The Skills of Information Professionals for the Analysis and Organization of Data. Conceptual Proposal and Application of *OpenAI* in Job Advertisements on the Web

José-Antonio Moreiro-González; Francisco-Carlos Paletta; Juan-Antonio Pastor-Sánchez; Barbara Coelho-Neves

Recommended citation:

Moreiro-González, José-Antonio; **Paletta, Francisco-Carlos**; **Pastor-Sánchez, Juan-Antonio**; **Coelho-Neves, Barbara** (2024). "The skills of information professionals for the analysis and organization of data. Conceptual proposal and application of *OpenAI* in job advertisements on the Web". *Profesional de la información*, v. 33, n. 5, e330501.
https://doi.org/10.3145/epi.2024.0501

Article received on September 24th 2024
Approved on October 02nd 2024

José-Antonio Moreiro-González
https://orcid.org/0000-0002-8827-158X
Universidade Federal da Bahia (UFBA)
Programa de Pós-Graduação em Ciência da Informação
Univ. Carlos III de Madrid, Depto. de Biblioteconomía y Documentación
Madrid, 126. 28903 Getafe (Madrid), Spain
jamore@bib.uc3m.es

Francisco-Carlos Paletta
https://orcid.org/0000-0002-4112-5198
Universidade de São Paulo
Escola de Comunicações e Artes
Av. Prof. Lúcio Martins Rodrigues, 443
Butantã, São Paulo - SP, 05508-020, Brazil
fcpaletta@usp.br

Juan-Antonio Pastor-Sánchez
https://orcid.org/0000-0002-1677-1059
Universidad de Murcia
Facultad de Comunicación y Documentación
Campus Universitario
30100 Espinardo (Murcia), Spain
pastor@um.es

Barbara Coelho-Neves
https://orcid.org/0000-0002-3429-7522
Universidade Federal da Bahia (UFBA)
Programa de Pós-Graduação em Ciência da Informação
Av. Reitor Miguel Calmon, 1946
Canela, Salvador (BA), Brazil
babi.coelho@gmail.com

Abstract

This study explores the evolving roles of library and information professionals (IPs) in data analysis and organization, particularly within interdisciplinary teams. It builds on the premise that data is pivotal for strategic business and institutional management, with IPs contributing through their expertise in information and communication technologies (ICTs), mediation with users, and mastery of taxonomies and semantic organizational structures (SOCs). The study examines job advertisements to determine the skills, experience, and training employers seek in IPs for data-centric roles. A total of 84 international job postings from platforms such as *ALAJobLIST*, *LinkedIn*, and *Glassdoor* were analyzed. The postings were XML-tagged to identify descriptions, competencies, and requirements. Using *OpenAI* tools, keywords were extracted, standardized, and grouped into thematic clusters and macro-clusters for analysis. The findings reveal that required skills span eight macro-clusters: business management; data management; information organization; data analysis and statistics; programming; web technologies; analytical creativity; and domain-specific expertise. IPs are valued for their technical and technological skills, particularly in navigating and applying standards for data description, visualization, and communication. Conclusions highlight that while IPs possess strong foundational skills, additional training in statistics and analysis platforms is critical to fully meet employer demands. Employers trust IPs for their ability to handle data-centric roles in competitive environments, emphasizing the importance of communication, leadership, and customer-oriented skills. The study underscores the need to enhance LIS curricula to better integrate advanced statistical training and entrepreneurial competencies. This research emphasizes the growing recognition of

IPs' contributions to data analysis and organization, underscoring their essential role in addressing the complexities of interdisciplinary data-driven projects.

Keywords

Information Science, Librarians, Information Professionals, LIS; Curricula, Job Market, Web Job Advertisements, Skills, Experience, Content Analysis, *OpenAI*, Data Analysis, Taxonomic Organization, Metadata, Semantic Organizational Structures.

1. Strategic Framework and Practical Setting

It is well known that data generates knowledge, that its processing follows a clear rationale and that its analysis is used to make the right decisions. These premises make data decisive for properly designing the strategies to be followed in business and institutional management and for supporting the development of all kinds of projects, especially research projects. The study of data is usually carried out by personnel from integrated STEM fields, who analyze data in any organization or company to discover patterns and trends and identify atypical situations (**Štuikys; Burbaitė**, 2024).

Information Science was created at a time of communicative and technological change, molded by various disciplinary contributions (**Bereijo**, 2019). Since then, computers have allowed access to data, processed records of research results, samples and evidence, with particular attention to lists of names and dates, codes and calculations that have become another support for systems by codifying and expressing information in a quantitative, objective and precise way (**Crisci; López-Armengol**, 1983). Information professionals' interest in data initially led them to take on the management of research projects (**Semeler *et al.***, 2019). Therefore, working with Big Data meant addressing the skills needed in Data Science and how to acquire them in order to continue managing users' information needs. But they could not be limited to scientific data, as analytical services extend to the informational value that reveals trends, from which all types of public administrations, companies, industries or non-profit organizations plan and make decisions (**Herndon**, 2021). The efficiency of organizations increases with the appropriate management of data, which must be processed continuously so as not to lose predictive capacity and allow knowledge to emerge. On the other hand, the digital transformation is changing the job profiles of information professionals. In this context, there is an increase in job offers to join teams of collaborators in diverse environments that are characterized by data analysis services, which require LIS degrees, among others. The opportunity is conditional on having the necessary skills to work as data analysts or librarians, taxonomists, information architects and knowledge, content or competitive intelligence managers (**Paletta; Moreiro-González**, 2021). Even though the offers do not reduce the qualification required to the academic record, it is also derived from the skills and experience acquired in relation to preparing, analyzing and managing the quality and use of data. Undoubtedly, this new profile has a clear informative condition, because as well as identifying, selecting and classifying data, it guarantees its traceability, safeguarding and access, based on experience in managing hierarchies and content objects and in applying metadata standards that describe the data and guarantee effective use. That's why it's essential to have a perspective so that collection, integration and processing preserve their context. However, many organizations do not have metadata or, when they do, it is poor or unstructured. Data is often scattered throughout the organization, with identical information appearing disaggregated and even managed with different vocabularies (**Agarwal; Dhar**, 2014).

The idea of recording the growth of library and information professionals involved in data management and access was identified by **Stuart** (2020), although focused on the challenges faced by science and data scientists. However, he emphasizes how libraries are increasingly involved in providing services between people and the gradually complex data landscape. This requires the right skills, now reinforced by the development of open-source software and tools that make it increasingly easy to write and share code. He even foresees the boost that big data analysis and artificial intelligence will bring to this end. However, he recognizes that his book only addresses the employment possibilities offered by the growing amounts of data (**Stuart**, 2020).

At the heart of these recent profiles is the analysis and care of data: its lineage, master data, quality, ethics, governance, architecture and modelling, along with granular classification and processing in taxonomies. It is always based on a design that supports accessibility and navigation of the data and products visualized, as well as providing engaging and meaningful experiences for users. It uses statistical methods and algorithms to group and machine learning to classify. Grouping is carried out using coefficients of similarity, plus others of difference, correlation and association (**Stuart**, 2020). Predictive methods and genetic algorithms are used for modelling and decision-making (**Griffey**, 2019). Although the data managed is not of the volume of Big Data, it is analyzed using artificial intelligence statistical resources such as machine learning (ML), deep learning and text analysis (**Semeler *et al.***, 2024). It is worth emphasizing the role of AI in automation and retrieval because, in addition to classifying with learning patterns, it operates between users and data sets whose availability, quality and accessibility it improves, as well as their processing and reuse.

To carry out these different tasks, there are many skills, competences and abilities that must be mastered for this professional intervention to be reliable. And, consequently, also for the training that enables it. Of course, this occupation of library and information professionals has no feet of clay. It is backed up by technical training that makes them ideal for building bridges between users and scientific, governmental or commercial data. From a business and institutional perspective, information professionals are considered to have the ability to analyze and evaluate data (**Hair *et al.***, 2009). In

response, data librarians and digital taxonomists manage and organize knowledge based on the theory of organization, retrieval and information systems. This allows them to work in various environments that require skills in metadata and the management of data and content objects. Metadata schemas and semantic organizational structures (SOCs) contribute to the control, accessibility, interoperability and therefore analysis and exploitation of data.

Organizing the results of data analysis involves going beyond statistical classification techniques to the use of SOCs to manage the governance, access and use of data and products. To do this, they rely on the tools offered by information and communication technologies (ICTs) to create and maintain taxonomies on the Web, whether for company intranets or for relationships with users and clients or with other organizations via the Internet. SOCs reflect the objectives of the intellectual organization of a field of knowledge or an institution, which leads them to be differentiated according to the type of entity or company and the data to be processed, although they are always very close to the real world whose semantics they represent and from the perspective of the users they serve.

