WEMC Data Exchanges, Access and Standards Special Interest Group (Data SIG) ICEM 2019 Workshop: *New data in energy applications, how to build efficient integrations and reduce renewable volatility planning* – A summary

When: Wednesday 26th June 2019 11:15-13:05 CEST
Where: DTU Lyngby Meeting Centre, Building 101, Room S1

Session Description:

Over the last 5-10 years there has been a significant increase in renewable energy capacity that has led to a change in a different kind of weather intelligence by the Energy Industry to balance the energy in the short term. Nowcasting information has become an essential ingredient. Over a similar time period there has been an explosion of nowcasting data available to weather service providers, some from more conventional routes such as satellites and radar and some from more unconventional sources such as mobile phones, vehicles etc. In order to work together for both the Energy and Meteorology sectors there is now a bigger challenge to handle new data types and handle these to deliver the services that are needed now and in the future. The purpose for this session, in cooperation with the Royal Meteorological Society, is to understand the requirement and explore how that can be achieved, focusing on the data aspects (collection, access, sharing, formats etc), which will be one of the main focuses of WEMC's Special Interest Group on Data Exchanges, Access and Standards. An outcome of the session is to establish recommendations for the benefit of both sectors.

Target attendees: Academia, research organizations, SMEs and end-users
Session lead (moderator): Liz Bentley, Chief Executive, Royal Meteorological Society

Summary of Outcomes (Liz Bentley/Shanti Majithia):

Collation of post-it notes (see appendix), discussion points and thoughts from the workshop in general reveal four key messages to take away, with suggested action points for the Data SIG in blue (Kit, unless otherwise stated).

1. A complete an audit of all the datasets / databases that people in the community are aware of and use would be useful. Capture information about whether data is free to use, format, data fields etc. Investigate the feasibility of setting up a portal (maybe on WEMC website) of all the datasets that exists, with the aim to achieve the above. Anna-Maria Sempreviva: Would it be it a repository of all data? Or a metadata catalogue of data sparse in different Institutions repositories?[1]
2. There was interest from those who attended the workshop for a common shared database (on a platform like IBM or Amazon Cloud) which can be freely available. WMO template could be a start. Evaluate the WMO template and assess to what it extent it could be used, either directly or as a guide.

3. Develop standards / international definitions of the raw or / and process data, including data collected including from mobiles etc. Review outcomes/thoughts from Data SIG webinar (13th June 2019) which can begin to address this.
4. Financial model to make data freely available and sharing cost neutral either via a subsidy, commission or state funding. Identify a person/persons/organisation/research working on this.

[1]A metadata catalogue is already used by WMO in the Arctic database http://arcticdata.met.no/metamod/search hosted by Met.no. This is how Anna-Maria Sempreviva got the inspiration for the Wind Energy data strategy. This is also along the EOCI policy i.e. "networking" in Europe data infrastructures/model repositories.

It is recognised that is important to ensure the above is future proof considering new data which Utilities/Energy/Environment industry require now but also in the future.

Talk Summaries:

Challenges to Improve Predictions in the Renewable Sector – Shanti Majithia (Director of Energy and Climate Advisory Services (UK) & RMetS Fellow)

There has been a significant increase in renewable capacity. Wind power has become an important source of energy generation around the world, with global capacity reaching over 600GW in 2018, with China accounting for 221GW. For solar, the total installed PV power capacity grew by 25% to 510 GW by the end of 2018. When looking back just a decade, the world's cumulative PV capacity increased by over 3,200% from 15.8 GW in 2008. By 2023, solar PV is expected to lead renewable electricity capacity growth, expanding by almost 580 GW, due to advance in technologies, a steep cost reduction curve and low generation costs compared o new nuclear and coal plants.

Using the UK as an example, there are a number of websites that allow you to see real time data for the power generation in the UK National Grid. One example is 'GridWatch' (https://gridwatch.co.uk/). Websites like this demonstrate how solar is becoming an increasing part of the energy mix. Kit: Another example which graphically shows historical power generation is http://grid.iamkate.com/. The graph for *Sources: All time (quarterly averages)* show substantial increases in both solar and wind in just this decade.

Example of solar energy prediction in the UK: managing short term forecast and risk management and minimising volatility in forecasting models – The crucial time period for improving forecast accuracy is between 2-6 hours. Solar radiation data from 53 Met Office stations report hourly, helping to generate the first forecast at 7:30am local time. Error predictions are useful, the control room is interested mainly in the 5th and 95th percentiles so they can compensate for uncertainties. In addition, having examples of where forecasts went wrong and where their were demand errors can help to reduce financial impacts, as an error of 100MW equates to £28m annually. Current meteorological information is not as accessible or affordable as it could be. More data sites are needed, but also different weather providers (than just the Met Office) at a higher temporal resolution than hourly. Access to both historical observations and forecasts at the same location would aid machine learning. There are a number of gaps in the meteorological data. For example, the is no obligation to submit data observations from solar farms but this may soon change.

