Relationship between the Metadata and Relevance Criteria of Scientific Data

GUILAN ZHANG (ID)

JIAN WANG (ID)

JIANPING LIU (ID)

YAO PAN (ID)

Author affiliations can be found in the back matter of this article

]u[ubiquity press

ABSTRACT

The purpose of this study is to explore the information retrieval process in scientific data and to better understand the concepts and internal relationships of metadata and relevance criteria. Qualitative and quantitative analyses were performed using interview and eye movement data from 36 subjects. The results show that users paid attention to 45 types of metadata and used nine relevance criteria to judge the relevance of scientific data. There was a complex relationship between the metadata and criteria, mainly manifesting as one stimulus–multiple responses and multiple stimuli–one response. Metadata associated with the relevance criterion of topicality is the most complex, which includes common metadata and subject-related metadata. Metadata associated with the other relevance criteria (such as quality and authority) has no obvious professional characteristics. What's more, because of the essential difference between scientific data and documents, users use different criteria. When retrieving data, users pay more attention to the availability of data and whether they can be further analyzed and processed. This study clarifies the concepts of metadata and relevance criteria as well as their roles in relevance judgments. In addition, this study deepens the understanding of the scientific data relevance judgments and their cognitive process and provides a theoretical basis for improving scientific data-sharing platforms.

1. INTRODUCTION

Scientific data are series of original data, processing data, and result data produced by scientists in the process of scientific research. Metadata are data about data that describes the properties of the data. The amount and usage of data are booming. Data in science, medicine, business, and other fields are predicted to soon reach critical mass (Christine, 2014). Mass scientific data resources are the basis of scientific research. Scientific data sharing is the key to realizing information value and data reuse, and it is an important way to promote the flow of scientific data among researchers so that these data may be transformed into scientific conclusions (Deng Zhonghua, 2017). To better share and reuse scientific research results, which avoids wasting research funds, scientific communities in all fields are building scientific data sharing platforms, which provide mass data resources for researchers. Before reusing data, users must assess the data's relevance. They seek assurance that the data can be understood, and they must trust the data (IM Faniel, 2010). In contrast to other information carriers (like literature, images and videos), scientific data are highly purpose, targeted, subject-related, and technical. In addition, specialist software tools are usually required to analyze the results of scientific data. From document retrieval to data retrieval, information types have substantially changed, and an urgent question is whether the user retrieval modes and strategies have change accordingly.

Relevance is the relationship between the task at hand and the information presented to us. It is a core concept in the field of information science. A relevance judgment determines whether such a relationship exists or not, and relevance criteria are factors that affect judgment. In the age of data, people needed to quickly find information related to their own needs from a large amount of information, so relevance studies have become increasingly important. Many previous studies have focused on relevance in various contexts in order to provide a theoretical basis for various information retrieve systems. The results show that relevance judgment differs when the research situation changes (Taylor, 2009).

There is a continuous and indivisible cognitive process that extends from gazing at information to stimulating the brain to use criteria to make a judgment and includes stimulus, attention, and memory extraction (Gao and Xiaoyun, 2003). When a study focuses on scientific data, a user's retrieval behavior may change. How users search for relevant scientific data still needs further study. To better understand relevance judgment, this study explores the relationship between metadata and relevance criteria. Its aim is also to determine the following:

(a) what types of metadata users pay attention to when retrieving scientific data;
(b) what criteria are used in scientific data relevance judgment and how they differ from document criteria;
(c) the relationship between metadata and relevance criteria and the rules this relationship follows.

By answering these three questions, we hope to make the following contributions. First, the concepts of metadata and relevance criteria will be defined clearly, enriching relevance research. Second, the IR community will gain a deeper understanding of how users make their relevance judgment decisions within a data context. Finally, the findings of this study will have implications on the design of data retrieval systems. A comparison of the relevance criteria used for different media and situations will guide designers of different IR systems. Because data users have different needs and motivations, investigating relevance criteria within different data contexts will help the design of systems that meet their needs.

2. LITERATURE REVIEW

2.1 RELEVANCE JUDGMENT

Relevance judgment was critical in relevance researches. Lots of relevance researches were based on relevance judgment process (Ingwersen, 2011). From those studies, we could see relevance judgments were considered highly complex and cognitive (Xie and Benoit, 2013). Relevance judgment results were dynamic and contextual rather than constant (Anderson, 2005; Saracevic, 2016). The factors affecting judgment were mainly divided into internal and external factors. Internal factors mainly included individual cognitive differences. External factors mainly included information types, task, and pressure. Relevance judgment was a continuous process, which was closely related to the user's cognitive processing. In order to analyze relevance judgment more

accurately, scholars added lens model and making-decision theory into study (Wang Peiling, 1998; Soo Young Rieh, 2002). They thought users used limited environment information to make rational judgment, and those information could be linearly weighted. In addition, scholars used SEM (Structural Equation Model) to quantitatively analyze the factors' weight on relevance judgment (Xu Calvin, 2006; Xiaolun Wang, 2014; Jianping Liu, 2019).

Wang Peiling (1998) proposed document selection model based on lens model, presenting a relatively clear relevance judgment process. Balatsoukas (2012) recorded user's AOI(Area of Interest) using eye movement devices and got web relevance criteria through deep-interview, making relevance judgment more intuitive and visual. The studies demonstrated that these judgments were not singular actions but were instead embedded in very diverse and complex search and research practices (Anderson, 2005).

2.2 METADATA

Metadata was data about data. Metadata contained a lot of information. In the paper, metadata was limited to data structures, including datasets names, relationships, field, etc. Scholars preferred to call it metadata, including title, authors, time, abstract, key words, etc. Marchionini (2009) thought metadata should facilitate sense making during the relevance judgment process and not act merely as information access points. Metadata's content and layout would affect users' judgment and satisfaction (Drori, 2003). So scholars designed a lot of experiments to study the influence of metadata on relevance judgement, including dynamic abstract (Paek et al., 2004), data thumbnail (Dziadosz and Chandrasekar, 2002), and metadata classification (Rele et al., 2005).

Panos Balatsoukas (2010) thought participants preferred metadata that were easy to understand and grouped into categories. In order to better understand users' concerns about metadata, Balatsoukas used eye-tracking device to measure metadata quantitatively. The results showed different cognitive efforts will lead to different relevance judgments. The main metadata they focused on included title, abstract, URL, etc.

Curtis Watson (2013) studied how middle school students judge the reliability and relevance of web information. The study found participants more liked topical metadata and reliable metadata. Users' cognitive level and perceived authority, webpage graphic design, writing style, and authors all could affect users' relevance judgment.

2.3 RELEVANCE CRITERIA

Schamber (1996) defined relevance criteria the factors that influenced the user's data relevance judgments. In the 1990s, many empirical studies had been carried out to identify document relevance criteria or factors in different problem domains. For example, Barry (1994) interviewed 18 academic users who had requested an information search for documents related to their work to categorize their relevance criteria. And finally she identified 23 categories in 7 groups. Magluaghlin and Sonnenwald (2002) asked 12 graduate students with real information needs to judge the relevance of the 20 most recent documents and identified 29 criteria in 6 categories. Other researchers who had done similar research were Park (1993), Cool (1993), Westbrook (2001) and so on.

In the 21st century, the information carriers were diversified, and the empirical research on the relevance criteria also developed from literature to image (Markkula & Sormunen, 2000; Sedghi, 2008, 2012; Hamid, 2010, 2016; Tsai-Youn Hung, 2018), WWW (Tombros, 2003, 2005; Crystal & Greenberg, 2006; Savolainen, 2006; Yang-woo Kim, 2014; Yung-Sheng Chang, 2018), music (Laplante, 2010; Inskip, 2010), video (Yang, 2010; Albassam, 2017), mobile commerce (Xiaolun Wang, 2013) etc.

In conclusion, it could be found that the relevance criteria research was consistent with the development of current mainstream information carriers. There were common criteria across different contexts and information carriers (Barry & Schamber, 1998; Xu, 2006; Saracevic, 2015). At the same time, different information carriers have unique relevance criteria (Zhang, Wang & Liu, 2018). Scientific data has now become an indispensable material for research, work and study. So, there have been studies on scientific data relevance criteria. Sabbata (2012) carried out a study on the geographic data relevance, and found that users dealt with geographic entity data differently from traditional data. Relevance criteria specific to geographic data emerged,

such as directionality, spatio-temporal, visualization. Gao Fei (2017) focused on the relationship between scientific data user relevance criteria and clues. Wei Caoyuan (2018) carried out a research on the relationship between the scientific data relevance criteria and perceived value, and its influence on the relevance judgment. The results showed that the relevance criteria promoted the formation of perceived value, and the scientific data retrieval behavior was similar to the purchase behavior of commodities. Zhang Guilan (2018) also conducted a classification study on the relevance criteria of scientific data.

