

The Role of the Information and Documentation Professional in Big Data Management for Scientific Research

Eder Ávila^(a)
Andreu Sulé^(b)

a) Institute for Library and Information Research. UNAM, <https://orcid.org/0000-0001-5439-0495>

b) University of Barcelona, <https://orcid.org/0000-0002-2467-3678>

Contact: Eder Ávila, eder@ibli.unam.mx; Andreu Sulé, sule@ub.edu

Received: 04 February 2025; **Accepted:** 22 July 2025; **First Published:** 15 September 2025

ABSTRACT

The role of Information and Documentation professionals (IDPs) in managing big data within the context of scientific research is analyzed. Big data has transformed research by offering new opportunities, such as accelerating results and fostering interdisciplinary fields. However, it also presents significant challenges related to data access, quality, privacy, and security. Information professionals contribute to various areas: data retrieval, organization, description through meta-data, and long-term preservation. These multifaceted contributions demand a sophisticated skill set that require not only technical skills (e.g., database management and programming languages) but also interpersonal and ethical competencies that are essential for working in multidisciplinary teams. However, it should not be assumed that all Information and Documentation professionals inherently possess these qualities; rather, facing such challenges demands a continuous effort to cultivate intellectual openness, collaborative practices, and ethical responsibility. By fostering an open data culture and supporting responsible management practices through international standards for interoperability and reuse, Information and Documentation experts play a key role in adding value, structure, and ethics to scientific research in the digital age.

KEYWORDS

Big data; Documentation; Scientific research; Information professional.

Introduction

The nature of big data and its exponential growth has generated — and continues to generate — organizational and management needs in scientific research that have a direct impact on Information and Documentation professionals. Although the omnipresence of digital data and advanced information technology may have given a bigger role to professionals working in the fields of engineering, computer science, and statistics (Almeida et al. 2015), information and documentation professionals have a place on multidisciplinary teams working with big data on academic research projects. Tasks where information and documentation professionals can contribute their knowledge and skills to the development of data-driven research include: searching for and retrieving the most relevant data for each study; cleaning, transforming, and reconciling that data to enable more precise analysis; using or developing ontologies that allow machines not only to process data but also to interpret it and interconnect it with other datasets; and correctly and accurately describing data collections produced by the research center using consolidated metadata standards.

The aim of this article is to explore these and other tasks specific to Information and Documentation professionals working with the big data used on academic research projects. To this end, it begins with a definition of big data, to establish a conceptual framework and define the terminology used. This is followed by a discussion of the opportunities and challenges that big data poses for academic research, emphasizing some of the considerations and precautions to keep in mind when working with this type of data. Finally, an outline is offered of the different tasks where information and documentation professionals can add value to data-driven academic research.

Methodology

A literature review was conducted with the aim of identifying and analyzing information resources that address the theoretical foundations related to the role of information and documentation professionals in big data management within the context of scientific research. To achieve this objective, scientific articles, book chapters, and case studies were reviewed. The resources were obtained from prestigious academic databases such as Elsevier, SpringerLink, IEEE, and Emerald, as well as from institutional repositories like that of National Autonomous University of Mexico (UNAM) and specialized websites, including institutional portals of the European Union and UNESCO. This methodology ensured a comprehensive and rigorous approach to the analysis of the proposed research topic.

The search strategies within these sources focused on the following terms: big data management, scientific research, and information and documentation professionals. It is important to note that, although the selected terms allowed for the retrieval of resources addressing these topics in their titles, a detailed review of each resource was necessary to identify its disciplinary focus. This is because the conceptual notion of the information and documentation professional, as well as the activities related to big data management within the context of scientific research, encompasses multiple study variables, ranging from technical and pragmatic aspects to epistemological considerations.

Therefore, resources published between 2016 and 2024 were selected, taking into account technological and chronological criteria, as the phenomenon under study is constantly evolving due to the increasing number of publications analyzing its impact across various contexts and disciplines. Resources published outside this period were chosen based on their level of contribution and relevance to the analysis of the research topic proposed in this study. This approach ensured that the literature review was up-to-date and representative of the most recent trends and advancements in the field.

In addition to the above, additional filters were implemented to guarantee the quality and relevance of the selected resources. Priority was given to documents that presented an interdisciplinary approach, as big data management in scientific research requires the integration of knowledge from areas such as data science, computer science, library science, and knowledge management. Similarly, special attention was paid to studies offering critical or innovative perspectives on the role of information professionals in this emerging field.

Finally, the resources were organized into thematic categories to facilitate comparative analysis. These categories included: theoretical foundations of big data, applications in scientific research, competencies and roles of information professionals, and ethical and technical challenges. This structuring allowed for a deeper and more systematic understanding of the topic, contributing to the development of a relevant and well-grounded theoretical framework.

Definition and history of Big Data

The term “big data” first appeared in 1997 in an article by NASA researchers Michael Cox and David Ellsworth (1997). In this article, the authors highlighted the difficulties of storing, processing, and managing the increasingly large volumes of data required for their research with the technologies of the time.

