N. G. Surovtseva

Electronic document: to the problem of identification

The article considers the development of approaches to identify electronic documents and their classification in the area of records management and archival science. The ability to transfer documents that never existed on paper for permanent storage to archive makes us re-identify scanned documents as electronic copies of documents. The article also includes current approaches to identify the databases and web-documents as electronic documents in the conditions of their archival storage.

Key words: record, electronic document, classification of electronic documents, records management, archival science.

A considerable amount of papers of specialists in the area of records management and archival science is focused on the study issues of electronic documents. However, the overwhelming majority of them relates to the practical aspects of electronic documents functioning in the systems of electronic records management, or to the digitization issues and further storage and usage of electronic copies of documents on the conventional storage media in archives.

Two questions still remain relevant: firstly, which documents exactly should be assigned to electronic, meaning the problem of identifying electronic documents remains, and, secondly, which definition should be given to the electronic document[[1]](#footnote-1). The last one will depend directly on solving the first problem.

As it is known, this term began to be most actively used in Russia since the late 1990s. The term “machine-readable document” used before that time and other terms derived from it became absolutely unsuitable owing to the fundamental change in the technology of work with documents, caused by personal computer use.

Computers as tools for calculation, processing and preservation of information hardly affected the established processes of documentation and organization of work with documents: their implementation resulted in the unification of document’s form on paper, ensuring data reading by means of computing. Only technologies that ensured the visualization of document by human and the possibility of its transmission with no loss of content and appearance using a personal computer allowed to turn the page in automation of work with documents, allowed to create, transmit and store document with no use of paper.

So, electronic document owes its very appearance to creation and use in the management of modern information and communication technologies and information systems, which includes electronic records management system. The introduction of the term required its definition, and therefore the identification of its entity: what is electronic document? Is it a special type of document or just a document’s form, a record on machine storage medium? Describing the situation in solving this issue as long ago as at the end of XXs, M.V. Larin noted that “electronic document” term arises with the arrival of personal computer, as a record possessing new, different from traditional ones characteristics, but the relevant definition did not appear. There was not a common definition abroad as well2.

From our point of view, this phrase contains the entire root of the problem: the attempts to give a definition to the electronic document through its characteristics, differentiating it from a traditional record, inevitably bring researcher to the area of information technologies, through which the electronic document appears. For example, M.P. Bobyleva included the requirements for designing record, time of its transportation, capabilities of reference and research work, convenience of information perception, an ability to work with several documents at the same time and to highlight a fragment in the process, an ability to make a copy of the record, to ensure the preservation and accessibility of the document, data protection, etc. to the characteristics, differentiating electronic document from the paper one. It allowed author to make a conclusion on the presence of multidimensional, systematic differences between a paper and electronic document, and the necessity to consider electronic documents and electronic records management as qualitatively new systemic phenomenon3. All these characteristics are substantial, but mostly for electronic records management, since they are specificities of the technology of work with documents, not the entity characteristics of the electronic document.

Taking into account the inextricable connection of electronic document’s emergence with modern information and communication technologies, the experience in studying electronic document through the study of specificities work technology with electronic documents in electronic records management system, through the study of electronic records management, is the most common. That is how the whole variety of definitions appeared, where the specificities of storage medium and formal and logical structure of document, and other technological characteristics are reflected, but all of them appeared to be poorly applicable for records management sphere4.

Meanwhile, electronic document definition, as we have already mentioned, will depend directly on which information objects will be called by this term. Thus, one of the most important problems on the way of establishment of theoretical approaches to studying of electronic document is a problem of typology, or, at least, systematization and classification of electronic documents. It is noteworthy that the relevance of this problem was clearly understood by the researchers and specialists in the area of record-keeping and archival science as long ago as at the end of XXs. It is no coincidence that when defining the program of promising scientific researches for the period from 2001 to 2010, prepared by VNIIDAD in 1999, there was envisaged a topic “Typology of electronic records”5. Despite the fact that “electronic document” concept entry is connected with arising need for the term, allowing to define the documents, passing all phases of their lifecycle – from the creation to elimination or transmission to the eternal archival storage – in electronic form6, it is obvious that this term was used (and is still used) to describe not only such am information object, but for documents copies on conventional storage media scanned or digitized and transmitted into electronic form.

