Jan Trąbka

Cracow University of Economics

ENTERPRISE CONTENT MANAGEMENT PLATFORMS – CONCEPT UPDATE, ROLE IN ORGANIZATION AND MAIN TECHNOLOGIES

Introduction

Enterprise Content Management (ECM) is now one of the key IT tools implemented in enterprises. This manifests itself in the remarkable dynamics of the market, its constantly increasing value and main players, as described by the analysts of Gartner*, a research and advisory company, in its annual Magic Quadrant for Enterprise Content Management report [GART12]. The concept itself, as well as its technological elements have been evolving over the last twenty years. This evolution and the current role of content management in enterprises are presented in the first two chapters of this paper. The third part is devoted to the main components and technologies used to create ECM platforms. The conclusion contains a brief analysis of the ECM world market and points to its future development trends. This paper is an introduction to the discussion on the problem of discovering new tools for system analysis and design which can be used to implement ECM platforms in their present form.

ECM on the map of enterprise information systems

As regards the information needs of economic organizations and the development of management information systems, since the 1970s enterprises have been implementing MRP/ERP-class systems, which ensure planning and control

* Gartner – a research and advisory company founded in 1979 in the United States of America. It specializes in the strategic use of technologies and technologies management. Gartner is currently present in 85 countries, has 6,100 employees, including 1,460 research analysts and consultants.

of their activities as far as material, financial and human resources are concerned. Systems belonging to this class support the execution of organization's main processes in a consolidated, uniform working environment for all the employees. The development of management methods such as Supply Chain Management (SCM), which allows for collaboration between supply chain partners by means of WANs, Internet, EDI, mobile devices and automatic identification technology, forced the opening of ERP-class systems to new e-business channels for the performance of economic activities [MOLL05]. The rise of the competitiveness on the market, increasing number of data processing sources and the complexity of the processes, brought about the need to create dedicated tools for the analysis, inference and economic decision-making support – Business Intelligence (BI) systems. It should be pointed out that ERP systems provide information services focused on enterprise's physical resources and this is the context within which the data is stored in their databases. It constitutes a great part of the information resources, however, not the whole of it. What remains are the processes supporting the fourth strategic enterprise resources, namely information. After being transformed into knowledge it becomes a decisive tool in the rivalry on the competition market. In order to determine which tools an economic organization will need to manage its information resources, one should examine how they are acquired, stored and processed into knowledge which is then used by employees in processes execution and decision-making. To specify these requirements it is advisable to refer to the Japanese theory of knowledge management in an organization, which creates the notions of explicit knowledge (that one can codify and convey by the means of formal and systematized language) and tacit knowledge (individual, context specific, difficult to formalize and transfer) [NOTA00]. The authors of the theory concentrated on demonstrating the transformation between the two aforementioned kinds of knowledge, which becomes the motive force of the organizational knowledge creation. The knowledge creation process has been divided into the following stages: socialization, externalization, internalization and combination. Within the framework of this model one may indicate a few information components supporting the respective stages:

- groupware,
- information repository management (in electronic and non-electronic forms),
- management of web-based information publishing,
- workflow,
- automation of business processes,
- social communication.

The aforementioned set of tools provided in the form of an integral information platform has now constituted itself in the concept of ECM tools. The approach ordering information tools, which originates from the knowledge management process, has also been described in [ZIEM07a] within the context of information technologies in tacit knowledge management. The fact that ECM as a concept linking the tools listed has not been mentioned there, points to its dynamic evolution and growing popularity over the last few years. For the purpose of this paper the author quotes the definition of ECM published by the Association for Information and Image Management (AIIM), a global organization providing conceptual and technological standards for tools used in an organization information management*. According to the most recent definition "Enterprise Content Management is the strategies, methods and tools used to capture, manage, store, preserve, and deliver content and documents related to organizational processes. ECM tools and strategies allow the management of an organization's unstructured information, wherever that information exists" [AIIM14]. Pointing to the place ECM systems take on the map of enterprise information systems (Figure 1) one can find that their working area covers the management of mostly unstructured information, stored in electronic and non-electronic form in any of the enterprise's sectors. Thus, they complement the ERP and other structured data oriented operating systems. At the same time, along with BI-class analytical systems, which support the creation of operating and strategic knowledge about an organization, ECM systems constitute the basis for developing comprehensive tools for knowledge management in an organization. Enterprise Information Management (EIM) systems are one of these tools. Kampffmeyer [KAMP13] quotes the following definition from Gartner: "Enterprise Information Management is the combination and integration of ECM with BPM (Business Process Management) and BI (Business Intelligence)". The integration of information from various sources within an organization and from the outside of it belongs to corporate portals, also referred to as the Enterprise Information Portals (EIP). These are defined by Ziemba [ZIEM07b] as follows: "(…) they are applications which grant organizations access to the internal and external information sources and provide their users with personalized, according to their needs, information indispensable for decision-making". Corporate portals join systems which consolidate, manage, analyze and distribute information within the organization and beyond its boundaries. The author mentions here the constituent systems: Con-

* AIIM – founded in 1943, is a global, non-profit organization that provides independent research, education and certification programs to information professionals. AIIM currently has 80,000 members.

tent Management System (CMS), BI, data warehouse and Document Management System (DMS). Since, as has been already presented, ECM includes and considerably expands the CMS and DMS components, one can assume it is the first class of systems that is the pillar of corporate portals development.