Despite this initial emphasis on volume as an identifying element of big data, with the development and expansion of the Web and other data-generating technologies and services, its characterization became more complex. Doug Laney (2001), a computer engineer at the Gartner consulting firm, added two more “Vs” to the definition of the concept: velocity and variety. Velocity refers to the real-time creation of data (e.g., social network content), while variety points to the fact that data can come from different sources, in different formats, and at different levels of structuring. Subsequently, other authors (Bansal et al. 2020) have suggested incorporating other “Vs” into the characterization of big data, such as veracity, value, and variability. Lupton (2015) even proposed replacing all these Vs with 13 Ps that consider the sociocultural dimensions of big data: portentous, perverse, personal, productive, partial, practices, predictive, political, provocative, privacy, polyvalent, polymorphous, and playful.

However, as Kitchin and McArdle (2016) have shown, there is no characteristic profile that fits all big data, but rather different “types” of big data. There are datasets that are very large but not created in real time, and conversely, there are datasets created in real time that are not very large. After examining the nature of 26 datasets described in the literature as big data, the authors conclude that the two characteristics that truly differentiate big data from small data are velocity and exhaustiveness, understood as the ability to collect data from the entire system rather than just a sample.

Finally, it is worth noting that while initially the data collected was mainly related to the physical world of applied sciences, over time —and especially with the advent of social networks— the scope of data has expanded into social science research. Big data on human behavior has become a new source of knowledge for many researchers, unimaginable just a few decades ago. The use of this new resource raises sensitive questions related to data privacy and ethics in handling personal information.

Impact of Big Data on Scientific Research

The big data explosion has obviously had a huge impact on the field of academic research. In the last 15 years, every discipline has, to differing degrees, seen quantitative and qualitative changes to the possibilities of research work and the methodologies used to conduct it (Abdalla et al. 2024; Tosi et al. 2024).

Initially oriented towards computer science, finance, healthcare, and marketing, the scope of big data applications has expanded over the years to include disciplines as diverse as biology (ElSayed et al. 2021), astronomy (Klapp et al. 2024), education (Fischer et al. 2020), and the social sciences (Chen 2018).

As early as 2007, Jim Gray, a Microsoft researcher and a pioneer in database software, pointed out the significant impact that computing was having on the practice of science, a transformation that Gray called a “fourth paradigm” (Lynch 2009), represented by the integration of the three previous paradigms (theory, experimentation, and large-scale computation) in a context of data-intensive science. Bell (2009) traces the characteristics of the fourth paradigm back to Johannes Kepler’s exploitation of Tycho Brahe’s astronomical observations catalog for the discovery of the laws of planetary motion.

Based on this premise, the main opportunities that big data offers scientific research include:

- Acceleration of research: the availability of vast amounts of data and the ability to analyze it quickly and efficiently makes it possible to obtain results that previously required months or years of data collection and processing. Examples include the discovery of new drugs (Latha Bhaskaran et al. 2022), optimization of agricultural productivity (Misra et al. 2022), and support for disease diagnosis (Leopold et al. 2020).
- Emergence of new research fields: big data has enabled many scientists to tackle challenges that were previously insurmountable. In the field of the social sciences there are many examples due to the availability of extensive data on human behavior. Whether through data generated by social networks (Angskun et al. 2022) or by sensors (e.g., detecting people’s movement in a public space) (Jang et al. 2024), scientists can now formulate hypotheses and predictions that were previously unimaginable.
- Interdisciplinary collaboration: the complexity of capturing, organizing, and exploiting big data necessarily requires collaboration with computer scientists and statisticians at a minimum. Beyond this “forced” collaborative work, the capacity for exploiting big data allows many researchers to tackle projects drawing on related data sources from different fields. In public health, the spread of infectious diseases is often analyzed using big data from social networks, electronic medical records, and mobile services (Huang et al. 2020).

Similarly, in the book value chain —encompassing publishers, booksellers, and librarians— big data helps improve efficiency in the supply chain, increase the discoverability of books, and better understanding readers' needs to offer them the best content and services (Attanasio 2022).

However, these opportunities are not without challenges for researchers. As will be discussed below, some of these challenges can be addressed by information and documentation professionals:

