ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

Electronic records management: statistical assessment of metadata quality and its impact on search efficiency

Dazhi Zhu[1,*], *Qing Yan*[1], *Junlin Ji*[1], *Shi Chen*[1], *Shuhong Lin*[1] *and Yuan Wang*[1]

[1] Information and Communication Branch of Hainan Power Grid, Haikou, Hainan, 570203, China

SUMMARY: *In this paper, a Drosophila optimization algorithm based on linear decreasing step size and logical chaos mapping is firstly proposed as a metadata search method for electronic archives, which optimizes the stability of the FOA algorithm by introducing techniques such as logical chaos mapping theory. The quality assessment system of electronic archive metadata is constructed from three aspects, namely, form, content and utility, and an electronic archive evaluation model based on AHP-cloud model is established to realize the intelligent management of electronic archives. The results show that the method proposed in this paper not only fully takes into account the ambiguity and randomness of the evaluation results, but also makes the evaluation results more reliable, and its order in the similarity of the comprehensive cloud of evaluation results are: excellent>good>middling>poor>poor, and the distribution range of the comprehensive cloud is almost in the range of the "excellent" evaluation grade, which is extremely good and practical. Practicality. The completeness of data collection of this paper's e-file evaluation model is improved by nearly 40% on the 50th day compared with the comparison model, which effectively improves the operational efficiency of the system. In addition, the e-file evaluation model in this paper has the highest check accuracy and completeness of information of 98.38% and 94.32% respectively, and its comprehensive evaluation results are more accurate.*

KEYWORDS: *DSLC-FOA; cloud model; AHP algorithm; electronic archives; data evaluation system; search efficiency*

1 Introduction

In the context of the era of the gradual popularization and development of the Internet, digitalization has become an inevitable development direction in the field of information media, and the release and transmission of information is increasingly carried out in the form of data with the help of the network [1]. With the popularization and development of computer network communication technology, data transmission has become a relatively simple and fast process, and information access has become convenient and efficient in the context of digitization [2]. While digitalization brings convenience, it also brings some negative effects, as certain individuals or groups take advantage of the convenience of big data on the Internet to arbitrarily steal or copy the contents of unauthorized dissemination, causing serious losses to some enterprises and individuals [3, 4].

As a product of the digital era, the use of electronic files in archival work is becoming more and more widespread, electronic files may exist in various forms such as pictures, images,

*zhuz_1126@163.com
https://doi.org/10.65102/is2026530
Published: 30 April, 2026

characters, etc., and are more convenient than paper files in terms of storage and utilization [5-7]. Electronic archives, because of its characteristics arising from the computer system, requires computers and other equipment to complete the reading, writing and transmission, and has a certain dependence on computer technology and equipment [8]. In the whole life cycle of the archives, its content and form is not stable enough, resulting in electronic archives of the “four (authenticity, integrity, validity and security)” requirements are difficult to ensure that in the process of receiving, storage, utilization, transfer of the content can easily be tampered with, so that the value of its reduced [9].

The ease of modification of electronic information, and the instability of the storage medium increase the difficulty for archivists to manage electronic records [10]. In order to be able to solve these problems, archivists and record keepers need to evaluate their traditional management methods in light of new technologies. Ragaisis et al. explored Lithuania's establishment of a unified generic model for electronic documents covering their complete life cycle, and its development of the National Electronic Records Information System (NERIS) as a key outcome of its preparations for e-document applications, whose core functionality lies in the ability to receive official e-documents signed by a qualified e-signature and to permanently guarantee the document's completeness, authenticity, irrefutability, and long-term usability [11]. Using a descriptive qualitative approach, Caroline et al. examined the status of the implementation of a dynamic archival information system in a regional archives and library office in an institution where the implementation of the dynamic archival system covered archival recording, control, distribution, storage, and reduction, but due to the limited number of digitization managers, not all of the dynamic archives were digitized [12]. Benmakhlouf and Chouaou systematically describe the core features and operation mechanism of OnBase electronic document management system, analyze the functions of the system in realizing the efficient management of massive documents, files and archives, and find that the system can effectively support the digitization of documents and archives, promote the overall digital transformation of business processes, and guarantee the secure access to documents within the organization [13]. Las Johansen et al. concluded that the main problems of traditional archive management include insufficient storage space and difficulties in archive retrieval and monitoring, so they designed and developed an electronic document archiving and management system using the Sashimi model of the system development lifecycle to cope with the shortcomings of traditional archive management and to improve management effectiveness [14].

In the research field of e-records management, the concept of whole-process management is regarded as a comprehensive, systematic and process-control-oriented management strategy, aiming at constructing a management system covering all management activities of e-records. Li et al. designed an effective archive storage and compression system for network management, which realizes efficient compression and storage of text and image archives through a two-stage compression process, aiming to improve the efficiency of archive management with practical application value [15]. Badran and Hamoud propose a web-based electronic records management system that supports faculty members in uploading and updating their personal records of scholarly activities in real time and automatically archives details such as the time and place of acquisition of results and changes in title; Compared with the traditional paper-based management mode, the system significantly reduces the reliance on physical storage that is fragile, occupies space and is easy to be lost, significantly reduces manpower consumption, and at the same time achieves a fundamental improvement in the convenience and accuracy of information retrieval [16]. Falatiuk et al. describe the core architectural style and key technology selection for building a distributed e-record system, systematically propose the overall concept, terminology system, data model and core responsibilities that an e-record

system should satisfy, and recommend a collection of functions to realize these responsibilities; By comparing and analyzing different software architecture styles, the study finalized the use of a combination of microservice architecture and event-driven architecture as the optimal path to build an e-filing system [17].

