Nano Workshop report

Authors: Paula Carroll, Alireza Nouri

UCD Energy Institute held a workshop on 22nd February 2021 on the theme of Smart Grid Research Data - how to share research data, ensure GDPR compliance, reduce the risk of divulging potential innovation insights, and related issues. The webinar focused on how different stakeholders such as governments, industry and local energy communities can share and organise their data respecting the privacy of users and citizens.

This nano workshop was held online due to the ongoing Covid-19 restrictions. It followed the in-person PANTERA workshop in Dublin on 2nd December 2019. As noted in D5.2, participants at the 2019 workshop identified "access to data to build and prototype data driven models" as a significant challenge. The outcomes from that workshop in 2019 concluded:

1. There is significant activity and interest in Smart Grid R&I across industry and academia in Ireland.
2. There is a demand for access to relevant smart grid data to allow building of prototype models.
3. The deregulation of the market, and GDPR rules make cooperation and data sharing among actors in the electricity space difficult.
4. The proposed PANTERA platform offers an opportunity for data and knowledge sharing.
5. The PANTERA regional desks offer an opportunity for networking for the Irish Smart Grid R&I community, participants expressed interest in follow up activities.
6. The desired functionalities could be of use and truly supportive to the R&I community within the interactive platform.

In response to these conclusions, the nano-workshop aimed to showcase some successful data rich smart grid research projects. Expert speakers discussed their work followed by an open discussion facilitated by moderators Dr Paula Carroll and Dr Alireza Nouri from the UCD Energy Institute.

Webinar Format

Speaker briefing notes were provided to each of the panellists in advance of the webinar explaining the PANTERA objectives. The speakers covered a range of themes: the need for data to prototype data driven models; a case study in the role data played in informing renewable energy policy; commercial and legal aspects of data in research and innovation projects; and on overview of the PANTERA project. The webinar agenda was as follows:

1. Introduction (Facilitator Dr Paula Carroll, UCD Energy Institute)
2. Prof Eleni Mangina UCD Energy Institute, "*Why Machine Learning Needs data to unlock the potential of Smart Grids*."
3. Jim Scheer, Sustainable Energy Authority of Ireland "*Unlocking energy data to achieve decarbonisation targets*." Experiences of accessing Smart Meter Data, case study of gas and electricity data usage, based on an example of 200 homes.
4. Dr Stephen Donoghue "*Commercial and Legal Difficulties for Data Sharing in Energy Markets*"– Legal Aspects and Challenges; tips and advice; collaboration agreements
5. Dr Venizelos Efthymiou, FOSS, PANTERA opportunities for data sharing via EIRIE platform
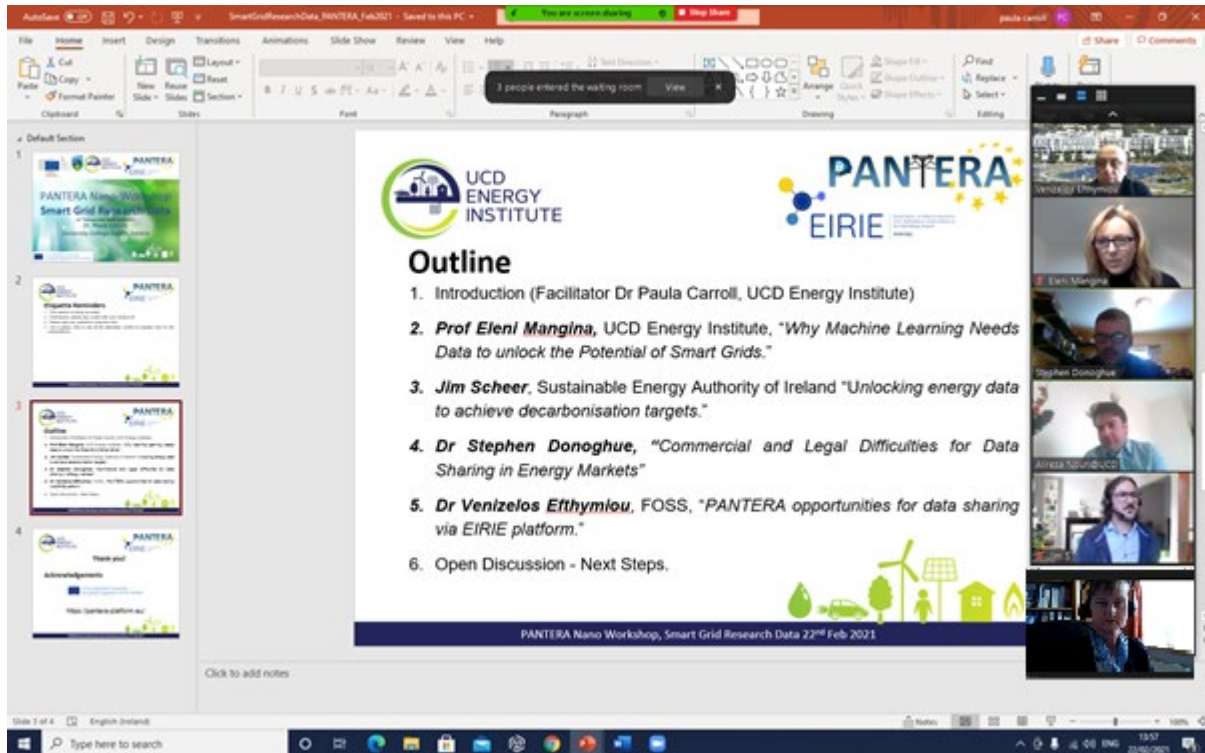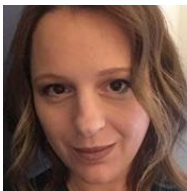6. Open Discussion - Next Steps.

