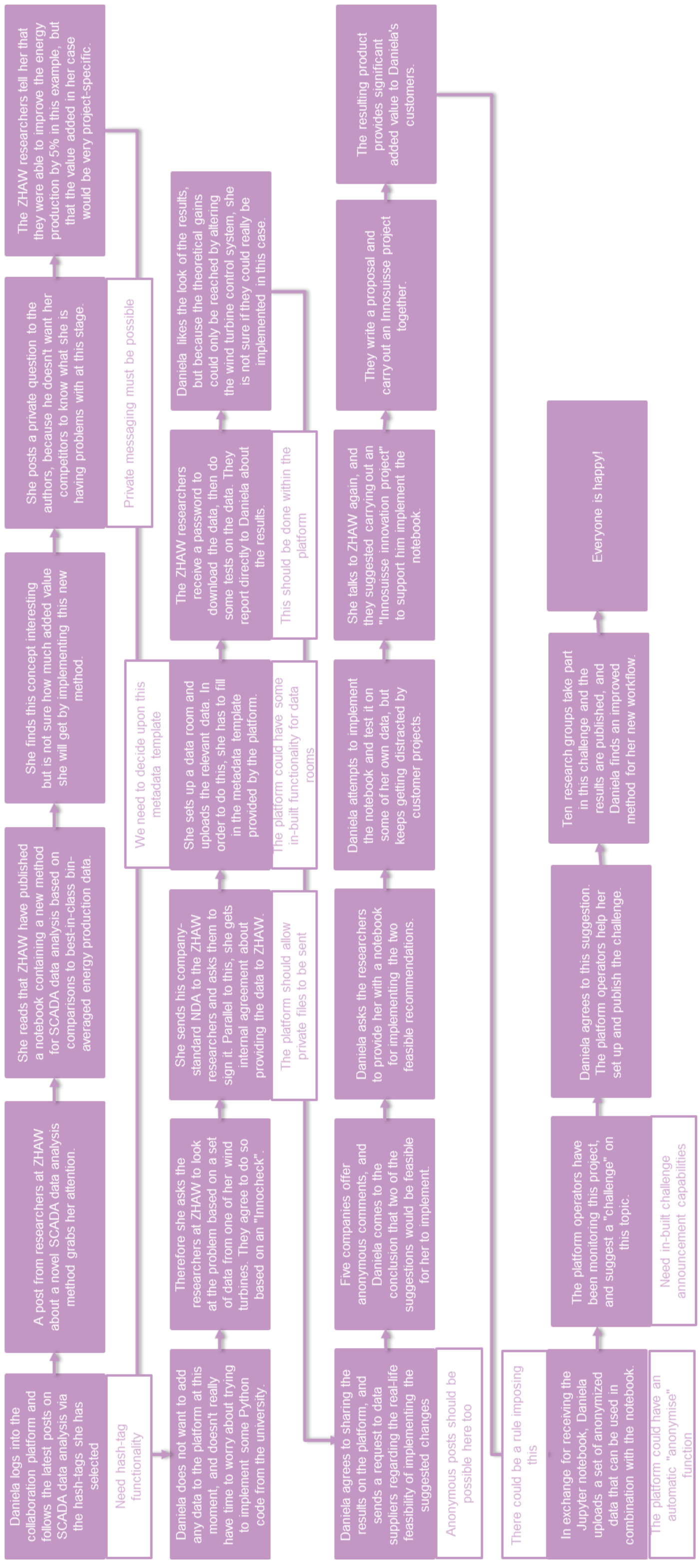


Figure 6. User journey 2b: Technology developer / data scientist (data platform)

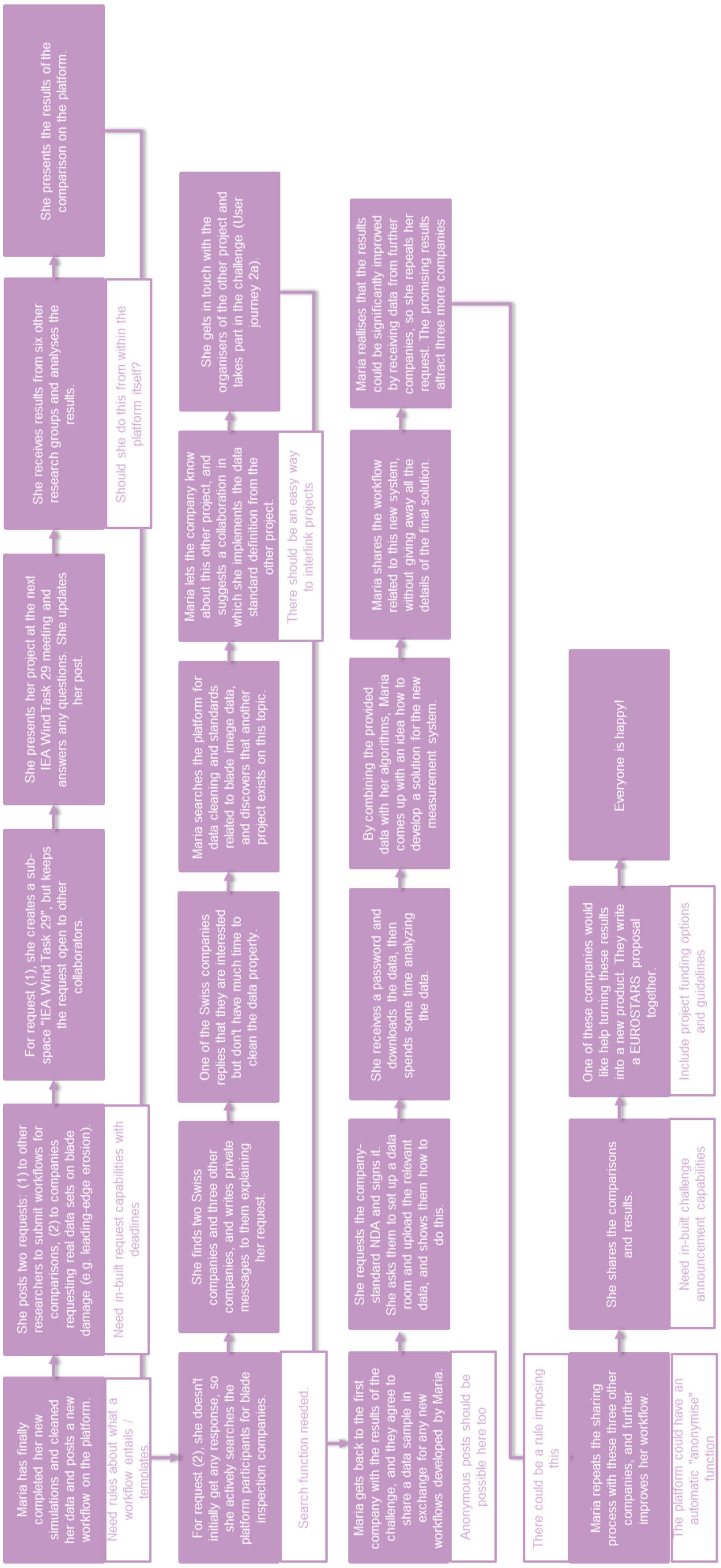


Figure 7. User journey 3: researcher.

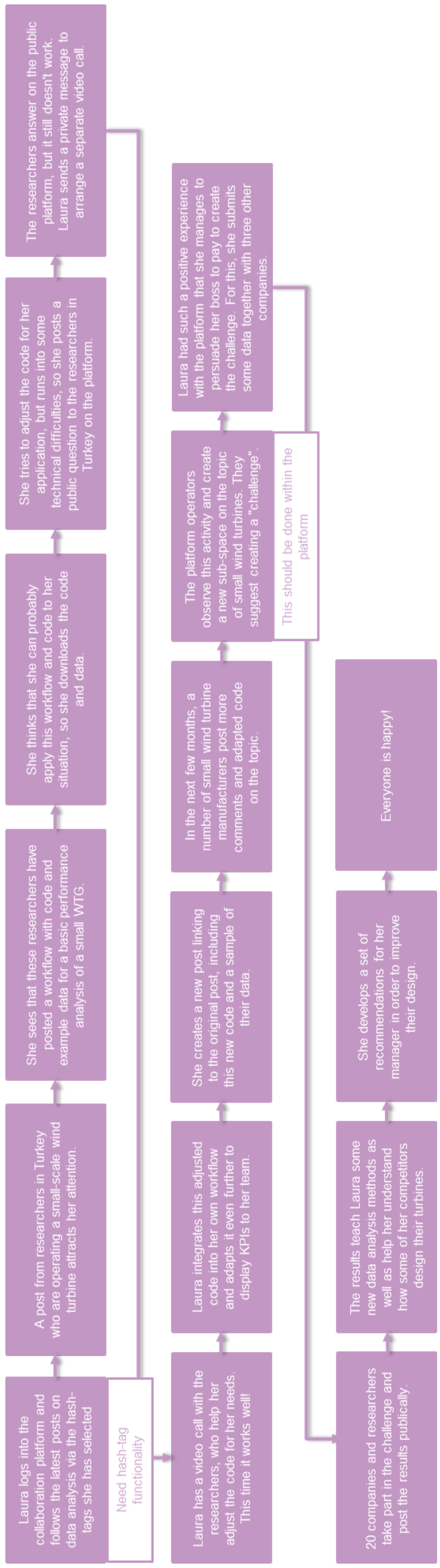


Figure 8. User journey 4: small wind turbine (WTG).

2.4 User interviews

A total of 34 user interviews were carried out with both Swiss and international potential users as shown in Table 3. In these interviews, we presented the latest concept and asked the following questions:

- What sort of data do you deal with?
- What are your main challenges related to handling this data?
- What are your motivations for sharing data and algorithms?
- What is your feedback about the platform? How could you imagine using it?
- Would you be interested in submitting a use case for the next stage of development?

Table 3. Summary of all the user interviews.

Date	Organisation	Organisation type	Discussion length
21.10.2020	ENBW	International company (DE)	1 hour
26.10.2020	Orsted	International company (DK)	1 hour
02.11.2020	Meteotest AG	Swiss company	10 minutes
02.11.2020	Basler & Hofmann	Swiss company	½ hour
17.11.2020	IWB	Swiss company	1 hour
18.11.2020	NREL	International research group (USA)	1 hour
19.11.2020	DTU	International research group (DK)	1 hour
19.11.2020	DNV GL	International company (DE)	1 hour
19.11.2020	Uni Stuttgart	International research group (DE)	1 hour
20.11.2020	EDP Renewables	International company (PT)	1 hour
20.11.2020	Emerson	Swiss-based international company	1 hour
23.11.2020	SKF	Swiss-based international company	1 hour
23.11.2020	NRG Systems	International company (USA)	1 hour
23.11.2020	Ins Santa Catarina	International research group (BR)	1 hour
24.11.2020	ZHAW	Swiss research group	1 hour
24.11.2020	Fraunhofer IEE	International research group (DE)	1 hour
27.11.2020	WinJi AG	Swiss company	1 hour
30.11.2020	GE Renewables	Swiss-based international company	1 hour
30.11.2020	Hitachi ABB	Swiss-based international company	1 hour
01.12.2020	Sulzer Schmid	Swiss company	1 hour
01.12.2020	Skypull SA	Swiss company	1 hour
01.12.2020	Open Data CH	Swiss company	1 hour
01.12.2020	Gilytics AG	Swiss company	1 hour
02.12.2020	LM Wind Power	International company (DE)	½ hour
02.12.2020	EKZ	Swiss company	½ hour
04.12.2020	Nispera AG	Swiss company	1 hour
07.12.2020	Aventron	Swiss company	1 hour
14.12.2020	EDP Renewables	International company (PT)	1 hour
14.12.2020	Anerdgy AG	Swiss company	1 hour
18.12.2020	BKW	Swiss company	1 hour
11.01.2020	RES	International company (UK)	1 hour
14.01.2020	Susi Partners	Swiss company	1/2 hour
21.01.2020	Fraunhofer IWES	International research group (DE)	1 hour

The main findings and corresponding actions are summarised in Table 4 in terms of the overall concept, the business model and marketing.

Table 4. Summary of findings from user interviews.

Findings ("concept")	**Actions**
Some users really believe that sharing data and code improves their innovation, whereas some are not open to sharing and want to use it as a commercial platform	Cater for both: (1) open part, (2) commercial part.
How do we "make" people interact and not just download the data one-way?	Provide them with incentives: • Showcase their work. • Compare codes/data. • Find partners • Provide an automatic anonymizing service, provide templates / data standards • "Choose" them as moderators (get them to pay)
How do we make the "private" part more attractive than the current situation?	• Access to "everyone" in one place. • Provide standard NDAs? • Moderate actively between open and private spaces.
Some organizations may want to actually use the platform for data storage and visualisation	How to include a simple GUI? (open source?)
The effort of data providers should be minimized.	Meta-data template? Tbd!
Users need to be able to see what data can be used for / how workflows can be applied	Label data: what is it suitable for?
Entries need to be divided into categories such as "challenges", "solutions", "offered", "requested", etc.	Define labels
The content needs a good quality control	Include quality control in the tasks of the platform owners.
The availability of external data needs to be monitored and maintained	Can the platform track availability and accessibility of external data?
Data protection needs to be dealt with	Find out what needs to be done (Google Cloud is certified according to ISO/IEC 27001)
The challenge providers are interested in an automatic evaluation of the submitted results	Automatic evaluation could be a later development.

