Data Synergy in times of crisis

Dr Sarah Higginson
Environmental Change Institute
University of Oxford
South Parks Road, Oxford, OX1 3QY, UK
Email: sarah.higginson@admin.ox.ac.uk

Dr Marina Topouzi
Email: marina.topouzi@ouce.ox.ac.uk

Dr Gesche Huebner
Email: g.huebner@ucl.ac.uk

Catherine Jones
Email: catherine.jones@stfc.ac.uk

Dr Michael Fell
Email: Michael.fell@ucl.ac.uk

Abstract

We face multiple, interlinking crises, all of which require the collection and sharing of quality data to understand them. Sharing data is good practice for responsible research and often a funder requirement. However, many projects still fail to deliver on the FAIR data sharing principles (that data be Findable, Accessible, Interoperable, Reusable). Whether qualitative or quantitative, high quality data collection, management and analysis is a crucial foundation for excellent, socially relevant research, particularly when it is interdisciplinary and the assumptions underpinning single disciplinary ontologies and methodologies might be contested.

Energy research produces interesting, specific data challenges: 1) the prevalence of large-scale consortia means many institutions are involved; 2) the multi-disciplinary approach favoured in such consortia results in a wide variety of domain standards and expectations; 3) as an applied area of study, energy researchers often collaborate with commercial partners, who may restrict data sharing.

Building on the authors' experiences of data management in RealValue, an H2020 project, and two UK-based consortia, the Centre for Research into Energy Demand Solutions (CREDS) and UK Energy Research Centre (UKERC), plus learning from a recent expert workshop of research system stakeholders, this paper will explore the concept of 'data synergy'.

Data synergy, a term coined during RealValue, describes data from multiple stakeholders, sources or disciplines that, when combined, are more valuable than any of the sources on their own. It has four dimensions – resources/ time, people, methods/ metrics, and technology – and considers data collection, sharing and management a socio-technical process that balances these dimensions.

The aim of this paper is to elucidate a set of principles and processes that will guide the international energy community moving forward, ensuring we are able to meet future challenges quickly based on FAIR data, whatever the project focus or methodology.

Introduction – the literature and context

The importance of sharing research data is well established. It can improve the productivity of research by allowing more researchers to generate insights for the same data collection effort, and avoid duplication in data collection. It also contributes to greater transparency, allowing other researchers to check and attempt to replicate analysis, ultimately improving quality (Tenopir, 2011, Piwowar, 2007). Nowhere is this more important than the field of energy demand research, where generating reliable insight quickly is a vital part of addressing the climate crisis. However, while significant work has gone into developing principles and infrastructures to support data sharing, too often energy research data still ends up being inaccessible to other researchers. This paper explores the reasons for this and presents a range of recommendations for how the situation can be improved. First, however, it provides the wider research policy context and an overview of previous work.

Data Sharing - the Policy Context

As the authors are based in the UK, the paper starts with the UK context before considering other worldwide initiatives. Research Councils UK (RCUK, the precursor to UK Research and Innovation, UKRI) published Common Principles on Data Policy in 2011, building upon existing Research Council policies. These principles cover a range of topics including the importance of making publicly funded data available and discoverable, following best practices for data management, complying with ethical or legal requirements and giving credit to those who produced data that is being re-used. This effort was further complemented by the issuance of Guidance on Best Practice in the Management of Research Data (RCUK, 2015). Individual Council policies were tailored to address specific issues and expectations within their respective domains. Jones (2012) provided an historical overview of the progression of UK funder policies, noting the necessity of translating these policies into tangible

infrastructure and incentive structures. In 2016, the UK took a significant step towards endorsing open research data through the Concordat on Open Research Data, which was jointly supported by RCUK, the Higher Education Funding Council of England (HEFCE), Research England, and Universities UK. This commitment to open research data has been sustained, as demonstrated by the current research data guidance provided by UK Research and Innovation (UKRI): *Making your Research Data Open*[1] and *Policies and Standards on Data Management and Sharing*[2]. The expectation is that UKRI funded research will be as open as possible.

At the same time the EU were developing their approach to research data from the programmes that they fund. The current policy (EU, 2019) expects data to be "as open as possible and as closed as necessary" and is funding the European Open Science Cloud to provide infrastructure for data sharing across different domains and communities and to provide a data portal[3]. In the Horizon2020 programme, data management plans are expected to be produced and kept up to date (EU, 2017).

Similarly, Australia and the United States have been actively engaged in promoting open research data sharing, with a focus on increasing transparency, facilitating collaboration, and maximising the societal and economic benefits of research. In Australia, there is a strong focus on open access and data sharing for publicly funded research, with the government endorsing the Organisation of Economic Cooperation and Development (OECD)'s expanded recommendation on data access. The Office of the National Data Commissioner oversees public sector data use, emphasising government service delivery, policy formulation, and economic benefits. Australia also actively engages in international data-sharing partnerships, influenced by political, economic, and legislative factors. In the United States, federal agencies like the National Institutes of Health (NIH)[4] and the National Science Foundation (NSF)[5] have policies encouraging data sharing and open access to research data in order to maximise public benefit. Initiatives like Data.gov provide access to numerous federal datasets, reflecting the government's commitment to open data and transparency.

While details may depend on the type of data or the location of the funder, the approach internationally, is that research data should be available to a wider community than the original creator of the data, within ethical and legal restrictions.

Data infrastructure in the UK

Data infrastructure is an overarching term, encompassing both the technical underpinning and the socio-cultural aspects to support data sharing. In the UK, some funders, in particular the National Environment Research Council (NERC) and Economic and Social Research Council (ESRC), have supported the development of physical data infrastructures, such as through the provision of domain focussed data centres, for example the Centre for Data Analytics (CDA), the UK Data Service (UKDS) and the Energy Data Centre (EDC). There is also a network of Research Performing Organisation (or university) institutional data repositories.

For energy researchers, there are two domain-specific repositories: UKDS for social science/mixed methods research and the EDC for more quantitative outputs. However, researchers also may choose to use their local University repository, general services such as Zenodo or Figshare, or NERC repositories if their research is more environmentally based. As energy researchers are multidisciplinary they may also choose to deposit their data in various disciplinary repositories, and this leads to a fragmented landscape for those trying to discover energy data.

