WORKSHOP AGENDA

Workshop Name:	IEA Wind Task 32 Workshop on the **e-WindLidar** initiative		
Date of Workshop:	03-10-2018	Time:	10:00 – 17:00 (CET)
Workshop Facilitators:	Nikola Vasiljevic, Tobias Klaas	Location:	Riso, Denmark

1. Workshop Objective

This workshop intended to provide researchers, manufacturers, practitioners and end-users of lidar data:

- A detailed overview about FAIR data principles and their application on lidar data
- An exchange of experience about the everyday challenges in dealing with data
- An introduction to the current status of the common lidar data format proposed by the "e-WindLidar" initiative and the possibility to actively contribute to its further development
- The possibility of testing the open-source lidar data conversion tool "lidaco" that has been developed within the e-WindLidar project, which is able to convert raw data from many lidar systems to the common data format under supervision of the developing project team and describe data with meta information
- Extend the e-WindLidar working group with new members

2. Workshop Attendees

Name	Contact Information	Affiliation
Nikola Vasiljević	niva@dtu.dk	DTU Wind Energy
Tobias Klaas	tobias.klaas@iee.fraunhofer.de	Fraunhofer IEE
João Correia Lopes	jlopes@fe.up.pt	UPORTO
Daniel Gomes	danfergo@gmail.com	UPORTO
Elliot Simon	ellsim@dtu.dk	DTU Wind Energy
Norman Wildmann	norman.wildmann@dlr.de	DLR
Tobias Ahsbahs	ttah@dtu.dk	DTU Wind Energy
Robert Menke	rmen@dtu.dk	DTU Wind Energy
Antoine Larvol	al@windarphotonics.com	Windar Photonics
Rozenn Wagner	rozn@dtu.dk	DTU Wind Energy
Arezoo Zamanbin	arza@dtu.dk	DTU Wind Energy
Pedro Santos	paas@dtu.dk	DTU Wind Energy
Leonardo Alcayaga	lalc@dtu.dk	DTU Wind Energy
Neil Davis	neda@dtu.dk	DTU Wind Energy
Alex Forsting	alrf@dtu.dk	DTU Wind Energy
Julien Tissot	jtissot@leosphere.com	Leosphere
Peter Clive	peter.clive@woodplc.com	Wood Group - Clean Energy
Sean Daly	spdaly86@gmail.com	DNV GL
Andrew Clifton	andrew.clifton@windfors.de	WindForS

3. Workshop Agenda	
Time	Topic
09:30 - 10:00	Welcome coffee and danish pastry
10:00 - 10:10	Introduction to the workshop (Nikola Vasiljevic & Tobias Klaas)
10:10 - 10:40	Exercise 1: ● In groups of 5 state your everyday challenges working with data ● Collect feedbacks on post-its
10:40 - 11:00	Let's be FAIR: Introduction to the FAIR principles (Nikola Vasiljevic)
11:00 - 11:40	Exercise 2: ● In groups discuss pros and cons of FAIR principles ● Collect your feedbacks on post-its
11:40 - 12:00	Making LIDAR data FAIR (Tobias Klaas)
12:00 - 12:40	LUNCH BREAK
12:40 - 13:40	Exercise 3: ● In groups discuss pros and cons of the presented generic format *What is missing? What needs to be improved? Do we need it?* ● Collect feedbacks on post-its
13:40 - 14:00	Lidaco - lidar data conversion tool (Daniel Gomes & João Correia Lopes)
14:00 - 15:30	Exercise 4: ● Hands-on with Lidaco (convert your own data set)
15:30 - 15:50	COFFEE BREAK
15:50 - 16:05	e-WindLidar toolbox - towards community based tools (Nikola Vasiljevic)
16:05 - 16:20	Presentation of WindSP platform (João Correia Lopes)
16:20 - 16:50	Brainstorming next actions (Andy Clifton): Communication, workshops, prioritization of tools development
16:50 - 17:00	Wrap-up (Nikola Vasiljevic)