Findings ("concept")	**Actions**
We are solving "real" problems, and not "nice to haves" like in hackathon	Make this differentiation clear
Data protection may be an issue for some European users	A European Google cloud may not be enough. Check this.
How exactly does the virtual environment work?	Need more clarity on the virtual environment
The platform owners should help users prepare and anonymise their data	
The platform should not produce MORE work!	Think about how to interface with existing platforms such as Predix, EDP, etc.
Emphasise the importance of knowledge sharing and multiplying	Include this in slogan?
Larger companies will not want to share their core knowledge	Focus on smaller companies? On recruitment?
Federated learning is very interesting	Include in future development plan
How to measure the platform's success?	Number of users? Number of engagements? Number of downloads? Number of solutions?
How to get data cleaning algorithms?	?
Some users don't know how they could even benefit from machine learning	Include teaching and learning
Is anonymizing the data really enough? What is actually meant by "anonymizing"?	Think about this
Include live events / webinars	Host the webinar interview series on the platform
Need to connect code to a Git repository	Can just be linked (or are we providing an internal service?)
Challenge providers have certain requirements on data quality, results format etc.	Should some of these rules be hard-wired or can each challenge provider name their conditions?
It is very important for users to understand how to grant access / control rights?	How to grant access / control rights?

Findings ("business model")	**Actions**
Commercial users are willing to pay for data / results if they get a real benefit from them	Pay for getting data / results based on quality / number of solutions
Larger companies have very complicated and slow processes: hard to find the responsible person	Work out who the right people are and discuss the business model with them
Users are willing to pay for posting a challenge	Pay small amount for posting challenge plus extra for each result
Define "pain point" and "commercial purpose". The benefit needs to be larger than the extra effort!	To do, including estimation of running costs and quantification of benefit.
Users see the benefit of using the platform for recruitment	Include recruitment in the business model and on the diagram
Users may be willing to actually pay to be a moderator, as they gain leads (e.g. €5,000)	Consider this, although this is very much a marketing perspective and I don't know if technical people would see this benefit
Commercial users my even be willing to pay to use the platform (all information in one place)	Consider this
Anyone who pays anything needs to clearly see the benefits.	Create an info graphic based on the business model
Use the platform to catalogue the performance of users to solve particular problems (better than e.g. AWS because it's industry-specific).	Do we even have to define the problems? The users will define them themselves? Maybe we should define categories?
Finance challenges / events through company sponsoring	E.g. CHF 5,000; 10,000 and 15,000
The challenges could be run on Kaggle	Think about this
Users may be willing to pay for "my portfolio"	Include in business model

Findings ("marketing")	**Actions**
Make it clear that it is dangerous NOT to share (e.g. COVID)	Include in marketing plan
Potential users would like to see some examples of actual topics.	Create a new presentation including example.
Why should we trust a platform operated by an unknown Swiss university?	Get larger companies on board, include collaborators in public presentation

As well as providing us with some very helpful feedback, many organisations would like to develop real use cases with us in the next development stage, as summarized in Table 5. The two use cases at the top (EDPR and Fraunhofer IWES) are the two that we plan to start with in the next phase, due to the very high interest of these two organisations. See Chapter 5 for more details.

Table 5. Summary of planned use cases.

Organisation	Use case	Their submission	What they want	Notes
EDPR (PT)	Several problems available! → Challenge	Challenge with SCADA data	Results to "sell" the platform internally	Start immediately!
Fraunhofer IWES (DE)	Data standardisation	Challenge	Organise a collaboration	Start immediately!
BKW (CH)	Under-performing WTG	Question (open) with SCADA data (private)	Analysis results (and code)	They may not want this request to be public
Anerdgy (CH)	Small WTG performance analysis	Question (open) with SCADA data (open)	Analysis results (and code), communication with other companies, GUI	They are open to publishing all data
OST/ETH (CH)	Digital Twin	Standard classes, structures etc.	Tbd	
Aventron (CH)	Pattern recognition wind turbine performance	Question (open) with SCADA data (private)	Analysis results (and code)	
EKZ / Nispera (CH)	Under-performing WTG	Question (open) with SCADA data (private)	Analysis results (and code)	Together with Nispera
Hitachi ABB (CH)	Tbd	Tbd	tbd	Large company, slow processes
Sulzer Schmid (CH)	Blade inspection	Tbd	Tbd	
Winji AG (CH)	SCADA data	Question (open) with SCADA data (private)	Analysis results (and code)	
Skypull SA (CH)	Airborne data	Tbd	Tbd	
Voliro (CH)	Drone measurement data	Tbd	Tbd	
ENBW (DE)	Tbd	Tbd	Tbd	
Orsted (DK)	Tbd	Tbd	Tbd	
Ins Santa Catarina (BR)	Teaching	Video of lectures	Tbd	
NRG Systems (US)	Measurement data		Tbd	
RES (UK)	Tbd (operational data)	tbd	Tbd	

2.5 Mini-demonstrator

In this section we first introduce the main functionalities of the mini-demonstrator, then introduce the example entries we have created, and finally summarise the main findings from discussing the mini-demonstrator with the interviewees.

2.5.1 Main functionalities

Parallel to the user journeys and user interviews, we developed a mini-demonstrator together with Stacker Group, based on their existing software, Relight. The mini-demonstrator therefore already includes some of the functionality discussed in the previous sections, in particular:

- Basic Wiki-platform with access control, text-based posts, threaded conversations and user-defined labelling/filtering.
- Definition of open and private spaces with access control.
- Embedding links and videos in the posts.
- Content search engine indexes and displays related content.
- Post "liking" and ordering in the overview according to popularity.
- Direct data storage and linking with access control on Google Cloud.
- Live streaming of events.

This allowed us to already show the basic functions to some of the interviewees. The feedback has been extremely positive!

For example, Figure 9 shows the sign-up page on the left and the start page on the right, with two example spaces: the open space called *SwissWindEnergy* and a private space that this user has access to on the right called *PrivateCollaborationSpace*. Users can define their own public or private spaces. On entering the public space, the user is presented with the view shown in Figure 10, a place where the community can post, collaborate, and communicate, including advanced features for tagging, topic'ing, commenting, posting video, and many other features. Users can post code, data, videos and explanations, and can have threaded conversations with peers, as shown in Figure 11 (left). An example of the content search engine indexing and displaying related content is shown in Figure 11 (right), enabling the community build a network of interconnected content. The search engine shows content related to a challenge posted to the community - including all replies. This allows all available data, code solutions and algorithms on a given topic to be catalogued in one place.

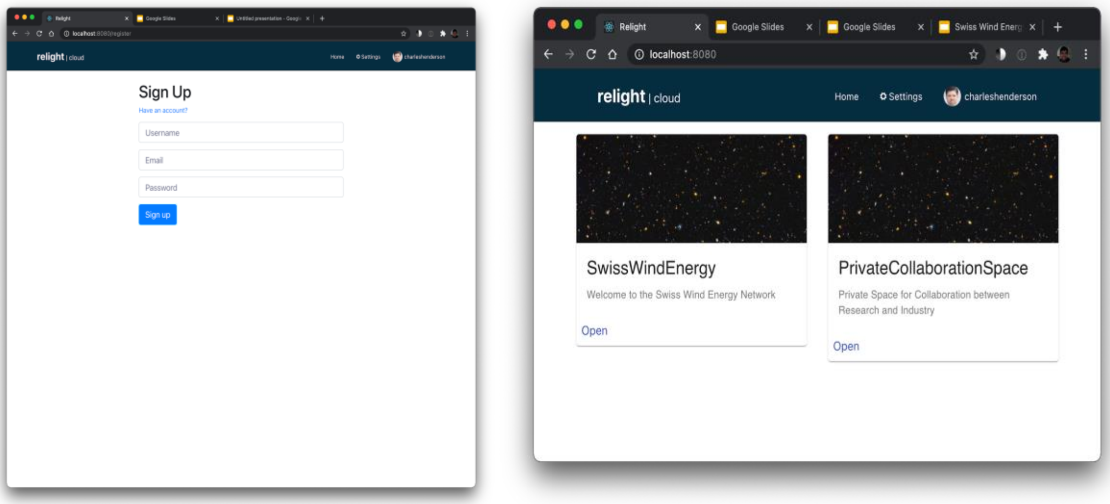


Figure 9. Mini-demonstrator: Sign-up page (left) and start page (right).

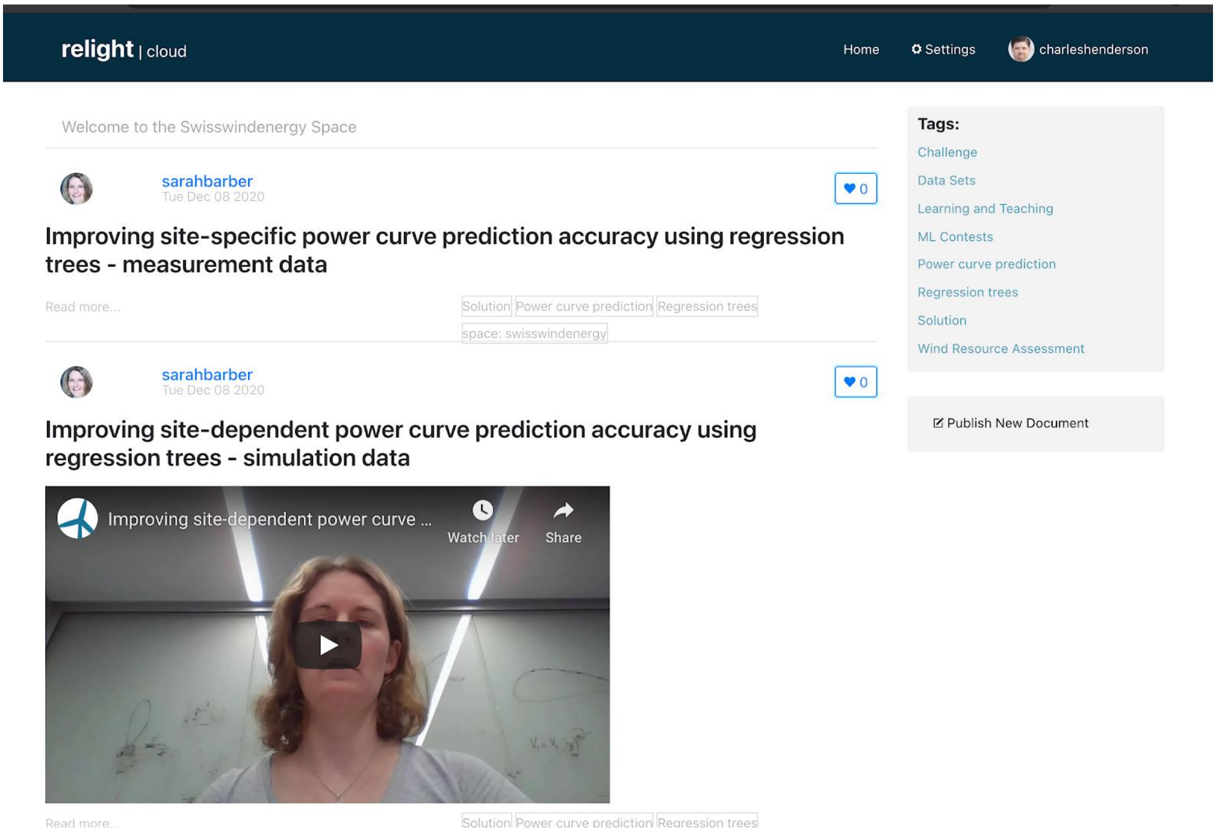


Figure 10. View of user interface.

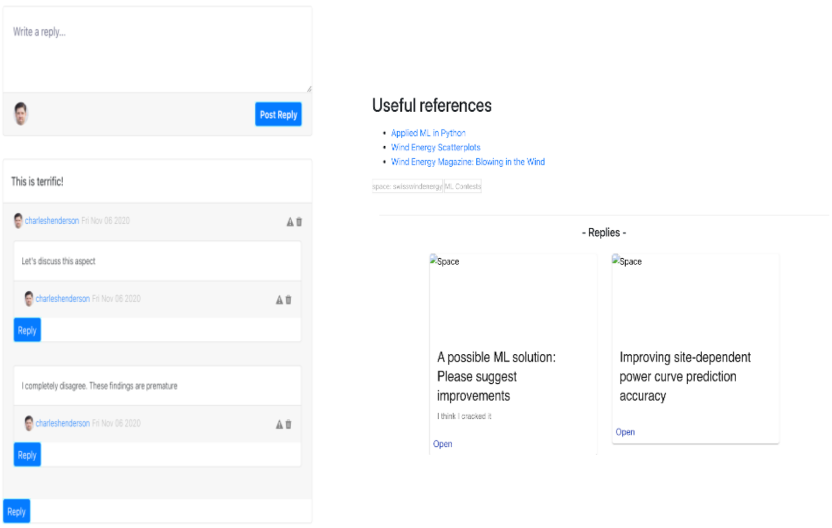


Figure 11. Interaction possibilities: threaded discussion (left), referenced solutions (right).

The functionalities that we plan to implement in the future are summarised in Chapter 5.

2.5.2 Example entries created

In order to demonstrate and explore the capabilities of the platform, we created some example entries of our own as shown in Figure 12. The first table shows the two "challenges" and related "solutions" that we created, and the second table summarises the datasets, two of which are stored on the platform and the rest of which are linked to their existing location.

Table 6. Summary of example entries for the mini-demonstrator.