At the same time, these studies have revealed some limitations. Firstly, different scholars put forward different relevance criteria, and there were great differences in classification. For example, in the image relevance criteria study, Youngok (2000) presented 9 criteria and Hung (2005) presented 12 criteria. Secondly, The expression of terms is vague, and the relevance criteria of the same meaning have different expressions in different studies, for example, accuracy and reliability, utility and usefulness (Xu, 2006). An important reason for this limitation was that scholars didn't have a consistent understanding of the concept of relevance criteria and didn't make a clear distinction between metadata, clues and criteria. Wang Peiling (1994) proposed a document selection model based on the lens theory, presenting a relatively clear relevance judgment process. In the model, the literatures' information elements and relevance criteria were clearly defined. Information elements provided clues for users, and the relevance criteria were the product of cognitive processing in the mind. Balatsoukas (2012) recorded the user's AOI (area of interest) in the retrieval process through eye tracker, and explored the web page relevance criteria through in-depth interview, making the relevance judgment process more intuitive and visual.

Scholars have done a lot of research on metadata and relevance criteria. Various empirical studies proved that both metadata and relevance criteria affected user relevance judgment. And there must be a relationship between metadata and relevance criteria. But there were few studies focused on this point. So the paper would concentrate on the relationship between metadata and relevance criteria to help better understand relevance judgment process.

3. METHODS

3.1 CONCEPT MODEL AND RESEARCH HYPOTHESIS

Strictly speaking, relevance does not behave; people behave (Saracevic, 2015). Scholars have carried out a large number of experiments to explain relevance by observing and describing a user's behavior in the relevance judgment of information. Relevance judgment is an information cognitive process (Gwizdka, 2014). Wang Peiling proposed a document selection model in which users process the information to form relevance criteria. David Bodoff's integrated model of browsing and search relevance argues that users make judgements after focusing on document characteristics. According to cognitive psychology, in the retrieval process, a user's eyes will constantly focus on metadata from the outside world. The metadata then stimulates the user's brain to process the received information.

Psychologist Egon Brunswik proposed the lens model to solve the problem of limited human rational judgment. The environment information that the user pays attention to is the lens, and the perceived stimuli form the clue. The objects that the clue reflect in the memory are mental representations. Based on this theory, the concept model was proposed (**Figure 1**). Scientific data metadata make up the object environment information, which can then stimulate users. Relevance criteria are mental representations, which are the users' responses.
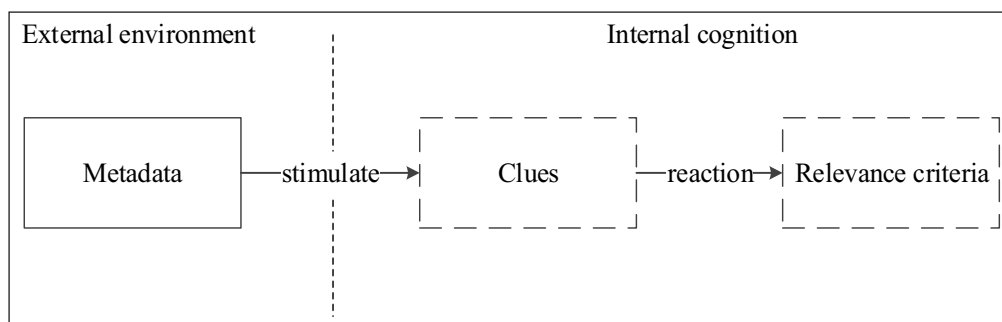

Figure 1 Concept model.

This study hypothesizes that metadata and relevance criteria are joined through clues. The objective stimuli the users feel is the clue, and the responses that the brain forms after processing clues are relevance criteria. Relevance criteria are influenced by external objective metadata information. The scientific data metadata form the independent variable, and the relevance criterion is the dependent variable. This study mainly explores the relationship between the independent variable and dependent variable.

Variable definitions

(a) Scientific metadata describe and interpret the content, attributes, and characteristics of scientific data. They are objective and real.

(b) Clues are the interaction product between the objective world and the subjective individual, and are the intuitive reactions formed by brain responding to the metadata stimulus.

(c) Relevance criteria: These criteria are created by the further processing of clues, and they are the factor that influences users to make relevance judgments.

3.2 EXPERIMENT AND METHODS

The purpose of this experiment was to explore the relationship between scientific data metadata and relevance criteria. The experiment evaluated each type of metadata and its fixation dwell time, the relevance criteria users employed to make relevance judgment, and the relationship between them. The study combined contextual experiments with interviews to obtain data (*Figure 2*). An eye tracker was used to record each user's fixation behavior during retrieval, and screen-capture video recorded the user's browsing and clicking behavior. The information processing in each user's mind was obtained through video playback and in-depth interviews.
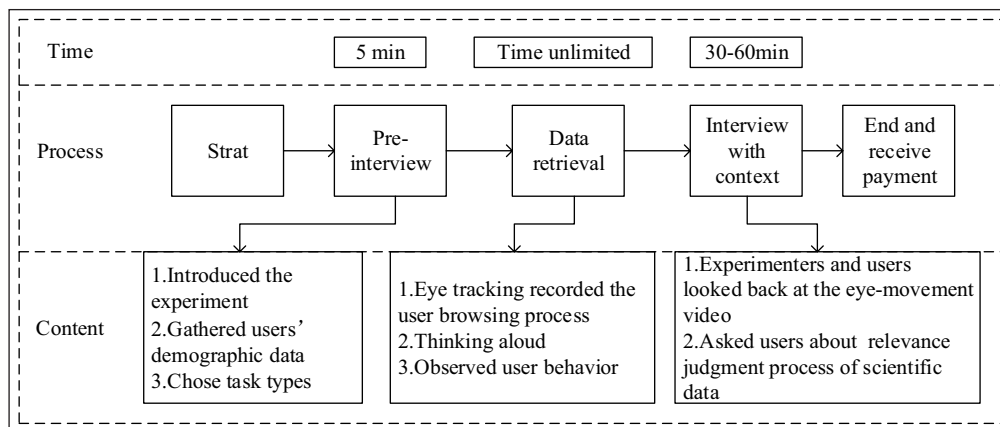

Figure 2 Contextual experiment process.

In our contextual experiment, first, each user's professional background and how frequently they used scientific data were obtained through a questionnaire. Next, the researchers introduced the eye movement equipment and explained the experiment. Then, users started searching for data according their topic, and this was recorded by video. Finally, the researchers interviewed the users while replaying the video.

3.2.1 Participants

The participants' choice followed three principles. First, participants often retrieve scientific data. Second, participants took part in experiment voluntarily. Finally, the retrieval task didn't involve scientific secrets. By referring to the number of subjects in previous studies and the amount of tasks involved in later data processing, we finally selected 36 participants.

Questionnaires were handed out to students of a data analysis course to recruit appropriate subjects. According to their answers, 36 graduate students who often used scientific data sharing platforms (such as NCBI,[1] NBS,[2] or the national meteorological data network[3])

---

1   National Center for Biotechnology Information: *https://www.ncbi.nlm.nih.gov/.*

2   National Bureau of Statistics of China: *http://www.stats.gov.cn/enGliSH/.*

3   *http://data.cma.cn/site/index.html.*

were chosen to participate in our experiment. They were majors in agricultural economics, crop science, regional development, biological science, feed nutrition, and environmental development. They were between 22 and 30 years old. The participants represented the young scientific data retrieval group.

We promised to respect participants' privacy and that the data will only be used for research. At the end of the experiment, the participants were paid.

3.2.2 Task

Scientific data is highly specialized. Different research areas have different data platforms. In order to meet their actual retrieval needs, participants were given the right to make their own choices. Participants chose research task that they were interested in. And they determined the data sharing platform needed to complete the task. At the same time, each participant must retrieve at least 3 related scientific data to keep the task challenging.

The participants searched for scientific data according to their usual retrieval habits, without any limits on the type and number of data sharing platforms. And search time was not limited so that the participants would feel no pressure to complete the task.

3.2.3 Data Collection

Eye movement collection
Participants' eye movements were captured through an eye-tracking device (EyeLink 1000 plus) as they searched for scientific data. The device had a 17-inch screen with the eye tracker embedded in it and permitted a 250-Hz sampling rate with gaze point accuracy down to 0.15°. Before the retrieval, each participant's eyes were adjusted to ensure the accuracy of their fixation points. In general, a saccade lasts no more than 100 ms (Duchowski, 2007). Hence, we set the minimum fixation to 200 ms, which is the average time people need to read when solving problems (Rayner, 2009; Lorigo, 2008). This means that a steady fixation must last more than 200 ms (Balatsoukas, 2014).

Interview data collection
After data retrieval, participants were invited to participate in semi-structured interviews based on the video of the retrieval process. The questions were as follows.

(a) Platform choices and data relevance. Why did you choose this platform? Were the data relevant or satisfactory? Did the data support the task? If not, what other data did you need?

(b) Data choices process. What information did you pay attention to during browsing? Why did you focus on this information? How did this information affect you? Why did you click on these data?

(c) Inactivity after attention. Why didn't you choose to click on the data that you already focused on? What information made you decide not to click on them?

