DATA SCIENCE: FROM ACADEMIA TO INDUSTRY

Making impact in the energy sector

Stephen Haben
Digital And Data Consultant (ESC)

Samuel Hinton
Senior Data Scientist (ARENKO)

JULY 2022

CONTENTS

1. THE URGENT NEED FOR INNOVATION 2

2. SUMMARY AND RECOMMENDATIONS 3

2.1. TECHNICAL KNOWLEDGE: SUMMARY & RECOMMENDATIONS 3

2.2. TRAINING AND SKILLS GAP: SUMMARY & RECOMMENDATIONS 4

2.3. COLLABORATIONS: SUMMARY & RECOMMENDATIONS 5

2.4. ACCESSIBLE AND REPRODUCIBLE RESEARCH: SUMMARY & RECOMMENDATIONS 6

2.5. CODE DEVELOPMENT FOR ACADEMICS: SUMMARY & RECOMMENDATIONS 7

2.6. INDUSTRIAL SUPPORT FOR ACADEMICS: SUMMARY & RECOMMENDATIONS 8

3. INTRODUCTION 9

3.1. INFORMATION COLLECTION METHODOLOGY 11

3.2. OUTLINE OF THE REPORTS 12

4. TECHNICAL KNOWLEDGE IMPACT 15

4.1. TYPES OF IMPACT 15

4.2. CHALLENGES TO MEASURING AND IDENTIFYING IMPACT 16

4.3. CHALLENGES OF PRODUCING IMPACTFUL RESEARCH: DATA AND FUNDING AVAILABILITY 18

4.4. CHALLENGES WITH UTILISING IMPACTS 20

4.5. WAYS TO IMPROVE KNOWLEDGE TRANSFER IMPACT 22

5. TRAINING AND THE SKILLS GAP 27

5.1. DATA SCIENCE SKILLS GAPS 27

5.2. GRADUATES FOR BUILDING CAPACITY 28

5.3. OPPORTUNITIES AND NEEDS 29

6. BRINGING IT ALL TOGETHER 31

6.1. SCENARIOS 31

6.2. FUTURE WORK 33

7. BIBLIOGRAPHY 34

1. THE URGENT NEED FOR INNOVATION

This document is designed to help the energy sector and academics find better ways to collaborate and share knowledge of the latest data science research, and lead to a more rapid transfer of findings and solutions from academic research into industry and operational impact.

To decarbonise the energy system will require intense co-ordination across, academia, government, regulation, and the wider industry. The recent price hikes have also demonstrated that consumers are particularly vulnerable due to the UK's dependence on conventional generation and the market coupling of the electricity and gas prices. A considerable contribution of the UKs carbon emission outputs is from heating and transport sectors with some recent figures reporting contributions up to 37% and 27% respectively. Electrification is viewed as a major solution to reduce these emissions.

Unfortunately, increased electrification will also exacerbate and increase costs and demand on the network. To keep costs contained and reduce the impact on the network will require utilisation of smart and advanced solutions such as:

- Variable and dynamic pricing and tariffs
- Distributed demand side response of high demand
- Co-ordinated control of storage, generation, and demand.

These interventions will mean that data and digital will be critical to the delivery of a Net Zero energy system as outline in the recent Energy Digitalisation Taskforce (Sandys, et al., 2022). This includes smart meter data, smarter control and coordination, improved interoperability capabilities, and more decentralised control. The future energy system will require management and control of millions of devices and assets, but despite the increased visibility of the network through advanced monitoring and smart meters, very little discernible difference has been made to the customers and little has been done to mollify one of the biggest single price hikes in recent memory.

To truly enable a smart and fair energy system, the sector desperately needs to utilise advanced and innovative data science. Despite substantial development in these areas already, millions of pounds of innovation funding, and considerable research and trials, very little has transferred, or been productionised, within the energy sector. Despite the extant knowledge base, not enough risk is being taken to demonstrate these new products and services.

The recent price increase has pushed millions more households into fuel poverty and it is urgent that the sector acts within the next few years to help decarbonise the energy system, but also support the most vulnerable in society. It is vital that algorithms, knowledge, and skills already developed are utilised to implement life saving products and services, and that requires engaging, and collaborating with the learnings and outputs from academia who have spent years developing the tools which can enable a whole host of energy and cost saving solutions.

This document, and the supplementary reports, are intended as guides into the opportunities and challenges for industry and academics to get the most out of their relationships and research, from better utilisation of the published work, better sharing of code, to working together and building capacity.

2. SUMMARY AND RECOMMENDATIONS

Researchers and academics within universities are a valuable resource for innovative and cutting-edge research, but the sharing of knowledge and learning is often difficult. The area has been studied in detail more generally, and the Catapults themselves are the result of trying to better support cross-sector work. Despite optimising the sharing of ideas, tools, and skills the landscape is still fraught with challenges and limitations.

This report serves to contribute further learnings in this area by exploring academic-industry impacts via the specific area of data science within energy systems. This more concentrated approach hopefully provides some tailored insights into some of the challenges, blockers, and opportunities in these areas.

Through a series of in-depth interviews with experts in energy data science from academia and industry, we have compiled information on six important topics around bridging the academic-industry gap. These include:

- **Technical Knowledge:** How to better share with industry the technical knowledge being generated by academics.
- **Training and the Skills Gap:** How universities may be able to help narrow the worrying skills gap in data science in the energy industry.
- **Collaboration:** What are the ways to get the most out of collaboration?
- **Accessible and Reproducible Research**: How can academia help to make sure that research can make the biggest real-world impact?
- **Code Development for Academics:** What are some of the vital coding skills that academics should be developing to better prepare them for working in industry?
- **Industrial Support**: How can industry help support the research which is most important for their sector?

The interviews are the driving force behind the content in this report, but we also include some higher-level related literature which help to better understand and support the context and narrative. Further to this our colleagues in Arenko have performed a short paper review to better understand the difficulties and barriers in making research both accessible and reproducible.

Each of the sections below summaries some of the findings, lessons, and recommendations we found. We then go on to describe "Technical Knowledge" and "Training and the Skills Gap" in more detail in the following sections of the report. The other four components are supporting mechanisms and are described in separate supplementary reports which are shared with this main report.

2.1. TECHNICAL KNOWLEDGE: SUMMARY & RECOMMENDATIONS

One of the main topics of this paper is concerning the value created from academia and how it can be more easily discovered and better utilised by industrial organisations. Academia excels at developing advanced techniques and novel ways to solve common problems. Thus, the knowledge generated by universities has the potential to revolutionise some of the practices in industry, create new trend setters, and allow companies to keep momentum going by generating new tools and products.

Despite this potential, there are many challenges in making academic outputs visible and useful to industry. Many industrial organisations have very little time to spend on research, and knowledge from papers and talks from academia are often not accessible or easy to reproduce. It is also hard to assess what the state-of-the-art research is from the industry perspective since there is no common benchmarking model and different datasets are used in each study.

Further to this, academia excels when it is focused on novel and new problems. This can be hampered however, by the peer-review process used for deciding funding, which can be disadvantageous to risky research. However, even if the right research is funded, it is very difficult to assess what the impacts are as they are hidden or are only expected to be applied in 5-10 years' time.

Through our interviews, the following were some recommendations to help improve the sharing and implementation of academic outputs:

- **Better tracking of industrial impacts:** An investigation is required into how to best track academic impacts within industry. Currently, the situation is based on self-reporting from academics and partners within industry and may only require a simple letter of support. Better tracking of impact will lead to better understanding of what funding models are effective.
- **Improved accessibility of academic outputs:** There should be a continued drive to improve accessibility to research via open access publishing, and further, academics in applied areas should consider creating additional accessible abstracts which can describe the work to a wider audience.
- **Open data and science:** Industry focused research should emphasize reproducibility, clarity, sharing code and utilising open data. This requires incentive changes within academia where code sharing is not rewarded or recognised, although some journals are starting to require a more open research-based approach.
- **Improved discoverability of research:** Academia should hold industry focused events to reach the wider users of their tools and research. They should try to share in newsletters and industrial focused magazines etc., and if possible, use public forums for dissemination.
- **Utilise focused data science challenges:** Data science competitions have proven to be popular for engaging a wide range of participants, can improve benchmarking of important problems, and encourage data sharing and further innovation. They can also be used to support recruitment.

The industry impact of technical knowledge developed in academia is explored in detail in Section 4 in this document.

2.2. TRAINING AND SKILLS GAP: SUMMARY & RECOMMENDATIONS

Increased digitalisation across society means that digital skills are in high demand but in short supply, and this is particularly true in data science. The situation is expected to be similar within the energy sector, and our qualitative research supports this view. The vital training required to support the UK achieving Net Zero is therefore in direct competition with many other, often quite attractive, big data industries.

Academia is one obvious source of highly skilled individuals with technical and transferable knowledge, and experience who could help fill this gap, and pathways from academic research into industry was a major topic raised in our interviews.

Collaborations are a common way for industry and academia to share knowledge and resources, but they are also a desirable way for industry to identify proven individuals who have the skills they need to build capacity within their organisation. These close collaborations also provide a mechanism for universities to better understand the skills they may need to be providing training for.

This report finds the following recommendations in order to better utilise academia in supporting the emerging skills gap in digital and data:

- **Investigation of the skills gap:** There is a need to understand the nature of the data science skills gap within the energy sector. This can help identify which specific skills are needed, which training may be needed, and where are the best opportunities. Energy Systems Catapult are one such organisation but there will be support required from across the sector to understand the complete landscape. One particularly important area identified in our research is coding skills which will be discussed in more detail in one of our supplementary reports (Haben, &, Hinton, 2022c).
- **Coordination across the sector:** Better co-ordination between industry and academia is required to develop a curriculum of necessary skills and knowledge which are needed in industry and where they could fit with the teaching schedule of universities.
- **Data to support training:** Training individuals with the necessary skills means that data needs to be available to enable practitioners to develop the real testable skills required to implement them within an operational setting.
- **Industry driven training:** Industry should promote practical experience to help upskilling academics such as secondments, internships, training programs. Teaching a limited module within university courses is unlikely to completely bridge the data skills gap.

This subject is explored in detail in Section 5 in this document.

2.3. COLLABORATIONS: SUMMARY & RECOMMENDATIONS

There are many ways universities and industry can work today and the best approach will depend on the desired outcome. Not all organisations work regularly with academia, and even those that do may not know about the various collaboration mechanisms. Each of these collaboration options may have different advantages and disadvantages, depending on the objectives of the partnership. However, when a less appropriate approach is taken there can be dissatisfaction from both sides especially since there is often a misalignment between academia and industry with respect to expectations, operational practices, and final output.

In many cases a collaboration may be pursued with the wrong partner or when the timing is not right. In other cases, one of the partners may lose interest or their involvement shrinks over the course of a longer project. This is often the case from industrial side where the objectives have changed due to different business directions, or the project has converged towards more theoretical objectives.