Figure 1: Getting ready to admit the participants to the Nano Workshop. Top to bottom: Dr Venizelos Efthymiou, Prof Eleni Mangina, Dr Stephen Donoghue, Dr Ali Nouri, Mr Jim Scheer and Dr Paula Carrol.l

Speaker Bios

Prof Eleni Mangina is a **professor is the School of Computer Science**, University College Dublin, Ireland. Eleni carried out her PhD work at the University of Strathclyde (UK), Department of Electronic and Electrical Engineering, working on Agent-based applications for intelligent data interpretation. Her lab operates at the intersection between applied Artificial Intelligence (VR/AR; Data Analytics; UAVs; Information Systems) and a portfolio development within interdisciplinary applications (i.e. Engineering [Energy Sector] and Educational Systems.

Jim Scheer has over 18 years' experience working in the field of policy analysis and development related to environmental issues. He joined SEAI in 2007 and is currently **Head of Department (Data and Insights)** responsible for energy statistics, modelling, behavioural economics, and finance at SEAI. He holds a Professional Diploma in Advanced Management Performance (Smurfit Business School), MSc. Economic Policy Studies (Trinity College Dublin), BSc. Environmental Science (Flinders University, South Australia). Jim is passionate about getting people connected to the need for climate action now.

Dr Stephen Donoghue is the **UCD Research and Innovation Case Manager**. Stephen is responsible for managing the identification, protection and commercialisation of University College Dublin research outputs within the environment area. He has a background in biotechnology, software and law and has significant experience in managing and transacting data.

Dr Venizelos Efthymiou worked for Electricity Authority through various posts including Executive Manager Networks / Distribution System Operator of Cyprus from 1979 to 2013. He is serving the Steering Committee of ETIP SNET, ETIP PV, SET Plan and Horizon 2020 Programme Committee. He is the **Chairman of FOSS Research Centre for sustainable energy and PANTERA consortium lead** of the University of Cyprus and CIGRE Cyprus

Webinar Motivation

Cities are central to a worldwide climate action plan and the implementation of net-zero development strategies [2] Developing smart cities is an important part of the solution for a sustainable future of our society. Smart cities also provide energy flexibility, storage, and energy generation capabilities, either as infrastructure (e.g., building envelopes, thermal storage, batteries, electric vehicles) or by means of the associated energy systems (e.g., heat pumps, boilers, cogeneration, renewables). To achieve the necessary transition to a low carbon economy in cities, it is essential to increase energy systems integration and to push efficiency performance levels significantly beyond current limits by 2050. Such an ambitious object can be achieved only if more stakeholders can share their private usage datasets without violating privacy rights. It is therefore paramount to be acquainted with anonymization data processes and the terms of the EU GDPR.