3.3 EYE TRACKER

The eye tracker hypothesis – that the tasks users choose are determined what they see –is based on brain–eye consistency (Nielsen, 2010). The examination of eye movements (such as the number of fixations and fixation length) has been used in psychology and cognitive science research as a means of understanding the processes of reasoning and decision making (Rayner, 2009). Since then, the movements of our eyes have been studied to uncover the relationship between eye movement and cognitive processes as well as to identify how visual stimuli affects us and influences the decisions we make (Jacob and Karn, 2003). Relevance judgment is a cognitive process, which is difficult to observe and measure. To better study relevance judgment, an eye tracker is a good solution, because cognitive processes such as mental effort and attention can be inferred using eye movement data such as saccades and fixations. In an information retrieval context, the number and length of fixations have been used to study the attention and energy users have focused on search results lists and web pages. The experiment proved that cognitive effort is the highest for partially relevant documents and lowest for irrelevant documents (Gwizdka, 2014, 2015; CT Yang, 2011). Bucher (2006) studied attention patterns in the process of news content selection with an eye tracker, and the results show that some obvious visual stimuli (prominent pictures or graphics) actively attract attention. Using eye movement devices,

Papaeconomou (2008) studied how users with different learning styles use relevance criteria to judge the usefulness of web pages. Balatsoukas (2010, 2010) studied relevance criteria usage in the relevance judgment process using an eye tracker. The results showed the effects of ranking order and metadata (title, summary, and URL) on the use of relevance criteria. Wenjing Pian (2016) used an eye tracker system to capture participant eye movements and found that people focus on different information and used different criteria in three types of use contexts.

An eye tracker is a good tool for recording the data information input of users in the retrieval process. This study combines eye tracker data with interview data, thus bringing together cognitive and behavioral approaches in the study of relevance judgment behavior within the context of user–search engine interaction.

3.4 DATA ANALYSIS

3.4.1 Eye movement data collection

This research focuses on the metadata to which users pay attention. Similar metadata (e.g., "title," "abstract," and "name") can be grouped together As shown in *Figure 3*, each column was treated as the same area of interest (AOI) in the data list (e.g., "entry name," "protein name," or "gene name"). The scientific data AOIs users paid attention to were labeled by Data Viewer, which is a commercial eye movement analysis tool. The data processing removed extraneous data such as post-click comments, residual comments, errors, and ads. A total of 3,359 final AOIs were obtained, which were divided into 45 types of metadata, such as "name," "data content," "title," "keyword," "author," "publish time," "links," and "data time." The recorded eye-movement data include the dwell time as well as the number and percentage of fixations.
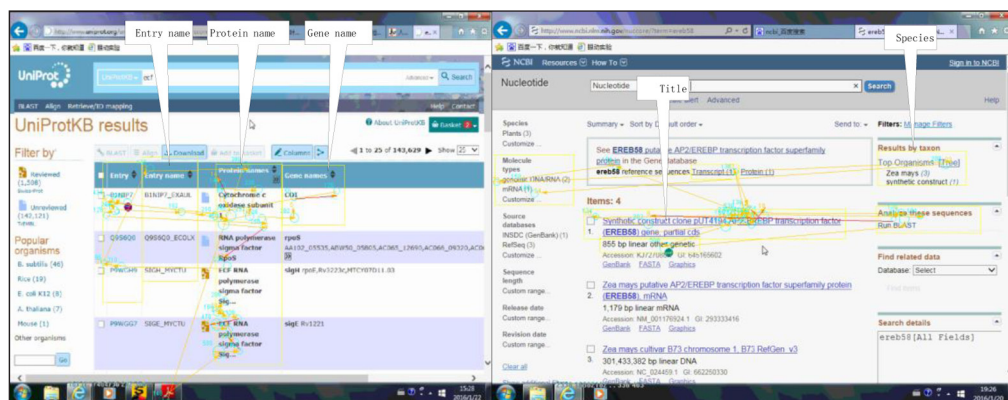

Figure 3 Division of AOIs.

3.4.2 Interview data analysis

Interview data were coded by three coders using NVivo 11 to ensure coding consistency and objectivity. The coding process included three stages. First, coders discussed and designed a coding table through precoding. Second, the interview data were coded according to the coding table (Appendix 1). Finally, relational nodes were coded by combining interview and eye movement data.

Node coding

This experiment involved three variables based on the concept model: metadata of interest, clue responses to presented information, and the relevance criteria used in the relevance judgment (*Table 1*). The code was divided into five tree-like nodes: criteria, clues, data type, databases, and metadata. According to the interview content and AOIs, the secondary nodes were constantly revised (Appendix 1).

Table 1 Example coding.

INTERVIEW DATA	METADATA	CLUES	CRITERIA
This is the voltage and also the condition. We need to compare which one works better.	experimental method	more effective	
We want to retrieve humidity, temperature. I want these indices.	Name	match my study	topicality
Do not need to pay, agricultural academy Intranet can enter.	Share level Free	I can share it whether it is free	availability

Through coding, a total of 376 criteria nodes were obtained, including criteria such as *topicality*, *authority*, *quality*, *currency*, *availability*, *standardization*, *usability*, *convenience*, and *comprehensiveness*. A further 320 clue nodes were obtained, including 66 node types such as "better" or "can't be opened." Finally, 628 information elements were obtained that included 45 node types such as "title," "name," "abstract," and "data time" (**Table 2**).

Table 2 Number of coding nodes.

NODE	MATERIAL SOURCES	NUMBER
criteria	33	376
clues	32	320
data type	22	31
databases	18	43
metadata	33	628

Relationship node coding

We found there were some relationship between metadata and criteria in a certain sentence. Metadata always come first, then criteria will come. And the sentences reflect the cognitive process of scientific data. So the study encoded the relationship between the metadata and criteria (**Table 3**).

(a) **Auxin** is the **keyword**. The term is **gene transcription protein**. Then, I found the **protein name** that **I'm looking for** in the **title.** So, the data is **relevant**.

(b) We need the **data from 2010 to 2020**. But these **data are too old**, and this source does not have data from 2010. We would have to spend money to access these data, which is **more trouble**.

Table 3 Example relation code.

	METADATA	RELATIONSHIP	CLUES	RELATIONSHIP	CRITERIA
1	Key word Title Name	stimulate	I'm looking for	process	Topicality
2.1	Data time	stimulate	Time span Update time	process	Topicality Currency
2.2	Cost	stimulate	Spend money to buy	process	Convenience

Metadata, clues, and relevance criteria were linked together through relational node coding in order to link user attention with a series of cognitive responses. Cross node analysis was performed on relational nodes to obtain weights. For instance, in **Table 4**, which shows an example of partial cross analysis, when users saw data time, they thought about the *time dimension* seven times, thought the *data are new* once, and thought the *data are old* three times.

Table 4 Example of partial cross analysis.

METADATA CROSSOVER NODES	CLUES	DATA TIME	AUTHOR	PUBLISH TIME
time dimension		7		
time is new		1	4	
time is too old		3	2	
time span		7		
continuous data		4		
difficult to obtain		1		
research content is similar			1	
Skepticism			1	
better known			2	

4. RESULTS

Through the contextual experiment and interview, the following results were obtained. (1) Users mainly paid attention to 45 types of scientific metadata, and used a total of nine relevance criteria to make relevance judgments when searching for data. (2) The conceptual model was validated. In a relevance judgment, the clues were the stimulus felt by the users when seeing the metadata, and the relevance criterion was the response formed by the clues.