- Data discovery: finding out whether the necessary data for a research project exists and where it is located is not always an easy task. Although there are an increasing number of repositories operated by governments (Data.gov, datos.gob.es, data.gov.in, Dades obertes, etc.), international organizations (World Bank Open Data, UNData, etc.), and institutions (Re3Data), and even thematic repositories (Dryad) and dataset search engines (Google's Dataset Search), many scientists are unaware of them or find them difficult to use.
- Data characteristics: beyond locating the data, rigorous research requires knowing the basic characteristics of the data: who created it, when it was created and for what purpose, what tools and methods were used to collect it, what organization published it, what fields make up the dataset, what formats it was published in, what access rights apply to it, and any other information necessary for its correct use and interpretation. Gathering data (especially from platforms such as social networks) raises important ethical and privacy challenges that must be addressed from the very beginning of the data lifecycle. In this regard, the data scientist has a key responsibility to ensure that data has been collected transparently, respectfully, and with full awareness of the rights of the individuals involved. It is not only about protecting data once it has been stored but about adopting an ethical stance from its origin. Without comprehensive information, scientists are at risk of basing their research on irrelevant data. Hence the importance of describing datasets with internationally consolidated metadata standards such as Dublin Core or DataCite Schema.
- Data quality: once it has been located and its relevance has been verified, scientists need to check the quality of the data. It is estimated that 80% of a data scientist's work is spent on cleaning processes (Subirats Maté et al. 2019; Ciavotta et al. 2022), the most common tasks being removing duplicates, correcting errors in data types and forms, normalizing controlled values, and ensuring the integrity of missing data. Additionally, if data is being used from different sources, data cleaning becomes even more important, as the goal of achieving a unique and coherent dataset can be hindered by the different structures and forms of each source. Beyond technical considerations, it is essential to critically assess data representativeness, especially in the field of social sciences and official data. Large volumes of data do not, by themselves, ensure a comprehensive or equitable view of reality, as they often originate from biased sources, exclude certain social groups, or reflect structural inequalities. This lack of representativeness can distort analyses and lead to misleading conclusions if not critically addressed. Therefore, Information and Documentation professionals must be aware of these limitations and promote data collection and analysis practices that embrace diversity and inclusion.
- New knowledge and skills: for scientists, the main challenge of working with big data is not its volume but the need to acquire knowledge and skills specific to data science that were often not required before. While it is true that developing advanced knowledge and

skills in statistics (probability, inference, multivariate analysis, etc.), mathematics (linear algebra, calculus and numerical analysis, graph theory, etc.), and computing (programming languages such as R or Python, databases, data visualization, etc.) can be a titanic task for a single person, it is equally true that a researcher intending to work with big data needs at least a basic grounding in these areas to form part of a multidisciplinary team and work side by side with statisticians, mathematicians, and computer scientists.

- New infrastructures: the three basic Vs of big data (volume, velocity, and variety) require specific computing architecture. To efficiently store, process, and analyze large volumes of data (in some cases generated in real time), systems with scalable or distributed storage capacity (computer clusters) and intensive processing (multiple CPU cores and large RAM capacity) are needed. High-speed networks are also advisable for the rapid transfer of data between system components. Obviously, this type of architecture comes at a cost that only large research centers can afford, and while cloud computing services can reduce the size of this investment, it is still significant (Abdalla 2022; Abdalla et al. 2024).
- Data security and privacy management: while security refers to all the tools and technologies necessary to monitor any type of attack, theft attempt, or other type of security breach (Das et al. 2021), privacy specifically relates to the handling of personal data, which is especially sensitive in research related to fields such as healthcare, finance, or social networks. It is important for researchers to be familiar with tools and techniques such as firewalls, antivirus software, anonymization, and encryption, so that they can work together with cybersecurity specialists to develop a comprehensive security plan for the project and understand the legal framework in which it operates (the General Data Protection Regulation in the case of the European Union). Cases such as the Cambridge Analytica scandal (Lapowsky 2018) highlight the potential impact of improper or malicious handling of personal data.

The Role of Information and Documentation Professionals

In the context described above, Information and Documentation professionals can play a crucial role. Their experience and skills in searching for, organizing, storing, analyzing, and managing information mean that they could make a valuable contribution to the management of the big data generated in research (Gbaje 2024).

According to Ohaji, Chawner, and Yoong (2019), data librarians play an essential role in the improvement of data accessibility and preservation in academic libraries and research institutions. Their professional profile enables them to offer services such as secure data storage, advice on open data policies, and training in data management. Additionally, their user-oriented focus places them in a privileged position to adapt to the objectives, needs, and limitations of any project. Garoufallou and Gaitanou (2021, 421) and Morriello (2020) have highlighted the importance of Big Data in the evolution of library services, the challenges that still need to be addressed, and the need for new skills for librarians.

Table 1 presents an outline of these skills, highlighting capabilities that are essential for responsible data management, such as effective communication, mastery of technological tools, and familiarity with intellectual property regulations and licenses.

Skills	Description
Interpersonal and behavioral characteristics	Formal communication skills, e.g., writing technical documents and case studies. This also includes adaptability to new experiences and always keeping updated on new developments.
Contextual knowledge of the institutional environment	Knowledge of funding policies of agencies that support scientific research. This includes an understanding of ethical procedures in the scientific process, disciplinary research methods, scientific communication, intellectual property, access methods, copyright, Creative Commons licenses, and royalties.
Data usage	An understanding of types of data (quantitative, qualitative), metadata schemas (Dublin Core, DataCite Schema), unique identifiers (DOI), and digital data preservation.
Knowledge of data manipulation technologies	Technical knowledge that includes an understanding of programming languages (Python, SQL, Java), database design and structuring, APIs for data retrieval, user-centered design, natural language processing, the Internet of things, and big data.
Conceptual skills	An understanding of data management principles, including data collection, organization, storage, preservation, and access. Knowledge of ontologies and controlled vocabularies to model and represent knowledge in a structured way. The ability to communicate complex information through effective data visualizations. Basic knowledge of artificial intelligence techniques such as machine learning and natural language processing to extract value from data.
Management	The ability to plan, organize, and execute complex data management projects. Effective communication skills to interact with users from different disciplines and explain technical concepts clearly and concisely. The ability to work in multidisciplinary teams and collaborate with other information professionals, researchers, and developers. Leadership skills to guide and motivate teams in data management projects.