Previous research has looked at the Brazilian information science labor market affected by the digital transformation, both in the public and private sectors. Now the cycle closes with a broadly international study of vacancies not just for Information Science (IS) graduates or professionals, much less for a single profile of archivist, librarian or information manager. The persuasion to do this stems from the epistemological approach of regrouping knowledge, the most stable principle of which is that if the nature of information is interdisciplinary, so are its applications (**Saracevic**, 1995). The information professions are located directly at the intersection of ICT skills and applications and therefore reflect the inevitable consequences of the digital transformation that has profoundly renewed professional paradigms and practices. In this sense, it is necessary to investigate whether specific Information Science techniques have continuity and how they are used in conditions determined by the need for competences common to different sectors of activity. This is because work is carried out in teams and interdisciplinary spaces in integrated tasks that require competences located at a crossroads of domains (**Moreiro-González**, 2017).

As a result, digital management itself leads to a combination of areas to meet the transversal nature of content and competences, both in terms of training and professional application. Furthermore, in this context, any study must be suited to working in interdisciplinary organizational structures with good technological resources. Consequently, these aspirations are reflected in the need to analyze the ability to manage large amounts of data, omnipresent attention to the user and institutional purposes, advanced handling of ICT tools and an in-depth understanding of the application environment. It is necessary to understand the profiles to be addressed in a corporate or institutional strategy in which communication and digital content management are combined. Furthermore, in order to find out how information professionals contribute to data analysis, the processing engine in this article is the *OpenAI* application programming interface, as it facilitates the procedures for analyzing the knowledge collected from both the basic contributions and the organization of the competences identified. There is no better way to include algorithms in data processing. If Information Science is going through an epistemological moment of reunification of knowledge, its professionals are obliged to collaborate and help understand complex facts, even more so in a context of scientific and technological transformation. Interdisciplinarity is the conceptual and labor reference and therefore also the basis of university education. Educational content needs to be approached with a comparable response from universities that cannot afford outdated curricula.

> In order to find out how information professionals contribute to data analysis, the processing engine in this article is the *OpenAI* application programming interface, as it facilitates the procedures for analyzing the knowledge collected from both the basic contributions and the organization of the competences identified

The proposal presented here makes sense as it seeks to discover the reasons that underpin companies 'and organizations' confidence in the good performance of library and information professionals in terms of analyzing and organizing data. This implies specific demands from companies and organizations that hold data on these professionals in order to include them among the decisive factors in their relationships with clients and users. In short, it implies rigorous monitoring of the value that librarians and information professionals bring to analyzing and managing information data from texts and other content objects. On the other hand, it means focusing on the basis of the security that leads them to participate in competitive tenders against trained competitors or specialists in basic sciences. If we want to give a real contrast to this exposition, confirmation comes from deepening these and other hypothetical arguments in the study of the LIS labor market related to the analysis and taxonomic organization of data. This is the context of this proposal, which aims to obtain generalist value from applied principles.

1.1. Taxonomic Representation of Data

The purpose of the data analysis process is not only descriptive and diagnostic, but also predictive. Data is often represented, structured and managed in taxonomies that harmonize categorization with visualization (**Gani *et al.***, 2016). This ratifies them as a systematizing prototype, even if far from the initial terminological character. In this way, the schematization of data can

generate new knowledge, but it is not always analyzed correctly to produce a visual synthesis that facilitates interpretation.

It's about discovering patterns and trends, facilitating research or prospective decision-making (**Husamaldin; Saeed**, 2020), which requires the collaboration of data and metadata managers, along with the researchers themselves or those responsible for information management. To interpret data, it is essential to know the characteristics of the organization in which the analyzed data applies. When the data is corporate, it is understood from the business rules. Defining a taxonomy assertively is fundamental when it comes to providing relevance for data labelling and improving exploration, access and intuitive understanding by users and customers. Its service to very different organizations means that only by knowing the specific field is it possible to distinguish what is relevant, to a greater extent in Big Data.

The most common approach is to visualize the data in a graphical set, a successor to the taxonomic tree (**NISO**, 2005), which recognizes what is most significant and guides navigation through the data, relationships and statistical trends, if they lead to minimizing uncertainties and enabling interpretation. Knowledge graphs configure the processing architecture from the Semantic Web standards that provide ontological character and grant interoperability to the metadata and linked data (**Bonatti *et al.***, 2019). Developers are strategic intermediaries who ensure direct communication between organizational units and the interoperability of symbolic resources: objects, knowledge, information, terms or data, to provide transparency and traceability of decision-making processes (**Lomax**, 2019).

1.2. Context and Objectives

As ICT improves processing, there is a progressive convergence with Data Science, which is why the demand for academic librarians trained in data has had to be proven (**Khan; Du**, 2017; **Federer**, 2018). It was also noted in job ads that, in just five months, the corporate data librarian-scientist qualification offered 12 vacancies in Brazil, while there were 11 vacancies for data analyst-manager and 6 for project data analyst-manager (**Paletta; Moreiro-González**, 2021). These figures indicate that organizations are integrating information professionals into interdisciplinary teams for data management, due to their skills in description, analysis, organization and dissemination. In addition, by offering professional opportunities to graduates from different backgrounds, we can confirm the extent to which the spectrum of options open to information science graduates. As well as knowing the characteristics of training, experience and occupation involved in heterogeneous training, however difficult it may be to act as a data analyst (**Oliver *et al.***, 2019). This co-occurrence is favored by the strategic purpose of SOCs to characterize contexts of use and meet real, practical needs.

The general objective is to verify the penetration of characteristic profiles of data analysis in the professional skills of Information Science. As a cognitive acquisition, it involves verifying the requirements established by companies to be data analysts and systematizes. As well as understanding how employers define an Information Professional who works with data management and analysis through the qualifications they declare and list, in which skills specific to data analysis and management coexist. They also perceive their organization and management through taxonomic schemes and other distinctive SOC of a field, company or entity.

2. Methodology

Content analysis is used to verify the evidence behind the selection of 84 full-text ads on international job portals. The advantages of analyzing web ads for understanding the activities carried out by Information Professionals have been highlighted in previous work (**Paletta *et al.***, 2021). Now, the ads must coincidentally contain data analysis, along with taxonomy or another SOC and even, as explicit training, any of the courses in Information Science. *LinkedIn* was initially chosen as a source, not only because it is a large network for job offers, but also because of the existence of the *Taxonomy and Ontology Community of Practice* discussion group, whose administrators specialized in data management. In addition, *ALAJobLIST* is used, which offers the advantage of directing ads to professionals or graduates in Information Science; as well as *Glassdoor Jobs*, which is international in nature and offers detailed descriptions, and the Brazilian site *Catho*, Table 1. The searches were carried out in the last month of 2022 and the first three months of 2023 using the terms 'Data analysis and taxonomist', 'Data analysis and taxonomy', 'Data taxonomy', 'Data catalog architect' and 'Data analysis and organization'.

Table 1: Sources of the Advertisements.

Professional Network	No of Ads	%
LinkedIn	33	39,28
Glassdoor Jobs	31	36,90
ALAJobLIST	11	13,09
Catho	9	10,71
Total	84	100

Only vacancies that coincide data analysis with taxonomic organization are mentioned. In the period covered, another 97 *LinkedIn* advertisements offered vacancies for data analysts, mainly for modelling, which require a degree in STEM (science, technology, engineering and mathematics), not Information Science, so they were left out. *ALAJobLIST* published 79 offers for data analysis and visualization, but less than half were related to data organization using taxonomies

or other SOCs. As the processing is done in English, some of *Catho*'s offers have been translated. In addition, 34 duplicate job offers were eliminated. The sample selected is considered sufficient to be representative and to obtain the necessary reproducibility to make its content reliable (**Krippendorff**, 2018).

The analysis was carried out in two phases: the first was tabular, enumerative and descriptive of the results and the entities identified automatically. The second qualitatively evaluates the content of the ads after processing. These seven stages take place in the first phase:

Step 1: After obtaining the ads in text format and in English, the documents are structured according to the following XML schema to make them homogeneous and expressive before analysis (**Berelson**, 1952):

- <content>: Root element of the XML document.
- <content>→<main>: Individual advert.
- <content>→<main>→<description>: General description of the advert, including title, place of publication and information text.
- <content>→<main>→<tasks>: Tasks to be carried out in the job offer.
- <content>→<main>→<requirements>: Qualifications and competences related to the job offer.
- <content>→<main>→<requirements>→<qualifications>: the qualifications required are specified in different lines.
- <content>→<main>→<requirements>→<competences>: the competences required are specified along different lines.

Step 2: The *OpenAI* API is used to extract the original keywords from the skills required in each job offer. The text-davinci-003 template is applied with the following prompt:

Extract keywords from the following list of competencies in a job offer and display them separated by semicolons.