Building a relationship in the first place, can come through many routes but this requires initiatives on both sides of academia and industry. Academia needs to ensure visibility and accessibility of their research, and sharing their code is another way to demonstrate the skills within a team and increase visibility. The nature of industry-academic relationships are often finite length projects. This can be good when pursuing a specific goal or looking to recruit but does not optimise access to the wealth of experience, knowledge, and talent within the universities.

Finding a collaboration in the first place can also be difficult especially when academics do not make their work very visible or accessible. There are many examples of long-term relationships which demonstrate how prosperous good working relationships can be. Alternatively, there are concerns that funding and projects are consistently being siloed into the same relationships again and again. This could lead to narrow focus and limited innovation as a host of ideas and out-the-box thinking is being missed from other resources.

Below are some recommendations which came out of our interviews:

- **Mutual understanding of culture and objectives:** To foster more fruitful collaborations, academia and industry must spend more time better understanding the culture, objectives, and practices of the other. This can help improve funding models, develop more business-ready solutions, and set more realistic expectations. This series of reports, especially the supplementary report on collaboration, aims to improve some of this understanding and the different collaboration mechanisms.
- **Consider longer term projects:** Industrial partners willing to fund academic research should propose/prioritise longer term projects and commitments, where possible. This is especially true when considering collaboration with postdoctoral researchers who have greater experience than students. However, since postdocs often have many other commitments, it is worth considering funding which gives time to pursue both the work and the other obligations.
- **Consider strategic partnerships:** In addition to individual projects, universities and companies need to nurture more long-term strategic partnerships which allow long term consistency but also allow flexibility to respond to new challenges and business opportunities. This clearly supports the industrial partner but also means the academics get to respond to the latest and upcoming challenges, providing opportunities for novel and new research.
- **Diversify academic models:** Some of the universities provide research opportunities with less focus on publications. Hence, in addition to more traditional academic roles, they provide an option to other more applied research and projects who can focus on knowledge transfer and more industrial impact. This may also help support an additional an external revenue income stream for universities.
- **Reduce suboptimal collaborations:** In some cases, collaborations are unnecessary but may be pursued for convenience, revenue and supporting other targets. However, this may hinder other more useful workstreams. Both industry and academic partners should ensure the collaborations are worthwhile before jumping in. It would be better to go separate ways then force the relationship for small amounts of funding.

This topic is explored in more detail in the supplementary report on industrial and academic collaborations (Haben, & Hinton, 2022a).

2.4. ACCESSIBLE AND REPRODUCIBLE RESEARCH: SUMMARY & RECOMMENDATIONS

Companies use academic outputs for a variety of purposes, this could be to help solve a problem which they are currently working on, to generate new models for their applications, or to create entirely new products for their business. In addition to the inaccessibility of academic research due to paywalls and lack of visibility, academic outputs may not be accessible due to the way they are written. This includes the complexity of their contents (there can be sometimes a certain pride in making language and methodologies more complicated than necessary) or they may be extremely lacking in the details such as how they have trained, selected, or applied their models or techniques. This means to understand the contents, it may require employees who are particularly knowledgeable in the topic, or some extrapolation and assumptions of the material.

As part of this topic in this report we also conducted a short literature review of papers in the area of price forecasting, from this and the interviews we wrote some guidance on how papers and outputs should be written and what they should contain. Some of the best practices include:

- **Data availability:** Papers should be released with the data or should use open data which should be linked to the manuscript. It is vital to help reproduce and test the results, and helps others to develop the methods further. Improvements in making data more open is a major way this can be supported.
- **Code sharing:** Can be a game changer in revealing the underlying method and providing key insights which are not apparent in the paper. This can be particularly useful if the paper is largely theoretical, and the implementation details are not included. Although it is rare for an organisation to use a code directly within their business, it is usually a good starting point for them to develop their own code or insights.
- **Accessible, assessable, and reproducible papers:** As well as being open access, papers should be written so that all details of the methodology are clear. A simple rule of thumb is that a method should be reproducible from the contents alone. This may not always be possible due to any methods which include some degree of randomness, but the results should be robust enough that this does not affect the overall outcomes. This ties into the prior point, where papers which include simple instructions on how to reproduce their results using a provided repository, represent a vast increase in utility for industry bodies.

This topic is explored in more detail in the supplementary report on accessible and reproducible research (Haben, & Hinton, 2022b).

2.5. CODE DEVELOPMENT FOR ACADEMICS: SUMMARY & RECOMMENDATIONS

Although university graduates and researchers are excellent at many technical skills, a large skill gap exists around programming, software development, and coding practises. Additionally, the tooling, frameworks, and technology which surround this area have been highlighted as areas requiring improvement for those wanting to contribute or transition into industry roles. Some training and practice would be extremely beneficial and make students and academics much more desirable to future employees. However, often these skills are not strongly developed within universities and hence it may not be possible to teach purely without support from industry. Some of the key recommendations from our research include:

- **Collaborative learning:** Close collaborations are beneficial for key coding skills. They can be a way for both sides to keep up to date with the latest knowledge, skills, and needs. In particular, industry partners will be able to inform academia about the latest technologies and software practices they are using.
- **Industry driven coding training:** Industry should provide or support key training in the area of coding practice and technologies. Since software development and technology stacks rapidly change, outsourcing this teaching may be a much more efficient way for universities to keep up to date with the latest training. At the very least universities should utilise industry feedback to help develop such courses. Guest lectures by industry professionals are another option.
- **Universities should increase focus on software skills:** Many students moving into research-focused careers within the energy domain graduate with minimal coding experience from their coursework. While most institutions offer coursework on

programming, software design, and similar, these courses are often left to be electives, are not encouraged sufficiently, or are out of date with modern practises.

- **Code publishing refocus within academia:** An increased emphases from academics on producing higher-quality code products as first-class outputs from research would provide students with portfolio pieces, increase their coding skills, drive better engagement with industry partners, and provide metrics that could be used to demonstrate research impact. This may require a culture change in academia where publishing of code is properly acknowledged as a key academic output and increased in value relative to the traditional manuscript publishing.

This topic is explored in more detail in the supplementary report on code development for academics (Haben, & Hinton, 2022c).

2.6. INDUSTRIAL SUPPORT FOR ACADEMICS: SUMMARY & RECOMMENDATIONS

If industry wants to take full advantage of the knowledge and tools developed by academia, then it needs to provide resources and problems that are the entire foundation for the research universities conduct. Industry are face-to-face with the most important problems as the energy sector digitalises. The toughest of these problems need to be shared together with their domain expertise with academia if they are to be able to support and develop innovative solutions.

For researchers to extend or improve on the current best practices requires having some indication of what those are. A lack of transparency of what is being used in industry can hinder academics in developing improvements. Of course, in many cases there may be worries about IP being shared but perhaps there is some compromises where at least roadblocks could be shared, or NDAs can be used to prevent commercially sensitive information from being exposed. It would be helpful for organisations to inform the authors of particular tools or methods, when their work is being utilised within the company.

Finally, aside from technical challenges, data is probably one of the most valuable commodities which industry can provide. Without real-world data it is impossible for academics to properly validate and test their models and techniques.

The following are the recommendations, on industry support, identified as part of this report:

- **Shared problems from industry:** Industry needs to be open with what some of their greatest current and future challenges are so that researchers can begin tackling and making progress with them. Industry domain knowledge is vital for identifying these problems and to properly define all the necessary parameters and constraints, ensuring that practical and realistic solutions are produced.

- **Shared and open data:** Industry should consider what data they can make available for general use. If possible, they should share other important dependent data such as location, pricing, asset information which may help academics create more realistic models.

- **Open real-time data and access:** Where applicable, industrial organisations should provide access to useful historical and live data via modern APIs which conform to open data standards and are discoverable on data API marketplaces and collections. Improving the accessibility and integration of data within various programmes and algorithms will lead to more innovative and up-to-date research, and help demonstrate the value in your data.

- **Improve feedback and transparency:** Where possible industrial organisations should provide academic contracts or open access to advanced data products like forecasts. This will enable academics to assess their models against the state-of-the-art and suggest advanced improvements or extensions. When quantifiable, academic outputs that provide industry value should be reported to the research team. This demonstrates the practicality and performance of their models and can motivate researchers to focus on problem solving which makes real world impact.
- **Innovation and value demonstration through competitions:** Industry should sponsor data science challenges through existing platforms to drive academic engagement on a specific topic, and to help develop competitive benchmarks.

This topic is explored in more detail in the supplementary report on the some of the key ways industry can support academic research (Haben, & Hinton, 2022d).

2.7. OUTLINE OF THE REPORTS

The summaries and recommendations above are investigated in more details in the following text and the supplementary reports. This contains much more detail about our approach, the methodology for collecting the information, and further literature which we found useful.

This report focuses on the impact academia can make in terms of knowledge sharing and skills training. We begin in Section 3 by motivating the need for improved collaborations and knowledge sharing by highlighting the benefits academic research has for both society and the wider economy. We then discuss how we collate the information and list the experts who helped us achieve a deeper understanding of the current academic-industry knowledge sharing landscape for data science in energy systems. The main bulk of the report is then split into two main sections focusing on knowledge transfer and skills in Sections 4 and 5 respectively. Finally, we conclude in Section 6 with illustrative examples of how the learnings in these reports can be applied to achieve greater impact from the research and outputs from academia.

The impacts discussed in this report can be helped or hindered by four mechanisms: collaboration; making research accessible and reproducible; proper coding skills development; and industrial support. Each of these topics is discussed in further detail in its corresponding supplementary report and highlights our interviewees experiences and knowledge in these areas. They contain an overview of each area, some of the main challenges, and some approaches and/or tools which can help support them. Each report then finishes with the recommendations which will help achieve the optimal impact within industry.

3. INTRODUCTION

The energy sector, especially regarding the areas of digitalisation and data, is progressing rapidly. This will require increased upskilling of individuals, new algorithms and tools to optimise the use of data, and research into the emerging challenges (Sevilla, et al., 2022).

To achieve a modern digitalised energy sector, it will be valuable to have close working relationships and collaborations between academia, research organisations and energy companies. Businesses often interact with academia to benefit from the innovative research and new insights being produced within universities to help get ahead of their competitors, but also to make new connections and strengthen their reputations (Hughes & Kitson, 2013). Universities can also provide a resource of graduates who are highly trained in technical subjects, such as computing and data science, and can help businesses bridge the skills gap (DuBois, 2020), (Miller & Hughes, 2017), (Ramachandran & Watson, 2021). A recent detailed policy paper by the UK Government's Department for Digital, Culture & Sport has emphasized the demand and gaps for data skills within the UK (UK Government, 2021).

There are also many benefits for academics to work with industry. Industry can be a valuable source of real data from which academics can develop and test models, and they can also provide interesting issues that can be the foundation of quality research problems. Further, connection with industry can also ensure that universities are training their students to join the workforce with knowledge and understanding of the best practices and relevant technologies. As mentioned in (Frontier Economics, 2014) "survey evidence has found that a significant amount of private sector output and innovation is thought to depend critically on public funding of academic research" and "There is a considerable literature which suggests that links between academic researchers and business can be useful in the innovation process".

