

## Article

# Sustainable Insights for Energy Big Data Governance in China: Full Life Cycle Curation from the Ecosystem Perspective

Ming Zeng, Yanbin Xu \*, Haoyu Wu, Jiaxin Ma and Jianwei Gao

School of Economics and Management, North China Electric Power University, Changping District, Beijing 102206, China; zengmingbj@vip.sina.com (M.Z.); why614613@126.com (H.W.); mancepu@163.com (J.M.); gjw102206@126.com (J.G.)

\* Correspondence: baizuxuyanbin@163.com

**Abstract:** With the development of the Energy Internet and the Internet of Things, diversified social production activities are making the interactions between energy, business, and information flow among physical, social, and information systems increasingly complex. As the carrier of information and the hub between physical and social systems, the effective management of energy big data has attracted the attention of scholars. This work indicates that China's energy companies have carried out a series of activities that are centered on energy big data collection, as well as development and exchange, and that the energy big data ecosystem has begun to take shape. However, the research on and the application of energy big data are mainly limited to micro-level fields, and the development of energy big data in China remains disordered because the corresponding macro-level instructive governance frameworks are lacking. In this work, to facilitate the sustainable development of the energy big data ecosystem and to solve existing problems, such as the difficult-to-determine governance boundaries and the difficult-to-coordinate interests, and to analyze the structure and mechanism of the energy big data ecosystem, data curation is introduced into energy big data governance, and a paradigm is constructed for sustainable energy big data curation that encompasses its full life cycle, including the planning, integration, application, and maintenance stages. Key paradigmatic issues are analyzed in-depth, including data rights, fusion, security, and transactions.

**Keywords:** big data; energy big data governance; energy big data ecosystem; full life cycle curation



**Citation:** Zeng, M.; Xu, Y.; Wu, H.; Ma, J.; Gao, J. Sustainable Insights for Energy Big Data Governance in China: Full Life Cycle Curation from the Ecosystem Perspective. *Sustainability* **2022**, *14*, 6013. <https://doi.org/10.3390/su14106013>

Academic Editor: Thanikanti Sudhakar Babu

Received: 14 March 2022

Accepted: 9 May 2022

Published: 16 May 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Decades of informatization construction in China have achieved data collection for the production, transmission, transaction, consumption, and other aspects of the energy industry. Thus, during energy power system construction and operation, massive data resources are accumulated. With the development of energy technology and the advent of the digital economy, the developing data resources have attracted the attention of numerous energy companies and scholars in China.

In terms of development strategies, China's energy enterprises are represented by the SGCC and China Energy, which have issued corresponding big data development strategies, as shown in Table 1. In terms of business models, the business models of energy big data companies usually include value propositions, business system design, market expansion, risk control, etc. [1,2]. In addition, Chen et al. proposed that the enterprise-level energy big data business models can be evaluated from the aspects of profitability, customer value, strategic positioning, etc. [3]. Chen et al. constructed an evaluation system for energy big data business models from the perspectives of the economy, technology, the environment, and society [4]. In terms of standard establishments, International Electrotechnical Commission Joint Technical Committee 1 (IEC JTC1) and other international organizations have issued a series of specifications for the basic standards for big data in the power industry [5]. Zhang et al. and Han et al. proposed a standard system for power transmission and transformation projects and for public data publication from data

description, data utilization, and association models [6,7]. In terms of information technologies, non-relational databases are considered to be an important solution for storing and managing large-scale data in energy companies. Data indexing, scheduling, and log management are important research directions in the field of information technology for energy big data governance [8,9].

**Table 1.** Energy big data development strategies of China’s energy companies.

Companies	Associated Development Strategies
State Grid Corporation of China (SGCC) (Beijing, China)	Build a world-leading energy internet company and enhance company momentum by improving the digitalization of grid operations and company business operations.
China Energy Investment Corporation (China Energy) (Beijing, China)	Actively carry out big data applications and governance to achieve goals related to “platform development, digital operations, ecological collaboration, industry chain collaboration and intelligent production”.
China Petroleum & Chemical Corporation (Sinopec Corp.) (Beijing, China)	Access internal and external data widely, create a “data + platform + application” development model, expand to a new space for digital economic growth, and create a new engine for the company’s high-quality development.
China Huaneng Group Co., Ltd. (Beijing, China)	Build a unified intelligent energy data center covering all of the industrial sectors of the company to achieve both the up- and the downstream data penetration of the industrial chain as well as energy production and consumption.
China Southern Power Grid Company (Guangzhou, China)	Accelerate digital transformation and digital grid construction to achieve interconnection patterns for a digital government, national industrial internet, and the energy industry chain.

To sum up, due to the lack of a concept for the full life cycle and ecosystem, the current research and governance of energy big data is still oriented towards specific enterprises or industries. Specifically, China’s energy big data governance work remains largely fragmented from the macro-perspective of the industry. Furthermore, systematic and scientific governance models of energy big data are lacking. Significant data barriers exist between companies, specialties, businesses, and departments. These barriers result in potential data risks, such as non-repeatable losses, low value density, and outdated values. Therefore, based on the concept of the full life cycle and the ecosystem, the construction of a guiding energy big data governance paradigm from the top-level perspective has become the key to promoting the data connection and value co-creation. Such an effort would support data value exploration and energy big data ecosystem construction and would promote an energy revolution that is driven by the digital economy.

In response to the above problems, this paper studies the element structure, the interaction mechanism between elements, the system operation, and the evolution mechanism of China’s energy big data ecosystem and further points out the development trends and critical challenges of China’s energy big data ecosystem. Secondly, this paper presents a paradigm for creating a full life cycle curation of sustainable energy big data that includes planning, integration, application, and maintenance. Furthermore, four key issues involved in the full life cycle curation are analyzed. These research results can provide a relevant basis and reference for the sustainable development of energy big data in China from the perspective of the macro-strategies and top-level design.

## 2. Status Review and Trends of Energy Big Data in China

### 2.1. Status Review of Energy Big Data

#### 2.1.1. Development

Data analysis methods and data application scenarios differ in the field of energy big data. Based on these differences, the development of energy big data can be divided into three periods: the sprout period (before 2010), the boom period (2010–2016), and the innovative development period (after 2016), as shown in Figure 1 [10–12].

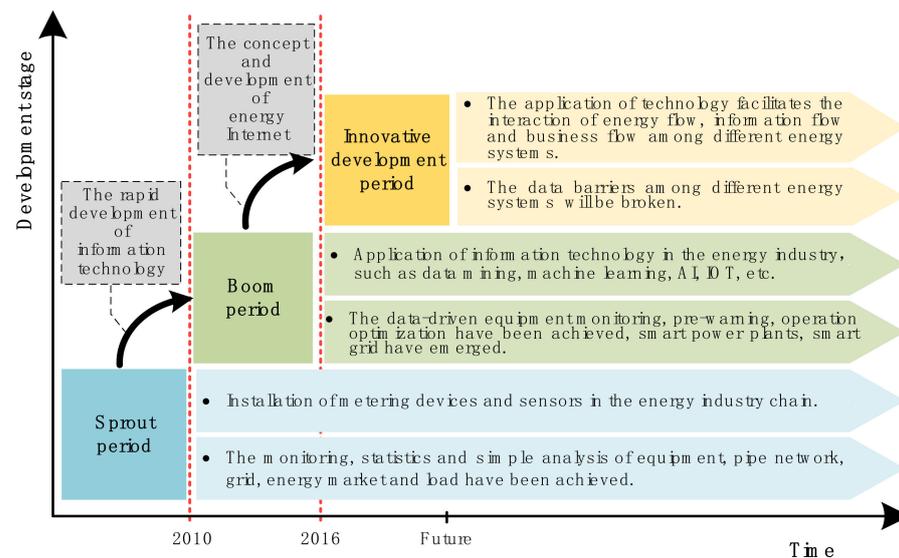


Figure 1. Development of energy big data.

During the sprout period, energy data—which include device parameters; operating data; trading prices; and the volume data for electricity, coal, and gas—could only be obtained by distributing metering equipment, installing sensors, and maintaining a manual record. These data were processed using statistical and visual processing methods to monitor and analyze the equipment, pipe networks, energy markets, and demand-side behavior. Furthermore, such analyses provided support for equipment operation, market bidding, and business development strategies by providing an understanding of the operational state of the equipment, energy transaction price trends, user preferences, etc.

The boom period witnessed the rapid development of information technologies such as data mining, machine learning, and artificial intelligence. Therefore, it became possible to further analyze and learn from energy big data. The advent of technologies such as the Internet of Things and advanced sensors permitted the timely collection of system-, equipment-, and even component-level data. Furthermore, equipment condition monitoring, pre-warning systems, and operation optimization driven by energy big data developed gradually. Meanwhile, research on and the construction of smart power stations, smart grids, and smart devices increased rapidly. These technologies and achievements significantly improved the security, reliability, and efficiency of the energy systems.

During the innovative development period, the concept of the energy internet was proposed. This concept deepened the integration of energy, business, and data flows during the production, transmission, distribution, trading, and consumption of energy sources such as electric, heating, and gas systems. The data among different energy systems were effectively connected. Additionally, emerging technologies—such as cloud computing, situational awareness, and deep-learning techniques—were applied and popularized. This process began to promote the transformation and upgrade of energy systems. Meanwhile, integrated energy system scheduling optimization and integrated energy market transaction decision optimization, etc., became possible [13–15].

### 2.1.2. Research

#### (1) Energy big data technology

Energy big data technology can be divided into four categories: acquisition, integration, analysis, and presentation technologies [14,15].

Data acquisition can be achieved through databases, networks, and devices. Common technologies include Oracle, NoSQL, web crawler technology, APIs, smart meters, and device sensors.

Data integration refers to the integration of distributed heterogeneous data from various sources and with different formats to form valuable, consistent, and available data resources. Data integration technology systems mainly include data extraction, cleaning, transformation, and storage.

Data analysis refers to the process of discovering hidden laws and knowledge through the analysis of massive, high-dimensional, and heterogeneous data. In terms of algorithms, the main data analysis technologies involve clustering, regression, association rule mining, deep neural networks, etc. In terms of modes, the main data analysis technologies include Spark, MapReduce, HaLoop, etc.

Data presentation refers to the use of image processing and other methods to transform data into visual forms to allow the information and rules contained therein to be displayed. Data visualization is an important data presentation technology that can be regarded as a product of the interaction of scientific visualization, information visualization, and visual analysis. Data visualization includes visualizations of relational, statistical, and spatial data based on the specific type of data being displayed.

#### (2) Applications of energy big data

Energy big data are widely used in energy industry planning, operation, consumption, and policy making [16–20].

In planning scenarios, the distribution laws of energy resources and loads in the spatiotemporal dimension can be mastered through the correlation and spatiotemporal analyses of energy big data, such as renewable energy resources data, meteorology data, geographic data, macroeconomic data, and multi-energy load data. Additionally, these laws can provide critical guidance for energy station site selection, installed capacity determination, and pipe network topology design, etc.

In operation scenarios, awareness of the generator real-time shift peak load capacity, real-time transmission capacity, and equipment real-time operating conditions provides a valuable reference for operation optimization and pre-warning systems. This awareness can be obtained by carrying out a trend analysis based on monitoring the operational data of the energy system, equipment, pipe network, and node load under various working conditions.

In consumption scenarios, the correlation between a user's energy consumption behavior and meteorological, economic, demographic, energy price, and other data provides the theoretical support to create a user portrait. This correlation can be revealed through an energy big data analysis and provides the support for multi-timescale and multi-energy load predictions (e.g., medium- and long-term heat-load forecasting, ultra-short-term power-load forecasting, etc.), energy market transaction decisions, demand-side management, etc.

In policy-making scenarios, energy big data analyses can help to reveal the internal correlations between the effects of policy implementations and local energy consumption structures, economic development levels, industrial structures, and other factors. These can then be used to provide suggestions for the establishment and improvement of policy mechanisms. For example, an energy big data analysis can calculate and predict renewable energy outputs, regional load demand, and the trans-regional transportation costs and capacities of renewable energy sources. Thus, these analyses provide important references for the development of renewable energy trans-regional trading mechanisms and for the formulation of renewable quota policies.

### 2.1.3. Policies

In recent years, many important economies, including China, have focused on promoting the development of the digital economy, increased the development of big data to a national strategic height, and introduced a number of relevant policies to promote the development of energy big data. Tables 2 and 3 list the policies related to energy big data in China and in typical major economies, respectively.