Archivists were the first to turn to the classification issues of electronic documents. At the turn of XX-XXI centuries a whole variety of papers appeared, in which there were proposed classification schemes for electronic documents of the archive, or there were considered specificities of work with the certain types of electronic documents in the archive7. A little later, specialists in records management got engaged with this issue, considering the specificities of managing different types of electronic documents8. The first synthesis article allowing to consider different issues of work with electronic document over its lifecycle was “Electronic documents in management” paper of Larin M.V. and Ryskov O.N., published in 2005. The analysis of “electronic document” term use in foreign and Russian practice allowed authors to attribute the following categories of information objects to electronic documents:

* Electronic documents themselves, which lifecycle takes place in the electronic environment only;
* Electronic (digital) copies of records on paper and other media;
* Databases (registries, lists, cadasters and other);
* Web-documents9.

It should be mentioned that the basis for building these classifications, attributing one or other information object to electronic document was, in many ways, in studying foreign experience, recommendations of International Council of Archives for work with electronic documents. It was proposed by many regulations of foreign countries to include e-mail messages, databases management (in which registries, filing cabinets and record books on paper were transformed) into the records management sphere, which shows the necessity of information preservation of different sites and social sites.

Moreover, the identification of the above-mentioned information objects as electronic documents corresponds to the term “document”, adopted in Russian records management, as the information recorded on the medium with the references allowing to identify it, and also to the term “archival document”.

Thus, based on the studying of foreign experience, at the beginning of the century in Russia there was formed a consistent for record-keeping and archival science classification of electronic documents, which could include not only the records with their entire lifecycle in electronic environment, but also electronic documents copies on conventional storage media, databases, which can be considered as the complicated electronic document in case of their conversion into a set of files, and, for example, web-sites of organizations – sources of gathering. Such approaches to classify records remained until the present time10. However, the adoption of new regulations in the sphere of work organization with documents required a clearer selection of the electronic documents group from the list of above-mentioned information objects.

Earlier in record-keeping as well as in the archive specialists worked not so much with documents, the entire lifecycle of which takes place in the electronic environment, but more with scanned or digital copies of documents 11. In so doing, both documentalists and archivists created such electronic documents (electronic documents copies) in their information system and solved the problems of their further storage and functioning on their own.

Audio-visual records are the exception, their creation, transmission and storage on digital media does not cause so much trouble, because they have no connection with the information system (like text and image documents), and, in fact, there was a change of medium type for them, and also scientific and technical documentation, which specificity we do not consider in this article.

The approach allowing to attribute electronic document copy to electronic documents was also relevant in the early stages of automated electronic records management systems development, when in the overwhelming majority of cases throughout the lifecycle document from the electronic environment flowed to the paper at the time of its signature, and vice versa. And if we had not considered electronic copies of paper documents as electronic documents, we would have had electronic records management systems without electronic records. Meanwhile textual managerial documents continued to be received for storage to state and municipal archives on paper.

Such a mixed lifecycle practice of a document continues today. However, maybe we are ready, at least at the level of the state management, for the realization of such a model of electronic records management, which allows to completely abandon paper document. This was facilitated by the development of regulations base in the sphere of electronic record management, as well as the quality of information infrastructure and the level of functionality of the modern electronic records management systems.

With the adopted changes for Record-Keeping Rules in the Federal Executive Bodies of Power12 two objects in record-keeping included in electronic records management system were selected: electronic document and electronic document copy. With that the term “electronic document” ended up being tightly related to non-paper technology of document creation in the information system, and the presence of electronic signature. In the same aspect electronic document (textual) is defined in the adopted Rules of Organization Archives Work, which transmission to storing should be performed in the container format, which composition includes content and electronic document metadata, electronic signatures files and a visualized copy of the textual electronic document in the PDF/A format13.