Within the next 2-5 years, there are a number of data sources, exchanges and processes likely to become available which would be helpful with solar and wind prediction. Research particularly at Sheffield Solar (University of Sheffield) has the potential to help energy companies and met service providers to work together, leading to better data and therefore better forecasts. There is also an opportunity to gather data from not just conventional routes (e.g. satellites and radar) but also unconventional routes such as smartphones and vehicles, both which can help improve nowcasting.

Both the energy and meteorology sectors need to work together to address this challenge for better high quality and new data types to deliver the services that are needed now and in the future. What should data requirements or a design specification look like? This would be a useful topic for discussion. Also, it is key to work together and narrow the gap between data providers and data users, considering huge investment already committed in the renewable sector (see World Energy Outlook April 2018).

References/Acknowledgements (not cited in-text):

- *Thank Alasdair Bruce previously of National Grid UK and National Grid to share information their solar prediction background.*
- *SunReign Power Market: (http://www.powermarket.net/#how-it-works)*
- *Solcast: https://solcast.com.au/*
- *Meniscus: http://www.meniscus.co.uk/wp-content/uploads/2019/01/Case-study-Short-term-solar-irradiance-predictions-and-impact-on-PV site-revenue.pdf*

Link to Presentation PDF

How and Why Data Sharing Should Be Developed – Anna Maria Sempreviva (DTU Wind Energy)

The advancement in the renewable energy field needs innovative solutions in terms of products, services and business models for securing a stable flow of energy to society. This needs better integration of data from different sources. However, stakeholders generally consider the data they collect as assets which can give them a competitive advantage, and therefore the data remains non-accessible and often secret. To instil confidence on the Open Data and FAIR (Findable, Accessible, Interoperable and Re-usable) data principles in H2020 projects, the European Commission, declared that data must be as open as possible and as closed as necessary.

After initial diffidence, Open and FAIR data principles are increasingly considered with interest, as it is becoming clear that sharing data can multiplying the value of owned data by allowing re-use for multiple applications; more-over sharing data would allow making the workflow from idea to innovation more efficient, thus faster.

In her talk, Dr. Sempreviva introduced why and how data sharing should be developed and generalized and presents a possible strategy for controlled data sharing within the FAIR and “as open as possible and as closed as necessary” principles.

The idea is simple and based on Information and Organization Retrieval basic concepts: metadata and taxonomies. Tagging Research data allows organizing data in networks of searchable thematic metadata catalogues distributed at different levels, while data are kept accessible “on demand”. Data users with ideas can connect with data owners with data, data is findable through a faceted search, but the access of data is controlled.

A possible development includes a ‘data marketplace’ with different business models for data sharing. This strategy has the potential to make stakeholders visible through data, and to support cross-fertilization by connecting stakeholders from different sectors and disciplines.

Publication - *Open Science: Sharing Data, Tools and Workflows. A Strategy to Inspire Efficient Collaboration (Sempreviva et al. 2019)*

Link to Presentation PDF

Data Difficulties, Dealing with Issues, and Potential Paths Forward to Improve the Wider Use of Data – *Sue Ellen Haupt (NCAR/WEMC)*

Sharing data can aid situational awareness and enable prediction for the good of all. It 'takes a community' beginning with the end user, all the way back to the research community and funding organizer, for effective and productive data sharing. What really needs to happen to connect the research to the end user is some sort of operations translation communication.

A challenge that arises from such a range and volume of data, is that it comes in very disparate forms. Some are point data, some are gridded data, some are pixel data. There are large volumes of data. Inconsistent timestamps are an issue. For example, whether UTC has been used versus a local time. And if in local time, does it include changes for daylight savings in those parts of the world that go to daylight savings time? Time zone issues are important. For example, in the United States some utilities span multiple time zones and it has been observed that the time zone on the data depends on which time zone it has downloaded in rather than the location of a particular plant. These challenges amongst others demonstrate it is critical to have standardized data.

Not addressing data issues can increase costs, raise frustrations and degrades the quality of forecasts. Amongst the benefits of doing so are the aiding of data simulation training, AI algorithms, helping to initialize real time algorithms and verify models and helping with the smooth real-time operation of wind and solar plants (by knowing information on temperatures, exact cloud cover, etc). It is clear that by bringing in all these types of data in the meteorological community and especially in the applications in the energy interest industry, and now with the 'internet of things' (something Haupt and NCAR have working on for a while) there is an opportunity to bring in large amounts of real time data and integrate it in ways that provide decision support.