**Table 2.** China's policies regarding energy big data.

Implementation Time	Departments	Strategies	Related Content
2015	The State Council	<i>Outline of Action to Promote the Development of Big Data</i>	Propose the planning and construction of an information resource base and security protection systems in important fields, including the energy field; accelerate data collection, analysis, and open sharing; and promote the integration of data resources
2016	Ministry of Industry and Information Technology	<i>The Big Data Industry Development Plan (2016–2020)</i>	Promote big data applications in energy and other industries; ensure data information security in energy and other industries as a key project
2016	National Development and Reform Commission, National Energy Administration	<i>Guiding Opinions on Promoting the Development of "Internet+" Smart Energy</i>	Propose the development of the energy big data service industry; improve energy big data business and supervision systems; encourage cooperation between internet companies and energy companies to explore the commercial value of energy big data; and promote innovation in energy big data applications
2017	Drafting committee directly led by the Political Bureau of the CPC Central Committee	<i>Reports of the 19th National Congress of the Communist Party of China</i>	Propose promoting the deep integration of the big data, energy, and other real economy sectors; strengthen the construction of power grids and other infrastructure networks; and support the optimization and upgrade of traditional industries
2020	Ministry of Industry and Information Technology	<i>Guidance on the development of industrial big data</i>	Propose promoting the digital transformation of energy and other important industries while considering data aggregation, data sharing, data applications, data governance, and data industry development as key tasks

**Table 3.** The policies of major economies regarding energy big data.

Economies	Relevant Policies	Related Content
The United States	Digital government Strategy, Grid 2030, Federal Big Data Research and Development Strategic Plan, Federal Data Strategy, Critical and Emerging Technologies National Strategies, etc.	<p>In the early 2010s, the United States proposed exploring the potential value of data by building an open and shared digital platform.</p> <p>In the mid-2010s, the United States further enhanced investment and guidance on big data-related projects in various fields, including the energy industry. These projects included the development and application of big data technology, the establishment of data sharing and protection mechanisms, etc.</p> <p>In the early 2020s, the United States proposed standardizing the big data governance model, technical standards, resource library construction, and ethical framework in various fields. The positioning of big data begins to change from “technology” to an “asset”.</p>
European Union	European Digital Agenda, Digitising European Industry, Shaping Europe’s digital future, 2030 Digital Compass, Action Plan on the Digitalisation of the Energy Sector, etc.	<p>In the early 2010s, the EU proposed that all of society should enjoy the benefits brought by the digital era. Data market construction, the development of digital technology standards, and data security enhancement were identified as the critical directions for development directions in the future.</p> <p>In the mid-2010s, the EU proposed speeding up the process of digital infrastructure construction and increasing the digital innovation capacity of various industries by relying on Digital Innovation Hubs, thereby accelerating the process of European industrial digitalization.</p> <p>In the early 2020s, the EU proposed accelerating the digitalization process from the industry level rather than at the enterprise level.</p>
The United Kingdom	Data Capability Development Strategic Plan, UK Digital Strategy, Industrial Strategy: building a Britain fit for the future, National Data Strategy, etc.	<p>In the early 2010s, the UK proposed attaching importance to the cultivation of talents in the field of big data technology and improving the legal and institutional system to ensure that data could be shared securely.</p> <p>In the mid-2010s, the UK proposed strengthening the development and protection of big data in the industrial field and further improving digital infrastructure in the energy sector.</p> <p>In the early 2020s, the UK proposed fully exploring the economic value of big data and promoting the cross-border flow of data under the premise of ensuring security.</p>
Japan	Basic Law on the Promotion of Public and Private Data Utilization, Manifesto to Create the World’s Most Advanced Information Technology Nation, Strategic Implementation Framework for Japan’s Industrial Internet Value chains, Science and Technology Basic Plan Sixth Edition, etc.	<p>In the mid-2010s, Japan proposed accelerating the application of big data technology in agriculture, industry, and other fields. In addition, Japan proposed promoting the reasonable opening and circulation of public and private data through the formulation of laws and standards.</p> <p>In the early 2020s, Japan proposed a top-level framework for the development of the industrial internet in the digital era and promoting technological research and innovation in big data. Additionally, Japan proposed accelerating industrial transformation and upgrading by promoting the application of big data technology in energy industries.</p>

Table 3. Cont.

Economies	Relevant Policies	Related Content
Russia	2017~2030 Russian Federation Information Society Development Strategy, Russian Federation digital economy planning, Development goals and strategic tasks of the Russian Federation until 2024, Decree on the Development Goals of the Russian Federation up to 2030, etc.	In the mid-2010s, Russia proposed positioning the digital economy as one of 12 priority areas for development by 2024 and analyzing and managing big data as a key production factor; Russia also proposed promoting a revolution in production methods, technology, and equipment in various fields through digital infrastructure construction and big data technology talent training and application. In the early 2020s, Russia proposed positioning digital transformation as one of the five national development goals and increasing investment in big data technology.

Comparing Tables 2 and 3, it can be seen that many of the world’s important economies have positioned digital transformation as an important direction for development in the future and have actively promoted the application of big data technology in important fields related to national development, including the energy industry, agriculture, etc. In contrast, China has clarified the production factor attributes of big data and has actively carried out relevant policies and strategies; however, the progress in terms of top-level framework design, governance models, governance specifications, and big data sharing mechanisms is lagging behind.

2.2. Evolutionary Trends of Energy Big Data—The Energy Big Data Ecosystem

In the field of ecology, a natural ecosystem refers to a dynamic system that is composed of biological populations and their living environment. Individuals in the biological population live in a dynamic balance of interactions and mutual influences, where energy flow and material exchange constantly occur between different individuals. The expansion of the digital economy has further highlighted the attributes of the production factors and assets of data. Further exploration into the potential benefits and values in energy big data has attracted the attention of China’s energy industry (Table 1). As shown in Figure 2, the energy big data ecosystem has begun to take shape. This section analyzes the characteristics of the energy big data ecosystem from three aspects: ecosystem elements, the interactions between elements, and the ecosystem’s operating mechanism.

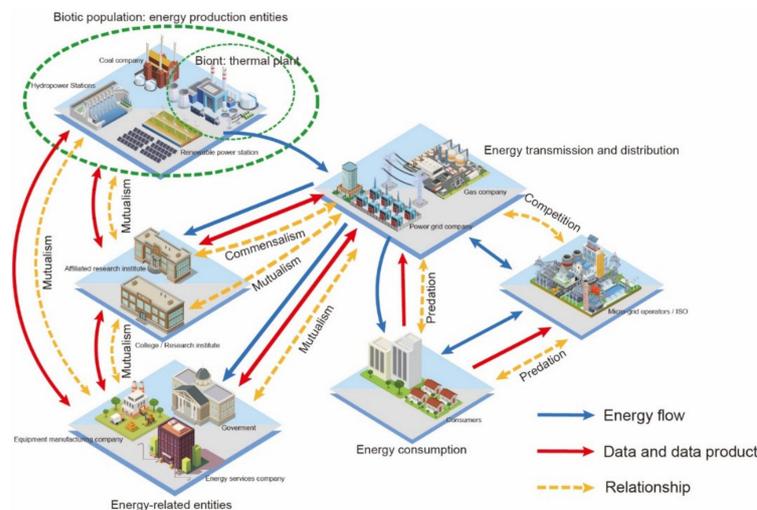


Figure 2. Schematic diagram of the energy big data ecosystem.

### 2.2.1. Elements of the Energy Big Data Ecosystem

From the perspective of element composition, the elements in the energy big data ecosystem mainly consist of energy production companies, equipment manufacturers, power grid companies, scientific research institutions, and other entities in the energy industry. Each single company or institution can be regarded as an individual. Similarly, companies or institutions with the same attributes can be regarded as a population. Together, all the entities can be regarded as the biocoenosis, while the data can be regarded as the inanimate matter and energy in the ecosystem.

From the perspective of the functions of these elements, in accordance with data generation, development, application, and extinction, the companies or institutions that are involved in each stage can be regarded as producers, primary consumers, secondary consumers, or decomposers in the energy big data ecosystem. A given entity in the energy big data ecosystem can switch between the roles of the producer or the consumer in different situations [21]. Take, for example, an energy production company. When collecting the historical data of units and generators, the company is acting as a data producer. When the company adopts an energy-saving transformation and optimizes the operation of the unit according to these historical data, the company is acting as a data consumer. When deleting or destroying unit operation historical data, the company is acting as a data decomposer.

### 2.2.2. Interaction between the Elements

Cooperation and competition between different individuals are universal phenomena in natural ecosystems. Correspondingly, companies, institutions, and other entities in the energy big data ecosystem attempt to maximize their own benefits. Thus, they choose cooperation or competition strategies that center on data collection, utilization, and development. These choices can be seen as predation, competition, commensalism, or mutualism [22].

For example, when power grid companies analyze user characteristics by collecting consumption data from users under various conditions, the relationship between the company and its users can be regarded as predation. When both a power grid company and an ISO (independent system operator) analyze user characteristics by collecting power consumption data, their relationship can be regarded as competition. When a power grid company makes development plans based on public data that have been published by the government and the government analyzes the macroeconomic trends via the electricity consumption data collected by the power grid company, the two entities are cooperating with each other, but they are not interdependent. In this case, the relationship between the power grid company and the government can be regarded as commensalism. In another example, a research subsidiary of a power grid company uses data collected by the power grid company for research. The research is used to support the power grid company's formulation of a development strategy. Thus, the relationship between the power grid company and the research subsidiary can be regarded as mutualism.

### 2.2.3. Ecosystem Operation and Evolutionary Mechanism

From a macro-evolutionary mechanism perspective, an energy big data ecosystem is dynamic and open, and it undergoes continuous evolution. Its overall evolution and development are driven by both external motivation (policy release, technological innovation, etc.) and internal motivation (the pursuit of profit maximization, the expansion of the digital service business, etc.) [23]. During this evolutionary process, the elements of the energy big data ecosystem will experience generation (startup), development, maturity, and finally, due to changes in external motivation, limited resources and timeliness; the system elements will then undergo decline and extinction, completing the entire life cycle [24,25]. Taking data as an example, collection, cleaning, and mining can be regarded as the development and maturation processes. Meanwhile, the need to destroy or recollect data due to the degradation of data values caused by technological or market changes can be regarded as the decline and extinction processes.

From the perspective of microcosmic operation mechanisms, the activities of populations in natural ecosystems promote the flow of matter and energy between producers and consumers. Similarly, entities in the energy big data ecosystem conduct a series of activities. These include data collection, processing, development, sharing, exchange, and transaction. Such activities promote the continuous flow of data between data producers and consumers. During the process, on the one hand, the value of the data continues to grow. On the other hand, the value-added data can provide considerable benefits for the data consumers at all levels. Thus, such mechanisms effectively maintain the operation and dynamic balance of the energy big data ecosystem. Consider power grid companies as an example. Power grid companies desensitize the electricity consumption data collected from users and develop primary data products. Scientific research institutes purchase the corresponding data products from power grid companies to meet their research needs and to produce data-driven demand-side management optimization models, namely advanced data products through data mining and analysis. To improve economic benefits and to reduce operational costs, power grid companies purchase data models from scientific research institutes to support the formulation of operational strategies. Above all, the value added to data—from raw data to primary and advanced data products—is achieved through the data flow processes among users, power grid companies, and scientific research institutes. Corresponding benefits are created for data consumers along the way.

### *2.3. Critical Challenges to the Sustainable Development of the Energy Big Data Ecosystem*

Although the energy big data ecosystem has already taken shape, its sustainable development still faces three challenges.

#### *2.3.1. How to Conduct an Ecosystem-Oriented Top-Level Design of Data Governance*

The entities in an energy big data ecosystem spontaneously carry out activities such as collection, development, and transaction that center on the energy big data that are currently available (Section 2.2). However, the operation of the ecosystem as a whole remains disorderly when viewed from a macro-perspective. Some policies and research studies have been created and conducted in relation to energy big data (Section 2.1); in particular, the Chinese government has issued a series of policies to promote the development of the energy big data industry. However, the current research on energy big data primarily focuses on technology and applications at the data level, including acquisition, analysis, integration, model development, scheme optimization, decision support, etc.

In summary, proposing an ecosystem-oriented and instructive top-level design based on the corresponding methods is the primary challenge to the sustainable development of energy big data.

#### *2.3.2. How to Define the Boundaries of Energy Big Data Governance*

The inputs of equipment cost, labor cost, management cost, and large-scale data are critical for achieving the goals of energy big data governance, which include data mining and data value-added and long-term data preservation. However, these activities involve various types of data, such as equipment testing data, equipment operation data, equipment design parameters, pipeline operation data, energy trading and consumption data, meteorological data, and macroeconomic data. This extremely large number of data will reach the petabyte level. The amount of energy data needed for this data governance effort and its potential costs are difficult to measure. In addition, the development and operation of the data systems and data platforms of each entity are independent to some extent. Therefore, structured, semi-structured, and unstructured data—including numeric values, images, and natural language texts—will be generated objectively. The significant complexity and heterogeneity of energy big data will further increase governance costs.

In conclusion, defining the boundaries of energy big data governance will provide important guidance for achieving a balance between goals and costs and will help to foster the sustainable development of the energy big data ecosystem.