Table 1. Skills of Information and Documentation professionals for handling big data in the context of scientific research. Source: Based on Semeler and Pinto (2020). Available at: <https://www.scielo.br/j/tinf/a/LNskQZxLgCVPRvbbFqN-9vZz/?lang=en#>

The skills of Information and Documentation professionals mentioned above qualify them to take on a range of tasks aimed at handling data in the context of scientific research, where it is possible to establish various methodologies and techniques for studying data, taking into account its nature and type. According to Morriello (2020, 3), in some countries, associations such as CILIP (United Kingdom) have identified data science as a key area for librarians, highlighting competencies in data management and curation, machine learning, and data analysis. Similarly, entities like the Australian National Data Service (ANDS) have outlined a precise profile for the data librarian, emphasizing their role in managing and preserving research data. Under this premise, Figure 1 presents the tasks that an information and documentation professional can perform in the field of data management within the context of scientific research.

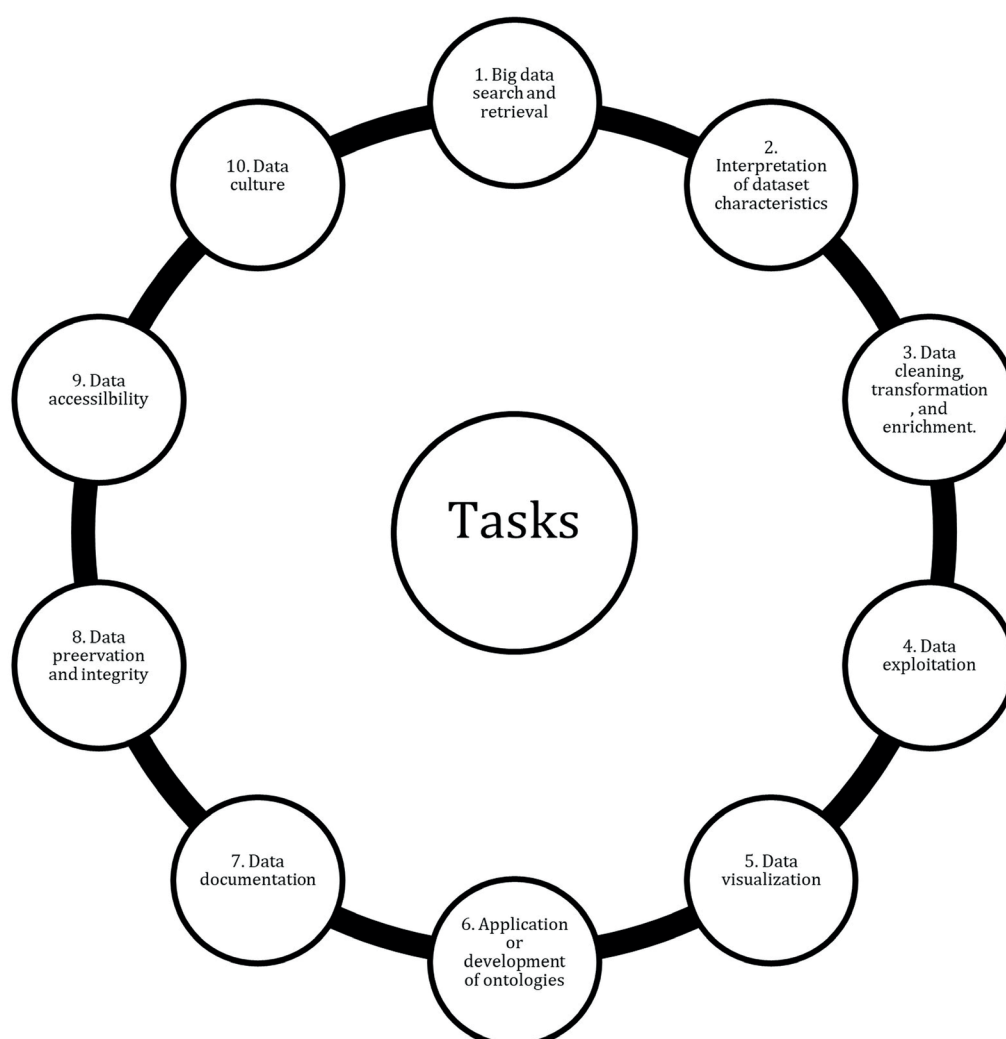


Figure1. Tasks of the Information and Documentation professional for data management in the context of scientific research. Source: own elaboration.

These tasks include the following:

1. Big data search and retrieval. The current abundance of data and the variety of available sources (from open repositories and institutional databases to social networks and restricted access sources) have created a significant challenge for researchers. The diversity of formats, data quality, access levels, and platforms is making the retrieval process increasingly complex. This context demands information and documentation professionals with advanced skills in search techniques, the use of Boolean operators, data mining, and the selection of specialized tools to optimize retrieval when dealing with masses of information (Chapman et al. 2020; Morriello 2020).
2. Interpretation of dataset characteristics. With the development of advanced metadata schemas such as DataCite Schema or Data Catalog Vocabulary (DCAT), datasets can include a very detailed level of description (e.g., authorship, creation context, applied methodologies,

formats, and limitations), which is key to their interpretation and proper use. However, this wealth of information is often too vast for the researcher to interpret every element and use it in their analysis. Here, the information and documentation professional plays a crucial role in breaking down and contextualizing these metadata, facilitating interpretation and ensuring that the data is accessible and understandable to the research community. According to Sanjuán (2019, 7), analyzing data always involves reflecting on it, regardless of the research project's methodological approach.