Regarding the research on the retrieval efficiency of electronic archives, Ong et al. found that the existing manual archiving methods have drawbacks such as slow retrieval, document fragility and chaotic filing, and therefore designed a set of electronic archiving system aimed at guaranteeing the authenticity of the documents, enhancing the access efficiency and preservation security; the results showed that the system can significantly improve the effectiveness of archive management and provide support for the digital transformation of the government's administrative processes [18]. Imaniyati et al. pointed out that educational institutions are transitioning from traditional manual management to digital systems, which significantly improves the accessibility, transparency, and sustainability of archives, and that effective e-records management not only enhances institutional responsiveness, but also meets the needs of modern educational development [19]. Ur Rahman and Alhaidari proposed and developed a system that integrates institutional data resources into a digital library in the form of "digital objects". The system organizes the data through a structured approach and extracts meaningful information from it. Each digital object is recorded with a quantitative description known as "metadata". During retrieval, the system employs a filtering strategy to extract relevant information that best matches the query from the knowledge base [20]. In summary, as far as the research on enterprise e-file management process is concerned, although some scholars have explored the problems existing in the enterprise e-file management process, most of the researches are more focused on a single link of the e-file management process, and relatively few researches are specifically aimed at the quality of the e-file metadata and the efficiency of retrieval, especially in the specific application practice, the lack of empirical researches and practice cases makes these strategies and suggestions of the feasibility and effectiveness are difficult to be fully verified.

In order to solve the problems of low efficiency of information search and poor optimization of algorithms in the existing electronic information management system, this paper proposes an improved Drosophila optimization algorithm based on linear decreasing steps and logical chaotic mapping for searching the data of electronic files. The algorithm changes the fixed steps of Drosophila optimization algorithm to linear decreasing steps to improve the accuracy of Drosophila optimization algorithm. At the same time, logical chaos mapping is used to improve the stability of the Drosophila optimization algorithm. After that, in order to realize the statistics and assessment of the quality of e-file metadata, an e-file data quality assessment system containing six indicators, such as accuracy, and completeness, is constructed. Then the hierarchical analysis method and cloud theory are used to construct a cloud model for the assessment of e-file management data; and the metadata quality and search efficiency of an e-file are scientifically assessed by taking the example data M of the management process of the e-file as an example.

2 Metadata processing techniques based on electronic records management

2.1 Data search method based on DSLC-FOA algorithm

2.1.1 Electronic archives information management system design

In this paper, a Fruit Fly Optimization Algorithm (FOA) based on linearly decreasing steps and

logical chaos mapping (DSLC) is proposed to solve the problem of electronic archive information management with the following innovations.

(1) Changing the fixed steps of FOA to linearly decreasing steps is very effective in improving the accuracy of the fruit fly optimization algorithm in order to avoid falling into local optimization.

(2) The adoption of DSLC-FOA can weaken the sensitivity of the initial conditions of the optimization solution, reduce the influence of the initialization parameters on the optimal solution of the e-file information management, and improve the stability of the FOA.

The e-file management system architecture is shown in Figure 1. As shown in the figure, the e-file includes all the material data and the informationized data of the whole process of the machine from purchasing, testing, warehousing, receiving, ledger, inspection, maintenance and testing to scrapping. Through the host, virtualizer, memory and network software, to build "wired + wireless" data collection platform, to achieve full coverage of the electronic file data.

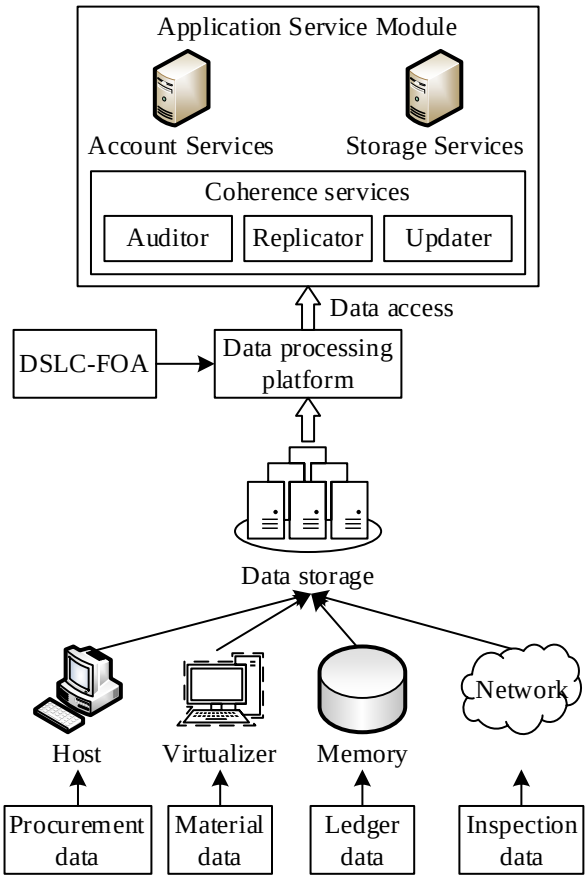

Figure 1: The Architecture of Electronic Records Management System

2.1.2 DSLC-FOA modeling

In this paper, a DSLC-FOA is proposed to optimize the design of e-file information management, and the whole algorithmic model improvement includes the following three parts.

(1) Linear Decreasing Step

In the processing of FOA, the first improvement step is to change the fixed step to a linearly decreasing step in order to avoid falling into local optimization and improve the accuracy. The specific improvement is as follows:

$$S' = S - \frac{S\left(G_{\max} - 1\right)\alpha}{G_{\max}} \tag{1}$$

where: S is the initial iteration step; S is the actual step; $G_{\max}$ is the maximum number of iterations; α is a data parameter with a fixed value of 0.8.

(2) Introduction of logical chaos mapping theory

In order to reduce the impact of the initialized e-file information parameters on the solution and enhance the stability of FOA, the second improvement step is to utilize the logical chaos mapping theory. The expression is as follows:

$$x(n+1) = ux(n)[1 - x(n)], x(n) \in [0,1] \tag{2}$$

where: n is the number of iterations for e-file information retrieval; u is the chaos control e-file information parameter. The above e-file information search system exhibits chaotic behavior. The formula for calculating the chaotic variables of e-file information is as follows:

$$C_{x(n+1)_i} = 4C_{x(n)i}\left[1 - C_{x(n)i}\right] \tag{3}$$

where: $C_{x(n)i}$ is the i th chaotic variable after n iterations of e-file information retrieval. The system is in a chaotic state when $C_{x_i} \in \backslash[0,1\backslash]$ and $C_{x(n)_i} \in \{0.25, 0.50, 0.75\}$. $x_i \in \backslash\left[a_i, b_i \backslash\right]$ can be changed by C_{x_i} through Eqs. (4) and (5) as follows:

$$C_{x_i} = \left(x_i - a_i\right) / \left(b_i - a_i\right) \tag{4}$$

$$x_i' = a_i + C_{x_i}\left(b_i - a_i\right) \tag{5}$$

where: C_{x_i} is the chaotic variable of e-file information; x_i is the value of the i th e-file information chaotic variable, which is converted to a regular variable after chaotic mapping transformation; b_i and a_i are the upper and lower limits of the value of x_i.