Challenge	Solution	Notes
Improving site-dependent power curve prediction accuracy	Improving site-dependent power curve prediction accuracy using regression trees - simulation data	This challenge was created by Sarah Barber, the solution refers to a solution she has developed, the code and data are included.
	Improving site-dependent power curve prediction accuracy using regression trees - measurement data	This challenge was created by Sarah Barber, the solution refers to a solution she has developed. The data is included, the code is not yet available.
Optimal choice of wind model or method	A new method for the optimal choice of wind model	This challenge was created by Sarah Barber, the solution refers to a solution she has developed, a description and link to a paper are provided.
	Comparison metrics simulation challenge	This challenge was created by Sarah Barber, the solution refers to a solution she has developed, a description and link to a paper are provided.
	A new method for the optimal choice of wind model - further studies	This challenge was created by Sarah Barber, the solution refers to a solution she has developed, a description and link to a paper are provided.

Data	Format	Notes
Coastal operating wind farms	Stored on platform	We downloaded this from Zenodo[1] and uploaded onto the cloud for demonstration purposes
Regression tree simulation data	Stored on platform	Test and training data created through simulations as part of the "regression tree" solution listed in the previous table
RAVE data	Link only	In a next phase, more information about the metadata as well as links to published work using the data should be provided. Contact the data owners.
ORE Catapult Operational Data	Link only	In a next phase, more information about the metadata as well as links to published work using the data should be provided. Contact the data owners.
Orsted Offshore Operational Data	Link only	In a next phase, more information about the metadata as well as links to published work using the data should be provided. Contact the data owners.
EDP Open Data	Link only	In a next phase, more information about the metadata as well as links to published work using the data should be provided. Contact the data owners.
https://eo-winds.net/2020/08/13/offshore-wind-farms-production-time-series-an-introduction-to-entso-e-and-elexon-bmrs-datasets/	Link only	Link to interesting article

[1] https://zenodo.org/record/1475197#.YA7kchYxnb0

The first challenge is shown in Figure 12 followed by one solution in Figure 13. Here it can be seen how easy it is for participants to enter challenges and receive, share and further discuss solutions.

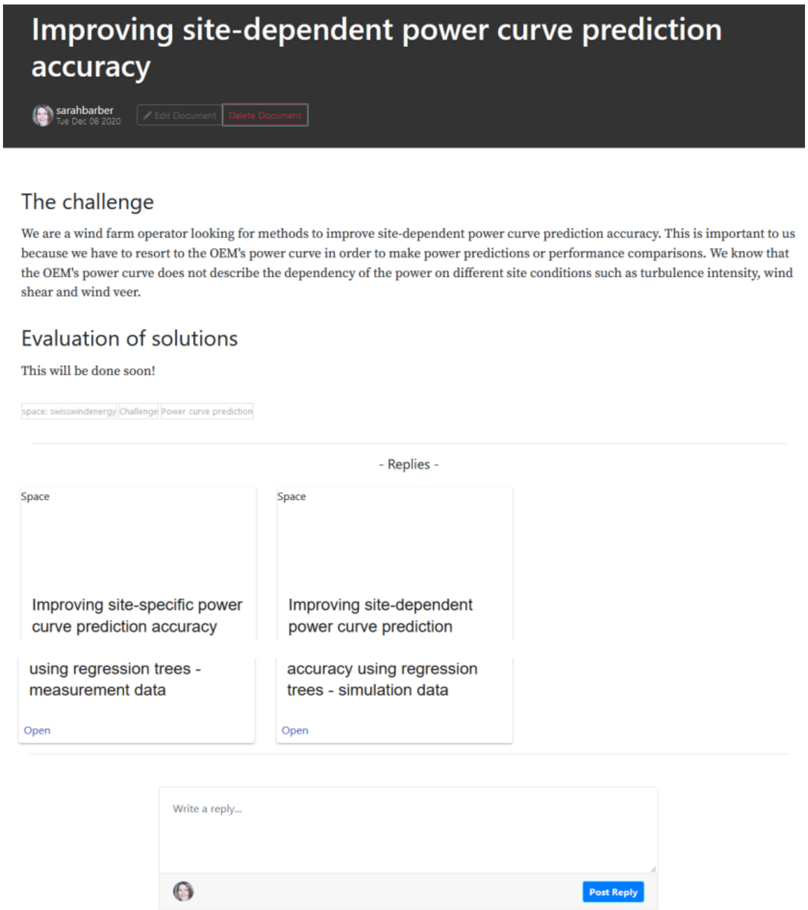
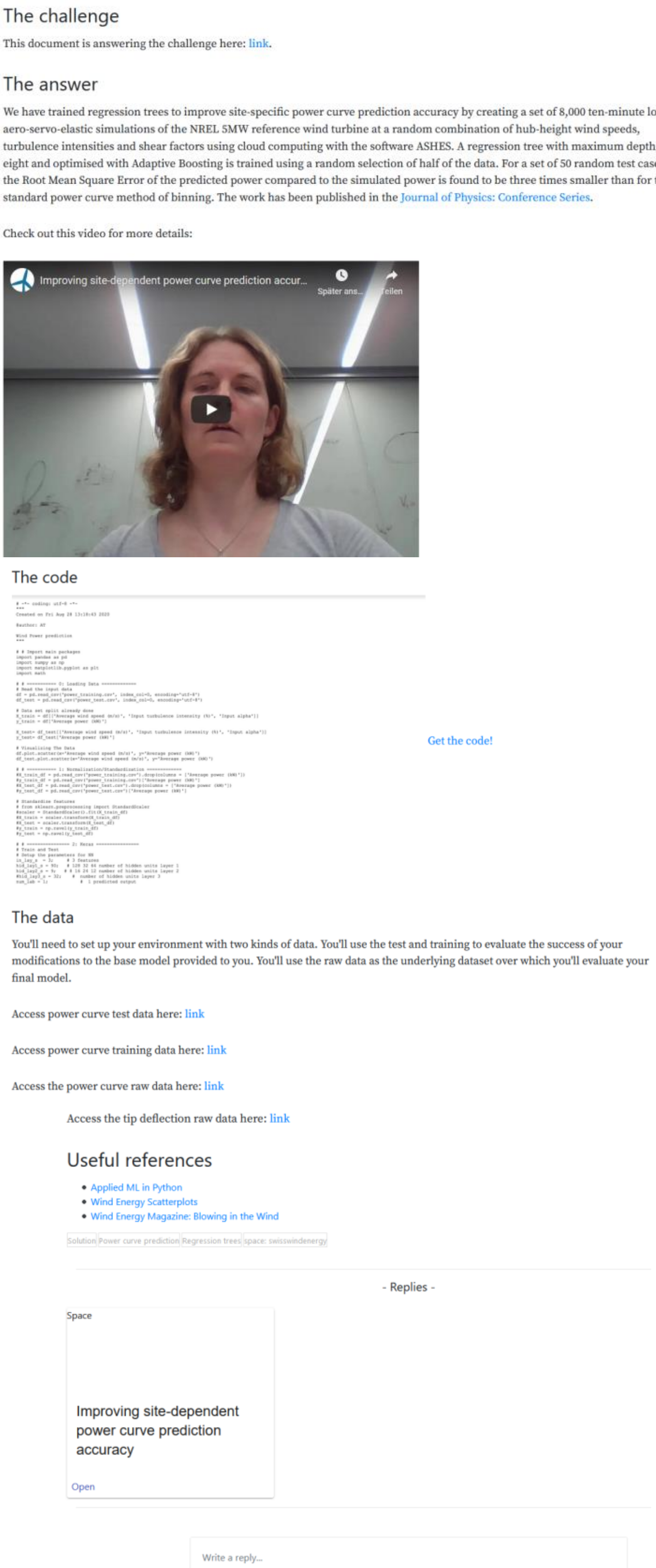


Figure 12. Example challenge.



The challenge

This document is answering the challenge here: link.

The answer

We have trained regression trees to improve site-specific power curve prediction accuracy by creating a set of 8,000 ten-minute long aero-servo-elastic simulations of the NREL 5MW reference wind turbine at a random combination of hub-height wind speeds, turbulence intensities and shear factors using cloud computing with the software ASHES. A regression tree with maximum depth of eight and optimised with Adaptive Boosting is trained using a random selection of half of the data. For a set of 50 random test cases, the Root Mean Square Error of the predicted power compared to the simulated power is found to be three times smaller than for the standard power curve method of binning. The work has been published in the Journal of Physics: Conference Series.

Check out this video for more details:

Improving site-dependent power curve prediction accur...

Später ans...

Teilen

The code

```
# -*- coding: utf-8 -*-
"""
Created on Fri Aug 28 13:18:43 2020

@author: AT

Wind Power prediction
"""

# # Import main packages
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import math

# # =========== 0: Loading Data ============
# Read the input data
df = pd.read_csv('power_training.csv', index_col=0, encoding='utf-8')
df_test = pd.read_csv('power_test.csv', index_col=0, encoding='utf-8')

# Data set split already done
X_train = df[['Average wind speed (m/s)', 'Input turbulence intensity (%)', 'Input alpha']]
y_train = df['Average power (kW)']

X_test= df_test[['Average wind speed (m/s)', 'Input turbulence intensity (%)', 'Input alpha']]
y_test= df_test['Average power (kW)']

# Visualizing The Data
df.plot.scatter(x='Average wind speed (m/s)', y='Average power (kW)')
df_test.plot.scatter(x='Average wind speed (m/s)', y='Average power (kW)')

# # =========== 1: Normalization/Standardization ============
#X_train_df = pd.read_csv("power_training.csv").drop(columns = ['Average power (kW)'])
#y_train_df = pd.read_csv("power_training.csv")['Average power (kW)']
#X_test_df = pd.read_csv('power_test.csv").drop(columns = ['Average power (kW)'])
#y_test_df = pd.read_csv('power_test.csv")['Average power (kW)']

# Standardize features
# from sklearn.preprocessing import StandardScaler
#scaler = StandardScaler().fit(X_train_df)
#X_train = scaler.transform(X_train_df)
#X_test = scaler.transform(X_test_df)
#y_train = np.ravel(y_train_df)
#y_test = np.ravel(y_test_df)

# # ================ 2: Keras ================
# Train and Test
# Setup the parameters for NN
in_lay_s  = 3;     # 3 features
hid_lay1_s = 90;   # 128 32 64 number of hidden units layer 1
hid_lay2_s = 9;   # 8 16 24 12 number of hidden units layer 2
#hid_lay3_s = 32;    #  number of hidden units layer 3
num_lab = 1;          #  1 predicted output
```

Get the code!

The data

You'll need to set up your environment with two kinds of data. You'll use the test and training to evaluate the success of your modifications to the base model provided to you. You'll use the raw data as the underlying dataset over which you'll evaluate your final model.

Access power curve test data here: link

Access power curve training data here: link

Access the power curve raw data here: link

Access the tip deflection raw data here: link

Useful references

- Applied ML in Python
- Wind Energy Scatterplots
- Wind Energy Magazine: Blowing in the Wind

Solution | Power curve prediction | Regression trees | space: swisswindenergy

- Replies -

Space

Improving site-dependent power curve prediction accuracy

Open

Write a reply...

Figure 13. Example solution.

2.5.3 Summary of mini-demonstrator findings

The main insights from the development and discussion of the mini-demonstrator are summarised in Table 7. Combined with the findings from the user journeys and user interviews, we developed a final concept as described in Chapter 3 and a list of next steps as described in Chapter 5.

Table 7. Summary of findings from mini-demonstrator.

Category	Actions
Structure	Create a label called "type" (challenge, solution, approach, dataset, etc.), "topic" (power curves, etc.) with filtration based on labels.
	Identify topics and moderators for these topics: keep everything in one space and show moderator when you click on the topic (focus on O&M for now).
	Standardise pick lists, custom labels (the moderators merge or divide tag names).
	On problem definition, need to see the datasets that are applicable to it. The problem needs to state its requirements (tag types).
	"The platform" has sub-spaces, e.g. "main space", "The Swiss Wind Energy R&D Network", "IEA Wind Task xx". The "Platform" is a separate space.
	Use "favourites" to determining sequencing (done by the moderator of each task).
	Should the data always be included inside a post labelled data?
	How to grant access / control rights?
Business plan	Pay to moderate a topic.
	Pay for a challenge.
	Companies pay for publishing "commercial solutions"/advertising (one free).
Other	Indexing and cataloging the data / metadata.
	Rules on how to use, e.g. always link the challenge? Provide templates for creating documents?

3 Final concept

In this chapter, we present the final concept of the platform by discussing the technicalities, the business model, then introduce some examples of applications and finally examine the value added for and in Switzerland.

3.1 Technicalities of the platform

3.1.1 Overview

An overview of the concept is shown in Figure 14. The concept we have developed is the first central collaboration platform in the wind energy industry that provides real incentives for participation. It is a Wiki- and cloud-based platform with threaded discussions and searchable labels, allowing users to post requests, challenges and contests on one hand (data-owners), and answers, solutions and submissions on the other hand (algorithm/method-owners) on key wind energy topics, as well as datasets, lectures, information, standards, data standards, and more. The platform is divided into separate "open" and "private" sections, allowing users to share and showcase information on their own terms. It contains an optional docker environment for running and reproducing entire workflows in a standardised manner. The platform does not seek to replace any existing platforms; its power is in providing a central location for solving the most important wind energy industry challenges through sharing and exchanging selected information. For example, the datasets do not have to be stored on the platform (even though they can be), but a link or API to an existing platform can be provided, code can be either uploaded directly or provided via links to existing GitHub repositories, existing videos can be embedded, etc.. As this platform will provide users with a real incentive for participation, it will be able to sustain itself long-term by generating revenue from user membership fees, advertising, and fees for particular features. We have changed the name of the platform to the wedowind.com platform, with a slogan "Accelerating innovation through knowledge sharing".