3. Data cleaning, transformation, and enrichment. Raw data generally suffers from problems such as duplicates, empty fields, typographical errors, and inconsistent values that hinder its analysis and exploitation. Using tools like OpenRefine, information and documentation professionals can perform exhaustive cleaning processes to correct and homogenize the data. These tools also allow the integration of additional data through external API queries (such as Wikidata) to enrich the information with specific details (e.g., geolocation, thematic categories, or unique identifiers), maximizing its value and usefulness in multidisciplinary research (Ciavotta et al. 2022; Ridzuan and Zainon 2019).
4. Data exploitation. Exploiting big data requires not only a basic understanding of the data itself but also advanced skills in statistics and programming to extract patterns, perform predictive analyses, and apply methods such as data mining or machine learning. An information and documentation professional with knowledge in these fields can act as a key advisor or collaborator, helping researchers derive meaningful conclusions from complex datasets and discover relationships that might otherwise go unnoticed. In recent years, numerous methods, techniques, and technologies have been developed to help organizations make better data-based decisions and extract value from the data they collect. These are used for capturing, processing, storing, managing, and analyzing data in order to exploit it more effectively and obtain greater benefits from it (Curto 2016).
5. Data visualization. The interpretation of large volumes of data is greatly facilitated by visualization. Tools like Tableau and Power BI allow the creation of graphs, maps, and interactive dashboards that translate complex data into clear and understandable visual representations. With a mastery of these tools, information and documentation professionals can design intuitive visualizations that not only effectively communicate findings but also allow researchers to interact with the data and explore various perspectives. According to Tufte (2007), the visual representation of data helps avoid misinterpretations. In this respect, graphs must be faithful to the data and avoid visual manipulations that distort reality.
6. Application or development of ontologies. Ontologies offer a valuable means of structuring and relating data in a way that is readable by both humans and machines. An ontology can be used to classify data and connect different datasets, allowing researchers to identify underlying relationships, to make inferences, and to automate analysis processes. This approach is especially useful in the field of artificial intelligence and natural language processing, where data must be semantically understood to make complex discoveries. An ontology models the knowledge of a particular domain, which could be useful in another application; effective sharing of ontologies is therefore very important for data reuse (Alfaifi 2022).

7. Data documentation. In an environment where massive data is intended to be reused, proper documentation is essential. Creating accurate descriptions based on standardized metadata schemas such as DataCite Schema, Data Catalog Vocabulary (DCAT), or CEDAR, or on more generic ones like Dublin Core or Schema.org (Guha, Brickley, and Macbeth 2016), ensures that data is accessible, understandable, and reusable in the future (Ávila Barrientos 2020). This correct and precise identification of datasets facilitates their citation in scientific texts, in addition to supporting a variety of data management activities and the development of reproducible, transparent, ethical, and high-impact research practices. Moreover, to foster a culture of data integration, scientists need to be convinced that preparing their data for online publication is an effort worth making (Brase 2010).
8. Data preservation and integrity. Data curators should also implement strategies that ensure the longevity and sustained integrity of data, in order to guarantee its validity and relevance for future research (Morriello 2020). This work is essential to ensure that published data has the same level of value, reliability, and accessibility as a conventional academic publication. According to Condon, Simpson, and Emanuel (2022), research data integrity is crucial for scientific rigor, and “although [the] inability to reproduce or replicate a study’s findings is not always a research data integrity issue, it is an important factor in successful attempts.” Furthermore, data reuse depends on the integrity of research data and the reliability of the original dataset.
9. Data accessibility. In the context of data repository management, information and documentation professionals have to design and develop digital systems to store and share research data securely and efficiently. They need to ensure that repositories function properly and are updated regularly, and to implement access and use policies that will guarantee data protection while promoting its reuse. The FAIR principles (findability, accessibility, interoperability, reusability) highlight the need to make data and other digital assets more accessible. Such accessibility is a key fundamental principle for bringing data into a domain with clearly defined access rules. There are three main components to the application of this principle: access protocol, access authorization, and metadata longevity (Wise et al. 2019).
10. Data culture. There is a need to promote a data culture by raising awareness in the scientific community about the importance of data management and open science. This can be done by providing researchers with training on the development of good data management practices and fostering collaboration between researchers and information and documentation professionals to optimize the data lifecycle. Data policies, mandates, and initiatives at the international level reveal the value of research data, as well as the importance of being able to share and reuse it. Ultimately, the promotion of a data culture gives meaning to any institutional action to coordinate the work carried out by researchers (Melero 2018; Morriello 2020).

In addition, Attanasio (2022) points out that the emergence of big data has led to a transformation in the way metadata is generated, shifting from a model where a human validates relationships between entities to a system where machines create connections by tracking digital events. This automation presents an opportunity to improve information management but also raises challenges in terms of data control and ethics.