(3) Data integration processing based on cloud computing technology

The specific steps for obtaining e-file information data are as follows: in the process of integrating data management, let the overall e-file information management data set be X, and in the item set X, k data are selected for the construction of the first frequent itemset Y_1, and the principle of its data selection is: $P \geqslant Q$. where P is the probability of data seeking; Q is the minimum support of the itemset. The Y_1 construction expression is specified as:

$$Y_1 = X(k - y) \tag{6}$$

where: y is the constructed data item in the frequent item set Y_1. The candidate set of constructed data C with its second frequent itemset Y_2, where the constructed data items in the candidate set C need to be selected from the first itemset Y_1, and the selection benchmark is the two-dimensional itemset E with infrequent items removed; The second frequent term set Y_2 then selects the data in the candidate set C, which is selected based on the principle of $P < Q$. Its constructive expression is specified as:

$$C = Y_1(k-y) - E \tag{7}$$

$$Y_2 = C(k-x) \tag{8}$$

where: x is the constituent term in the frequent term set Y_2. When the first frequent itemset Y_1 and the second frequent itemset Y_2 of the constituent items, satisfy $x = y$, the two are merged, and similarly the other frequent itemsets are merged, and finally the construction of the candidate set is completed.

$$\begin{cases} Y_1 + Y_2 = X(k-y) + C(k-x) \\ Y_n + Y_{n+1} = X\left(k-y_i\right) + C\left(k-x_i\right) \end{cases} \tag{9}$$

where: Y_n is the nth frequent term set; Y_{n+1} is the $n+1$th frequent term set; y_i is the constituent term in the nth frequent term set Y_n; x_i is the constituent item in the $n+1$th frequent item set Y_{n+1}.

Then as in the previous step, the nth frequent itemset of the data is constructed; the association rules are examined in the constructed nth frequent itemset with the association rule confidence formula:

$$\text{Confidence}\left(Y_n > Y_{n+1}\right) = P\left(Y_n \mid Y_{n+1}\right) \tag{10}$$

where: Confidence is the confidence level, obtain n frequent item sets of association rules with a preset value less than the confidence level, and dynamically obtain the data that meets the rule.

2.2 Evaluation system construction of e-file metadata quality

The ISO 8000 data quality series standard fills the gap between the ISO 9000 quality management series standard and data products, which is an internationally recognized global data quality standard. There is no specialized data quality management document in the archive field, this study refers to ISO 8000, GB/T 36344-2018, and proposes a framework for archive data quality assessment based on the conceptual connotation of archive data quality, and combs to illustrate the indicators of archive data quality assessment from the three dimensions of form, content and utility. The assessment of archival data form refers to the assessment of archival data quality with respect to the external form performance of archival data; the assessment of archival data content refers to the assessment of archival data quality with respect to the specific content of archival data; the assessment of archival data utility refers to the assessment of archival data quality with respect to the degree to which the archival data can be provided to the users, and the assessment system of archival data quality is shown in Table 1. The assessment system of archival data quality is shown in Table 1.

Table 1: The Evaluation System of the Data Quality of Archives

Dimension	Metric
Archive data format	Normalization
	Integrity
Archive data content	Accuracy
	Safety
Utility of archival data	Chronergy
	Serviceability

2.2.1 Archival data form dimensions

(1) Standardization. Normativity is used to assess whether the data structure, data format, data type, data value domains, etc. of archive data are in line with domestic and international standards and the provisions of the system's preset programs. The extraction and expression of complex electronic archive metadata requires semantic and association-oriented specifications as a basis, and thus metadata specifications generally adopt XML as their default description format.

(2) Integrity. Integrity is used to assess whether the archival data maintains a unified and holistic state, and whether data entities and data attributes are missing. The semantics of archival data is expressed by a formalized language that follows a certain syntax. In order to ensure the understanding of the content, context and structure of the archive, it is necessary to maintain the semantic integrity of archival data in the archive management process.

2.2.2 Archival data content dimensions

(1) Accuracy. Accuracy is used to assess whether archival data objectively and truthfully reflect the facts of the archival record. The accuracy of archival data includes both the accuracy of the original data collected or created, and the accuracy of the processes through which it is stored, transmitted, and operated.

(2) Security. Security is used to assess whether the content of archival data involves personal privacy and state secrets, and whether necessary measures have been taken to ensure that the data are under effective protection and legal utilization. Archival data not only involves personal privacy and organizational secrets, but may also be related to state secrets and social stability, and is at a higher level of protection in the entire data system, making it necessary to set up a targeted security protection system to manage archival data.

2.2.3 Archival data utility dimensions

(1) Timeliness. Timeliness is used to assess whether changes in archival data occur in a timely manner in response to the use of the target resource. The timeliness of archival data is expressed in terms of time period as the extent to which the number or frequency distribution of archival data records within a certain timeframe meets the business requirements; in terms of point in time as the extent to which the number of archival data records based on timestamps, the frequency distribution, and the response time meets the business requirements; and in terms of temporal ordering as the relative temporal relationship between archival data elements.

(2) Availability. Usability is used to assess whether archival data can be accessed and understood. Openness of archival data is an important initiative for upgrading the services of archival institutions. Under the condition of ensuring that the open types, open formats, and open permissions of datasets and data interfaces have systematic regulations, consideration can be given to opening up archival data that has already passed the closure period and is not within the scope of confidentiality. At the same time, the degree of organization and development of

archival data has a direct impact on the user's perception of the quality of the data, and indirectly affects the role played by the data and the results produced.