Following the development of RCUK policy and guidance, data management plans were adopted as a tool to support researchers in considering data management issues before data is generated/collected. From the start training of researchers has been identified as key, as discussed by Ward (2011), who states that many of the issues identified for effective data management planning are "a people-problem not a technical-problem". However, Smale's (2020) review of the development and efficacy of Data Management Plans (DMPs) concludes that, although they are used as a proxy for data management planning, they have not had as much impact as had been hoped on research processes and actors in the ecosystem. A survey conducted by Doonan (2020) among US institutions, found that data sharing training was not fully integrated into their academic programs.

---

[1] Making your Research Data Open: https://www.ukri.org/manage-your-award/publishing-your-research-findings/making-your-research-data-open/

[2] Policies and Standards on Data Management and Sharing: https://www.ukri.org/who-we-are/mrc/our-policies-and-standards/research/data-management-and-sharing/

[3] https://data.europa.eu/en

[4] National Institutes of Health (NIH): https://www.bing.com/search?pglt=41&q=NIH+america+data&cvid=c48d4f76b9e04f7f80fb330657fdcc89&gs_lcrp=EgZjaHJvbWUyBggAEEUYOdIBCDY2MjlqMGoxqAIAsAIA&FORM=ANNTA1&PC=U531#

[5] National Science Foundation (NSF): https://www.bing.com/search?q=nsf+america+data&cvid=6a4c05329fd84aa9977bd332b4384bf7&gs_lcrp=EgZjaHJvbWUyBggAEEUYOdIBCDU3NzdqMGo0qAIAsAIA&FORM=ANAB01&PC=U531#

In the UK, the UK Data Service (UKDS) leads in providing data management training for social scientists to support them in depositing their data[6]. However, significant variation exists in data sharing practices across sectors and regions. Since 2011, various research domains have adopted distinct standards and expectations for data sharing, influenced by the types of data collected and their legal requirements. For example, domains like Engineering, lack well-established community standards due to potential intellectual property issues with commercial partners, while research involving human subjects and requiring stringent ethical approval, often imposes data sharing restrictions and time delays. As Matthews (2015) highlights, to ensure data remains discoverable and usable over the long term, it is essential to provide persistent identifiers, metadata, and domain-specific ontologies. Alongside the training of researchers has been the development of a new professional role, in the form of data managers, or data stewards (synonymous terms), to support the process of data management. Wendlehorn (2023) discusses the aspects of this role in more detail.

In addition to the development of high-level data infrastructure policies, and supporting processes driven by funders and institutions, there were also researcher-led initiatives emphasising the importance of data sharing. This led to the publication of the FAIR Data principles (Wilkinson, 2016), followed by subsequent efforts to apply these principles to research software (Barker, 2022). The FAIR principles[7]state that data should be Findable, Accessible, Interoperable and Reusable, establish the need for both human and machine-readable metadata and highlight the importance of ensuring licensing and use of domain standards for describing the content. These principles do not imply that the data should be open, rather determining that the way data are accessed should be public. By adopting the FAIR principles, research data becomes more valuable as re-use and reproducibility is a fundamental part of the creation and storage. Another research-led initiative is the Research Data Alliance[8] (RDA), whose mission is to "Build the social and technical bridges to enable open sharing and re-use of data" and which works through multidisciplinary & multinational working groups to create standards and best practice guidance. It currently has 14,139 members and 108 groups. Similarly, the Transparency, Reproducibility, and Quality (TReQ) principles (Huebner, Fell, and Watson, 2021) encourage open data as one of the crucial tools in conducting more transparent, reproducible and high-quality research.

Despite domain variations, it is now well established that sharing data and having others use and cite that data, enhances research careers and supports reproducibility by enhancing the transparency of the research. The sixth RCUK Common Data principle[9] is acknowledgement that data use, publishing and enabling the citation of data has gone from a recommended change as discussed in (Lawrence, 2011) to an increasingly standard practice, as identified by the survey of Dryad[10], an open data publishing platform, where articles undertaken by (Mayo, 2016) show an increase in formal data citation. However, data is still under-represented in formal assessment, for example, the latest UK Research Evaluation Framework results for UK universities (UK REF2021) show that out of 185,353 submitted outputs, only 31 were described as datasets, so there is still a long way to go.

Having set the historical context for data sharing, the paper will now reflect on its practical implementation in two prominent UK energy consortia, the UK Energy Research Centre (UKERC)[11], and Centre for Research into Energy Demand Solutions (CREDS)[12].

Methodological Approach

The insights and recommendations discussed in this paper come from a number of stands of evidence, experience and forms of engagement. This section starts by discussing the practical experience of supporting data sharing within two UK energy consortia and then discuss the outcomes of a workshop of systemic stakeholders on the subject, organised by the authors.

The data management experience of two UK energy consortia

The UKERC consortium, in its twentieth year, conducts interdisciplinary research focused on sustainable future energy systems. Their work addresses the challenges and opportunities arising from the shift towards a net-zero energy system and economy. The current fourth phase, 2019-2024 was an £18 million consortium involving 150 researchers in 20 institutions. Similarly, CREDS, a five-year (2018-2023), £19.5 million research consortium involving 150 researchers from 24 institutions, conducted interdisciplinary research with a vision to make the UK a leader in understanding the changes in energy demand needed for the transition to a secure and affordable, zero-

[6] UK Data Service (UKDS) https://ukdataservice.ac.uk/learning-hub/research-data-management/
[7] GO FAIR Initiative - GO FAIR: go-fair.org
[8] RDA | Research Data Sharing without barriers: rd-alliance.org
[9] RCUK has been replaced by UKRI and their data principles can now be found here: Data management and sharing – UKRI
[10] Dryad is an open data publishing platform and a community committed to the open availability and routine re-use of all research data: https://datadryad.org/
[11] UKERC: https://ukerc.ac.uk/
[12] CREDS: https://www.creds.ac.uk/

carbon society. This included investigating the technical, social, and governance challenges related to demand reduction, flexible energy consumption, and the utilisation of decarbonized energy sources.