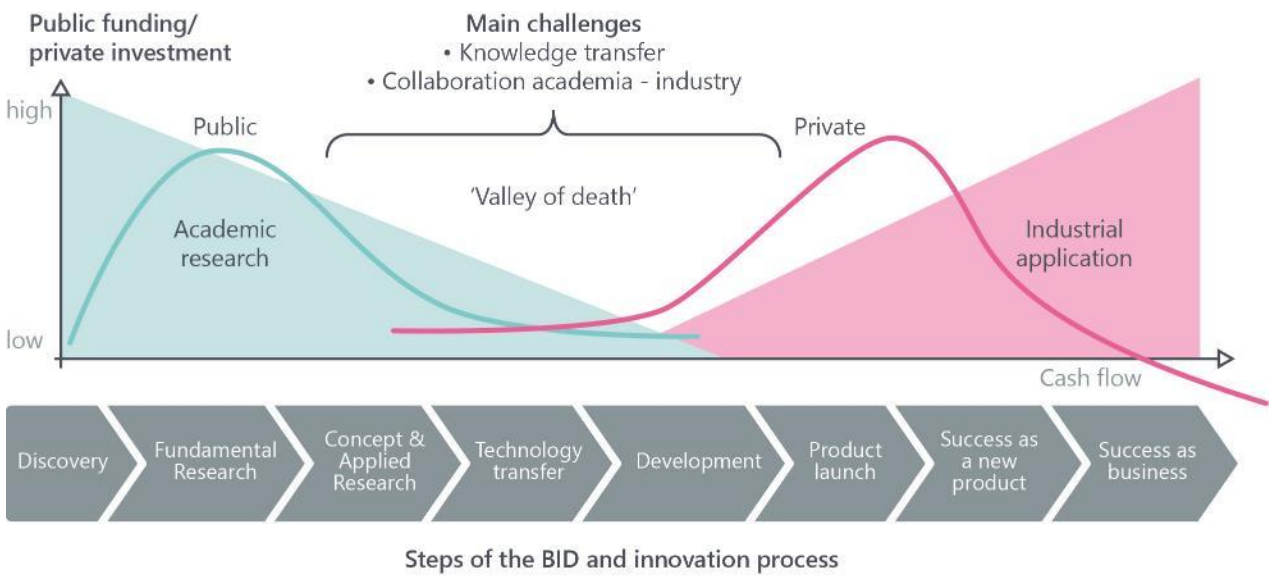

Figure 1. Valley of Death, reprinted and adapted from (Chirazi, Wanieck, Fayemi, Zollfrank, & Jacobs, 2019), under CC BY 4.0 license.

Despite the potential opportunities that academic research provides, relatively little is utilised within the industry, whether it is shared code, or the learnings published in research journal articles. Often it is not easy to find or access these learnings, but even if they are available, many industrial organisations don't have time to thoroughly investigate or read the papers, never mind, translate them to their own business. Collaboration barriers may also make it more difficult for academics to push for industry change and modernisation (ways of working, data access etc.).

Problems of translating research into impact are not new, the "Valley of Death (see Figure 1), i.e. the phase between research and utilising that innovation in business as usual, has been written about extensively (e.g. (Gulbrandsen, 2009) (Chirazi, Wanieck, Fayemi, Zollfrank, & Jacobs, 2019), (Rossini, 2018)). It may be expected that the gap between academia and industry will be less pronounced within data science and artificial intelligence areas since many of the methods and techniques are very applied and centre around data collected in the real world. However, concerning specifically AI topics, a recent article by *The Data Science & AI Section of the Royal Statistical Society* analysed the responses of 284 from their community of practicing data scientists and AI specialists on their thoughts of the UK AI councils recent AI roadmap (UK AI Council, 2021) and stated that "knowledge transfer between academia and companies is not currently working" (The Data Science & AI Section of the Royal Statistical Society, 2021).

Within the power sector a workshop on "Translating Power Systems Research into Practice ("From the lab into the field")" chaired by Prof. Keith Bell, and Prof. Daniel Molzahn as part of the 21st Power Systems Computation Conference (PSCC 2020)[1] touched upon some of the barriers to bridging the valley of death, many discussions touch upon some of the topics in this report, namely:

- The misalignment between industry and academia, where the timelines and objectives can be very different. In academia the focus is on papers and striving to be academically thorough whereas in industry easier and faster solutions may be more desirable
- The skills gap. In academia the focus can be to use more technical language and methods, but there is a need to try and improve communications skills (presenting and writing). There is also a need to consider the importance of domain knowledge, understanding that there is still lots to learn from those practitioners and engineers who've been in the industry for several years.

To the authors knowledge, and those we interviewed, there is very little investigation into considering the challenges of academia-industry collaborations and knowledge sharing within the energy sector for data science driven research. The aim of this report, is to investigate the current landscape, opportunities, barriers, and recommendations to improving the impact made by academia within industry by collating desktop research and conducting interviews with experts in the field from industry practitioners, applied researchers, funding bodies, and other prominent figures in this area. An additional unique aspect of this research is the investigation into tools and resources which can help facilitate the learning, assimilation and dissemination of academic research and impact to businesses.

3.1. INFORMATION COLLECTION METHODOLOGY

The research in this report has been conducted primarily through semi-structured one-to-one interviews with experts of data science in energy systems. We were interested in academics who worked a lot with industry, and those in industry who had some experience working with academics. The amount of interaction for the industrial partner was also of interest, since small amounts of collaboration may indicate there were potential issues with their approach and we were interested in what they may be. From the university side, we mainly interviewed permanent academic staff, but also some postgraduate students who were heavily involved in industry work. We also considered other areas which had significant relevance to this topic, including Innovate UK,

[1] A full recording of the event can be found here: https://www.youtube.com/watch?v=7OrzJ8X7RXg

a major funder for collaborative research, and Electric Data Solutions, an SME focused on providing analysis of impact. The interview research gathered here is naturally UK centric however, we have also included some international interviews from individuals with particularly successful academic-industrial collaborations.

The focus of the questions was around a number of topics including:

- Experience with industry as an academic, or vice versa.
- How some collaborations were formed.
- Particularly successful collaborations or examples of major impact.
- Challenges or blockers with partner interactions.
- How some challenges were solved.
- Any tools that they found useful in improving collaborations.

A major component, and potential blocker, of collaborations and impact is accessible and reproducible research (see (Haben, & Hinton, 2022b)). Hence as part of this research, the team at Arenko performed a review of papers within a specific subset of data science in energy (in this case electricity price forecasting) to investigate some of the best (and worst) practices in published papers which make it difficult to understand, disseminate, or replicate the methodologies demonstrated within them. Arenko, being practitioners of advanced algorithms, are advantageously positioned to assess these papers within an industrial context.

In addition to the interviews and paper reviews we also performed some brief desktop research to discover tools to facilitate knowledge sharing, write production and/or reusable code, write reproducible research, or discover other relevant research articles. Based on this, we found very little that discussed the specific challenges of translating a relatively applied discipline of data science to industrial, and practically focused, energy applications. It should be noted that this report is primarily focused on the outputs of the interviews and the paper reviews as a basis of understanding and providing evidence for the recommendations given and should act as a companion to other sources of information. It is not intended as an extensive review of the current literature.

We hope that this report will provide some assistance at addressing these challenges different levels of the process from individual researchers to innovators, to decision makers and funding bodies.

We would like to thank the following individuals for their time and help in putting these reports together:

Clem Attwood, John Barton, Keith Bell, Ayrton Bourn, Jethro Browell, Julien Caudron, Richard Dobson, Elisa Evans, Thomas Falconer, Fabio Giunta, Jose Ortiz de Lanzagorta Gonzalez, Miha Grabner, Danica Vukadinović Greetham, Peter Grindrod, Andy Hadland, Catherine Jones, Jack Kelly, Jesus Lago, Angeliki Loukatou, Robyn Lucas, Adriana Martins, Yush Mittal, John Moriarty, Aidan O'Sullivan, Pierre Pinson, Zoya Pourmirza, David Richardson, Colin Singleton, Martin Szomszor, Frank Tutu, Dimitrios Vardouniotis, Rafał Weron, Sam Young.

3.2. OUTLINE OF THE REPORTS

This report investigates, via detailed interviews and desktop research, the problems, and potential solutions, for academic-industry relationships in the energy sector specific for data science solutions. This report is split into two main parts:

1. Academic Outputs and Impact
2. Mechanisms to support impact

The first part on Impact is detailed in this report and describes the two main ways that academia provides impact within industry. Namely through either providing knowledge, or through providing particular, technical skills. One of the main outputs of academia is through the technical learnings created via research. This generated knowledge is shared primarily through papers, demonstrators, presentations at conferences, and sharing code, and industry can utilise this to innovate, compete and excel within the data science areas of their sector. The other major output from academia is producing technically skilled individuals. With the major skills gap in the energy industry, especially in data related fields, universities can be a valuable resource in filling this gap.

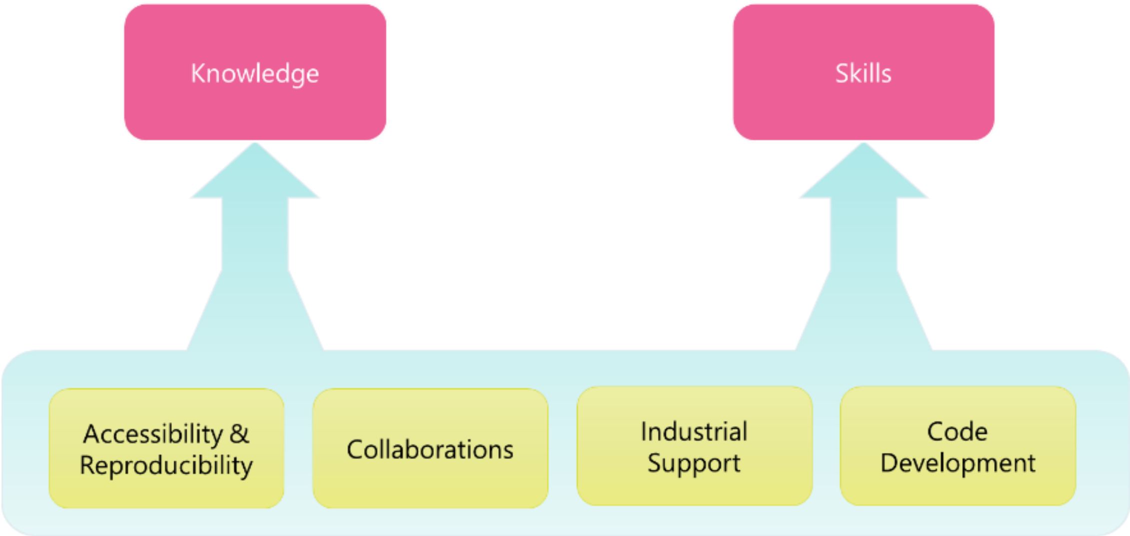

Figure 2. Illustration of the structure and the dependencies for our reports on academic driven impact within industry. The two main methods of impact are knowledge and skills (in pink) and these are supported by four supportive mechanisms (yellow). Each of the supportive mechanisms is explored in more detail in supplementary reports which accompany this main report which focuses on the impacts.