As well as providing a central location for collaboration, a key innovation of this platform is related to the separate "open" (in green on the figure) and the "private" (in blue) sections of the platform, combining to give the "collaborative" part (in yellow). This allows users to share information on their terms. It gets away from the idea that sharing has to involve losing competitive advantage, by allowing organisations to share only the information that they believe will provide them with a benefit. For example, an organisation would be free to follow the threads on a particular interesting topic in the "open" section, and then could contact someone who has been working on a relevant topic and request a private collaboration. They could then create a "private" space, only providing access to their information to selected people, on their own terms (e.g. once an NDA has been signed). The information can be stored on the platform, but does not have to be, if the organisation does not want to make use of the optional virtual environment. The entire collaboration then takes place in this "private" space. The results can then be shared with the community if the partners want to showcase their work, discuss further questions, work collaboratively on a particular challenge or define a "contest" for the community, etc. Furthermore, a description of the available information could be entered into the "open" section without actually making the information available to the entire community, in order to encourage requests and collaborations. As well as commercial collaborations, other "private" spaces may include IEA Wind Tasks, network spaces and other working groups.

This provides all user types with a significant added value, including:

- Getting real problems solved.

- Getting access to the state-of-the-art in relevant wind energy topics as well as real industry problems.
- Getting access to reproducible code, workflows and algorithms related to specific challenges and/or data sets.
- Getting access to relevant and usable data.
- Making use of data-controlled private spaces to keep data safe and to collaborate only with selected partners, boosting innovation through data and algorithm benchmarking and comparisons.
- Sharing experiences and learning on relevant topics.
- Gaining visibility and showcasing work.
- Finding international – or local – project partners, and collaborating with the best data scientists / members of the wind energy community.

Belonging to a small country without a large wind energy programme, Swiss organisations benefit additionally as follows:

- Increased international visibility of Swiss organisations and their solutions via The Swiss Wind Energy R&D Network.
- Increased learning from international players.
- Improved international collaboration.
- Easier access to state-of-the-art technology and solutions from the global community.
- Specific benefits from the machine learning solutions provided on the platform due to the difficulty of analytically describing the effects of mountainous conditions such as complex terrain, turbulence, icing, etc. on wind turbine behaviour

Further specific details are discussed in the next sub-sections.

3.1.2 Background and access control

The platform will be accessed at www.wedowind.com. Please note that currently the mini-demonstrator is available at https://relight.cloud/space/swisswindenergy, and access can be given on request. The platform is run on Google Cloud. We have a Google Cloud account and use this for platform administration. It can be run on servers in Europe, if required for data protection reasons. The public can view the "open" part of the platform but cannot comment, post or download data. Users have to request access to the platform. Registered users can comment, post and download in the open space, and access selected private spaces. Registered users can create their own private spaces, e.g. commercial collaborations, IEA Wind Tasks, network spaces. Only entry creators can edit their own documents.

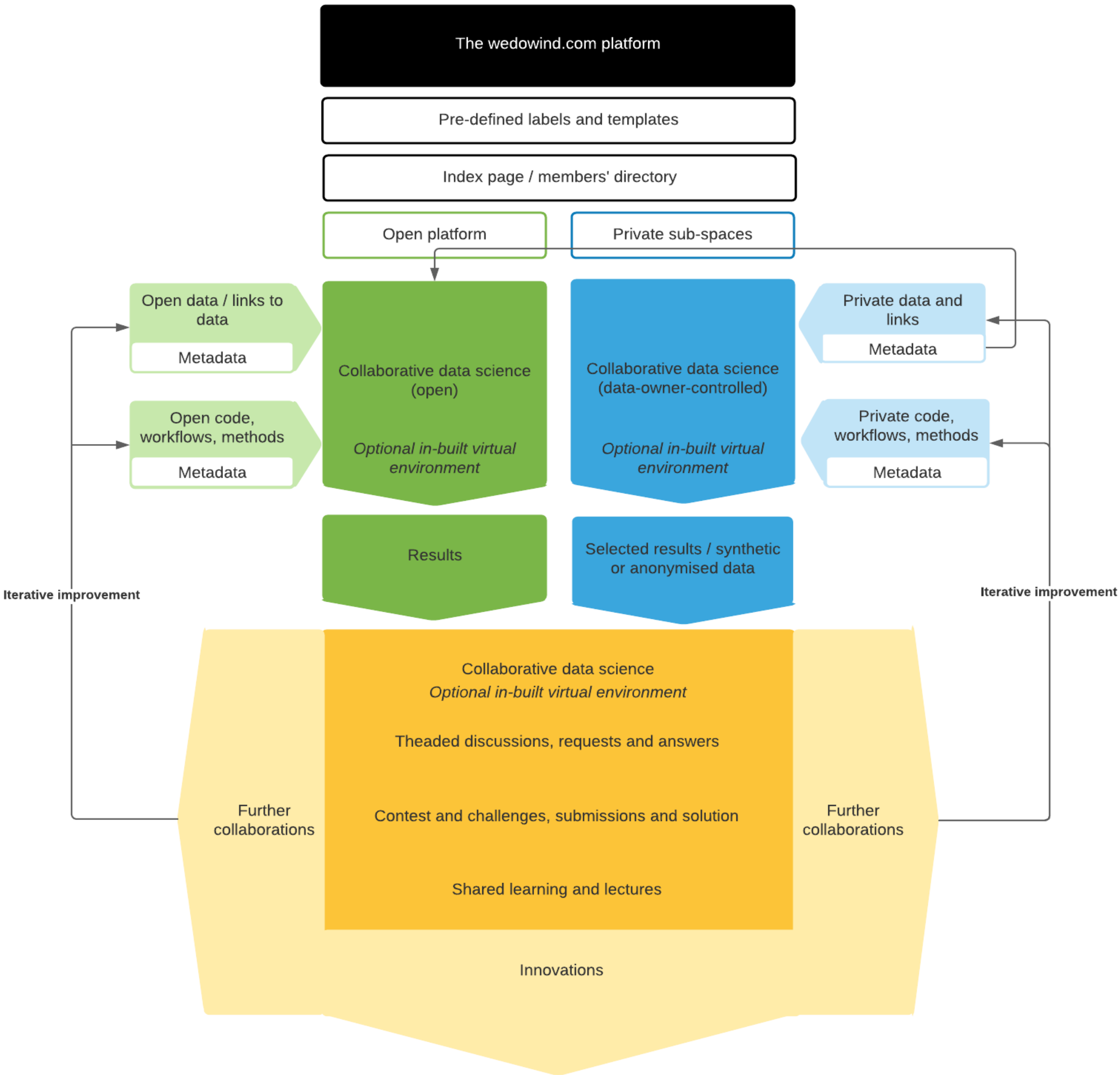


Figure 14. Overview of final concept.

3.1.3 Structure and labels

When the user logs onto *www.wedowind.com*, they will see the Open Platform as well as the private platform to which they have access, as shown in Figure 15.

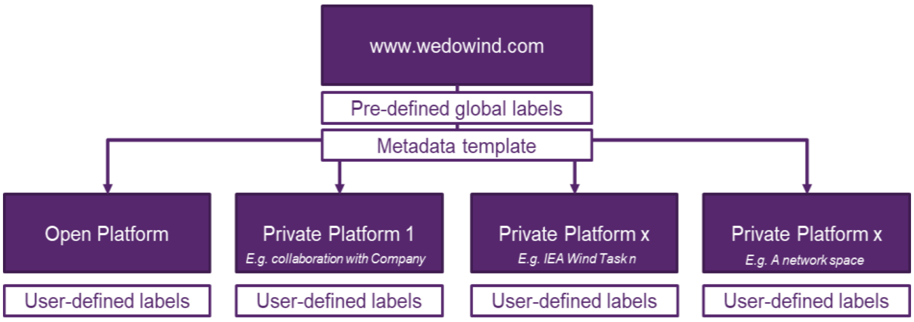


Figure 15. Structure of the platform.

Within each platform, the user can create various document types, for which templates are provided. Each document created is labelled with both pre-defined labels and user-defined labels, as shown in Figure 16. The pre-defined global labels consist of one Type and at least one Theme. The proposed document Types are shown in Table 8 (top), but are expected to change as the platform is further developed. Generally, the idea is that users can search and filter entries according to their labels. For example, the user can view all contests and challenges that have been posted related to power curve prediction, or all the requests and datasets related to filtering of SCADA data. This allows them quick access to state-of-the-art knowledge, to the correct project partners and collaborators, to reproducible workflows, to learning resources, etc. and ultimately fostering innovation. An example of a template (for the Type "challenge" is shown in Figure 17.

The proposed Themes are shown in Table 8 (bottom). For each Theme, an industry expert will be chosen as a Theme moderator, with the following roles:

- **Quality control of the entries:** ensuring that the users use the templates and labels correctly, and ensuring that the entries are relevant, preventing spamming and selling.
- **Consolidation and tidying of user-defined labels**: regular checking and consolidation of user-defined labels, e.g. when similar labels with slightly different spelling are used (e.g. Data Set, dataset, data).
- **Suggestion and creation of contests:** monitoring of activity on the platform and active communication with the users in order to help / encourage them to create relevant content.
- **Active management of running contests:** ensuring that the contests are running smoothly and encouraging users to submit their results on time.
- **Communication with platform owners:** regular updates on the general running of the platform.

Several people have even indicated a willingness to pay us for acting as moderator, in order to increase their visibility and reputation within the industry (see Section 3.2 for further details).

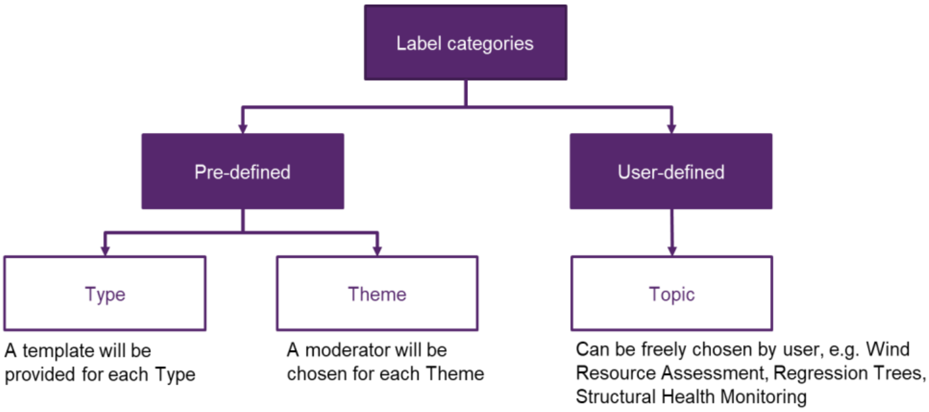


Figure 16. Label categories.

Table 8. Type and Theme categories.

Label category	Label name	Description	Example
Type (pre-defined)	Challenge	A challenge defined by a user	How do I improve the accuracy of power curve prediction?
	Solution	The response to a challenge	This is a method we developed…..
	Request	A request defined by a user	
	Answer	The response to a request	
	Contest	A contest defined by a user	Develop a new ML method with the lowest RMSE for predicting power performance on this provided dataset by a certain date with xx prize money.
	Submission	The answer to a contest	This is the resulting code and results.
	Dataset	A dataset	This can be a link or the data itself, the metadata has to be defined according to the metadata template
	Learning	Any entry enhancing the learning of the other users	A video lecture
	Information	Any other information	A link to an interesting article, funding guidelines
	Live	A live event	A webinar interview series
	Standard	A description of a standard	
	Taxonomy	A description of a taxonomy	
	Template	A template for each Type	

Label category	Label name	Description	Examples
Theme (pre-defined)	Reducing_OPEX	Decreasing operating costs of wind farms	Monitoring, data analysis, maintenance, forecasting, logistics, Airborne Wind Systems.
	Asset_management	Improving asset management of wind farms	Monitoring, data analysis, maintenance, forecasting, logistics, Airborne Wind Systems.
	Planning	Decreasing planning costs of wind farms	Wind resource assessment, wake effects, atmospheric boundary layer, wind measurements, environmental impact, financing, risk reduction, logistics, complex terrain.
	Offshore	Special topics only applicable to offshore	Floating technologies, marine environments, offshore structures.
	End_of_life	End-of-life issues including recycling	Lifetime estimation, remaining lifetime assessment, recycling, materials.
	CAPEX	Reducing WTG investment costs	Wind turbine design, optimization, aerodynamics, aeroelastics, component design, blade testing, structural issues, upscaling.
	Social	Increasing social acceptance	Wildlife detection, noise monitoring and reduction, reducing impact on environment, behavioural issues, financing concepts.
	Grid	Maintaining grid stability and reliability due to the fluctuating wind resource	Storage, smart grids, microgrids, energy system modelling.
	Small wind	Small wind turbines < 100 kW	Monitoring, design