In UKERC's fourth phase, data management planning was a key performance indicator and was managed by the Energy Data Centre (EDC). EDC is a capability within UKERC and provides a discovery and curation service for the community focussed on energy data, projects and grey literature publications. They support UKERC researchers and provide information on data management, sharing and FAIR data on their website. EDC has a metadata record for UKERC4[13], data collated regardless of location, which thereby holds a complete record of all UKERC data outputs.

EDC supported UKERC colleagues to ensure each internal project created a data management plan. Data management and data availability were tracked and progress reported, not standard within the funded community. The EDC adopted the approach "As open as possible, as closed as necessary" and recommended that data be shared through deposits in a domain repository such as the EDC or the UK Data Service (UKDS), where potential re-users would expect to find it. Data managers from EDC supported UKERC themes through presentations at theme meetings and one-to-one meetings to discuss data management concerns.

As a result, data has been deposited with the EDC and UKDS from completed theme projects. A small number of projects are not expected to deposit data at all, due to the nature of the research (e.g. elite interviews). At the conclusion of Phase 4; it will be easier to discover and re-use the outputs than those from previous phases; moving data sharing forward in the UKERC community. The destination for the UKERC4 data is 39% to the EDC, 25% to the UKDS, 7% cannot be shared, 11% has 3rd party data with other restrictions and 18% is undecided/unknown.

In CREDS, archiving data was a contractual requirement, with non-compliance risking payments being withheld. Initially, CREDS formulated a comprehensive data management plan for the entire program and embarked on the recruitment of a data manager. This proved challenging, requiring three recruitment rounds to recruit someone who stayed only 18 months before moving on to better prospects. Nevertheless, CREDS successfully created a data management template and provided training, enabling each project to develop a data management plan, a task that most projects completed.

Subsequent to the data manager's departure, an internal Research and Data Quality project was initiated. This project engaged representatives from each of the nine thematic areas as 'Quality Champions' to ensure widespread communication across the consortium, and delivered regular presentations during Whole Centre Meetings. The project's primary objectives were:

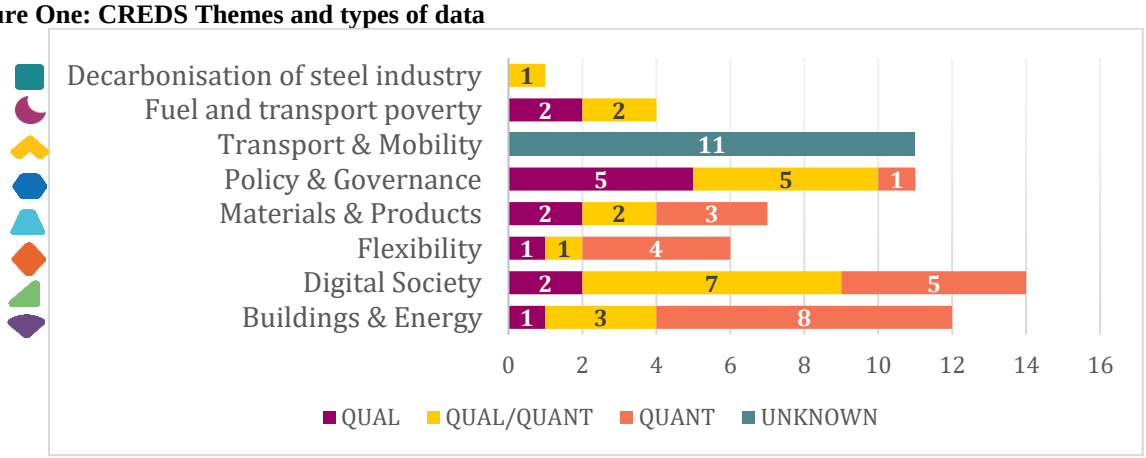
1. Promoting Transparency, Reproducibility, and Quality (TReQ) in data. This culminated in the creation of a series of six professionally filmed videos[14] covering: Principles, Pre-registration, Reporting Guidelines, Pre-prints, Open Data and Code, and a Checklist on how to document usage of these tools. These resources were disseminated through meetings, newsletters, and a paper addressing the enhancement of good quality energy research practices (Huebner, G., Fell, M, 2020).
2. Collection and cataloguing of the data used within CREDS, utilising the data management plans, the outputs of which are discussed below (see Figure One).
3. Improving the awareness of our community about data issues[15], done through training and involving a 'quality champion' in the project from each of the nine CREDS themes.
4. Engaging with the UK Data Service (UKDS), with whom CREDS data is archived. As a result, they conducted two tailored training sessions to enhance researchers' skills, complemented by a range of training resources available on their website and personalised assistance to individual researchers. Data submissions to UKDS undergo a stringent review process, including necessary corrections and quality control measures to ensure data submissions meet established standards.
5. Collaborating with UKERC and UKDS to enhance the archiving and discoverability of CREDS data.
6. Engaging with the wider energy research ecosystem, which took the form of a workshop and is discussed below.

The output of objective two above, 'collection and cataloguing of the CREDS data', provides insight into the diversity within energy consortia. Figure One diagrams the results of the data catalogue process on the basis of the returned Data Management Plans in 2022. It shows the nine themes, each with their unique CREDS symbol and colour. The graph shows how many projects each theme undertook (a total of 66 overall, though this is less than the final number of projects because this analysis was done before the end of the programme). The colours in the graph denote whether the methods used in each project in each theme were qualitative, quantitative or mixed, with one theme unknown because they did not return their data. It is worth noting the wide range of subject areas covered by the themes, each of which was interdisciplinary in its own right; all of which have their own ontologies, languages and cultures; and most of which were cross-institutional. Equally diverse was the range of data, from

[13] Metadata record for UKERC4: http://www.ukerc.rl.ac.uk/
[14] Six TReQ Videos: https://www.creds.ac.uk/treq-video-introduction/
[15] CREDS/ UKERC Data Sharing Briefing Note: https://doi.org/10.5286/UKERC.EDC.000971

secondary and primary sources, including quantitative (for example Smart Meter Research portal data, the Health Survey for England data, Secondary analysis of time-use data, Literature review of buildings retrofit policy, and a UK-Germany case study), and qualitative (for example interviews with National organizations and local authorities, focus groups with transport stakeholders and workshops with representatives from the construction industry).

