FAIR Metadata Standards for Low Carbon Energy Research—A Review of Practices and How to Advance

August Wierling [1,*], Valeria Jana Schwanitz [1,2], Sebnem Altinci [3], Maria Bałazińska [4], Michael J. Barber [5], Mehmet Efe Biresselioglu [3], Christopher Burger-Scheidlin [5], Massimo Celino [6], Muhittin Hakan Demir [3], Richard Dennis [7,8], Nicolas Dintzner [9], Adel el Gammal [10], Carlos M. Fernández-Peruchena [11], Winston Gilcrease [12], Paweł Gładysz [13], Carsten Hoyer-Klick [14], Kevin Joshi [15], Mariusz Kruczek [4], David Lacroix [16], Małgorzata Markowska [4], Rafael Mayo-García [17], Robbie Morrison [18], Manfred Paier [5], Giuseppe Peronato [19], Mahendranath Ramakrishnan [20,21], Janeita Reid [22], Alessandro Sciullo [12], Berfu Solak [3], Demet Suna [5], Wolfgang Süß [23], Astrid Unger [5], Maria Luisa Fernandez Vanoni [10] and Nikola Vasiljevic [24]

[1] Department of Environmental Sciences, Western Norway University of Applied Sciences, 6856 Sogndal, Norway; valerias@hvl.no
[2] The Create Centre, The Schumacher Institute, Bristol BS1 6XN, UK
[3] Sustainable Energy Division, Izmir University of Economics, Balçova, Izmir 35330, Turkey; sebnem.altinci@ieu.edu.tr (S.A.); efe.biresselioglu@ieu.edu.tr (M.E.B.); muhittin.demir@ieu.edu.tr (M.H.D.); berfu.solak@ieu.edu.tr (B.S.)
[4] Central Mining Institute, 40-166 Katowice, Poland; mbalazinska@gig.eu (M.B.); mkruczek@gig.eu (M.K.); mmarkowska@gig.eu (M.M.)
[5] AIT Austrian Institute of Technology, 1210 Vienna, Austria; michael.barber@ait.ac.at (M.J.B.); christopher.burger-scheidlin@ait.ac.at (C.B.-S.); manfred.paier@ait.ac.at (M.P.); demet.suna@ait.ac.at (D.S.); astrid.unger@ait.ac.at (A.U.)
[6] ENEA Agenzia Nazionale per le Nuove Tecnologie, l'energia e lo Sviluppo Economico Sostenibile, 00196 Roma, Italy; massimo.celino@enea.it
[7] Copenhagen University Library, University of Copenhagen, 1165 Copenhagen, Denmark; ricd@kb.dk
[8] The Royal Danish Library, 1219 Copenhagen, Denmark
[9] Faculty of Technology, Policy and Management, Delft University of Technology, 2628 ZC Delft, The Netherlands; n.j.r.dintzner@tudelft.nl
[10] European Energy Research Alliance, 1000 Bruxelles, Belgium; a.elgammal@eera-set.eu (A.e.L.); ml.fernandez@eera-set.eu (M.L.F.V.)
[11] CENER National Renewable Energy Center, Sarriguren, 31621 Navarra, Spain; cfernandez@cener.com
[12] Dipartimento di Culture, Politica e Società, University of Turin, 10153 Torino, Italy; gregorywinston.gilcrease@unito.it (W.G.); alessandro.sciullo@unito.it (A.S.)
[13] Faculty of Energy and Fuels, AGH University of Science and Technology, 30-059 Kraków, Poland; pawel.gladysz@agh.edu.pl
[14] German Aerospace Center, Department of Energy Systems Analysis, Institute of Networked Energy Systems, 26129 Oldenburg, Germany; carsten.hoyer-klick@dlr.de
[15] Centre for Urban Science and Engineering, Indian Institute of Technology Bombay, Mumbai 400076, India; kevinjoshi9888@gmail.com
[16] Université de Lorraine, CNRS, LEMTA, F-54000 Nancy, France; david.lacroix@univ-lorraine.fr
[17] CIEMAT Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas, 28040 Madrid, Spain; rafael.mayo@ciemat.es
[18] Independent Researcher, 10627 Berlin, Germany; robbie.morrison@posteo.de
[19] Idiap Research Institute, 1920 Martigny, Switzerland; giuseppe.peronato@idiap.ch
[20] Reiner Lemoine Institut, 12489 Berlin, Germany; mahendranath.ramakrishnan@rl-institut.de
[21] FH Upper Austria, 4600 Wels, Austria
[22] Fraunhofer Institute for Energy Economics and Energy System Technology, University of Kassel, 34109 Kassel, Germany; janeita.reid@student.uni-kassel.de
[23] Karlsruhe Institute of Technology, Institute for Automation and Applied Informatics, 76344 Eggenstein-Leopoldshafen, Germany; wolfgang.suess@kit.edu
[24] Department of Wind Energy, Technical University of Denmark, 4000 Roskilde, Denmark; niva@dtu.dk
* Correspondence: augustw@hvl.no; Tel.: +47-57676240

Abstract: The principles of Findability, Accessibility, Interoperability, and Reusability (FAIR) have been put forward to guide optimal sharing of data. The potential for industrial and social innovation is vast. Domain-specific metadata standards are crucial in this context, but are widely missing in the energy sector. This report provides a collaborative response from the low carbon energy research

community for addressing the necessity of advancing FAIR metadata standards. We review and test existing metadata practices in the domain based on a series of community workshops. We reflect the perspectives of energy data stakeholders. The outcome is reported in terms of challenges and elicits recommendations for advancing FAIR metadata standards in the energy domain across a broad spectrum of stakeholders.

1. Introduction

Metadata is needed for concise knowledge communication of the energy transition to its stakeholders. The concept of metadata lends itself to the idea of maintaining a laboratory notebook: data would represent the outcome of a specific experiment, a physical notebook that contains how and what data was collected from a given project, and a database is the library where all notebooks from various projects are stored. All this data is connected to some context, which is captured by metadata here represented by a card index. Metadata enables navigation within and across databases. In other words, data is inevitably linked to its metadata, and without metadata, data loses usefulness [1]. Hence, a credible and traceable documentation of knowledge about the energy system is not possible without metadata.

Metadata for low carbon energy research is essential as it provides the context for data used to decide on energy transition pathways. Metadata needs to enable reproduction and validation from the research community's perspective, and metadata needs to be comprehensive (trackable) for society. This is an integral part of science. Metadata supports transparency of energy transition processes, the validity of the basis for decision-making, and collaboration across disciplines and societal groups. Metadata, therefore, also serves the purpose of communicating data between its users. As the research and policy questions connected to the low carbon energy transition are interdisciplinary and complex, it is crucial to understand the energy system as a whole and in its parts (refer to Georgescu-Roegen [2] for further reading on the Principle of the Whole). This implies that data describing the historic, present, and future course has to be in a suitable format that allows scaling, analysis, and comparing heterogeneous information with different granularity levels, may differ across user groups, which is why it is essential to identify them beforehand so that the richness and detail of metadata can be defined.

Wilkinson et al. [3] stress that 'Metadata needs to meet domain-specific standards'. However, current metadata practices in the energy domain fall short of these aspirations. There are no data standards to link to units, technologies, methodologies, and infrastructures. This report is a collaborative work and examines this gap in more detail in the following sections. It reviews, explores, and tests metadata practices and related governance principles for the low-carbon energy domain. Building on a series of community workshops and discussions with dozens of researchers from June 2020 to December 2020, we elicit lessons learned and provide recommendations for advancing FAIR (meta)data standards for low carbon energy research. (We use the phrase '(meta)data' as a shorthand for 'metadata and data'.) Thereby FAIR stands for Findability, Accessibility, Interoperability, and Reusability of data [3]. Metadata plays a vital role in all the FAIR guiding principles. Out of the 15 criteria, 13 are explicitly referring to metadata.

While metadata needs to be accessible by anybody, research data in energy science covers a wide range in openness, ranging from open data, to shared data, and closed data. While open data is understood as data 'that can be freely used, re-used and shared by anyone for any purpose' [4], shared data may be widely accessible, but have some conditions about its use. Closed data is limited to a well defined, usually small group of users. Given the commercial interest of energy stakeholders, closed data is not uncommon in the energy domain. Moreover, the General Data Protection Regulation (EU GDPR) may

be required to limit access to the data. While there are good reasons for the protection of commercial interests and private data, the European Union encourages its member states to foster that data should be 'open by design and open by default' [4]. In any case, the FAIR principles comprise the entire range of openness by understanding accessibility as 'accessibility under well defined conditions' and linking re-usability to 'clear and accessible data usage licenses' [3]. Moreover note that metadata may even just play an important internal role for a research team by enabling a proper documentation of the data context.

The report is structured as follows. Section 2 outlines the approach used to engage the low carbon energy research community and other experts. We believe the method to be useful for replication. Section 3 is presenting insights from workshop discussions, structuring them along with reflections on the needs of user groups (Section 3.1), methods on how to advance metadata standards from the perspective of humans and machines, and technical points of departure in the energy domain (Section 3.2). These sections review and evaluate current metadata practices and standards in the field. The report concludes with lessons learned and recommendations on advancing metadata standards for low carbon energy research (Section 4). That section also expands on the role of openness for data.