4.1 SCIENTIFIC DATA METADATA AND RELEVANCE CRITERIA

4.1.1 Scientific data metadata

As shown in *Figure 4*, 45 types of metadata and their dwell times were obtained. The longest dwell time was for "name" (430,638 ms). The percentage of total dwell time spent on "name" was 14.65%. The other metadata with the most dwell times were "data content" (8.2%), "title" (7.25%), "keyword" (7.15%), and "data time (6.73%). The results in *Figure 5* further show that topicality was mainly invoked when the users saw metadata such as "name," "data content," "title," and "keyword." Hence users spent most of their energy on topical judgment (Balatsoukas, 2012). These metadata were common across all fields of study.
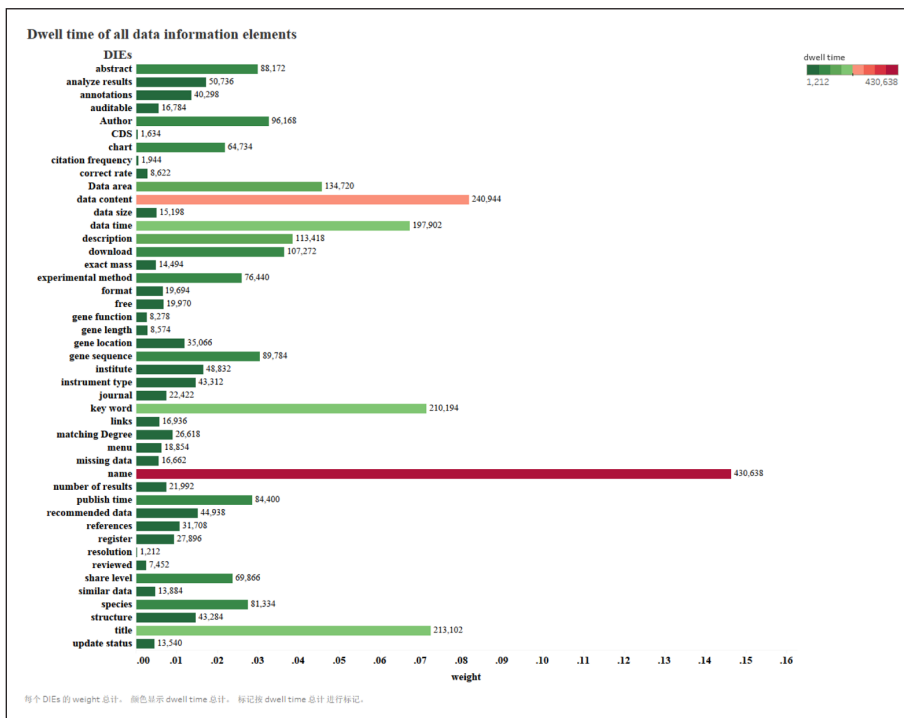
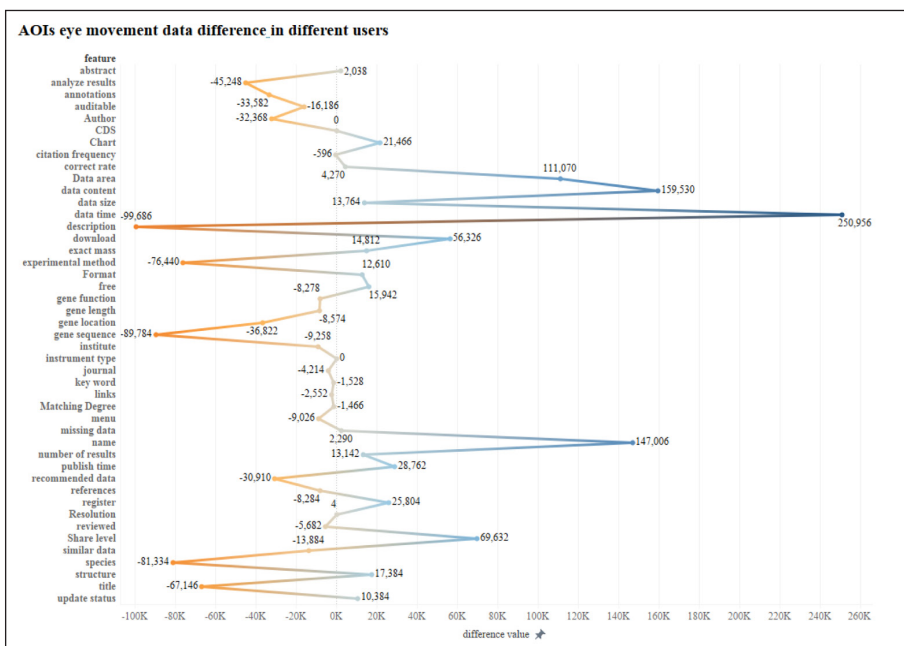Users paid less attention to metadata with shorter dwell times. The shortest dwell time was for "resolution" (1,212 ms), which accounted for 0.04% of the total dwell time. In addition, the dwell times for "CDs,"[4] "citation frequency," "reviewed," and "gene function" were respectively 1,634 ms, 1,944 ms, 7,452 ms, and 8,278 ms, which accounted for 0.06%, 0.07%, 0.25%, and 0.28% of the total dwell time, respectively. These metadata varied according to field of study. Users majoring in meteorological remote sensing paid attention to "resolution." Users majoring in biological genetics paid attention to "reviewed" and "gene function."

Therefore, a comparative analysis was made on the different fields of study of the users. The users majoring in bioscience, crop science, and feed research were classified as the "experimental" group, because their main data were obtained from laboratory experiments, and the purpose of data retrieval was mainly to compare those data with their own research data. The users majoring in agriculture economics, meteorology, and regional development research were classified as the "investigational" group, because their main data were collected from data-sharing platforms such as NBS, NCBI, and the National Earth System Science Data Sharing Infrastructure.

The results of this comparative analysis show that there are significant differences in the dwell times of some subject-relevant metadata, such as "data time," "data content," "gene location," "gene sequence," "gene function," and "experimental methods." Investigational users paid more attention to "data time and "data content" because data from the search were their research objects and the research always had certain requirements with respect to region and time. Experimental users paid more attention to "gene location," "gene sequence," and "gene function," which are subject-relevant metadata. Moreover, they performed a lot of laboratory experiments, so "experimental methods" also was paid more attention.

At the same time, the study also found that there are no significant differences in metadata common to the two groups such as "title," "keyword," "abstract," "number of results," and "institute." These metadata are essential for relevance judgment, regardless of the field of study.

4.1.2 Scientific data relevance criteria

Nine relevance criteria were obtained through data coding: *topicality*, *availability*, *quality*, *completeness*, *authority*, *currency*, *convenience*, *usability*, and *standardization* (*Table 5*).

Table 5 Criteria definitions.

CRITERIA	
Topicality	The data is consistent with the user's research, such as data related in terms of content, time, and region.
Availability	The user can obtain the data without any external factors (i.e. no access permission, no download links, high prices, etc.).
Quality	The quality of data, for example, whether the data is accurate, correct, and valid.
Standardization	The data classification system and collection process are consistent with national requirements.
Authority	Users can trust this data, mainly referring to a person or an organization that publishes influential data.
Comprehensiveness	The data has full coverage, or the data is complete without missing any elements.
Convenience	It is convenient to retrieve, obtain, and use the data.
Usability	The data can be used without cognitive limitations or formatting problems.
Currency	The data is valuable to the research and valid for only a certain period of time, such as the publication date is recent, or not outdated.

Topicality was the most frequent criterion, and was 44.80% of all criteria nodes. *Availability* and *quality* made up 12% and 11.20% of all criteria nodes, respectively. Their usage adds up to about 70% of all usage (*Table 6*). The use of *topicality*, *availability*, and *quality* accounts for the majority of usage, and the remaining six criteria also play a significant role in the final relevance judgment. However, frequency was not consistent with importance. For example, in the interview, one user mentioned that "I am a student of geography; I want to see if the format is correct." When the *currency*, *usability*, *standardization*, and other criteria did not meet the users' needs, users could decide that the data were irrelevant, even if *topicality* and *quality* were satisfied well.

---

4   CDs: Coding Sequence, a term of structural genomics.

RELEVANCE CRITERIA	NODES	PERCENTAGE
Topicality	168	44.80%
Availability	45	12.00%
Quality	42	11.20%
Completeness	32	8.53%
Authority	29	7.73%
Currency	20	5.33%
Convenience	16	4.27%
Usability	15	4.00%
Standardization	8	2.13%
Total	375	100.00%

Table 6 Percentages of relevance criteria nodes.

The relevance criteria vary with respect to type of information carrier. The relevance criteria of documents were the longest and most comprehensively studied. Scientific data and documents are both generated in scientific research activities, serve scientific research, and are constantly presented. However, there are also some differences between them. Documents contain mature knowledge that has been extracted from scientific data by researchers. Therefore, a comparison between documents and scientific data more clearly shows the changes caused by the essential nature of different information carriers.

Barry, Schamber, Wang Peiling, Saracevic, Taylor and others have discussed the concepts of document relevance criteria and their usage. However, because their discussions took place in different contexts, the relevance criteria were different, and there is not yet any consensus about what a set of criteria should contain. The document selection model studied by Wang Peiling is the most similar to the concept model studied in this paper. Therefore, the document relevance criteria derived in Wang's research were compared with the scientific data relevance criteria derived in this study.

As shown in *Table 7*, Wang Peiling proposed 11 document relevance criteria, and this article proposes nine scientific data relevance criteria. A comparison of the two studies shows that there are three unique criteria for scientific data: *comprehensiveness*, *standardization*, and *convenience*. *Accessibility* is strengthened in importance and *novelty* disappears. Compared with the criteria for documents, the criteria for data had higher *purpose*, *pertinence*, and *practicality*, but also poorer *substitutability*. The purpose for users to retrieve scientific data was mainly to support their own research analysis or conclusions, which requires high accuracy and consistency. As some users mentioned, "I study the grain output in the past ten years, and there is no 2009 output in this data set. I cannot perform the next analysis without this data. I need to find it through other channels." Therefore, the comprehensiveness of the data affected the users' relevance judgment. Moreover, each industry has its own data requirements, and each data platform, unit, or laboratory has its own requirements for data. The irregularity of

DOCUMENT RELEVANCE CRITERIA		SCIENTIFIC DATA RELEVANCE CRITERIA	
Topicality	65.30%	Topicality	44.80%
Quanlity	9.40%	Availability	12.00%
Scope	9.40%	Quality	11.20%
Novelty	5.30%	Completeness	8.53%
Currency	2.80%	Authority	7.73%
Discipline	2.70%	Currency	5.33%
Source	1.70%	Convenience	4.27%
Authority	1.10%	Usability	4.00%
Special need	1.00%	Standardization	2.13%
Time cost	0.80%		
Availability	0.20%		

Table 7 Comparison of document and scientific data relevance criteria.

data severely restricts data sharing and usage. Hence, the standardization of data also affects the user's relevance judgment. These problems do not exist in document relevance judgments, because the information transmitted by documents is broader than the information in data, and they serve different purposes in scientific research.