2.3 Archival metadata quality evaluation model based on AHP-cloud modeling

2.3.1 Hierarchical analysis to determine weighting factors

Hierarchical analysis (AHP) is a multi-criteria decision-making method that combines qualitative and quantitative analysis. The steps are as follows:

(1) Establish a hierarchical structure containing the target layer, criterion layer, and decision-making layer, and have the experts compare them two by two according to their relative importance, and construct judgment matrices respectively:

$$A=\left[a_{ij} \right]_{n\times n} \tag{11}$$

where a_{ij} is the importance of the i th factor relative to the j th factor.

2) Calculate the maximum eigenvalue $\lambda_{\max}$ and eigenvector ω of the judgment matrix, and normalize to obtain the weight vector:

$$\lambda_{\max}=\sum_{i=1}^{n}\frac{(AW)i}{nWi} \tag{12}$$

3) Conduct consistency test. After obtaining $\lambda_{\max}$, calculate the consistency ratio CR according to the corresponding value of RI, if $CR<0.1$, then it meets the requirements, otherwise the judgment matrix needs to be adjusted until it passes the consistency test:

$$\begin{cases} CI=\dfrac{\lambda_{\max}-n}{n-1} \\ CR=\dfrac{CI}{RI} \end{cases} \tag{13}$$

where n is the matrix order.

2.3.2 Construction of a comprehensive assessment cloud model for e-records management

U is a quantitative domain expressed in exact numerical terms, and C is a qualitative concept on U. If the quantitative value $x\in U$ and x is a one-time random realization of the qualitative concept C, there are: $\forall x\in U,U\to[0,1],x\to\mu(x)$. The distribution of x over U is called a cloud, and a single x is called a cloud droplet.

A cloud model characterizes a concept by three numerical features Ex,En and He. Expectation Ex is the point that best represents the qualitative concept, the closer to the expectation, the more concentrated the cloud droplets are; Entropy En indicates the uncertainty of the state in which the cloud model is in, the higher the entropy, the greater the degree of ambiguity; Hyperentropy He reflects the entropy of the entropy, the higher the hyperentropy, the greater the degree of epiphenomenal dispersion. There are two kinds of cloud generators, forward and reverse, the forward cloud generator can produce quantitative cloud

droplets based on the numerical eigenvalues of the cloud, and the reverse cloud generator is to transform the specific data into qualitative concepts expressed by numerical eigenvalues, and then realize the mutual mapping between the qualitative and the quantitative, which effectively makes up for the shortcomings of the traditional methods in dealing with uncertainty.

1) Determine the risk evaluation system and standard cloud

According to the 5-level scale method, the risk level of electronic records management is divided into five levels: “low risk, low risk, medium risk, high risk and high risk”. Set the quantitative domain of risk level $U \in [0,10]$, according to the principle that the larger the score, the higher the risk is quantified into five score intervals, so as to facilitate the calculation of the standard affiliation cloud model:

$$\begin{cases} x_i = \left(x_{\max} + x_{\min}\right)/2 \\ En_i = \left(x_{\max} - x_{\min}\right)/2\sqrt{2\ln 2} \\ He_i = k \end{cases} \tag{14}$$

where k is a constant, usually taken as 0.1.

2) Determine the evaluation cloud of each evaluation index of the index layer

Experts are invited to score each risk indicator according to the risk level divided, combined with the actual situation of each type of risk, and calculate the risk evaluation cloud of each indicator through the inverse cloud generator as well as Matlab software:

$$\begin{cases} Ex = \overline{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i \\ En = \sqrt{\dfrac{\pi}{2}}\dfrac{1}{n}\sum_{i=1}^{n}\left|x_i - Ex\right| \\ He = \sqrt{S^2 - En^2} \end{cases} \tag{15}$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - Ex\right)^2 \tag{16}$$

where x_i is the expert scoring value and n is the number of samples.

3) Calculate the digital characteristics and similarity of the comprehensive evaluation cloud

The weight coefficients of various types of risks are calculated with the evaluation cloud of each indicator to get the parameters of the comprehensive evaluation cloud model. Calculate the similarity between the evaluation cloud and the standard cloud according to the following formula, and the largest risk level obtained is the closest risk interval of the indicator:

$$\begin{cases} Ex = \sum_{i=1}^{n} Ex_i W_i \\ En = W_i\sqrt{\sum_{i=1}^{n} Ex_i^2} \\ He = W_i\sum_{i=1}^{n} Hx_i \end{cases} \tag{17}$$

$$\mu_k = \exp\left(-\frac{\left(X_i - Ex_i\right)^2}{2En_i^2}\right) \tag{18}$$

where x_i is the expert scoring value and W_i is the weight value.

3 Statistical assessment of metadata quality and search efficiency analysis of electronic archives

3.1 Analysis of the results of the statistical evaluation of the metadata quality of electronic archives

3.1.1 Evaluation results synthesized for cloud computing

In order to verify the usability of the method proposed in this paper, some of the collected experimental data of an electronic archive management department is selected as the object of study, and these statistics contain a total of 10 keywords, which are represented by 1~10, respectively. Table 2 shows the partial scoring results of each evaluation index after treatment. The results show that the scoring results of 6 indicators on 10 keywords after processing are basically above 90 points, which shows that the comprehensive cloud computing results of this paper's model are better.

Table 2: Partial scoring results of each evaluated indicator after processing

Metric	1	2	3	4	…	10
Normalization	100.00	100.00	100.00	100.00	…	100.00
Integrity	90.99	90.88	90.92	91.07	…	91.02
Accuracy	94.87	94.94	94.97	94.96	…	94.92
Safety	97.92	97.96	97.96	97.90	…	98.00
Chronergy	97.97	97.96	97.95	97.97	…	97.95
Serviceability	100.00	100.00	100.00	100.00	…	100.00

The numerical eigenvalues of the cloud model of each evaluation index can be obtained by using the inverse cloud generator, and the numerical eigenvalues of the cloud of evaluation results of the indexes are shown in Table 3. The numerical eigenvalue of the comprehensive cloud of evaluation results is obtained as A6 (93.25, 0.8147, 0.2209), and the comprehensive cloud of evaluation results is shown in Figure 2. As can be seen from the figure, the cloud droplets of the evaluation result synthesized cloud are basically distributed between 91 and 95, and when the evaluation result value is taken as 93.25, the degree of affiliation is 1 at this time, and at this time, 93.25 best represents the quality condition of the data set M.