### 2.3.3. How to Balance the Relationship between Various Interests in the Ecosystem

From the perspective of the relationships between the macro-system and micro-individuals and to realize the value-added, sharing, long-term preservation, and reuse of energy big data, the energy big data ecosystem requires all of the data sources (micro-individuals) to publish and publicize relevant data. However, to protect their own interests and rights, the micro-individuals in the ecosystem tend to maximize the relevant benefits by disclosing as few of their own data as possible.

From the perspective of the relationships between micro-individuals, as production and management activities increase, energy, business, and capital flows become increasingly frequent between the individuals in the ecosystem. Meanwhile, the data footprints and flow directions become increasingly complex. In addition, because the role of the micro-individual changes as the specific scenario changes (Section 2.2.1), the purpose and goal of each micro-individual's participation in the energy big data governance system also changes continuously. This situation causes the interactions and relationships between the individuals who are involved in data governance to become more complex. Overall, it is challenging to balance the relationship between the rights and benefits of micro-individuals to guide healthy competition and cooperation between the micro-individuals in the ecosystem and to facilitate the sustainable development of the energy big data ecosystem.

## 3. Research Methodology

### 3.1. Data Curation Theory

As early as the 1990s, European and American countries, represented by the United States and the United Kingdom, carried out a series of studies on long-term digital preservation. The goals of these studies were to avoid the data losses caused by technological updates or mismanagement and to ensure the authenticity and integrity of the data [26]. As research has progressed, people have found that negative data preservation will most likely lead to the emergence of dark archives. Therefore, how to ensure long-term data access and utilization by reforming management modes has gradually become an important research topic in library science, information science, and other fields.

In 2001, the Joint Information System Committee (JISC) proposed the need to establish a specialized organization to lead research and work in the digital field while focusing on the preservation and management of digital resources [27].

In 2002, Jim Gray emphasized the timeliness of data preservation in *Online Scientific Research Data Management, Publication and Archiving* and proposed the concept of data curation.

In 2003, a research report by the National Science Foundation (NSF) in the United States pointed out that the absence of scientific and effective data management mechanisms would lead to data failure risks or even data losses. Thus, research on relevant management mechanisms should be strengthened to ensure that data are usable in the future.

In 2004, the Digital Curation Center (DCC) was established and proposed defining the term data curation, which is different from data management. Data curation emphasizes "data preservation, sorting, maintenance and value-added work in the whole life course of data and research at all stages".

After almost two decades of development, data curation is considered to be the most effective method for maintaining the security and authenticity of data resources and achieving long-term data reuse. Beyond library science and information science, research and applications focusing on data curation theory have also attracted the attention of the biomedical and physical sciences. Figure 3 shows the literature on data curation collected by the Web of Science in recent years.

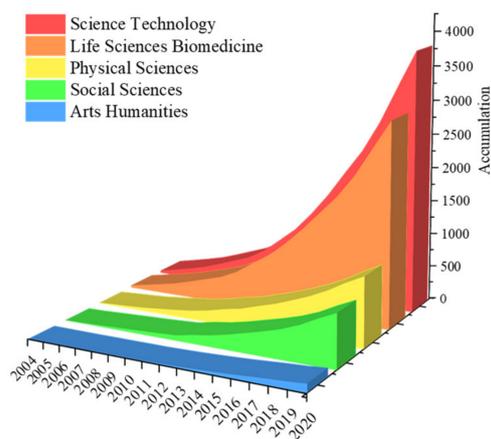


Figure 3. Published literature relevant to data curation collected by the Web of Science.

### 3.2. Data Curation Methods

Since the proposal of data curation, several institutions have proposed corresponding conceptual models based on the definition and core concept of data curation. These models combine the characteristics of various fields, as shown in Table 4 [28–34].

Table 4. Conceptual models of common data curation.

Conceptual Model	Characteristics	Model Elements
DCC model	Mainly applicable to the governance of research data.	Mainly includes creation or reception, appraisal and selection, ingestion, preservation, action, storage, access, reuse, and transformation.
OAIS model	Primarily applied for the long-term preservation, retrieval, and reading of digital resources in information systems.	Consists of five main modules, including ingestion, archival storage, data management, access administration, and preservation planning.
CENS model	Primarily applied to the governance of experimental data.	Includes nine main elements: experimental design, instrument calibration, data collection, data cleaning, calculation and derivation, data integration, data analysis, results publication, and data preservation.
DAMA model	The impact of functional and environmental elements on data governance is fully considered.	Primarily consists of a functional framework and environmental framework. The functional framework includes strategy, organization, roles, and standards. The environmental framework includes culture, practices and methods, technology, and principles.
CALib model	Mainly applied for data governance in university libraries.	Consists of three main submodels: enabler, scope, and implementation and evaluation. The enabler includes strategic objectives, roles and functions, technology, and other elements. The scope includes three elements: a value creation layer, a value guarantee layer, and a basic data layer. Implementation and evaluation include the implementation method, maturity evaluation, and auditing.
DGI model	This model is largely used for data governance in business operations. It provides auxiliary support for decision-making tasks such as enterprise management and business expansion.	Consists of three main submodels: rules, people and organization, and processes. Specific elements include mission and vision, objectives, stakeholders, data managers, data rules and definitions, and ten other elements.
Gartner 6 phases model	This model has process management characteristics and is mostly used for data governance in business operations.	The model is divided into four parts: specification, planning, construction, and operation. The main elements of these parts include demand analysis, data architecture design, data quality monitoring, and data access auditing.

### 3.2.1. Functions and Effects of Data Curation

Combined with Table 4, the various conceptual data curation models vary across the applicable scenarios and model elements. In essence, when data curation theory is applied to data governance, the following functions and effects can be achieved:

- (1) In accordance with certain principles and processes, data curation constructs a top-level design covering the period starting from data generation to extinction. Data curation defines the work that should be carried out at each stage to ensure the orderly development of data governance and provides a guiding paradigm for data governance. The correctness and timeliness of the data governance activities and decisions can be improved significantly.
- (2) Data curation can establish standardized data governance norms through supervision, guidance, evaluation, and other tasks. Meanwhile, data curation can effectively encourage countries, industries, or companies to form a standard system that includes data access scope, data format, communication protocol standards, data naming, data update frequency, and other dimensions. Therefore, data quality can be improved from the dimensions of standardization, completeness, accuracy, timeliness, accessibility, and other dimensions, and large-scale multisource heterogeneous data from different institutions and industries can be integrated under the premise of effectively controlling governance costs.
- (3) Data curation can effectively clarify the roles of related individuals during the data governance and provide answers that allow individuals to take actions at specific stages, in specific situations, and for specific purposes. In this way, the interests of all the individuals can be balanced, cooperation can be organized effectively on the basis of protecting the rights of individuals, and access permissions can be guaranteed for individuals.
- (4) Data curation involves persistent work throughout the entire data life cycle. Through data product creation and data service innovation, data curation promotes the continuous development and maturation of data. Simultaneously, it provides broader, more reliable, and more valuable data and services to the individuals who are involved in data governance.

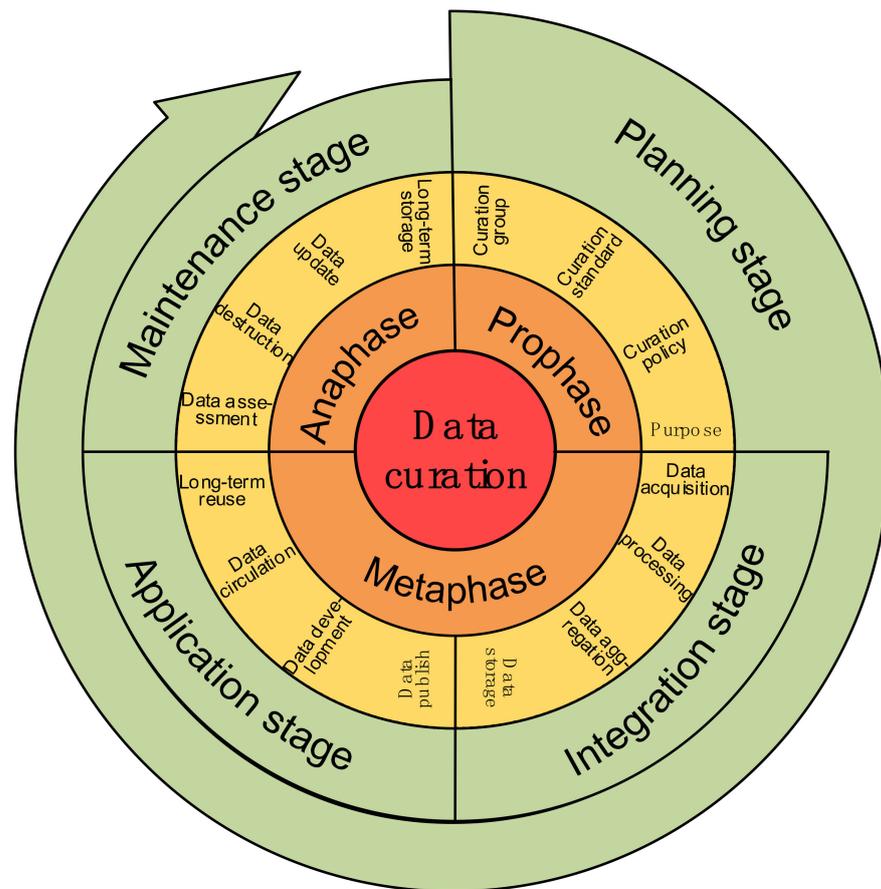
### 3.2.2. The Framework of Data Curation

Although various conceptual models of data curation exist (Table 4), the data curation framework can be divided into four parts from the viewpoints of the definition and effect of data curation, including planning, integration, application, and maintenance, as shown in Figure 4.

The planning stage aims to guide the implementation of data governance through effective systematic planning work. The entities participating in data governance and their objectives are complex and diverse. Therefore, during the data curation process, the planning stage is usually oriented based on the data and business needs of various entities, and this stage defines purposes, workgroups, boundaries, etc. Finally, it guides the development of data governance macroscopically by constructing a strategy.

The purpose of the integration stage is to integrate data resources using corresponding technologies to provide basic support for data applications. To ensure data consistency and efficient data sharing, the tasks that are usually carried out during this stage include data acquisition, data fusion, data aggregation, identification, and storage processing. The goal is to achieve heterogeneous data integration and, ultimately, to construct data resources for long-term accessibility.

The application stage explores the potential value of data through data publication, data development, data analyses, etc. This stage is usually guided by the needs of the data consumers and involves cleaning, mining, desensitization, encryption, large-scale storage, etc. The purpose is to achieve the long-term utilization and reuse of data resources in the form of freely available data transactions, data exchanges, etc.



**Figure 4.** The data curation framework.

In the maintenance stage, data are maintained according to the assessment results to ensure the quality of the data resources. This stage is considered the feedback node of data curation. During this stage, the data quality and the curation achievements are evaluated from the standpoints of accuracy, completeness, availability, timeliness, and scalability. Thus, the overall effort and state of the curated data can be obtained. In addition, weaknesses and deficiencies in curation and directions for improvement can be accurately identified.

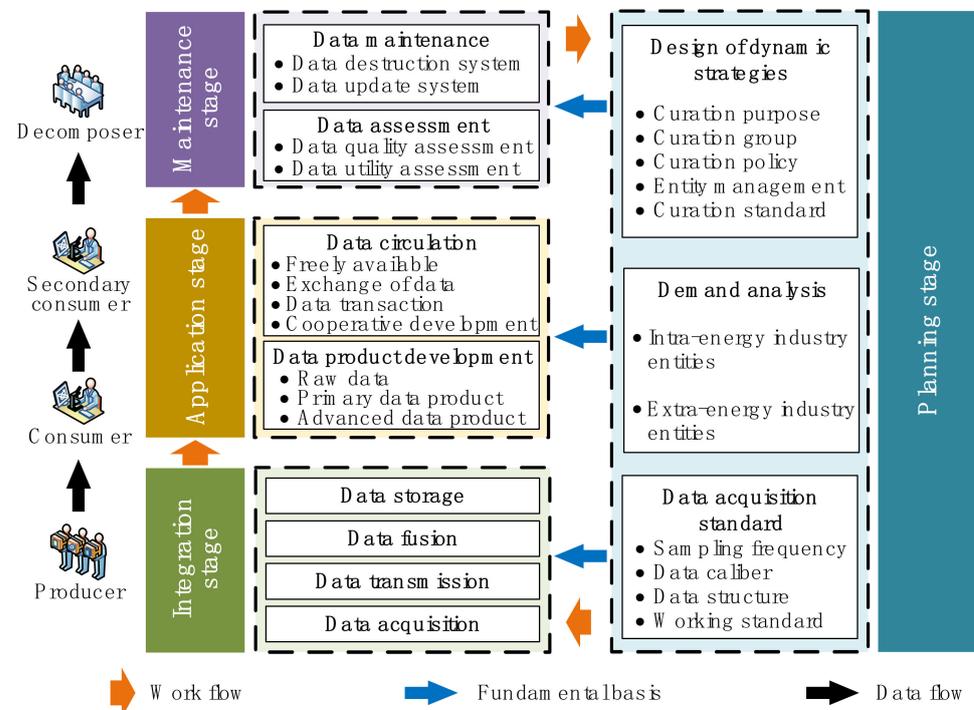
#### 4. A Paradigm for the Sustainable Full Life Cycle Curation of Energy Big Data

##### 4.1. General Framework of the Full Life Cycle Curation Model for Energy Big Data

Comparing the functions and effects of data curation (Section 3.2.1) and the critical challenges in the sustainable development of energy big data ecosystems (Section 2.3) indicates that data curation theory has good compatibility and application potential in the field of energy big data governance. By combining the basic concepts of data curation theory (Section 3.2) and the characteristics and challenges of the energy big data ecosystem (Sections 2.2 and 2.3), the curation model proposed in this paper covers the entire life cycle of energy big data and includes four main stages: planning (Section 4.2.1), integration (Section 4.2.2), application (Section 4.2.3), and maintenance (Section 4.2.4).