The two main components of impact are supported primarily by four supportive mechanisms which we featured prominently in our research as illustrated in Figure 2. In part two we will dive into each of these supportive mechanisms which are:

- How industry can help support academia with their research
- How academia can better support technical knowledge exchange through more accessible and reproducible research
- How to create quality collaborations between industry and academia, and
- How code developed from academic research could be better utilised.

Collaboration is by far one of the most important ways to support both the utilisation of knowledge and skills. Working together is a common way to recruit reliable personnel for the organisations we spoke to, but also provides the industry with direct experience applying some of the state-of-the-art technologies and methods. We discuss this process as well as some of the challenges and potential solutions in the first supplementary report (Haben, & Hinton, 2022a). To maximise the sharing of the technical outputs and knowledge from universities requires research to accessible and reproducible (outlined in our second supplementary report, see (Haben, & Hinton, 2022b)), but in the case of algorithms and models, this can be achieved by the proper development and sharing of codes and algorithms (See supplementary report (Haben, & Hinton, 2022c)). To create useful learnings in academia requires that academics know the major problems that industry is struggling with, and has realistic data and knowledge of the problems. These topics will be covered in the final supplementary report on Industrial Support for Academia (Haben, & Hinton, 2022d).

This main report focuses on the main impact of technical knowledge (Section 4) and skills (Section 5) and in Section 6 we finish with some examples of how the learnings and recommendations can be utilised to improve the collaborations and learning dissemination between academia and industry.

4. TECHNICAL KNOWLEDGE IMPACT

This section introduces and discusses the first major output and potential impact from academia: Technical Knowledge. Arguably, the primary objective of a university is to expand the sphere of human knowledge and understanding. In many cases the focus is not on directly applicability or utilisation (especially in more theoretical areas such as pure mathematics) but to develop theories and hypothesis which may or may not have a direct impact on society, technology, or the economy. Nevertheless, a large proportion of research, is focused on real-world applications and can have major impacts on human civilisation. This is particularly true in emerging areas such as data science where many technologies are in early-stage development since certain techniques and approaches have not been possible due to computational and resource constraints. This section focuses on the technical knowledge coming out of universities and their research asking:

- Why are research outputs important for industry?
- What are some of the challenges in sharing and assimilating this information?
- What are some of the ways this knowledge sharing, and facilitation can be improved?

4.1. TYPES OF IMPACT

Academia and industry have a lot to offer each other and can help bridge the gap in each other's resources and knowledge. For example, academia has the potential time to focus in detail on specific technical challenges and can test out lots of different tools and methods that may not be possible within the faster-paced industrial organisation. On the other hands industry has access to real-world problems, financing, valuable domain knowledge and data, all of which are required to ensure there are interesting and well-defined problems for academics to explore, generate new advanced models and push the boundaries of human knowledge.

The knowledge contained within academia can give companies a competitive edge. Indeed, many successful organisations have embraced the latest technologies and learnings, this includes those start-ups who spin out of the university environment utilising the transformation of the methods they founded/developed as academics. The majority of R&D is predictable and a necessity as part of a business. It is the unknown unknowns which possess the potential for game-changing innovation. Riskier or more novel research is a means for a company to hedge against other competitors in their field (Grindrod, 2020). Further it may be that academia can provide momentum in a particular area that the industrial organisation is lacking. A network of university collaborations can be insurance against rapid change.

To summarise, academia can provide a wealth of tools and information which can help industrial organisations in a variety of different areas, including:

- Bring new ideas with very skilled individuals. This may also create some internal tension within the company, driving competition.
- Getting known expertise: individuals in companies are rarely benchmarked since they are working on ideas which cannot be shared hence it is rare that wider comparisons are possible.
- They allow companies to keep momentum within their own organisation while exploring new ideas and avenues, alternatively it allows companies to repurpose individuals.
- New methods and models for solving old problems as well as evidence for the effectiveness of various techniques.
- Testing the viability of new (and potentially risky) opportunities and products.

- Developing readily available code demonstrating their models.
- Optimal ways of utilising data.
- Alternative and improved ways to manage risk.

There are many examples of successful incorporation of academic learnings within the energy industry. The CREST model has been used by many organisations for testing network modelling for the past 25 years. Another interviewee of ours described how their work on price forecasting was incorporated by external organisations, and Countinglab, a spin-out from the University of Reading, who were part of our interviews have also been successful in many sectors of data science including energy. In 2020, the National Grid launched their Carbon Intensity App a successful data science collaboration between industrial and academic partners. In addition, increasing numbers of projects from the Network Innovation Allowance and Network Innovation Competition utilise data science via academic collaborations to test and demonstrate advanced and sophisticated algorithms and models. Data Science is increasingly being applied in many energy applications from fault detection, to solar nowcasting, to flexibility services. This means there is a real need for strong technical data science skills for already sparse pool of individuals (see Section 5 on the Skills Gap).

Although the greatest need is in those organisations with limited technical resources, even those with very strong data science and analytical teams, may have limitations on what can be learnt and achieved within the stricter time and resource limits of the business. In the energy sector there has been a relatively slow uptake of data science skills, especially in larger organisations, whereas smaller organisations often have the skills but not always the resources to deploy the more technical and advanced models being presented in the academic research. Academia can provide an extra resource in all those cases to do the important foundational and fundamental research and testing.

4.2. CHALLENGES TO MEASURING AND IDENTIFYING IMPACT

Before discussing the difficulties in incorporating or sharing academic research into industry, it is worth understanding what would be considered impact and how to assess it. Impact is measured differently depending on whether you are in academia or in industry. In industry the measure of impact can be relatively well defined via their key performance indicators (KPIs) or targets that they must meet. The success of an organisation can therefore be measured by financial gains over the year, number of collaborations or projects won, etc. For the more technical KPIs they can be facilitated by the learnings from academia.

For academia the situation is less clear. Although Universities have minimal funding sources, the (increasing) majority of their income in the UK comes from tuition fees and to a lesser extent from research funding (EPSRC, BBSRC, etc.) and funding body grants. Further to this, each UK university is also subject to the Research Excellence Framework (REF), an assessment which evaluates (every six years) the impact of the research produced in each department and is used to inform the selective allocation of funding for research. There are many resources for tracking academic impact, including numbers of publications, the impact factor of journals, peer reviews, indexes for publications (such as h-index), and amount of funding obtained (Publons, 2022), (Clarivate, 2022), (Elsevier, 2022). Although important for universities outputs, for our purposes we are interested in industrial impact.

Industrial impact has becoming increasingly important within universities. For this reason, a major component of the REF is the "impact beyond academia". The UK is not alone in this, and many countries have included some form of impact assessment of academic outputs (Williams & Grant, 2018).

The REF has increased the focus on impact but still has many limitations not least that, as described from those we interviewed, it can distract from the main purpose of universities which is teaching and producing novel research. Further to this, the evidence required by the REF can be quite small, and very often a simple letter of recognition from a friendly industrial partner can be utilised to illustrate the impact your research has made. In other cases, it can be difficult to actually produce measurable outcomes especially when the research's intended impact is not short term, and the benefits will only be observed much further into the future (which is perhaps to be expected from novel and advanced research outputs). As a result, the impact statements may be misrepresentative of their true value and impact. It is difficult to assess where research is actually used within industry as a company revealing their approaches risks exposing intellectual property or methodology to competitors. Hence, even if some method has been revealed or known to be utilised it may have significant modifications to that used in practice.

There are many other possible approaches for tracking industrial impact put forward by our interviewees, but they acknowledge that each had their own limitations or flaws. These are as follows:

- Tracking patents and licenses.
- Number of industrial collaborations, or percentage of portfolio with industrial collaborations.
- Code reuse – although difficult to know how much is implemented.
- Subsequent funding – this doesn't measure direct impact and the connection between successful subsequent funding may be tenuous.
- Match funding contributed by industry in collaborative R&D projects. Similarly, income from consultancy.
- Industrial Co-authors: This could act as a surrogate for the strength of links and collaborations between academics and the industrial personnel. This is attractive however as it also may be a mechanism in which to improve industrial engagement in academic research and take a personal stake in the work. It could also be used as a possible indicator for industrially useful research.
- Employment: How many researchers were employed based on the work? This clearly demonstrates that the organisation was pleased with the partners they work with but, as will mentioned in Section 5, this may not tell us anything about the actual research impact. Further finding this information is not straightforward but could be supported by universities tracking and publishing alumnus information.
- Awarded tenders: Industry could also identify where the research was used in application for other tenders.

There are some tools for tracking some of this impact (UKRI, 2022), (Clarivate, 2022) and funded projects through UK research council grants now must fill in impact as part of the Research Fish online system. This tracks information on publications, subsequent funding, collaborations, IP and licensing, and awards etc. However, this is primarily dependent on annual reports from the main principal investigators of the projects, and only includes visible impacts. However as mentioned, most companies do not openly advertise the research, tools or resources that they use as this may have commercial implications. This highlights a common problem encountered in our investigation: **it is very difficult to accurately measure the impact of academic research**. Impact evaluation is often not considered early enough in a project design, nor is enough (if any) of a project's budget allocated for these purposes.

Not understanding the true impact made by the research and/or how to track research is a major drawback in incorporating academic research into industry. If it cannot be measured properly then it is impossible to tailor research to make the maximum impact. Further, if we were able to track the impact, we could also explore optimal ways to create targeted secondary impacts and benefits which could then be included in the design and implementation of collaborative projects. The next section will explore the difficulties in making impact in industry with academic research. However, bearing in mind that it is difficult to produce objective measures of impact in general, the information below is based on the anecdotes and experience of the interviewees.

4.3. CHALLENGES OF PRODUCING IMPACTFUL RESEARCH: DATA AND FUNDING AVAILABILITY

Academic research can be incorporated into industry in indirect or direct ways. Indirect impact can be via reading the published articles from research, or utilising code/tools that the academic has developed and shared through their Git repository. Alternatively, academics can work directly with industry via funded projects (including PhDs) or through consultation. This may involve developing a tool, or through fundamental research of a question that is of interest to the company. Direct research has the advantage that the focus will often be strongly related to a problem that the organisation and similar organisations are working on. Thus, the industrial representatives have an opportunity to shape and direct the research to their specific needs, anchored to real domain knowledge. The vast (and increasing) number of academic outputs provide a wealth of opportunities but it may be difficult to find relevant and good quality research. The difficulties for academia to make impact in either direct or indirect ways is explored in this section.

One of the main challenges is creating an opportunity to do relevant research in the first place. Two core roadblocks highlighted in our interviews are funding and data availability. A lot of research in universities is conducted via postdocs or PhDs with temporary contracts and supervised/directed by permanent staff such as the lecturers and professors. Hence **funding is an essential requirement for research to progress**. Permanent staff will of course also perform research, but they typically have several other responsibilities including teaching, supervision, administration and obtaining the funding itself. Funding can often come from research council grants, specific funding calls, or directly funded from an industrial partner. An advantage of industrial funding is that there can be a direct negotiation between two interested parties to achieve a mutually beneficial project, and, in addition, there is no competition with others for limited funds. For this reason, building a strong relationship with an industrial partner can have benefits for years or decades. Building a collaboration between industry and academia is discussed in much more detail in the supplementary report on this topic (Haben, & Hinton, 2022a).