4. Pre-work/Preparation (documents/handouts to bring, reading material, etc.)		
Description	Prepared by	Links
e-WindLidar draft report	Nikola, Tobias, Joao, Daniel, etc.	https://drive.google.com/open?id=1n5nYugbd8Q5WnXxsJ8MypC7KSCJJ1Uw-
Lidaco installation guide	Joao, Tobias	https://drive.google.com/open?id=1327NdFUP4kW3IHH1dG3sTXk5LvmOnnBwTt07OtfZnFk

MEETING MINUTES

Workshop Name:	IEA Wind Task 32 Workshop on the **e-WindLidar** initiative		
Date of Workshop:	03-10-2018	Time:	10:00 – 17:00 (CET)
Workshop Facilitators:	Nikola Vasiljevic, Tobias Klaas	Location:	Riso, Denmark

1. Agenda and Notes, Decisions, Issues

The workshop particularly focused on the group work in the following plenum sessions:
1. Everyday challenges dealing with (lidar) data
2. FAIR data principles
3. Brainstorming next key steps to tackle in the lidar community

We kick off with the first plenum session. The workshop participants were divided into three groups which were moderate by:
Andy Clifton (AC group)
Tobias Klaas (TK group)
Nikola Vasiljevic (NV group)

Everyday challenges dealing with (lidar) data

The group feedbacks on challenges were collected using post-its where two out of three groups rated their challenges on the scale from Small to Big (Challenges). The photos of post-its per group can be found here:
AC group
TK group
NV group
The summary of feedbacks per group:
AC group grouped data challenges into four clusters from the perspective of data user:
- **Metadata level**
At this level we are missing appropriate metadata to understand what data are about and how to use them
- Data handling level
At this level the group identified issues on how and in what appropriate format measurements should be recorded. Also, how to deal with different data structure for various lidar configurations (e.g., LOS, Scanning, etc.). Furthermore, it was pointed out that we probably need a universal/single way to flag missing data (Null or NaN) and provide timestamps. Also, the group indicated challenges they face in dealing with with corrupted data (incomplete datasets), or merging several datasets.
- Data processing level
In dealing with data the group pointed out that it is not that trivial to perform filtering of LOS data (i.e., what is good LOS speed what is bad). Same goes for the retrieval of physical values,i.e. wind vector, from LOS speed measurements and how to deal with volume averaging. Notion and definition of the rage gate position can be also somewhat ambiguous.

- Delivery level
It is uttermost critical to know who are your data users, since probably the level of details regarding the provided data product will be different from one use case to another.

TK group provided challenges from data creators perspective and group them into Big and Small challenges with an Intermediate step:
Small challenges
What is a proper way to flag missing data? Is it NULL or NaN? The data format should agree on a standard way of flagging missing data.
Different standards for time stamps! Let's agree on one! The group proposes to use the ISO 8601 time format.
Intermedium challenges
Naming conventions, can we call same things with one name? Is it LOS speed or Radial velocity or Velocity?! There is a need for a common standardized vocabulary.
There is a need for guidance on how to use new data that is not described within the standard. That aims e.g. on how to generate names for new variables or attributes in order to make the data usable.
What is appropriate metadata structure? Generally speaking the group sees the description of Metadata as a challenge as it is often unstructured, incomplete, inconsistent and processing steps are often not described well.
Big challenges
Information on probe volume and CNR are needed not just LOS speed.
Data have tendency to be unstructured.
Campaign metadata are detached from data and it is often hard to gather all relevant information that might also be relevant for data intepretation.
We need distinct data restriction levels (encryptions) in order to ensure that confidential data is secured. This is of particular importance for data that is provided to third parties but might also be relevant in big companies when the data is provided to other departments.
We need simple APIs to access (pull) and to push data to and from the lidar devices.
Since every producer has it own way of storing data and there are also different naming conventions it is critical to agree on possibility one way of doing this, potentially an intermediate standard/exchange standard (e.g., GPX for GPS data) so it becomes easier to combine data from different lidar devices.
It is seen as a big challenge to combine the data from different devices for one measurement campaign and to carry out an overall analysis.