Participants at the PANTERA Dublin workshop identified a gap in the availability of appropriate data to build and test smart grid product prototypes in such a context, we are aiming to organise an open forum for smart grid stakeholders such as citizens, government, industry, and academia to facilitate data sharing and data processing techniques. This series of PANTERA nano-workshops will directly focus on how different stakeholders such as governments, industry, and local energy communities can share and organise their data respecting the privacy of users and citizens. The workshops will also provide a forum to explore opportunities among stakeholders to develop new data-driven solutions for sustainable smart cities and communities. Four different areas have been identified as pivotal for the smart grid open data community: (i) community data, (ii) energy-efficient and smart-ready buildings data, (iii) local production from RES, (iv) infrastructure, finance, and enabling policy-related data.

Having a shared data platform is a paramount step to tackle the enormous number of multi-sectoral and multidisciplinary challenges we need to face to achieve the target of a more sustainable future society. The nano-workshop will provide a forum where industry, universities, citizens can foster collaborations and organise new projects and prepare data sharing pipeline, and address the role the PANTERA platform could play as a portal to or server for identified data streams.

Webinar Summary

Dr Mangina noted that Big Data is driving down the cost of power generation and enabling sharing and recycling of energy. Smart Meter data exhibit the Big Data five V's of Volume, velocity, variety, veracity and value. She remarked that access to suitable data is a stumbling block to allow demand response and flexibility in the power system. She queried if power companies can make anonymised data available to unlock the value of flexibility. Open issues include the lack of interoperability and data sharing standards. She gave several examples where data had been used to create data driven control strategies.

Mr Scheer describe a project that accessed and assessed electricity and gas billing data to evaluate a home retrofit scheme. He noted that the data was acquired prior to the current GDPR regime, but that it was used to extract insights and create an evidence base for further policy design.

Dr Donoghue spoke about the principles of copyrighting and noted that the processing of data adds structural value, as it can be copyrighted and licensed. Similarly, computer code is considered copyrightable. He discussed the different types of data analysis and whether a dataset could be reverse engineered to identify a natural person. He outlined the CC license types as standardised ways to grant copyright to data or other creative works.

Dr Efthymiou described the objectives of the PANTERA project and its progress to date. He highlighted that the EIRIE platform is expected to be available in Spring 2021.

There was strong interest in each of the presentations, with participants submitting questions via the chat and raising their hand to speak.