There are several resources for collaborative funding opportunities. Most recently and relevant to this report is the Strategic Innovation Fund (SIF) developed between Innovate UK and Ofgem. The SIF focus on the Electricity System Operator, Electricity Transmission, Gas Transmission and Gas Distribution sectors and aims to bring a diverse set of collaborators together to work on innovation projects. Similarly, there is Ofgem's Network Innovation Allowance (NIA) or Competition's (NIC) which involve consortiums and collaborations of partners on network focused projects. These and other sources of funding are typically competitive with many universities competing on the same calls. A limitation of these funding models is that the available funds will not cover all areas of interest and therefore there may only be a few sources of funding for an academic's specific skill set. Putting together grants is a very time-consuming process, and most of these grants are unsuccessful (Herbert, 2013). Further, when a product/application is seen to be fully developed and finally commercialised funding may dry up.

Another common statement from interviewees is the overly deep focus on peer-review as the core criteria for judging the majority of funds and grants. Submissions are scored by experts in the area according to a marking criterion and any poor scores or, in many cases, a single bad review will mean the application will fail. Evidence suggests, peer-review may be an unreliable indicator of a projects quality (Jerrim & de Vries, 2020). Secondly it may not be the best indicator of success (Fang, Bowen, & Casadevall, 2016). Finally, as pointed out by our interviewees, the peer-review process, reduces the success for riskier, and hence highly unconventional but potentially innovative ideas (Bendiscioli, 2019). This was one of the reasons for the Advanced Research and Invention Agency (ARIA) programme and was identified by one interviewee as an alternative model which would allow such research to take place (BEIS, 2021). The ARIA programme is focused on riskier research and does not run-on peer review since it was found that "extra layers of approvals and review in the funding system, while well intentioned, can stifle the creativity and dynamism of scientists". Instead, they will run a programme manager model who can distribute funds across projects of their choosing. It is not clear what impact this will have on energy systems and/or data science research.

This links to a common issue discussed by academic interviewees where they felt funding models left large gaps in research for important areas. There was discussion of too much focus on fast returns on investment and wanting something "high tech, small, fast, cheap and efficient" to quote one interviewee. Another interviewee mentioned that sometimes the focus is too much on entrepreneurs and producing the "next Microsoft" rather than trying to take something from low technology readiness to high. Not everything is commodifiable in energy systems but perhaps this highlights the philosophical gap between industrial and academic approaches and aims. Academic timelines are much longer and the outputs from their research may not be useful until much further in the future. In contrast, industry often wants relatively quick gains and results. One way to optimise the gains from academic research, is for financing to be available which enables longer term projects which are focused on industrial problems. Postdoctoral researchers are often the most suited for developing sophisticated models, but they are often in insecure positions with short contracts varying from a few months to a couple of years. This instability makes it difficult for long term relationships to be maintained or for a problem to be extensively investigated. However, even with funding longer term projects these timelines are often misaligned with industry. By the time a PhD or long-term research project is finished the company may have moved on and implemented a solution which is solved to the extent that they want. These issues will be discussed more in the supplementary report on Academic-industry collaborations (Haben, & Hinton, 2022a).

Data is another major blocker for research. Even if funding is available, open data sources are relatively sparse. The few open data sources that are available become overused or are quickly out of date, especially given the fast developments within data science. A good example of this is the Irish Smart meter data ((CER), 2012) which is one of the earliest smart meter datasets of reasonable size that was made available publicly. It was ideal for the burgeoning research community who utilised this data extensively (Haben, Arora, Giasemidis, Voss, & Greetham, 2021), but the data is now nearly a decade old, only contains two years of data, and was bias towards particular household sizes with much of them subject to tariff interventions. Thus, there is biases in the data which make it difficult to generalise to other situations or countries.

Collaborations can make access to data much easier, however there are usually restrictions on how to use the data, and what, if any, is allowed to be used within an academic publication (the major impact currency of academics). Further, the data is, of course, not allowed to be shared beyond the specific research group. This means that it is impossible to validate or recreate the results from any research published. Industrial interviewees have also mentioned that it can often not be worth their trouble to extract and pre-process the data, especially with teams which are particularly busy.

Finally, regarding data, even if it is openly available, the accompanying documentation is often poor, and has varied complexity. As an example, consider a common difficulty for energy systems research: weather data access and utilisation. Weather observations and forecasts are vital for so many applications in energy systems, but the data is often difficult to get hold of, may be difficult to understand for non-experts, and can come in formats (such as NetCDF files) which may be unfamiliar to many even in data science domains. In this area, academia has an advantage, since several numerical weather centres provide free access to a whole host of weather data. This illustrates another vital way academia can support industry via the access to additional services and resources. In this example, they can de-risk an organisation from purchasing expensive weather products until after they have been tested.

The data issues are not isolated to academia. There are many issues across the sector as highlighted by the Energy Data Taskforce (EDTF) (Sandys, et al., 2019). As the recommendations from this piece of work are assimilated across the industry, we hope to see more and more data being available which will help innovators both in academia and industry.

If data and funding are not an issue, then the core objective of the research is to **choose and define the problem** so that both academics and industrial partners can benefit. This is not trivial since both academia and industry often have different purposes and objectives and can both fail to understand each other's viewpoints. Universities can all too often ignore the industrial application and problems and can view partners as simply a source of financing. Similarly, industry can often ask for solutions to problems which perhaps do not optimal utilise the academic partner. However, there are clear benefits (as mentioned above) to the partnership when the problem and resources are aligned.

A common problem from the academic perspective is knowing what are the interesting and pertinent questions that industry is working on. Collaborative projects are a great way that a problem can be defined and refined between the partners. However, if there is no collaboration then how can an academic understand which problems to focus on? Interviewees mentioned a few routes they take to develop new ideas including going to industrial conference and reviewing papers. Also mentioned were sandpits, where academics and "practitioners" can come together to generate cross-disciplinary research projects. However, despite showing promise it has been suggested that researcher funding does not often align to enable the projects which are developed.

4.4. CHALLENGES WITH UTILISING IMPACTS

If the research is properly directed and aligned with industry there is a higher chance that the outputs will be realised within the organisation. Often the research isn't realised within the company due to a host of issues including the following highlighted in our interviews:

- The research is a future application for the business (which is unsurprising as research is often about new products/models/methods).

- The research was not focused on the practical application and went in a direction not suitable to the organisation.
- The research simply failed. As with any risky endeavour there is no guarantee the outcomes are as expected. In fact, this is a major reason for doing the research in the first place.
- The organisation doesn't have the necessary skills to integrate the modelling. The fundamental research may not be produced in a way that can be easily incorporated into the operational system of the company and since this is often advanced research it may not be easily translated by a current employee.

Beyond these reasons, there is also a limitation with simply collaborations centred around an individual project(s). They do not necessarily foster deep and fulfilling partnerships and can be simply seen as "just a project" by both the university and the company. More value and impact may come from developing long-term strategic partnerships rather than focus on short-term transactional projects. However, this may require changes of culture and attitudes within both industry and academia. This will be discussed more in the supplementary report on collaboration (Haben, & Hinton, 2022a).

Industrial corporations can also utilize the research outputs indirectly through utilising e.g. papers and codes which are produced by academics. Papers are often used by some organisations as a way to solve particular operational problems, e.g., solve a particular optimisation. In these cases, there isn't necessarily a need to find some advanced, novel or even interesting result. There are a variety of ways that organisations can find and keep up with the latest research, however not many organisations utilise these resources. Popular methods include active email groups (such as OpenMod Initiative and Powerswarm), through data science newsletters, webinars, and attending conferences.

Keeping up with research is much more complicated than finding it for an industrial organisation. Although many of the organisations we spoke to have reading clubs and other knowledge sharing events many could not find the time to read papers or had other difficulties with assimilating research outputs. The most common issues were with accessing the papers in the first place, how to identify which papers to investigate, and how to implement what is in them.

Accessing research papers is a major problem, even within academia. Much of the research is behind paywalls and cannot be accessed without a subscription. Authors may not even be able to access some of their own papers as the copyright is handed over to the publishers and may lie behind a subscription. In many cases, institutions pay twice whereby they pay to publish their article and then again to access it. Access to papers within the organisations we talked with was a mixed bag, with some having some access and others no subscriptions. The increased use of open access journals has helped to increase the access of academic research and many journals also allow a preprint (a version of the paper published usually without the final changes as asked for by the reviewers) to be made available (e.g., through arXiv or ResearchGate). Alternatively, many individuals contact the authors directly who are often permitted to share preprints. Open access has its own difficulties, in particular the prices are very high to publish in such journals, open access has also led to an increase in so-called predatory open-access journals who will take advantage of the pressures within academia to build an extensive list of publications.

If papers can be accessed, the increasing amount of research outputs can make it difficult to select and assess what papers are worth utilising the limited time available to industrial personnel. Academic papers typically have a short abstract which outlines the main conclusions of the paper. However, these are often focused on theoretical academic results and may not be decipherable, or may not immediately explain the relevant information, to an industrial energy data scientist. A more accessible industrially focused abstract was touted as a welcome addition to applied papers which would help filter what would be useful for a practitioner.

In some cases, papers are chosen by industry personnel based purely on the fact that they are relatively famous or illustrate or represent one of the earliest versions of a new algorithm. In other cases, the papers were chosen as they focused on a specific problem that the person was working on, or they hoped had a solution to a particular difficulty they were having (one example given being an optimisation algorithm which was not converging). However, in terms of more exploratory research many interviewees found it difficult to identify a suitable and appropriate paper for a problem they were having. Many other different criteria were given to us for how individuals selected papers:

- Reputation of the authors, either from previous work, citation records, or status (e.g. professor, or senior lecturer).
- A paper being cited by other good quality papers or authors with good reputations.
- Quality of the writing. Red flags such as non-conventional notation or use of non-standard benchmarks can indicate the writer(s) are novices in the area.
- Number of citations. However, in cases of large numbers of unnecessary self-citations means that the reliability of the paper should be questioned.

The final major issue was the quality of the paper itself, with methods being unclear or the results not easy to replicate. This and the other issue of accessibility will be covered in more detail in supplementary report on accessible and reproducible research (Haben, & Hinton 2022b).

4.5. WAYS TO IMPROVE KNOWLEDGE TRANSFER IMPACT

Our interviews uncovered many potential ways to improve both the amount and quality of knowledge sharing impact from the research performed in academia. The following subsections highlight the main topics in this area.

4.5.1. IMPROVED SHARING OF CODE AND TOOLS

A major resource that academics can share are software tools. For example, a highly utilised code is the BMRS wrapper developed by Ayrton Bourn which can extract data from Elexon’s Balancing mechanism reporting service.