Table 3: The Digital Eigenvalue of the Evaluation Result Cloud

Metric	*Ex*	*En*	*He*
Normalization	100.00	0.0000	0.0000
Integrity	90.38	0.7569	0.1940
Accuracy	94.89	0.8087	0.2588
Safety	93.29	0.8671	0.2457
Chronergy	96.33	1.2597	0.6819
Serviceability	100.00	0.0000	0.0000

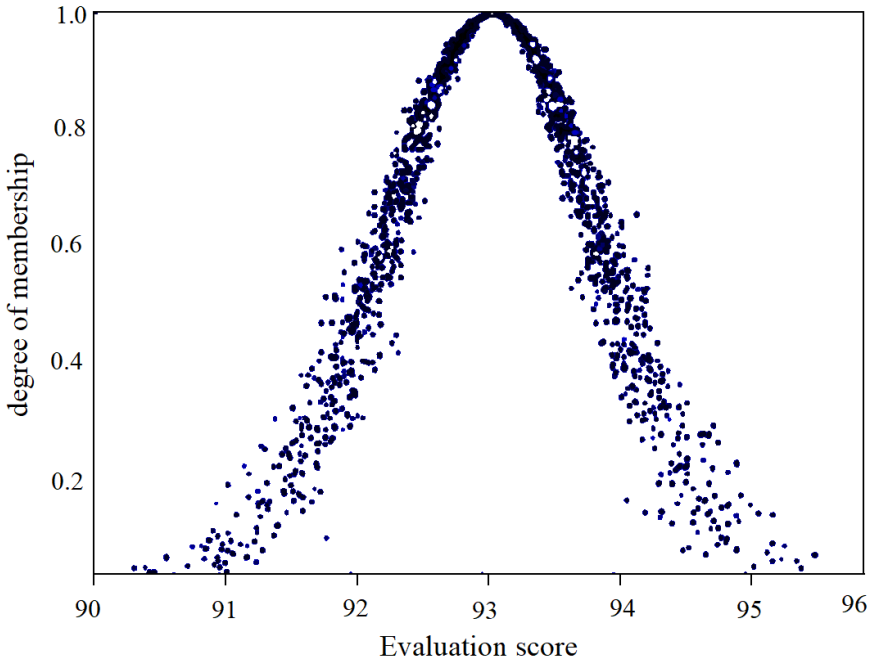

Figure 2: Evaluation result composite cloud map

3.1.2 Evaluation indicator weights

The weights of the evaluation indicators can be determined using the commonly used hierarchical analysis method. Based on the relative importance of each evaluation index, a judgment matrix is constructed to find the weights of accuracy (x_1), completeness (x_2), standardization (x_3), security (x_4), usability (x_5), and timeliness (x_6) indicators.

Eventually, the indicator weight vector W can be derived as:

$$W = \begin{bmatrix} \text{Quasigroups} \\ \text{Completeness} \\ \text{Normality} \\ \text{Safety} \\ \text{Serviceability} \\ \text{Timeliness} \end{bmatrix} = \begin{bmatrix} 0.3815 \\ 0.1809 \\ 0.1197 \\ 0.0453 \\ 0.0901 \\ 0.1825 \end{bmatrix} \tag{19}$$

The maximum eigenvalue of the judgment matrix $\lambda_{\max} = 7.5003$, then the consistency index CI=0.0629, and the consistency test judgment index CR=0.0629<0.1, which meets the requirement of matrix consistency, i.e., the weight vector W sought above is valid.

3.1.3 Similarity calculations

According to the cloud similarity calculation method, the similarity value between the evaluation result synthesized cloud and each evaluation level cloud can be calculated, and the similarity between the evaluation level synthesized cloud and the evaluation level cloud is shown in Table 4. Figure 3 shows the distribution of the evaluation result synthesized cloud and each evaluation level cloud. From the table, it can be seen that the evaluation result synthesized clouds are ranked in similarity as: excellent>good>medium>very poor>poor. It can also be visualized from the figure that the distribution range of the evaluation result composite cloud is

almost entirely within the range of the evaluation grade “excellent”. Therefore, the data quality status of the example data M can be considered as excellent.

Table 4: Similarity between Comprehensive Cloud and Grade Cloud in Evaluation

Order of evaluation	Terrible	Bad	Secondary	Good	Outstanding
Similitude	4.1979e-28	8.5621e-36	5.5814e-21	5.6309e-05	0.8185