2. Method to Enable Community-Wide Reflection on Metadata

This report results from discussions within the low carbon energy research community organized through a series of online workshops between June and December 2020. The consortium of the EERAdata project has developed the method for community engagement [5]. This project aims to create and establish a FAIR and open data ecosystem for low carbon energy research, for which the (small) project consortium is 'going where the data are'. EERAdata strives to transfer best practices in FAIRification developed in other communities to the low carbon energy research domain. To this end, a series of six workshops are conducted, following the concept of defining, implementing, and sustaining FAIR practices. The second workshop brought together experts for FAIR and open data with energy domain specialists with the aim of creating a shared understanding of FAIR and open energy (meta)data. This report presents a white paper on the matter. The recommendations will be taken up in the new transversal Joint Program 'Digitalization for Energy' of the European Energy Research Alliance (EERA) [6]. While all workshops were designed to gather and reflect inputs from data scientists and data stewards (from the field and beyond), the December workshops were dedicated to compiling content and preparing a first version of the current paper as a community effort.

Figure 1 presents an overview of the different seminars held. Additional details are also provided in the Supplementary Materials. A community-wide initiative was chosen as a suitable litmus paper for various reasons. Firstly, the advancement of metadata practices that examine existing gaps that hinder low carbon energy research affects the global community. Existing gaps and needs cannot be looked at myopically as a researcher's problem but rather as a global problem. They will continue to negatively affect the FAIR and open data ecosystem in low-carbon energy if urgent steps are not taken to change the current narrative and practices. Secondly, constructive inputs from a wider community open up transparent discussions about metadata and tap into a diversified knowledge base. All broad community stakeholders should form resilient solutions for the current issues that hinder metadata standards. Thirdly, expert practitioners' input enabled non-specialists to understand recent research on specific topics better and generate further interest that may influence the longed-for change.

All workshops were delivered in an online format. Participants could contribute to the working draft of this paper via online collaboration tools that were always available and accessible. Participants were invited to join writing teams and discussion groups to share comments, provide useful research from credible sources, and add their contributions from expert analysis sessions featured throughout the workshops.
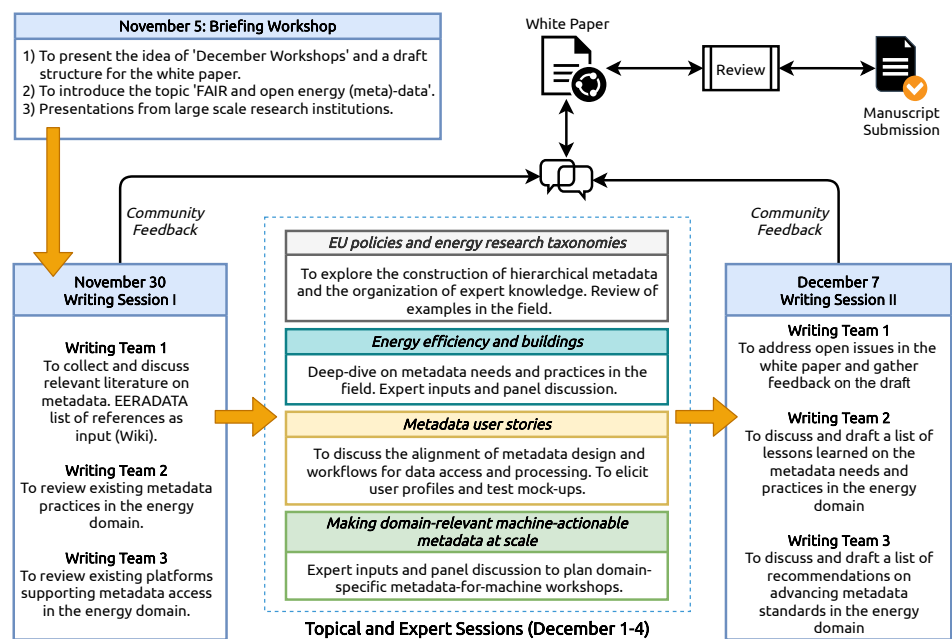
Figure 1. Method for drafting the community report. Details on the design of workshops and links to materials are provided in the Supplementary Materials.

3. Discussion of Results

3.1. Needs and Gaps for Users of Low Carbon Energy (meta)Data

3.1.1. Users of Low Carbon Energy (meta)Data and their Needs

As already discussed, metadata provides data to communicate the nature of the underlying data between data users, including data creators and data re-users. The extent of context needed naturally depends on the needs of the data users. Thus, a definition of the required richness and depth of metadata as well as the conditions for accessing (meta)data is inseparable from a stakeholder and needs analysis. The group of energy metadata users is diverse, ranging from different academic disciplines to the general public. Besides, some stakeholders have been identified from the general point of view of scholarly communication [7], where energy research is one topic from many. In detail, we have identified the following user groups, their use of energy data, and the need for specific metadata to accommodate these needs:

1. **Researchers** —*(a) Energy domain experts* create data from scratch, re-use existing data to curate, aggregate, analyze, and publish data. They come from various disciplines. As re-users of data, they need support from metadata in searching and finding data. They require context information with high granularity and rich provenance to assess data quality and relevance, precise information on usage rights, and intellectual property rights (IPR) requirements. *(b) Interdisciplinary scientists* inform themselves on (other) expert knowledge, re-use data, aggregate and analyze data, and publish data. They require context information for searching and finding data. They need information on the context of data and provenance (mostly on aggregated levels) and precise information on usage rights and IPR. Shared and open data fall within the interest of this group.

2. **Science funders of energy R&D activities** inform themselves of the results of funded research and projects. They need to monitor, adjust, and plan funding policies and principles to better direct R&D investments and to ensure the impact of policy measures [8–10]. Evaluation agencies are a key mediator in this regard as they compile existing (meta)data and develop evaluation tools. Another important objective of funders is to unlock the potential of added value by exploring knowledge generated by research, exploiting and expanding its use [11]. Therefore, funding agencies also require the Data Management Plan (DMP) to track data's provenance and future

maintenance. Typically, this application is at a high level of aggregation, and information should be easy to understand and disseminate. Last but not least, a pivotal interest is that the funding agency be acknowledged in the metadata.

3. **Planners and decision-makers** (incl. energy market regulators, security coordinators, policy-makers) inform themselves on expert knowledge. They may reuse data, analyze some data, and publish aggregated data and decisions. This involves a middle to high level of aggregation, information on the context of data, and provenance. An important aspect is the metadata information on aggregated data. While this group has an interest in open data policies, they are also obliged to protect private data pertaining to citizens.

4. **Energy and other industries** (incl. technical and operational planners and decision-makers, energy market operators). They inform themselves on expert knowledge and re-use and analyze some data on all aggregation levels. They require information on the context and provenance of data. Important is metadata information on aggregated data. To assess the commercialization potential, legally secure information on usage rights and IPR are needed.

5. **General public**. The group informs itself to adjust behavior and practices (e.g., energy consumption behavior, voting in elections, engagement as prosumers, activists or citizen-scientists). They may re-use data to tell others and pose questions leading to research activities. A high level of aggregation is needed to make data easy to understand and navigate.

6. **Data scientists** (incl. data engineers, software and algorithm developers). This group codes, tests, and validates software with existing data. Specifically, data engineers (who may be domain specific or agnostic) are key builders of (meta)data pipelines and back-end solutions. Data scientists may re-use (meta)data to design scientific workflows considering interoperability with other tools and demonstrate applications. They need concise metadata to integrate data sources within the software tool, machine-actionable open file formats, agreed standards, terminologies, and interoperability protocols. Being in legally-compliant control of private data is of particular interest for this user group.

7. **Publishers, librarians, and data curators** publish, store, and archive research data. They may re-use data to link them to metrics such as access statistics and to cross-reference. They need metadata regarding ownership/authorship of results or the size of the data and information such as keywords to provide searchability in their publications and archives.

Note that in practice, a data stakeholder may take two or more roles at the same time. Some institutions may also provide several functions simultaneously by, e.g., bringing together basic researchers, interdisciplinary scientists, public relation specialists, and data custodians.

The data activities and metadata needs identified above can be categorized into broader topics to illustrate the user groups' differences. Typical activities by data stakeholders in the life-cycle of data are: (1) creating and collecting data, (2) actively re-using data (e.g., creating new data or products), (3) passively re-using data (e.g., by distributing the information consumed), (4) accessing information, (5) publishing data, and (6) saving and archiving data (data curation). Note that the order of activities might differ between the user groups depending on the actual workflow. The needs can be organized roughly under the topics: (1) to identify data context and data organization (including standards applied), (2) to retrieve (meta)data, (3) to manage and process information, and (4) to inform on ownership, funding, IPR, commercial protection, and data protection regulations (such as European GDPR). Note that these categories are not entirely exclusive and may support each other in providing the required context. They are similar to some of the purposes of metadata found in Haynes [12].