Academics are ideally positioned to produce code based on their own work, especially if it is highly technical, and it can provide deeper insights into how the algorithms may work which may not be clear from the accompanying methods paper. It is also a way to make major impact within industry as it provides an example of a complete framework for the model. Another example of a popular academic tool used in industry is the CREST thermal electric demand model, which was especially invaluable given the sparse amount of data in this area.

There are several issues with sharing code. Academia is usually not set up to write or maintain code and there is no incentive for an academic to do so. Code is viewed as the means to an end (a scientific publication), instead of the end itself, and is thus not treated as a first-class output. Or, to put it another way, code publishing is not viewed as particularly valuable within the university system where journal papers have the most currency.

As will be discussed in the supplementary report on code development for academics (Haben, & Hinton, 2022c), proper coding practices used in industry are not typically implemented in academia nor is there any peer review. For these reasons there is hesitance in directly utilising academic code in industry as it hasn't gone through the rigorous of testing and review that their own code has. For these reasons, many industry interviewees we spoke to typically rewrote the implementation, where using the provided code as reference material is often clearer than using a paper as reference. This may minimise the uptake of such code although may slightly reduce the coding effort required from the industry side. Improving the inclusion of citation indexes for code may be one way to encourage further uptake and highlight the impact that the work is producing.

Some of these practices are changing, albeit slowly, within academia and now code submissions are more common than they used to be. Code repositories are being populated by savvy students as a way to advertise their experience and skills to future employees.

4.5.2. FACILITATION THROUGH IMPROVED PRACTICES

Academics making more impact within industry is clearly a desirable move for both sectors but as discussed there are many issues with this model. To solve the issue will require both sides making changes but perhaps also different funding models and new mechanisms.

As will be discussed in the supplementary report on collaborations (Haben, & Hinton, 2022a), to ensure the best possible outcomes a better understanding of the aims, needs and limitations of each other's organisation is vital. Academia should not identify industry simply as a source of funding and data, and, if wishing to work together, should earnestly ensure that they are focused on the problem of interest to the organisation, i.e. solve the problem in a way the business wants. Industry should understand that an important output from academia will be publications and an impact narrative, and should therefore try to be supportive and ensure that they are flexible with their knowledge and data (whilst not infringing any security or IP concerns).

To achieve this, good scoping is needed, since academics don't always understand which aspects can hinder productionising their outputs, they should take into account some of the practicalities of their methods and approaches before diving in. A well scoped project should merge the best of the industrial organisations' domain knowledge and the technical know-how of the researchers.

If the academic researchers are required to produce a prototype code, then they should design this to fit in with the current operations as close as possible or to enable it to be easily integrated and/or understood by personnel in the organisation. This could mean utilising Excel or some other enterprise software which the company uses, or feel more trust in using, since it may be difficult to judge the reliability of open-source software or tools they are unfamiliar with. Further to this, it is important to include the right people for the project. Production level code should be delivered by those such as software developers with experience in these areas. This may require universities to consider including less research focused staff to include in their projects. It may require working closely with start-ups and/or other external capabilities. Of course, training students and researchers to produce more suitable code is also desirable (Code Development for Academics report, (Haben, & Hinton, 2022c)).

It should be noted that any research project naturally has risk, the research may end up leading to dead ends, or disprove the usefulness of a particular technique in a particular application. Any such risks should be identified early in the project. Such research projects should certainly not be part of something which is intended to be made operational without proof-of-concept preliminary research being performed first.

4.5.3. MAKING RESEARCH DISCOVERABLE AND ACCESSIBLE

When sharing industry relevant research there are many ways that academics can make their work more accessible to companies. The easiest way to do this is by publishing in open access journals, although these can be expensive and may not be possible unless funding is available. In these cases, academics should at least try to publish an open access preprint which most journals allow in some form or another and which is very close to the final published version.

There are two further relatively easy approaches which can further increase the impact of industrial focused research papers. First, it is important and responsible for academics to advertise the work in channels popular with industrial organisations, for example, the mailing list such as Powerswarm which have many industrial members, LinkedIn groups, and they should present their work at major conferences and events attended by industrial organisations (these typically have discount rates for students and universities, or may even be free if they are giving a presentation). This is also a perfect opportunity to build your networks, and collaborations with companies who are deeply aligned with your work.

The second way is to ensure there are accessible summaries to the researcher. Discussions with industrial interviewees consistently stated this as a desirable way to help them sift through the ever increase amounts of research materials. A more industry-friendly versions of an abstract would be a simple progression in the right direction. This would support the many journal reading clubs that many of the interviewees host. Other ideas also discussed include:

- clear tagging of the industry application the research is relevant to,
- summaries of the general (practical) learnings, and
- trying to utilise more accessible language rather than obscure technical jargon.

There were also suggestions for more application-focused journals which do not have to abide by the strict and often overly technical solutions which often hinder the publication of more practically focused and empirical research. Other disciplines have already implemented open-access journals focused on practitioners, e.g. (Cadotte, Jones, & Newton, 2019). Finally, as discussed above, code sharing with a paper is a common way to increase interest in a paper and allows testing and reproducibility of the methods. Although, this almost certainly won't match the standards required to implement operationally it will offer deeper insights into the methods and some of the more nuanced issues which cannot be fit within the body of the main manuscript.

Authors may also consider releasing data with their paper as this has shown to increase the citations compared to papers without data (Colavizza, Hrynaszkiewicz, Staden, Whitaker, & McGillivray, 2020). Many journals, (e.g. Royal society) require authors to publish the data they use and this can only help improve the reproducibility of the research being produced.

Academic papers are often written without full details on how to implement, this makes them impractical for use within industry. For academics interested in their research making real impact, they should try to ensure their research is reproducible from the published paper. Authors should ensure they include proper benchmarks, try and include open data, share their data, and address many of the issues as identified in our supplementary report on accessible and reproducible research (Haben, & Hinton, 2022b). Power system demonstrators[2] and living labs can be a powerful way to illustrate the real-world performance of an algorithm or tool and increase the confidence that the research translates from theory to practical implementation.

---

[2] For example, the Power Networks Demonstration Centre: https://pndc.co.uk/

A few other approaches to help improving utilisation of research were discussed by our interviewees, including:

- Creating a scoring metric for researchers and papers (for example as used on ResearchGate) to assess outputs which may be worth spending extra time on. It should be noted however that any metric will always be subject to gaming, and it is not unheard of for increased academic impact to be manufactured.
- Including more details on the how the model and code is run: how much data is required for it to produce accurate results? How long does it take to run, and what are the potential bottlenecks when scaling? What are the difficult bits to code?
- Utilise a standard set of benchmarks, this allows comparisons across papers against a common benchmark (see next section 4.5.4).
- Tagging the research with application specific keywords.

4.5.4. IMPROVED BENCHMARKING THROUGH DATA SCIENCE CHALLENGES

Research papers are often used to promote a new or novel technique, and hence in many cases an inappropriate or weak benchmark is used to emphasize the improvement produced by the new approach. Even where a competitive benchmark is utilised, different papers use different benchmarks, with different data, and/or different partitions of the data into training, model selection and testing, making comparisons virtual impossible in standardised ways.

A suggestion from many of our participants was to utilise open competitions to help develop better benchmarks and state-of-the-art models. The competition environment has been utilised successful within Kaggle and other platforms, especially for forecasting (Hyndman, 2020), and the authors themselves have hosted[3] or engaged[4] in competitions for energy applications.

The Codabench open-source platform was specifically designed with the explicit aim to produce competitive crowdsourced benchmarks (Xu, et al., 2021). Competitions have several benefits: they can be designed to combat a very wide range of problems from forecasting to the Netflix collaborative filtering prize and can vary in size, complexity and prizes. The Learning to Run a Power Network (L2rpn) challenge looked at the relatively complex task of real time grid operations using AI. Competitions also have a lot of flexibility in format with many allowing for code submission-type challenges, which means pre-trained models can be included. If submissions are open source this enables others to easily pick up the models and develop for their own use. Framing a problem as a well-defined data science challenge allows the option of comparison in a controlled environment with various constraints. Of course, this means that any competition is likely to be an oversimplification of the real-world problems.

Another drawback of these static data science challenges is that over familiarity on the data and the problem is likely to lead to overtraining. Further it is likely that the utilised data is too short to demonstrate the general performance of a model. One suggestion to reduce these effects was to begin introducing a rolling challenge where new live data is continually added allowing new models to be introduced without overtraining issues. Such an endeavour would require continual Not all problems are amenable to the simplified environment of a data science challenge however,

[3] Presumed Open Data Science Challenge joint with WPD: https://www.youtube.com/watch?v=SetYxiM3rf0
[4] Global Energy Forecasting Competition: http://www.drhongtao.com/gefcom

many adjacent problems can be addressed and tested benchmarks emerging from competitions offer a leg up to new players in different areas, encouraging innovation and new solutions.

5. TRAINING AND THE SKILLS GAP

5.1. DATA SCIENCE SKILLS GAPS

A core purpose of higher education is training, courses, and the transfer of knowledge. As students in areas adjacent to data science move up the academic hierarchy the knowledge becomes more specialised and the focus is more on research skills rather than textbook knowledge, and examinations. For example, a PhD is (depending on the subject) strong evidence that the individual knows how to practice science, i.e., make observations, generate evidence, test hypotheses, and make decisions based on data.

Research and data science skills will be particularly useful for the energy sector especially as organisations digitalise and collect increasing amounts of data. Indeed, PhDs have been a common source of recruitment for industry, and many students are hired based on their direct work with the companies who sponsor their research. Further, for knowledge about more advanced techniques and approaches, academia is one of the few sources for producing individuals in these areas since students can spend an entire PhD (3 or 4 years) diving deep into a subject, eventually emerging as experts in a particular area.

However, there are many organisations concerned with the emerging skills gap. Research into data science employment have shown the many organisations are struggling to fill data science positions. One report from 2020, showed there was three times the job postings for data scientists as there was job searches. Harvey Nash's 2021 survey of senior technology decision makers, the *Digital Leadership Report,* showed big data and analytics skills were highlighted by organisations as being the second most in-demand skills for which businesses were struggling to recruit (Harvey Nash Group, 2021). One academic interviewee who works with many industrial partners in the energy sector highlighted that data savviness was low, even within relatively technical companies. This is consistent with the UK governments recent 2021 policy paper on *Quantifying the UK data skills gap*, (UK Government, 2021), which showed (see Table 11 within the report) that programming was an area identified with the biggest gap between what was viewed as important by employers and which students believed they would have good skills in in the future.

Focusing more on the topic of this report, within the energy sector, the role of digital and data is growing more important and there is a need for the corresponding skills. In particular, the energy sector has a lower digital maturity than many other industries. For example, a recent report (Burnett, Cousins, Baxter, & Bloice, 2021) considered the oil and gas sector stated that the "current approaches to on-the-job learning do not appear enough to ensure the digital skills required by the data community stay current".