In this case textual managerial electronic documents not existing on paper and earlier transmitted to the archive only on paper will be transmitted to the archive storage. In that regard, today the clarification of the term “electronic document” becomes relevant for the archivists as well.

Under the current circumstances, in our view, it would be useful to use the term “electronic document” regarding textual managerial documents only in relation to records, which entire lifecycle flows in the electronic environment, and to use the term “electronic document copy” regarding scanned or digital record copies, as it is enshrined under the Rules of Record-Keeping. Otherwise, we will have to clarify which electronic documents exactly are described every time: initially created as electronic document or scanned (digital), because the technologies of work with them will have some differences14.

A few words should be said about other groups of information objects, which were above attributed to electronic document.

As we recall, databases were allegedly attributed to electronic document. The reasons which led to the possibility of such identification are, in our view, two factors. Firstly, databases grew from lists, registries, cadasters and other, which were earlier kept on paper and were acknowledged as documents. It has to be said that such acknowledgement was well reasoned, despite the fact that such documents were constantly complemented with new records. Since the creation of such documents had a discrete, meaning divided, intermittent character, at this particular moment they possessed all characteristics of the document: credibility, unity, authenticity and suitability for use.

Secondly, storage of recorded information on machine media began exactly from the machine-readable data storing, frequently numerical, statistical. Their transmission for long-term storage was performed together with the related hardware and software. That meant the transmission of databases in the truest sense of the word with the relevant DBMS. Such technology was declared ineffective since the last century. As a result there was adopted a technology of data storage separate from DBMS without any structural and logical relations, as a set of files15.

Thus, database storing is more appropriate to be considered as the storage of a set of documents, which composition is largely determined by the list of formed reports, and the creation of such set of documents is possible at any moment of interest. But in this case it will be about not a database storing, but a set of separate documents (copies), which can be created on paper as well as in electronic form.

Let us now turn to so-called web-documents, among which are documents placed on the sites pages in the Internet. From our point of view, for the sphere of operational document management or record-keeping, records management of the organization’s official site is not an actual problem at all, because all the information presented on it will be reflected in the documents of organization, and site management is a function of organization. The information will be transmitted to the site from some document. The official site is virtual information standing close to the consumer as much as possible and containing background information, which should be constantly updated.

The problem of information reservation from the sites present in the Internet was brought up by the international archival and library communities at the turn of XX-XXI centuries.16 Its practical solution was carried out by both in different ways, due to the differences in approaches to gathering of archives and libraries. For the archivist techniques and methods of storing “virtual reality” should not break “the principles of organization of the archival science, approved by a century of experience and a tradition”17. One of the basic principles is carrying out the value examination when selecting documents for the archival storage, meanwhile the basis of gathering of libraries is the receipt of the required copy of document. This difference has a crucial importance for the development of this direction in libraries and archives.

A significant experience of sites data selection for storage was accumulated at the beginning of our century by the Central Archive of Records on Electronic Media in Moscow (CAREMM). In our view, a very significant experience of web-sites data storage is a formation and collection description of web-documents “Hostage Taking in Moscow's Dubrovka Theater on October 23-26, 2002”18. The aim of this work was a selection, storage and description of the document’s set published in the Internet on this topic. There were targeted sites, types and a theme of records for collection, basic principles of systematization, a created working group of 2 persons – archive workers, “who one by one formed the collection during the entire working day”. With that, when downloading the documents from the forum, for example, there was performed a text editing (removal of obscene expressions, foul language). There was carried out a parallel work on the procedures of preservation and description. Authors describe in details the difficulties they faced when applying traditional approaches of documents systematization to the web-sites, which they were not able to fully embody, and also all the technical problems related to the migration from medium to medium. Thus, it becomes clear that this is not about a collection of web-records, but a creation of a new information product based on the Internet sites information. As a result of this work, there was created a new author’s information resource.