Applying the categories above, Figure 2 visualizes the extent of activities and needs for each user group. We distinguish between a low engagement (L), an intermediate

level (M), and a pivotal movement (H). As discussed above, the needs of each user group usually come with different levels of granularity. For simplicity, we distinguish only fine (F—symbolized by a circle) and coarse (C—symbolized by a pentagon) levels; when a user group requires specific types of metadata at the different levels, we show it as both (B—symbolized by a diamond). The choices for the values reflect the results of workshop discussions and impressions from the literature review. This figure confirms the variation in typical activities with data, e.g., scientists mainly create research data. In contrast, all other user groups mainly re-use data in analysis, aggregation, and publication. It underlines the role of standards as a basis for mutual understanding between the user groups, accommodating their interests and backgrounds.
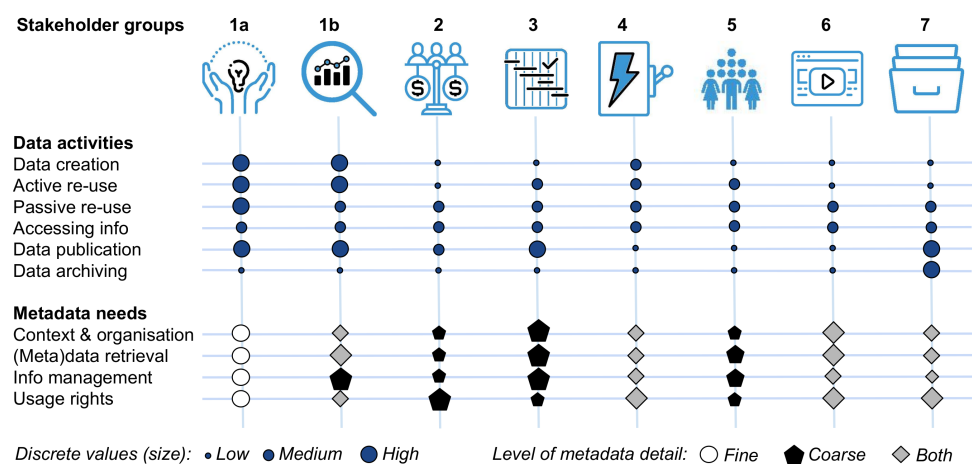
Figure 2. Data activities and metadata needs for each stakeholder group. 1a: Energy domain experts, 1b: Interdisciplinary scientists, 2: Funders of energy R&D activities, 3: Planners & decision-makers, 4: Energy & other industries, 5: General public, 6: Data scientists, and 7: Publishers, librarians, and other data curators.

3.1.2. Gaps for Users of Energy (meta)Data

The section provides a review of gaps between metadata needs and existing (meta)data practices. We consider the user groups identified in the previous section.

Most of the literature is narrated from the perspective of the researcher (domain experts, interdisciplinary, and other scientists). Limitations and gaps related to data and metadata are significant impediments for research activities [13,14]. The primary issues raised pertain to findability and accessibility issues of the FAIR principles [15–17]. For an enriched research cycle, the metadata and data should be descriptive, complete, easy to find, accessible, and machine parsable. Data that lacks descriptive and full metadata reduces the resource's reusability and diminishes its value. Further limitations derive from a lack of completeness (e.g., missing years from time series, incomplete sampling, etc.), lack of a user-friendly database, and (meta)data formats not being machine-actionable. The data itself should fit with the domain norms regarding semantics and packaging. In addition to this, metadata should, where possible, be provided in a Linked Data or RDF format, such as JSON-LD.

The factors mentioned above decrease the ability to locate data while at the same time disrupting the pace of research, slowing adoption and use by other users [18,19]. Metadata and data should be registered or indexed in searchable sources that academics and other data stakeholders can easily access. Although a significant amount of data is continuously generated, only a small proportion is necessarily available in academic databases that are easy to find and access [20]. In this sense, the lack of an appropriate format to store the data is a significant barrier since a considerable amount of information is not deposited in sustainable and widely-known repositories [15]. Moreover, the available data is not always linked to metadata that conforms to any metadata schema.

From a theoretical perspective, particularly in qualitative research, data is prone to inaccuracies, omissions, and inconsistencies [21]. This weakens the reuseability of the data. Interdisciplinary research may be affected by a broader set of data and metadata-related challenges. For example, the interoperability criteria of the FAIR guiding principles has crucial importance and its lack means that essential information is merely absent from specific datasets, making academic research needlessly time-consuming to process multiple datasets, costly, labor-intensive, and error-prone when processing multiple datasets [22]. A difficulty for processing data in different academic fields is linked to the following deficiencies: (1) lack of controlled vocabularies and terminologies, (2) lack of agreements on controlled vocabularies and terminologies, (3) lack of agreement on formats in which controlled vocabularies and terminologies are expressed, (4) lack of linking between existing controlled vocabularies, (5) lack of agreement on formats for metadata, and (6) lack of agreement on metadata schemas.

Specific gaps faced by funders of energy R&D activities (incl. evaluation agencies) are connected with their need to monitor and evaluate the efficiency of the provided funds. Therefore, access to (meta)data, the richness of documentation of activities and results, as well as the interoperability are current key issues. For example, the literature points to the significance of research output in showing which funds are appropriately utilized to enhance socioeconomic progress and to assess the impact of policy support measures [8,9]. While some evaluation tools are available [23], their proper application suffers from a lack of coverage, appropriate levels of detail, as well as missing interlinkages between the deliverables of research projects.

Data and metadata are significant ingredients and drivers of the policy process. Evidence from the literature highlights that data sharing is essential for policymakers to carry out necessary activities spanning planning, monitoring, decision-making, and implementation with accurate scientific data [16]. In this context, planners and policymakers—as well as analysts preparing policy-relevant reports—benefit from shared and/or open data platforms. Simultaneously, it is imperative to transparency that policymakers should share data openly to achieve successful cooperation between academic and policy-making levels. Transparency is essential to gain social trust and engage the public in the policy-making process [19,24]. However, as the literature review shows, governance problems or insufficient skills can pose a problem for the data's openness [25]. Moreover, one main category of policymakers' issues in managing data has been identified as legal problems [24]. Ethical and security concerns and additional workload are essential reasons for the difficulty in achieving openness—even in metadata—at the policy-making level [26]. As policymakers' decisions affect the public and industry significantly, they should engage private sector representatives and citizens in the decision-making processes, and the data generated should be accessible at every level [21]. Another related topic that has received interest from policy data stakeholders is the accessibility and openness of government-run sites being deemed essential to increase communication between municipalities and citizens [17].

The energy and other industries is another main user group of low carbon energy (meta)data, especially against the background of the increasing digitization of the energy sector. High quality metadata is a prerequisite to new business concepts such as smart demand response, the operation of small-scale distributed renewables, smart charging of electric vehicles, and the development and management of energy storage solutions. The literature stresses the general lack of common standards for sharing (meta)data in the industry [18,20,22]. Without such standards, data is lost and interoperability is limited if not disabled, causing forgone opportunities for performance optimization [27], demand forecasting [28], the monitoring of energy use [29], and the creation of energy management frameworks [27]. The matter is even more pressing when aiming at a global adoption of circular economy concepts, in which the flows of resources, waste, and—consequently —data produced by different participants of an industrial symbioses need to be integrated. The coordination of such a deeply linked industrial infrastructure requires collaboration

based on FAIR (meta)data. Finally, unresolved issues around the sharing of (meta)data and intellectual property rights need to be settled.

Traditionally, the general public requires (meta)data that can be trusted to judge the accountability and transparency of governmental measures [30–32]. Cantadora et al. [33] emphasize the general public's data needs as a prerequisite for citizen participation in policymaking or community initiatives. The authors discuss the OECD's Open Government Data project that relies on making the governments' data available to the general public for transparency and accountability.

Of increasing concern is the protection of privacy rights in the advent of large-scale metering and the collection of consumer information. Currently, (meta)data are not designed to prevent the inference of sensitive personal information nor do they support the regulation of discrimination practises based on customer segmentation [34]. The authors also point out that the need for fine-grained data that allows the tracking of choices of individual household members is at odds with the requirement for data aggregation due to data protection. Moreover, with the application of aggregation procedures, the task of following the flow of data becomes demanding.

With the opening of energy markets for new actors, the general public's role broadens from a passive participant (who receives information and takes IT-informed consumption choices) to active engagement in the energy transition. Gendron and Killian [35] refer to the concepts of a 'data citizen' and 'information democratization', whereby they argue that citizens lack knowledge on concepts like data agility and data literacy. For example, Anhalt-Depies et al. [36] discuss knowledge gaps around privacy rights when participating in citizen-science activities. The empowerment of citizens regarding FAIR (meta)data knowledge and competences is also important for steering interest in the general public to take up new professions, such as energy data stewards and data analysts. Lhoste [37] investigates possibilities of engaging civil society in R&D with the help of specifically tailored research systems. Dunnett et al. [38] is a best practice on harmonizing data for solar and wind farm locations by sourcing information from the collaborative OpenStreetMap project.

The specific gaps data scientists are confronted with are presented in regular surveys conducted by Appen, an artificial intelligence (AI) platform. Their annual Data Scientist Report presents the results of interviews with more than 100 participants on data management practices. The larger share of participants represents technology companies with sizes in a spectrum of 100 to more than 10,000 employees. The results from 2017 [39] show that data scientists spend most of their time (51%) on data curation, i.e., the collecting, labeling, cleaning, and organizing of data. At the same time, only 19% is devoted to data analysis and modeling. This imbalance could be rectified with the help of FAIR (meta)data. 51% of the respondents characterize the data they work on as unstructured data, pointing to the need of accommodating heterogeneous data. Regarding the type of data, 91% of the data scientists use is text data, 33% image data, 11% audio data, and 20% video data showing that (meta)data have to comply with various multimedia standards. While several approaches are applied occasionally, agreements on selected standards are missing. A subsequent Appen White Paper [40] highlights several trends. The first one is the increasing number of AI projects and the growing need for high quality data in parallel. Also, the growing significance of data privacy and related ethical issues is even more paramount because data-driven AI is now increasingly applied in decision-making. The 2018 survey also reports on the perception of algorithmic bias. Indeed, 9% of data scientists believe that there is no algorithmic bias, whereas 75% think there is less algorithmic bias than human (data scientist) bias. This result may support the machine actioning of the data while there is at the same time a danger of amplifying bias coming from training sets that are not representative or of poor quality [41]. Rich metadata is the way forward to reveal and track potential data quality issues.

