An Approach to Decision Support System Usage for Data Storage Configuration Variant Selection

O.V.Andriichuk, S.V.Kadenko,V.V.Tsyganok

Institute for Inforamtion Recording of the National Academy of Sciences of Ukraine, Kiev, Ukraine

Abstract. Electronic record management systems (including archives and libraries) should meet a large set of requirements, which can be described by tangible and intangible criteria. If digital data storage is needed for a library/archive, its configuration should be clearly defined at preliminary development stages. Tangible criteria can be represented quantitatively, by specific values of certain parameters. Intangible ones (reflecting, for instance, non-functional requirements) should be described by expert estimates, since there are usually no quantitative values to describe them. The paper suggests an approach to data storage configuration selection using multi-criteria decision making (MCDM) support methods, based on MoReq requirements and hierarchical storage management concept. MCDM support technology used allows selecting optimal data storage configuration, meeting both tangible and intangible requirements, in every specific case.
Keywords: data storage, configuration, decision-making, alternative, criterion, expert estimation.

1. Introduction

Electronic record management system (EDMS) development envisions 6 stages (pre-development stage, featuring pre-project research, analytical stage, when the requirements are formulated, project stage, when technical design is built, development stage (when the software is actually created), testing stage and implementation stage). At the pre-development research and analytical stages, when the terms of reference are formulated, it is extremely important to ensure their maximal compliance with the customer's requirements.

The task of formulating the terms of reference is placed on the project workgroup, which should include analysts representing the customer and analysts, representing the executor. Together they are to work out the terms of reference, ensuring long and sustainable functioning of the EDMS in future.

Typical requirements to electronic record management systems (ERMS) (a more general category) and electronic document management systems (EDMS) are specified in the Model Requirements (MoReq, 2001). Most requirements

Received: 31 March 2012 / Accepted: 20 November 2012 ISSN 2241-1925

described in this document are functional ones, which can be described by ether Boolean (for instance, "The ERMS system must provide facilities to manage input queues") or quantitatively expressible (such as, for instance, maximal number of metadata attributes supported by the system) parameters.
These parameters (factors or criteria) are often called "tangible" ones (Saaty, Jiao/Tseng etc).
In order to describe non-functional requirements (presented, mostly, in section 11 of MoReq standard) it is necessary to formulate intangible factors or criteria.
Digital libraries and, particularly, archives (also representing types of EDMS) have long life spans, so these non-functional requirements have vital importance for EDMS of these types, sometimes they can be considered even more critical than functional ones. In order to evaluate the extent of future EDMS compliance with non-functional requirements experts should be involved in this EDMS development stage. The more adequately the terms of reference (considering both functional and non-functional requirements) are formulated, and the closer they are to the customer's needs, the less time and money will be spent on re-constructing the future EDMS during succeeding development stages, and during its implementation.

2. Description of expert technology-based approach to EDMS terms of reference development process

It is suggested that the possible variants of future EDMS configuration (to choose from) are to be formulated and evaluated by analysts from both customer's and executor's sides – members of the project group. They will act as experts in this context (according to our terminology).
Again, usage of expert technology can facilitate prevention of future EDMS mismatches with the customer's needs, since experts are most competent specialists in the given field and they are most likely to be capable of predicting these future possible mismatches and considering them while formulating the requirements.
As it has already been noted, requirements to the future EDMS include functional and non-functional ones. Examples of functional requirements include security-related requirements, access-related requirements, requirements to data retention and disposal, searching, retrieval, rendering etc.
According to MoReq "…non-functional requirements often are difficult to define and measure objectively, it is nevertheless valuable to identify them so that they can be considered, at least at a high level". Examples of non-functional requirements are: ease of use, performance, system availability, ability of the system to adapt to media degradation, hardware and format obsolescence. Particular attention should be given to non-functional requirements reflecting hierarchical storage management concept (the more often the data are addressed, the higher the hierarchy level they are stored at – in each specific case these issues are to be defined by experts). Examples given here are general and each of them can and should be decomposed into a set of more specific ones.
Non-functional requirements represent a weakly-structured subject domain (i.e., it is problematic to define their interconnection and quantitatively describe

them). So, one of the most convenient way to describe it in the most thorough manner would be to use a hierarchical approach, i.e., formulate a hierarchy of requirements (Saaty, 2008), including both functional (tangible) and non-functional (intangible) requirements.