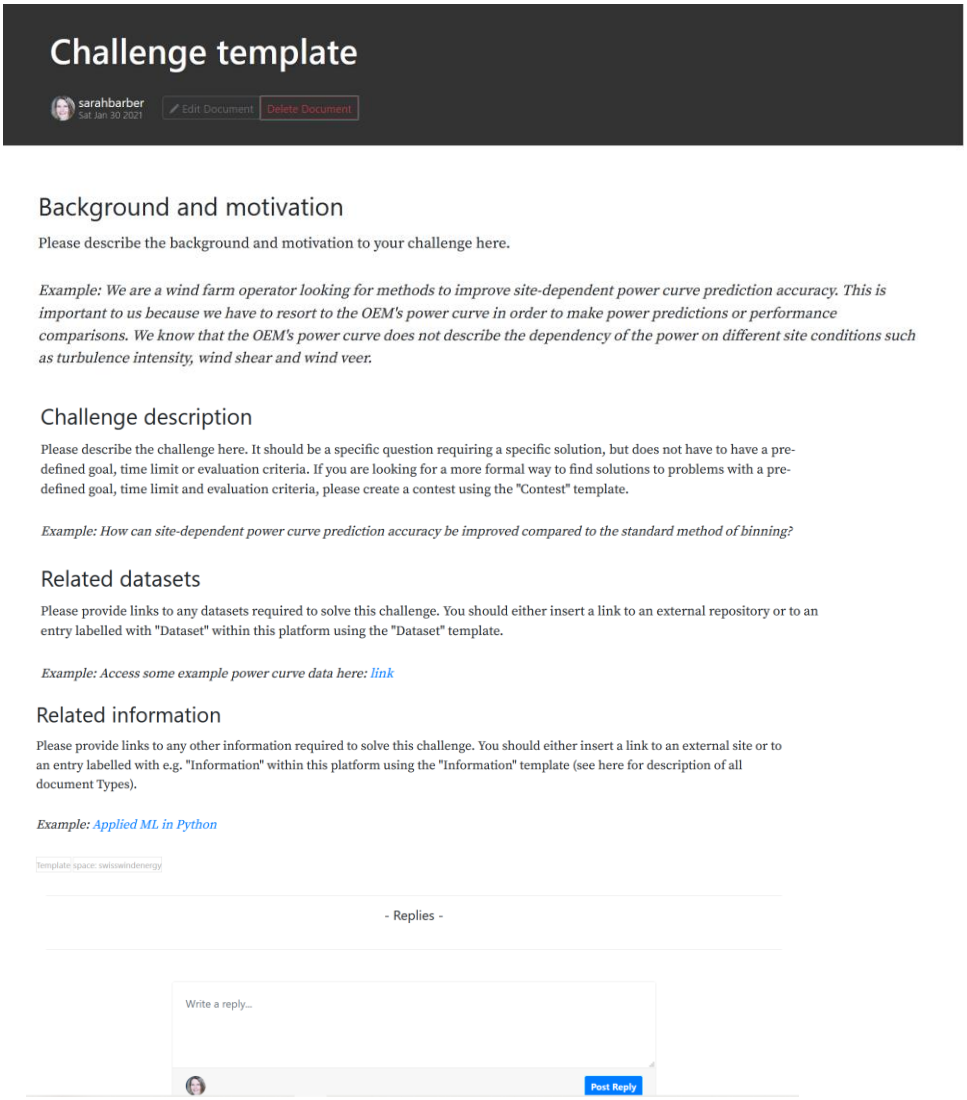


Figure 17. Example template.

3.1.4 Definition of metadata and taxonomies

The wind energy sector uses a wide range of data, physical models and mathematical models for different purpose throughout the entire wind turbine and wind farm lifecycle. There are many different interfaces involving many different parties that require this data to be shared. Metadata gives context to data sets, and is crucial for meaningful and productive analysis. Metadata may encompass administrative, descriptive, or structural data. Entire data sets, such as the turbine power curve, can be considered metadata. Uncertainty information can also be considered to be metadata. Metadata requires some kind of standard to ensure successful data sharing, even within a closed, or "private" wind energy project. For data sharing and collaboration, the requirement for such a standard becomes even stronger.

Although some efforts have been made towards a wind energy data sharing standard (as described in Chapter 2), **there is currently no agreed-upon standard for most wind energy data**. A standard metadata approach would benefit the wind energy sector by allowing for greater ease in sharing data.

The platform will support users on this topic by providing metadata templates, which will be defined by the relevant experts in the area of data standards. We have good contacts with these experts through our activities in IEA Wind Task 43. We plan to motivate the experts to do this in an innovate way – we plan to actually to run a "challenge" on the platform as one of the first planned use cases on the topic, asking the community to come up with a "solution" themselves (see Chapter 5). This will allow us to demonstrate the value of our platform but at the same time actually result in a solution for defining metadata improving data sharing, improving the effectiveness of the platform itself!

3.1.5 Ensuring data confidentiality / definition of IP rules

As mentioned in Section 3.1.1, data confidentiality is ensured on this platform as follows:

- The public can view the "open" part of the platform but cannot comment, post or download content (data/code/workflows etc.).
- Users have to request access to the platform: registered users can comment, post and download in the open space, can access private spaces governed by the data owner, and can create their own private spaces, e.g. commercial collaborations, IEA Wind Tasks, network spaces. Only entry creators can edit their own entries.
- The data-owner controlled private spaces are run by the data-owners according to their own requirements. They can store the data on the platform and run algorithms in the virtual environment, or they can provide links or APIs to their data at another location. They can give selected people access on their own terms, for example only after signing a confidentiality agreement or only via a password.

3.1.6 Virtual docker environment and federated analytics

Code sharing will be possible in three different ways:

1. Pasting code blocks directly into documents or by creating direct links to GitHub or GitLab repositories (currently possible in the mini-demonstrator).
2. Automated extraction of Jupyter notebooks into the format required for the platform, enabling easy code publishing.
3. Inclusion of a docker solution, which is already being developed by Stacker Group to complement its Relight platform. This involves users downloading a docker image and developing code in a standardised environment and publishing the final version of the code on the platform. The docker distribution will include a standard operating environment with Jupyter and other python packages that are either generic or wind specific and designed to facilitate a standard and reproducible workflow.

Furthermore, federated analytics has a high potential to be applied on this platform in the future. The majority of data scientists would like to be able to have access to real data from wind fleets, but many owner-operators are reluctant to share this data. As a result, a key obstacle to the development of machine learning models has been the difficulty of accessing reliable training data. Federated models for machine learning have presented a promising approach to solving this problem. Private spaces within the platform will allow for a decentralised/federated approach to sharing, balancing the dual need for sharing and privacy preservation.

3.2 The business model

One of the key requirements of the platform is that it is self-sustaining, can cover its own running costs long-term and can be scaled up to cater for the global wind energy market, as well as other markets in the future.

In this section of the report, we firstly discuss various possibilities for the business model based on the results of the user interviews, and then conduct a risk analysis. The final business model has not yet been defined – this will be developed in the next phase as shown in Chapter 5.

3.2.1 Business model possibilities

For the consideration of business model possibilities, we have developed a Value Proposition Canvas that includes the following topics:

- Customer profile: Customer jobs; Gains; Pains.
- Value proposition: Products and services; Gain creators; Pain relievers.
- Revenue streams.

The results are summarised in Figure 18. We have split the customers into two groups: (1) Companies, who own or have access to data and either (a) develop data analytics solutions for internal use and/or external customers or (b) utilise external data analytics solutions, and (2) Researchers, who develop data analytics solutions and require access to relevant data and problems. In some cases we even differentiate between small and large companies, where "small" refers to companies with limited internal resources for data analytics and "large" refers to companies with internal data analytics team. The analysis allowed us to define more clearly what the platform should offer the users:

- A wiki-style posting of workflows, code and data.
- Ensure reproducibility by combining data, code and workflows.
- Allow codes and data to be compared and developed.
- Access to cloud data storage (a certain amount is free).
- Access to cloud computing (a certain amount is free).
- Access to research / industry partners and customers.
- Sub-spaces for public or private collaborations.
- Access to real industry problems via contests, challenges and requests.

Together with the user interviews, this then allowed us to define the following possible revenue streams:

- **Users pay for a "contest":** one of the key parts of the platform is the ability for users to create entries that are labelled as "contests", "challenges", "requests", etc. as described in Chapter 2. The "submissions", "solutions" and "answers" that are created as response to these posts allow the platform content to be searched and filtered according to the key relevant problems in the industry. The "contest" entries are special cases of "challenges" that have a specific goal, deadline, evaluation criteria and prize money associated with the solution quality, similar to the *Kaggle* platform. Our user interviews indicate that companies are prepared to pay the platform operators for posting a "contest", either via a fixed price, a percentage of the prize money, or a percentage of the number of "submissions".

- **Users pay for "solutions":** As described in Table 8, "challenges" refer to a problem defined by a user, such as "*How do I improve the accuracy of power curve prediction*?". As opposed to "contests", challenges do not have fixed goals, deadlines, evaluation criteria or prizes. However, potential users have expressed willingness to pay for useful results, or "solutions" to their challenges.
- **Users pay for "my portfolio":** The platform allows users to showcase their work by summarising all their contributions in one place. This allows them to build up their reputation and more easily find project partners, customers or even jobs. Companies have expressed a high interest in the potential of the platform for recruiting (one of *Kaggle*'s main value propositions) as well as increasing their visibility and creating new contacts.
- **Users pay membership fees depending on organisation type (less for individuals and universities):** This is the most simple solution, and requires users to pay monthly or yearly fees depending on their membership type. This idea is used for many existing platforms including Slack, LinkedIn and Dropbox, and is usually accompanied by some free functionalities over a limited time period. This could be combined with any of the other revenue streams.
- **Credit system:** This idea involves a type of "marketplace" in which users trade in tokens – they gain credits for adding contributions (e.g. posting a solution) and lose credits for taking information (e.g. downloading a data set), and possibly pay money to compensate credits. This idea is very much in trend at the moment, but one of the challenges is to define the credit unit fairly (i.e. how many MB of data are equivalent to how many lines of code) and how to take into account the quality of the content rather than just its size or length. Furthermore, this system may actually reduce collaboration rather than increase it, due to the effect "A Fine is a Price" described in a study of day-care centres[9]. This shows that a payment cannot compensate for undesired behaviour (i.e. collecting children late), and can even lead to an increase in this undesired behaviour because the payment is perceived to compensate for the behaviour. In the case of this platform, paying for information (e.g. data) through money or credits could lead to organisations then not bothering to contribute anything useful at all.
- **Users pay to be a Theme moderator:** As mentioned in Section 3.1.3, some potential users have indicated willingness to pay to be named a Theme moderator, expecting to profit in this role from increased visibility and networking opportunities.

It can be seen that there are a number of interesting possible revenue streams for this platform; however, there is more work required to establish exactly what the users are willing to pay for (and how much). This is part of the planned future development discussed in Section 5.

3.2.2 Risk analysis

Our risk analysis is summarised in Table 9, with the risks in the first column and the future mitigation strategies in the second column. We have already attempted to minimise risks through the user interviews; however, they all have potential to be reduced even further in the next development stage. These strategies have been included in the development plan in Section 5. We would particularly like to emphasise the importance of acting quickly and starting the next phase of development before someone else does so.

[9] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=180117

Table 9. Risk analysis.

Risk	Future mitigation strategies
Not enough participants at the start	Start with specific "challenges", invite key people who are especially interested (see Section 5)
	Offer incentives for users who "acquire" their colleagues
	Create a "marketing presentation" with example
Use cases don't work as expected	Plan more use cases than required, invite pre-defined users to contribute
The platform does not work as expected	Start with a few users, adding more users in planned stages
The functionalities for the MVP aren't enough to attract users	Plan resources for continuous development of changes and requirements that come up as we add users
The requirements of the users are not met	Build up the platform slowly, adding more users in planned stages
Participants don't stay long-term	Communicate with the participants during the next development phase
Participants don't contribute long-term	Continuously talk to participants to ensure that the concept provides incentives to contribute long-term
	Offer a trial period to all users
Not enough transfer from the "private" to the "open" part	Focus on communicating the collaboration and networking benefits
General resistance to sharing information from some companies	Initial focus on the organisations with low resistance / high potential
Users may be put off by Google Cloud	Use a server in the EU, discuss the topic further with users
Negative effects due to the strong dependency on Stacker Group	Further discussions with other companies, definition of cooperation contract as soon as possible
Someone else develops a competing platform	Start with the next development stage immediately!

Value Proposition Canvas

Revenue streams
Users pay for a contest / proportion of prize money / number of submissions
Users pay for contest solutions
Users pay for "my portfolio"
Users pay membership fees depending on organisation type (less for individuals and universities)
Credit system
Pay to be a theme moderator

Value Proposition		Customer Profile	
Products and services	**Gain creators**	**Gains**	**Customer jobs**
Collaboration and data sharing platform including: Wiki-style posting of workflows, code and data Ensure reproducibility by combining data, code and workflows Allow codes and data to be compared and developed Access to cloud data storage (a certain amount is free) Access to cloud computing (a certain amount is free) Access to research / industry partners and customers Sub-spaces for public or private collaborations Access to real industry problems via contests, challenges and requests	**Companies:** Define "challenges", "requests" and "contests", link to replies in a thread Provide templates for posting reproducible code and workflows Provide metadata templates Allow users to showcase their work Provide learning resources Allow threaded conversations Provide access to data storage and computing (certain amount is free) Select users to moderate themes (they pay) **Researchers:** Allow data to be easily added via templates / allow links and storage Provide a template for well-defining the challenges, requests and contests Reduce the barrier for entering content via templates Provide a platform for easily embedding learning videos Provide access to data storage and computing (certain amount is free) Allow threaded conversations	**Companies:** Get their problems solved Improve their methods by: - accessing reproducible code and workflows (including for data cleaning) - accessing partners/researchers/recruiting - accessing learning resources / following the state-of-the-art - sharing and comparing workflows/data Access storage and computing resources in one place Improve visibility / get more partners and customers by: - showcasing their work and their ability to solve problems - being selected as a theme moderator **Researchers:** Access relevant data in one place Access real problems Access potential partners Access learning resources Access storage and computing resources in one place Compare and develop algorithms with other users	**Companies:** Data cleaning and sorting Data analysis / data science Developing recommendations / making decisions Ensuring competitive advantage Finding research partners Comparing and improving their methods Selling their services (some companies) Ensuring visibility as industry leaders **Researchers:** Defining real industry problems Finding industry partners for new projects Developing relevant and innovative workflows / methods Developing new scientific methods Publishing and communicating work
	Pain relievers	**Pains**	
	Larger companies: Allows data cleaning workflows to be added / provide examples Provides metadata templates and data standards information Allows private rooms to be created and enable easy communication between users Provides an incentive for users to invite their colleagues Allows private rooms to be created / automated anonymisation feature **Smaller companies:** Allows private rooms to be created and enable easy communication between users Allows data and problems to be easily shared Provides metadate templates and data standards information Provides an incentive for users to invite their colleagues Allow users to showcase their work Provide a platform for easily embedding learning videos **Researchers:** Provide metadata templates Define "challenges", "requests" and "contests", link to replies in a thread	**Larger companies:** Spend too much time on data cleaning and preparation Too much data in different formats Increasing pressure to reduce costs / improve performance / stay innovative No overview of state-of-the-art research Not easy to share data due to internal resistance / unwillingness **Smaller companies:** Not enough resources/time to implement advanced analytics Increasing pressure to use data to improve performance Data in different formats No overview of state-of-the-art research No access to research partners No understanding of how ML could help them **Researchers:** Not enough data available to develop their methods and workflows Lack of understanding of the "real" problems No access to industry partners	

Figure 18. Value Proposition Canvas.