After processing each job offer with this prompt, a text file is obtained in which the 1,554 keywords extracted from the competences of all the ads appear on different lines and separated by semi-colons. After avoiding the various repetitions, 1,126 unique keywords were obtained.

Step 3: The keywords are processed with a *Python* script to obtain the following statistics:

- Name of keywords
- Total frequency of occurrences (the same keyword can appear several times in the same advert) in the same advert).
- Number of different adverts in which the term appears.

The resulting CSV file is manually reviewed to group the synonyms. As an example, 'categorization and classification' was used as a synonym for these keywords: Categorization; categorization; Categorization design; classification; Classification; Classification definitions; Classification schemes; Classification Systems; Classifying.

102 keywords that didn't make sense were discarded. This brings the number of keywords selected to 1,024. The keywords are then taken to their canonical form (compound terms, singular and plural) and the corresponding preferred terms (synonyms) are defined between words with the same or very similar meaning. The total number of synonyms is 480, which represents an average of 2.13 keywords per preferred term.

Step 4: In addition, the *OpenAI* API is used to identify Clusters from the grouping of keywords obtained in step 2. To do this, apply the prompt:

Identify up to a maximum of 20 thematic clusters from the following list of keywords.

When *OpenAI* returned the number of clusters, it was asked to assign each of the keywords to one or more of them (up to a maximum of three). The number of clusters proposed by *OpenAI* was 36, which it was decided to limit to 20 after careful analysis and for reasons of processing time efficiency. Why wasn't it asked to do this with synonyms? Because we wanted to maintain the original semantic meaning of the adverts and because the mapping between the keywords and the synonyms had already been done, so it was automatic to work out the clustering of synonyms from the clustering of keywords.

Step 5: After obtaining the list of subject groups, the *OpenAI* API is used again to group each of the keywords from step 2 with each of the subject groups obtained in step 4. To provide some flexibility, each keyword could be assigned to two subject groups using the prompt:

For the following keyword: {Key_word}

assign one of these categories (maximum 2 categories): {Subject_categories_list}.

Show only the assigned categories separated by ';'. Don't include any explanation.

Processing is therefore carried out on a case-by-case basis by keyword. Using the original keywords in this step, rather than their synonyms, allows for greater flexibility if, at some point, it is deemed desirable to define different synonyms. On the other hand, step 5 is carried out using the thematic groups from step 4, rather than working directly with broader

thematic areas. The latter option always makes it easier to define several areas to combine the different groups.

Step 6: manually unify the clusters into 8 macro-clusters. For example, the Data management field unifies the following thematic clusters in terms of coherence and cohesion: Data management; Databases; Project management; Security and privacy; Government and regulations.

Step 7: Finally, a *Python* script is used to calculate aggregated statistical data on the frequency of ads by thematic group and area.

Step 8: The interpretation of the data obtained focuses on the behaviour of job offers in the field of data analysis for Information Professionals.

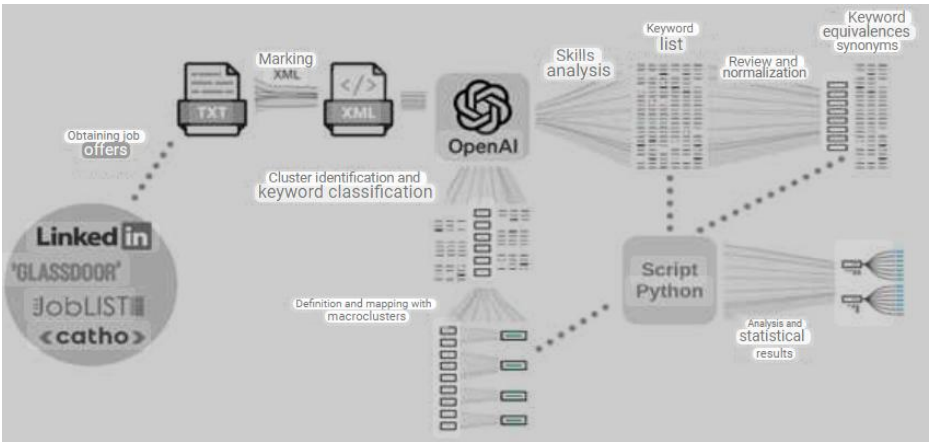

Figure 1: Data Processing Flow.

3. Descriptive Results

This first phase identifies the general characteristics of the sample studied by means of annotated tables, as well as observing the behavior of the circumstances of place, formation, position and corporation.

Table 2: Countries or Regions where the Professional Activity is Carried Out.

Country	No of jobs opening
USA	27
England	14
Brazil	12
Canada	5
Spain	4
Latin American	3
Colombia	2
France	2
North America	2
United Kingdom	2
Others	11
Total	84

Others includes 11 countries with a single occurrence: Argentina, India, Ireland, Kazakhstan, Mexico, Poland, Singapore, South Korea, Sweden, Switzerland and Turkey. There is an internal bias in the case of Brazil, as it includes the 9 advertisements that appeared in the *Catho* network, plus three from other sources. Six vacancies (7.14%) are offered for remote work, three in the USA, two in Latin America, posted from the USA, and one in Colombia.

Table 3: Names and Numbers of University Degrees in LIS.

BIS		MIS	
Archives	2	Digital Information Management	4
Information Management	4	Knowledge organization	2
Information Science	8	Information Management	3
Information studies	4	Information Science	9
Library and Information Science	10	Library and Information Science	8
Library Science	12	Library data	2
Records Management	4	Library Science	8
		Library services	2
		Metadata	2
PhD		Taxonomy/Ontology	4
Information Science	1	Web-metrics	2
Total	45		43

The ads present a wide variety of designations for university courses in LIS, Library and Information Science, which Table 3 shows as bachelor's degree (BIS); master's degree (MIS) and PhD. However, the curriculum is no longer just the academic degree, but the accumulation of skills and experience acquired. That's why, in addition to the degree, 35 of the programmers require years of experience, 3 years on average for bachelor's degrees and 4.5 years for master's degrees. Up to four of them validate the GIS for equivalent numbers of years of postgraduate experience. Two ads require studies to have been carried out at an accredited college or university. In two others, an ALA-accredited master's degree. Four others allow a time-equivalent degree in Information, Data Science or Taxonomy, as an equivalent combination of training and experience. Six times it is stated that 'master's degree preferred'. Sometimes the difference is not in the studies undertaken, but in statistical skills (6), programming skills (2) or skills in handling specific software (5).

Table 4: Community Organizations and Information Centers Originating Positions.

Community Organizations	Information Centers
British Horse Society	Bibliotecas *IBRM* (São Paulo-BR)
Epilepsy Foundation. MD-USA	*McLeod Business Library*, VA-USA
Lausanne Switzerland Information Technology	*The Felipe Herrera Library, IADB*
Metropolitan Council. Saint Paul, MN-USA	*The National Archives*, England (2 job offers)
Observatorio de Drogas de Colombia	*Raymond A. Mason School of Business*, VA-USA
	SOAS (Library Records) University of London
	Stockholm Environment Institute
	University of London
	University of Nevada
	Wilfrid Laurier University, ON-CAN (2 job offers)
Total: 5	Total: 12

The offers for information centers record the impact of libraries and archives from an entrepreneurial perspective of their services, objectives and performance, Table 4. Detailed university archives and libraries manage, link and preserve digital assets by processing data from their resources, especially from research projects and activities, as well as evaluating the institution's performance and facilitating informed decision-making. Thanks to the provision of services from digital objects with linked metadata to share their experience externally and dialogue with user feedback. Although it is true that they require profiles in the LIS field in 23 of the offers (26.1%).

Table 5: Companies Originating Job Offers.

Leading contracting companies	Companies with more than one position	
	Company	Job openings
3CON	*HP*	5
Airbus	*BairesDev*	4
Amazon	*Bloomberg*	3
AstraZeneca	*Expedia*	3
Boeing	*JPMorgan Chase Bank*	3
Compass UOL	*Schneider Electric*	3
Deliveroo	*Curated Real experts*	2
Deloitte	*Netflix*	2
Delta		
Dow Jones		
HireTalent		
IDB Inter-American Development Bank		
INL		
MediaMonks		
Meta		
Petrobras		
PwC		
Resideo		
Standard & Poor's Global 100		
UBS		

In the 79.7% of advertisements that identify the contracting companies, they are large multinationals, Table 5. The fact that they are hired by such large companies underlines the confidence in their ability to develop processes related to the analysis, accessibility and quality of data and its taxonomic organization, in addition to the management of digital assets. They work in all companies and sectors. 4% of the ads are for a "Confidential company".

The three offers from *Schneider Electric* are identical, although they were made separately for activities in France, Poland and Spain. The software company *BairesDev* is based in Virginia (USA), although its job advertisements were two for Latin America and one each for the Brazilian and Colombian sites.

Table 6: Professional Positions Names in Information Centers.