A detailed compilation of gaps from the point of view of publishers, librarians, and data curators can be found in Gregg et al. [7]. Summarizing their findings, publishers are faced with the transition from traditional cataloging and metadata compilation to more

automated forms. Also, proprietary concerns may lead to the in-house development of metadata concepts. Librarians are engaged in creating metadata, consuming metadata, and preserving metadata at the same time. As consumers, they rely on researchers and publishers' high-quality metadata and fail if gaps are encountered from that side. Bascones and Staniforth [42] even state that "due to poor quality metadata supply combined with complex and continuous processing, librarians rarely know the status of their library holdings and collections of e-resources," affecting user experience, user satisfaction, and cost calculations. Of course, poor quality metadata should be corrected rather than preserved, but procedures, agreements, and guidelines are lacking. Finally, data curators are faced with the wish to assign persistent identifiers to highly granular data against the background of limited financial and administrative resources. Indeed, the lack of acknowledgment for data curation and the allocation of corresponding resources is a gap that can be observed for all energy data stakeholders.

3.2. Advancing Metadata Practices—Practical Explorations

3.2.1. Human Perspective on Metadata Navigation

When knowledge was codified in traditional encyclopedias, users oriented themselves within that given framework of knowledge organization. However, with the growing amount of experience and increased fragmentation, new ways of organizing and codifying knowledge are needed to enable users to search, understand, and publish data. To tap that potential, metadata must be appropriate and well-structured, according to how specific users view the domain's knowledge. That may differ across groups of users and within a group, e.g., acknowledging different schools of thought or disciplinary differences.

The user groups of low carbon energy data discussed in the preceding section present different views on the energy domain, leading to how knowledge is encoded in general and specifically through metadata. Ontologies and taxonomies are an essential tool to tell what the data is about. They organize knowledge with the help of concepts and relationships between them, allowing researchers to navigate the information and to capture the context of data. Ontologies are understood here following Studer et al. [43] as a "formal, explicit specification(s) of a shared conceptualization." Note that ontologies are shared between users in the communication of data. Having many stakeholders with various needs and interests may lead to fragmentation into several niche ontologies and the need for standardization and integration to come to a mutual understanding. Taxonomies are a specific form of ontology where all concepts are arranged hierarchically as sub- and super-concepts. Ontologies allow for richer knowledge representation, including different types of generic classes (e.g., materials, buildings, renewable energy generators), which are linked by multiple relations between them (e.g., steel is used to build wind turbines) [44,45]. Ontologies support the communication between humans, between machines, as well as between humans and machines. However, while humans typically have a predefined set of concepts, e.g., from being educated in energy science, devices need to learn about concepts in the form of agreed-upon, persistent labels.

In the energy domain, no unified approach to manage the domain knowledge exists. Instead, several ontological schemes have been developed, leading to numerous 'ontology islands' that only partially match each other. Consequently, metadata from different ontologies are primarily unrelated, potentially inconsistent, and not harmonized, preventing the navigation and reuse of heterogeneous data. In view of these difficulties, the idea of multiple application-specific metadata schemes is gaining ground. Table 1 presents a list of ontologies that have been developed within the energy domain.

Table 1. Ontologies identified in the energy domain.

Name	Field of Application	Reference
BOnSAI	Smart building ontology for ambient intelligence	[46,47]
Building Integrated PV	Urban energy planning for building integrated PV	[48]
DEHEMS	Energy consumption behavior of households	[49]
Electr. Markets Ontology	Supporting agent-based simulations	[50]
EnArgus Ontology	Reporting on energy projects	[51]
EnergyUse ontology	Collaborative web-based framework	[52]
DEFRAM	Energy efficiency measures	[53]
OEMA ontology network	Energy management ontology	[54]
OEO	Open energy ontology by the energy modeling community	[55]
OntoPowSys	Power system ontology for eco industrial parks	[56]
OntoWind	Wind Energy Ontology	[57]
ProSGV3	Smart energy consumption	[58]
SAREF4ENER	Smart household appliances and flexibility	[59]
SARGON	Intelligent energy domain ontology	[60]
SEAS Ontology	Generalisation of the Semantic Sensor Ontology	[61]
SEMANCO ontology	Low carbon infrastructure urban planning	[62]
SUMO	Upper foundation ontology owned by IEEE	[63]
ThinkHome	Energy Resources Ontology	[47,64]
WETAXTOPICS	Wind Energy Taxonomy of Topics	[65,66]
WEP	Wind Energy Parameters	[67]

To explore ontologies' role in constructing hierarchical metadata and to brainstorm on solutions for overcoming the problems of ontology islands, a practical exercise was undertaken in the workshop session 'EU policy and energy research taxonomies' (see Supplementary Materials). Participants explored existing ontology islands in the energy domain to discover knowledge organization conflicts between given ontologies and their understandings. Second, participants were asked to identify options for linking ontology islands, attempting to unify the subdomains into a consistent whole. The practical exercise focused on the European Union's ontology frameworks and researcher-based ontologies for representing knowledge in the energy system with different granularity levels. One important lesson learned from the workshop discussions is that matching different ontologies is laborious. Therefore, preference should be given to well-established concepts over less established ones, e.g., backed-up by a larger institution or community-agreed standards.

Participants in the workshop agreed overwhelmingly that it is relatively straightforward for specialists in a sub-domain to develop standard terminology and its codification by an ontology with high granularity. The difficulty lies in defining sub-domain boundaries. The problem is to keep a specific level of detail for terms and concepts (tailored to user needs) while matching them at the same time with those that exist outside. The problem is a lack of agreements on terminology, concepts, and even philosophical approaches. According to Collier and Mahon [68] and Sartori [69], empirical and theoretical concepts should be delineated to avoid (or at least be aware of) conceptual stretching that is to "broaden the meaning—and thereby the range of application—of the conceptualizations at hand ... i.e., to vague, amorphous conceptualizations [with the net result that . . .] our gains in extensional coverage tend to be matched by losses in connotative precision. It appears that we can cover more only by saying less, and by saying less in a far less precise manner" [69]. An example from the workshop is the term 'asset', which is understood and operationalized differently across disciplines.

Ontology islands need to be overcome to enable integrated knowledge frameworks and the coexistence of different ideas seems contradictory at first sight. However, ontologies are not limited to 1:1 relations because they also allow 'is similar to' or 'is connected to' connections. The problem also concerns the impossibility of replicating the complexity of reality—the number of links of one subdomain to others is simply too large. Even aligning

a few sub-domains is beyond manual editing. However, algorithms can navigate vast amounts of relations as long as they are semantically linked.

First attempts have been implemented to generate and align ontologies based on, e.g., automated natural-language processing [70]. Therefore, an automatically generated ontology was among those studied in the workshop, namely EuroSciVoc. This vocabulary organizes scientific topics that were extracted from the CORDIS database on EU projects. Analyzing energy-related entries, the classification fails in reproducing sufficiently the understanding of experts and links between subjects. For example, one would expect 'energy crops' as an entry within agricultural sciences or 'energy history' being part of humanities/history.

As knowledge about any subject continually changes with time, the knowledge representation should capture these changes. An example is the rising importance of new energy technologies. The dynamic update does not only concern the content of some metadata but requires review, creation, and deletion of metadata terms over time. Therefore, implementing metadata with the help of ontologies means that specific concepts and relations have to be regularly reconciled with reality for timeliness and relevance. However, how best to enable dynamic and automated updates of ideas and connections is still a work in progress and was seen as a significant issue by the workshop participants. Approaches from the 'Continuous capture of metadata' project [71] may establish a support infrastructure. AI should be backing the process since the amount of information involved is impossible to screen by humans.

During the discussions, it became evident that selective knowledge representation leads to a bias. As with data itself, the choice of relevant metadata may be in danger of prejudice, particularly selection bias [72]. To some extent, this is natural since populating an ontology is usually not done as a randomized process but involves human choice based on prescribed knowledge [73]. The sequential character of updating the ontology may lead to self-amplified effects resulting in substantial deviations from reality. Transferring this to automated procedures, e.g., machine learning, may even increase this danger if the training data is not carefully screened for bias [41]. To avoid selection bias, a systematic review of the ontologies by experts and all relevant stakeholders should be undertaken. To this end, open metadata is essential. However, as the evaluation of metadata practices reveals, it is not uncommon that also metadata is closed behind paywalls (Section 3.2.3). Also, aligning newly created ontologies to existing ones may help to minimize bias. In this regard, the initiative for an Open Energy Ontology (OEO) is an important step forward. It supports co-editing, and hence, a pluralistic view on critical terms and concepts, minimizing possible bias. Its design process is documented on GitHub, and agreements are reached by consensus at open developer meetings [55].

3.2.2. Machine Perspective on Metadata Navigation

Recently, more and more routine tasks are automated due to the advances in software engineering. This also applies to many of the activities considered in the previous sections. Data search and literature review now rely extensively on algorithms. Humans are assisted if not replaced by algorithms with their own needs for clarity of communication regarding the exchange and re-use of (meta)data. Typically, strict enforcement of standards, terminologies, and protocols is needed to implement persistent and unique identifiers. Only then will it be possible to interlink, analyze, and aggregate data at the massive scale required in tomorrow's machine-based energy infrastructure.

Up to this point, we did not elaborate much on the importance of enabling the navigation of data and metadata from the machines' perspective. The rationale for allowing machine-actionability is that the amount and complexity of energy data have surpassed the threshold of manageability by humans. Machine support is needed for finding, organizing, and analyzing data. Indeed, a rapidly increasing number of datasets in the energy domain qualify as 'big data' in terms of volume, velocity, variety, and veracity. The trend will intensify with the continued increase of digitization of the energy sector and beyond.

An illustrative example is the amount of data generated from household-level electrical load measurements, which easily exceeds one billion data readings when monitoring the inhabitants of a medium-sized European city over a period of two years [74].