Figure One: CREDS Themes and types of data



Data sharing workshop

An expert online workshop, Improving Data Sharing in Energy Consortia[16], was held in October 2023, and aimed to get the research ecosystem into the room. Having started to collaborate, CREDS and UKERC decided to deliver the workshop jointly in order to bring together the different experiences of the authors, as outlined in Table One, which tracks how the diverse experiences of the authors ultimately resulted both in the workshop and, subsequently, this paper.

Each of the authors had experiences prior to working in their respective consortia that made them determined to pursue this agenda and improve data practices in their consortium. Not only was this somewhat achieved, but the two consortia were able to collaborate as well. It is hoped that the fruits of that collaboration, including this paper, lay the foundation for further work to be done by the new Energy Demand Research Centre (EDRC)[17], which follows on from CREDS and, hopefully, a fifth phase of UKERC.

Table One: Summary of learning journey that led to this paper

<table>
<tr><td rowspan="4">Authors' experiences of data management over a period of 10-20 years</td><td>UKERC, EDC</td><td>CREDS, Oxford</td><td>CREDS, UCL</td></tr>
<tr><td>Historical overview of the progression of UK funder policies. Contribution made to the RCUK policy guidelines (Jones, 2012).</td><td>Data Synergy concept developed during RealValue (Higginson et al., 2018).</td><td>TReQ tools developed and paper written. (Huebner et al., 2017)</td></tr>
<tr><td rowspan="2">EDC takes on data management in UKERC. Projects produce Data management plans.</td><td colspan="2">CREDS hires data manager and produces data management. Archiving through UKDS who offers training and bespoke support.</td></tr>
<tr><td colspan="2">Data project set up by researchers from Oxford and UCL. Recruits quality champions and catalogues data. Six TReQ videos and a working paper produced Huebner, Fell (2020).</td></tr>
<tr><td colspan="3">Enhanced discoverability of CREDS data project (CREDS, UKDS, UKERC). Joint CREDS/ UKERC expert workshop arranged, facilitated by Cultivate Innovation Ltd. Data collected using FAIR framework and analysed using Data Synergy Framework. Recommendations written up in report and blog. Joint paper written for eceee.</td></tr>
</table>

[16] Improving Data Sharing in Energy Consortia: Summary of Workshop Outputs: https://doi.org/10.5286/UKERC.EDC.000970

[17] EDRC: https://www.edrc.ac.uk/

The workshop was accompanied by a *Briefing Note* (Jones & Higginson, 2023) to set the context, summarise the common lessons learned in CREDS and UKERC, and form a starting point for the workshop. It highlighted six common lessons learned/areas for improvement on data sharing:

1. There is a need for specialist data managers to support data processes, especially in large consortia, but they can be hard to recruit and retain.
2. It requires time and effort for researchers to share and archive data but incentives are lacking in many domains, and expectations and skills vary widely across different disciplines and domains. The difference in domain expectations in respect to data is most often overlooked in interdisciplinary collaborations.
3. Consortia need to set expectations for staff and any additional funded teams or projects (such as through flexible funding). This includes setting up systems like induction materials, guidance on how data should be cited, project reporting frameworks that explicitly include data processes, and ways of giving credit to and recognising those who excel.
4. The interactions between employing institutions and large consortia can complicate matters. For example, ethics processes are conducted by institutions but will affect a consortium's ability to share and archive data.
5. Key stakeholders such as publishers and funders have a vital role to play. Their expectations and requirements around data sharing and archiving could provide a huge incentive to researchers to prioritise this area, and is already starting to have an impact. However, carrots need to be provided as well as sticks.
6. Particular types of data can be complicated to share but pose very different challenges. On the one hand there are qualitative data, such as elite interviews where stakeholders might be identifiable, while on the other, quantitative data such as models or new types of data containers pose significant technical challenges. Starting to set norms and standards across the energy research field would be helpful for both researchers and data managers trying to navigate this complexity.

Having set the context with the *Briefing Note*, the workshop was co-designed with and facilitated by Cultivate Innovation Ltd. and brought together 39 expert stakeholders from the energy research ecosystem, including consortium leads (n=14), data managers (n=4), researchers (n=14), publishers (n=1), and funders (n=6). Its aim was to explore lessons related to data sharing in the energy research domain from the perspective of different stakeholders and to develop recommendations based on these lessons.

A visual tool (Mural) was used to capture responses from participants during group discussions. During the workshop responses were arranged using the FAIR framework but this proved inelegant because the principles apply to data rather than learnings. The report, and this paper, are therefore organised using the Data Synergy framework (Higginson, et al. 2018). This was developed during RealValue[18], a large consortium of industry and academic partners from different countries and research backgrounds who collaborated around diverse sets of data collected in real-world settings. The framework describes data from multiple stakeholders, sources, or disciplines that, when combined, are more valuable than any of the sources on their own. It considers data collection, sharing, and management as a socio-technical process that balances these dimensions. The framework has four key 'pillars' for effective data sharing:

- **Time/ Resources:** The need for synchronisation between research objectives and project management, as well as the need to provide time and resource for the tasks required.
- **People:** The impact of different actors in the process, and their skills, experiences and knowledge.
- **Methods/ metrics:** The potential for consistent metrics or a set of principles that facilitate project comparisons and are sympathetic to both quantitative and qualitative research traditions.
- **Technology:** This includes both testing novel technologies and data collection technologies.

The report, *Improving Data Sharing in Consortia* (Colechin, 2023), was based on a thematic analysis of the data and was shared with participants from the workshop. It highlighted lessons learned, key barriers to data sharing, and recommendations on the prerequisites for effective data sharing.