Part of the requirements can be taken straight from MoReq and included into the hierarchy once and for all, since, it is a standard for all kinds of EDMS and these requirements will constitute a universal part of the hierarchy.

It should also be noted that requirements set forth in MoReq are closely connected with each other (there are a lot of cross-references within the standard text) and, consequently, allow easy interpretation in the form of hierarchy graph. But, at the same time, MoReq allows adding new requirements for specific EDMS configurations. And it is to perform this task of adding new configuration-specific requirements to (or deleting them from) the hierarchy the experts should be addressed.

The project group would, naturally, include specialists from different sub-domains (subject domain experts, business-users, IT specialists, developers, analysts, testers, management representatives etc). So, each group of criteria (requirements), belonging to the sub-domains, is to be formulated by adequate specialists from the project group. Also, these very specialists are to evaluate the compliance of future EDMS configuration to the customer's requirements, related to their respective sub-domains. After the hierarchy formulation is completed, all requirements (criteria) from the hierarchy are to be weighted by the experts. After that all possible EDMS configuration variants are to be evaluated according to all criteria, and aggregate estimates (which can be also called relative efficiencies, or ratings) of configuration variants as to their compliance with the customer's needs are to be calculated.

3. Software description and a hypothetical example

The complex target-oriented expert evaluation technology, briefly described above, is implemented in "Solon" decision-making support software family, developed by the Laboratory of Decision-making systems (DSS) in the Institute for Information Recording of the National Academy of Sciences of Ukraine (http://www.dss-lab.org.ua/Main.html). It should be stressed that "Solon" DSS family is targeted at facilitating expert decision-making support in *any* weakly-structured domains. Multi-criteria expert choice of the most suitable EDMS configuration variant is just one of the suggested applications of the DSS. "Solon" DSS allows facilitating all the above-mentioned decision-making process stages with its user-friendly interface. Let us consider the DSS functioning on a hypothetical example where five alternative EDMS configuration variants are evaluated.

In this example the hierarchy of requirements was built by the paper authors (based on MoReq and general EDMS development concepts) but in reality it is assumed to be built by respective experts.

The EDMS whose alternative configurations are evaluated in this case is targeted at managing the hydro-meteorological information archive. The archive itself is meant to facilitate collection and storage of hydro-meteorological information as well as informational support of research tasks. Volumes and types of data to be managed in the archive are:

- 50 000 digital magnetic tapes with data on environmental conditions for several decades' period, 1500 Tb of satellite data on magnetic optic mediums, 1000 Tb of other informational materials.
- 200 millions of various paper documents in A0-A4 formats, 400 millions of photos, including documents on photo mediums, books and manuscripts.

The total archive volume is 6 Pb. Annual volume growth rate is approximately 10 Tb, including 1 Tb of ground-level hydro-meteorological measurements and 8 Tb of satellite data.

The key tasks of archive creation are: digitalization of paper mediums; data migration to modern storage devices; providing access to different data types; facilitation of web-technologies' implementation and multi-level (hierarchical) data storage.

Based on these tasks the following configuration variants were formulated:

<u>Configuration 1</u>: 1 magnetic tape library IBM TS3500 (http://www-03.ibm.com/systems/storage/tape/ts3500/), and disc data storage system IBM DS8300, management server based on IBM system z9 BC managed by OS z / vm and zLinux, archive data management software IBM DB2 Content Manager OnDemand and IBM Tivoli Storage Manager (http://www-01.ibm.com/software/tivoli/products/storage-mgr/productline/compare.html).

<u>Configuration 2:</u> 40 archival optical storage devices ELAR NSAM 7000-BD (http://ncm.ru/nsm_bd.shtm), disc data storage system IBM DS8300, control server on the basis of IBM system z10 EC managed by OS z / vm and zLinux, document archiving and management software Saperion.