An illustration of the new way of integrating (meta)data is the following. Energy grids can be connected to devices, such as IoT sensors, that collect significant amounts of metadata and place them in real-time storage [75]. This is valuable for grid operators to better manage operations and energy suppliers to assist them in smartly adjusting energy supply with demand. Metadata through machine-to-machine data exchange and advanced IoT sensors can help make more accurate weather and load predictions, which is key to ensuring effective renewable energy integration. The combination of electrical and IoT infrastructures metadata with AI is the fundamental aspect of cloud-based energy management systems [76]. The participation of end-users is critical for the energy cloud as it connects and coordinates excess electricity with different end-users. This AI coordination can modify peak storage capacity based on energy generation and consumption [28]. While the grid of end-users is connected to a broader public distribution network [77], the exchange of information between various automation devices is realized by interconnecting communication networks [78].

When done correctly, machine-actionability enhances the communication of metadata and supports the efficient processing and integration of data from multiple sources. The FAIR principles have already been designed with machine-actionability in mind. Indeed, the FAIR principles are developed with idea of serving machines first, and then humans with (meta)data. The systematic assigning of persistent identifiers to metadata and the linking of metadata is thereby central. For this purpose, well-established semantic web technologies exist (e.g., RDF—Resource Description Framework, OWL—Web Ontology Language, and LOD—Linked Open Data). However, the use of these technologies is demanding for researchers outside of the IT community, exceeding the researcher's capacities (in terms of time and skills). Consequently, activities for machine-actionability need funding, training, guidance by specialists, and supporting tools. In the end, implementing machine-actionability has to be enabled by involving the different (human) user groups of the domain-specific data. The process of implementing machine-actionability is still an ongoing discussion with limited experience gathered so far in the energy domain. However, templates, software tools, and organizational support exist, ready to be customized for and with the energy research community. For example, linked data and RDF are well known and established, but simple tools are missing, so that anyone can leverage on these technologies. A panel discussion as part of the International FAIR Convergence Symposium 2020 was organized to start the community's conversation. Led by the initiative GO-Fair [79], so-called Metadata-for-Machines (M4M) workshops for the energy community were discussed, whereby input was provided from experiences gained in other organizations. M4M workshops intend to kick-start FAIRfication efforts in a community, generating minimally viable metadata components that are modular, reusable, and extended later on. The process starts with the community experts that develop a set of informal terminologies. Then domain experts, with the help of data stewards, go through a formalization process by using metadata templates to describe both the structure and the semantics of metadata. Moreover, care is taken that the results are compliant with FAIR principles. Online tools such as the CEDAR workbench [80] are available to make the metadata machine-actionable (i.e., the tools convert the templates into semantic components), while sheet2rdf [81] and excel2rdf [82] are available for simple creation of machine-actionable controlled vocabularies. In one of the first M4M workshops related to the energy domain [83], the wind energy community created and employed machine-actionable metadata templates for datasets [84]. A light weight yet comprehensive metadata specification, such as the one proposed by the Frictionlessdata Framework, has also been adopted by energy-related projects [85].

The process of documenting the implementation of FAIR principles in a community is known as the 'FAIR Implementation Profile (FIP)' [86]. In summary, the FIP is a collection of FAIR implementation choices made by a specific community for the FAIR principles. Once

realized, FIPs provide guidance and orientation on FAIR infrastructures for different data communities. Overall, the above process of implementing machine-actionable metadata is echoed by the EERAdata project (data stakeholder involvement—linking to IT experts —developing and agreeing on common standards—FAIRification through support tools). The panel discussion also revealed essential insights for the planning of energy-specific M4M workshops. Most importantly, the community's broad involvement is crucial to ensure that the design of metadata templates reflects energy science's interdisciplinary character.

The metadata repositories resulting from FIPs are known as the network of FAIR Data Points. They provide software that allows data owners to expose their data's metadata to the public for seamless access by humans and machines alike. When created by trusted communities, other communities' reuse and uptake can accelerate the overall implementation and convergence of the FAIRification of (meta)data.

3.2.3. Advancing Metadata Practices—Technical Aspects

We expand on the practical approach towards defining metadata for low carbon energy research by focusing on the functions that metadata should serve different groups of users as discussed in Section 3.1. Typically, a range of standards is implemented to provide these functionalities. The various standards to deliver the desired functionality are denoted as a metadata concept—we call implementing such a vision a metadata practice. To understand the extent of performance, we analyze existing metadata practices in the low carbon energy research community, applying a set of the following evaluation criteria:

- **Richness** relates to the extent and completeness of metadata, supporting the usefulness of metadata. For example, only 12 identifiers for administrative data is used in the repository 'figshare', whereas 100 identifiers are requested for the 'dataverse'.
- **Consensus** concerns the level of agreement in the community and through the institutional backup. For example, a consortium working on the concept has a more extensive backup than an individual, supporting metadata's credibility. Alternatively, a community may have formally approved a practice, or the practice may even have been adopted by standard-setting institutions such as CODATA. The continuous curation further increases consensus.
- **Accessibility and transparency** concern the ease of using metadata for different stakeholders. This criterion supports usefulness and findability, answering the following questions: Does metadata support the empowerment of many users, e.g., by providing information on varying levels of granularity? How good is the quality of data documentation? Are open standards supported? Specifically for machines: Are persistent and unique identifiers used? Are they retrievable by standard search engines?
- **Linked metadata** supports compatibility and interoperability between data. This criterion concerns how well metadata is interlinked, structured, and grounded, allowing navigation at different levels.
- **Functional implementation** should be up to standard, supporting authorization and authentication. Recommended are the ubiquitous use of persistent identifiers, a high degree of modularity, and the inclusion of licensing information. JSON-LD and XML-RDF are recommended metadata formats.

The above criteria to evaluate metadata practices were derived from reviews and discussions in the workshops (Section 2). It is important to note that equal emphasis has been put on assessing human-readability and machine-actionability of metadata. The requirements were tested against the extent to which they align with standards proposed in the literature [12,87–95]. The Supplementary Materials offers additional details on the development of the evaluation criteria as well as the results from the assessment of metadata practices within the community (Supplementary Materials, Table S1).

The workshop discussions led to identifying several metadata practices and platform developments in the energy domain, most of them being in-house developments. Twelve of the more mature metadata practices were chosen for assessment. These are (1) metadata strings of the Open Energy Platform, (2) the Standard International Energy Product Classi-

fication, (3) the Energy Identification Code scheme (ENTSO-E), (4) the CityGML Energy Application Domain Extension, (5) the Common Information Model by the International Electrotechnical Commission, (6–11) metadata models of the projects Hotmaps, EMODnet, ShareWind, PV-GIS, NOMAD, SEMANCO, and (12) the Materials Experiment and Analysis Database. The reader should note that the assessment of the metadata practices is not meant to scrutinize these pioneering efforts. Instead, the evaluation is used to analyze the current state of the art and to identify common prevailing issues.

We find a large diversity in compliance by applying the evaluation criteria for metadata practice as presented in the list above. Many of the rules show a high granularity in the field of application (standards on richness). Still, they do not link systematically to areas outside of their scope (criteria on linked metadata). This limits the possibility of traversing the metadata for broader reuse. Besides, the existence of these metadata standards is usually unknown outside of the particular sub-domain. We also found that some of the metadata practices are already supported by standard-setting organizations. An example is the Common Information Model by the International Electrotechnical Commission. However, this example shows at the same time that the metadata information is only available under commercial licensing and represents a major impediment, preventing the possibility of critically scrutinizing and co-editing the standard (accessibility and transparency criteria). On the opposite end of the spectrum lies the Open Energy Platform, encouraging broad community involvement. The initiative is currently the broadest and most comprehensive approach towards metadata in the energy domain. The biggest challenge for all metadata practices in the energy domain is, however, machine-actionability. Many of the current metadata practices are limited to documentation for humans in the form of, e.g., widely-used spread sheets and PDF files. A proper functional implementation does not require the use of semantic web technologies but would nonetheless benefit from these technologies. For further details, see also Table S1 in the Supplementary Materials.

4. Conclusions

4.1. Lessons Learned

A common ground for metadata is essential for communication and validation of results in the energy research domain. The importance of standards will increase with the sustained digitization of energy sector, creating the demand for both human-readability and machine-actionability of data. The stakeholders of energy (meta)data form a diverse group who uses data differently and have specific needs with varying expectations on the coverage, granularity, and accessibility of (meta)data. Also, effective communication of data within and between user groups requires domain-specific metadata, emphasizing the need for agreed-upon standards and user group specific implementations. Multiple conceptual approaches and metadata practices exist. Many of them are, however, islanded solutions and connecting them is a big challenge. Moreover, the domain-specific translation of the FAIR data guiding principles is a new research area and few mature best practices exist. Science has reached a threshold where data curation is becoming so resource-intensive that an individual researcher becomes overburdened with this task, rendering the organization of a community-wide process to FAIRify domain (meta)data nothing but wishful thinking. The workshop discussions and literature review revealed several specific challenges, which we highlight below. The last section expands the conclusions of this community white paper with the following recommendations.

- **Reaching agreements on common standards** requires a synthesis of established terminologies, vocabularies, and data formats to create rich and linked, but still comprehensible, metadata that caters to the needs of all stakeholder groups. Metadata concepts to track the domain- and user-specific workflows are still under development. Practices on licensing and access control are uncommon. While the functional implementation should adhere to emerging solutions found in the ICT community (e.g., the widespread use of persistent identifiers, semantic web technology), they need to tie with established practices in the energy domain and match existing IT competencies.