As shown in Figure 5, the planning stage is the starting point for energy big data curation. In this stage, the top-level design of the curation work is completed from the aspects of the design of dynamic strategies, a demand analysis, and a data acquisition standard. This stage constructs the overall guidance for the entire curation model. The integration stage collects and integrates the data of the producers in the energy big data ecosystem to provide rich data resources for data applications. In the application stage, the data value is explored and shared through data utilization and long-term reuse. This is an important

step in which data function as a production factor to achieve value externalization. The maintenance stage is the feedback and closed-loop point of the overall curation model. This stage updates or destroys the data according to the data quality and the results of the utility assessment.



**Figure 5.** Framework of the full life cycle curation model for energy big data.

Based on the above, the connection of the workflows in the four stages forms an energy big data governance paradigm that covers the entire data life cycle: generation, integration, preservation, application, reuse, transaction, destruction, and update (Figure 5); provides support for long-term storage, reuse, and updates for energy big data; and realizes the sustainable and dynamic governance of energy big data.

## 4.2. Implementing the Full Life Cycle Curation of Energy Big Data

### 4.2.1. Step 1: Planning Stage for the Full Life Cycle Curation of Energy Big Data

As shown in Figure 6, the planning stage involves three parts: dynamic strategy design, demand analysis, and data acquisition standards [35].

Dynamic strategy design aims to complete the top-level design of the full life cycle curation model for energy big data. The development of the energy big data ecosystem is in its infancy. Therefore, in a manner different from the traditional data management strategy solidified in the planning stage, energy big data curation work should be based on the ecosystem's entity structure (e.g., data producers and consumers), the demands of data consumers, different data types, etc. Then, combining the top-down expert-driven model with the bottom-up participant-driven model, dynamic curation strategies and models are constructed from the perspectives of the curation purposes, curation groups, curation policies, curation standards, and ecosystem entity management.

Reasonable data curation boundaries are the key to the sustainable development of the full life cycle curation of energy big data. Investment in the sensors, meters, servers, and other equipment involved in data access, perception, acquisition, transmission, storage, and analysis can significantly increase the cost of energy big data curation. Therefore, a demand analysis that is aimed at potential data consumers can effectively promote the precise connections between supply (data and data products) and demand. This work provides a reference for determining the data curation boundaries and for controlling and

compressing curation costs. In addition, potential data consumers can be divided into intra- and extra-energy industry entities. Intra-energy industry entities include thermal power plants, power grid companies, gas companies, etc. Extra-energy industry entities include governments, research institutions, colleges and universities, etc.

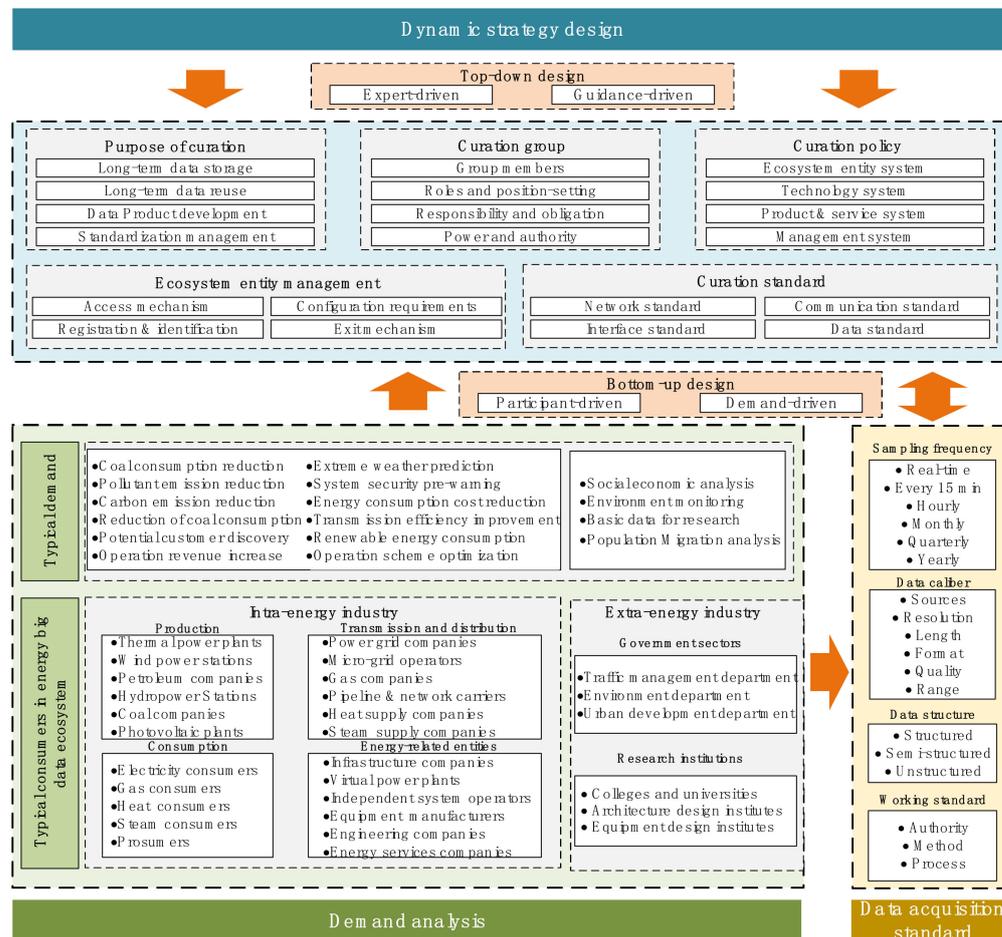


Figure 6. Schematic diagram of the planning stage.

In the data acquisition standards layer, based on the curation strategy and demand analysis results, the data acquisition standards for data curation can be formulated according to dimensions such as data sources, data structure, sampling frequency, etc.

Specifically, to coordinate the diverse demands and goals of the entities in the energy big data ecosystem (Figure 6) and to complete the curation flow and workflow, such as data curation boundary determination, strategy and plan making, standard system settings, etc. (Figure 5), a curation working group needs to be formed during the planning stage to handle the connection between the data curation process flow, workflow, and the goals of sustainable energy big data curation from the full life cycle of the data. The curation working group can be formed through various modes, such as through direct curation by the government, the collaborative autonomy of ecological entities under government supervision, or the establishment of third-party public institutions (industry associations, research institutes and universities, etc.).

#### 4.2.2. Step 2: Integration Stage of the Full Life Cycle Curation of Energy Big Data

The integration stage aims to build a data asset pool that aggregates multiple data sources and that can be continuously accessed and scheduled during the application stage. As shown in Figure 7, this stage can be divided into four layers: data acquisition, data transmission, data fusion, and data storage [36].

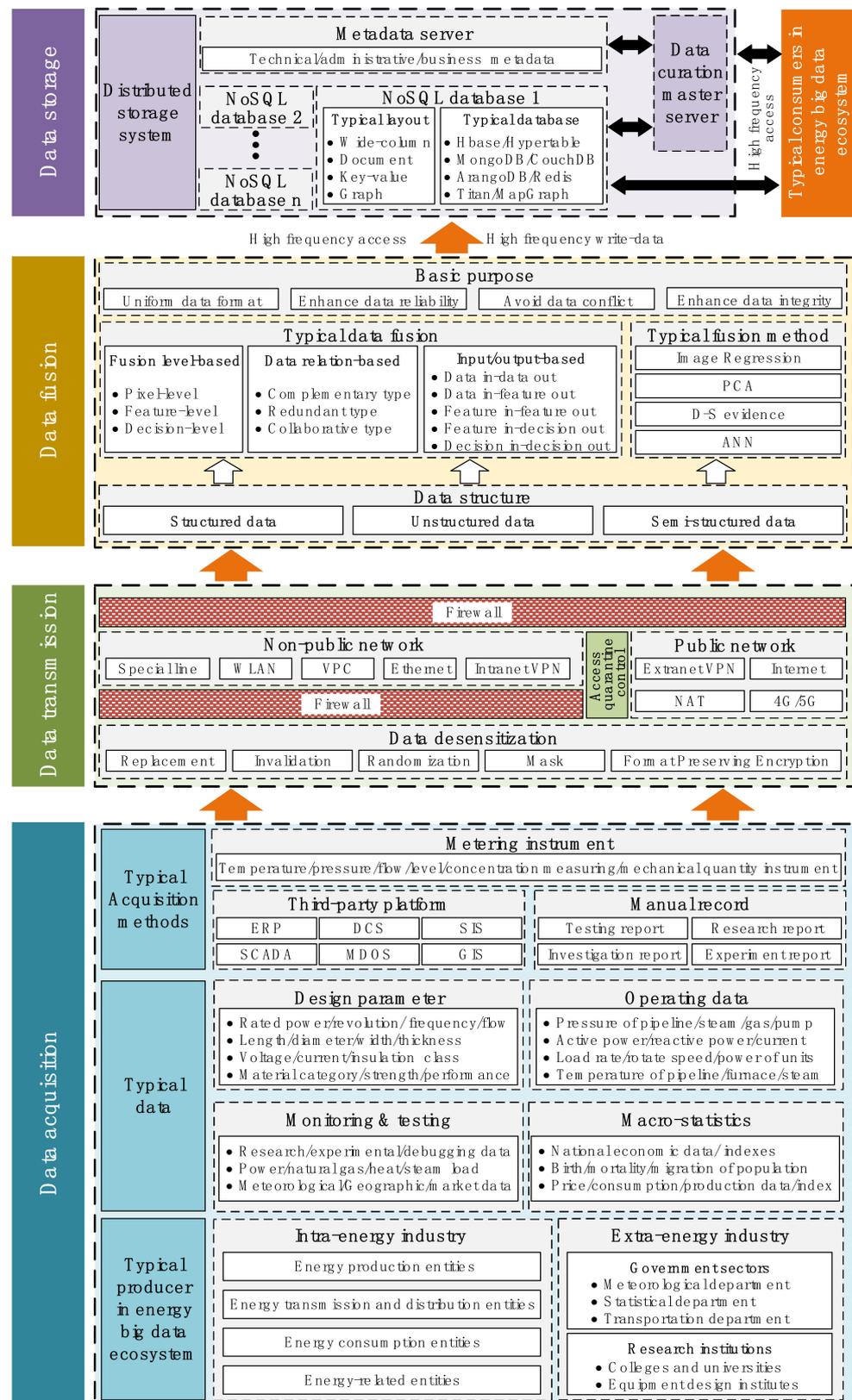


Figure 7. Schematic diagram of the integration stage.

The data are obtained from equipment, grids, the energy market, the macro-economy, the weather, etc., by sensor systems, manual recordings, and third-party platforms. The acquired data, which are extracted from relatively independent data sources and systems

belonging to different entities, are heterogeneous. Guided by the strategies and standards established in the planning stage, the data fusion layer conducts data-mapping and fusion operations to coordinate the data format and semantic conflicts for the structured, semi-structured, and unstructured data that have been extracted from various data sources, such as numbers, natural language, images, etc. This approach can shield the underlying differences that exist among data sources that are caused by multivariate heterogeneity and can improve the value density of the data.

Furthermore, it is difficult for traditional, centralized, relational data-storage models for data management to adapt to the requirements of large-scale and strongly heterogeneous data. Instead, a “logically unified, physically distributed” storage model can be applied for the large-scale storage of energy big data. For example, large-scale storage can use nonrelational database technology (NoSQL) to fragment the integrated data to corresponding node servers at the edge of the energy big data ecosystem. Furthermore, a distributed server cluster can be utilized to build a unified virtual data asset pool that effectively meets the data access and data query requirements at different nodes during the application stage while reducing data storage costs [37,38].