Analysis and discussion

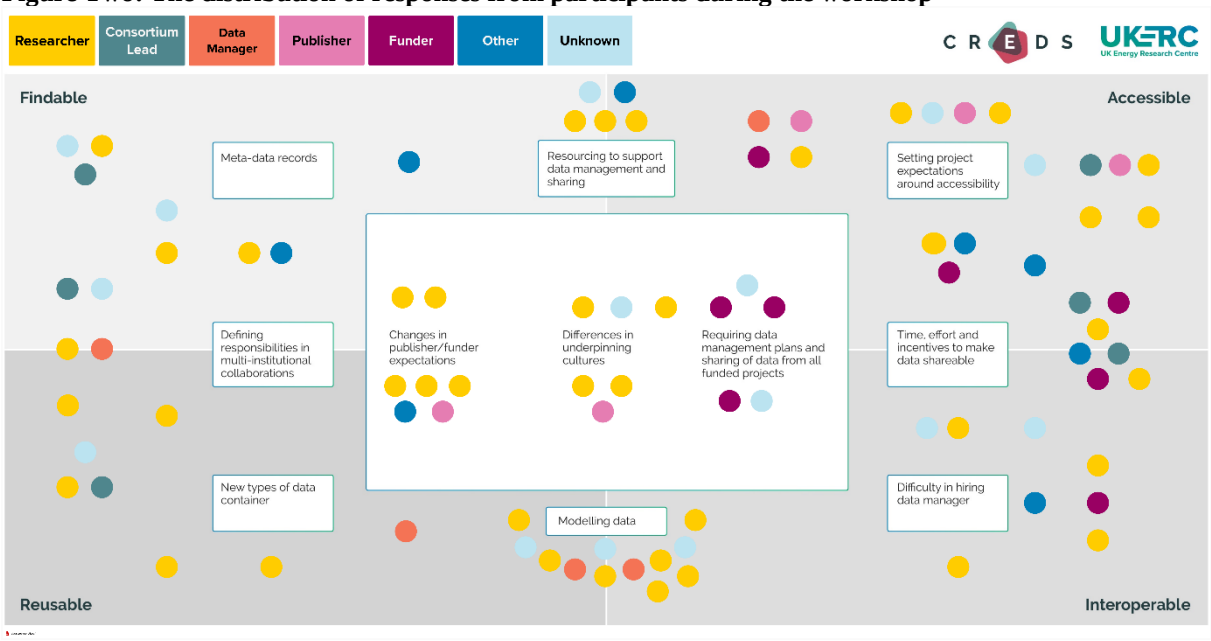
This section brings together the practical experience from CREDS and UKERC and the findings from the workshop to further explore the issues raised and to suggest some recommendations for change.

The *Briefing Note* allowed the lessons from CREDS and UKERC to be taken into the workshop and tested against the experiences and priorities of other ecosystem stakeholders: funders, publishers, data managers, consortium leaders and researchers. Figure Two shows how they responded by capturing a summary Mural board, which is divided into the four FAIR principles. Stakeholders self-identified as to their sector and were divided into small mixed groups to discuss and respond to the lessons coming out of CREDS and UKERC, which appear in the text

[18] RealValue: https://cordis.europa.eu/project/id/646116

boxes. Each of the dots represents a lesson about, or barrier to, effective data sharing from a colour-coded stakeholder.

Figure Two: The distribution of responses from participants during the workshop



For the reasons explained above, the workshop report and this paper analyse the results of the workshop using the data synergy framework, as follows.

Time and Resources

The most significant lesson and barrier related to resources – the lack of time and resource available to researchers in general, as well as the tendency to underestimate the amount of time and resource required for sharing data, combined with a lack of incentives and poor oversight means it is seldom a priority. It can be difficult for an individual to quantify the benefits of the time investment, but if the consortium is clear about the data sharing expectations, both for what should be shared and what is not required, then researchers and data managers can concentrate on research data that add value to the energy community. While consortia cannot themselves provide professional credit to researchers for the outputs, valuing the sharing of data and recognising the contribution of those involved creates a virtuous circle.

As already alluded to, energy research is often done in large consortia undertaking research within a wide variety of energy-related disciplines. Each consortium will have provided information on overall data management as part of their funding bid, but setting effective standards and expectations to support and monitor the outcomes of all the different projects within the consortium requires significant work, which such consortia may lack the skills to do. For example, the difficulty of managing ethics processes, collaboration agreements and data management protocols across different institutions in large consortia is often not accounted for. Greater recognition of data as a valuable output in its own right by both institutions and research centres, supported by funder policies to share data, would improve the situation. Data management planning early in the project lifetime is essential to understand what data can be shared and will be useful, and making sure all the resources, systems and permissions are in place for this to happen. Guidance from institutions or even funders on how to produce data management plans may support researchers and help to improve skill levels, which can be patchy.

It is worth noting that, even in CREDS and UKERC, which were well resourced with large core teams and dedicated data management support, sharing data was a challenge. With this in mind, it was pointed out by workshop participants that much research is not done in consortia but rather conducted as small (sometimes very small) projects. Participants stressed the need for systems to be ‘fit for purpose’ – that they remain as simple as possible and flexible, able to adapt to changing circumstances or the differing needs of different researchers. Crucially, they should allow researchers to think about the cost/benefit analysis involved in preparing data for sharing, because data sets are not equally valuable, there is significant effort in creating enough metadata and context for data, and it is not possible to prepare even the most valuable data for all possible uses. There is also, of course, an energy cost to storing data and it is worth asking how long data should be stored if it is not being used.

People

Closely related to the first pillar, 'people' constituted an important set of lessons and barriers. The lack of incentives and clarity about the benefits of data sharing have already been mentioned. Perhaps even more relevant, however, is the need to put systems in place such as adequate training (including in doctoral training centres), proper induction processes so that everyone is clear about expectations (including those joining later in a consortium's lifetime), clear leadership from senior members of staff (who should act as role models in this regard) and, ideally, support from a data manager or, at least, a data centre (such as UKDS or EDC). Such support would also help to demystify the process, which is often a barrier. Where there is resource to hire a dedicated data manager, the wide range of tasks required in the energy space needs to be recognised so they are given a clear steer on how to operate, and compensation needs to be adequate enough to make the role attractive.

The roles of different types of stakeholders in the research ecosystem is also relevant here. Funders can set clear but flexible frameworks with good guidance. Publishers have a lot of power to require that data is shared but should also recognise the constraints researchers face, particularly in small projects. They can provide additional incentives for data sharing by accepting 'short/data papers' or even publishing 'data descriptors' as a key outputs of a research project, rather than a by-product. This could incentivise researchers to publish their data. Both funders and publishers could support research that uses secondary data, rather than always prioritising novel and original research. This would certainly help to make the data that is shared more useful. The Economic and Social Research Council's (ESRC) responsive mode call for secondary data analysis set to close in September 2024[19] is welcome in this context.