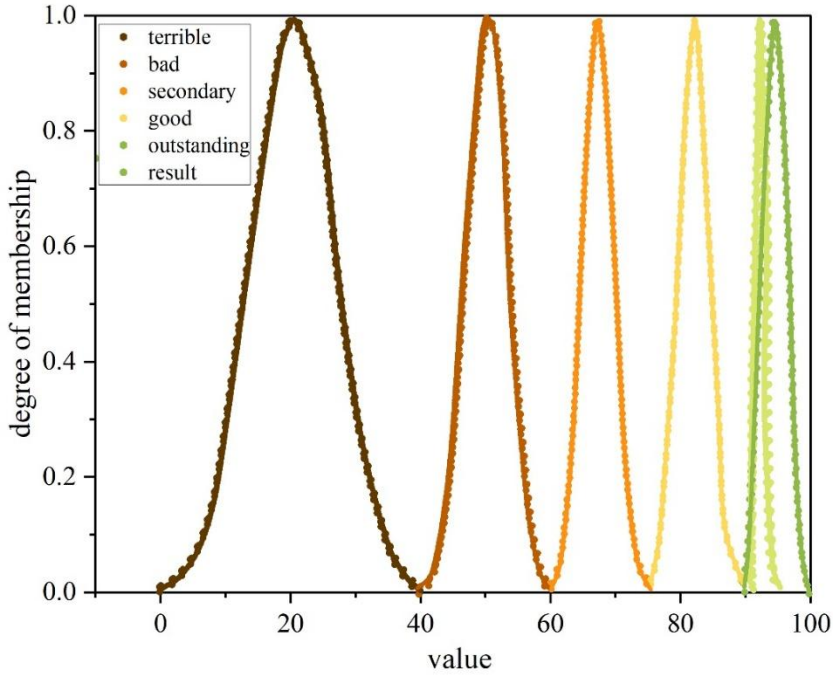

Figure 3: Distribution map of comprehensive cloud and evaluation grade cloud

3.1.4 Analysis of results

In order to further validate the usability of the cloud model-based test data quality evaluation algorithm, this paper adopts the commonly used fuzzy comprehensive evaluation method to evaluate the data quality of the example data M, and compares and analyzes the results of the two. To carry out the fuzzy comprehensive evaluation, this paper adopts the triangular and ascending half trapezoid of the affiliation function. The weights of the evaluation indexes and the affiliation matrix do fuzzy operation, and the result matrix of the fuzzy comprehensive evaluation can be obtained as F=W'*R=[0.6917,0.4755,0.0468,0.0000,0.0000], and the affiliation degree of the data quality level based on the fuzzy comprehensive evaluation is shown in Table 5. According to the principle of maximum affiliation, the quality of the test data obtained by the AHP-cloud modeling method is “excellent”. Compared with the traditional fuzzy comprehensive evaluation method, the data quality evaluation method based on AHP-cloud model proposed in this paper takes into account the fuzzy and random nature of the evaluation results, and at the same time, it can better avoid the subjective arbitrariness defects of the traditional method, which makes the evaluation results more credible.

Table 5: Membership Degree of Data Quality Grade of Electronic Records

Order of evaluation	Terrible	Bad	Secondary	Good	Outstanding
Membership value	0.0000	0.0000	0.0468	0.4755	0.6917

3.2 Analysis of the results of the search efficiency evaluation of electronic archive metadata

3.2.1 Comparison of completeness of data file collection

The search efficiency of this paper's retrieval system and the traditional retrieval system (Lucene-based e-archive retrieval system) for e-archive management is analyzed. The results of data archive collection completeness comparison are shown in Figure 4. It can be seen that the data archive collection completeness of the e-archive evaluation model proposed in this paper is higher, while the Lucene-based e-archive retrieval system is designed to collect less data archive completeness. At the 50th day, the maximum values of completeness of e-archive data collection by this paper's retrieval system and the traditional retrieval system are 100% and 60.03%, respectively.

The main reason for this difference is related to the division of hardware component performance modules by the hardware design in the paper. In the data collection module, the model in this paper controls the precise location of the archive data and transforms the data search mode, optimizes the system search structure, and promotes the enhancement of the system's fast retrieval performance. Simplify the data operation steps in data storage to facilitate system operation, improve the internal data information, and adjust the structure ratio. Synthesize the search characteristics of the archive data to achieve accurate retrieval and content analysis of the data. However, the Lucene-based electronic archive retrieval system design does not have this step operation, the collection and detection of data is less effective, and the system data archive collection is less complete.

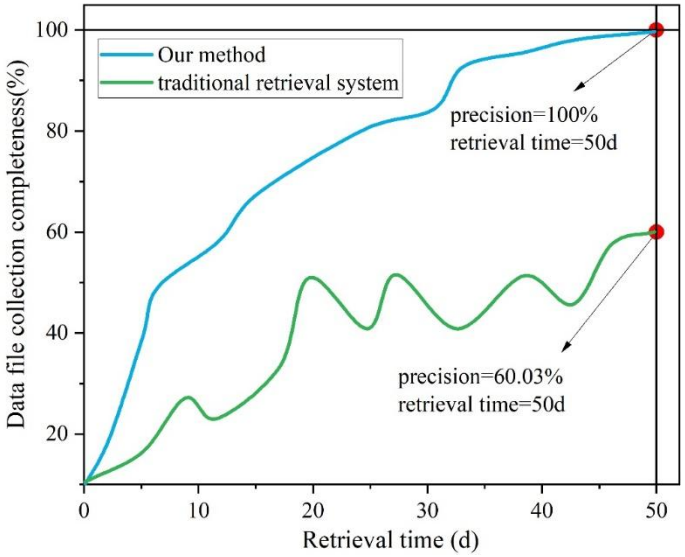

Figure 4: Comparison of completeness of electronic records data collection

3.2.2 Comparison of search accuracy

The results of the retrieval accuracy comparison are shown in Figure 5. The comparison shows that the retrieval accuracy of the e-file evaluation model proposed in this paper is 100% when the retrieval time is 800 s. The retrieval accuracy of the Lucene-based e-file retrieval system design is 60%, and it is obvious that the retrieval accuracy of this paper's model on the e-file data is much larger than that of the traditional retrieval model. With the increasing retrieval time, the design system of this paper has been improving the retrieval accuracy of e-file data, and has been located above the traditional retrieval system design. The design system controls

the data within a certain range of manipulation, preventing the interference of external data and the leakage of internal data, and is able to obtain more accurate retrieval data results and improve the retrieval accuracy of the system. This is because the system software design continues to enhance the basic performance of the data, convert the internal data operation mode, adjust the data state, and according to the data results of the data optimization parameters, the archive data for centralized control operations, continue to enhance the operational performance of the data system, reduce the time required for the operation, which in turn improves the efficiency of the system operation. While ensuring that the data system is in a safe operating state, the internal data of the software system is always updated to meet the operational requirements of the system.