Therefore, the data librarian emerges as a key figure in research and teaching, playing a crucial role in the curation and management of research data. However, the impact of big data and artificial intelligence has redefined how data and metadata are generated and used.

Conclusions

Information and documentation professionals play a pivotal role in the efficient management of massive data within the landscape of contemporary scientific research. Their expertise in organizing, structuring, and curating extensive datasets ensures not only their quality and accessibility but also enhances the reproducibility and integrity of research outcomes. In an environment increasingly driven by complex data analysis, these professionals contribute significantly to the success of multi-, inter-, and transdisciplinary projects by enabling seamless information flow and facilitating informed decision-making.

Beyond supporting data search and retrieval, their responsibilities extend to ensuring long-term data preservation and fostering its ethical and responsible reuse. By implementing rigorous metadata standards, data governance frameworks, and interoperability practices, they help mitigate risks associated with data loss, misinterpretation, and security breaches. Additionally, their role in promoting open science principles strengthens collaborative networks and democratizes access to knowledge, fostering innovation across diverse fields of inquiry.

To remain at the forefront of these evolving demands, information and documentation professionals must continuously develop competencies in data management, digital preservation, and emerging technologies. Mastery of artificial intelligence, machine learning, and data visualization techniques will allow them to provide strategic insights and add substantial value to research in the era of big data. By embracing these challenges, they position themselves as indispensable partners in the advancement of knowledge and scientific discovery.

The article provides a broad overview of the opportunities, challenges, skills, and tasks faced by Information and Documentation Professionals (IDPs) in the context of big data and scientific research. While the proposed lists are valuable and comprehensive, there is a need to further explore their practical application and to establish priorities within each category.

One relevant line of development involves using the skills list as a basis for empirical studies aimed at identifying training gaps in specific contexts. This could support the design of professional development programs more closely aligned with actual workplace demands.

Similarly, the task list provides a useful framework for investigating, through surveys or field studies, the real roles played by IDPs in concrete settings such as research centers, academic libraries, or data management units. This approach would allow for an assessment of whether the roles defined in theory match those practiced in everyday professional contexts.

Finally, it is recommended that future versions of the analysis be structured according to strategic priorities, distinguishing between critical, desirable, and emerging tasks. This would not only facilitate its use by institutional managers but also help IDPs align their professional development with key competencies for the contemporary digital environment.

References

- Abdalla, Hemn Barzan. 2022. "A Brief Survey on Big Data: Technologies, Terminologies and Data-Intensive Applications." *Journal of Big Data* 9 (1): 107. <https://doi.org/10.1186/s40537-022-00659-3>.
- Abdalla, Hemn Barzan, Ardalán Hussein Awlla, Yulia Kumar, and Maryam Cheraghy. "Big Data: Past, Present, and Future Insights." In *CTCNet '24: Proceedings of the 2024 Asia Pacific Conference on Computing Technologies, Communications and Networking*, 1–15. Chengdu, China, September 2024. <https://doi.org/10.1145/3685767.3685777>.
- Alfaifi, Yasser. 2022. "Ontology Development Methodology: A Systematic Review and Case Study." In *2nd International Conference on Computing and Information Technology (ICCIT)*, January 25–27, 446–50. <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9711664>.
- Almeida, Maria B., Renato R. Souza, and Renato M. A. Baracho. 2015. "Looking for the Identity of Information Science in the Age of Big Data, Computing Clouds and Social Networks." In *Proceedings of the 4th International Symposium of Information Science (ISI)*, Croatia. https://www.researchgate.net/publication/273448682_Looking_for_the_identity_of_Information_Science_in_the_age_of_big_data_computing_clouds_and_social_networks.
- Angskun, Jiratchaya, Siriporn Tipprasert, and Thara Angskun. 2022. "Big Data Analytics on Social Networks for Real-Time Depression Detection." *Journal of Big Data* 9 (1). <https://doi.org/10.1186/s40537-022-00622-2>.
- Attanasio, Pietro. 2022. "New Challenges in Metadata Management between Publishers and Libraries." *JLIS.it* 13 (1): 1–22. <https://doi.org/10.4403/jlis.it-12777>.
- Ávila Barrientos, Eder. 2020. "Análisis de los Principios para la Descripción de Datos de Investigación mediante DataCite Metadata Schema." *Anales de Documentación* 23 (1): 1–14. <https://doi.org/10.6018/analesdoc.400341>.
- Bansal, Mohit, Inderveer Chana, and Siobhán Clarke. 2020. "A Survey on IoT Big Data: Current Status, 13 V's Challenges, and Future Directions." *ACM Computing Surveys* 53 (6): 1–59. <https://doi.org/10.1145/3419634>.
- Bell, Gordon. 2009. "Foreword." In *The Fourth Paradigm: Data-Intensive Scientific Discovery*, edited by Tony Hey, Stewart Tansley, and Kristin Tolle, XI–XV. Redmond, Washington: Microsoft Research.
- Bhaskaran, Krithika Latha, Richard Sakyi Osei, Evans Kotei, Eric Yaw Agbezuge, Carlos Ankora and Ernest Domanaanmwi Ganaa. 2022. "A Survey on Big Data in Pharmacology, Toxicology and Pharmaceuticals." *Big Data and Cognitive Computing* 6 (4): 161. <https://doi.org/10.3390/bdcc6040161>.
- Brase, Jan. 2010. "DataCite: A Global Registration Agency for Research Data." In *RatSWD Working Paper*, 149. Berlin: Rat für Sozial- und Wirtschaftsdaten (RatSWD). <https://www.econstor.eu/bitstream/10419/43624/1/638240131.pdf>.
- Chapman, Adriane, Elena Simperl, Laura Koesten, George Konstantinidis, Luis-Daniel Ibáñez, Emilia Kacprzak, and Paul Groth. 2020. "Dataset Search: A Survey." *VLDB Journal* 29 (1): 251–72. <https://doi.org/10.1007/s00778-019-00564-x>.