Haupt and her colleagues at NOAA have performed parallel research that demonstrate when multiple data sources are utilised from, say, wind farms, then publicly available forecasts are better. This benefits the entire community including the renewable energy community. Various studies including some that NCAR has done, have shown that the NWP model can do better when we have very local data.

Haupt finished by presenting some recommendations. For instance, utilities should start sharing their data with national centres so that the national centres can improve their forecast for everybody, including the energy industry. Historical data needs to be saved as it is valuable for training machine learning algorithms. Data needs to be recorded in standardized formats such as using UTC timestamps to bring them in line with meteorological timescales, and having more data is going to improve the situation for all stakeholders, especially end-users in the energy industry.

Link to Presentation PDF

Appendix 1 – Post-it note responses to the 'Future challenges/opportunities' discussion:

- Resilience – What data is needed to improve resilience? *Including short term resilience from extreme weather events and over longer timescale as our climate changes.*
    - Citizen science
    - Co-located, high frequency (5 min or finer) weather and power observations, available in real time
    - Backup of data
    - Historical memory data, old fisherman observation
    - Humidity, dust, quality demand
    - Real-time data, quality filtered, historical trends, forecasts with uncertainty estimations
    - Precipitation data, global historical variation, elevation/altitude, demographic information/location, alternative renewable, wind speed
    - Human intelligence
    - Bring along failure and operations/maintenance data as well
    - IOT collecting data worldwide
    - Collection of failures/problems or errors and solutions to overcome them
    - Sensors on cell phones – pressure/tmp, other devices
    - Long-term climatological
    - Smart meter data, citizen science
    - Historical storms
    - Historical data, meteorological observations, power system
    - Document disaster (common people): Taking photos for example is proof of an extreme event happening
    - Homogenization of data to see if climate change or because of e.g. relocation
    - Vulnerability
    - Extreme events, meteorological, power data
    - Power station, location, flood resilience, topographical data, rainfall data, flood maps
    - Electricity grid, network topology
    - Real time data, power, meteorological at same location
    - Short-term: Upstream data, how to harden the data transfer?
    - Detailed installed capacity, worldwide energy use
- Identifying gaps – Opportunities to reduce gaps that exist between different communities *(Including knowledge gap between weather data providers and data users. Data sharing. Datasets. Data formats. Data standards / Knowledge transfer, networking, facilitating transfer of information, pilot studies - proof of concept etc)*
    - Glossary
        - Definition used in different
        - Linking different communities taxonomy
        - Use existing glossaries

  - Sharing software tool
    - Link to other formats
  - Avoid complex downloading
  - Automatic downloads
  - Dissemination
  - Missing globally agreed procedures
  - DAT market plan with a business model
  - Trust
    - Public/private partnership
- Observational data – What does the future hold? *(List of observational data that is coming online or already available but not used. / List the potential use of this data to the energy sector.*
  - Climate data explosion
  - Should met data be available for free?
  - Lack/hard to find met data
  - Production data not available
  - Location data?
  - Data hub: Amazon/IBM
  - Precipitation data calibration
  - Working in silos
  - Cloud computing will help to share data
  - Metered data
  - Data protection
  - Share data
  - Find data accessible - no
  - Cloud source data in meteorology to connect to computing
  - Energy data not easily available
  - Table/not easy to communicate
  - Wiki idea
  - Tornado data
  - Not all measurements stored and saved
  - Global agreement
    - Public/private partnerships
    - Licenses
    - Share but not with everybody
    - Trust/collaboration
    - Taxpayers
    - WMO

Appendix 2 – Post-workshop ideas/thoughts/responses etc…

Thoughts from Anna-Maria Sempreviva (13/8/2019):

Collation of the 'post-it' notes: Further subcategories must be created as there is inhomogeneity. For example in *"What data is needed to improve resilience?".* We find data (variables) and Actions (Backup of data, dissemination…) and others e.g. Trust, Citizen science.

I have a comment on *"Investigate the feasibility of setting up a portal (maybe on WEMC website) of all the datasets that exists, with the aim to achieve the above.":* I think here we should explain which type of database: is it a repository of all data? or a metadata catalogue of data sparse in different Institutions repositories? Actually, the latter, the metadata catalogue is already used by WMO in the arctic database http://arcticdata.met.no/metamod/search hosted by Met.no. That is how I got the inspiration for the Wind Energy data strategy. This is also along the EOCI policy i.e. "networking" in Europe data infrastructures/model repositories.

This document was finalised on 2nd September 2019 Any further ideas and thoughts please email Kit at kit.rackley@wemcouncil.org.

Members of the WEMC Data SIG are encouraged to use the discussion forum at: http://www.wemcouncil.org/wp/forums/forum/data-sig/ (WEMC member log-in required).