Documents contain a large amount of information. Even if the original text cannot be obtained, the main or key information can be obtained from the abstract. However, data are different. The ultimate goal of users to retrieve data is to obtain and use data. If the data cannot be used, their value will be greatly discounted. Therefore, the weight of availability increases in data relevance judgment.

In conclusion, the difference in the behaviors of users lies in the essential difference between scientific data and documents. Documents are laden with knowledge, whereas scientific data are laden with facts. Knowledge is something that human beings can directly process cognitively, but facts cannot be processed this way. Humans need to process the data using instruments such as Power BI Desktop or CDAT. Therefore, when retrieving data, users pay more attention to the availability of data and whether they can be further analyzed and processed.

4.2 RELATIONSHIP BETWEEN SCIENTIFIC DATA METADATA AND RELEVANCE CRITERIA

As a whole, the relationship between metadata and relevance criteria can be summarized as one stimulation to multiple responses and multiple stimulations to one response. There is an intermediate element – clues – between metadata and relevance criteria. The users must first experience the stimulation presented by the metadata, and this stimulation consists of clues. Then, users process the stimulation to form the relevance criteria. The concept model was verified by the experiment.

The relationships and weights among metadata, clues, and criteria are visualized in *Figure 6*. The same metadata produced different stimulations through users' eyes. For example, when users see "name," one might respond with *matches my study*, for example, "this index is the main content of my research" (Participant 22). Someone else might respond with *fits my needs*, for example, "according to my research, I'm looking for wheat, but there's very little about wheat" (Participant 7). When users see the "data time," one might respond that *data are old*, for example, "only the 2013 digital version is available, which is too old" (Participant 22). Another person might respond with *difficult to obtain*, for example, "the latest data are hard to get" (Participant 24). It highly depends on the user's cognitive workspace, which is closely related to work experience, research direction, and the user's understanding of his/her problems.

As users responded to different stimulations, the relevance criteria invoked in the brain also changed. When users could only download data from 2013, they used *currency* to judge the relevance of the data. When users thought that data were difficult to obtain, they used *availability* to judge the relevance of the data.

Different metadata will stimulate users to employ the same relevance criteria for relevance judgment. When users focused on metadata such as "title," "abstract," "keyword," "name," "data content," "description," or "species," *topicality* was stimulated. For example, for "species" to *topicality*: "I'm looking for a related species, but I don't see it here." (Participant 3); for "keyword" to *topicality*: "I directly searched for bagasse, but I only saw an item about sweet potato. I thought this study would be similar to mine, so I clicked on it." (Participant 21); and for "title" to *topicality*: "I read the title and it is not related to my search." (Participant 25). When users focused on items such as "auditable," "journal," "author," and "institute," *authority* was stimulated. For example, for "auditable" to *authority*, "Auditable data[5] is authority." (Participant 7); and for "journal" to *authority*: "The journals have great reputation. We might use data from very famous journals."

Metadata related to *topicality* are the most complex and include subject-irrelevant metadata[6] and subject-relevant metadata[7] in different fields. Metadata related to other relevance criteria

---

5   Auditable data: data audited by a third party.

6   Subject-irrelevant metadata: metadata are irrelevant to the subject, such as "title", "key word", "abstract", "publish time" and so on.

7   Subject-relevant metadata: metadata are relevant to the subject, such as "gene location" and "gene length" is relevant to genetics; "data format" and "data area" is relevant to geography.
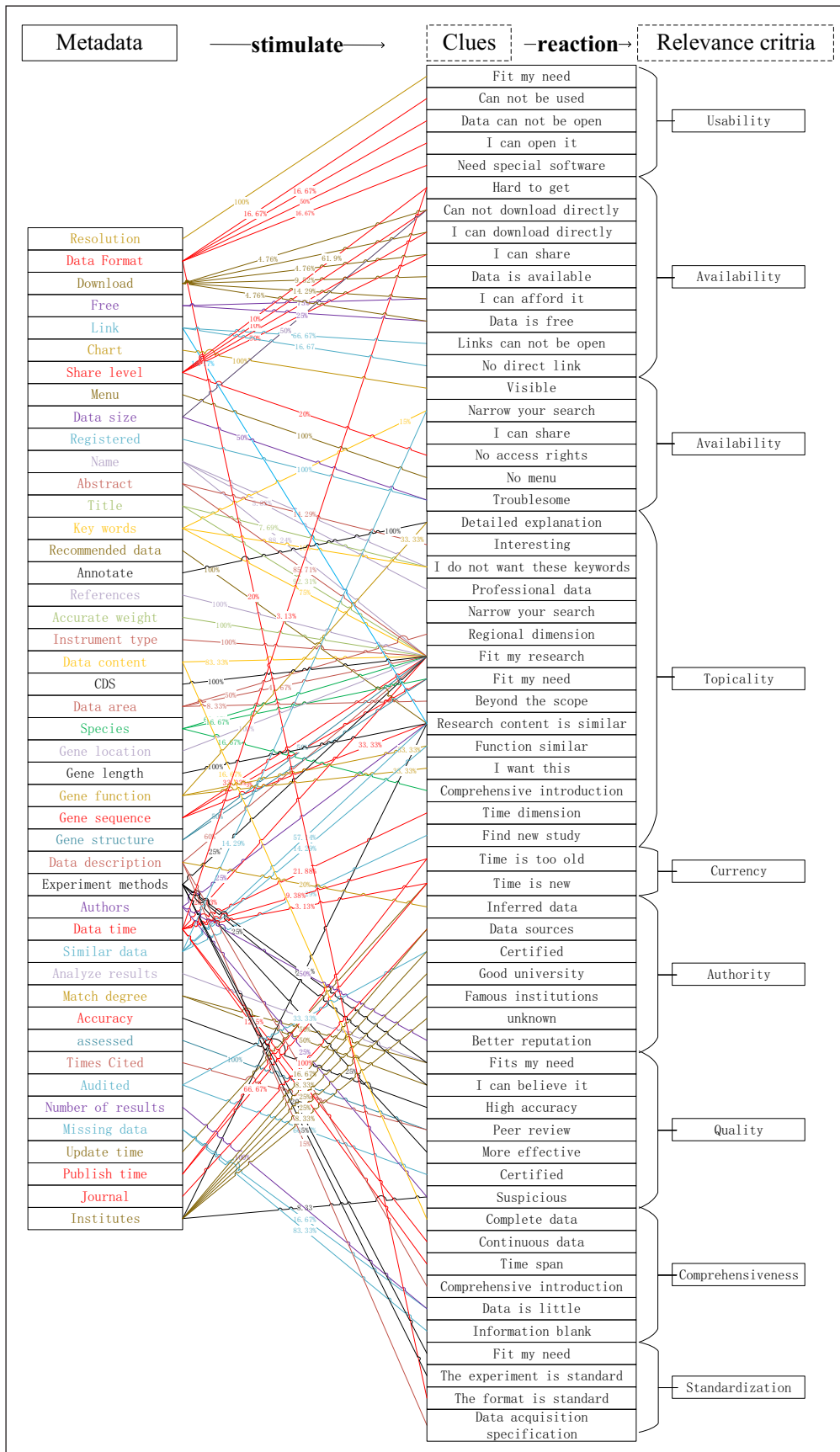
Figure 6 Relationships among metadata, clues, and relevance criteria.

(such as *quality*, *authority*, and *availability*) only include subject-irrelevant metadata. Twenty-five types of metadata stimulated users to use *topicality* to make the relevance judgment, and these types can be divided into three categories. The first category is subject-irrelevant metadata, and this category includes metadata like "name," "title," "keyword," "abstract," "annotation," "author," "recommended data," "institute," "links," "description," and "similar data." The second category is metadata related to meteorology, agricultural economics, and remote sensing and includes "data area," "data content," and "data time." The third category is metadata related

to biology, genetics, and engineering and includes "gene function," "gene length," "gene location," "structure," and "gene sequence." However, nine metadata types stimulated users to use *quality* when making relevance judgments: "analyze results," "auditable," "author," "citation frequency," "correct rate," "experimental method," "institute," "matching degree," and "reviewed." Five metadata stimulated users to use *authority* when making relevance judgments: "author," "journal," "institute," "auditable," and "description." Metadata associated with *quality* and *authority* were irrelevant to subject. Hence, in the relevance judgments, the differences in the metadata of different groups were mainly reflected in *topicality*.

Using the dwell times of metadata (*Figure 4*) and relationships among metadata and criteria (*Figure 6*), the times spent on the nine relevance criteria were calculated (*Table 8*). *Topicality* took the longest, accounting for 65.7% of the total time, followed by *availability* and *quality*, which accounted for 6% and 5.89% of the total time, respectively. The criteria dwell times represent users' effort in scientific data retrieval.