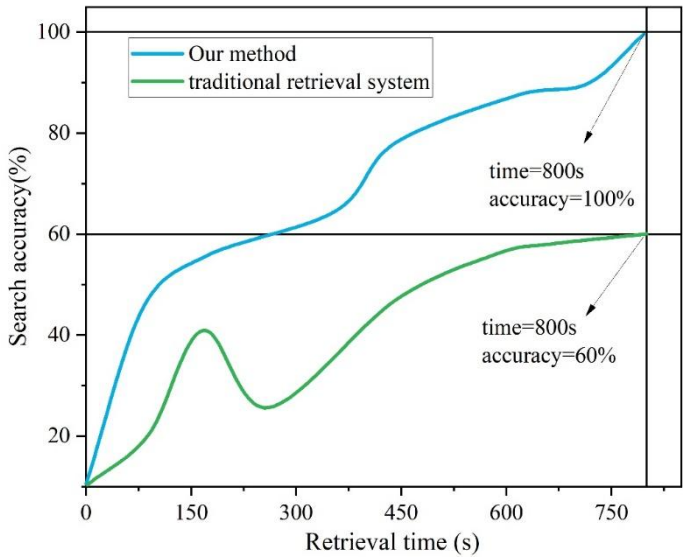

Figure 5: Search accuracy comparison results

3.2.3 Analysis of the results of the comprehensive assessment of the electronic records management system

In order to verify the usability of the designed electronic archive data retrieval system in the application, it is necessary to design the test environment and test method in the system testing, and utilize the traditional retrieval system as a control group for testing. In the testing process of the system, weights and similarity formulas can be utilized to aid in comparison when the thresholds are the same.

In order to ensure the usability of the experimental results, the built-in inference machine of the participant and Jena is uniformly used to ensure the consistency of the parameters in the retrieval process. In order to save hardware consumption during the testing process, PCs with higher hardware configurations are used to act as different location roles in multiple clusters. This facilitates the exchange of information data in the retrieval clusters and enables the interchange of IP addresses and port information.

In the performance testing of the system, the search performance of the retrieval system is mainly measured. The retrieval system performance is generally quantified by using the search accuracy rate P and the search completeness rate R, and the search accuracy rate is calculated by the formula:

$$P = \frac{D_r}{D_{Ra}} \tag{20}$$

where D_r denotes the number of relevant files retrieved from the e-file data based on the keywords, D_{Ra} denotes the number of all files retrieved, and the formula for the search rate is:

$$R = \frac{D_r}{D_a} \tag{21}$$

where D_a denotes all the keyword-related files in the electronic archive data. However, it should be noted that the search rate and the search accuracy rate are constrained by each other, and if comparing the retrieval performance of the two systems, it is necessary to synthesize these two indexes to make a comparison, so a F value is introduced to judge the retrieval effect:

$$F = \frac{2RP}{R+P} \tag{22}$$

In the above formula, the larger the value of F, the better the retrieval performance of the retrieval system. Under the above test environment and judging conditions, the retrieval system designed in this paper and the Lucene-based e-file retrieval system are tested respectively, and the retrieval results of the amount systems are compared.

(1) Keyword retrieval results of different systems

In the system performance test, the information content of the adopted electronic archive data is divided, mainly into data and title, and the keywords for searching are weighted, and the obtained search threshold ranges from 0.02 to 0.50. In the process of system testing, the selected electronic archive data set contains a large number of keywords, and in the retrieval performance test, one keyword is entered for each test, and the results of the retrieval system designed in this paper and the Lucene-based electronic archive retrieval system are used as comparisons, respectively. During the test, the keyword retrieval results of different systems are shown in Table 6. The results show that among the 10 keywords, the minimum values of search accuracy and search completeness of this paper's retrieval system are 96.31% and 88.79%, respectively; while the traditional retrieval system is 77.06% and 41.32%. In electronic file management, there is a big difference between the two retrieval systems in terms of the accuracy and completeness of the keywords.

Table 6: Results of keyword retrieval in different systems

Keyword Number	Our retrieval system		Traditional retrieval system	
	Precision ratio(%)	Recall ratio(%)	Precision ratio(%)	Recall ratio(%)
1	97.29	90.13	85.41	47.56
2	97.42	94.32	84.12	41.89
3	98.38	92.64	81.20	55.17
4	96.31	88.79	84.35	44.92
5	96.97	91.90	77.06	44.26
6	97.87	94.18	82.00	46.26
7	98.08	94.30	83.52	41.32
8	97.57	92.88	83.67	42.98
9	97.16	90.72	84.87	51.43
10	97.99	89.87	85.27	44.38

(2) Comparison of the results of the retrieval performance of the two systems

According to the results above, the F-value situation in the two retrieval systems can be obtained as shown in Figure 6. From the results in the figure, it can be seen that the retrieval system designed in this paper in the actual application test, the comprehensive evaluation index F-value are greater than 90, while the F-value of the Lucene-based electronic archive retrieval system is between 55.29 and 65.70, which can be seen in this paper's system of retrieval performance is much higher than that of the traditional model, and further proves that this paper's system of retrieval effect is excellent.

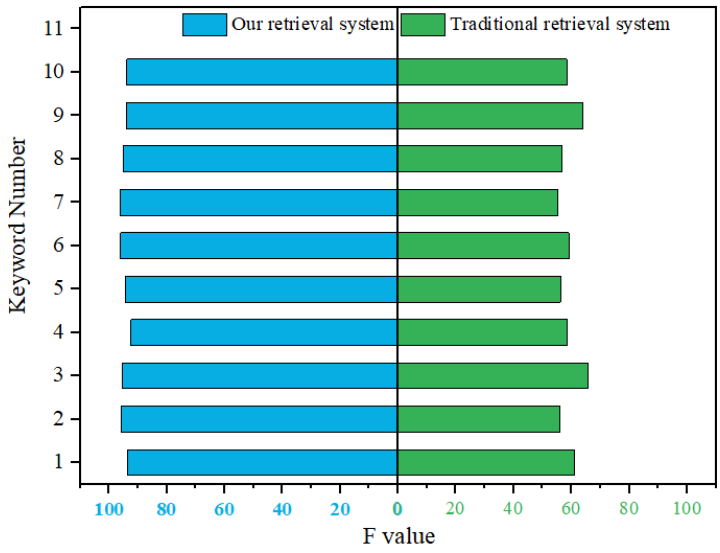

Figure 6: F-value in two retrieval systems

4 Conclusion

In this paper, a data retrieval method based on DSLC-FOA algorithm is firstly proposed, on the basis of which an e-archive data quality assessment system is constructed, and the quality of e-archive metadata is evaluated by combining the hierarchical analysis method (AHP) and cloud modeling. The results show that:

(1) The evaluation results of the comprehensive cloud of the model in this paper are basically distributed between 91 and 95, with better fuzzy and randomness, and when the value of the evaluation results is taken as 93.25, it represents the best quality condition of the data set and the evaluation results are more credible.