Institutions need to align their systems to make data sharing as simple as possible. Importantly, many ethics processes still insist on data only being seen by the immediate project team, or destroying data after a few years (which precludes archiving), and may forget to remind researchers to get permission from participants to share or archive data (as distinct from using it for publication). Tools such as well-designed consent forms, impact statements and ethics declarations are helpful here. The balance between protecting participants and sharing data, and the requirements of funders, need to be more clearly communicated in some institutions.

Given the lack of data sharing across the energy community at the moment, research centres have an opportunity to lead the way, as CREDS and UKERC have tried to do. Setting standards and expectations is a key driver - consortia are a proxy for the community and researchers are very influenced by the communities to which they belong. Being transparent about what they learn as they do so, and sharing those learning stories as is being done in this paper, is invaluable. Creating a cross-consortia forum for sharing expertise in supporting energy researchers is an idea that is being explored. There is also an opportunity for consortia (and fora) to signpost relevant training such as Huebner et al. (2020).

Research partners, particularly commercial ones, may need quite a lot of persuasion to share data, or may refuse to do so at all. Researchers wanting to work with them will need to remember that they may need to sign non-disclosure agreements, and that gaining access to commercially sensitive data takes a long time and requires detailed negotiations, a set of skills in itself.

The interactions between all these stakeholders is, of course, complex as each has different motivations and operational norms. The researcher is key to enabling the collection/creation of the data so that it is able to be shared, however they fit into a wider ecosystem. Continuing to have the sorts of 'whole system' conversations started in this workshop in order to deepen mutual understanding and cooperation would seem to make sense.

To enable data sharing expectations to be realised, consortia need to value and invest in data management professionals and ensure that the data planning process is not considered to be administrative but adds value to the ensuing data.

Methods/ Metrics

The development of the Data Synergy pillars included a recommendation to develop consistent metrics that would allow comparisons of data across contexts. It also suggested the need for protocols to establish conventions for collecting and sharing data, both quantitative (e.g. what to capture, how often, where and by what means, or what scales to use for variables such as age, income and cost) and qualitative (e.g. suggesting sets of questions to frame semi-structured interviews around particular subjects, like comfort, control or convenience). There is the potential for consistent metrics or a set of principles that facilitate project comparisons and are sympathetic to both quantitative and qualitative research traditions. It is acknowledged that the Interoperable principle within FAIR is the hardest to achieve and within communities of practice, coming towards standards and common ontologies is one way of moving towards improvements. Such consistent metrics and concomitant sets of data sharing principles would facilitate quality data and enable comparisons across different project's findings, saving countless researcher hours spent in cleaning data. Meta-data has a role to play here too, particularly when seeking to deliver against FAIR principles of data sharing.

As has been pointed out above, energy research faces a significant challenge due to its interdisciplinary nature and multiple research approaches, data sources and data repositories across different communities and disciplines. Even the purpose for which data is collected and shared varies widely. For example, engineers might focus on the

[19] ESRC call: https://www.ukri.org/opportunity/esrc-responsive-mode-secondary-data-analysis-round-two/

repeatability of an experiment, requiring input data. On the other hand, modellers may see inputs and outputs as being as important as the modelling process itself. For certain types of models such as stochastic models, reproducing outputs from inputs is non-trivial, making it necessary to store all model outputs which can involve huge amounts of data.

Energy models offer a particular area of challenge in data sharing as they blur the difference between data and software, and are an example where the data may be meaningless without the software as discussed in (Davenport, 2020), The RCUK guidance suggested that output data need not be kept if input data and the code were stored. This causes issues if the modelling or simulation has some sort of randomisation involved. The energy community is a large producer and user of models but common standards for what to archive to enable FAIR data and reproducibility have yet to be agreed, an important concern for sharing and FAIR-ness in the energy community which needs to be resolved and addressed. Energy consortia have the opportunity to be involved in this setting of standards.

Meanwhile, 'experiments' in the social sciences are generally non-repeatable by definition, and data sharing is complicated by the need to ensure participant anonymity and confidentiality. This, however, does not preclude the possibility of developing some protocols. Data sharing would certainly help strengthen qualitative studies which often have small sample sizes due to the expense of collecting data and are less often shared due to the time-consuming nature of preparing them for archiving and the need for anonymity.

Technology

Some consideration was given by workshop participants to the role of technology in data sharing. The key 'technology' in data sharing is data management plans (DMPs), which should be in place at the start of projects, and then consulted and reviewed throughout the lifetime of the project. CREDS and UKERC were both empowered to support the creation of such plans, and to monitor progress through to project completion, in a way that a funder or an academic institution may not have the resource to do. This has a financial impact on the programme but has demonstrated value by the data which is now available. While there is work on making DMPs machine actionable through a Research Data Alliance Interest Group [20], and thus less labour intensive to maintain for researchers, this is not embedded in the energy research culture.

There was also some discussion in the workshop around more technical tools to ensure that data is secure when sharing across institutions, and the use of proprietary software like SPSS for making data sharable. There are tools available for institutions to automate the data management process such as DMPOnline[21] and this might be an area where consortia can come together to influence tools. However, the technical discussions mainly focused on the potential use of artificial intelligence approaches to make use of less organised data sets and the use of Large Language Models for data discovery. The challenges around sharing and preserving data generated by AI is an emerging topic and relevant to energy demand work (Rozite et al, 2023).

Participants also moved beyond the creation of data sets to discuss analysis of when and how data is accessed. This process is easier if the data sharing has a clearly defined purpose and the data itself is able to be standardised. Participants in the workshop highlighted the Cambridge Crystallographic Data Centre (CDDC) as an organisation that had been particularly successful in this regard. Researchers who produce this type of data deposit it with the CDDC as part of the publication process. While data used in the spectrum of energy research cannot be standardised in the same way as crystal structures, perhaps there is an opportunity for subsets of the community to agree on common standards to optimise the potential for data sharing.