3.3 Examples of usage of platform

There are many different ways users may want to use this platform. In Figure 19 we give eight examples of possible applications as follows:

- Large wind farm owner/operator:
  - CHALLENGE: "How can we make our workflow for xyz more efficient (using example data abc)?", resulting in suggested SOLUTIONS (reproducible workflows with data, code and descriptions) in private or public spaces.
  - CONTEST: "Develop a workflow for solving xyz with dataset abc by date ddd." (with prize money), resulting in SUBMISSIONS by the deadline in private or public spaces.
- Smaller wind farm owner/operator:
  - CHALLENGE: "Why is wind turbine number x under-performing in this set of data?", resulting in suggested SOLUTIONS in private or public spaces.
  - REQUEST: "How can I use ML to help sort and clean all my different datasets?", resulting in ANSWERS in terms of links, videos, threaded discussions, suggested workflows.
- Researcher:
  - REQUEST: "Has anyone got data xyz for doing abc?", resulting in ANSWERS in terms of links to data (on platform or external).
  - SOLUTION: "We have solved problem xyz like this" (connect to "My portfolio"), resulting in REQUESTS for collaboration in open or private spaces.
- Innovative technology provider / manufacturer:
  - CHALLENGE: "How can we make our workflow for xyz more efficient (using example data xyz)?", resulting in suggested SOLUTIONS in private or public spaces.
  - SOLUTION: "We have solved problem xyz like this" (connect to "My portfolio")., resulting in REQUESTS for collaboration in open or private spaces.

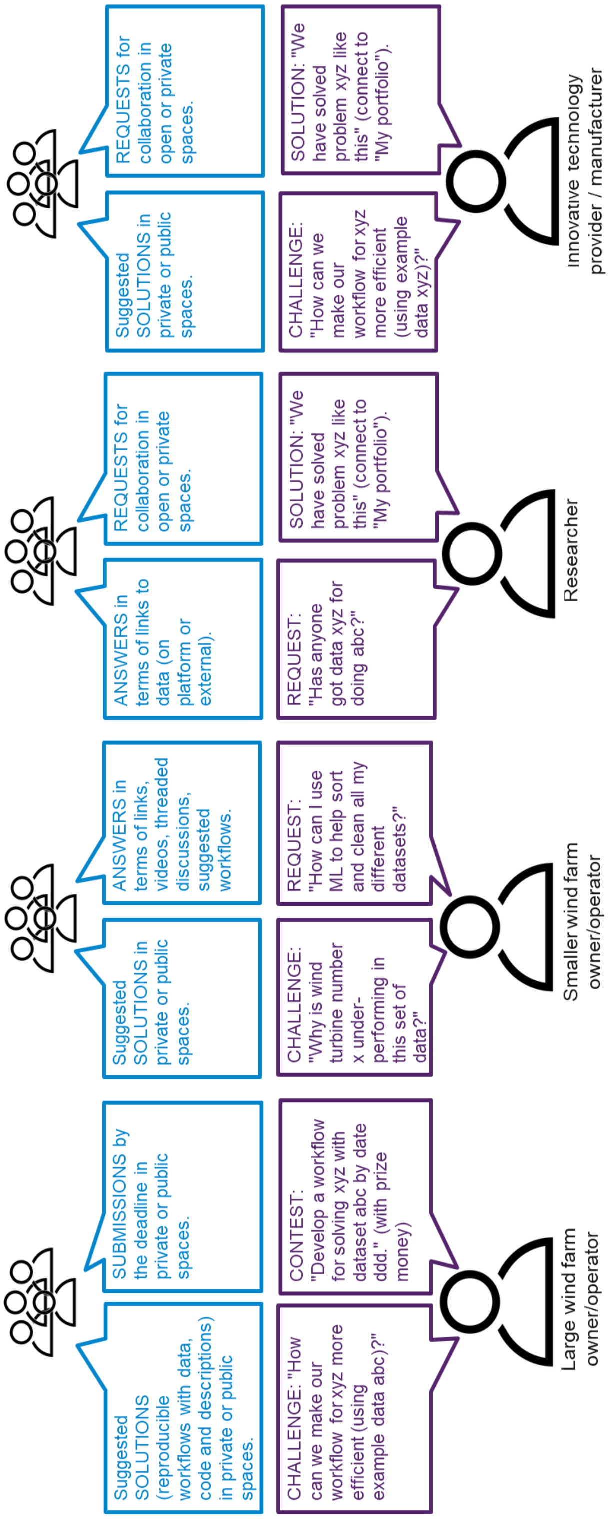


Figure 19. Examples of usage of platform.

3.4 Value added for and in Switzerland

In this section, we first look at the added value for all users, Swiss and international, and then the specific benefits to Swiss organisations and Switzerland in general. Finally, we review the commitment of Swiss partners to contributing to this project and explain why The Swiss Wind Energy R&D Network is the best partner for this project.

3.4.1 Benefits for all users

The added value that the platform provides to all users, Swiss and international, is summarised below:

- Users can **get real problems solved** on the platform by (a) searching for challenges and solutions on particular topics, (b) posting contests, challenges and requests, (c) winning relevant partners for collaborations.
- Users can **stay informed on state-of-the-art** relevant topics, real industry problems and solutions and utilize this knowledge to improve their own products and services.
- Users can **boost innovation** by accessing and further developing reproducible code, workflows and algorithms related to specific challenges as well as relevant and usable data sets.
- Users can **make use of data-controlled private spaces** to keep data safe and to collaborate only with selected partners, improving innovation through data and algorithm benchmarking and comparisons.
- Users can **multiply their knowledge and learn new things** by sharing experiences and learning on relevant topics.
- Users can gain **visibility** and **showcase** their work, ultimately winning new partners, customers and employees.
- Users can **easily find international or local project partners**, and collaborate with the best data scientists / members of the wind energy community.

3.4.2 Additional benefits in Switzerland

The additional benefits for Swiss organisations, because of the limited scope wind energy in Switzerland, include:

- Increased **international visibility** of Swiss organisations and their solutions via The Swiss Wind Energy R&D Network.
- **Strengthening and increased focus of Swiss knowledge, products and services** through improved collaboration between Swiss organiations (e.g. between industry and research).
- Improved **international collaboration**, and ultimately an **easier market entry** in the global wind energy market.
- Increased **learning** and **technology transfer** from international players.
- Easier **access to state-of-the-art technology and solutions** from the global community.
- Specific benefits from the **machine learning solutions** provided on the platform due to the difficulty of analytically describing the effects of mountainous conditions such as complex terrain, turbulence, icing, etc. on wind turbine behaviour

The benefits that can be transferred to the Swiss Energy Strategy 2050 include:

- The whole platform concept can be transferred to other energy-based applications.
- Tokenomics and incentive schemes.
- Attract local multi-disciplinary Swiss talent to provide state of the art solutions to world-wide wind energy problems.
- Anonymous computing: sharing data without compromising data privacy and run compute algorithms and models without compromising copyrights and authorships.

Other benefits to Switzerland in general include:

- **Improved acceptance of wind energy** in the general public due to more media coverage and exposure to attractive and innovative projects.
- Improved **visibility and reputation of Switzerland** as an innovate country in the area of renewable energy, a globally relevant topic.
- **Technology transfer** from the participating organisations to organisations and stakeholders in other areas (e.g. industries (drones, space, watches, PV, etc.), knowledge areas (machine learning, AI, big data, analytics, smart materials, smart monitoring, etc.), governmental organisations, the general public, etc.).
- Future potential to **expand the concept to many other areas relevant to Switzerland**, including in societal, social and energy applications (we already have contact with organisations such as opendata.ch, Sociolution GmbH, utilities, etc).

3.4.3 Commitment of Swiss partners

It can be seen in Section 2.4 that we involved many Swiss partners in the user interviews, although it was equally important for us to talk to international players due to the international nature of the solution.

We have had a **very positive response** to this platform from Swiss partners, with all the organisations showing a **high level of interest**. However, some are more likely to actually use the platform than others. We have quantified our assessment of the likelihood with a number from 1 to 3, where 3 means that we are actively working on a use case together for the next development phase, 2 means that they expressed an active interest to contribute and 1 means that more interaction is required to persuade them of the benefits. It must be noted that the only organisations with a score of 1 and 2 were ones we talked to right at the beginning of the process, when the concept wasn't as far developed as it is now. We are quite sure that a further conversation with these organisations would transform their score a 3.

The 22 Swiss partners we talked to were (including Management Board members of the Swiss Wind Energy R&D Network as well as the user interviews), with our estimated contribution likelihood scores:

- IWB (utility) **[3]**
- EKZ (utility) **[3]**
- BKW (utility) **[3]**
- GE Renewables (component and wind turbine provider) **[2] – large company with slow processes**

- Hitachi ABB (power electronics component provider) **[2] – large company with slow processes**
- Meteotest AG (wind resource assessment company) **[1] – benefits of ML not clear.**
- Basler & Hofmann (wind energy consultant, formerly New Energy Scout) **[1] – ability to contribute data not clear.**
- WinJi AG (platform and analytics provider) **[3]**
- Sulzer & Schmid Laboratories (blade inspection company) **[3]**
- Skypull SA (Airborne Wind System developer) **[3]**
- Open Data CH (non-profit association) [not relevant]
- Gilytics AG (GIS service provider). **[3]**
- Nispera AG (platform and analytics provider) **[3]**
- Aventron (wind farm investor / owner) **[3]**
- Anerdgy AG (small wind turbine manufacturer) **[3]**
- Susi Partners (investor) **[2] – further discussions necessary**
- Cognite (digital services) **[3] (no official interview carried out)**
- ZHAW (university) **[3]**
- ETHZ (university) **[3]**
- OST (university) **[3]**
- EPFL (university) **[2] – further discussions necessary**
- FHNW (university) **[2] – further discussions necessary**

During the development of this concept, it wasn't possible to talk to all relevant players in Switzerland. However, these 22 interviews have convinced us that the platform has high potential to benefit many Swiss organisations. **The Swiss Wind Energy R&D Network has identified more than 200 organisations (research groups, start-ups, SMEs, technology providers, associations, etc.) who are directly involved in wind energy in Switzerland, who would benefit directly from the platform**. Furthermore, these organisations are well balanced between data providers (utilities, technology developers) and solution providers (universities and some technology providers), leading to benefits from both sides of the platform.

The Swiss Wind Energy R&D Network currently has more than 100 members and is in the process of defining its strategy for 2021. We aim to concentrate our activities on the topic of data sharing and open innovation, using it as an opportunity to (a) encourage Swiss organisations to use the platform and (b) to continuously improve it. This includes the Innosuise Event Series, for which we have already received funding, as well as the Innosuisse Innovation Booster, for which we are applying this year.

3.4.4 Why The Swiss Wind Energy R&D Network?

The reasons why The Swiss Wind Energy R&D Network is the ideal organisation to be leading this project are summarised below:

- Switzerland stands for **quality and innovation**, and is therefore the ideal country to launch and run such a high-quality and innovative platform. The hurdle to persuade users of our quality and innovation is not expected to be high.
- Switzerland is located in the **centre of Europe** and has a large amount of experience coordinating multi-national and interdisciplinary projects. We therefore have easy access to people and organisations who can support us and give us advice on the successful development of this project.
- **The Swiss Wind Energy R&D Network** is a non-profit association representing many different stakeholders with different skill sets and backgrounds. We therefore understand the needs of a wide range of stakeholders very well and are not influenced by any business or money related factors. Any money we make from this platform will be used for its further development or for general activities of The Swiss Wind Energy R&D Network.
- We have **many contacts with relevant international partners** and work on many different relevant topics, including in several IEA Wind Tasks, in particular in Task 43: Digitalisation. We are therefore in a position to influence the entire community.
- We have shown that this platform is something that the community wants. Existing attempts have failed because they have not provided a large enough incentive for data sharing. We have recognised this and developed a solution in a very short time, whilst everyone else is still just talking about it. **We need to at immediately and implement this platform before someone else does it.** Switzerland is the ideal country to do this in, and we very much hope to gain financial support for the next development stage as soon as possible (see Chapter 5 for more details).