<u>Configuration 3:</u> Data storage system architecture EMC Centera, representing redundant array of independent nodes (RAIN) Storages and servers. Access and storage nodes, included into the architecture, represent servers of Intel platform with ATA discs. Servers are connected with each other through internal LAN, and they also have Ethernet for external connection. All nodes are working under control of Linux OS modification. Search within the EMC Centrea is conducted by Centra Seek and Chargeback Reporter (http://www.emc.com/products/detail/hardware/centera.htm).

<u>Configuration 4:</u> 10 magnetic tape libraries Quantum Scalar i2000 (http://www.quantum.com/ServiceandSupport/SoftwareandDocumentationDownloads/S2K/Index.aspx), storage system based on magnet discs COPAN 400M Native MAID (http://www.sgi.com/products/storage/maid/400M/specifications.html) by SGI company, control server based on IBM system z9 BC managed by OS z / vm and zLinux, QStar HSM software.

<u>Configuration 5:</u> 70 libraries based on UDO or magnetic optic mediums Plasmon G638 (http://www.dataarchivecorp.com/udo-plasmon_g-638.htm). Storage systems on magnetic discs COPAN 400M Native MAID (one cabinet) by SGI company. Control server based on IBM system z9 BC managed by OC z / vm i zLinux. Software of document archiving and management system Saperion.

A hypothetical hierarchy of criteria according to which alternative configuration variants are estimated is built as follows. The top (zero) level of the hierarchy includes only its main goal – "building an effective archival data storage system". The first hierarchy level includes its immediate sub-goals (or sub-criteria) – "fulfilling the requirements to data storage system", "low cost of system creation", "approval by higher management", "time required for data storage system creation". These goals are further decomposed; fourth and fifth levels of the hierarchy are comprised by criteria, reflecting MoReq requirements (including non-functional ones). In general, the hierarchy includes about 90 criteria, interconnected with links. The hierarchy structure is shown on Figure 1.

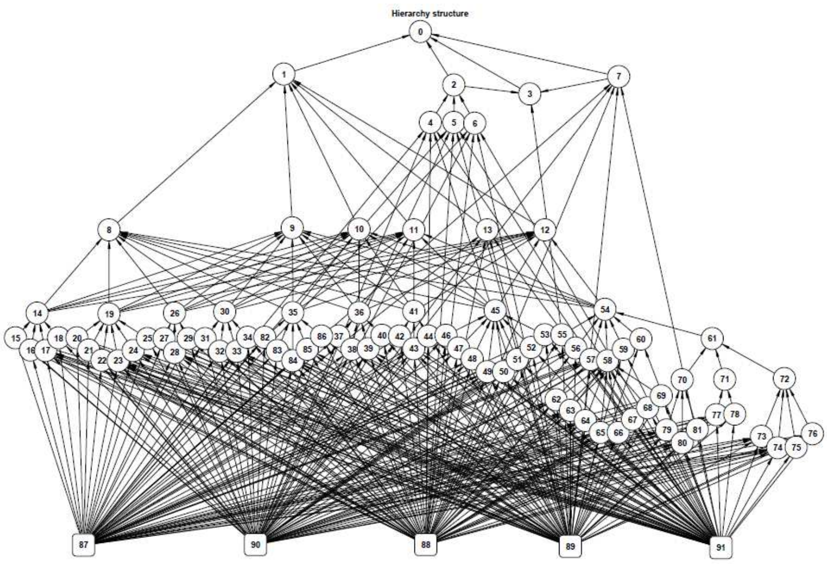

Figure 1. Criteria Hierarchy Structure

Relative influence of each criterion upon its "ancestor" in the hierarchy graph is estimated by experts. When all the influences are estimated, relative efficiencies (ratings) of alternative configuration variants are calculated.

Complete list of criteria (requirements) included into the hierarchy, and their respective numbers, can be found in Table 1.