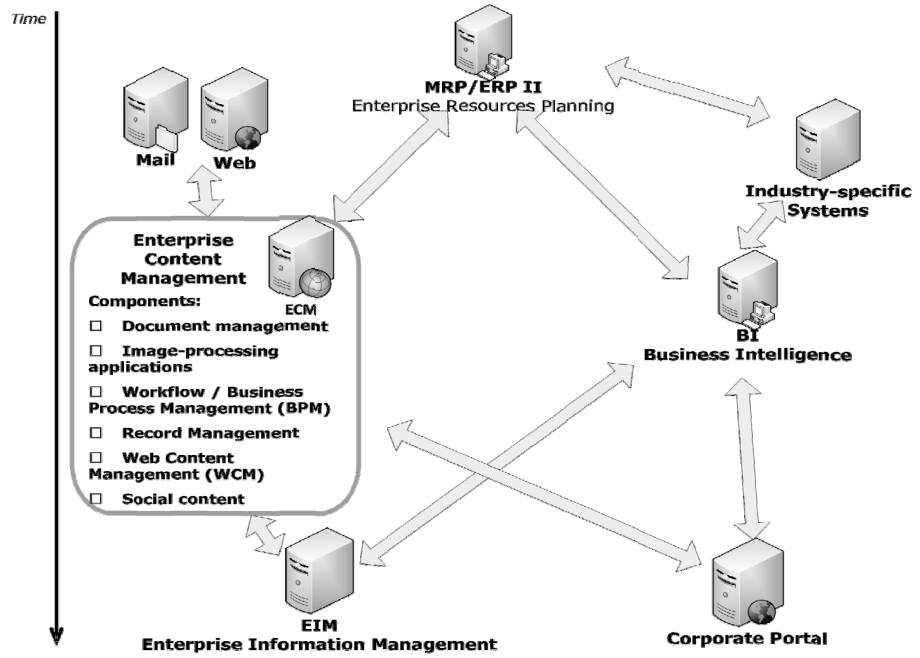

Fig. 1. Enterprise information systems

Source: Own elaboration.

ECM systems are a collection of components which have their own functionalities (e.g. a document repository managing electronic document life cycle), however, they are not dedicated to any specific functions or processes of an enterprise. It is by the implementation that enterprises decide on the process and the way in which the selected components will be used. For this reason, in this paper, ECM platform has been characterized as a set of tools for creating functionalities a given organization needs. When analyzing examples of ECM platforms implementations one can indicate the enterprise functional areas they most often support: office documents circulation with a central document repository, project management, quality processes management (including ISO), companies and capital groups documentation management, project and contract management, budgeting, financial documents circulation, payroll and HR processes, IT services management (helpdesk), management of intranet, internet and extranet websites.

The evolution of the concepts of "content" and "ECM"

The problems with clear classification of content management tools result from the evolution of the concept's interpretation and have been observed among publicists over the last ten years. The evolution is connected with gradual expansion of the meaning and its range along with the development of information technologies, especially the Internet, as well as with the needs of enterprises that these technologies are used by. Many authors, also Polish ones, just to mention Ćwiklicki [ĆWIK03] and Adamczewski [ADAM07], have discussed the etymology and evolution of the concept of content, which is the reason why this paper presents solely the technologies whose appearance substantially influenced the current form of the ECM platforms. At the same time the evolution presented shows the gradual expansion of the concept's range. The author's own idea of the concept is proposed at the end of this part of the paper.

Towards the end of the 1990s, along with the development of internet technologies, the concept of content was associated with what could be presented on a WWW site and the site's creation and updating processes. Adamczewski [ADAM07] analyzing the development of the Content Management System (CMS) conception quotes the following: "Content management has been an urgent problem since the time when various kinds of organizations and institutions had to develop a formula for dealing with the overload of information, as spreading and presenting information were essential for their functioning. What arouses the commentators' greatest interest is the problem of publishing content on WWW sites and it is this aspect of CM that has become the primary subject of most of the books on this area" [ROCK03]. One can see strong emphasis put on the stage of content storage and distribution – distribution by means of the Internet. The conception of workflow systems was born in the same time period, that is at the end of the 1990s [TRAB13]. Their role was to deal with circulation of documents, information and tasks, so they concentrated on the stage of content creation and distribution inside of an organization, with a particular stress on electronic document as basic means for content storage. As the workflow systems developed, the following technologies were included as their components: scanning, OCR, cataloguing services (dealing with document repositories), indexing, searching, and XML. The last one mentioned allows to normalize and send electronic documents between various information systems. Over the last years we have been watching the increasing role of social media defined as a set of internet and mobile applications, which enable creation and exchange of user-generated content [KAHA10]. This set of applications consists of: blogs (Twit-

ter), collaborative projects (Wikipedia), music-, photo- and video-sharing services, so called content communities (YouTube), social networking sites (Facebook) as well as all sorts of instant messaging clients and voice communication services (based on the VoIP technology). In order to show the significance of social media use in an enterprise, one should look at the latter as at a community of employees which every day in a number of ways of communication, exchanges various content. The process of gathering content using social media tools, aims not only at streamlining communication but also at creating essential organization knowledge bases.