- **Limited functionality of supporting tools and infrastructure**. If an energy research team aims at tackling the problem of interoperability and machine-actionability of their (meta)data, very limited or unspecific top-down support is available. For example, one of the main projects of the European Cloud—EUDAT Collaborative Data Infrastructure—offers support for integrated data services. However, none of the seven core communities covers low carbon energy research, and none of the use cases of the European Open Science Cloud includes best practice examples from the field. (Compare with the information about the core communities at https://eudat.eu/news/eudat-keeps-engaging-communities-all-you-need-to-know-on-the-eudat-data-pilots, (accessed on 1 September 2021)). At the same time guidance is needed to address the lack of resources and skills. In consequence, the reluctance to commit to engage with new data curation technologies appears well founded.
- **Durable solutions** for the energy domain and beyond. Since the context of data is given by the current domain knowledge, metadata is inevitably dynamic in nature. Therefore, metadata needs to be updated to align with the current state of knowledge. Moreover, the relevance and appropriateness of data descriptions has to be regularly verified. No practices are established yet to enable dynamic and machine-supported updating.
- **Data curation literacy across user groups**. The lack of FAIR data curation skills is an issue for all stakeholders of (low carbon) energy data. Indeed, the level of competence ranges from IT-enthusiasts to technology skeptics. Also, little experience exists to map specialist-terminology with non-specialist terminology to facilitate the broad reuse of (meta)data.
- **Organizing a community approach**. For advancing (meta)data practices, a framework able to address and involve all stakeholders is needed. A suitable framework should have a high visibility in the community to ensure legitimacy and to promote the uptake of standards. The challenge is that the implementation of such a framework is beyond the possibilities of a few funded projects, because it requires a long-term and consistent commitment.
- **A challenge to fund data curation activities**. Data curation as well as skill building is resource-intensive for individual researchers and institutions alike, but available funding is clearly not matching the needs.
- **Lack of acknowledgement for data curators**. Data curation is not acknowledged in the same way as data processing and analysis. Despite its relevance, this task is largely seen as inferior. Even more, contributions by data curators are generally less visible (including suitable journals to publish on the matter).

4.2. Recommendations

Advancing FAIR (meta)data practices for low carbon energy research data supports the energy transition and enables the sector to tap the potential of the knowledge economy. However, the explorations in this community report reveal a number of challenges. We address these with the following recommendations:

1. **Augment and synthesize existing metadata practices in the energy domain**. The starting point to reach agreements on common standards is the identification, extension, and integration of existing approaches. This includes the search for best practices in other domains. In this way, domain-specific attempts can be matched with emerging solutions. At the same time, researchers are picked up where they are, and their contributions utilized. With this white paper, we invite interested researchers and existing initiatives in the field to organize a joint conference with the Open Energy Modelling Community, the Wind Energy Community, and the EERAdata project, among others.

2. **Establish platforms that gather and document domain-specific metadata practices**. The recommendation is to build a knowledge base with supporting tools, guidelines, best practices, FAIR implementation choices, and links to (meta)data repositories. The

hub should be anchored in a number of widely frequented energy community websites. The platform should also offer contacts to domain experts and other FAIR data curators (e.g., linking energy-related working groups of top-down initiatives). This promotes the use of general metadata standards such as Dublin Core and also supports a energy-specific implementation of FAIR (meta)data principles that aligns with the general trends. The FAIR Implementation Profiles [86] can serve as an inspiration.

3. **Build a FAIR energy data network and foster long-term commitment**. A starting point is to utilize existing international and EU-level networks in the field. The idea is to link up with periodic conferences and workshops, which are mainstream meeting places for the energy community already. Established international umbrella organizations that target the wide range of energy data stakeholders from researchers to policy-makers, industries, funders, publishers, and so forth, appear as a good choice. One candidate is the European Energy Research Alliance (EERA), working at the interface of European policy-making, national governments, the research community and industry. Members of EERA engage in a broad range of research and cross-cutting activities facilitated in 18 Joint Programs. In particular, the recently established transversal program 'Digitalization of Energy' may serve as a suitable means to kick-off community activities.

4. **Develop a multi-year work program** actively engaging low carbon energy data stakeholders to (1) estimate the resources needed to implement and sustain FAIR data activities from the perspective of user groups, (2) coordinate the development of community practices and consensus on standards, (3) organize workshops to advance unresolved issues (incl. the open/closed data spectrum, incentive problem of data curation, business opportunities, data protection and GDPR, data licensing, and so forth), and (4) promote the building of FAIR data competences and skills.

5. **Advocate for broadening funding programs and encourage industry funding**. The lack of funding for FAIR data activities is a major barrier to advance community practices, build the necessary skills, and increase awareness for the need for data curation. It is important to actively approach funders to motivate the funding of on-the-job training and the development of new university education on energy data stewardship for instance. The industry should also be targeted, and public-private partnerships seem a way forward in this respect.

We conclude this community report on the advancement of (meta)data practices for low carbon energy research with a quote by Scott [96] that "Metadata is a love note to the future". In the same spirit, we hope to inspire the readers. It is the hour for pioneers to discover largely uncharted territory.

References

1.	Kahn, R.; Wilensky, R. A Framework for Distributed Digital Object Services. Available online: http://www.cnri.reston.va.us/k-w.html (accessed on 1 December 2020).
2.	Georgescu-Roegen, N. *The Entropy Law and the Economic Process*; Harvard University Press: Cambridge, MA, USA, 1971.
3.	Wilkinson, M.; Dumontier, M.; Aalbersberg, I.; Appleton, G.; Axton, M.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L.B.; Bourne, P.E.; Bouwman, J.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 160018. [CrossRef] [PubMed]
4.	European Union. Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information—PE/28/2019/REV/1. *Off. J. Eur. Union* **2019**, *L 172*, 56–83.
5.	Schwanitz, V.J.; Wierling, A.; Paier, M.; Biresselioglu, M.E.; Celino, M.; Bałazińska, M.; Fernandez Vanoni, M.L. Towards a FAIR and Open Data Ecosystem in the Low-Carbon Energy Research Community (eeradata), Project Funded by the European Union's Horizon 2020 Programme. Available online: https://www.eeradata.eu/ (accessed on 17 February 2021).
6.	Transversal Joint Program 'Digitalization for Energy' of the European Energy Research Alliance (EERA). Available online: https://www.eera-set.eu/component/projects/projects.html?id=183 (accessed on 17 February 2021).
7.	Gregg, W.J.; Erdmann, C.; Paglione, L.A.D.; Schneider, J.; Dean, C. A literature review of scholarly communications metadata. *Res. Ideas Outcomes* **2019**, *5*, e38698.
8.	Lane, J.; Bertuzzi, S. Measuring the Results of Science Investments. *Science* **2011**, *331*, 678–680. [CrossRef] [PubMed]
9.	Bloch, C.; Sørensen, M.P.; Graversen, E.K.; Schneider, J.W.; Kalpazidou, E.; Aagaard, S.K.; Mejlgaard, N. Developing a methodology to assess the impact of research grant funding: A mixed methods approach. *Eval. Prog. Plan.* **2014**, *43*, 105–117.
10.	Álvarez-Bornstein, B.; Bordonsa, M. Is funding related to higher research impact? Exploring its relationship and the mediating role of collaboration in several disciplines. *J. Inform.* **2021**, *15*, 101102.
11.	O'Kane, C.; Zhang, J.A.; Cunningham, J.A.; Dooley, L. Value capture mechanisms in publicly funded research. *Ind. Market. Manag.* **2020**, *90*, 400–416.
12.	Haynes, D. *Metadata for Information Management and Retrieval: Understanding Metadata and Its Use*, 2nd ed.; Facet Publishing: London, UK, 2018.
13.	Rumble, J.; Broome, J.; Hodson, S. Building an International Consensus on Multi-Disciplinary Metadata Standards: A CODATA Case History in Nanotechnology. *Data Sci. J.* **2019**, *18*, 1–11.
14.	Schauppenlehner, T.; Muhar, A. Theoretical Availability versus Practical Accessibility: The Critical Role of Metadata Management in Open Data Portals. *Sustainability* **2018**, *10*, 545.
15.	McQuilton, P.; Batista, D.; Beyan, O.; Granell, R. Helping the Consumers and Producers of Standards, Repositories and Policies to Enable FAIR Data. *Data Intell.* **2020**, *2*, 151–157. [CrossRef]
16.	Guo, S.; Yan, D.; Hu, S.; An, J. Global comparison of building energy use data within the context of climate change. *Energy Build.* **2020**, *226*, 110362. [CrossRef]
17.	Kopackova, H.; Michalek, K.; Cejna, K. Accessibility and findability of local e-government websites in the Czech Republic. *Univ. Access Informat. Soc.* **2009**, *9*, 51–61. [CrossRef]
18.	Hirth, L.; Mühlenpfordt, J.; Bulkeley, M. The ENTSO-E Transparency Platform—A review of Europe's most ambitious electricity data platform. *Appl. Energy* **2018**, *225*, 1054–1067. [CrossRef]
19.	Francis, D.J.; Das, A.K. Open Data Resources for Clean Energy and Water Sectors in India. *J. Libr. Inform. Technol.* **2019**, *39*, 300–307. [CrossRef]
20.	Holmegaard, E.; Johansen, A.; Kjærgaard, M.B. Towards a metadata discovery, maintenance and validation process to support applications that improve the energy performance of buildings. In Proceedings of the IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops), Sydney, Australia, 14–18 March 2016; pp. 1–6.
21.	Oakleaf, J.; Kennedy, C.; Baruch-Mordo, S.; Gerber, J.; West, P.; Johnson, J.; Kiesecker, J. Mapping global development potential for renewable energy, fossil fuels, mining and agriculture sectors. *Sci. Data* **2019**, *6*, 101. [CrossRef] [PubMed]
22.	Mishra, S.; Glaws, A.; Cutler, D.; Frank, S.; Azam, M.; Mohammadi, F.; Venne, J.S. Unified architecture for data-driven metadata tagging of building automation systems. *Autom. Constr.* **2020**, *120*, 103411. [CrossRef]
23.	Fahrenkrog, G.; Polt, W.; Rojo, R.; Tubke, A.; Zinocker, K. RTD Evaluation Toolbox. Assessing the Socio-Economic Impact of RTD Policies, Prepared by Joanneum Research and the Institute for Prospective Technological Studies (JRC-EC), IPTS, Seville, 2002. Available online: https://publications.jrc.ec.europa.eu/repository/handle/JRC23461 (accessed on 10 September 2021).
24.	Morrison, R. Energy system modeling: Public transparency, scientific reproducibility, and open development. *Energy Strategy Rev.* **2018**, *20*, 49–63. [CrossRef]
25.	Martin, S.; Foulonneau, M.; Turki, S.; Ihadjadene, M. Risk Analysis to Overcome Barriers to Open Data. *Electron. J. e-Gov.* **2013**, *11*, 324–388.
26.	Pfenninger, S.; DeCarolis, J.; Hirth, L.; Quoilin, S.; Staffell, I. The importance of open data and software: Is energy research lagging behind? *Energy Policy* **2017**, *107*, 211–215. [CrossRef]
27.	Benndorf, G.A.; Wystrcil, D.; Réhault, N. Energy performance optimization in buildings: A review on semantic interoperability, fault detection, and predictive control. *Appl. Phys. Rev.* **2018**, *5*, 041501. [CrossRef]
28.	Li, S.; Yang, J.; Fang, J.; Liu, Z.; Zhang, H. Electricity scheduling optimisation based on energy cloud for residential microgrids. *IET Renew. Power Gener.* **2019**, *13*, 1105–1114. [CrossRef]