#### 4.2.3. Step 3: Application Stage of the Full Life Cycle Curation of Energy Big Data

The application stage is key to exploiting the value in and sharing of energy big data. Such applications should follow the FAIR data principles of “findable, accessible, interoperable and reusable” [39]. As shown in Figure 8, the application stage includes a data product development layer and a data circulation layer [40,41].

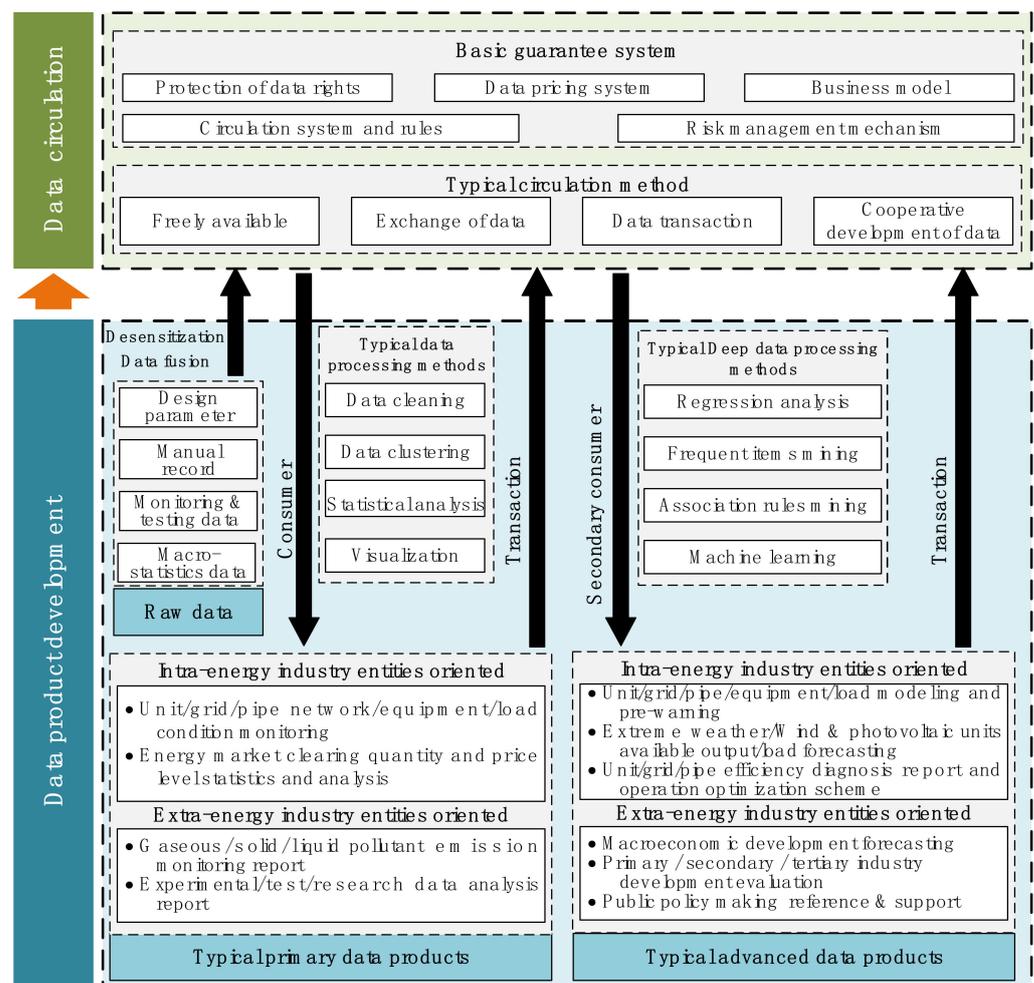


Figure 8. Schematic diagram of the application stage.

The data product development layer can be divided into primary and advanced data product development. Primary data products for circulation can be obtained through the cleaning, clustering, and visualization of the raw data that are imported and collected from energy suppliers and equipment, energy consumers and equipment, third-party platforms, etc. Some examples are energy market reports, pollutant emission monitoring reports, research data analysis reports, etc. [42]. Advanced products—such as load modeling, load forecasting, system security pre-warning, extreme weather prediction, operation optimization schemes, macroeconomic development forecasting, etc.—can be obtained through the in-depth development of primary data products.

Compared to traditional energy systems, data and data product circulation are important for achieving sustainable data applications and data reuse. To avoid the potential legal risks of data circulation and to restore the commodity attributes of data products [43], it is necessary to analyze the interest demands of energy consumers, energy suppliers, and governments as both data producers and consumers [44]. A frame of reference is needed for key issues in data circulation, such as subject matter confirmation, the price formation mechanism, and the division of responsibilities and rights. Therefore, a basic guarantee system should be constructed according to the data pricing system, business model, and risk management mechanism dimensions, among others.

#### 4.2.4. Step 4: Maintenance Stage of the Full Life Cycle Curation Model of Energy Big Data

Considering the risk of data value degradation, the primary and advanced data products in the energy big data ecosystem need to be maintained regularly. As shown in Figure 9, the maintenance stage is divided into a data assessment and a data maintenance layer.

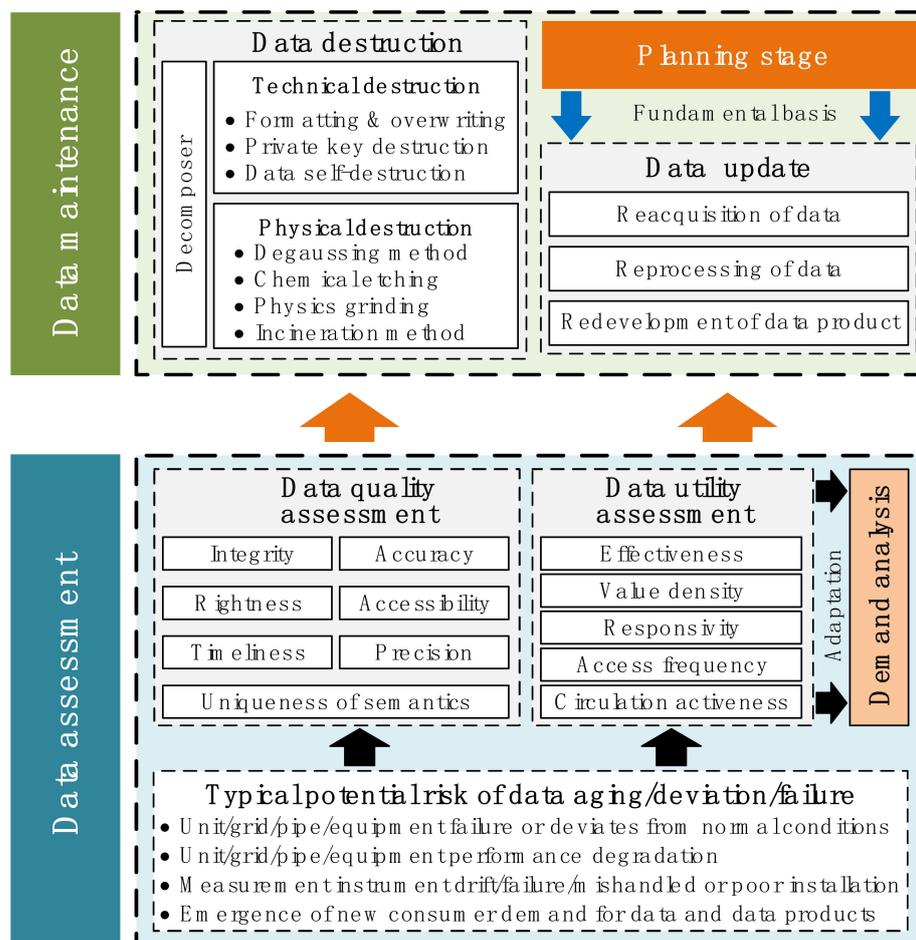


Figure 9. Schematic diagram of the maintenance stage.

The data assessment layer includes both quality and utility assessments. On the one hand, considering the potential risk of data aging, bias, and error due to deterioration in the operational characteristics of energy equipment and pipeline networks and the drift or failure of metering or sensing devices, quality assessment focuses on the quality of the data in terms of value, format, meaning, and structure to ensure the accuracy, integrity, accessibility, and semantic uniqueness of the data in the energy big data ecosystem [45,46]. On the other hand, given the enrichment of the data economy and data business systems in the energy big data ecosystem, any existing primary and advanced data products will not be able to adapt to the growing needs of different data consumers. Moreover, based on a full understanding of the needs of data consumers and considering the potential risk of data asset value degradation, a utility assessment should be conducted in terms of effectiveness, value density, responsiveness, and access frequency [47].

The data maintenance layer includes data update and destruction. According to strategies and standards formed in the planning stage, the data will be updated, replaced, and expanded through data reacquisition, reprocessing, and redevelopment based on the results of the data assessment in the data update module. Considering data security and privacy, data destruction mainly focuses on data that have reached the end of their life cycles, such as expired or obsolete data (such as historical power market transactions, expired nodes, and redundant backup data). Technical means, such as overwriting, secret key destruction, and physical destruction, are used to eliminate the data stored in the application, platform server, cloud server, etc. [48].

## 5. Challenges and Key Issues in the Full Life Cycle Curation of Energy Big Data

### 5.1. Data Rights

Data constitutes the critical production factor and has become the “oil” of the digital economy era. However, current data product development processes treat data ownership and use as completely separate states. This unclear definition and division of data rights may transform the real producers and owners of the data in the energy big data ecosystem into the “tenants” of the data controllers [49]. Clarifying the ownership of data-related rights and interests is an important aspect of data sharing, development, and transactions, and it is key to facilitating the sustainable development of energy big data ecosystems. Due to the different natures of data sources, energy big data can be divided into nonpublic data and public data [50].

#### 5.1.1. Nonpublic Data Rights

Nonpublic data are primarily generated by a series of activities, such as those through the production, operation, and consumption of various entities in the energy ecosystem. Rights to nonpublic data mainly include personality and real rights [51].

Personality rights research on data evolved from traditional privacy rights research [52]. Considering that energy big data include a substantial amount of sensitive information from actual data producers and given the progress in data mining and information technology techniques, it may be difficult to achieve totally irreversible data anonymization and desensitization [53]. Therefore, unlike traditional privacy rights, personality rights to data in the digital economy era should not be limited to defining and protecting private data. In addition, protecting the intent and rights of both the natural and the legal people who are the actual data producers should be emphasized regarding whether the data may be disclosed [54].

Real rights refer to the privilege of managing, using, and benefiting from data and highlight that data are a type of property in the digital economy era. As data are objective electronic records that contain information, the actual producer of the data should indeed enjoy corresponding property rights [55]. However, because the data controller or user expends money and labor to collect, process, and develop the data, the value of the data as a commodity increases accordingly [56]. Thus, adopting traditional exclusive real rights to grant complete property rights to either the original producers or the controllers may result

in blocked data circulation, reduced enthusiasm for data development, data monopolies, and other phenomena, thus leading to the distortion and failure of the data market [57]. Therefore, how to coordinate the conflict between the original data producers and the data controllers with regard to data real rights has become a key to ensuring the sustainable development of data ecosystems [58,59]. Some scholars have proposed addressing the rights conflicts between the original producers and the controllers by setting up a dual real rights mode (assigning nominal ownership to the original producer and actual ownership to the controller) or by independently creating restrictive data producer rights. However, the above approaches may lead to an ambiguous division and attribution of real rights and may even conflict with existing copyrights and intellectual property rights, thus aggravating disputes over data property rights [60,61]. The structure of data real rights: “ownership + usufruct”, which are based on the idea of the division of rights in property law, can be reasonably derived and distributed between the original producer and multiple controllers without violating the traditional real rights framework [62]. Ownership constitutes the most comprehensive control over data; it includes the privilege of possession, use, income, and disposal. Ownership clarifies the attribution of data as objectively existing in insubstantial objects and provides an important legal basis for data availability and circulation. As a derivative right of ownership, usufruct stresses the privileges of data control, development, and transfer. Usufruct provides an important legal basis under which processors can control data. Thus, it can effectively improve the data allocation efficiency, which is a valuable resource in the digital economy era [53].

#### 5.1.2. Public Data Rights

Public data—such as geographic, meteorological, municipal, and macroeconomic data—are strategic resources that are related to national security and social economic development. In the process of openness and circulation, research on public data rights has focused on its sovereignty attributes. As an extension and representation of national sovereignty in the digital economy era, the sovereignty of public data, which includes the rights of possession, jurisdiction, use, and disposal, has a certain exclusivity [63]. Strengthening the awareness of data sovereignty is conducive to improving the protection and development of national strategic data resources. However, it is worth pointing out that overemphasizing data sovereignty may aggravate the zero-sum game, foster antagonistic behavior between countries in the digital field, and hinder the flow and allocation of digital resources between countries [64].