NV group also provided challenges from data creators perspective which ranged from Small to Big challenges:
Small challenges
The group identified issues such as the complexity of the lidar measurement process which in principle can lead complex data sets that potentially could be not easy to understand by end users. Lidars can create large quantity of data, therefore there is a need for proper approach in data compression especially if that data are to be sent over mobile networks. It is important to think about the future of data processing, for example how our data structure should be in order to survive test of time especially if

we look at more and more present integration/interfacing of different data streams coming from various sensors/services (Big data problem). We do face issues that same data are stored and multiple locations.

Intermedium challenges
On contrary to what was stated that currently the lidar data are complex to use, the group pointed out that we might face the misuse of lidar data if they are too neat. In other words, if the lidar community 'elevate' data to the level at which they become too EASY to be use we might face pitfalls in which lidar data becomes wrongly interpreted by for example data scientist who have very little or no domain (i.e. lidar) knowledge. In contrast to this, today we do face often challenge in which end-users are 'scared' to use lidar data since they are not trivial to process. With the current common practices used to store and describe data their reusability in long-term is somewhat questionable. As pointed in other groups, there is no consensus on vocabulary, i.e. naming and defining terms used in the lidar community. We use different time formats, sometimes data formats are not readable. The way we filter data depends from one person to another, no consensus what is the 'right' way in doing this.

Big challenges
The volume of data we collect today is massive, data catalogues are missing which would greatly simplify identification of 'right' periods to further analyze data (creating insight and value from the data). We have lack of background information. We are not aware on how data were produced, which leads to the question on how to judge the data quality (we rely on the reputation on the data creator). We have issues with poorly structured data format, lack of metadata.

FAIR data principles
In the plenum session that followed the FAIR principle presentation the groups composition and setting remained unchanged. The group participants rated pros and cons of the FAIR principles using Post-Its. The photos of post-its per group can be found here:
AC group
TK group
NV group

AC group:
CONS
We need first movers who will take the risks and immerse into the application of the FAIR principles. The adoption and application of the FAIR principles require time (setup time). Tools to convert data to be in accordance to the FAIR principles need to be developed as well as the tools that work with FAIR data. The adoption of the FAIR principles will be associated/require a steep learning curve. There are knowledge gaps, how many people are actually familiar with the FAIR principles? There is requirement for trust in the community, (if I make my data findable and/or accessible am I exposing too much of my work? Can I as a company actually do this?). There is a large challenge on how to motivate in our case lidar vendors to apply the FAIR data principles. To shift

our work to new paradigma time is needed, thus there are costs associated with this (business as usual does not bring any additional costs at least today).

PROS

The adoption of the FAIR data principles can significantly accelerate entrepreneurialism, i.e. development of products/services. Instead of coding one-ofs we can build code, thus tools, which are reusable by us and by others who are working with the same data format. Due to potentials of registering permanent identifiers (DOIs) for both data sets and code there is a direct mechanism for getting (community) reward. This motivates activities. You are getting more outputs for your euro (i.e. avoiding waste of time/efforts on for example data massaging). Adopting the FAIR principles is a key to align with other communities. It will expose data and lead to the faster starting off with activities. Being FAIR / INTEROPERABLE represents boost sales of products/services.

There is a political dimension that bridges or i.e. weights pros and cons FAIR principles adoption.

TK group
CONS

The adoption of FAIR principles doesn't mean that required work in analysing the data will be easy, especially when we think of complex datasets.
There is a misunderstanding that FAIR mean OPEN. Selling the FAIR principles at high level, e.g. entity management, is foreseen to be hard.
An idea of interoperability sound as an utopia. There is fear that by going for the FAIR data format information about the instrument characteristics would be lost.

PROS

Data creators (finally) are getting visible in the community. Since data are to be visible this pushes boundaries in new fields beyond say wind energy or atmospheric science. FAIR is easy to sell at the low level. FAIR means that data can be used not only by humans and machines but even by ALIENS!