Archives	Libraries
Assistant Archives Officer	Business & Data Analytics Librarian
Information & Records Management Specialist	Data Librarian
Project Archivist (Library Records)	Digital Asset Librarian
Information management	Head of Cataloguing, Taxonomy and Data
Content Indexer	Librarian (2)
Digital Asset Specialist	Librarian Associate
Digital Media Manager (2)	Library Data Analyst
Document Management Specialist	Resource Librarian
Information/Data Management Analyst (2)	Taxonomist and Metadata Librarian
Information Generation Specialist	Technical Librarian
Information and Resources Specialist	

What are the occupations of Information Professionals? (Table 6). Although they retain their original main designation, even six of these names are dedicated to processing, organizing and visualizing business and commercial data. Even so, almost 80 per cent of the remaining job titles do not use the traditional terms for Information Professionals, which directly reflects the changes in the services provided. Particularly at the crossroads of companies in commerce and ICT, other terms are favored in the title to develop combined activities within open processes. Transversal profiles for transdisciplinary teams where graduates from different fields of knowledge using good technological means participate to create, process, organize and manage digital data and information.

Table 7: Professional Positions Names in Corporations.

Data Analyst	Data Organizer	Information Architect
Business analyst	Engineering Linguist/taxonomist	Digital information architect/taxonomist (2)
Data analyst (4)	ESG Taxonomy Data Specialist	Content coordinator
Data analyst & taxonomist	Global grocery taxonomist	Content entry specialist media
Data analytics (Forensic)	Knowledge engineer/ontologist (3)	Content strategist (3)
Data and innovation manager	Ontology consultant	Information architect
Data curation specialist	Ontology/Taxonomy consultant	Data and information architect
Data governance analyst (2)	Taxonomist (6)	Data architecture intern
Data management specialist/taxonomist	Taxonomy strategist	UX Content designer (2)
Data labeling research	Operations/Project coordinator	UX Content strategist (2)
Data protection and privacy analyst	Product data management taxonomist	UX eCommerce content architect (4)
IT Business analyst	Product data taxonomist	
Knowledge analyst	Product taxonomy data	
Knowledge manager (2)	Product manager	
Optimization analyst	Senior semantic modeler	
Taxonomy transformation data analyst	Ontologist	
Technical advisor on sustainable finances	Vocabulary prototyping associate	
Text analyst		
Web analytics analyst (2)		

As already noted in the designations of information professionals in job postings related to digital transformation (**Paletta *et al.***, 2021), it is now confirmed that analyst, in 18 appearances, includes a very diverse expression. Not only the obvious commitment to data analysis, but also to products, texts or web. By close association with the 6 professional labels of knowledge analyst or manager. Undoubtedly, the selection, analysis and study of Web content objects is one of the most comprehensive and growing occupations of information professionals (**Agarwal; Dhar**, 2014).

Data organizer lists the names of those who manage and visualize knowledge in the field of data and products (**Nickerson *et al.***, 2013), while also including tools for communicating with other analysts, but especially with users and clients. This is complemented by the category of information architect with the designations of user experience (UX) analysts, as well as information and data architect, content coordinator, designer or strategist of content objects at the service of e-commerce.

Working together on mixed subjects and tasks requires competences and skills from different fields and backgrounds, as well as different areas of application (**Kennan**, 2016). In a labor scenario of digital transformation, the characterization of the connection to data networks and the real-time sharing of the knowledge produced is crucial. This means that professional performance does not only come from academic experience, but also from the set of skills and experience acquired.

4. Analyzing the Competences and Skills Required

The evaluative analysis is based on the 8 inferred macro-clusters, which unify the clusters of synonyms and therefore keywords enunciated by *Open AI*. Occasionally, the concept classifications resulting from this application may seem strange. But the proposals are respected because, even with some subjectivity, they can be understood within the framework of the suggested set.

Table 8: Macroclusters Descriptive Statistics.

Macrocluster	KW	KW No.	Synonyms No.	Job offers
Business and enterprise management	413	340	113	73
Data management	226	181	88	73
Information management, library science and KO	228	151	68	68
Data analysis and statistics	198	139	78	62
Programming and office automation	193	98	62	60
Web standards and technologies	194	116	68	55
Analytical skills and creativity	116	76	24	49
Domain-specific topics	47	48	42	23

The macroclusters are ordered by their frequency in the job offers. The frequency of keywords (KW) reflects the relative importance of each group about the object of study and the design of the search strategy. In terms of frequency and the number of keywords they represent, the macroclusters are fairly balanced. However, at the top is Business and enterprise management, which almost doubles the next and contains 25.57 per cent of the frequency of keywords and 29.59 per cent of their number. This shows the business and commercial nature of most of the offers. At the bottom, Domain-specific topics have, in both cases with 2.9% and 4.1%, less than half the frequency of that one. From then on, the number of synonyms approaches each other, although certain differences remain at the extremes. This closeness increases in the frequency with which the macroclusters appear in the ads.

Synonyms No. is the total frequency or number of times a synonym or favorite term appears in Job offers. While the simple frequency of synonyms (Syn. single_freq) is the number of times the synonyms of a macrocluster appear in the ads. It coincides with the frequency of appearance in different job offers. They even appear several times in the same advert. 35 of the synonyms belong simultaneously (polyhierarchical) to two macroclusters in Table 8.

4.1. *Business and Enterprise Management Skills* Macrocluster

This macro-cluster achieves high performance in advertisements that demand people capable of managing and controlling the activities and resources of companies and organizations. This means having the tools, behaviors and skills applicable to business operations and strategy, marketing and, of course, teams and the workplace. The macrocluster is included in 73 different job offers (86.9%). It has been classified into 5 synonyms intermediate groups (clusters).

Table 9: Synonym Clusters of *Business and Enterprise Management Skills* Macrocluster.

Cluster	Job offers
Business and enterprise management	62
Organizational, workplace and interpersonal skills	62
Language skills	14
Leadership and management skills	8
Education and enterprise management	1

The skills used in day-to-day operations are critical to business success and are integrated into the Business and Enterprise Management cluster. The advertisements seek the best candidates with skills that are critical to understanding the business and its boundaries. Combine business analysis of tools and data infrastructure with best practices to achieve the performance of business operational processes, along with the improvement of technical procedures and strategies. Business intelligence, project modelling and management, and operations research skills facilitate organizational change in a global and hybrid context. There are also some more specific technical skills, such as financial management, consultancy, contract and tender management and accounting. The managerial ability reaches the highest importance among information professionals in **Choi and Rasmussen** (2009), also in **Khan *et al.*** (2015), along with leadership. Not to mention the importance given to resources that focus on marketing and branding strategies, as well as familiarity with e-commerce and merchandising that promote a good user experience. These last qualities also concern the macrocluster of web standards and technologies.

For jobs with strong interpersonal relationships, there is a high demand for organizational and communication skills, the ability to work in a team and leadership skills. The search for candidates with organizational and procedural skills is therefore not a broad group, but an essential one. The ability to manage time, prioritize urgent activities and optimize processes is essential when dealing with large workloads. To ensure effective and productive human interaction, **Erozkan** (2013) and **Zhang and Chen** (2023) emphasize the role of communication and interpersonal relationships, which is supported by the results of this analysis. Communication skills are emphasized as they enable participants to share, collaborate, empathize and establish correct relationships in personal interactions. Therefore, organizational, workplace and interpersonal skills are closely related to the social skills of analytical skills.

Employers are looking for professionals who can communicate both orally and in writing for marketing and advertising, external reporting, business news, and even academic and technical communication. With these skills, professionals gain confidence in effective relationships with users, consumers and suppliers, but also within the team and organization and with colleagues. **Goulding *et al.*** (1999), and **Erozkan** (2013), rank them among the most required in job advertisements.

Particular attention should be paid to the cluster of language skills, as there are advertisements that require knowledge of the English language. Others, operating in an international context with English as the vehicle language, do not even consider it. Knowledge of French and Spanish is exceptionally required, but only once for each language.

Achieving good interpersonal skills involves other people skills such as culture, impartiality, humility and maturity. Sharing experiences is encouraged, to which positive attitudes, sensitivity or tact can be added to achieve effectiveness when communicating with or assisting customers and suppliers (**Meganck *et al.***, 2020). The ability and conviction to promote diversity, equality and inclusion is very important in the terms and conditions of most tenders. As well as a culture integrated approach to sustainable development.

Many of the skills required to manage data demand precise and complex action. Candidates are required to have multidisciplinary training that allows them to generate social, economic and cultural value in companies. As well as keeping up to date with their initial academic skills, they must keep up to date with company management when they join and then with further training. At the same time, they must have pedagogical resources to channel training activities from methodologies based on theoretical reflection and practical application that favor learning for real services.