Work on embedding FAIR principles in all parts of the research eco-system is an ongoing challenge. A particular challenge for UK multidisciplinary energy researchers is the fragmented repository landscape. There is an opportunity to use standards and techniques to harvest and ingest information from services automatically. As a start to this process, colleagues at EDC and CREDS worked on automating the creation of metadata-only records in the EDC, based on the information held in the UKDS, to improve the discoverability of CREDS outcomes. The EDC is also part of the FAIR-IMPACT EU project[22], which aims to support repositories to become FAIR-er.

Recommendations

The practical lessons learned by CREDS and UKERC, and the expert workshop discussions led to a set of recommendations, listed in the workshop report and reproduced below in Table Two. These recommendations identify key themes, which are also evident in the discussion above. Again, while many of the key concerns are not specific to the domain, the opportunity for consortia to set expectations is highlighted. As is also made clear by the organisation of the table, the main foci of the recommendations concern people and process.

Table Two: Recommendations arising from the joint CREDS-UKERC experience

[20] Research Data Alliance Interest Group: https://www.rd-alliance.org/groups/active-data-management-plans.html

[21] DMP online: https://dmponline.dcc.ac.uk/

[22] FAIR_IMPACT EU project: https://fair-impact.eu/2nd-open-call-support-opens-30-august

People	Process
• The data sharing experience of others is important – we need to secure better access to data for sharing, and make the methods, tools and **guidance** on how to do this more available. Energy consortia have a role in setting expectations and developing/pointing to resources. • Skills and knowledge in the area of data management vary widely across the energy community, partly because of the involvement of so many different domains. **Training** is required to improve researcher awareness of the value of data sharing and to improve their data management skills. Energy consortia can provide/ host this training and have a role to play here, emphasising the domain aspects of data management. • A **peer network** for data managers would be useful to enable sharing of best practice and identify areas to work on together to embed FAIR data and Open Research practices within researcher's activities. Building on existing Energy Consortia collaboration activities, such as the Cross-Consortium Engagement Meeting (CCEM) would get this process started. • Providing metadata and good quality data indicators takes time. Managing data across multiple institutes, ethics teams and collaboration agreements can be complex. The different disciplinary domains common to energy consortia may have different standards that need to be met. All of these require **expertise, attention and resourcing**.	• The energy community is a large producer and user of models in a wide variety of areas and common standards for what to archive to enable FAIR data and reproducibility have not yet been agreed. Such **protocols** would be helpful to discuss. The energy research specific issues for sharing the outputs of energy models should continue to be highlighted. • Creating better data sets requires them to be more **highly valued**. Institutions need to take the value of data more seriously, funding activities effectively, rewarding individuals for taking an active role, and recognising the importance of workload management. Energy consortia should help to set this framework as part of their culture. • **Data Management Plans** are an essential starting point and should be in place at the beginning of projects. However, to make most effective use of them, they should also be flexible, with appropriate mechanisms in place to reflect and learn as change occurs. For large energy research centres in particular, proposals need to budget for a data manager, recognising that this is an important role requiring appropriate remuneration to secure quality personnel. • **Not all data are equal**. The skills of the data manager should provide guidance and support to help discriminate between the value of different data sets and prioritise management effort accordingly.

Drawing on the experiences of CREDS and UKERC, the recommendations coming from the workshop and the discussion above, a set of data sharing expectations, or principles is proposed below. It is hoped that these are flexible enough to cover the wide variety of contexts in energy research but that they also lay the foundation for a shared vision for what data sharing could be in this community. As such, these data management expectations might be included as part of any induction given to new members of energy consortia. It may be that these principles have relevance in an international context too.

Proposed consortia data sharing principles

While some data sharing considerations depend on the type of data being collected, the authors recommend these expectations as a starting point:

- **Assume that data will be shared.** Using the "as open as possible, as closed as necessary" approach sets the expectation that data will be shared unless there is an (articulated) research reason not to do so.
- **Put the data where people will look for it.** Critical mass is important and having data located next to similar data will aid discovery.
- **Encourage conscious data management decisions.** A process that facilitates and records explicit decisions about how the data will be managed and shared, is better than making implicit decisions which may impact on the ability to share data in the future.
- **Ensure the ethics process is supportive of data sharing.** Consider how the ethics process can support future data sharing and what adjustments can be made to enable this at the start of the process.

- **Consider reproducibility and transparency of the research process from the start.** Gathering the provenance of the data process cannot be done effectively in retrospect.

Conclusion

As can be seen from the introduction, data management and sharing policy and principles have been established for decades but, in the energy domain, there is not much research data available, which leads to the question: *Is Energy Research special?*

The authors have identified specific challenges for large multi-institutional, multidisciplinary energy-focussed programmes in setting up data management frameworks that enable effective data sharing. Key to resolving these challenges is a shared understanding of the importance of sharing research data, making the process of doing so clear and ensuring credit is received for this work. As highlighted by the Data Synergy Framework, the final Recommendations and the Data sharing Principles in the paper above, these are socio-technical challenges. Improving data sharing requires changes throughout the research ecosystem; but energy research consortia can support this process and some key principles that could be adopted are suggested.

The data management process is not a perfect solution and can be considered to be bureaucratic, but the conscious decisions that flow from thinking about its key aspects will help to ensure data is generated in a way that enables sharing, where appropriate and possible. Using these techniques CREDS and UKERC have enabled research data to be shared. Setting up a peer network between data managers working with consortia would enable expertise to be shared and provide a stronger platform for feeding back into UKRI policy. This policy is currently being reviewed and advocating for the issues recognised within large consortia will strengthen this process, enabling it to flex to the requirements of all different researcher types.

Reflecting on the challenges energy research brings and recognising that large scale energy consortia are somewhat special and not always aligned to expectations, there exists the capacity to improve data sharing and synergy across the energy research community. At a time when multiple crises look to be addressed by energy research, it can be expected both that the relevance of good data will become increasingly important, making its discoverability ever more vital, and that such research will come under increasing scrutiny, making the quality of the processes governing it essential to get right.