NV group
CONS

There is a question are the FAIR principles only applicable/useful for open data? Due to the required efforts in adopting them ideally there should be a big FAIR data platform for data creators to simply push there data which comes in various formats. By default the FAIR principles are too high level / too general. What do they actually mean in practice? How to implement them? To adopt them we need to 'seek' experts from databases. Being FAIR means that the community has to have a lot of trust. We need set of data access policies or potentially to encode/decode data for security reasons. Do metadata help to understand how data were recorded/created?

PROS

Only publishing metadata could be sufficient to generate new business opportunities. E.g., if an entity A is up to publish metadata cards for specialized measurements they done across the globe immediately they can be recognized as an experts in that specific measurement technique. Orphaned data or data that has been long forgotten

can get exposure leading to generation of new knowledge, products, services. We avoid repetition. The FAIR data are more resilient and robust. The FAIR principles provide a generic way for describing data. Being interoperable means that the lidar technology would gain even better acceptance in the wind industry since it can be interface to much larger systems.

Making lidar data FAIR plenum

The plenum session regarding MAKING LIDAR DATA FAIR, i.e. outcome of the e-WindLidar project, was done around the table. The workshop participants were asked where and how they can contribute to the development of the lidar community FAIR data principles. In the plenum our key areas were identified where the work is needed regarding the FAIR lidar data principles, which are:
1. Lidar data products standardization
2. Lidar data conversion tool to common format(s)
3. Lidar data processing tools (e.g., filtering techniques that work with common format)
4. Dissemination of the e-WindLidar initiative / FAIR lidar data

Per workshop participant these were feedbacks

Norman Wildmann stated that would comply to the standard, since he already uses NetCDF, he would like to contribute to the LIDACO development, he could contribute with tools for creating higher level data products. He also raised several questions such as 'Is the adoption of the FAIR principles and work on the community based standards worth time?', 'Do we know if there are other people who are not on the workshop could be of help?', 'Who are people and other domains (e.g., airport safety) that can push the e-WindLidar initiative beyond the wind energy community'.

Julien Tissot sees Leosphere in the role of an entity that can disseminate the initiative results and help in pushing the lidar data standards beyond the wind energy community especially if their R&D team is happy with the proposed standard since they are currently in the process of switching from Matlab to Python.

Tobias Ahsbahs would like to help in designing data structure particularly from the perspective of compatibility of the structure with the existing Python libraries such as for example XARRAY.

Robert Menke would like to help in testing of Lidaco and he would like to share his tools for data processing.

Antoine Larvol implied that Windar can implement the new data standards directly into their lidar, they are open to implement whatever standards/protocols the community find prefered. Similar to Leosphere he sees that Windar can help in disseminating the lidar data standards via their customers.

Elliot Simon would like to test LIDACO and push the converted datasets into online data catalogue, help building first examples of the LIDACO application and also to convert his data processing tools to work with the new data standard while helping in the dissemination of the e-WindLidar initiative.

Arezoo Zamanbin will encourage WindScanner customers to use new standard and improve the WindScanner software to store data in the new format.

Leonardo Alcayaga see his contribution with data processing tools that work on top of the lidar data standard.

Sean Daly sees that he can contribute to the data processing toolbox development. He intends to push the e-WindLidar initiative through DNV GL.

Peter Clive stated that he and his team in Wood Group will test the conversion tool and see how it works with some datasets they can make accessible with the community. Probably there is a good chance of implementing the lidar data standard into the new lidar Wood Group is developing.

Andy Clifton: sees his role in disseminating the e-WindLidar activity through Task 32. He will initiate developing and hosting of the glossary of (lidar) terms which will lead to standardization of lidar vocabulary and definition of terms. Also, he will find an appropriate person from Uni Stuttgart to develop reader for Halo lidar they operate and to check with Wood Group whether the current Halo format is equivalent with Gallion lidar.

Pedro Santos: would like to contribute to the lidar data standardization, test the current converter and build data processing toolbox.

Tobias Klaas and **Nikola Vasiljevic** would like to remain key contact points and main locomotives regarding the e-WindLidar initiative keeping the community on track regarding the lidar data standardization, tool box development and initiative dissemination.