CRITERIA	NODE NUMBER	PERCENTAGE	CRITERIA	DWELL TIME	PERCENTAGE
topicality	168	44.80%	topicality	1932354	65.70%
availability	45	12.00%	availability	176395	6.00%
quality	42	11.20%	quality	173161	5.89%
completeness	32	8.53%	convenience	170221	5.79%
authority	29	7.73%	completeness	169633	5.77%
currency	20	5.33%	authority	139940	4.76%
convenience	16	4.27%	currency	122595	4.17%
usability	15	4.00%	standardization	40100	1.36%
standardization	8	2.13%	usability	16933	0.58%
total	375	100.00%	total	2941332	100.00%

Table 8 Percentage of criteria frequencies and dwell times

$$\text{criteria dwell time} = \sum \text{metadata dwell time}_i * \text{weight}_i$$

Regression analysis of the two groups data show a linear relationship (*Figure 7*), $R^2 = 0.967$, $P = 0.000$. The correlation is significant at the level of $a = 0.05$. The use frequency of relevance criteria is positively correlated with the effort expended in relevance judgments. For example, *topicality* was recorded 168 times in interview data, and users also spent the most energy on topical-based relevance judgment.
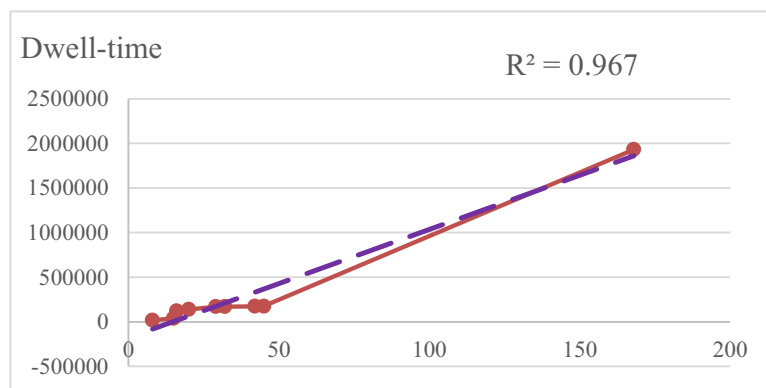

Figure 7 Regression analysis of two groups data.

5. DISCUSSIONS

The study found nine scientific data relevance criteria, namely, topicality, availability, quality, standardization, authority, comprehensiveness, convenience, usability and currency. Most previous studies had focused on documents and web pages, with a few on images and video. The research situation involved work, life, entertainment, etc. Research subjects included professors, students, doctors, journalists, etc. The results showed there was an overlap between the relevance criteria mentioned in this study and the previous studies with new criteria

emerged from the data analysis (Sarah Albassam, 2018). The appearances of new criteria were directly related to the essence of information carriers. For example, images selection needed to consider resolution and size. Documents selection needed to consider languages. Web pages selection needed to consider link security and information reliability. For scientific data, comprehensiveness and standardization were two unique criteria. Comprehensiveness focused on the continuity and integrity of data in time and regional sequences. Standardization focused on data classification system and statistical methods. Because scientific data had a strong professional, domain, and practicability, users had identified the need of data before the retrieval, without inspiration. This was why many scholars mentioned novelty in their studies, but it did not appear in this paper. At the same time, the study cannot ignore the influence of external situation and user's cognitive on the change of relevance criteria. Audrey Laplante suggested that although research had found that some of the relevance criteria (quality and authority) found in documents and web pages still applied to music environments, there would be some unique music relevance criteria. As Saracevic pointed out, relevance research cannot be separated from the situation, and should consider the dynamic interaction between the internal and external factors of the situation.

More than one scholar tried to generalize a set of relevant criteria across different dimensions, but without success (Schamber, 1996; Bales and Wang, 2005). There were two reasons, one reason behind this difficulty was that different studies had various labels and definitions for similar relevance criteria and the grouping/categorization of the findings also varied among different studies. Another challenge in comparing relevance criteria studies was that various methodologies had been applied in relevance criteria literature (Maglaughlin and Sonnenwald, 2002; Savolainen and Kari, 2006). Relevance was a multidimensional, dynamic process, and information carrier was only one of the dimensions. Here, we could be sure that the change of information carrier will certainly cause the change of relevance criteria.

Different users may have different responses when receiving the same stimulation, and may have the same response when receiving different stimulation. In the Document selection model put forward by Wang Peiling, a document (distal object) was represented by a set of document information elements (metadata) as clues. Document information elements were processed to judge a document on several criteria. This study enriched and expanded document selection model, which not only clearly defined the concepts of information elements and clues, but also explored the corresponding relations between them. Panos Balatsoukas and Gao Fei also used eye tracker to study relevance criteria, but they all focused on the fixation of metadata and the usage of relevance. They ignored the relationship between metadata, clues and criteria in the information processing process. According Hochberg's view of perception, participants were able to perceive completely different shapes of the same physical stimulation. His perception fundamentally determined his answers to questions about shape, motion, size, depth, etc. In the relevance judgment, clues were the cognitive reflection after the user perceived the external information, which related to the user's cognitive workspace. For example, when seeing a data from 2008, somebody said that was consistent with study time and somebody might think the time was too old. The reason was user's different needs and cognitive abilities. Human perception involved inferences when it came to recognizing something, and this recognition pattern explained why what we know determined what we see. Through analysis, it can be seen that users' cognitive workspace played a crucial role in the process from receiving to perceiving and processing information. The workspace was a relatively stable cognitive state formed under the long-term working and retrieval environment.

In relevance judgment, it was not only necessary to study how relevance criteria affect the judgment process, but more importantly, how metadata affected the relevance judgment process through criteria. The interpretation of informational clues provides a novel approach to deepen empirical research on how people use information content (Savolainen, 2010). Such research efforts would provide opportunities to take one step closer to the goal proposed by Gerstberger and Allen (1968), that is, to explore "the actual process of using the information". Topicality was the fundamental criterion, and users spend the most energy on it and use it most frequently (Abe Crystal; Rahayu A Hamid; Sedghi, 2013; Sarah Albassam, 2018). Student used titles, summaries, and connectedness to topic as prime metadata when making web pages judgment (Watson, 2013). Users used titles, key words, and abstract to topic as prime metadata of relevance when making documents judgment. And users used name, data area, data content, species and so

on to topic as prime metadata of relevance when making scientific data judgment. So even if users used the same criteria to judge relevance under different information carriers, there were differences in the metadata they paid attention to. Only by understanding the fixation differences of these metadata can we better improve the scientific data sharing system.

Combined with eye movement data and interview data, it was found that the frequency of criteria usage was positively correlated with the amount of attention spent on it. The study used eye tracker collecting eye movement data, such as the number and length of fixations, which could reveal a more accurate picture of the cognitive effort spent by users during the relevance judgment process (Balatsoukas, 2012). Users spent the most attention on topicality, more than 60%. The other relevance criteria took less than 10% attention. This was consistent with the anchoring adjustment strategy in the judgment decision. At the beginning, attention will be focused on the topicality, which was an anchor. Subsequently, other criteria were insufficient adjustments to this anchor, like accessibility, quality, authority, etc. (Reid Hastie, 2004).

6. CONCLUSIONS

The main purpose of this paper was to explore metadata, relevance criteria, and the relationship between them. An eye tracker recorded the attention paid to metadata by users during the retrieval. Relevance criteria usage were obtained from interviews. The combination of quantitative data obtained by the eye tracker (fixation duration) and qualitative data obtained via interviews (relevance criteria, clues and other nodes) makes these research results convincing.

Users pay attention to 45 metadata when retrieving scientific data. The 45 metadata can be divided into subject-irrelevant metadata and subject-relevant metadata. Subject-irrelevant metadata includes "name", "key words", "abstract", and so on. And there are no significant differences in subject-irrelevant metadata between investigational users and experimental users. Subject-relevant metadata includes "gene location", "gene length", "resolution", and so on. Investigational users paid more attention to "data time and "data content" because data from the search were their research objects and the research always had certain requirements with respect to region and time. Experimental users paid more attention to "gene location," "gene sequence," and "gene function," which are subject-relevant metadata.

Nine relevance criteria for scientific data were found in the study, respectively, topicality, availability, quality, completeness, authority, currency, convenience, usability, and standardization. Because of the essential difference between scientific data and documents, users use different criteria. Documents are laden with knowledge, whereas scientific data are laden with facts. Knowledge is something that human beings can directly process cognitively, but facts cannot be processed this way. Humans need to process the data using instruments. Therefore, when retrieving data, users pay more attention to the availability of data and whether they can be further analyzed and processed.

When retrieving scientific data, different users may have different responses when receiving the same stimulus or the same response when receiving different stimuli. The metadata stimulating *topicality* are the most complex and include subject-irrelevant metadata and distinctive subject-relevant metadata. The metadata stimulating other criteria (such as *quality* and *authority*) have no obvious subject-relevant characteristics.

This paper analyzed the process of relevance judgment for scientific data from the perspective of information cognitive processing. The concepts of metadata, clues, and relevance criteria were clearly defined through a situation experiment combining eye tracking experiments with interviews.

This paper provided a theoretical and empirical basis for the next stage in the study of the normal form equation of scientific data relevance judgment based on the lens model. The practical significance of this study is that it enables the more targeted improvement of a scientific data sharing system, such as changing the presentation of pages and providing personalized services for users with different needs. This research not only determined the metadata that users mainly care about and the relevance criteria of scientific data that are frequently used, it also found the corresponding relationships among them.