Finally, the leadership and management skills focus on managing groups, projects, time and priorities, and taking responsibility for decision making. The occurrences reflect their relevance because of the importance to business and corporate management of the skills identified for leading teams, building communities, persuading, inspiring, monitoring, building consensus, balancing tasks and, of course, managing the team, objects and resources. Executive and senior management are occupations at the highest level of organizational management.

4.2. *Information Management, Library Science and Knowledge Organization* Skills Macrocluster.

Table 10: Macrocluster Synonyms Found in 3 or more Job Offers.

Synonyms	Job Offers	Synonyms	Job Offers
Taxonomies	33	Database management	6
Metadata	14	Searching	6
Categorization and classification	12	Controlled vocabularies	5
Ontologies	12	Knowledge management	5
Information architecture	10	Libraries	4
Content management	9	Archives	4
Knowledge organization	8	Cataloging	3
Outlook	8	Document management	3
Adobe	6	Editing	3

This macro-cluster brings together 81 synonyms in two clusters. The majority are included in the first, Information Management and Librarianship, as only 5 appear in the Information Management cluster. It is logical that the synonym Taxonomy should have a high frequency, as it is one of the elements of the search strategy used to select the advertisements, Table 10. The most frequently preferred terms in this group are also part of other clusters. Thus, Taxonomy, Metadata and Database management in Data management; Content management and Searching in Web standards and technologies; and Ontologies in Web standards and technologies. This reconfirms the transdisciplinary confluence of Information Science when it comes to analyzing and organizing data, information and knowledge (**Bereijo**, 2019).

The job requirements show that data management and organization are based on information retrieval theories. But the applications and work environments vary as much as the recruiting companies and the professional positions on offer, which most often require experience and expertise in metadata, knowledge management and digital objects. In this case, taxonomy is the preferred term among a group of synonyms that provide semantics to data in order to understand differences in meaning and promote the effectiveness of its analysis, interoperability and reuse. But also, to design, deploy and govern institutional assets through an architecture that models and values corporate information (**Federer**, 2018). Although in general terms organizations seek to connect metadata with content objects, whether data, people, texts, products or services, which must be interpreted based on the needs of customers-users, sometimes internal. As well as understanding the cognitive model of the companies requesting candidates capable of supporting the desired objectives.

As a result, requesters will be prepared to define information architecture in the form of site maps, linked taxonomies, metadata schemas, structured vocabularies, content models, ontologies or content relationship diagrams that add value to clients and the organization. Data management requires the ability to design, implement and govern key institutional assets through an architecture that integrates all corporate data and determines service attributes. This involves skills related to KO with information architecture, ICT and *UX Design*. In an operating context in which it is necessary to master the value and principles of data models (**Given; Olson**, 2003). Among the different types of data companies offer are those related to commercial products, whose exact definition and differentiation, designation of attributes and categorization openly demand precision in their classification.

Therefore, the most demanded skills are to create, improve and manage granular taxonomies of e-commerce products, to help companies create more effective catalogues, as well as to refine the user experience (**Papadimitriou *et al.***, 2012) in a correct navigation when searching for objects and data, especially through mobile devices. This is why 34 of the disclosures require experience in designing and managing data taxonomies of commercial and business research objects. Although the most common period of professional experience is three years, five of the vacancies in this group require five years and another seven years. But that's not all, as the ability to use taxonomies is required twice for organizing project data, as well as for coordinating library knowledge. Another nine opportunities call for expertise in updating or improving taxonomic data modelling. This is complemented by 26 advertisements which, in a broad sense, place this correct management of data or big data as a skill derived from the mastery of ontologies, knowledge graphs, classification schemes and other SOCs, as **Mayr *et al.*** (2016) have pointed out. And linked in the announcements to previous training in semantic tools and standards (RDF, OWL, SKOS, SPARQL...) covered in other macroclusters.

“The most demanded skills are to create, improve and manage granular taxonomies of e-commerce products, to help companies create more effective catalogues, as well as to refine the user experience in a correct navigation when searching for objects and data, especially through mobile devices”

Professional experience is also requested in 24 of the advertisements concerning the relationship between taxonomies and metadata schemes, as a function of identifying and distinguishing data, its origin, attributes and contextual structure, as well as determining its integration, interoperability and archiving. This ability to organize data allows candidates to preserve the consistency of the taxonomy with the needs of the organization, making it readable for scientists and for users and consumers. For data to be meaningful, it is essential to know the context of its origin and use, which gives it its true meaning. This is the great contribution of the skills of information professionals to its analysis and interpretation (**Seadle; Havelka**, 2023). Of great affinity with many of the attributes addressed in the macro-cluster of competencies presented in Table 11.

4.3. *Data Management* Skills Macrocluster

Table 11: Synonym Clusters of *Data Management* Macrocluster.

Cluster	Job Offers
Data management	53
Project management	33
Databases	20
Governance and regulations	18
Security and privacy	4
Product Information Management (PIM)	3

Data management skills can be understood as *Master data* or *Digital asset management* (DAM) since they apply to the management and sharing of digital objects in institutions and companies. In most cases, this is data related to commercial transactions whose life cycle is managed by companies through *product data management*. This is a function that helps organize and bring together in a single system all the data and information related to the different development processes and workflows. With sufficient qualification and management of resources to maintain data in different formats over time with the same quality, accuracy and integrity. *Data integration* is the responsibility of *Data governance and regulations,* as it guarantees the *Extract, Transform, and Load (ETL) processes* which aim to ensure that data is up-to-date and valid for the purposes for which it was created. This always includes the standardization of systems, *best practices*, *data privacy and compliance*, and the consistency of the policies on which decisions are made.

This macrocluster also involves skills for searching for standards and planning project management. For the collection, organization, storage and documentation of research data management, as **Kennan** (2016) rightly points out, among other work responsibilities of librarians. Project management requires creativity and interpersonal skills, but above all, managerial and organizational skills, when it is not common for PI's to have been trained in these profiles, as **Kinkus** (2007) notes. Therefore, to become qualified, potential project managers need to be trained. Even if, on four occasions, they try to recruit professional managers and even have a personnel development center. In project management, emphasis is placed on training to describe the content of databases with metadata, which is highly relevant to improving the performance of operations in big data scenarios.

The presence of relational *database* specialists ensures that the logical modeling of tables, views and data indexes is independent of physical storage structures. This capability extends to the query languages used in this cluster to access and manage organizations' data, both structured (SQL, its different standards and technologies or *Neo4J*) and unstructured (*NoSQL*, *Dynamo DB*), both (*PostgreSQL*) and other relational databases (*IBM DB3*, *WRDS*). These programs could be covered in the *Programming and office automation* macrocluster. If they are included in *Data management,* it is because*, in* addition to manipulation, candidates need to plan DB entities and relationships, as someone is required to manage them alongside the data manager. Using data management skills in the workplace means ensuring the use of the database, an understanding of *database* design concepts and access to documentation on the organization's processes.

4.4. *Data Analysis and Statistics* Skills Macrocluster

Table 12: Synonym Clusters of *Data Analysis* and Statistics Macrocluster.

Cluster	Job offers
Data analysis and statistics	51
Artificial intelligence and machine learning	28
Statistical analysis and modelling	11
Mathematics	6
Big data and massive data processing	4

Table 12 shows the importance of experience in data analysis methodology and tools. Working as a data analyst requires mastering technical skills in mathematics, statistics, programming and databases. These collection and analysis tools and techniques aim to identify and extract useful information, model the data and communicate it to executives and the team. An affinity with numbers that is associated with statistical analysis methods, algorithms, with a solid basis in algebra and linear calculus, probability, equations and integrated transformations, *machine learning* and *AI*, always linked to statistical software. These are extended in a small way to six of the advertisements with a competent *Mathematics performance*, aimed at training in the management of theoretical statistics, formal logic *and* set theory, and which includes *Mathematical and theoretical topics and Mathematical modelling and optimization*.

Undoubtedly, the theoretical foundations of statistics and mathematics along with technological skills are basic to data analysis (**Hardin; Horton**, 2017). So, what kind of training in mathematics would be ideal for a data scientist with a degree in Information Science and little formal training? It is certainly different from the training that would be desirable for a mathematician. But it's not as strange as it may seem, because quantitative aspects have always been addressed in IS, as **Brookes** (1980) showed, in the analysis of information issues, especially social ones, and in providing a mathematical description of information spaces. And the number of applications that require frequency statistics and logarithmic laws (**Bawden**, 2008) in metric research that applies mathematics and statistics is growing.