#	Goal formulation	#	Goal formulation
0	Creation of a System for Archive Data Storage	46	Management of Non-electronic Records
1	Meeting the Requirements to Data Storage System	47	Hybrid File Retention and Disposal
2	Financial Costs of Data Storage System Creation	48	Document Management
3	Support of Data Storage System Creation by Top Management	49	Workflow
4	Existence of Standard Decisions	50	Electronic Signatures
5	Existence of Specialists with Previous Working Experience	51	Encryption
6	Intensity of Each Specialist's Participation	52	Electronic Watermarks etc
7	Time Needed for Data Storage System Creation	53	Interoperability and Openness
8	System Unity Principle	54	Non-Functional Requirements
9	Development Principle	55	Ease of Use
10	Complexity Principle	56	Performance and Scalability
11	Standardization Principle	57	System Availability
12	Universality Principle	58	Technical Standards
13	Principle of New Tasks	59	Legislative and Regulatory Requirements
14	Classification Scheme	60	Outsourcing and Third Party Management Data
15	Configuring the Classification Scheme	61	Long Term Preservation and Technology Obsolescence
16	Classes and Files	62	Hardware
17	Volumes	63	Operating Systems
18	Maintenance of the Classification Scheme	64	Production Standards of User Interface
19	Controls and Security	65	Relational Database Management Systems
20	Access	66	Network Protocols and Operating Systems
21	Audit Trails	67	Implementation of Encryption on Different Levels
22	Backup and Recovery	68	Exchange Standards
23	Tracking Record Movements	69	Applied Development Interface and Developer Kits
24	Authenticity	70	Media Degradation
25	Security Categories	71	Equipment Obsolescence
26	Retention and Disposal	72	Format Obsolescence
27	Retention Schedules	73	Format Migration
28	Review	74	Emulation
29	Transfer, Export and Destruction	75	Technology Conservation
20	Capturing Records	76	Alignment of Data and Software
31	Capture	77	Equipment Monitoring
32	Bulk Importing	78	Data Migration to New Modern Media
33	Types of Documents	79	Observance of Adequate Conditions of Media Storage, Usage and Processing
34	E-mail Management	80	Facilitating Scheduled Media Replacement
35	Referencing	81	Saving Several Document Copies and Their Comparison
36	Searching, Retrieval and Rendering	82	Class Identification
37	Search and Retrieval	83	Folder Identification
38	Rendering: Displaying Records	84	Volume Identification
39	Rendering: Printing	85	Document Identification
40	Rendering: Other	86	Extract form a Document
41	Administrative Functions	87	Configuration 1 (project)
42	General Administration	88	Configuration 2 (project)
43	Reporting	89	Configuration 3 (project)
44	Changing, Deleting and Redacting Records	90	Configuration 4 (project)
45	Other Functionality	91	Configuration 5 (project)

Table 1. List of criteria (sub-goals, or requirements)

The relative weights (ratings) of five configuration variants mentioned above (and included into the hierarchy at the lowest level as projects with numbers 87-91) are shown in Table 2.

Configuration number	1	2	3	4	5
Relative weight	0.2153	0.2361	0.1875	0.1667	0.1944
Configuration rank	2	1	4	5	3

Table 2 Ratings (relative weights) of alternative configuration variants

As we can see, Configuration 2 is the best one according to the specified criteria.

4. Conclusions

An expert decision-making support technology for evaluating EDMS configuration variants at pre-development stage is suggested. It proves to be particularly effective while evaluating EDMS configuration variants as to their compliance with non-functional requirements.

The technology allows capturing and considering mutual influences and interconnection of different requirements (represented by Boolean, quantitative and qualitative criteria).

The relative ratings of alternative configuration variants can be calculated using an algorithmically built estimate aggregation function, incorporating Boolean values, quantitative values and (in case feedback is present in the hierarchy graph) – iterative calculation of ratings.

One of the possible directions of future research is extension of the described approach to other EDMS development standards, beside MoReq.

References

MoReq Specification: Model Requirements for the Management of Electronic Records (2001). Retrieved November 2, 2011, from http://www.cornwell.co.uk/moreq.html.

Saaty, T. L. (2008). Relative Measurement and Its Generalization in Decision Making; Why Pairwise Comparisons are Central in Mathematics for the Measurement of Intangible Factors; The Analytic Hierarchy/Network Process. *RACSAM 102 (2)*, 251–318.