29. Vikhorev, K.; Greenough, R.; Brown, N. An advanced energy management framework to promote energy awareness. *J. Clean. Prod.* **2013**, *43*, 103–112. [CrossRef]

30. Fahey, R.A.; Hino, A. COVID-19, digital privacy, and the social limits on data-focused public health responses. *Int. J. Inf. Manag.* **2020**, *55*, 102181. [CrossRef]

31. Hallinan, D.; Friedewald, M.; Mc Carthy, P. Citizens' perceptions of data protection and privacy in EU. *Comput. Law Secur. Rev.* **2012**, *28*, 263–272. [CrossRef]

32. Brunnschweiler, C.; Edjekumhene, I.; Lujala, P. Does information matter? Transparency and demand for accountability in Ghana's natural resource revenue management. *Ecol. Econ.* **2020**, *181*, 106903. [CrossRef]

33. Cantadora, I.; Cortés-Cediel, M.E.; Fernández, M. Exploiting Open Data to analyze discussion and controversy in online citizen participation. *Inf. Process. Manag.* **2020**, *57*, 102301. [CrossRef]

34. Veliz, C.; Grunewald, P. Protecting data privacy is key to a smart energy future. *Nat. Energy* **2018**, *3*, 702–704. [CrossRef]

35. Gendron, J.; Killian, D. Data citizens: Rights and responsibilities in a data republic. In *Data Democracy—At the Nexus of Artificial Intelligence, Software Development and Knowledge Engineering*; Batarseh, F.A., Yang, R., Eds.; Academic Press: London, UK, 2020; pp. 9–28.

36. Anhalt-Depies, C.; Stenglein, J.L.; Zuckerberg, B.; Townsend, P.A.; Rissman, A.R. Tradeoffs and tools for data quality, privacy, transparency, and trust in citizen science. *Biol. Conserv.* **2019**, *238*, 108195. [CrossRef]

37. Lhoste, E.F. Can do-it-yourself laboratories open up the science, technology, and innovation research system to civil society? *Technol. Forecast. Soc. Chang.* **2020**, *161*, 120226. [CrossRef]

38. Dunnett, S.; Sorichetta, A.; Taylor, G.; Eigenbrod, F. Harmonised global datasets of wind and solar farm locations and power. *Sci. Data* **2020**, *7*, 130. [CrossRef]

39. The State of AI and Machine Learning 2017. Available online: https://appen.com/resources/whitepapers/ (accessed on 3 February 2021).

40. The State of AI and Machine Learning 2018. Available online: https://appen.com/resources/whitepapers/ (accessed on 3 February 2021).

41. Ntoutsi, N.; Fafalios, P.; Gadiraju, U.; Iosifidis, V.; Nejdl, W.; Vidal, M.E.; Ruggieri, S.; Turini, F.; Papadopoulos, S.; Krasanakis, E.; et al. Bias in Data-driven AI Systems—An Introductory Survey. *arXiv* **2020**, arXiv:2001.09762.

42. Bascones, M.; Staniforth, A. What is all this fuss about? Is wrong metadata really bad for libraries and their end-users? *Insights* **2018**, *31*, 41. [CrossRef]

43. Studer, R.; Benjamins, R.; Fensel, D. Knowledge engineering: Principles and methods. *Data Knowl. Eng.* **1998**, *25*, 161–198. [CrossRef]

44. Arp, R.; Smith, B.; Spear, A.D. *Building Ontologies with Basic Formal Ontology*; MIT Press: Cambridge, MA, USA, 2015.

45. Glauer, M.; Booshehri, M.; Emele, L.; Flügel, S.; Förster, H.; Frey, J.; Frey, U.; Hastings, J.; Hofmann C.; Hoyer-Klick, C.; et al. The Open Energy Ontology ESWC 2021 Conference Resources Track. Available online: https://openreview.net/forum?id=ofZ5BBg2lSe (accessed on 23 December 2020).

46. Stavropoulos, T.G.; Vrakas, D.; Vlachava, D.; Bassiliades, N. BOnSAI: A smart building ontology for ambient intelligence. In *WIMS '12 Proc. of the 2nd International Conference on Web Intelligence, Mining and Semantics*; Burdescu, D.D., Akerkar, R., Bǎdicǎ, C., Eds.; Association for Computing Machinery: New York, NY, USA, 2012; pp. 1–12.

47. Lork, C.; Choudhary, V.; Hassan, N.U.; Tushar, W.; Yuen, C.; Ng, B.K.K.; Wang, X.; Liu, X. An Ontology-Based Framework for Building Energy Management with IoT. *Electronics* **2019**, *8*, 485. [CrossRef]

48. Ouhajjou, N.; Loibl, W.; Anjomshoaa, A.; Fenz, S. An Ontology-based Urban Energy Planning Support: Building-Integrated Solar PV. In *Architecture, Engineering, and Construction*; Mahdavi, A., Martens, B., Scherer, R., Eds.; CRS Press: London, UK, 2014; pp. 543–550.

49. Shah, N.; Chao, K.M.; Zlamaniec, T.; Matei, A. Ontology for Home Energy Management Domain. In *Digital Information and Communication Technology and Its Applications, Proceedings of the International Conference DICTAP 2011*; Cherifi, H., Mohamad Zain, J., El-Qawasmeh, E., Eds.; Springer: Cham, Switzerland, 2011; pp. 337–347.

50. Santos, G.; Pinto, T.; Vale, Z.; Praça, I.; Morais, H. Electricity markets ontology to support MASCEM's simulations. In *Highlights of Practical Applications of Scalable Multi-Agent Systems. The PAAMS Collection: International Workshops of PAAMS 2016*; Bajo, J., Ed.; Springer: Cham, Switzerland, 2016; pp. 393–404.

51. Oppermann, L.; Hinrichs, E.; Schade, U.; Koch, T.; Rettweiler, M.; Ohrem, F.; Plötz, P.; Beier, C.; Prinz, W. EnArgus: Zentrales Informationssystem Energieforschungsförderung. In *GI Edition Proceedings Band 246 45. Jahrestagung der Gesellschaft für Informatik—INFORMATIK 2015*; Cunningham, D.W., Hofstedt, P., Meer, K., Schmitt, I., Eds.; Gesellschaft für Informatik eV.: Bonn, Germany, 2015; pp. 1717–1730.

52. Burel, G.; Piccolo, L.G.S.; Alani, H. EnergyUse—A Collective Semantic Platform for Monitoring and Discussing Energy Consumption. In *The Semantic Web—ISWC 2016*; Groth, P., Simperl, E., Gray, A.J.G., Sabou, M., Krötzsch, M., Lecue, F., Flöck, G.Y., Eds.; Springer: Cham, Switzerland, 2016; pp. 257–272.

53. Blomqvist, E.; Thollanderb, P.; Keskisärkkä, R. *Energy Efficiency Measures as Linked Open Data*; IOS Press: Lund, Sweden, 2014.

54. Cuenca, J.; Larrinaga, F.; Curry, E. A Unified Semantic Ontology for Energy Management Applications. WSP/WOMoCoE@ISWC, 86–97. Data Scientist Report. 2017. Available online: https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport.pdf (accessed on 6 December 2020).