In terms of data sovereignty, in the process of making data publicly available, two different solutions have been proposed by the United States and the European Union at the present stage.

Although the United States stresses that cyberspace is a global commons, the U.S. position is that countries around the world should minimize interference in the flow of data across borders to achieve liberal data sharing. The *Clarifying Lawful Overseas Use of Data Act* issued by the U.S. not only grants corresponding rights and makes it convenient for the U.S. government to acquire public data outside of its the territory through overseas US-owned electronic communication companies and computing service companies but also sets a series of draconian rules regarding access to U.S. public data by foreign governments. In part, the U.S. adopts a double standard on the issue of public data sovereignty mechanisms and the cross-border flow of public data [65]. In contrast, the European Union (EU) position is that data sovereignty should be defined based on the location where the data are physically stored, emphasizing that the data produced in a country should be stored locally and that management and protection should be strengthened in cross-border data circulation scenarios [66]. To this end, the EU has issued a successive series of policies and laws, including the *General Data Protection Regulation*, *The White Paper on Artificial Intelligence*, and *A European Strategy for Data*. These papers clarify the authorization, transmission, evaluation and supervision, and disposal mechanisms and

the rules that should be followed when granting public data access and sharing to help ensure the integrity and independence of data sovereignty [67].

Competition exists in relation to public data protection and regulation and the sharing and liberalization that must be balanced with regard to the sovereignty of the cross-border flow of public data. To do so, on the one hand, countries should seek opportunities for competitive cooperation; follow win–win cooperation principles; promote the construction of multilateral cooperation mechanisms, laws, and regulations for public data co-governance; and urge countries to assume the corresponding responsibilities and obligations while still guaranteeing their rights to enjoy public data development. On the other hand, the development and application of the blockchain, data watermarking, and other digital technologies to establish public data sovereignty, access authorization, data traceability, data destruction, and other scenarios should be accelerated to provide technical support for the sustainable development of public data co-governance involving multilateral cooperation.

## 5.2. Data Fusion

The increased diversity of data sources means that energy big data ecosystems will accumulate large volumes of heterogeneous data in the spatiotemporal dimension. The fusion processing of data sources and their data at various stages can provide strong support for improving data quality and perfecting knowledge graphs. Data fusion can be subdivided into intra- and inter-source types.

### 5.2.1. Inter-Source Data Fusion

During the period in which equipment performance degradation can be ignored, the measurement data, soft measurement data, or related characteristic indexes (such as system carbon emissions, unit coal consumption, and equipment efficiency) of a single data source (such as a microgrid system, a generator, or equipment) in the energy big data ecosystem should be stable within a certain range over the same sampling period, the same typical day, or under the same working conditions.

However, many factors, such as the transient states and unsteady operations of systems and equipment, differences between meter installation positions, and meter drift or faults, may lead to data deviations or even serious errors. Consequently, the historical data of the corresponding data source for the same sampling period and obtained during the same typical day or working conditions should be fused along the time dimension to remove incorrect data, reduce data conflicts, and improve data quality and reliability [68].

### 5.2.2. Intra-Source Data Fusion

In general, entities in the energy big data ecosystem only collect and preserve the data related to their own needs. Therefore, data fusion among data sources in the energy big data ecosystem can achieve interaction and supplementation between different data sources. Thus, data fusion effectively alleviates the “isolated island of information” phenomenon and provides support for knowledge graph improvement, the exploration of the potential value of data, and data service expansion. Considering that the development, modeling, and operation of the information systems of entities in the energy big data ecosystem are relatively independent (e.g., the supervisory control and data acquisition (SCADA) systems used by power grids, the supervisory information system (SIS) used by power plants, and the meteorological data operation system (MDOS)), the heterogeneous data generated by multiple data sources must be fused along the spatial dimension at three levels: pixel, feature, and decision [69].

Among these, pixel layer fusion collects original data from associated data sources through the design and authorization of interface interaction and communication specifications, as well as through the construction of a unified metadata mapping framework. For example, historical datasets are constructed by integrating operating data from equipment with the same type but located in different regions and operating under different working conditions. Feature layer fusion is primarily aimed at extracting multidimensional features

such as coordinates, power, temperature, and pressure from the original data and using them to construct a conceptual model or knowledge map through semantic data fusion. For example, the equipment operating model for the overall working conditions based on the historical dataset can be constructed through feature layer fusion. Decision-level fusion aims to obtain a decision set that has high consistency and that is robust and generalizable by fusing individual decisions from multiple data sources. For example, the fusion of the operation strategies extracted from equipment operating models for the overall working conditions can provide a reference for the design and manufacture of the equipment and its operation optimization.

### 5.3. Data Security

Energy big data are an important asset and function as an information carrier in the digital era. Ensuring the security of data circulation is a key issue in constructing the full life cycle curation of energy big data. Blockchain technology has the characteristics of openness, traceability, and immutability. Thus, it can be used to create smart contracts that support the establishment of secure, reliable, and sustainable energy big data ecosystems from two dimensions: data encryption and access permission administration [70–73].

#### 5.3.1. Data Encryption

Ensuring the consistency, accuracy, and security of primary and advanced data products is the basic requirement of energy big data circulation. When a node or entity in the energy big data ecosystem receives and confirms a data-sharing or transaction request, a trigger can automatically take action. Such actions include invoking a corresponding smart contract to encrypt the data to be accessed according to a preset algorithm. Then, the encrypted data and address will be generated and published to the energy big data ecosystem. Subsequently, other nodes or entities can decrypt the data to obtain the plaintext (unencrypted string) alone by obtaining the corresponding smart contract. Such strategies effectively reduce the risk of data leakage or tampering during circulation [72].

#### 5.3.2. Access Permission Administration

To reduce the risk of illegal access to energy big data, access permissions can be controlled via secret key management, accessing entity control, and the scope of accessible data control.

In terms of secret key management, data owners and users (consumers) can obtain a unique-identity secret key by sending a registration application to the certificate authority (CA), and the key can provide support for entity encryption and identification. In addition, data owners can use encryption algorithms such as SM4 to encrypt data files, data addresses, and the symmetric keys that provide support for data encryption and identification [74].

From the point of view of accessing entity control, the access list contract is designed so the data owner can adjust the access list contract by invoking functions. Moreover, the use of an access list contract can gain the capability of accessing entity control by flexibly adding or canceling the authorization of the accessing entities.

In terms of the scope of accessible data control, based on the registration contract and data contract, constructor functions can be used to write the secret keys, identity information, and the data of each node or entity into the blockchain. Furthermore, a one-to-many mapping relationship can be constructed between registration contracts and data contracts, which allows the data owner to modify the authorization list to adjust the range of accessible data [73].

### 5.4. Data Transaction

Data transactions allow data to participate in the allocation of social resources by having them act as a new production factor. Data valuation, transaction subject matter, and price mechanisms are three of the key issues that are in data transactions.

#### 5.4.1. Data Valuation

Data valuation is important for pricing the subject matter involved in data transaction. However, data are a nontraditional asset and a nonentity. The subjective nature and uncertainty of its utility, the high cost of initial production but extremely low replication costs, and other characteristics make it difficult to quantify the value of data assets. Consequently, no standardized or unified data valuation method has been formed yet. By considering the characteristics of data assets and drawing on the valuation methods for both intangible and tangible assets, three methods can be used for data valuation: comprehensive evaluation-based, cost-based, and comprehensive utility-based data valuation.

Comprehensive evaluation-based data valuation builds an evaluation system that represents the data quality, scale, function, reputation, risk, and other dimensions. Then, corresponding experts or data consumers can conduct qualitative or quantitative evaluations based on the evaluation system indicators. The analytic hierarchy process (AHP), entropy weight method, and other methods are used to determine the weight of each indicator. The value of the data assets can be determined by calculating the total evaluation score. Furthermore, this method can overcome the subjective impact of data asset utility on the value evaluation [75].

The cost-based data valuation method makes a quantitative valuation. In doing so, it considers the market leverage effect, inherent value, and other factors based on calculating the labor, equipment, and other costs needed to replace each dataset at the time of the data valuation. This method is able to counteract the data value dynamics [76].

The comprehensive utility-based method first evaluates the cost of data replacement and then represents the data value by superimposing the difference in benefits after and before data products are applied by data product consumers, as shown in Equation (1). This method is better able to overcome the influences of data asset utility uncertainty and subjectivity on data valuation and to reflect the value of the data as a commodity while reflecting the cost of the data assets (we assume that the data subsets are mutually exclusive).

$$V_U(D) = \sum_{i=1}^N c_i + \sum_{i=1}^N P_{i,j} \cdot \Delta U_{i,j} \quad (1)$$

where  $V_U(D)$  is the value of data set  $D$  based on the comprehensive utility-based method,  $c_i$  is the replacement cost of  $i$ , which is a subset of  $D$ , and  $P_{i,j}$  and  $\Delta U_{i,j}$  are the probability and difference of benefit generated by data consumer  $j$ 's consumption of data product  $i$ .

#### 5.4.2. Subject Matter

Clarifying which data transactions are being traded—that is, the subject matter of the data transactions—is the prerequisite to setting the prices of the data transactions and constructing a data transaction market. At present, most studies have diversified definitions of data subject matter. These include raw data, information, model algorithms, decision methods, technical services, etc. In essence, this is a paid transfer of rights between buyers and sellers. Consequently, combined with the analysis in Section 5.1, the subject matter of data transactions is data rights, which mainly include the rights of ownership, use, and usufruct.

Ownership is the right that allows one to have the most comprehensive control over data. When the transaction subject matter is ownership, the data buyer will obtain the right to possess, use, derive benefit from, and dispose of data. The production of data assets involves multiple entities. Therefore, when “rights of use” is the subject matter of the transaction, on the one hand, after the purchase, the purchaser can use the data without changing the nature of the data assets. On the other hand, ownership disputes between entities with regard to the same data set can effectively be avoided to a certain extent [77]. The situation is different from the rights of use, which is the case when the transaction subject matter is usufruct. Then, the data purchaser gains the right to utilize the data and generate income from the data (i.e., by mining and reprocessing the purchased data).

In summary, different subject matter can be regarded as methods of disassembling data rights. That is, the subject matter of the transaction presents a certain combination of rights. Therefore, in combination with the data valuation from Section 5.4.1, the Shapley value can be calculated based on Equation (2) to realize the value allocation and accounting of different types of subject matter, thus providing support for the value formulations of different types of subject matter.

$$V_r = \sum_{r \in R} \frac{(r - |R|)! (|R| - 1)!}{r!} \cdot [V(R) - V(R - \{i\})] \quad (2)$$

where  $R$  is the set of subject matter (data rights),  $r$  is the subject matter that is a subset of  $R$ , and  $V_r$  is the portion of the total value that can be attributed to it.

#### 5.4.3. Pricing Mechanisms

A scientific and reasonable price mechanism should reflect the supply and demand relationship of the market effectively, and to a certain extent, it should guarantee the benefit and profit spaces of the buyers and sellers. Considering the complexity of data value quantification and the dependence of price formation on specific trading scenarios, forming a unified and standardized pricing mechanism is difficult [78,79].

From the perspective of transaction price formation, the price of data transactions can be determined through bilateral negotiation, usage measurements (statistics such as the number of data accesses and data flows, etc.), public auctions, and listing transactions [80–82]. Bilateral negotiations, public auctions, and listing transactions are widely used and are applicable to the three types of subject matter transactions mentioned in Section 5.4.2. However, price formation based on the use of statistics is typical on a per-time basis and is applicable to repeatable transaction scenarios. Therefore, this method is only applicable to rights of use [77].

From the perspective of product pricing methods, the price of the subject matter can be determined by pricing methods that are based on expected revenue, information entropy, game theory, and market querying [83–85].