Chen, Shu-Heng, ed. 2018. *Big Data in Computational Social Science and Humanities*. Cham: Springer.

Ciavotta, Michele, Vincenzo Cutrona, Flavio De Paoli, Nikolay Nikolov, Matteo Palmonari, and Dumitru Roman. 2022. "Supporting Semantic Data Enrichment at Scale." In *Technologies and Applications for Big Data Value*, 19–39. Cham: Springer. <https://doi.org/10.1007/978-3-030-78307-5>.

Condon, Patrick, Jennifer Simpson, and Michael Emanuel. 2022. "Research Data Integrity: A Cornerstone of Rigorous and Reproducible Research." *IASSIST Quarterly* 46 (3). <https://doi.org/10.29173/iq1033>.

Cox, Michael, and David Ellsworth. 1997. "Managing Big Data for Scientific Visualization." *ACM SIGGRAPH* 97 (1): 21–38. https://www.researchgate.net/profile/David-Ellsworth-2/publication/238704525_Managing_big_data_for_scientific_visualization/links/54ad79d20cf2213c5fe4081a/Managing-big-data-for-scientific-visualization.pdf.

Curto, Joan. 2016. *Fundamentos de Big Data: Habilitando la Explotación de Datos Complejos*. UOC. https://openaccess.uoc.edu/bitstream/10609/136326/2/Fundamentos%20y%20usos%20del%20big%20data_Mo%C2%BFdulo2_Fundamentos%20del%20big%20data%20Habilitando%20la%20explotacio%C2%BFn%20de%20datos%20complejos.pdf.

Das, Pradeep Kumar, Harihar K. Tripathy, and Siti Aisyah M. Yusof. 2021. *Privacy and Security Issues in Big Data: An Analytical View on Business Intelligence*. Cham: Springer.

ElSayed, Ibrahim A., Kamal ElDahshan, Hoda Hefny, and Eman K. ElSayed. 2021. "Big Data and Its Future in Computational Biology: A Literature Review." *Journal of Computer Science* 17 (12): 1222–8. <https://doi.org/10.3844/jcssp.2021.1222.1228>.

Fischer, Christian, Zachary A. Pardos, Ryan S. Baker, Joseph Jay Williams, Padhraic Smyth, Renzhe Yu, Stefan Slater, Rachel Baker, and Mark Warschauer. 2020. "Mining Big Data in Education: Affordances and Challenges." *Review of Research in Education* 44 (1): 130–60. <https://doi.org/10.3102/0091732X20903304>.

Garoufallou, Emmanouel, and Panorea Gaitanou. 2021. "Big Data: Opportunities and Challenges in Libraries, a Systematic Literature Review." *College and Research Libraries* 82 (3): 409–28. <https://crl.acrl.org/index.php/crl/article/view/24918/32750>.

Gbaje, Ezra. 2024. "Big Data Analysis: An Emerging Paradigm in Information Management." *Nigerian Library Association Kogi State Chapter 2021 Conference*, Kogi State University Ayingba, Kogi State, September 6. https://www.researchgate.net/publication/384838489_Big_Data_Analysis_An_Emerging_Paradigm_in_Information_Management.

Guha, Ramanathan V., Dan Brickley, and Steve Macbeth. 2016. "Schema.org: Evolution of Structured Data on the Web." *Communications of the ACM* 59 (2): 44–51. <https://doi.org/10.1145/2844544>.

Huang, Haibo, Zhaohui Peng, Haibo Wu, and Qiong Xie. 2020. "A Big Data Analysis on the Five Dimensions of Emergency Management Information in the Early Stage of COVID-19 in China." *Journal of Chinese Governance* 5 (2): 213–33. <https://doi.org/10.1080/23812346.2020.1744923>.

Jang, Seongmin, Juhyun Bae, and Younghoon Kim. 2024. "Street-Level Urban Heat Island Mitigation: Assessing the Cooling Effect of Green Infrastructure Using Urban IoT Sensor Big Data." *Sustainable Cities and Society* 100. <https://doi.org/10.1016/j.scs.2023.105007>.

Kitchin, Rob, and Gavin McArdle. 2016. "What Makes Big Data, Big Data? Exploring the Ontological Characteristics of 26 Datasets." *Big Data and Society* 3 (1). <https://doi.org/10.1177/2053951716631130>.