4 Conclusions

The conclusions of this work are:

- We have developed an innovative concept for the first central collaboration platform in the wind energy industry that provides real incentives for participation. The platform does not seek to replace any existing platforms; its power is in providing a central location for solving the most important wind energy industry challenges through sharing and exchanging selected information.
- It is a Wiki- and cloud-based platform with threaded discussions and searchable labels, allowing users to post requests, challenges and contests on one hand (data-owners), and answers, solutions and submissions on the other hand (algorithm/method-owners) on key wind energy topics, as well as datasets, lectures, information, standards, data standards, and more. The platform is divided into separate "open" and "private" sections, allowing users to share and showcase information on their own terms. It contains an optional docker environment for running and reproducing entire workflows in a standardised manner.
- As this platform will provide users with a real incentive for participation, it will be able to sustain itself long-term by generating revenue from user membership fees, advertising, and fees for particular features.
- Through more than 30 user interviews and development of a mini-demonstrator, we have identified a very large interest in and high demand for a self-sustaining platform that provides a central location for wind energy industry and research to share and collaborate on reproducible workflows, data and code whilst respecting confidentiality requirements.
- There is a very high potential for this platform and we are committed to continuing its development as quickly as possible, as described in the final chapter of this report. We plan a bottom-up approach, starting with the definition of a Minimum Viable Product and further developing the platform as we introduce more users and better understand their needs, as described in the final chapter.
- The wind energy community benefits from this platform by enabling collaborations, leading to more efficient problem solving of the key challenges, ultimately reducing the cost of wind energy.
- Swiss organisations benefit from the increased opportunities for relevant learning and technology transfer from international and national players as well as collaborations that enable an easier entry in the global wind energy market.
- The Swiss Energy Strategy 2050 could benefit from transfer of the ideas and concepts developed here to the entire Swiss energy branch, including tokenomics, incentive schemes, anonymous computing, data sharing and computing without compromising copyrights and authorships. We are already involved in plans to apply the solution to other fields, as well as for the Swiss Wind Atlas. The platform has particularly high potential for applications with a wide range of stakeholders.
- Other **benefits to Switzerland** in general include the improved visibility and reputation of Switzerland as an innovate country, technology transfer to other fields, including in societal, social and energy applications, and the improved acceptance of wind energy in the general public due to more media coverage and exposure to attractive and innovative projects.

5 Outlook and next steps

The future development of this platform will follow the standard software development strategy of quickly releasing a Minimum Viable Product (MVP) and gradually building up features based on the experience gained with a gradual increase of use cases and users. In this chapter, we first introduce the planned MVP, examine the time plan and finally estimate the required development costs.

5.1 MVP

The MVP has been defined as shown in Figure 20. It will have some basic **functions**, including a sign-up for users (already existing), a unique URL and Space Name, a rich text editor (already existing), a code of conduct, the pre-defined labels as well as the pre-defined templates discussed in Section 3.1.3. We aim to focus on two **use cases** at the start, and are already in dialogue with these two organisations about their implementation. EDPR plan to provide a "contest" on a specific topic and data set provided by them. They already have experience in this and are very keen to work with us. Their international reputation will help us attract users. Fraunhofer IWES intend to provide a "challenge" on the topic of data standards and metadata. This is a key topic of interest identified by many potential users, without a complete answer. We intend to use the results of this challenge to improve the functionality of the platform itself (see Demonstrator 1 below). It should be noted that we are purposely choosing well-known international partners for the MVP, in order to attract more users and increase the benefit for the Swiss partners, who will introduce their use cases starting with the Demonstrator 1.

For the MVP, the platform will contain two separate spaces for the two use cases. This does not reflect the reality of the platform once it is fully in operation, but it will allow us to easily monitor and influence the development. **Users** will then be invited to contribute to the platform, and provide "submissions" to the contest and "solutions" to the challenge, as well as any other relevant contributions. We will invite the entire international community to sign up for the chance to contribute to this unique opportunity, and then select users who we think will contribute positively to its development. The **content** will thus be restricted to content related to the two use cases.

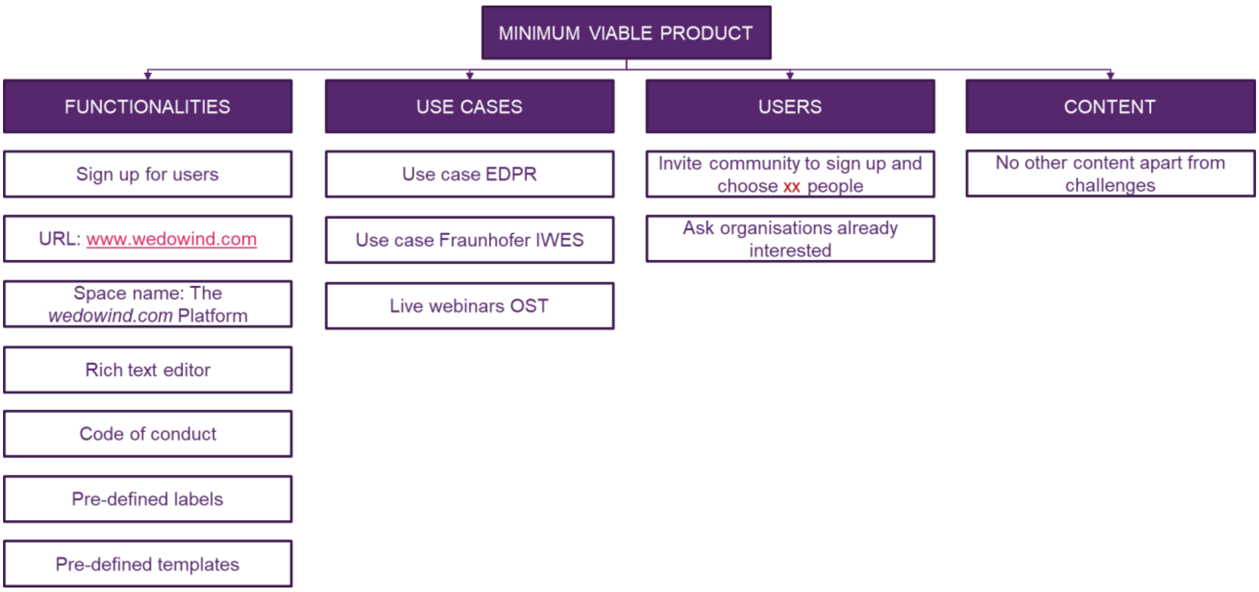


Figure 20. Description of MVP.

5.2 Further development

5.2.1 First stage: 2021-2022

Our time plan for the further development of the platform until the end of 2022 is shown in Figure 21. We are already in the process of implementing the changes required for the MVP, and plan to run the MVP use cases in March. We then plan three further versions of the demonstrator, each time taking one month to implement the results of the previous use cases. As well as implementing new functionalities based on the continuous learnings, there are some functionalities that we have already penciled in as being important, based on the user interviews. These include the following:

- **Rules on data standards / metadata / FAIR:** this is very important for providing users with an added value, saving them a large amount of time in data cleaning and preparation. Our idea is to use the results of the Fraunhofer IWES use case described above in the MVP to improve the functionality of the platform itself. This could be newly dataset templates and metadata rules, for example.
- **Data protection guidelines:** we would like to provide users with a clear and transparent guideline on the data protection measured used in the platform, so that the platform becomes a trustworthy "safe space" for them.
- **Connecting to Github:** a versioning system for links to Github repositories.
- **All templates and rules:** as well as the initial templates for the MVP, we plan templates and rules for all other document Types.
- **Content ordering via favourites:** the entries will be ordered according to their popularity, in order to make the most relevant entries more visible.
- **Strategy for integrating existing platforms:** we want to make it easy for organisations to integrate existing platforms directly into our platform via APIs, rather than just providing a link to them.
- **Creating private spaces:** users need to be able to create and administrate their own private spaces.
- **User profiles:** users will be able to create a profile so that other users know who they are communicating with.
- **Virtual environment:** this contains an optional docker environment for running and reproducing entire workflows in a standardised manner. This is vital for ensuring reproducibility and enhancing innovation.
- **User portfolio / showcasing:** this will allow users be create their own portfolios, essentially just summarising all their entries, in order for them to "show off" their skills and experience in order to help them find collaboration partners and for recruiting purposes.
- **Guidelines for accepting users:** in order to register on the platform, users will have to sign up and request to participate. The platform operators will choose whether to accept them or not. This will be done in order to maintain a high quality of entries and to avoid spamming or selling (LinkedIn can be used for that). We aim to develop guidelines in order to ensure a fair and transparent decision.
- **GUI integration:** some potential users have indicated they would like to use the platform for data storage and visualisation. In order to do this, a GUI functionality will have to be integrated.

- **Implement payments:** once we start to charge money for using the platform, or for certain functionalities, we will need a method to make and organise payments. Relight has already got a solution for this, which would need to be developed for our needs.

An overview of our current plan to implement these functionalities is shown in Figure 22. It should be noted that these may change at any time based on what we learn from the use cases.

Further details of the planned use cases, users and content for Demonstrators 1, 2 and 3 are shown in Figure 23, Figure 24 and Figure 25. Again, these may change as we go along, but give us some guidance on the expected workload. We are already discussing the use cases with the organisations mentioned; however, we are constantly carrying out new discussions and developing new ideas for use cases, so more are expected to be added to the list. We are also planning to add more content ourselves, using existing data, papers, codes and workflows that have already been published in different locations.

At the end of 2021, we plan for a "first launch" to a restricted set of users without any strong regulation of use cases, which we then plan to organically develop during 2021. There are some functionalities requiring longer development (such as implementing the virtual environment or introducing federated analytics, as well as ones not yet known to us) that will be developed in this phase as well. Then at the end of 2022 (indicated by the star), we intend the final launch to the public. Parallel to this, we will further develop the business plan and fully define the revenue streams and payment model.

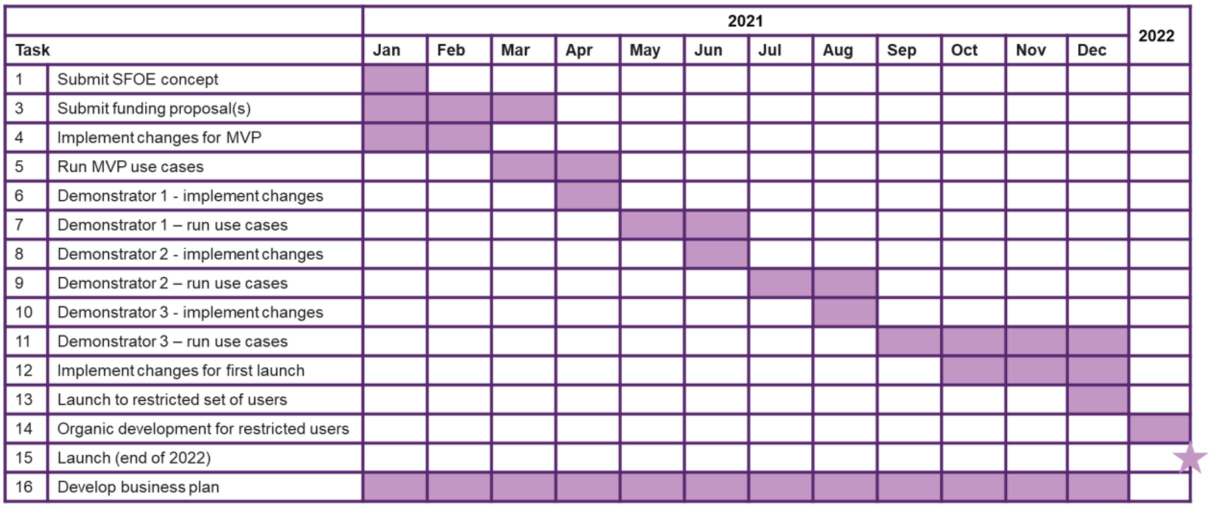


Figure 21. Time plan of future development.

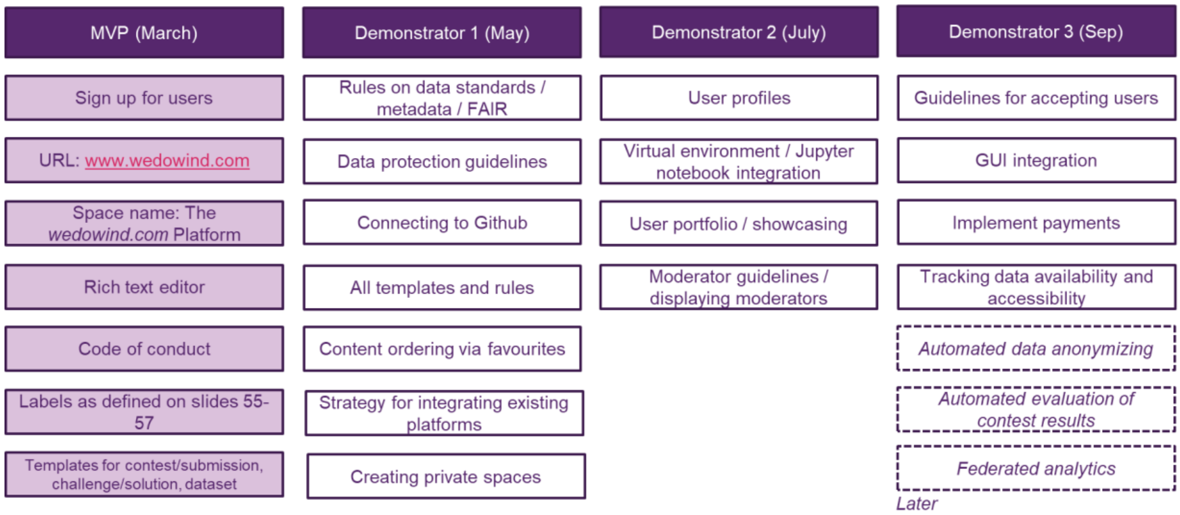


Figure 22. Planned functionalities.