Participants analysis

There were 85 registrations for the webinar. Figure 2 depicts the countries of the participants' institutions, whereas figure 3 shows the primary interest of the registrants.
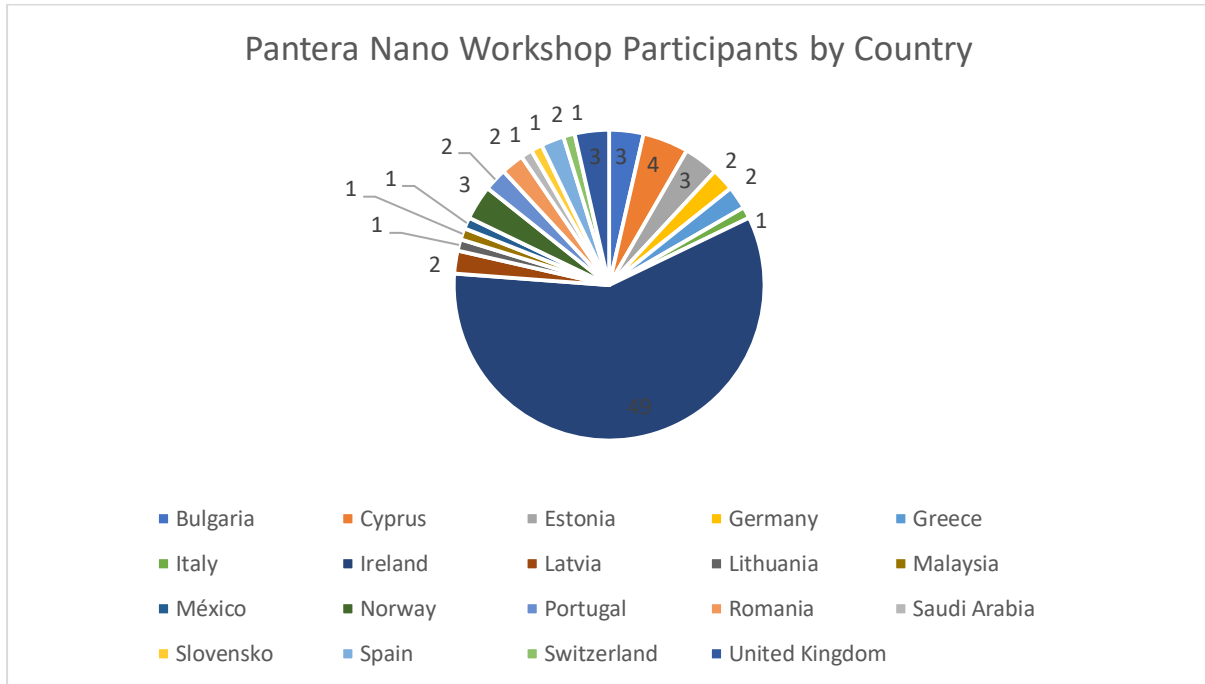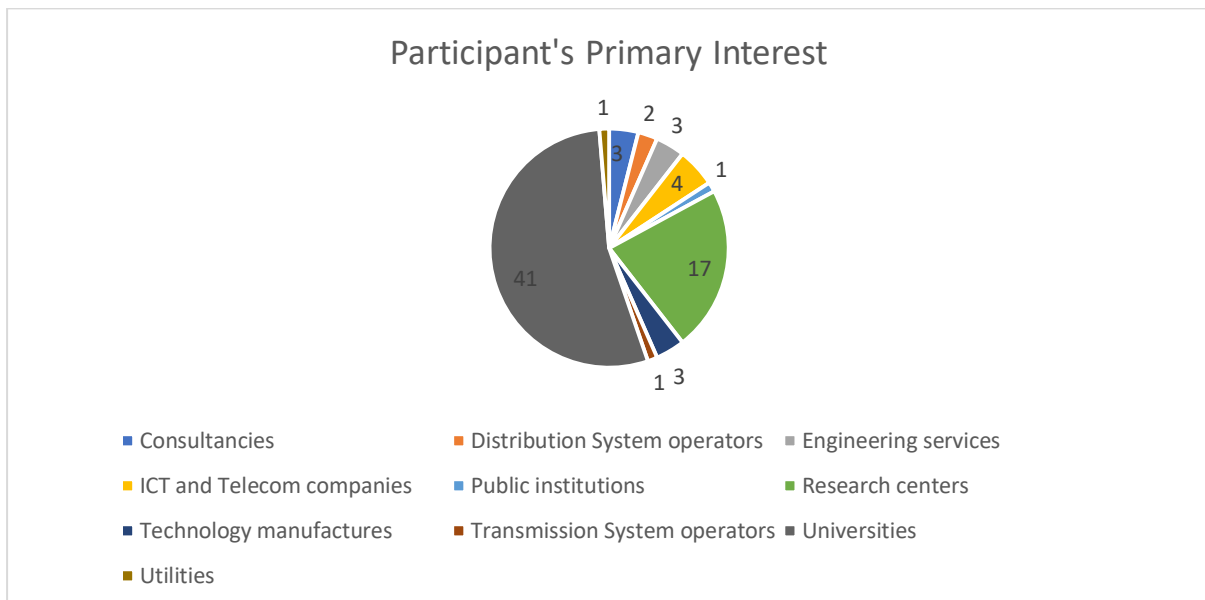

Figure 2: Participant Registration by Country

Figure 3: Participant's Primary Interest

Summary of Queries and Answers

During the presentations, some questions were raised by the participants. The respective speaker answered these questions (orally after the presentations or within a text chat that was available to all participants). The summary of some questions and answers is provided here, as well as abbreviations, which are explained for more clarity.

Questions for Eleni Mangina:

Q1 I was just wondering; we have all these big successful deep learning models that are blowing everything else out of the water. But then you hear stories about some large models requiring as much energy to train as a town or small city. At what point are our models too big?

Another participant also followed up on this question: Good question! For instance, micro grids - with a decentralised control, or blockchain with multiple layers.

A1 (Eleni Mangina). Eleni discussed this point and also answered in the chat: The efficiency of the algorithms is a very important factor. See an interesting paper on how to estimate the EC (Energy Consumption) in ML (Machine Learning). https://doi.org/10.1016/j.jpdc.2019.07.007

Q2 Hi Eleni, is there any emerging standard for data gathering and inspections, schemas for data storage/cloud?

A2 (Eleni Mangina). There is current work on standards for Smart Grid and you can follow information updates at EU level at the following link:

https://www.cencenelec.eu/standards/Sectorsold/SustainableEnergy/SmartGrids/Pages/default.aspx

I strongly believe that more emphasis should be given to the data standards. Data sharing goes hand in hand with data standards.