Klapp, Jaime, Celia R. Fierro-Santillán, and Leonardo G. Sigalotti. 2024. "El Estudio de las Estrellas en la Era del Big Data." *CIENCIA ergo-sum* 31. <https://doi.org/10.30878/ces.v31n0a36>.

Laney, Doug. 2001. "3D Data Management: Controlling Data Volume, Velocity and Variety." *META Group Research Note*, 6.

Lapowsky, Issie. 2018. "Facebook Exposed 87 Million Users to Cambridge Analytica." *Wired*, April 4. <https://www.wired.com/story/facebook-exposed-87-million-users-to-cambridge-analytica/>.

Leopold, Jane A., Bradley A. Maron, and Joseph Loscalzo. 2020. "The Application of Big Data to Cardiovascular Disease: Paths to Precision Medicine." *Journal of Clinical Investigation* 130 (1): 29–38. <https://doi.org/10.1172/JCI129203>.

Liao, Hsin-Ta, Zhijie Wang, and Yang Liu. 2020. "Exploring the Cross-Disciplinary Collaboration: A Scientometric Analysis of Social Science Research Related to Artificial Intelligence and Big Data Application." *IOP Conference Series: Materials Science and Engineering* 806 (1). <https://doi.org/10.1088/1757-899X/806/1/012019>.

Lupton, Deborah. 2015. "The Thirteen Ps of Big Data." *The Sociological Life*, May 13. <https://simplysociology.wordpress.com/2015/05/11/the-thirteen-ps-of-big-data/>.

Lynch, Clifford. 2009. "Jim Gray's Fourth Paradigm and the Construction of the Scientific Record." In *The Fourth Paradigm: Data-Intensive Scientific Discovery*, edited by Tony Hey, Stewart Tansley, and Kristin Tolle, 177–83. Redmond, Washington: Microsoft Research.

Melero, Remedios. 2018. *Recomendaciones para la Gestión de los Datos de Investigación*. Maredata. <https://digital.csic.es/handle/10261/173801>.

Misra, N. N., Yash Dixit, Ahmad Al-Mallahi, Manreet Singh Bhullar, Rohit Upadhyay, and Alex Martynenko. 2022. "IoT, Big Data, and Artificial Intelligence in Agriculture and Food Industry." *IEEE Internet of Things Journal* 9 (9): 6305–6324. <https://doi.org/10.1109/JIOT.2020.2998584>.

Morriello, Rosanna. 2020. "Birth and Development of Data Librarianship." *JLIS.it* 11 (3): 1–15. <https://doi.org/10.4403/jlis.it-12653>.

Ohaji, Ifeanyi K., Brenda Chawner, and Pak Yoong. 2019. "The Role of a Data Librarian in Academic and Research Libraries." *Information Research* 24 (4). <https://informationr.net/ir/24-4/paper844.html>.

Ridzuan, Fakhitah, Wan Mohd Nazmee, and Wan Zainon. 2019. "A Review on Data Cleansing Methods for Big Data." *Procedia Computer Science* 161: 731–38. <https://doi.org/10.1016/j.procs.2019.11.177>.

Rossi, Renata, and Kenji Hiram. 2015. "Characterizing Big Data Management." *Issues in Informing Science and Information Technology* 12: 165–80. <http://iisit.org/Vol12/IISITv12p165-180Rossi1921.pdf>.

Sanjuán Núñez, Luis. 2019. *El Análisis de Datos en Investigación Cualitativa*. UOC. https://openaccess.uoc.edu/bitstream/10609/147145/2/MetodosDeInvestigacionCualitativaEnElAmbitoLaboral_Modulo5_ElAnalisisDeDatosEnInvestigacioCualitativa.pdf.

Semeler, Alexandre R., and Ana L. Pinto. 2020. "Data Librarianship as a Field Study." *Transinformação* 32. <https://doi.org/10.1590/2318-0889202032e200034>.

Subirats Maté, Lluís, Daniel O. Pérez Trenard, and María Calvo González. 2019. *Introducción a la Limpieza y Análisis*. UOC. <https://openaccess.uoc.edu/bitstream/10609/148647/1/IntroduccionA-LaLimpiezaYAnalisisDeLosDatos.pdf>

Tosi, Davide, Rrezon Kokaj, and Marco Roccetti. 2024. "15 Years of Big Data: A Systematic Literature Review." *Journal of Big Data* 11 (1). <https://doi.org/10.1186/s40537-024-00914-9>.

Tufte, Edward R. 2007. *The Visual Display of Quantitative Information*. Cheshire: Graphic Press.

Wise, John, Alexandra Grebe de Barron, Andrea Splendiani, Beeta Balali-Mood, Drashtti Vasant, Eric Little, Gaspare Mellino, Ian Harrow, Ian Smith, Jan Taubert, Kees van Bochove, Martin Romacker, Peter Walgemoed, Rafael C. Jimenez, Rainer Winnenburg, Tom Plasterer, Vibhor Gupta, Victoria Hedley. 2019. "Implementation and Relevance of FAIR Data Principles in Biopharmaceutical R&D." *Drug Discovery Today* 24 (4): 933–8. <https://doi.org/10.1016/j.drudis.2019.01.008>.