APPENDIX 1

criteria	no direct link	unipront
authority	no key word	**metadata**
availability	no limited	abstract
completeness	no menu	analyze results
convenience	no sense	annotations
currency	Out of place	auditable
standardability	pre-processing	Author
quality	professional	CDS
topicality	publication is good	Chart
usability	reduce the scope	citation frequency
clues	Regional dimension	cloudage
be discovered	area dimension	correct rate
better known	Related to protein information	Data area
cannont be used	Research content is similar	data content
cannot download directly	skepticism	data size
cannot open	someone recommend me data	data time
certificated	special software	description
complete data	technological improvement	download
comprehensive introduction	the discription of title	exact mass
continuous data	time dimension	experimental method
correct rate is high or low	time is new	Format
data be identified	time is too old	free
data is little	time span	gene function
data sources	Time to close	gene length
data was affected	track the study	gene location
depth	very troublesome	gene sequence
detailed explanation	visual	institute
difficult to obtain	whether I can afford	instrument type
every body use it	whether I can believe	journal
fit my needs	whether it is free	key word
functional similarity	Whether it's available or not	links
good university	**data type**	Matching Degree
guessed data	chromatogram map	menu
haven't checked	composition of feed	missing data
human review	gene	name
I can download	gene physical map	number of results
I can open it	geography	publish time
I can share it	meteorological data	recommended data
I want this	statistics	references
information blank	**databases**	register
interest	ABC	Resolution
larger organization	China meteorological science data sharing service network.	reviewed

little-known	CNKI	Share level
long time to check	FAO	similar data
match my study	GS Cloud	species
whether it matches	National Statistics Bureau	structure
more effective	NCBI	title
no cloud	soybean databases	update status

ADDITIONAL FILE

The additional file for this article can be found as follows:

- **Eye movement data sets.** The data set include RECORDING_SESSION_LABEL,EYE_USED,IA_AREA,IA_AVERAGE_FIX_PUPIL_SIZE,IA_DWELL_TIME,IA_DWELL_TIME_%,IA_FIXATION_%,IA_FIXATION_COUNT,IA_LABEL. All data come from 36 subjects, which were got through eye trackers. DOI: *https://doi.org/10.5334/dsj-2021-005.s1*

ACKNOWLEDGEMENTS

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Guilan Zhang *orcid.org/0000-0002-9153-3579*

Institute of Scientific and Technical Information of China, Beijing, China

Jian Wang *orcid.org/0000-0003-4958-7669*

Agricultural Information Institute, Chinese Academy of Agricultural Sciences, Beijing, China

Key Laboratory of Agricultural Big Data, Ministry of Agriculture and Rural Affairs, Beijing, China

Jianping Liu *orcid.org/0000-0002-1817-5373*

Agricultural Information Institute, Chinese Academy of Agricultural Sciences, Beijing, China

Key Laboratory of Agricultural Big Data, Ministry of Agriculture and Rural Affairs, Beijing, China

Yao Pan *orcid.org/0000-0003-2317-7442*

Institute of Scientific and Technical Information of China, Beijing, China

REFERENCES

Albassam, SAA and **Ruthven, I.** 2018. Users' relevance criteria for video in leisure contexts. *Journal of Documentation*, 74(1): 62–79. DOI: *https://doi.org/10.1108/JD-06-2017-0081*

Anderson, TD. 2005. Relevance as Process: Judgements in the Context of Scholarly Research. *Information Research: An International Electronic Journal*, 10(2): n2.

Audrey Laplante. 2010. Users' Relevance Criteria In Music Retrieval In Everyday Life: An Exploratory Study. *11th International Society for Music Information Retrieval Conference*, Utrecht, Netherlands, August 9–13, pp. 601–606.

Awang, H. 2016. Effects of relevance criteria and subjective factors on web image searching behaviour. *Journal of Information Science*, 43(6): 786–800. DOI: *https://doi.org/10.1177/0165551516666968*

Balatsoukas, P, O'Brien, A and **Morris, A.** 2010. Design factors affecting relevance judgment behaviour in the context of metadata surrogates. *Journal of Information Science*, 36(6): 780–797. DOI: *https://doi.org/10.1177/0165551510386174*

Balatsoukas, P and **Ruthven, I.** 2010, August. What eyes can tell about the use of relevance criteria during predictive relevance judgment? In *Proceedings of the third symposium on Information interaction in context*. ACM. pp. 389–394. DOI: *https://doi.org/10.1145/1840784.1840844*

Balatsoukas, P and **Ruthven, I.** 2010, October. The use of relevance criteria during predictive judgment: an eye tracking approach. In *Proceedings of the 73rd ASIS&T Annual Meeting on Navigating Streams in an Information Ecosystem-Volume 47*. American Society for Information Science. p. 73. DOI: *https://doi.org/10.1002/meet.14504701145*

Balatsoukas, P and **Ruthven, I.** 2012. An eyetracking approach to the analysis of relevance judgments on the Web: The case of Google search engine. *Journal of the American Society for Information Science and Technology*, 63(9): 1728–1746. DOI: *https://doi.org/10.1002/asi.22707*

Bales, S and **Wang, P.** 2005. Consolidating user relevance criteria: A metaethnography of empirical studies. *Proceedings of the American Society for Information Science and Technology*, 42(1). DOI: *https://doi.org/10.1002/meet.14504201277*

Barry, C. 1994. User-defined relevance criteria: an exploratory study. *Journal of the Association for Information Science & Technology*, 45(3): 149–159. DOI: *https://doi.org/10.1002/(SICI)1097-4571(199404)45:3<149::AID-ASI5>3.0.CO;2-J*

Barry, CL and **Schamber, L.** 1998. Users' criteria for relevance evaluation: a cross-situational comparison. *Information processing & management*, 34(2–3): 219–236. DOI: *https://doi.org/10.1016/S0306-4573(97)00078-2*

Bucher, HJ and **Schumacher, P.** 2006. The relevance of attention for selecting news content. An eye-tracking study on attention patterns in the reception of print and online media. *Communications*, 31(3): 347–368. DOI: *https://doi.org/10.1515/COMMUN.2006.022*

Chang, YS and **Gwizdka, J.** 2018. Relevance criteria dynamics: A study of online news selection on SERPs. *Proceedings of the Association for Information Science and Technology*, 55(1): 768–769. DOI: *https://doi.org/10.1002/pra2.2018.14505501108*

Choi, Y and **Rasmussen, EM.** 2002. Users' relevance criteria in image retrieval in american history. *Information Processing & Management*, 38(5): 695–726. DOI: *https://doi.org/10.1016/S0306-4573(01)00059-0*

Cool, C. 1993. Characteristics of text affecting relevance judgments. *Automotive News*, 17(4): 77–85.

Crystal, A and **Greenberg, J.** 2006. Relevance criteria identified by health information users during web searches. *Journal of the American Society for Information Science and, Technology*, 57(10): 1368–1382. DOI: *https://doi.org/10.1002/asi.20436*

Drori, O. 2003. How to Display Search Results in Digital Libraries-User Study. In *NDDL/VVEIS*. pp. 13–28.

Duchowski, AT. 2007. Eye tracking methodology. *Theory and practice*, 328.

Dziadosz, S and **Chandrasekar, R.** 2002. Do thumbnail previews help users make better relevance decisions about web search results? In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM. pp. 365–366. DOI: *https://doi.org/10.1145/564376.564446*

Faniel, IM and **Jacobsen, TE.** 2010. Reusing scientific data: How earthquake engineering researchers assess the reusability of colleagues' data. *Computer Supported Cooperative Work (CSCW)*, 19(3–4): 355–375. DOI: *https://doi.org/10.1007/s10606-010-9117-8*

Gao, D and **Xiaoyun, X.** 2003. *Cognitive psychology*. Shanghai: East China Normal University Press.

Gao, F, Lei, S and **Jian, W.** 2017. An Exploratory Research on the Relationship Between Agriculture Scientific Data User Relevance Clues and Criteria. *Library and Information Service*, 15: 72–80. DOI: *https://doi.org/10.13266/j.issn.0252-3116.2017.15.008*

Gerstberger, PG and **Allen, TJ.** 1968. Criteria used by research and development engineers in the selection of an information source. *Journal of applied psychology*, 52(4): 272. DOI: *https://doi.org/10.1037/h0026041*

Gwizdka, J. 2014, August. Characterizing relevance with eye-tracking measures. In *Proceedings of the 5th Information Interaction in Context Symposium*. ACM. pp. 58–67. DOI: *https://doi.org/10.1145/2637002.2637011*

Gwizdka, J and **Zhang, Y.** 2015, August. Differences in eye-tracking measures between visits and revisits to relevant and irrelevant web pages. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM. pp. 811–814. DOI: *https://doi.org/10.1145/2766462.2767795*

Hamid, R and **Thom, J.** 2010. Criteria that have an effect on users while making image relevance judgments. In *Fifteenth Australasian Document Computing Symposium*, School of Computer Science and IT, RMIT University, 8.

Hung, TY, Zoeller, C and **Lyon, S.** 2005, December. Relevance judgments for image retrieval in the field of journalism: A pilot study. In *International Conference on Asian Digital Libraries*. Berlin, Heidelberg: Springer. pp. 72–80. DOI: *https://doi.org/10.1007/11599517_9*

Ingwersen, P and **Järvelin, K.** 2011. *The turn: Integration of information seeking and retrieval in context*. Berlin, Germany: Springer Publishing Company, Incorporated.