55. Boosheri, M.; Emele, L.; Flügel, S.; Förster, H.; Frey, J.; Frey, U.; Glauer, M.; Hastings, J.; Hofmann, C.; Hoyer-Klick, C.; et al. Introducing the Open Energy Ontology: Enhancing data interpretation and interfacing in energy systems analysis. *Energy AI* **2021**, *5*, 100074. [CrossRef]
56. Devanand, A.; Karmakar, G.; Krdzavac, N.; Rigo-Mariani, R.; Foo Eddy, Y.S.F.; Karimi, I.A.; Markus Kraft, M. OntoPowSys: A power system ontology for cross domain interactions in an eco industrial park. *Energy AI* **2020**, *1*, 100008. [CrossRef]
57. Küçük, D.; Küçük, D. OntoWind: An Improved and Extended Wind Energy Ontology. *arXiv* **2018**, arXiv:1803.02808.
58. Gillani, S.; Laforest, F.; Picard, G. A Generic Ontology for Prosumer-Oriented Smart Grid. In Proceedings of the CEUR Workshop, Aalborg, Denmark, 7–11 July 2014.
59. ETSI: SmartM2M;Smart Appliances Extension to SAREF; Part 1: Energy Domain, (ETSI TS 103 410-1), 2017. Available online: https://saref.etsi.org/saref4ener/v1.1.2/ (accessed on 30 December 2020).
60. Haghgoo, M.; Sychev, I.; Monti, A. SARGON–Smart energy domain ontology. *IET Smart Cities* **2020**, *2*, 191–198. [CrossRef]
61. Lefrançois, M. Planned ETSI SAREF Extensions based on the W3C&OGC SOSA/SSN-compatible SEAS Ontology Patterns. In Proceedings of the Workshop on Semantic Interoperability and Standardization in the IoT, SIS-IoT, Amsterdam, The Netherlands, 20 November 2017.
62. Madrazo, L.; Sicilia, A.; Gamboa, G. SEMANCO: Semantic tools for carbon reduction in urban planning. In Proceedings of the 9th European Conference Product & Process Modelling, Reykjavik, Island, 25–27 July 2012.
63. Kott, J.; Kott, M. Generic Ontology of Energy Consumption Households. *Energies* **2019**, *12*, 3712. [CrossRef]
64. Kofler, N.J.; Reinisch, C.; Kastner, W. A semantic representation of energy-related information in future smart homes. *Energy Build.* **2012**, *47*, 169–179. [CrossRef]
65. Sempreviva, A.M.; Vesth, A.; Bak, C.; Verelst, D.R.; Giebel, G.; Danielsen, H.K.; Mikkelsen, L.P.; Andersson, M.; Vasiljevic, N.; Barth, S.; et al. Taxonomy and metadata for wind energy Research and Development. IRPWind Report 2017, Zenodo. Available online: http://doi.org/10.5281/zenodo.1199489 (accessed on 17 February 2021).
66. Vasiljevic, N.; Sempreviva, A.M. Wind Energy Taxonomy of Topics. Available online: http://data.windenergy.dtu.dk/controlled-terminology/taxonomy-topics/ (accessed on 17 February 2021).
67. DTU Wind Energy, Wind Energy Parameters. Available online: http://data.windenergy.dtu.dk/controlled-terminology/wind-energy-parameters/ (accessed on 17 February 2021).
68. Collier, D.; Mahon, J. Conceptual 'Stretching' Revisited: Adapting Categories in Comparative Analysis. *Am. Polit. Sci. Rev.* **1993**, *87*, 845–855.
69. Sartori, G. Concept Misformation in Comparative Politics. *Am. Polit. Sci. Rev.* **1970**, *64*, 1033–1053.
70. Chen, H.; Luo, X. An automatic literature knowledge graph and reasoning network modeling framework based on ontology and natural language processing. *Adv. Eng. Inf.* **2019**, *42*, 100959. [CrossRef]
71. C2meta. 2021. Continuous Capture of Metadata, University of Michigan. Available online: https://c2metadata.org/ (accessed on 12 January 2021).
72. Heckman, J.J. Selection bias. In *Encyclopedia of Social Measurement*; Kempf-Leonard, K., Ed.; Elsevier: Amsterdam, The Netherlands, 2005.
73. Thornberg, G.; Oskins, W.M. Misinformation and Bias in Metadata Processing: Matching in Large Databases. *Inf. Technol. Libr.* **2007**, *26*, 15–26. [CrossRef]
74. Murray, D.; Stankovic, L.; Stankovic, V. An electrical load measurements dataset of United Kingdom households from a two-year longitudinal study. *Sci. Data* **2017**, *4*, 160122.
75. Jiang, L.; Xu, L.D.; Cai, H.; Jiang, Z.; Bu, F.; Xu, B. An IoT-oriented data storage framework in cloud computing platform. *IEEE Trans. Ind. Inf.* **2014**, *10*, 1443–1451.
76. Govindarajan, R.; Meikandasivam, S.; Vijayakumar, D. Cloud computing based smart energy monitoring system. *Int. J. Sci. Technol. Res.* **2019**, *8*, 886–890.
77. Giordano, A.; Mastroianni, C.; Sorrentino, N.; Menniti, D.; Pinnarelli, A. An energy community implementation: The unical energy cloud. *Electronics* **2019**, *8*, 1517.
78. Burgio, A.; Giordano, A.; Manno, A.A.; Mastroianni, C.; Menniti, D.; Pinnarelli, A.; Scarcello, L.; Sorrentino, N.; Stillo, M. An IoT Approach for Smart Energy Districts. In Proceedings of the 2017 IEEE 14th International Conference on Networking, Sensing and Control (ICNSC), Calabria, Italy, 16–18 May 2017; pp. 146–151.
79. Go-Fair Website, Metadata for Machines. Available online: https://www.go-fair.org/how-to-go-fair/metadata-for-machines/ (accessed on 3 December 2020).
80. Gonçalves, R.S.; O'Connor, M.J.; Martínez-Romero, M.; Egyedi, A.L.; Willrett, D.; Graybeal, J.; Musen, M.A. The CEDAR Workbench: An Ontology-Assisted Environment for Authoring Metadata that Describe Scientific Experiments. In *The Semantic Web—ISWC 2017*; d'Amato, C., Fernandez, M., Tamma, V., Lecue, F., Cudré-Mauroux, P., Sequeda, J., Lange, C., Heflin, J., Eds.; Springer: Cham, Switzerland, 2017; pp. 103–110.
81. Vasiljevic, N. sheet2rdf: Automatic Workflow for Generation of RDF Vocabularies from Google Sheets (Version v0.1). Zenodo. Available online: http://doi.org/10.5281/zenodo.4432136 (accessed on 17 February 2021).
82. Vasiljevic, N.; Graybeal, J. excel2rdf: Automatic Workflow for Generation of RDF Vocabularies from Excel Sheets (Version v0.1). Zenodo. Available online: http://doi.org/10.5281/zenodo.4273159 (accessed on 17 February 2021).

83. DeiC. 2020. M4M for the Danish e-Infrastructure Cooperation. Available online: https://www.go-fair.org/2020/07/08/m4m-for-the-danish-e-infrastructure-cooperation/ (accessed on 17 February 2021).

84. Graybeal, J. Easy FAIR Metadata for Earth Science. Available online: http://bit.ly/fair-metadata (accessed on 17 February 2020).

85. Hotmaps Project. Available online: https://gitlab.com/hotmaps/ (accessed on 17 February 2021).

86. Magagna, B.; Schultes, E.A.; Pergl, R.; Hettne, K.; Kuhn, T.; Suchánek, M. Reusable FAIR Implementation Profiles as Accelerators of FAIR Convergence. OSF Preprints. Available online: https://doi.org/10.31219/osf.io/2p85g (accessed on 17 February 2021).

87. Lammey, R.; Mitchell, D.; Counsell, F. Metadata 2020: A collaborative effort to improve metadata quality in scholarly communications. In *Septentrio Conference Series*; No 1 (2018): The 13th Munin Conference; Scholarly Publishing: Tromsø, Norway, 2018.

88. Schneider, J. Metadata 2020: Updates and Plans, Research Data Alliance, Berlin 2018. Available online: https://rd-alliance.org/system/files/documents/RDA_Metadata2020.pdf (accessed on 28 December 2020).

89. National Information Standards Organization (NISO). A Framework of Guidelines for Building Good Digital Collections (3rd Edition). 2007. Available online: https://www.niso.org/sites/default/files/2017-08/framework3.pdf (accessed on 17 February 2021).

90. Wilkinson, M.D.; Dumontier, M.; Sansone, S.A.; da Silva Santos, L.O.B.; Prieto, M.; Batista, D.; McQuilton, P.; Kuhn, T.; Rocca-Serra, P.; Crosas, M.; et al. Evaluating FAIR maturity through a scalable, automated, community-governed framework. *Sci. Data* **2019**, *6*, 174. [CrossRef] [PubMed]

91. Király, P.A. *Metadata Quality Assurance Framework*; Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG): Göttingen, Germany, 2015.

92. Stiller, J.; Király, P. Multilinguality of Metadata Measuring the Multilingual Degree of Europeana's Metadata. In *Everything Changes, Everything Stays the Same? Understanding Information Spaces, Proceedings of the 15th International Symposium of Information Science (ISI 2017), Berlin, 13–15 March 2017*; Gäde, M., Trkulja, V., Petras, V., Eds.; Werner Hülsbusch Publishing: Glückstadt, Germany, 2017; pp. 164–176.

93. Bruce, T.R.; Hillmann, D.I. The Continuum of Metadata Quality: Defining, Expressing, Exploiting. In *Metadata in Practice*; Hillmann, D., Westbrooks, E., Eds.; ALA Editions: Chicago, IL, USA, 2004; pp. 238–256.

94. Ochoa, X.; Duval, E. Automatic Evaluation of Metadata Quality in Digital Repositories. *Int. J. Dig. Libr.* **2009**, *10*, 67–91. [CrossRef]

95. Kemp, J.; Dean, C.; Chodacki, J. Can Richer Metadata rescue research? *Ser. Libr.* **2018**, *74*, 207–211. [CrossRef]

96. Scott, J. The Metadata Mania. 2011. Available online: http://ascii.textfiles.com/archives/3181 (accessed on 17 February 2021).