Questions for Jim Scheer:

Q1. The consultation paper from the Dept of the Environment, climate and Comms for a Microgenaration Support Scheme - MSS, closed last week. In it, a micro-generator must have first met minimum energy efficiency. Will this be what your measurements be use as the standard please?

A1 (Jim Scheer). The researchers are providing insight to the government. The government is now asking experts and consumers, which way should we set the level of efficiency standards before we are going to support you on renewable technologies. That is important from different aspects. We only support the policymakers in the Department of Energy, but we are encouraging them not to avoid supporting people who are willing to take some actions.

Questions for Stephen Donoghue:

Q1 In a decentralised data structure, DLT (Distributed Ledger Technology)[1] for example, there is no such thing as ownership of data? But Agreement for data sharing. Right?

Weather data for example, is available from weather data API service provider and chargeable. We use such service from Germany in our development works. Will the Irish authorities going to license these types of services here?

A1 (Stephen Donoghue). There are public and private players in the provision of weather data. The open gov has access to free data under the cc (Creative Commons (standard) licences) by 4 licence[2].

---

[1] See https://en.wikipedia.org/wiki/Distributed_ledger for more information on DLTs

[2] See https://creativecommons.org/about/cclicenses/ for more information on the six CC licenses

Commercial providers will have commercial agreements and sometimes non-commercial agreements for university researchers.

Comment1 (Another participant). The issue smart Grid in NI (Northern Ireland) is facing, in recent years, is all about data ownership. Grid operator (energy supplier) installs the smart meter which produces the data for electricity use, or able to sell as prosumer. Consumer should have the claim of the ownership of the data is they are the energy producer and able to sell at "fair price" as the Irish government is consulting the public.

Q2 (Eleni Mangina). There should be at high-level decision on sharing data - it is consumers' data. Shouldn't these belong to the consumers? Shouldn't I be in a position to decide whether my data will be used?

A2 (Stephen Donoghue). In principle, I agree. If a person/consumer generates data they should be in a position to decide whether it should be shared, especially as it has personal data and privacy aspects associated with it. What makes the farmer's data movement in the US complex is that the farmers wish to get paid for the provision of technical data from machines they buy and use. Who "owns" that data?

Q3-1 It's hard to believe that car routing data can be really anonymised in a realistic manner... it's crazy that the data can be claimed like that.

A3-1 (Stephen Donoghue). Privacy rights vary enormously from jurisdiction to jurisdiction.

Q3-2 Surely that couldn't fly under any sort of scrutiny in the EU?

A3-2 (Stephen Donoghue). Probably not, but by purchasing the car you might be giving rights to the company to use some sort of anonymized version of the data.

Questions for Venizelos Efthymiou:

Q1 Great work Venizelos. Indeed, such platform would bring together and enhance the management and sharing of Smart Grid Networks around EU. Would this platform be open for private companies to contribute or only to research institutes?

A1 (Venizelos Efthymiou). Indeed, we know that data can come from anywhere, and that is why we're putting in place a process through which we will operate data managers, thus any one will have the possibility of uploading on the platform, but what will be uploaded is not going to be published unless it goes through an approval process, because as we said earlier, we want to generate the data that is dependable and usable by all. So yes, it is open to everybody and we have a process in place through which you will identify the source of the received data. We will modify if required the content and upload in the system, once it passes our internal policies and quality control.

Webinar Outcomes

According to the talks and discussions among the speakers, experts and participants around the webinar theme, there should be a rigorous methodology for research for the implications of data protection in Smart Grid. The standardisation of research data sharing (Eleni's talk), data ownership and privacy rights (Stephen's talk) are among the challenges that should be overcome to be able to exploit the customers' data, ranging from electricity demand to e-mobility and car routing (James's talk) and boost the research and innovation on the development of smart grids and promote the energy transition (Venizelos's talk).

The main conclusions of the workshop can be summarised under the following strengths and weaknesses in relation to Smart Grid Research Data as follows.

Strengths include:

1. The move toward open data means some open-source data is available that can be used to prototype models, and the SEAI is actively working to facilitate data availability from projects funded by its Research and Development Programme.
2. Synthetic data can also be created using statistical models.
3. Building archetypes can be used to support research.

Weaknesses include:

1. Lack of standardised data formats and interfaces.
2. Challenges to access to end-users' data even with end-user's permission.
3. Lack of a high-level decision-making committee on research sharing data.
4. Lack of legal and commercial knowledge with respect to data among the smart grid R&I community.