A pricing method that is based on expected revenue is a typical cost-oriented pricing method in the accounting field and can better reflect the seller's expectation and intent. This method is based on calculating the production cost, and it forms the data transaction price by setting a reasonable expected return rate for the seller, as shown in Equation (3). A pricing method based on information entropy is a typical product value-oriented pricing method that can reflect the uniqueness and effectiveness of data products as commodities. As information entropy is non-negatively correlated with the volume and effective information amount of data products, this method calculates the price of data transactions by constructing an appropriate nondecreasing contact function and by calculating the information entropy of data products as shown in Equation (4). A pricing method based on game theory is a market-oriented pricing method that addresses the information asymmetry between buyers and sellers, the depreciation and obsolescence of data assets, and other problems in data transactions. With this method, buyers and sellers take turns bidding and finally reach Nash equilibrium; then, the transaction price is formed. A transaction price based on the Rubinstein model is shown in Equation (5). The pricing method based on market inquiry is a typical consumer demand-oriented pricing method with anti-arbitrage, discount-free, and timeliness characteristics. This method generates pricing based on the view generated by any query behavior demonstrated by the buyer to achieve dynamic differentiated pricing [84].

$$p_\varepsilon(D) = \sum_{i=1}^D c_i^0 \cdot (1 + \varepsilon) \quad (3)$$

$$\begin{cases} p_E(D) \equiv I(E(D)) \\ E(D) = - \sum_{d_i \in D} P(d_i) \cdot \log_b P(d_i) \\ \forall x_1 \geq x_2, I(x_1) \geq I(x_2) \\ \forall x_1, x_2 \geq 0, I(x_1 + x_2) \leq I(x_1) + I(x_2) \end{cases} \quad (4)$$

where  $p_E(D)$  is the price of  $D$  based on the information entropy-based pricing method,  $I(\cdot)$  is a nondecreasing contact function,  $E(D)$  is the information entropy of  $D$ ,  $d_i$  is the subset of  $D$ ,  $P(d_i)$  is the probability that  $D$  contains  $d_i$ ,  $b$  is the base, and the data are measured while  $b$  is equal to 2.

$$p_G = p_{sell,min} + \frac{1 - \delta_{buy}}{1 - \delta_{sell} \cdot \delta_{buy}} \cdot (p_{buy,max} - p_{sell,min}) \quad (5)$$

where  $p_G$  is the price of  $D$  based on a game-based pricing method,  $p_{sell,min}$  is the lowest price acceptable to the data seller and can be represented by the cost of data,  $p_{buy,max}$  is the highest price acceptable to the data buyer and can be represented by the benefits of data consumption, and  $\delta_{sell}$  and  $\delta_{buy}$  are the discount factor of the seller and buyer. These factors represent the endurance of both the seller and the buyer when participating in the game.

## 6. Conclusions and Discussion

### 6.1. Conclusions

In the context of the deep integration of internet technology and the energy industry, the long-term preservation, reuse, and update of energy big data are important manifestations of the sustainable development of energy big data governance, and a scientific management paradigm is important for facilitating the sustainable development of energy big data governance. Based on this, this manuscript first analyzes the structure and mechanism of the energy big data ecosystem and proposes the challenges faced during the sustainable development of the energy big data governance. Secondly, based on data curation theory, this paper proposes a governance paradigm that covers the entire life cycle of energy big data. Finally, the key issues in the life cycle curation of energy big data are analyzed and discussed, including data rights, data fusion, data security, and data transactions.

Compared to the current practices and research, the main conclusions and contributions of this paper are as follows:

- (1) In terms of research, according to the data flow and life cycle, this paper shows that energy production enterprises, equipment manufacturers, and other entities constitute producers, consumers, secondary consumers, and decomposers in the energy big data ecosystem and that there are predation, competition, reciprocity, and other relationships among different entities around energy data. Different from ecology, a given entity in the energy big data ecosystem can switch between the roles of the producer and consumer in different situations. In contrast, previous research on energy big data governance is mainly oriented toward specific enterprises and industries. This paper provides a new perspective for energy big data governance research because it considers it from the macroscopic and systematic dimensions according to the ecosystem concept.
- (2) In terms of research methods, on the basis of analyzing applicability, this paper introduces curation theory, which originated from the fields of library science and information science, into energy big data governance. The model that was constructed in this paper based on curation theory can provide systematic and theoretical support for workflow organization, specification formulation, and interest coordination among different entities as well as full life cycle management in the process of energy big data governance, and this enriches the sustainable methods for energy big data governance.
- (3) In terms of research content, this paper presents three energy big data curation challenges, namely how to carry out the ecosystem-oriented top-level design for data

governance; how to determine the boundaries of energy big data governance; and how to balance the relationships among various interests in the ecosystem. In view of the above problems, this paper proposes a governance paradigm that covers the entire energy big data life cycle according to the planning, integration, application, and maintenance stages; furthermore, this research analyzes four key issues of full life cycle curation, including data rights, fusion, security, and transactions: (1) in contrast, non-public data rights emphasize personality and real rights, and public data rights emphasize sovereignty. (2) In order to improve data quality and accuracy, inter-source data fusion should be processed from the time dimension; in order to alleviate the “isolated island of information”, intra-source data fusion should be processed from the spatial dimension at the pixel layer, feature layer, and decision layer. (3) The encryption of the data itself, the data address, and the secret key can be achieved by using the blockchain to create smart contracts. The application of this technology can reduce the risk of unauthorized access and tampering, effectively enhancing data security. (4) The valuation and confirmation of the subject matter are the premise of data transaction. The three value evaluation methods, the comprehensive evaluation-based, cost-based, and comprehensive utility-based methods, can effectively reflect the objectivity and dynamics of data value; the essence of data transactions is the paid transfer of data rights. According to different transaction scenarios, the subject matter can be a combination of ownership rights, use rights, and usufruct rights. Based on this, diversified pricing mechanisms can be used for data pricing, including bilateral negotiations, public auctions, expected revenue-based pricing methods, etc. These pricing methods can effectively overcome the difficulties in formulating price mechanisms due to the complexity of the data value and subject matter.

## 6.2. Discussion

Data rights, fusion, security, and transaction are the key issues affecting the sustainable governance of energy big data. The manuscript analyzed and studied these key issues in terms of the composition rights of public and non-public data, inter-source and intra-source data fusion, data encryption and authority control methods, value evaluation and price mechanisms, etc. However, this manuscript mainly focused on the curation model construction and the key issues involved in curation flow from a theoretical perspective. In order to connect the energy big data curation paradigm constructed in this paper with the real energy big data system, future research and practices can focus on the following aspects:

- (1) In terms of laws and regulations, in addition to the “invisible hand” of the market, the development of energy big data in the digital economy era also needs the support of the “visible hand” of the government. In the future, relevant laws and regulations on data curation should be studied and formulated from the perspectives of rights protection, access authority, market mechanisms, arbitration methods, and risk control.
- (2) In terms of organizational mechanisms, considering the complex role positioning of the various entities in the energy big data ecosystem, the diverse governance goals, and the potential conflicts of interest, in the early stages, a joint curation working group should be set up by energy industry companies under government supervision and gradually promote the development of energy big data curation at the regional (provincial), industrial, and national levels.
- (3) In terms of professional talent cultivation, energy big data curation requires professionals. Unlike traditional data administrators, data curators need to assume more complex roles, including those of planners, policy makers, information technology specialists, and researchers [86]. In addition to the ability for data archiving and preservation, database management and maintenance, etc., energy big data curation requires considerable knowledge of the law and the energy industry. Such knowledge includes intellectual property law, integrated energy system planning

and operation optimization theory, energy demand-side management theory, etc. As interdisciplinary integration becomes a more prominent trend, it is necessary to establish and improve the education and cultivation for energy big data curation talents in the future.

- (4) In terms of information platform development, as an important physical carrier for the curation of energy big data, information platforms should realize the functions of access subject management, cross-system data transmission and integration, and data transaction aggregation on the premise of guaranteed security and privacy. How to design and develop a platform suitable for regional, industrial, national, and international levels in terms of communication protocols, interfaces and ports, model algorithms, etc., is one of the key issues for future research and practices.

**Author Contributions:** Conceptualization, M.Z. and Y.X.; methodology, Y.X.; software, J.M. and Y.X.; data curation, J.M.; writing—original draft preparation, Y.X., H.W. and J.M.; writing—review and editing, Y.X., H.W. and J.G.; visualization, H.W. and J.M.; supervision, M.Z. and Y.X.; project administration, M.Z.; funding acquisition, M.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Major Program of the National Social Science Fund of China (19ZDA081).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** This work was supported by the Major Program of the National Social Science Fund of China (19ZDA081).

**Conflicts of Interest:** The authors declare that they have no known competing financial interest or personal relationships that could appear to influence the work reported in this paper.

## References

- Guo, L.; Dong, J.; Chen, Z.; Bao, N.; Wang, Y.; Wu, C.; Wu, Y.; Xue, G. Business Model Evaluation of Energy Big Data Value-added Services Based on Entropy Weight-Topsis-Grey Correlation Method. *Sci. Technol. Manag. Res.* **2022**, *42*, 73–80.
- Wang, X.; Chen, A.; Li, J.; Zheng, C.; Pan, X.; Yang, Z. Research on Data Business Operation Mode Based on Energy Big Data Center. *Distrib. Util.* **2021**, *38*, 37–42.
- Chen, Q.; Liu, D.; Lin, J.; He, J.; Wang, Y. Business Models and Market Mechanisms of Energy Internet. *Power Syst. Technol.* **2015**, *39*, 3050–3056.
- Chen, R.; Li, H.; Peng, X.; Yang, J.; Dong, X.; Liu, S.; Li, X. Study on Evaluation Method for New Energy Big Data Service Project Applying Improved TOPSIS. *Electr. Power Constr.* **2021**, *42*, 126–134.
- Zhao, Y.; Yuan, S.; Chen, Y.; Qian, C.; Xu, H. Design of Internet power information management system for energy in the park based on big data analysis. *Autom. Instrum.* **2021**, *10*, 169–173.
- Zhang, X.; Tang, C. Study on the construction of the framework of metadata standards for government information in China. *J. Inf. Resour. Manag.* **2018**, *8*, 25–36.
- Han, W.; Wang, T.; Peng, J.; Wu, G.; Zhang, X.; Zhang, M.; Cao, N.; Ma, H. Research on Standard System Framework of Power Transmission and Distribution Project Data Management Technology. *Sci. Technol. Manag. Res.* **2018**, *38*, 224–229.
- Xu, Z.; Wang, Z.; Chi, R.; Hong, Y.; Chi, M.; Yi, W. Equalization Treatment Technique for Inverted Secondary Index Cluster of Quasi-real-time Data in Distribution Network. *Proc. CSEE* **2020**, *40*, 6494–6506.
- Sun, C.; Xiao, W.; Zeng, L.; Bai, J. Design and Implementation of Massive Surveillance Data Cloud Storage Service Model. *Geomat. Inf. Sci. Wuhan Univ.* **2020**, *45*, 1099–1106.
- Mu, L.; Cu, L.; An, N. Research and Practice of Cloud Computing Center for Power System. *Power Syst. Technol.* **2011**, *35*, 171–175.
- Gong, Y.; Lv, J. Application of Big Data Mining Analysis in Power Equipment State Assessment. *South. Power Syst. Technol.* **2014**, *8*, 74–77.
- Qu, H.; Pang, X.; You, M.; Xu, Z. Application value of power big data and analysis and design of sharing platform. *Manag. Adm.* **2017**, *7*, 104–108.
- Zeng, M.; Yang, Y.; Li, Y.; Zeng, B.; Cheng, J.; Bai, X. The Preliminary Research for Key Operation Mode and Technologies of Electrical Power System with Renewable Energy Sources Under Energy Internet. *Proc. CSEE* **2016**, *36*, 681–691.