In addition to the skills gap, the energy sector is competing with many other sectors for capable and skilled data scientists, from finance to healthcare. Energy has not been involved in data analytics as long as other sectors (e.g. insurance), nor had the access to a wealth of data (e.g. compared to social media companies like Facebook and Twitter, or Streaming services like Netflix and Spotify), nor the financial allure of industries such as Fintech. However, with the increased prevalence of smart meters and other advanced metering, new decentralised products and services, and the importance of climate change on society, energy may become a much more desirable occupation. However, to attract the best data scientists will require making sure more data is available, emphasizing to the wider community that there are interesting (and difficult) problems to solve, and ensuring there is sufficient financing to hire the top talent.

A contrary opinion about the skills gap was also found amongst our interviewees. In some of the organisations of the individuals we interviewed they had found they had no issues finding large numbers of skilled individuals with which to fill their technical roles. In contrast, they mentioned, it was at the top of the organisations, at the managerial levels where the skills gap was most noticeable. New techniques and knowledge are unknown to business leaders and hence they may not trust application of advanced data science models and methods. With this mindset, the required modernisation of an energy organisations digital and data systems may not be realised if a cynical approach is taken to new and upcoming approaches. Hence, a more holistic view of the skills gap may be required to fulfil their digital potential, where training focuses on all levels from technical practitioners, to managerial.

5.2. GRADUATES FOR BUILDING CAPACITY

Universities are of course a major potential source of the hard data skills that the energy industry will sorely need. The specific skills will obviously depend on the area (data science, computer science, mathematics, etc.) and the level (masters, PhD, postgrad etc. see the supplementary report on collaboration, (Haben, & Hinton, 2022a)). For those working in areas of data science there is the direct development of skills in coding, models, methods and techniques, and theory, whatever the level. For areas focused on data science, each level will provide some useful skills for an organisation trying to build their workforce. For a masters course, students will often gain a breadth of foundational knowledge and skills, and a general idea of the landscape. There is often a short research project at the end which will also likely test their ability to apply their skills in a relevant area. For PhD's the students will gain in-depth knowledge in the latest methods and techniques, and likely have a lot more experience in presentation, coding, and advanced models. Specifically, they will have very detailed knowledge in their dissertation topic area by the time they have finished. They will also likely have gained a broader understanding of the landscape, the major organisation, some of the core challenges, and developed a network of contacts.

At postdoctoral research level, the understanding and skills will be much broader. For those who've been in their position for a couple of years or more, the researchers will have gained a much wider understanding of the state-of-the-art, have deeper and wider networks and connections, and have learned to work quickly, being able to understand and applying the knowledge of advanced techniques to solve complex problems. They will also have gained a lot of communications experience, likely presenting at many conferences, teaching several courses, and taken on student supervision etc.

Even without a degree in the specific area of data science, many other skills can be beneficial from hiring postgraduates. Often, they have experience and skills in communication and storytelling, vital in data science for constructing narratives, and getting the information across to business leaders and non-technical experts. Others will also have experience (especially at postdoctoral level) writing bids and teaching and hence will have an extra set of skills and tools for helping develop and grow businesses. Since much of research at university also involves writing technical documents and manuscripts they will also likely have honed their skills in writing well-defined, evidenced, concise language.

Recruiting anyone is always a risk, however, as individuals move up the academic hierarchy from student to researcher there may be increasing evidence to demonstrate their skills and knowledge. One common way for organisations to de-risk recruitment via universities is to have worked with them before. Developing collaboration (see (Haben, & Hinton, 2022a)) with a university has multiple benefits beyond the project itself and one of those is the ability to observe demonstratable skills and work ethic from their partners. Organisations we talked to had recruited many students via collaborative projects.

A close collaboration between industry and academia can also ensure that, for those leaving academia, they have sufficient ability and necessary skills that are in demand in industry. For example, UCL has develop a master programme in Energy Systems and Data Analytics utilising a strong connection with major energy industry organisations to help shape the course. Such a course teaches both important domain knowledge and technical data science information, and many of their students go on to work in the energy industry or have even started their own energy analytics companies.

Although the technical nature of data science means that universities are an important source of these future skills it is worth emphasizing the importance of other skills and knowledge which are not the focus in many taught courses. Many interviewees (in both academia and industry) expressed concerns that students coming from academia are also missing other technical, and non-technical skills which would help them when they enter industry. These include skills in coding, the wider technology stack (e.g. Slack, Git, basic SQL databases etc.), standard coding practices (version control, code review, scalability etc.), communication, and perhaps the most important one mentioned by several participants: domain knowledge. The performance or quality of data science modelling may vary greatly, dependent on the domain knowledge of the practitioner. Students/researchers working in data science areas can have coding skills that range from self-taught beginners to advanced software engineering skills, however our interviews and investigations corroborate that those coming from academic backgrounds commonly lack solid fundamental programming skills. While those academics are likely to be able to produce advanced algorithms and models, they may not be written according to operational standards or incorporate important practices and tests required to validate the model and codes. The specific issues with coding skills are covered in supplementary report on code development for academics (Haben, & Hinton, 2022c).

Finally, on the topic of hiring academics it is worth noting that industry also needs to understand that hiring, say a PhD, is not always about the knowledge they bring, it is also about building capacity. PhDs and Postdocs can assimilate complicated new information relatively quickly and hence, with suitable training, they should pick up the latest technology stack skills relatively promptly. A way to facilitate this would be a readily available list of in-house training or suitable resources for new starters.

5.3. OPPORTUNITIES AND NEEDS

It has been recognised that “there is a need to increase opportunities for students to acquire relevant work experience during their studies” (Wilson, 2012). From our interviews, particularly for data science applications, it is recommended that proper industry level coding practices are taught and encouraged within courses and code sharing is encouraged within the academic research community. However, without a measurable impact for academics (unlike with their academic papers), this output is unlikely to get the attention it deserves (Section 4.5.1). Code development and sharing will be the focus of another supplementary report (Haben, & Hinton, 2022c).

Courses like those provided by UCL can provide a wide range of knowledge for a data scientist to excel in the energy industry. However, as-of-yet, in the UK, there are not many universities running courses on machine learning and AI specifically for the energy industry. An increase of these types of energy focused data science courses would help with skills development and teach students more of the fundamental domain knowledge that industry desires. Further it would also likely increase the number of students choosing to stay within the energy sector when they leave. It is also clear that having input from industry on designing these courses is vital. Such courses need to have the input from the industry to make sure the right skills are being taught, especially in a fast-moving industry such as data science.

For developing training course for industry-level coding skills it would be useful for academics to begin closer connections to industry, who can keep universities up to date with the latest software and technology stacks. One option is guest seminars from experienced industrial practitioners, who can directly develop students' skills and could also be another source of recruitment for the visiting organisation. Alternatively, there are organisations such which offer training, or offer review sessions, tips and guides for software development etc.[5]

Understanding the career paths of academics in industry would help the sector better identify how to recruit and retain skilled individuals and further develop their skills. Unfortunately, there is very little tracking of where individuals go. LinkedIn, for example was mentioned as a previously good source of such data but have since closed off access to their data. Perhaps alumni data collected by universities could be a source of this information, and insights could be generated from an independent body to help the wider sector?

Training goes both ways and some of those industrial participants interviewed mentioned attending courses at universities where the masters programme was opened (for a fee) to external organisations. This provides an opportunity for a two-way sharing of knowledge, with domain specific information and objectives from the industry side and training on up-and-coming techniques or even fundamental data science approaches being shared from the academic side. However, such courses or opportunities did not appear to be widely known. A central repository of such training courses could be useful.

[5] Some of those mentioned within our interviews included Software Sustainability Institute, Software Carpentry, and the Data lab.

6. BRINGING IT ALL TOGETHER

This report and the supplementary parts have all demonstrated many areas to improve transferring the knowledge and tools developed in universities into an industrial context. This section illustrates how these concepts fit into various scenarios where academic contribution can provide benefits to private organisations. We then investigate the next steps for this research.

6.1. SCENARIOS

Scenario 1: Aiming to solve a particular data science related problem

Your data science team are trying to upgrade a particular capability that is a core part of the business. To upgrade this feature means adding increased complexity which no-one in the team is particularly experienced in. Through some discussions at conferences and searching the academic literature they find that there are many researchers who have already worked in this area and have developed a body of knowledge which could be useful for you (Section 4). You discover that some of the papers are made available online (see supplementary report on “Accessibility and Reproducibility” (Haben, & Hinton, 2022b)) but the language is rather technical and although the researchers have released a rather primitive version of their code (supplementary report on “Code Development for Academics” (Haben, & Hinton, 2022c)) it isn’t completely clear to you or the team.

You decide to contact the team who are willing to help you develop their models into the code but in turn they are interested in demonstrating their work with real operational data and adding some of the more practical challenges into their models (Supplementary report on “Industrial Support for Academics” (Haben, & Hinton, 2022d)). This is the start of several future projects together (Supplementary report on collaboration (Haben, & Hinton, 2022a)) with the academic team getting insights to the current and future problems facing the industry, some data, and potentially some funding for students and postdoctoral researchers. For the industry they are now implementing some of the most advanced methods in the industry and are looking to recruit individuals, maybe looking at the universities’ masters program as a source of capacity building.

Scenario 2: Building a new capability within the data science team

A new (or possibly old if emerging from a non-energy sector) topic of advanced data science has been gaining prominence in your sector. Further investigation reveals that it hasn’t been studied in detail even within the academic literature (see supplementary report on “Accessibility and Reproducibility” (Haben, & Hinton, 2022b)) although there have been some small investigations within a couple of universities. You approach one university as you know that they have a reputation for good work in this area, having successfully completed some NIA projects which you saw presented at a large industry conference (Section 4.5.3).

The team suggest a research project in the area to test some the practicality of the approach. You have been asked to quickly demonstrate the viability and hence suggest a short masters research project (Supplementary report on collaboration (Haben, & Hinton, 2022a)). During the project some progress is made, and the technique looks hopeful as a mechanism for understanding the area. The student from the master’s project is interested in pursuing the research further and since you were happy with their work you agree to fund a PhD with the university. Over the next three years, your team provides support by helping to refine and frame the problem, provides data for testing and training the models and also part of the funding towards the project models (Supplementary report on “Industrial Support for Academics” (Haben, & Hinton, 2022d)). Although the final outputs from the PhD are not used, the learning and knowledge that is developed throughout is incorporated into some of your models and some pitfalls and dead ends are avoided due to testing performed by the PhD (see Supplementary report on collaboration (Haben, & Hinton, 2022a)). The PhD student also develops some of the code that finally goes into the operational models and this is made easier since it uses your own data and is guided by your needs (see, the supplementary reports on "Code Development for Academics" (Haben, & Hinton, 2022c) and "Industrial Support for Academics" (Haben, & Hinton, 2022d)).

By the end of the PhD the team is satisfied with the work performed by the student, and they offer them a position within the organisation working on products related to the research and models develop during the dissertation (Section 5.2). The PhD has demonstrated that the student deeply knows the subject (and in many cases may now be one of the experts in the field) and throughout the project has shown they work well within the organisation and the rest of the team.