On the other hand, within the multifaceted nature of professional skills for *Data analysis*, one of the biggest obstacles has been the perceived need to learn a programming language and to use advanced mathematics and statistics. Since the development of visual programming tools has made available low-code and no-code software that has levelled the mathematical approach field and allowed more social researchers, and of course Information Science researchers, to apply ML/AI analysis to datasets (**Hersh *et al.***, 2023). It has become a relevant skill to use intuitive programming tools that make it easier to explore data with little programming experience, create data models, evaluate data analysis results and make data science research reports, among others (**Hersh *et al.***, 2023). In any case, the offers demand experience and skills in handling software for *Data management* in Business intelligence, where *Excel*'s statistical functions stand out for their 24 occurrences and whose presence is repeated in the *Microsoft Office* cluster. The same goes for Power Bi's 7 complex spreadsheet applications. *Tableau* also appears four times. While other programs appear once or twice, such as *Stata*, *Qlick*, *Iterix*, *SPSS*, *Spark*, *Variety* or those for AI, such as *Clear Forest* or *Open Calais*.

4.5. *Web Standards and Technologies* Skills Macrocluster

Table 13: Synonym Clusters of Web Standards and Technologies Macrocluster.

Cluster	Job offers
Web standards and technologies	82
Web design and development	76
Google and related technologies	14

Web standards and technologies are recommendations published by the *World Wide Web Consortium* (*W3C*) that any web developer can use to build websites, such as HTML/XHTML and Cascading Style Sheets (CSS) as an easy channel for networking and exchanging all types of cultural, commercial, industrial, scientific and personal data. Moving into the area of business ethics and compliance. Companies and organizations are looking for people with expertise mainly in Semantic Web technology for content management, using formal semantic standards based on RDF (Resource Description Framework), SPARQL (SPARQL Protocol), RDF Query Language and, eventually, OWL (Web Ontology Language). As well as in the RDF application SKOS (Simple Knowledge Organization System) as a model to represent the basic structure and content of the KOS and its implementation in Pool Party Semantic Suite.

Content management and interactive web development are other sought-after skills, aimed at user satisfaction (*user experience*) and search engine optimization (SEO). Depending on the situation, this means having expertise in social media campaigns using innovative *crowdsourcing* platforms. Also, the enterprise collaboration platform *MS-SharePoint* or the editors *Figma*, *WordPress*, *Tableau* and *Miro* to analyze and visualize large amounts of interactive data dedicated to knowledge in an organization or company, summarizing it and presenting it in an easy-to-understand format (**Larson; Chang**, 2016). The list is completed by other content management systems with less presence in the ads: *Draw.io*, *Drupal*, *IBM DataStage*, *Inxight*, *iServer*, *Lucidchart*, *Refinitiv* or *Trello*. There's no shortage of *Microsoft* involvement with *Azure*, *MS-365*, *MS-Visio*, *MS Teams* and *MS Fast*.

The third cluster is grouped according to the criteria of performance on the different *Google* platforms and products. From *Google G Suite,* with typical office software, to *Google Cloud Platform* (GCP), with all the web development applications. There are various analytical tools that are managed. For example, *Looker studio* to convert data into customizable and informative dashboards and reports (**Lakshmanan**, 2017). *Google Analytics* to track the steps users take using data from different websites and applications. *Google Search Structured Data*, marking up structured data in *Google* searches. The combination of *Google* Analytics data with GNIP data from the main social networks. And *Google Sheets*, which could well be classified as data management software.

Table 14: Synonym Clusters of *Programming and Office Automation* Macrocluster.

Cluster	Job Offers
Microsoft Office	64
Programming	35
Microsoft Suite	13
Programming languages and querying	11

4.6. *Programming and Office Automation* Skills Macrocluster

Polyhierarchy is seen in 15 of its synonyms, derived from the fact that it is software applied to the procedures of other macroclusters or to data from specific fields of knowledge. This demonstrates the importance of technological skills, as their value is increased when added to skills in *Web standards and technologies* or those related to *Data management* software.

The advertisements emphasize the need for skills in using the *Microsoft Office* suite of applications to create, edit, store and transmit information via personal computers. Sometimes the skills are required for the whole package, but they are mainly specified for *Word* and *Excel*, the latter already mentioned above for statistical functions in data management. In addition to *PowerPoint*, *Access*, *VBA* and *Windows XP* (see Table 14).

Cluster *Programming,* on the other hand, incorporates specialized knowledge in applications and systems architecture and computer programming interfaces, cloud environments, graphic editors and technical strategies for product development. This includes architectural applications such as *TOGAF*, the *MTLAB* programming and numerical computing platform, the *FileMaker Pro* relational database multiplatform, the *VB.net* object-oriented programming language and even the application of semantic web technologies to the development of *Top Quadrant*'s business solutions.

It applies to the *Programming languages and querying* cluster, with skills in both *querying* (R) and the automation and analysis of statistical data (Java, *Microsoft c#*, *MATLAB*, *Python* and *Sparql*).

4.7. *Analytical Skills and Creativity* Macrocluster

Table 15: Synonyms Development of *Analytical Skills and Creativity* Macrocluster.

Synonyms	Total freq.
Attention to detail	20
Problem solving	18
Collaborative mindset	17
Simultaneous tasks	15
Creativity	15
Analytical skills	13
Meet goals	12
Critical thinking	11
Adapting to change	8

As this is a single cluster, its main synonyms are presented below. Some of its skills are compatible with those of the Organizational, workplace and interpersonal skills cluster, as shown by its three polyhierarchical crosstabs, Table 15. Analytical and creative skills appear in 49 ads (58.33%). As they form a single cluster, they have been divided into synonymous groups: analytical skills, which allow observing, investigating and interpreting a subject to develop complex ideas and solutions. Such as critical thinking and problem-solving skills to respond to ICT developments. Critical thinking increases the likelihood of achieving desired outcomes when applied to solving problems encountered along the stages leading to the achievement of goals (**Halpern**, 1998). And to have independent criteria in complex economic crossroads in the face of misinformation and possible contradictions. Then, being creative, being able to innovate and having initiative are highly valued characteristics in any organization as they prove the commitment and integration of its personnel (**Khan *et al.***, 2015). Finally, the willingness to adapt to change goes hand in hand with the desire to keep up with digital transformation. In the works of **Choi and Rasmussen** (2009), **Zhang and Chen** (2023), technological capability was one of the most sought-after capabilities to keep up with digital transformation. It is the way to respond to users' expectations and their access to a growing range of data and content objects, often requiring organizational change. It is approached from an attitude of collaboration in achieving common goals with colleagues in a group or service, and with attention to detail - intellectual curiosity about external events and new developments.

The development of working skills consists mainly of collaborative working skills, which reflect cooperation between members of a team or between different departments and with external partners. More specifically, the aim is to form cross-functional teams whose members have the necessary skills to solve complex situations independently. In this scenario, the participation of information professionals contributes to the achievement of the project objectives (**Semeler *et al.***, 2019). The advertisements mention the execution of complex processes in simultaneous tasks that require efficient, ethical and demanding attitudes of dedication and initiative in work. In several cases, the job is assigned to people with work experience in similar organizations and is even characterized as decentralized and non-hierarchical.

4.8. *Domain-specific Topics* Skills Macrocluster

Table 16: Synonym Clusters of *Domain-specific Topics* Macrocluster

Cluster	Job Offers
Science and academic publishing	26
Health and medicine	10
Artistic and creative pursuits	2
Social and behavioral sciences	2
Domain-specific topics	2
Education and academic publishing	1

In the *Science and academic publishing* cluster, the skills in climate change, sustainable development and sustainable finance stand out, which are intertwined with those in social and behavioral sciences and socio-environmental sciences, which, in turn, form a polyhierarchical with *science and academic publishing*.

There is a notable demand for the ability to supervise work written by university professors and researchers, from bibliographical surveys to peer-reviewed journals and even minutes of academic meetings. Finally, are the different publication models in academic communication, including the journalistic type, and the use of images in them, particularly when they are aimed at education. As well as the writing of research and development projects and their results, and even the application of the *Theory of Constraints* (TOC) to the solution of scientific problems.

The capabilities of *science and academic publishing extend to* the manual curation of *CHEMBL*, a chemical database of bioactive molecules with pharmaceutical properties, the *Reactome Biology* database and the *Uniprot* repository *of* protein data. Synonyms that are strongly associated in the offers to those of the *health and medicine* cluster which, from experience in health data management (**Beam; Kohane**, 2018), extends to Medicine, medical diagnostics, crisis intervention, Pharmacy and pharmaceutical products, health insurance, food safety, and even to veterinary issues in the e-learning platform *BHS Wise*.

This macrocluster also includes calls for expertise in specific subjects in the field of religion; experience in carrying out artistic and creative activities using the vector graphics editor *Sketch* and practice in sexual orientation.

5. Conclusions

Taking the skills requirements detected in job advertisements as a source, there is a double reason for Information Science (IS) to get involved in analyzing and organizing data:

- That IPs are listed in advertisements as possible candidates for the vacancies on offer. To fulfil the tasks sufficiently, advertisers are looking to integrate into their teams a professional with a wide variety of backgrounds, but whose qualifications are suitable for managing a wide variety of data and information. The PI, as a data and content analyst, teams up with other specialists, particularly in data science, STEM, administration and management, and communications, to share their knowledge and skills. At the same time as interpreting procedures and products in all divisions of the organization. There is a strong tendency to work with professionals from different backgrounds, which means working in a challenging interdisciplinary field.