14. Xue, S.; Lai, Y. Integration of Macro Energy Thinking and Big Data Thinking Part Two Applications and Exploration. *Autom. Electr. Power Syst.* **2016**, *40*, 1–13.
15. Xue, S.; Lai, Y. Integration of Macro Energy Thinking and Big Data Thinking Part One Big Data and Power Big Data. *Autom. Electr. Power Syst.* **2016**, *40*, 1–8.
16. Feng, D.; Zhang, M.; Li, H. Big Data Security and Privacy Protection. *Chin. J. Comput.* **2014**, *37*, 246–258.
17. Yu, X.; Xu, X.; Chen, S.; Wu, J.; Jia, H. A Brief Review to Integrated Energy System and Energy Internet. *Trans. China Electrotech. Soc.* **2016**, *31*, 1–13.
18. Wang, J.; Li, C.; Zheng, Y.; Chen, L. Design of power grid clean energy three base construction system based on big data platform. *Autom. Instrum.* **2019**, *5*, 44–47.
19. Li, L.; Xu, Z.; You, M. Power big data trading for the future smart grid. *Manag. Adm.* **2018**, *2*, 121–124.
20. Zhang, D.; Miao, X.; Liu, L.; Zhang, Y.; Liu, K. Research on Development Strategy for Smart Grid Big Data. *Proc. CSEE* **2015**, *35*, 2–12.
21. Sun, L.-L. Knowledge innovation-oriented knowledge ecosystem model construction for e-commerce enterprises. *Intell. Sci.* **2016**, *34*, 143–146.
22. Zhang, P.; Li, Q.; Zhang, J. A collaborative knowledge evolution model of supply chain enterprises based on ecological population perspective. *Intell. Sci.* **2016**, *34*, 150–153.
23. Wang, G. Digital scholarly journal knowledge ecosystem and its evolutionary motives. *Mod. Intell.* **2012**, *32*, 28–31, 43.
24. Li, T. Research on the mechanism of micro knowledge ecosystem operation in smart libraries. *Intell. Sci.* **2019**, *37*, 133–137.
25. Zhang, M.; Huo, C.; Wu, Y. A study on the evolution of knowledge ecosystem in international digital libraries. *Library* **2015**, *10*, 88–93.
26. Yang, H. UK data guardianship research results and their application in university libraries—A review of DCC construction. *Libr. J.* **2014**, *33*, 84–90.
27. Beagrie, N.; Pothan-Ariadne, P. The Digital Curation: Digital Archives, Libraries and e-Science Seminar. 2001. Available online: <http://www.ariadne.ac.uk/issue/30/digital-curation/> (accessed on 20 October 2021).
28. DCC. What Is Digital Curation? Available online: <https://www.dcc.ac.uk/guidance/briefing-papers/introduction-curation/what-digital-curation/> (accessed on 2 October 2021).
29. Laughton, P. OAIS functional model conformance test: A proposed measurement. *Program Electron. Libr. Inf. Syst.* **2012**, *46*, 308–320. [CrossRef]
30. Wallis, J.C. Moving archival practices upstream: An exploration of the life cycle of ecological sensing data in collaborative field research. *Int. J. Digit. Curation* **2008**, *3*, 114–126. [CrossRef]
31. DAMA International. *The DAMA Guide to the Data Management Body of Knowledge*; Technics Publications: New York, NY, USA, 2009; Volume 37.
32. Bao, D.; Fan, Y.; Li, M. Data governance and its framework in higher education libraries. *Libr. Intell. Work.* **2015**, *59*, 134–141.
33. Data Governance Framework. Available online: <https://datagovernance.com/data-governance-framework-components/> (accessed on 20 October 2021).
34. Xiao, J.; Feng, G. A comparative analysis of domestic and foreign data governance models. *J. Lit. Data* **2020**, *2*, 14–25.
35. Chu, J.; Wang, M. The strategy of open sharing of scientific data in the United States and the inspiration for China. *Intell. Theory Pract.* **2019**, *42*, 153–158.
36. Zhang, X.; Ming, X.; Yin, D. Application of industrial big data for smart manufacturing in product service system based on system engineering using fuzzy DEMATEL. *J. Clean. Prod.* **2020**, *8*, 121863. [CrossRef]
37. Wang, S.; Peng, Y.; Lan, H.; Luo, Q.; Peng, Z. Survey and Prospect: Data Integration Techniques. Available online: <https://kns.cnki.net/kcms/detail/detail.aspx?doi=10.13328/j.cnki.jos.005911> (accessed on 9 September 2021).
38. Liu, Y.; Cao, X. Research on performance optimization of distributed storage of massive video data. *Appl. Res. Comput.* **2021**, *38*, 1734–1738.
39. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 167–172. [CrossRef] [PubMed]
40. Darch, P.T.; Sands Ashley, E.; Borgman Christine, L.; Golshan Milena, S. Library Cultures of Data Curation: Adventures in Astronomy. *J. Assoc. Inf. Sci. Technol.* **2020**, *71*, 1470–1483. [CrossRef]
41. Yang, C.; Puthal, D.; Mohanty, S.P.; Kougiannos, E. Big-Sensing-Data Curation for the Cloud is Coming: A Promise of Scalable Cloud-Data-Center Mitigation for Next- Generation IoT and Wireless Sensor Networks. *IEEE Consum. Electron. Mag.* **2017**, *6*, 48–56. [CrossRef]
42. Vasily, B.; Brian, M. Data Curation Framework for Facilities Science. In Proceedings of the International Conference on Data Management Technologies and Applications, Reykjavík, Iceland, 29–31 July 2013; pp. 211–216.
43. Xing, H.Q. Mechanisms for Distribution and Realization of Personal Information Property Rights Under Background of Big Data Transactions. *Law Rev.* **2019**, *37*, 98–110.
44. Chen, Y.Y.; Wang, Y.Y. The Research on Stakeholder Interaction Relationship of Open Sharing of Scientific Research Data, Library Tribune. 2020. Available online: <http://kns.cnki.net/kcms/detail/44.1306.G2.20191218.1413.004.html> (accessed on 9 October 2021).

45. Balazinska, M.; Howe, B.; Suciu, D. Data Markets in the Cloud: An Opportunity for the Database Community. *Proc. VLDB Endow.* **2011**, *4*, 1482–1485. [CrossRef]
46. Stvilia, B.; Hinnant, C.C.; Wu, S.; Worrall, A.; Lee, D.J.; Burnett, K.; Kazmer, M.M.; Marty, P.F. Research project tasks, data, and perceptions of data quality in a condensed matter physics community. *J. Assoc. Inf. Sci. Technol.* **2015**, *66*, 246–263. [CrossRef]
47. Sun, L.L.; Yuan, Q.J. Research on Evaluation Index System of E-commerce Data Quality from the Perspective of Data Asset Management. *J. Mod. Inf.* **2019**, *39*, 90–97.
48. Chang, P. Simulation Research on Automatic Destruction Method of Life Cycle Controllable Data. *Comput. Simul.* **2019**, *36*, 371–375.
49. Xiang, F. I am the Alpha—On the ethics of human-machine. *Cult. Column* **2017**, *6*, 128–139.
50. Hsu, Y.C.; Hong, J.R. Caution on Urban “Public-Private Partnerships” and Private Control of Public Data: A Review of “Smart Cities in a Digital World”. *Int. J.* **2020**, *42*, 159–176.
51. Wiebe, A. Protection of Industrial Data—A New Property Right for the Digital Economy? *J. Intellect. Prop. Law Prot.* **2017**, *12*, 62–71. [CrossRef]
52. Wei, F.; Chang, Y. Review of domestic and foreign research on data tenure and analysis of development dynamics. *Mod. Intell.* **2017**, *37*, 159–165.
53. Shen, W. On data usufruct rights. *China Soc. Sci.* **2020**, *11*, 110–131, 207.
54. Artyushina, A. The EU is Launching a Market for Personal Data: Here’s What That Means for Privacy. MIT Technology Review. 2020. Available online: [https://www.researchgate.net/publication/343640324\\_The\\_EU\\_is\\_launching\\_a\\_market\\_for\\_personal\\_data\\_Here%27s\\_what\\_that\\_means\\_for\\_privacy](https://www.researchgate.net/publication/343640324_The_EU_is_launching_a_market_for_personal_data_Here%27s_what_that_means_for_privacy) (accessed on 9 September 2021).
55. Václav, J. Ownership of Personal Data in the Internet of Things. *Comput. Law Secur. Rev.* **2018**, *34*, 1039–1052.
56. Jansen, L. Private law protection of data property rights and interests. *Gansu Soc. Sci.* **2020**, *6*, 132–138.
57. Duch-Brown, N.; Martens, B.; Mueller-Langer, F. The Economics of Ownership, Access and Trade in Digital Data; Digital Economy Working Paper; 2017. Available online: <https://ssrn.com/abstract=2914144> (accessed on 9 September 2021).
58. Van Asbroeck, B.; Debussche, J.; César, J. Building the European Data Economy: Data Ownership. White Paper, Bird and Bird. 2017. Available online: [https://sites-twobirds.vulture.net/1/773/uploads/white-paper-ownership-of-data-\(final\)](https://sites-twobirds.vulture.net/1/773/uploads/white-paper-ownership-of-data-(final)) (accessed on 12 October 2021).
59. Banterle, F.; Data Ownership in the Digital Economy: An European Dilemma. EU Internet Law in the Digital Era. 2020. Available online: [https://link.springer.com/chapter/10.1007/978-3-030-25579-4\\_9](https://link.springer.com/chapter/10.1007/978-3-030-25579-4_9) (accessed on 20 October 2021).
60. Feng, G.; Xue, Y. From the rights regulation model to the behavior control model of data trust—An alternative way of thinking about the construction of data subject rights protection mechanism. *Law Rev.* **2020**, *38*, 70–82.
61. Yu, P.K. Data Producer’s Right and the Protection of Machine-Generated Data. *Tulane Law Rev.* **2019**, *93*, 859–929.
62. Cui, G.B. The theory underlying the limited exclusivity of big data. *Jurisprud. Res.* **2019**, *41*, 3–24.
63. Du, Y. Research on national data sovereignty in the era of big data. *Int. Obs.* **2016**, *3*, 1–14.
64. Feng, S. The data game and legal response in Tik Tok’s ban. *Orient. Law* **2021**, 74–89.
65. Zhang, X. Patterns and lessons learned from the rule building of data sovereignty and the rule building of data sovereignty in China. *Mod. Jurisprud.* **2020**, *42*, 136–149.
66. Pang, B.; Xuan, L.; Bai, Y.; Li, G. Status and characteristics of the EU’s construction of a data space governance rule system in the context of global data sovereignty game. *J. Inf. Resour. Manag.* 2021. Available online: <http://kns.cnki.net/kcms/detail/42.1812.G2.20201125.1513.002.html> (accessed on 22 October 2021).
67. Li, W. Digital sovereignty in Europe in the framework of strategic autonomy in 2020: A comprehensive acceleration. *Inf. Secur. Commun. Priv.* **2021**, *3*, 31–37.
68. Dong, W.; Tian, K.; Chen, Y.; Xu, Y.; Lan, M.; Zeng, M. Research on the evaluation method of integrated energy system based on game and evidence theory under energy internet. *Smart Power* **2020**, *48*, 73–80.
69. Zhang, Y.; Xie, H.; Mao, J.; Li, G. Research on multi-source data requirements and fusion methods for urban data portrait construction. *Intell. Theory Pract.* **2020**, *43*, 88–96.
70. Franks, P.C. Implications of Blockchain Distributed Ledger Technology for Records Management and Information Governance Programs. *Rec. Manag. J.* **2020**, *30*, 287–299. [CrossRef]
71. Wang, P.; Li, M.; Liu, X. The construction of credible ecology of document archive management from the perspective of blockchain. *Arch. Res.* **2020**, *4*, 115–121.
72. She, W.; Chen, J.S.; Liu, Q.; Hu, Y.; Gu, Z.; Tian, Z.; Liu, W. New blockchain technology for medical big data security sharing. *J. Chin. Comput. Syst.* **2019**, *40*, 1449–1454.
73. Wang, R.; Yu, S.; Li, Y.; Tang, Y.; Zhang, F. Medical blockchain of privacy data sharing model based on ring signature. *J. Univ. Electron. Sci. Technol. China* **2019**, *48*, 886–892. (In Chinese)
74. Ge, J.; Shen, T. Blockchain-Based Access Control Method for Energy Data. *Computer Applications*. Available online: <http://kns.cnki.net/kcms/detail/51.1307.TP.20210304.1108.004.html> (accessed on 12 October 2021).
75. Zuo, W.J.; Liu, L.J. Research on big data asset valuation method based on user perceived value. *Intell. Theory Pract.* **2021**, *44*, 71–77+88.
76. Shannon, C.; Yang, L.; Song, J. A CIME model design and implementation for data asset evaluation. *Comput. Appl. Softw.* **2020**, *37*, 27–34.

77. Li, C.; Wen, T. Research on the profitability model of big data trading in China. *J. Intell.* **2020**, *39*, 179–186.
78. Xiong, Q.; Tang, K. Advances in research on the boundary rights, transactions and pricing of data elements. *Dyn. Econ.* **2021**, *2*, 143–158.
79. Gkatzelis, V.; Aperjis, C.; Huberman, B.A. Pricing Private Data. *SSRN Electron. J.* **2012**, *25*, 1–15. [[CrossRef](#)]
80. Bocken, N.M.; Mugge, R.; Bom, C.A.; Lemstra, H.J. Pay-per-use business models as a driver for sustainable consumption: Evidence from the case of HOMIE. *J. Clean. Prod.* **2018**, *198*, 498–510. [[CrossRef](#)]
81. Bakos, Y.; Brynjolfsson, E. Bundling Information Goods: Pricing, Profits, and Efficiency. *Manag. Sci.* **1999**, *45*, 1613–1630. [[CrossRef](#)]
82. Cai, L.; Huang, Z.H.; Liang, Y.; Zhu, Y.Y. A review of data pricing research. *Comput. Sci. Explor.* **2021**, *15*, 1595–1606.
83. Li, X.; Yao, J.; Liu, X.; Guan, H. A First Look at Information Entropy-based Data Pricing. In Proceedings of the 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS), Atlanta, GA, USA, 5–8 June 2017; pp. 2053–2060.
84. Koutris, P.; Upadhyaya, P.; Balazinska, M.; Howe, B.; Suci, D. Query-based Data Pricing. *J. ACM* **2015**, *62*, 1–44. [[CrossRef](#)]
85. Cai, L.; Huang, Z.H.; Liang, Y.; Zhu, Y.Y. *Research on Data Transaction Pricing Based on Information Entropy*; Shanghai Jiaotong University: Shanghai, China, 2018.
86. Wang, F.; Shen, J. Progress of foreign data stewardship (Data Curation) research and practice. *Chin. J. Libr. Sci.* **2014**, *40*, 116–128.