References

Barker, M., Chue Hong, N.P., Katz, D.S. et al. (2022) Introducing the FAIR Principles for research software. Sci Data 9, 622 . https://doi.org/10.1038/s41597-022-01710-x

Colechin, M., and Quigley, C., (2023) Improving Data Sharing in Energy Consortia: Summary of Workshop Outputs. UKERC, https://doi.org/10.5286/UKERC.EDC.000970.

Davenport, H., Grant, J., Jones, CM. (2020) Data Without Software Are Just Numbers DataScience v19(3) DOI: https://doi.org/10.5334/dsj-2020-003

Doonan, A., Akmon, A., Cosby, E., (2020) An Exploratory Analysis of Social Science Graduate Education in Data Management and Data Sharing. International Journal of Digital Curation v.15(1) DOI: https://doi.org/10.2218/ijdc.v15i1.671

EU Open Access and Data management expectations from https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm, retrieved 1 February 2024

EU Open Science policy https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en, retrieved 1 February 2024

Higginson, S., Topouzi, M., Andrade-Cabrera, C., O'Dwyer, C., Darby, S., Finn, D. (2018). Achieving Data Synergy: The Socio-Technical Process of Handling Data. In: Foulds, C., Robison, R. (eds) Advancing Energy Policy. Palgrave Pivot, Cham. https://doi.org/10.1007/978-3-319-99097-2_5

Huebner, G.; Fell, M. (2020). TReQ Tools: How to Improve Transparency, Reproducibility and Quality in Energy Research. figshare. Preprint. https://doi.org/10.6084/m9.figshare.11663466.v1

Huebner, G.M., Fell, M.J., Watson, N.E. (2021) Improving energy research practices: guidance for transparency, reproducibility and quality, Buildings and Cities, Volume: 2 Issue: 1, Page 1–20, DOI: 10.5334/bc.67 https://journal-buildingscities.org/articles/10.5334/bc.67

Huebner, G., Nicolson, M., Fell, M., Shipworth, D., Elam, S., Hamner, C., Kennard, H., Johnson, C. (2017) 'Are we heading towards a replicability crisis in energy efficiency research? A toolkit for improving the quality, transparency and replicability of energy efficiency impact evaluations', in. ECEEE Summer Study 2017, Presqu-Ile de Giens, France.

Jones, C.M., and Higginson, S., (2023) CREDS/UKERC Data sharing workshop briefing note, https://doi.org/10.5286/UKERC.EDC.000971

Jones, S (2012) Developments in Research Funder Data Policy, International Journal of Digital Curation v.7(1) DOI: https://doi.org/10.2218/ijdc.v7i1.219.

Lawrence, B., Jones, C., Matthews, B., Pepler, S., Callaghan, S. (2011) Citation and Peer Review of Data: Moving Towards Formal Data Publication. International Journal of Digital Curation v.6(2) DOI: https://doi.org/10.2218/ijdc.v6i2.205

Matthews, B., Crompton, S., Jones, C., Lambert, S., (2015) Towards the Preservation of the Scientific Memory. International Journal of Digital Curation v.10(1) DOI: https://doi.org/10.2218/ijdc.v10i1.361

Mayo, C. Vision, TJ., Hull, EA., (2016) The location of the citation: changing practices in how publications cite original data in the Dryad Digital Repository. International Journal of Digital Curation v.11(1) DOI: https://doi.org/10.2218/ijdc.v11i1.400

Piwowar HA, Day RS, Fridsma DB (2007) Sharing Detailed Research Data Is Associated with Increased Citation Rate. PLoS ONE 2(3): e308. https://doi.org/10.1371/journal.pone.0000308

RCUK: Guidance on best practice in the management of research data (2015) retrieved from https://www.ukri.org/wp-content/uploads/2020/10/UKRI-020920-GuidanceBestPracticeManagementResearchData.pdf, 1 February 2024

RCUK, HEFCE, Wellcome Trust, Universities UK, Concordat on Open Research Data (2016) retrieved from https://www.ukri.org/wp-content/uploads/2020/10/UKRI-020920-ConcordatonOpenResearchData.pdf, retrieved 1 February 2024

Rozite, V., Miller, J., Oh, S. (2023) Why AI and energy are the new power couple, IEA Commentary, https://www.iea.org/commentaries/why-ai-and-energy-are-the-new-power-couple, 4 February 2024

Smale, N.A.,Unsworth, K., Denyer, G., Magatova, E., Barr, D., (2020) A Review of the History, Advocacy and Efficacy of Data Management Plans International Journal of Digital Curation v.15(1) DOI: https://doi.org/10.2218/ijdc.v15i1.525

Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, et al. (2011) Data Sharing by Scientists: Practices and Perceptions. PLoS ONE 6(6): e21101. https://doi.org/10.1371/journal.pone.0021101

Ward, C., Freiman, L., Jones, S., Molloy, L., Snow, K. (2011) Making Sense: Talking Data Management with Researchers, International Journal of Digital Curation v.6(2) DOI: https://doi.org/10.2218/ijdc.v6i2.202

Wendelborn, C., Anger, M., Schickhardt, C., 2023, What is data ship? Towards a comprehensive understanding, Journal of Biomedical Informatics, V140, https://doi.org/10.1016/j.jbi.2023.104337

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018. https://doi.org/10.1038/sdata.2016.18

UKRI. Publishing your research findings (2023) retrieved from https://www.ukri.org/manage-your-award/publishing-your-research-findings/making-your-research-data-open/ , retrieved 1 February 2024

Acknowledgements

The authors gratefully acknowledge support from:

- Centre for Research into Energy Demand Solutions, funded by the Engineering and Physical Sciences Research Council and the Economic and Social Research Council, grant reference number EP/R 035288/1.
- The UK Energy Research Centre (UKERC) funded by the UK Research and Innovation Energy Programme and Natural Environment Research Council, grant number EP/S029575/1.
- Energy Demand Research Centre (EDRC), funded by the Engineering and Physical Sciences Research Council and the Economic and Social Research Council [grant number EP/Y010078/1].
- Energy Demand Observatory and Laboratory (EDOL) and funded by the Engineering and Physical Sciences Research Council (EPSRC) Grant EP/X00967X/1

We would also like to thank the participants in our workshop for their insights and experience, and the reviewers for their comments.