Inskip, C, Macfarlane, A and **Rafferty, P.** 2010. *Creative professional users' musical relevance criteria.* 36(4): 517–529. Sage Publications, Inc. DOI: *https://doi.org/10.1177/0165551510374006*

Jacob, RJ and **Karn, KS.** 2003. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In *The mind's eye*. North-Holland. pp. 573–605. DOI: *https://doi.org/10.1007/978-3-642-03658-3_119*

Liu, J, Wang, J and **Zhou, G.** 2019. Understanding relevance judgment in the view of perceived value. *Library & Information Science Research*, 41(4): 100982. DOI: *https://doi.org/10.1016/j.lisr.2019.100982*

Lorigo, L, Haridasan, M, Brynjarsdóttir, H, Xia, L, Joachims, T, Gay, G and **Pan, B.** 2008. Eye tracking and online search: Lessons learned and challenges ahead. *Journal of the American Society for Information Science and Technology*, 59(7): 1041–1052. DOI: *https://doi.org/10.1002/asi.20794*

Maglaughlin, KL and **Sonnenwald, DH.** 2002. User perspectives on relevance criteria: a comparison among relevant, partially relevant, and not-relevant judgments. *Journal of the American Society for Information Science and, Technology*, 53(5): 327–342. DOI: *https://doi.org/10.1002/asi.10049*

Marchionini, G, Song, Y and **Farrell, R.** 2009. Multimedia surrogates for video gisting: Toward combining spoken words and imagery. *Information Processing & Management*, 45(6): 615–630. DOI: *https://doi.org/10.1016/j.ipm.2009.05.007*

Markkula, M and **Sormunen, E.** 2000. End-user searching challenges indexing practices in the digital newspaper photo archive. *Information Retrieval*, 1(4): 259–285. DOI: *https://doi.org/10.1023/A:1009995816485*

Paek, T, Dumais, S and **Logan, R.** 2004, April. WaveLens: A new view onto internet search results. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM. pp. 727–734. DOI: *https://doi.org/10.1145/985692.985784*

Papaeconomou, C, Zijlema, AF and **Ingwersen, P.** 2008, October. Searchers' relevance judgments and criteria in evaluating web pages in a learning style perspective. In *Proceedings of the second international symposium on Information interaction in context*. ACM. pp. 123–132. DOI: *https://doi.org/10.1145/1414694.1414722*

Park, TK. 1993. The nature of relevance in information retrieval: an empirical study. *The Library Quarterly*, 63(3): 318–351. DOI: *https://doi.org/10.1086/602592*

Park, YKS. 2014. User-based Relevance and Irrelevance Criteria during the Task Pursuing of Middle School Students. *Korean Society of Documentation and Information*, 48(3). DOI: *https://doi.org/10.4275/KSLIS.2014.48.3.055*

Pian, W, Khoo, CS and **Chang, YK.** 2016. The criteria people use in relevance decisions on health information: An analysis of user eye movements when browsing a health Discussion Forum. *Journal of medical Internet research*, 18(6). DOI: *https://doi.org/10.2196/jmir.5513*

Rayner, K. 2009. Eye movements and attention in reading, scene perception, and visual search. *The quarterly journal of experimental psychology*, 62(8): 1457–1506. DOI: *https://doi.org/10.1080/17470210902816461*

Rele, RS and **Duchowski, AT.** 2005, September. Using eye tracking to evaluate alternative search results interfaces. *In Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 49(15): 1459–1463. Los Angeles, CA: Sage CA, SAGE Publications. DOI: *https://doi.org/10.1177/154193120504901508*

Rieh, SY. 2002. Judgment of information quality and cognitive authority in the Web. *Journal of the American society for information science and technology*, 53(2): 145–161. DOI: *https://doi.org/10.1002/asi.10017*

Sabbata, SD and **Reichenbacher, T.** 2012. Criteria of geographic relevance: an experimental study. *International Journal of Geographical Information Science*, 26(8): 1495–520. DOI: *https://doi.org/10.1080/13658816.2011.639303*

Saracevic, T. 2015, May. Why is relevance still the basic notion in information science. In Re: inventing Information Science in the Networked Society. *Proceedings of the 14th International Symposium on Information Science*. pp. 26–35.

Saracevic, T. 2016. The notion of relevance in information science: Everybody knows what relevance is. But, what is it really? *Synthesis Lectures on Information Concepts Retrieval, and Services*, 8(3): i–109. DOI: *https://doi.org/10.2200/S00723ED1V01Y201607ICR050*

Savolainen, R. 2010. Source preference criteria in the context of everyday projects: Relevance judgments made by prospective home buyers. *Journal of Documentation*, 66(1), 70–92. DOI: *https://doi.org/10.1108/00220411011016371*

Savolainen, R and **Kari, J.** 2006. User-defined relevance criteria in web searching. *Journal of Documentation*, 62(6): 685–707. DOI: *https://doi.org/10.1108/00220410610714921*

Schamber, LJ. 1996. User Criteria in Relevance Evaluation: Toward Development of a Measurement Scale. *Asis Meeting*.

Schamber, L and **Bateman, J.** 1996, October. User criteria in relevance evaluation: Toward development of a measurement scale. In *Proceedings of the Annual Meeting-American Society for Information Science,* 33: 218–225.

Sedghi, S, Sanderson, M and **Clough, P.** 2008. A study on the relevance criteria for medical images. *Pattern Recognition Letters*, 29(15): 2046–2057. DOI: *https://doi.org/10.1016/j.patrec.2008.07.003*

Sedghi, S, Sanderson, M and **Clough, P.** 2012. How do health care professionals select medical images they need? *Aslib Proceedings*, 64(4): 437–456. DOI: *https://doi.org/10.1108/00012531211244815*

Taylor, A. 2009. Relevance criterion choices in relation to search progress. *Dissertations & Theses – Gradworks*, 26: 203–208. DOI: *https://doi.org/doi:10.7282/T3W959FQ*

Tombros, A. 2003. Searchers' criteria For assessing web pages. In *International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 385–386. DOI: *https://doi.org/10.1145/860435.860513*

Tombros, A, Ruthven, I and **Jose, JM.** 2014. How users assess web pages for information seeking. *Journal of the Association for Information Science & Technology*, 56(4): 327–344. DOI: *https://doi.org/10.1002/asi.20106*

Tsai-Youn, H. 2018. A Study on the Relevance Criteria for Journalistic Images. *Journal of Library & Information Science*, 44(2): 4.

Wang, X, Hong, Z, Xu, Y, Zhang, C and **Ling, H.** 2014. Relevance judgments of mobile commercial information. *Journal of the Association for Information Science and Technology*, 65(7): 1335–1348. DOI: *https://doi.org/10.1002/asi.23060*

Wang, P and **Soergel, D.** 1998. A cognitive model of document use during a research project. Study I. Document selection. *Journal of the American Society for Information Science*, 49(2): 115–133. DOI: *https://doi.org/10.1002/(SICI)1097-4571(199802)49:2<115::AID-ASI3>3.0.CO;2-T*

Watson, CL, Littledyke, M and **Parkes, M.** 2013. An exploratory study of students' judgements of the relevance and reliability of information. *https://hdl.handle.net/1959.11/13471.*

Wei, C, Jian, W and **Guilan, Z.** 2018. Research on Conceptual Model for Scientific Data User's Perceived Value Based on Grounded Theory. *Journal of Intelligence*, 5: 182–188. DOI: *https://doi.org/10.3969/j.issn.1002-1965.2018.05.028*

Westbrook, L. 2001. Faculty relevance criteria: internalized user needs. *Library trends*, 50(2): 197–206.

Xie, I and **Benoit, E, III.** 2013. Search result list evaluation versus document evaluation: similarities and differences. *Journal of Documentation*, 69(1): pp. 49–80. DOI: *https://doi.org/10.1108/00220411311295324*

Xu, Y and **Chen, Z.** 2006. Relevance judgment: What do information users consider beyond topicality? *Journal of the American Society for Information Science and Technology*, 57(7): 961–973. DOI: *https://doi.org/10.1002/asi.20361*

Yang, M and **Marchionini, G.** 2010. Exploring users' video relevance criteria—a pilot study. *Proceedings of the American Society for Information Science & Technology*, 41(1): 229–238. DOI: *https://doi.org/10.1002/meet.1450410127*

Zhang, G, Jian, W and **Jianping, L.** 2018. The Relationship Between Relevance Criteria and Target Information Type. *Journal of Intelligence*, 37(6): 171–179. DOI: *https://doi.org/10.3969/j.issn.1002-1965.2018.06.027*

Zhang, G, Wang, J, Zhou, G, Liu, J and **Wei, C.** 2018, October. Scientific Data Relevance Criteria Classification and Usage. In *Proceedings of the 2nd International Conference on Computer Science and Application Engineering*. ACM. p. 30. DOI: *https://doi.org/10.1145/3207677.3278010*

Zhonghua, D and **Yating, H.** 2017. Research on the development of China's scientific data sharing platform under the "Internet +" environment. *Information Studies: Theory & Application*, 40(2): 128–132.