- Due to their expertise in the information retrieval and knowledge organization techniques required by the announcements. IP contributes its technical expertise in metadata labelling to identify the origin of the data and the frequency of updates, keep it in its series and preserve it in the analysis phase. It is the way in which data is placed in the right context that gives meaning and coherence to the interpretation of its content. As for management, it involves the ability to represent data relating to the services and products of organizations and companies in different operating environments in a coherent way. Especially in big data, whose problem is not only its magnitude, but also its decontextualization. In response, a semantic approach to taxonomies can integrate the results of analyzing this data into information systems and decision-making according to the knowledge provided. This also guarantees the interoperability of vocabularies and metadata schemes to formalize their models and achieve compliance of taxonomies with data analysis.

Because of the business and commercial nature of the jobs on offer, *Business and Enterprise management skills* is one of the macroclusters with the most synonyms. It combines skills for analyzing data tools and infrastructures with best practices for the performance of operational processes, human and material resources and technical business

strategies, some of which are very specific. Business skills extend from economic fundamentals and market knowledge, which imply flexibility for new initiatives such as the user experience for e-commerce. Interpersonal skills are also of interest when it comes to connecting the units involved in an operation and customers. Communication skills, especially for transferring knowledge and the corporate image, are part of this group and are inseparable from linguistic expression skills and mastery of multimedia techniques.

Most of the qualities required in the ads corroborate the results of previous analyses for library and information professionals, so they do not represent a major change in their proposals, although they do represent a consolidation. This is the case with the managerial, operational and leadership skills that enable successful projects to be carried out and which have always characterized their professional performance, as well as those relating to relationships with users. At the same time, their information skills are conducive to the institutional changes required by the organizational context.

About the continuous accumulation of data and its performance in the administration, custody and distribution of digital objects, announcements require technical skills aimed at making data traceable, accessible, interoperable and reusable over a long period of time. *Data management*'s mission is to respond to the expectations and requirements of companies and institutions, and even funding agencies in the case of researchers, in line with the tradition of documentalists and academic librarians to have managerial and organizational skills. In this group, IS's contribution to data processing is determined by the hybridization with which data, information or knowledge is exploited, since, in addition to being put into practice in a specific field, it involves managerial skills which, in turn, are impossible without skills related to the use of databases and ICT. However, few educational institutions prepare them for this, so applicants need to have their university studies include training in these objectives or be trained in the workplace.

Multinational companies and organizations recognize the skills of information specialists in data analysis and management, as demonstrated by the fact that, among the macro-clusters of skills, *Information management, library science and knowledge organization* skills appear in third place by simple frequency of synonyms. This confidence is again based on knowledge of standards, formats and standardized methods of description, presentation and transmission. Advertisements therefore require experience and the ability to research and integrate data from different sources in order to improve the taxonomy of products and services to meet the current and future needs of customers and the company. With the opportunity to create semantic vocabularies suitable for big data, making them readable for users and data scientists. There are not many academic works available that provide a taxonomy for big data analysis techniques. Therefore, here the announcements are at the forefront of research into semantic data architecture.

While the data architect and data scientist are advanced profiles in terms of working with data and computer skills, the analyst profile is more compatible with a more flexible university education. The results show the importance of communication and language skills when interpreting data and giving it discursive coherence in relation to the organization's departments and teams and users. The advertisers' request for people with a university degree in the social sciences and humanities is based on this argument. It's perfectly possible for a IS graduate to work in these teams, but it doesn't come free. For they will certainly need good statistical and ICT skills of the kind already provided by many curricula with multiple current and practical subjects, if not acquired through master's or specialization courses, or even through on-the-job training and work experience.

The analysis platforms made a significant contribution to data processing through their intuitive interfaces. The platforms have given the necessary spreadsheets representative efficiency through clear, interactive and easy-to-share visualizations. In a way that optimizes the analytical power of the data contained in the spreadsheets.

The announcements cover the ICT field broadly, as it affects several of the macro-clusters into which the preferred terms are organized. They are those that involve technical skills and high knowledge of software for application in constantly changing procedures and processes. Digital skills are very close to web navigation skills, its structure and search tools and social networks. Obtaining these skills is, as noted above, covered by the various technology subjects taught on many IS courses. Technological training is therefore inherent to information professionals. Not only because it is an inseparable component of IS, but also because it has become so strong since the digital transformation that it has had a radical impact on competencies, with the consequent reflection in curriculum changes at undergraduate and postgraduate level and even in the very denomination of courses and professional profiles.

Having said that, it should be emphasized that the analysis of job advertisements is a tool that can be used in various ways to characterize the information professions and support decisions about their curriculum planning and, of course, about on-the-job training needs and other human resources issues.

However, there is still much to be explored in terms of the skills that information professionals need to have to be proficient in data analysis and management tasks. Along with the implications this will have for the professional identity of the profession and its councils and associations.

6. Funding

This research project has received support from *Fundação de Amparo à Pesquisa do Estado de São Paulo* (*FAPESP*).

References

Agarwal, Ritu; Dhar, Vasant. (2014). "Editorial—Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research". *Information Systems Research,* v. 25, n. 3, pp. 443-448. *https://doi.org/10.1287/isre.2014.0546*

Bawden, David. (2008). "Smoother Pebbles and the Shoulders of Giants: The Developing Foundations of Information Science". *Journal of Information Science,* v. 34, n. 4, pp. 415-426. *https://doi.org/10.1177/0165551508089717*

Beam, Andrew L.; Kohane, Isaac S. (2018). "Big Data and Machine Learning in Health Care". *JAMA,* v. 319, n. 13, pp. 1317-1318. *https://doi.org/10.1001/jama.2017.18391*

Bereijo, Antonio. (2019). "Theoretical Foundations of Library and Information Science. An Epistemological and Methodological Approach". *Balisages. La revue de recherche de l'Enssib,* v. 1. *https://publications-prairial.fr/balisages/index.php?id=85*

Berelson, Bernard. (1952). *Content Analysis in Communication Research*. New York: Free Press.

Bonatti, Piero Andrea; Decker, Stefan; Polleres, Axel; Presutti, Valentina. (2019). "Knowledge Graphs: New Directions for Knowledge Representation on the Semantic Web (Dagstuhl Seminar 18371)". *Dagstuhl Reports,* v. 8, n. 9, pp. 29-111. *https://doi.org/10.4230/DagRep.8.9.29*

Brookes, Bertram C. (1980). "The foundations of information science: Part II. Quantitative aspects: classes of things and the challenge of human individuality". *Journal of Information Science,* v. 2, n. 5, pp. 209-221. *https://doi.org/10.1177/016555158000200502*

Choi, Youngok; Rasmussen, Edie. (2009). "What Qualifications and Skills are Important for Digital Librarian Positions in Academic Libraries? A Job Advertisement Analysis". *The Journal of Academic Librarianship,* v. 35, n. 5, pp. 457-467. *https://doi.org/10.1016/j.acalib.2009.06.003*

Crisci, Jorge; López-Armengol, María-Fernanda. (1983). *Introducción a la teoría y práctica de la taxonomía numérica*. Washington: OEA. *https://www.researchgate.net/publication/50326234*

Erozkan, Atilgan. (2013). "The Effect of Communication Skills and Interpersonal Problem Solving Skills on Social Self-Efficacy". *Educational Sciences: Theory and Practice,* v. 13, n. 2, pp. 739-745. *https://files.eric.ed.gov/fulltext/EJ1017303.pdf*

Federer, Lisa. (2018). "Defining Data Librarianship: A Survey of Competencies, Skills, and Training". *Journal of the Medical Library Association : JMLA,* v. 106, n. 3, pp. 294-303. *https://doi.org/10.5195/jmla.2018.306*

Gani, Abdullah; Siddiqa, Aisha; Shamshirband, Shahaboddin; Hanum, Fariza. (2016). "A Survey on Indexing Techniques for Big Data: Taxonomy and Performance Evaluation". *Knowledge and Information Systems,* v. 46, n. 2, pp. 241-284. *https://doi.org/10.1007/s10115-015-0830-y*

Given, Lisa M.; Olson, Hope A. (2003). "Knowledge Organization in Research: A Conceptual Model for Organizing Data". *Library & Information Science Research,* v. 25, n. 2, pp. 157-176. *https://doi.org/10.1016/S0740-8188(03)00005-7*

Goulding, Anne; Bromham, Beth; Hannabuss, Stuart; Cramer, Duncan. (1999). "Supply and Demand: The Workforce Needs of Library and Information Services and Personal Qualities of New Professionals". *Journal of Librarianship and Information Science,* v. 31, n. 4, pp. 212-223. *https://doi.org/10.1177/096100069903100405*

Griffey, Jason. (2019). *Artificial Intelligence and Machine Learning in Libraries*. Chicago: ALA, Library Technology Reports.