The hiring of the student creates a stronger relationship with the university and now many new ideas, data and projects are shared with the academics who supervised the original project.

Scenario 3: Aiming to become leaders in an area of data science within energy

Your large cooperation has performed some horizon scanning and has identified several areas where the data science team should be progressing so as to stay competitive or even surpass their nearest rivals. The topics are relatively advanced and there isn't much spare capacity within the team to address the gaps. The data science team decide that one of the best ways to address these gaps and develop advanced skills is to work closely with a university (Section 5.2).

They do a study of the literature and investigate the various prominent universities and the head researchers related to the areas of interest and identify a university which would be a suitable research partner. They read some of their research to confirm the research is aligned with their aims online (see supplementary report on "Accessibility and Reproducibility" (Haben, & Hinton, 2022b)) and arrange to meet with the academics.

The discussion confirms that a long term relationship could be in their best interest and after running some shorter low risk projects (see the supplementary report on collaboration (Haben, & Hinton, 2022a), Section 3) they understand that to get the best out of the relationship they need to form a two pronged approach which address the research and the strategic alignments projects (Supplementary report on collaboration (Haben, & Hinton, 2022a), Section 5). They secure a few branches of work to fund a series of PhD's and postdocs who split their time across research and direct work for the organisation. This enables the academics to pursue longer term, risky research (which also fulfils the academic needs of publishing) as well as more immediate problems that affect the organisation (supplementary report on "Industrial Support for Academics" (Haben, & Hinton, 2022d), Section 3.3). In addition to this the organisation supplies real operational and other monitored data sets to the academics (supplementary report on "Industrial Support for Academics" (Haben, & Hinton, 2022d), Section 3.1). This means that the academics can provide real code, tools and solutions (supplementary reports on "Code Development for Academics" (Haben, & Hinton, 2022c)) and real-world validated results which may aid in their publications and future funding applications.

Their close working partnerships means they also get direct feedback on the academic research and its implementation within the organisation (supplementary report on "Industrial Support for Academics" (Haben, & Hinton, 2022d), Section 3.4) and these various narratives can feedback into the Universities' impact narratives which further support their future research council revenue (Section 4.2). The strategic relationship with the universities means they can also quickly pivot to problems important to the organisation as well as guide each other to key future areas.

6.2. FUTURE WORK

This document is only the start of many of our investigations. It is not intended to be a thorough literature review of the state of transferring impactful research and people to data science in the energy system. Instead, it has collected and shared some of the experience and wisdom from several experts in the area.

There are several actions and recommendations that this paper recognises and some forthcoming work from Energy Systems Catapult is looking at how we may address these matters. Some of the forthcoming work is as follows:

- In April 2022 there was the completion of the WPD data challenges. These three challenges provide an opportunity to assess some of the practicalities of challenges in addressing benchmarking methods (as discussed in section 4.5.4). Next steps for this will be to understand which problems are most suited for a challenge environment as well as see which problems need further development.
- Later in 2022 we will release a document on the results from our data science skills survey we launched in 2021. This will help to understand what skills are in abundance as well as which are in desperate need (Section 5). The next steps from this are to identify where academia can fill these gaps and what other training and resources may be useful. It would be worth understanding what skills in coding are desperately in need solutions (see supplementary reports on "Code Development for Academics" (Haben, & Hinton, 2022c)).
- Catalogue for Energy Data Projects (CoPED): After kicking off the initial phase, in Autumn 2022 ESC will be launching this platform which aims to visibility of the energy data focused projects. Thus aligned with the discussions in Section 4.5.3 we are trying to improve visibility and accessibility of research across the energy sector.

In addition, ever since we launched the Energy Data Taskforce (Sandys et al. (2019)), we are endeavouring to support industry in making their data open. As discussed throughout this and the supplementary reports, data is a key driver for academics and other innovators. We are also hoping to continually increase our ties to academia and bridge the gap between them and industry including through platforming via our "Value in Energy Data" webinar series, and working closely with universities to share our learning.

7. BIBLIOGRAPHY

CER Smart Metering Project, (2012). *CER Smart Metering Project - Electricity Customer Behaviour Trial 2009–2010 [dataset]*. Retrieved from www.ucd.ie/issda/CER-electricity

AI Council, (2021). *AI Roadmap*. Retrieved from https://www.gov.uk/government/publications/ai-roadmap

BEIS. (2021). *Advanced Research and Invention Agency (ARIA): policy statement*.

Bendiscioli, S. (2019). The troubles with peer review forallocating research funding. *EMBO Reports*.

Burnett, S., Cousins, F., Baxter, G., & Bloice, L. (2021). *The Impact of Digitalisation on Data Professionals*. Robert Gordon University, OGUK, RGU Energy Transition Institute.

Cadotte, M., Jones, H., & Newton, E. (2019). Making the applied research that practitioners need and want accessible. *Ecological Solutions and Evidence*.

Chirazi, J., Wanieck, K., Fayemi, P.-E., Zollfrank, C., & Jacobs, S. (2019). What Do We Learn from Good Practices of Biologically Inspired Design in Innovation? . *Applied Science, 9*. Retrieved from https://doi.org/10.3390/app9040650

Clarivate. (2022). *Advance research and accelerate real-world outcomes*. Retrieved from https://clarivate.com/industries/academia/

Colavizza, G., Hrynaszkiewicz, I., Staden, I., Whitaker, K., & McGillivray, B. (2020). The citation advantage of linking publications to research data. *PLOS ONE*.

DuBois, J. (2020). *The Data Scientist Shortage in 2020*. Retrieved February 8, 2022, from https://quanthub.com/data-scientist-shortage-2020/

Elsevier. (2022). *Scopus*. Retrieved from https://www.scopus.com/home.uri

Fang, F., Bowen, A., & Casadevall, A. (2016). NIH peer review percentile scores are poorly predictive of grant productivity. *Elife*. doi:doi: 10.7554/eLife.13323. PMID: 26880623; PMCID: PMC4769156.

Frontier Economics. (2014). *Rates of return to investment in science and innovation*. London: UK Department for Business, Innovation & Skills. Retrieved from https://www.gov.uk/government/publications/science-and-innovation-rates-of-return-to-investment

Grindrod, P. (2020). *Leading Within Digital Worlds: Strategic Management for Data Science*. Emerald Publishing Limited.

Gulbrandsen, K. (2009). *Bridging the Valley of Death: The Rhetoric of Technology Transfer*. Iowa State University. doi:https://doi.org/10.31274/etd-180810-527

Haben, S., Arora, S., Giasemidis, G., Voss, M., & Greetham, D. (2021). Review of low voltage load forecasting: Methods, applications, and recommendations. *Applied Energy*.

Haben, S., & Hinton S. (2022a). Academic and Industrial Collaborations, Data Science: From Academia to Industry – Supplementary Report. Energy Systems Catapult

Haben, S., & Hinton S. (2022b). Accessible and Reproducible Research, Data Science: From Academia to Industry – Supplementary Report. Energy Systems Catapult

Haben, S., & Hinton S. (2022c). Code Development for Academics Entering Industry, Data Science: From Academia to Industry – Supplementary Report. Energy Systems Catapult

Haben, S., & Hinton S. (2022d). Industrial Support for Academics, Data Science: From Academia to Industry – Supplementary Report. Energy Systems Catapult

Harvey Nash Group. (2021). *Digital Leadership Report.*

Herbert, D. B. (2013). Australia's grant system wastes time. *Nature, 495*. doi: https://doi.org/10.1038/495314d

Hughes, A., & Kitson, M. (2013). *Connecting with the Ivory Tower: Business Perspectives on Knowledge Exchange in the UK, .* Centre for Business Research, the UK~IRC and NCUB.

Hyndman, R. (2020). A brief history of forecasting competitions. *International Journal of Forecasting*.

Jerrim, J., & de Vries, R. (2020). Are peer-reviews of grant proposals reliable? An analysis of Economic and Social Research Council (ESRC) funding applications. *The Social Science Journal*. doi:DOI: 10.1080/03623319.2020.1728506

Miller, S., & Hughes, D. (2017). *The Quant Crunch: How the Demand for Data Science Skills is Disrupting the Job Market.* Burning Glass Technologies.

Publons. (2022). *Track more of your research impact*. Retrieved from https://publons.com/about/home

Ramachandran, K., & Watson, J. (2021). *Tech looks to analytics skills to bolster its workforce: Addressing the analysis talent shortage.* Deloitte.

Rossini, A. (2018). *Bridging the technological "valley of death"*. Retrieved from https://www.pwc.no/en/bridging-the-technological-valley-of-death.html

Sandys, L., Dobson, R., Brown, E., Graham, G., Lane, R., & Verma, J. (2019). *Energy Data Taskforce: A Strategy for a Modern Digitalised Energy System.* Energy Systems Catapult.

Sandys, L., Dobson, R., Verma, J., Johnston, G., Roberts, D., Leland, B., Haben, S., Ainsworth, E., Guinta, F., & Pearson, S. (2022). *Delivering a Digitalised Energy System: Energy Digitalisation Taskforce Report.* Energy Systems Catapult.

Sevilla, F., Liu, Y., Barocio, E., Korba, P., Andrade, M., Bellizio, F., . . . al., e. (2022). State-of-the-art of data collection, analytics, and future needs of transmission utilities worldwide to account for the continuous growth of sensing data. *International Journal of Electrical Power & Energy Systems*.

The Data Science & AI Section of the Royal Society, (2021). *The UK AI Strategy: are we listening to the experts?* Retrieved from https://datasciencesection.org/2021/09/13/the-uk-ai-strategy-are-we-listening-to-the-experts/

UK Government. (2021). *Quantifying the UK Data Skills Gap - Full report.* Department for Digital, Culture, Media & Sport. Retrieved from https://www.gov.uk/government/publications/quantifying-the-uk-data-skills-gap/quantifying-the-uk-data-skills-gap-full-report

UKRI. (2022). *Knowledge Exchange Framework*. Retrieved from https://kef.ac.uk/

Williams, K., & Grant, J. (2018). *A brief history of research impact: how has impact assessment evolved in the UK and Australia?* Retrieved from LSE: https://blogs.lse.ac.uk/impactofsocialsciences/2018/03/19/a-brief-history-of-research-impact-how-has-impact-assessment-evolved-in-the-uk-and-australia/

Wilson, T. (2012). *A review of business-university collaboration.* Department for Business, Innovation and Skills.

Xu, Z., Zhao, H., Tu, W.-W., Richard, M., Escalera, S., & Guyon, I. (2021). *Codabench: Flexible, Easy-to-Use and Reproducible Benchmarking for Everyone*.

OUR MISSION

TO UNLEASH INNOVATION AND OPEN NEW MARKETS TO CAPTURE THE CLEAN GROWTH OPPORTUNITY.

ENERGY SYSTEMS CATAPULT
7TH FLOOR, CANNON HOUSE,
18 PRIORY QUEENSWAY,
BIRMINGHAM, B4 6BS.

ES.CATAPULT.ORG.UK
@ENERGYSYSCAT