Sites selection closely resembles an implementation of the thematic request, preparation of which without a specific user and a target audience is a very difficult, costly and, mainly, a subjective question to a greater extent. And the completeness of information is very important for the researcher, because the site not corresponding to the criteria of external characteristics (web-design standard, style unity, image quality, etc.)19 may have the necessary information.

The situation is completely different in the system of library information resources. National libraries of many countries with the use of special software collect the information from all sites, which have a national domain in their URL-address (.ru for Russia). The participation of national archives in this activity is caused by their inclusion into the state system structure of information resources, and is mostly common for English-speaking countries, where the integration of archival and library spheres is much higher than in the Russian Federation.

Web-archiving in such form on the state level is performed within the responsibility for the preservation of the required copy of document20. Different strategies of data collecting are used on this (harvesting):

* Bulk Harvesting;
* Selective Harvesting;
* Event Harvesting.

It is worth noting that these are the strategies of harvesting, and not forming special information resources. Selective Harvesting is applied to the dynamic sites with the frequently changing information; when the Bulk Harvesting is impossible to carry out, it is performed with frequent intervals. And Event Harvesting is applied to the sites, which will presumably disappear as soon as the event ends. The result of such harvesting is a set of files, both textual and graphical, which is a statistical representation of the Internet pages, generally, only first and second level. After their systematization and processing (description) they can be introduced to the users. Thus, a new information resource is being formed.

These technologies have hardly been developed yet, and are only beginning to be used in the Russian Federation. The possibility and, mainly, feasibility of their usage in archive is a separate topic and cannot be considered within this article. It is important for us that web-document (Internet site) takes on a discrete view when stored, and becomes a set of files, meaning separate electronic documents (copies), moreover, document in their broadest understanding as the documented information, which, if gets to the archive, then not from the operational records management system.

Therefore, today we are not able to build a common consistent classification of electronic documents for record-keeping and archival science. In record-keeping those are legally relevant document, initially created in the electronic form, in the archive this is the documented information in the digital form, which can also be ultimately reduced to three groups: document, initially created in the electronic form, electronic document copy and electronic document copy on traditional medium. At the same time in the circumstances of unified lifecycle of the electronic document in record-keeping and archival science there is a necessity of its unambiguous classification as a document, which entire lifecycle from creation to storage in state (municipal) archive takes place in the electronic environment.

References

1 Larin M.V. Problems of Archival Storage of Digital Records on Behalf of a Citizen, Society, State // Electronic Information Preservation in the Information Society: Sourcebook of the international conference (Moscow, 3–5 October, 2011), M.: MCBS, 2012, p. 166.

2 Larin M.V. Several Problems of Managerial Document Evolution // Herald of an Archivist, 1999, n. 6, p. 43.

3 Bobyleva M.P. Administrative docflow: from paper to electronic one. Problems of theory and practice. M.: Pub. house MEI, 2010, pp. 21–32.

4 Yankovaya V.F. Electronic Record as an Object of Records Management // Herald of Volgograd State University. Ser. 2: Linguistics, 2013, n. 3 (19), pp. 229–235.

5 Electronic Records: Program of Scientific Researches (2001–2010) /

Rusarchive, VNIIDAD; led by M.V. Larin, 2000, p. 35. Dep. on records management and archival science n. 10595.

6 Documentation science: Textbook for students of higher education institutions /

M.V. Larin, Е.А. Pleshkevich,V.F. Yankovays and others; edited by M.V. Larin. M: Academy, 2016, pp. 50-51.