Hair, Josep; Black, William; Babin, Barry; Anderson, Rolf; Tatham, Ronald. (2009). *Análise multivariada de dados*. 6ª ed. Bookman Editora.

Halpern, D. F. (1998). "Teaching Critical Thinking for Transfer Across Domains. Dispositions, Skills, Structure Training, and Metacognitive Monitoring". *The American Psychologist,* v. 53, n. 4, pp. 449-455. *https://doi.org/10.1037//0003-066x.53.4.449*

Hardin, Johanna S; Horton, Nicholas J. (2017). "Ensuring That Mathematics is Relevant in a World of Data Science". *Notices of the AMs,* v. 64, n. 9, pp. 986-990. *https://doi.org/10.1090/noti1572*

Herndon, Joel. (2021). *Data Science in the Library: Tools and Strategies for Supporting Data-driven Research and Instruction*. London: Facet Publishing. *https://www.facetpublishing.co.uk/page/detail/data-science-in-the-library-by-joel-herndon/?k=9781783304592*

Hersh, William R.; Hoyt, Robert E.; Chamberlin, Steven; Ancker, Jessica S.; Gupta, Aditi; Borlawsky-Payne, Tara B. (2023). "Beyond mathematics, statistics, and programming: data science, machine learning, and artificial intelligence competencies and curricula for clinicians, informaticians, science journalists, and researchers". *Health Systems,* v. 12, n. 3, pp. 255-263. *https://doi.org/10.1080/20476965.2023.2237745*

Husamaldin, Laden; Saeed, Nagham. (2020). "Big Data Analytics Correlation Taxonomy". *Information,* v. 11, n. 1, pp. 17. *https://doi.org/10.3390/info11010017*

Kennan, Mary Anne. (2016). "Data Management: Knowledge and skills required in research, scientific and technical organisations." In: *IFLA World Library and Information Congress: 82nd IFLA General Conference and Assembly.* pp. 1-10. International Federation of Library Associations and Institutions. *https://library.ifla.org/id/eprint/1466*

Khan, Asad; Masrek, Mohamad Noorman; Nadzar, Fuziah Mohamad. (2015). "Analysis of competencies, job satisfaction and organizational commitment as indicators of job performance: A conceptual framework". *Education for Information,* v. 31, pp. 125-141. *https://doi.org/10.3233/EFI-150954*

Khan, Hammad Rauf; Du, Yunfei. (2017). "What is a Data Librarian?: A Content Analysis of Job Advertisements for Data Librarians in the United States Academic Libraries." In: *IFLA WLIC 2018 – Kuala Lumpur, Malaysia – Transform Libraries, Transform Societies in Session 139.* pp. 1-9. Education and Training, Science and Technology, and Continuing Professional Development and Workplace Learning. *https://library.ifla.org/id/eprint/2255*

Kinkus, Jane. (2007). "Project Management Skills: A Literature Review and Content Analysis of Librarian Position Announcements". *College & Research Libraries,* v. 68, n. 4, pp. 352-363. *https://doi.org/10.5860/crl.68.4.352*

Krippendorff, Klaus. (2018). *Content Analysis: An Introduction to Its Methodology*. Sage Publications. *https://doi.org/10.4135/9781071878781*

Lakshmanan, Valliappa. (2017). *Data Science on the Google Cloud Platform*. O'Reilly Media. *https://www.oreilly.com/library/view/data-science-on/9781491974551*

Larson, Deanne; Chang, Victor. (2016). "A review and future direction of agile, business intelligence, analytics and data science". *International Journal of Information Management,* v. 36, n. 5, pp. 700-710. *https://doi.org/10.1016/j.ijinfomgt.2016.04.013*

Lomax, Jane. (2019). "How to use ontologies to unlock the full potential of your scientific data – Part 1." *SciBite.Blog* (blog). *https://www.scibite.com/news/how-to-use-ontologies-to-unlock-the-full-potential-of-your-scientific-data-part-1*

Mayr, Philipp; Tudhope, Douglas; Clarke, Stella Dextre; Zeng, Marcia Lei; Lin, Xia. (2016). "Recent applications of Knowledge Organization Systems: introduction to a special issue". *International Journal on Digital Libraries,* v. 17, n. 1, pp. 1-4. *https://doi.org/10.1007/s00799-015-0167-x*

Meganck, Shana; Smith, Joshua; Guidry, Jeanine P. D. (2020). "The skills required for entry-level public relations: An analysis of skills required in 1,000 PR job ads". *Public Relations Review,* v. 46, n. 5, pp. 101973. *https://doi.org/10.1016/j.pubrev.2020.101973*

Moreiro-González, José-Antonio. (2017). "Superación de los límites marcados por la disciplinariedad en el ámbito de la información y documentación". *PontodeAcesso,* v. 11, n. 1, pp. 26-40. *https://periodicos.ufba.br/index.php/revistaici/article/view/23174/14679*

Nickerson, Robert C.; Varshney, Upkar; Muntermann, Jan. (2013). "A method for taxonomy development and its application in information systems". *European Journal of Information Systems,* v. 22, n. 3, pp. 336-359. *https://doi.org/10.1057/ejis.2012.26*

NISO. (2005). *Guidelines for the Construction, Format and Management of Monolingual Controlled Vocabularies*. Bethesda, Maryland: NISO Press.

Oliver, Jeffrey C.; Kollen, Christine; Hickson, Benjamin; Rios, Fernando. (2019). "Data Science Support at the Academic Library". *Journal of Library Administration,* v. 59, n. 3, pp. 241-257. *https://doi.org/10.1080/01930826.2019.1583015*

Paletta, Francisco-Carlos; Pastor-Sánchez, Juan-Antonio; Moreiro-González, José-Antonio. (2021). "Competíªncias e habilidades digitais requeridas aos profissionais da informaí§í£o nos anúncios brasileiros de emprego na web". *Profesional de la información,* v. 30, n. 1, pp. e300108. *https://doi.org/10.3145/epi.2021.ene.08*

Paletta, Francisco Carlos; Moreiro-González, José-Antonio. (2021). "A transformação digital e os impactos no mercado de trabalho: estudo dos anúncios de emprego na web para profissionais da informação no setor privado". *Information Research,* v. 26, n. 3, pp. 904. *https://doi.org/10.47989/irpaper904*

Papadimitriou, Panagiotis; Tsaparas, Panayiotis; Fuxman, Ariel; Getoor, Lise. (2012). "TACI: Taxonomy-Aware Catalog Integration". *IEEE Transactions on Knowledge and Data Engineering,* v. 25, n. 7, pp. 1643-1655. *https://doi.org/10.1109/TKDE.2012.54*

Saracevic, T. (1995). "Interdisciplinary Nature of Information Science". *Ciência da informação,* v. 24, n. 1, pp. 36-41. *https://doi.org/10.18225/ci.inf.v24i1.608*

Seadle, Michael; Havelka, Stefanie. (2023). "Information science: Why it is not data science". *Data and Information Management,* v. 7, n. 1, pp. 100027. *https://doi.org/10.1016/j.dim.2023.100027*

Semeler, Alexandre; Pinto, Adilson; Koltay, Tibor; Dias, Thiago; Oliveira, Arthur; González, José; Rozados, Helen Beatriz Frota. (2024). "Algorithmic Literacy: Generative Artificial Intelligence Technologies for Data Librarians". *EAI Endorsed Transactions on Scalable Information Systems,* v. 11, n. 2, pp. 1-11. *https://doi.org/10.4108/eetsis.4067*

Semeler, Alexandre Ribas; Pinto, Adilson Luiz; Rozados, Helen Beatriz Frota. (2019). "Data science in data librarianship: Core competencies of a data librarian". *Journal of Librarianship and Information Science,* v. 51, n. 3, pp. 771-780. *https://doi.org/10.1177/0961000617742465*

Stuart, David. (2020). *Practical Data Science for Information Professionals*. Facet Publishing. *https://doi.org/10.29085/9781783303465*

Štuikys, Vytautas; Burbaitė, Renata. (2024). "Introducing Data Science Concepts into STEM-Driven Computer Science Education." In: *Evolution of STEM-Driven Computer Science Education: The Perspective of Big Concepts.* Štuikys, Vytautas; Burbaitė, Renata (Eds.), pp. 217-245. Springer Nature Switzerland. *https://doi.org/10.1007/978-3-031-48235-9_8*

Zhang, Jin; Chen, Jianyao. (2023). "Skill analysis of library and information science professionals". *Journal of Librarianship and Information Science,* pp. 09610006231207656. *https://doi.org/10.1177/09610006231207656*