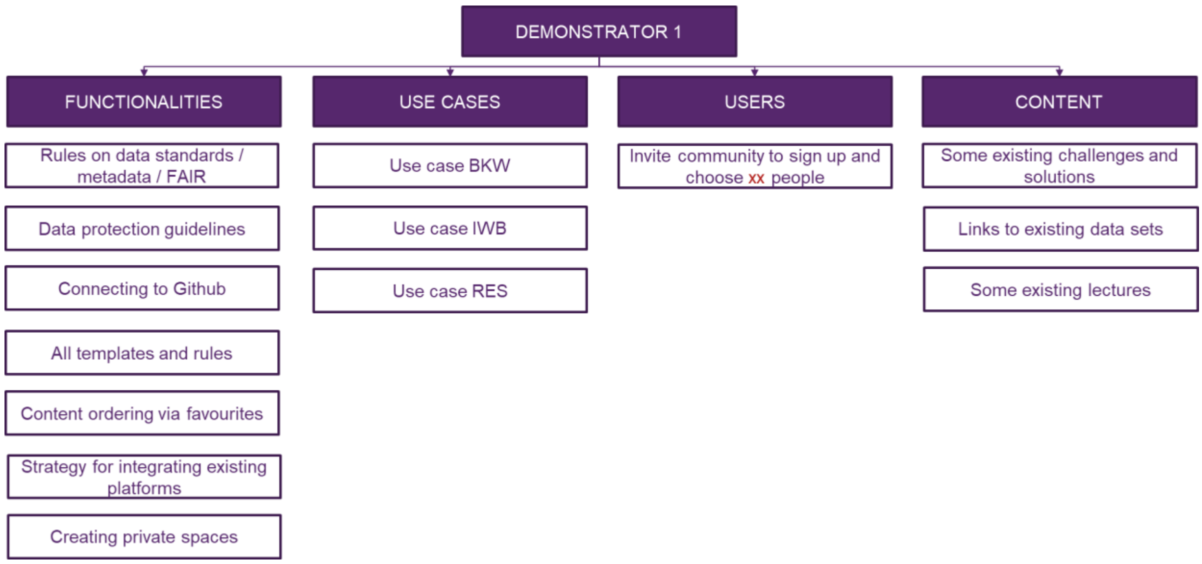


Figure 23. Development plan for Demonstrator 1.

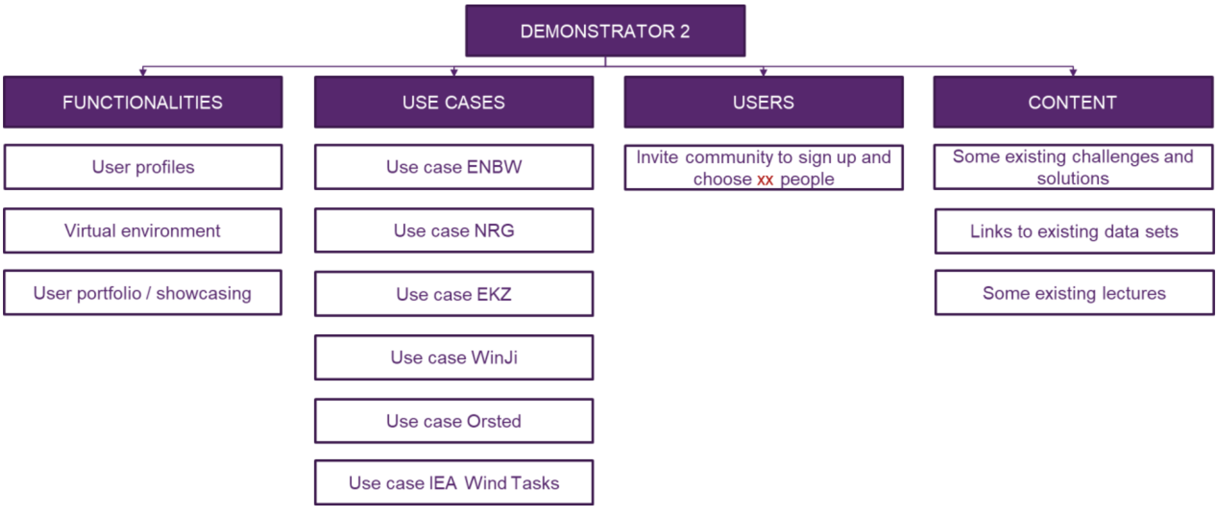


Figure 24. Development plan for Demonstrator 2.

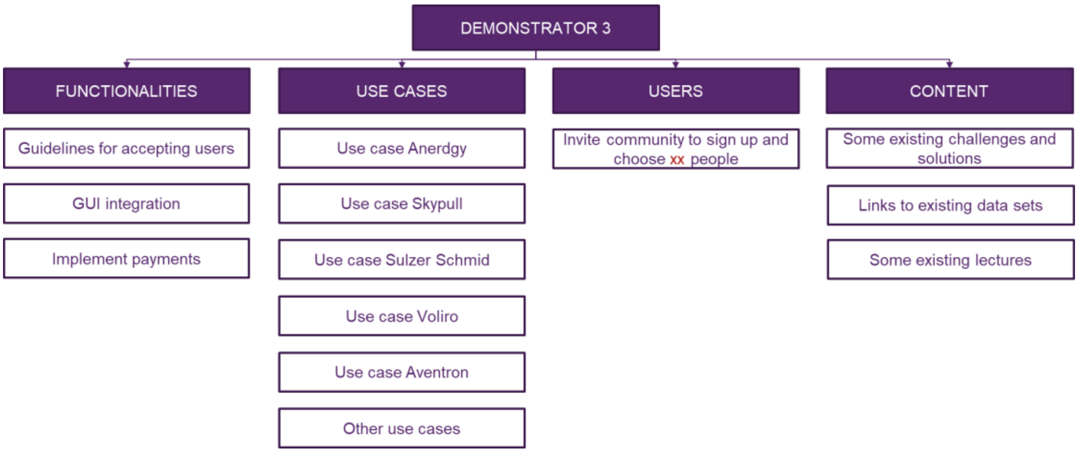


Figure 25. Development plan for Demonstrator 3.

5.2.2 Second stage: from 2023

Parallel to the developments described above, and in further development stages beyond the end of 2022, we will research and develop the following key topics:

- Automated data anonymising (to enable quick and easy data sharing).
- Automated evaluation of contest results (to enable fair and easy comparison of submitted results).
- Federated analytics (to boost data sharing by utilising data without accessing it directly).

This research and development will benefit this platform, but also can be transferred to many other applications.

Furthermore, there is a potential to expand the application of the platform to the following areas following the successful launch of the *wedowind.com* platform, which will be investigated and furhter considered during the course of the project:

- **The Swiss Wind Atlas Platform**: we are currently in the process of developing a concept for a new Swiss Wind Atlas Platform, which will be integrated into the wedowind.com platform as a private space. This will allow a continuous improvement of the Swiss Wind Atlas when further measurements, simulations and assimilation algorithms become available. As well as this, it will allow all types of stakeholders to interact with the atlas and contribute to helping the public understand how the atlas is created, helping them to develop fact-based opinions during a project development phase.

- **Wind energy acceptance:** the platform allows embedded videos and powerful visualisations. It has a high potential to be used in Switzerland for interacting with stakeholders to improve acceptance. Suisse Eole may be interested in using the platform to improve communication.
- **Other energy fields:** we have already discussed this platform with parties involved in other industries in Switzerland, such as the photovoltaic and the hydrogen storage industries. It appears that they are faced with similar challenges related to information sharing. Therefore we envisage an easy expansion to other industries.
- **Other general fields:** the strengths of the platform particularly come out when it is used in a multi-disciplinary way with many different type of stakeholders. The Relight platform from Stacker Group is already being applied in many US cities to enhance communication with the public. This is a potential future extension of this solution.

5.3 Development costs

Based on the development plan presented above, we estimate total costs of CHF 200'000 for the first two years, reducing to CHF 100'000 in year 3 and then continuing at this level for the long-term operation of the platform (staff and third-party costs). The detailed division of these costs are shown in the top part of Table 10. The third-party costs are planned for Stacker Group's time and license costs for developing and running the platform.

The expected income is shown in the bottom part of the figure. We expect to be able to start gaining income from users in the second year of the project, which should increase as the number of users and interactions increases. We expect to be able to raise about CHF 50'000 from other funding sources such as the private industry, Innosuisse or other organisations.

Please note that the significant in-kind contributions that have already been invested by OST, ETHZ and Stacker Group, as well as the in-kind contributions of the external partners during the development stage, are not included in this overview.

We would therefore like to request the following funds from SFOE: CHF 150'000 in 2021, CHF 125'000 in 2022, CHF 50'000 in 2023 and CHF 25'000 in 2024, amounting to an average of CHF 87'500 per year.

Table 10. Estimated costs and income.

Year	1	2	3	4	N
Staff costs					
Building up	CHF 50'000	CHF 50'000			
Use cases	CHF 50'000	CHF 50'000			
Running	CHF 50'000	CHF 50'000	CHF 50'000	CHF 50'000	CHF 50'000
3rd party costs					
Building up	CHF 50'000	CHF 50'000			
Running			CHF 50'000	CHF 50'000	CHF 50'000
Total costs	**CHF 200'000**	**CHF 200'000**	**CHF 100'000**	**CHF 100'000**	**CHF 100'000**
Income					
SFOE	CHF 150'000	CHF 125'000	CHF 50'000	CHF 25'000	
External funding	CHF 50'000	CHF 50'000			
Users		CHF 25'000	CHF 50'000	CHF 75'000	CHF 100'000
Total income	**CHF 200'000**	**CHF 200'000**	**CHF 100'000**	**CHF 100'000**	**CHF 100'000**

6 National and international cooperation

This project has involved a very large amount of national and international cooperation, as already described. The continued financial support of this project provides a unique opportunity to bring national and international cooperation possibilities onto a new level – not just in the wind energy industry, but for many other applications.

7 References

[1] A. Clifton, L. Kilcher, J. K. Lundquist and P. Fleming. Using machine learning to predict wind turbine power output. Environmental Research Letters, 8(2):024009, 2013.

[2] S. Barber and H. Nordborg, Improving site-dependent power curve prediction accuracy using regression trees, J. Phys.: Conf. Ser. 1618 062003, 2020.

[3] F. Pelletier, C. Masson and A. Tahan. Wind turbine power curve modelling using artificial neural network. Renewable Energy 89 207-214, 2016.

[4] Y. Zhang, H. Zheng, J. Liu, J. Zhao and P. Sun, An anomaly identification model for wind turbine state parameters, Journal of Cleaner Production 195 1214-1227, 2018.

[5] M. Schlechtingen and I.F. Santos. Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection. Mech. Syst. Signal Process. 25, 1849-1875, 2011.

[6] K. Leahy, L. Hu, I.C. Konstantakopoulos, C.J. Spanos, A.M. Agogino and T. J. O'Sullivan, Diagnosing and Predicting Wind Turbine Faults from SCADA Data Using Support Vector Machines. International Journal of Prognostics and Health Management, ISSN2153-2648, 2018.

[7] G. Helbing and M. Ritter. Deep Learning for fault detection in wind turbines. Renewable and Sustainable Energy Reviews 98 189–198, 2018.

[8] K. Kim, G. Parthasarathy, O. Uluyol, W. Foslien (Honeywell) S. Sheng and P. Fleming. Use of SCADA Data for Failure Detection in Wind Turbines. Energy Sustainability Conference and Fuel Cell Conference. 2011.

[9] L. Gigoni, A. Betti, M. Tucci and E. Crisostomi. A Scalable Predictive Maintenance Model for Detecting Wind Turbine Component Failures Based on SCADA Data.

[10] L. Tao , Q. Siqi, Y. Zhang and H. Shi. Abnormal Detection of Wind Turbine Based on SCADA Data Mining. Mathematical Problems in Engineering. Article ID 5976843. https://doi.org/10.1155/2019/5976843, 2019.

[11] F. Castellani, D. Astolfi, P. Sdringola, S. Proietti and L. Terzi. SCADA data mining techniques for efficiency and power assessment. Applied Energy 185 1076–1086, 2017.

[12] X. Wang, P.Guo andX. Huang. A Review of Wind Power Forecasting Models. Energy Procedia. Volume 12, 2011, Pages 770-778.

[13] Z. Lin and X. Liu. Wind Power Forecasting of an Offshore Wind Turbine based on High-Frequency SCADA Data and Deep Learning Neural Network. Energy. Volume 201, 15 June 2020, 117693.

[14] Z. Tia, X.W. Denga and H. Yang, Wake modeling of wind turbines using machine learning, Applied Energy 257 (2020).

[15] D. Stowell, M.D. Wood, H. Pamuła, Y. Stylianou and H. Glotin. Automatic acoustic detection of birds through deep learning: The first Bird Audio Detection challenge. Methods in Ecology and Evolution, Volume 10, Issue 3, March 2019, Pages 368-380.