An overview table has been composed based on the feedback from the workshop participants (see below)

Name	Data structure	Conversion tool LIDACO	Data processing tool box	Dissemination
Nikola Vasiljević	x	x	x	x
Tobias Klaas	x	x	x	x
João Correia Lopes				
Daniel Gomes		x		
Elliot Simon		x	x	
Norman Wildmann	x	x	x	x
Tobias Ahsbahs	x			x
Robert Menke		x	x	
Antoine Larvol				x
Rozenn Wagner				
Arezoo Zamanbin	x			x
Pedro Santos	x	x	x	
Leonardo Alcayaga			x	
Neil Davis	x			
Alex Forsting	x		x	
Julien Tissot				x
Peter Clive				x
Sean Daly			x	x
Andrew Clifton	x			x

Brainstorming actions follow-up

The groups were asked to propose actions in relation to the FAIR lidar data principles adoption / e-WindLidar initiative and position them with respect to Effort (easy to hard) and Benefits (low to high). Also the groups were asked to select three key actions that need to take place ASAP. The result of this last plenum session were presented via post-its. The photos of post-its per group can be found here:

AC group
TK group
NV group

AC group

Stated that it is necessary to test universal converter (Lidaco) and start working on upper level products (i.e., Level 3 - reconstructed wind field). We need to get more users and provide more examples since this will boost the dissemination and adoption of the e-WindLidar initiative.

TK group

In similar fashion TK group stressed importance of testing what we already developed, i.e. dissemination and engagement of new users. Get feedback from this tests and together with glossary of terms (lidar restricted vocabulary / taxonomy) agree on the final data standard / convention (i.e., e-WindLidar version 1.1). Proceed to Level 3 data products and setup next workshop on e-windlidar with a particular focus on lidar data filtering.

NV group

This group identified need to first generate 1.1 version of Level 2 standard through both study of the current data structure (abstract approach) and by testing lidaco (hands-on approach, which would need documentation of Lidaco). In parallel we should run dissemination of our ongoing work.

We agreed that the workshop participants will be divided into two work groups (WG) handling working on :
WG1 - Data structure (green in list above)
WG2 - Tools development and testing (yellow in list above)

WG2 entails two sub groups: (1) testing lidar converter and building readers, (2) developing and testing data processing tools.

Based on the feedbacks we formed following two working groups:
WG1
lead: Nikola Vasiljevic
members: Tobias Klaas, Elliot Simon, Tobias Ahsbahs, Arezoo Zamanbin, Neil Davis, Andrew Clifton and Norman Wildmann

Submitted by: Nikola Vasiljević

WG2

lead: Tobias Klaas

members: Robert Menke, Pedro Santos, Leonardo Alcayaga, Alex Forsting, Sean Daly, Daniel Gomes and Nikola Vasiljevic

All participants of the workshop should in their own way try to disseminate the work of e-WindLidar initiative.

It was agreed that the final version of the e-WindLidar report will be extended by new contributors who joined the initiative.

Workshop presentations available via this link:
https://drive.google.com/open?id=1kRwWaM9LVxr6oImR26SU-IdvXnEQh2zP

2. Action List		
Action	Involved	Due Date
Prepare e-WindLidar Slack to accommodate more participants	Nikola	20.11.2018
Invite people to e-WindLidar Slack	Nikola	20.11.2018
Glossary of terms, common vocabulary	Andy	
Lidaco documentation	Daniel	16.11.2018
Reader documentation	Tobias	16.11.2018
Division of people to work streams	Nikola & Tobias	
Examples of lidar data conversion + data processing (from zero to hero)	Nikola, Elliot, Tobias A., etc.	31.01.2019
Forming three work groups	Tobias & Nikola	19.10.2018
Check Jupyter Notebooks provided by Tobias Ahsbahs and put them on lidaco GitHub	Tobias	16.11.2018
Distribute follow-up e-mail with suggestions on work groups etc. / follow up tasks	Tobias & Andy	03.11.2018
Finalize e-WindLidar report and publish it on Zenodo	Nikola	31.12.2018

Submitted by: Nikola Vasiljević