7 Zalaev G.Z. Analysis and Classification of Electronic Documents // Herald of an Archivist, 1999, n. 2–3, pp. 62–65;

Problems of Classification of Electronic Records // Electronic documents and archives: Theory and practice: Materials of conf. (Moscow, 8–9 April, 1999). М., 1999, p. 1–4;

Several Questions of Creating Electronic Archives // Archival science and source science of the national history: Problems of interaction at the present time: Thesis of the reports and presentations on the third all-Russian conf., 25–26 February, 1999. М.: Scientific book, 1999, pp. 258–261;

Tikhonov V.I. Entity Characteristics, Composition and Classification of Electronic Documents // Documentation in information society: Electronic record-keeping and electronic archive: Reports and presentations on the VI international scientific and practical conf., 24–25 November, 1999, М., 2000, pp. 214–218;

Yushin I.F. Documentation of Virtual Reality and View on this Process from the Perspective of Source Specialist // Archival science and source science of national history, pp. 236–241.

8 Sokova A.N. Electronic Document and Electronic Archive // Record-keeping, 1999, n. 1, pp. 6–10;

Kuleshov S.G. “Electronic document” concept // Documentation in the information society: Electronic record-keeping and electronic archive: Reports and presentations on the VI international scientific and practical conf., pp. 54–57;

Ryskov O.I. Databases as a Special Group of Electronic Documents of Organization and the Most Important Information Resource of the Country // Record-keeping, 2004, n. 2, pp. 40–43;

Eremchenko V.A. On the Problem of Electronic Documents // Secretarial business, 2001, n. 3, pp. 31–33; and others.

9 Larin M.V., Ryskov O.I. Electronic Records in Records Management: Study guide. M.: Pub.: “At Nikita’s gate”, 2008, pp. 72–73.

10 Larin M.V. Electronic Records: Theory and Practice // Bulletin of the RSUH. Series: Documentation and archival science. Informatics. Protection of information and information security, 2015, n. 2, pp. 53-63.

11 Audio-visual, and scientific and technical documents were exceptional.

12 The Russian Federation Government Decree as of 26 April 2016, n. 356 “On Making Changes in the Rules of Record-Keeping in the Federal Executive Authorities”.

13 Rules of Storage, Generation, Accounting and Use Management of the Russian Federation Archive Fund Documents and Other Archive Documents in the State Authorities, Local Authorities and Organizations, approved by the Order of Ministry of Culture РФ as of 31 March 2015, n. 526 (registered by the Ministry of Justice on 7 September 2015 г., reg. n. 38830), par. 4.34.

14 This statement is not appropriate for audio-visual documents being received today for the storage in the state archives on the physically isolated media, as almost all them will be copies.

15 Tikhonov V.I., Yushin I. F. Electronic Documents and Archives: International Experience and Russian Reality: Analyt. Note // Electronic records and archives: Theory and practice, pp. 33–86.

16 Afanasyeva L.P. Problems of Value Examination of the Internet-resources Using the Example of Sites // Herald of an Archivist, 2005, n. 4, pp. 118–119;

Ryskov O.I. Web-Documents Management // Record-keeping, 2004, n.7, pp. 41–44.

17 Yushin I.F. Documentation of Virtual Reality, p. 237.

18 From the experience of forming and describing Internet-documents collection “Hostage Taking in Moscow's Dubrovka Theater on October 23-26, 2002” // CADEMM. М., 2004, p 16. Background inf. of VNIIDAD n.11601.

19 Afanasyeva L.P. Mentioned work, p. 127.

20 A number of Russian archives of audio-visual documents is responsible for documents storing by the Federal law as of 29.12.1994, n. 77 «On the mandatory copy of the document».

21 Jacobsen G. Web-Archiving in Denmark // International conference (Мoscow, 3–5 October, 2011) М.: 2012, pp. 112–113.

22 Kiy M.I. Web-Archiving: Current State and the Development Prospects in Russia // Scientific and technical information. Series 1: Organization and Methodology of Information Work, 2014, n. 10, pp. 1-9.

1. Surovtseva N.G., 2017 [↑](#footnote-ref-1)