(2) The electronic archive evaluation model proposed in this paper is higher than the traditional data archive retrieval model in terms of the completeness of archive data collection and retrieval accuracy, which proves that the model in this paper can better improve the operational performance of the system, improve the data monitoring structure, and has higher research value. In addition, this paper's model has a keyword retrieval rate of 96.31% and 88.79%, and its comprehensive evaluation index F-value is greater than 90, which is much higher than the comparison model.

Funding

This research was supported by the Digital Services (Personalization) Development Project (including Electronic Archive Management) (Project No: 072900HK24030042).

About the Author

Dazhi Zhu is from Hubei, China. He received his Master of Science degree in Electronics from Queen's University Belfast, UK. His expertise and research focus lie in the field of Smart Grid technology and application.

Qing Yan, from Ledong, Hainan Province, graduated with a major in Automation from Nanchang University, is a senior engineer specializing in grid informatization construction and big data applications. Research areas include computer science and its applications, microservices, hybrid clouds, and big data.

Junlin Ji, from Ledong, Hainan Province, graduated with a major in Software Engineering from Jilin University, is an engineer specializing in grid informatization construction and data management. Research areas include computer science and its applications, data management, and big data.

Shi Chen, from Shantou, Guangdong Province, graduated with a major in software engineering from Hunan University, is a senior engineer specializing in grid informatization construction and big data applications. Research areas include computer science and its applications, microservices, hybrid clouds, and big data.

Shuhong Lin, graduated with a bachelor's degree in Computer Science and Technology from Hainan University, and currently works as an engineer at Information and Communication Branch of Hainan Power Grid. His research interests focus on informatization and digitalization.

Yuan Wang, obtained a bachelor's degree in software engineering from the School of Computer Science at Central South University. Currently, she is an engineer at China Southern Power Grid Co., Ltd. in Hainan. Her research interests include information construction and planning.

References

[1] Roland, L., & Bawden, D. (2012). The future of history: Investigating the preservation of information in the digital age. Library & Information History, 28(3), 220-236.

[2] Szekely, I. (2017). Do archives have a future in the digital age?. Journal of Contemporary Archival Studies, 4(2), 1.

[3] Romansky, R. P., & Noninska, I. S. (2020). Challenges of the digital age for privacy and personal data protection. Mathematical Biosciences and Engineering, 17(5), 5288-5303.

[4] Adelaja, A. A. (2024). The Estimated Age Limit of Storing Data and its Impact on Data Sustainability. International Journal of Theory of Organization and Practice (IJTOP), 4(1), 59-74.

[5] Peace, T., & Allen, G. (2019). Rethinking Access to the Past: History and Archives in the Digital Age. Acadiensis, 48(2), 217-229.

[6] Patterson, C. (2016). Perceptions and understandings of archives in the digital age. The American Archivist, 79(2), 339-370.

[7] Hajtnik, T. (2019). Digital Age: Time To Transform of Public Archives. Atlanti+, 29(2).

[8] Jaillant, L. (2019). After the digital revolution: working with emails and born-digital

records in literary and publishers' archives. Archives and Manuscripts, 47(3), 285-304.

[9] Nicholson, B. (2013). The Digital Turn: Exploring the methodological possibilities of digital newspaper archives. Media History, 19(1), 59-73.

[10] Klareld, A. S., & Gidlund, K. L. (2017). Rethinking archives as digital: The consequences of" paper minds" in illustrations and definitions of E-archives. Archivaria, 83(1), 81-108.

[11] Ragaisis, S., Birstunas, A., Mitasiunas, A., & Stockus, A. (2012, July). Electronic Archive Information System. In DB&Local Proceedings (pp. 107-114).

[12] Caroline, D. A., Ismanto, B., & Rina, L. (2022). Implementation of digital archives using a dynamic archive information system. Jurnal Kajian Informasi & Perpustakaan, 10(2), 189-204.

[13] Benmakhlouf, H., & Chouaou, A. (2024). Electronic document, information, and archive management systems in economic institutions: A descriptive study of the onbase system. International Journal of Professional Business Review: Int. J. Prof. Bus. Rev., 9(6), 11.

[14] Las Johansen, B. C. (2017). Development of electronic document archive management system (edams): a case study of a university registrar in the Philippines. International Journal of Digital Information and Wireless Communications (IJDIWC), 7(2), 106-117.

[15] Li, J. (2022). Design of an effective archive management system with a compression approach for network information technology. Wireless Communications and Mobile Computing, 2022(1), 3503841.

[16] Badran, A. S., & Hamoud, A. K. (2024). Electronic Archive Management System: A Case Study in University of Basrah, Iraq. Southeast Europe Journal of Soft Computing, 13(2), 1-9.

[17] Falatiuk, H., Shirokopetleva, M., & Dudar, Z. (2019, October). Investigation of architecture and technology stack for e-archive system. In 2019 IEEE International Scientific-Practical Conference Problems of Infocommunications, Science and Technology (PIC S&T) (pp. 229-235). IEEE.

[18] Ong, D., Yanti, V. A., Sofyanty, D., & Kusumandari, S. (2025). Design of Archive Web Information System Electronic Documents in the Office Kotabaru District. Jurnal Sosial dan Sains (SOSAINS), 5(4).

[19] Imaniyati, N., Sobandi, A., & Adman, A. (2025). Optimizing electronic archive management through information and communication technology for educational SDGs advancement. Jurnal Cakrawala Pendidikan, 44(3), 714-729.

[20] ur Rahman, A., & Alhaidari, F. A. (2018). The digital library and the archiving system for educational institutes. Pakistan Journal of Information Management and Libraries, 20, 94-117.