Currently, the ECM platforms are tools combining the above-mentioned technologies in order to support content's entire life cycle and use it to create organization's knowledge.

Concurrently, due to the appearance of the technologies described, one can distinctly see other forms of information processing being included in the general concept of content. This relationship has been analyzed on the example of ECM definition published and updated by AIIM. The first AIIM definition published in 2005 [WIKI14] was: "Enterprise content management is the technology used to capture, manage, store, preserve, and deliver content and documents related to organizational processes". At that time, the focus was mainly on the electronic form of content, in particular electronic documents. The need to introduce records management, that is tracking the archiving and organization's sensitive information life cycle, forced widening the perception of the content in the context of information stored in paper documents (and managing their archives) as well as on other data storage media such as microfilms. In 2006 AIIM expanded its official definition with the following statement: "ECM tools and strategies allow the management of an organization's unstructured information, wherever that information exists". Finally in 2010 the previous statement was changed for: "ECM covers the management of information within the entire scope of an enterprise whether that information is in the form of a paper document, an electronic file, a database print stream, or even an email". The present version of the definition is quoted in chapter 2. Taking into account the hitherto discussion, this paper proposes a very broad understanding of the concept of content, namely as a collection of structured and unstructured data, information and knowledge available on data storage devices, that is databases records, documents, emails, messages sent through social media or sound and image recordings on traditional storage media, i.e. paper or microfilm. Metadata describing library attributes of the above-mentioned repertory (e.g. the authors, time of creation, versions, physical storage location in an organization, etc.) is an indis-

pensable element of the concept of content. Content management covers the content's entire life cycle, i.e. from the moment it is acquired or created through archiving to destruction.

ECM platforms operate on mainly unstructured content since structured content belongs to the field of ERP, CRM, BI and other industry-specific systems. The strategic goal of ECM platforms is to support the knowledge resources creation process in an organization.

The ECM platform functional components characteristics

According to the Gartner's analysts, ECM platform is a set of functions or applications used at each of the stages of content life cycle [GART12]. The components work together, however, they can be also used and sold separately. The components defined by the aforementioned authors are characterized as follows:

- **Document management** – its basic function is to manage the documents repository (including cataloguing and categorizing), indexing (for advanced search), version control, security rules and the other library services. The component also includes content replication between the repositories localized in different places within the information resources of an enterprise, including local repositories stored on the employees' personal computers or mobile devices.
- **Image-processing applications** are used to process images (e.g. electronic copies of paper documents) from acquiring, through cataloguing and archiving to using in the company circulation of information. The component has two functions:
  - it acquires the document, in this case it is scanning software and equipment, intelligent character recognition (OCR, HCR, ICR) and form processing technology (OMR); the function is realized by using native software or third party equipment and software components,
  - it can store the images of scanned documents in the repository as "just another" type of content which may be used in the process of document and task circulation.
- **Workflow/Business Process Management (BPM)** are functions dealing with the processes of the circulation of information and electronic documents, tasks assignment and tracking of their performance with parallel creation of audit trails for each of them. The minimum requirement is the circulation and accepting of the document according to the path. The more advanced tools are equipped with graphical interfaces to build workflow processes, which support also step-by-step and parallel processes. To describe business

processes many suppliers of ECM platforms support in their solutions script languages standards, e.g. BPEL. Thanks to this technology platform users can use specialized software to model organization's processes, which will be then transferred to and executed on the ECM platform.

- **Records management** refers to the realization of long-term policy of important information storage, irrespective of its form, be it electronic or traditional. The key function of these tools is to control the records life cycle from the time of their acquisition through to their planned disposal. What is important here are the storage places, access rights as well as a complete access history including documents stored only in non-electronic form. The minimum requirement is the control of business critical documents on the basis of record retention schedule (data retention in telecommunications is the policy of storing communication records, e.g. telephone calls, emails etc.). Legal requirements for this component are determined by the law of a given country. The most well-known legal regulations in this domain are: US Department of Defense Directive 5015.2-STD, Victorian Electronic Records Strategy (VERS) and Model Requirements for the Management of Electronic Records (MoReq2). MoReq was widely used in throughout the European Union.
- **Web Content Management (WCM)** is a set of functionalities supporting users in Web-published content management. The basic requirement for this component is that the content published must be stored in the organization's central repository. The component includes the functions of content creation by means of templates, distribution schedules and workflow with additional elements for change management. The assumption is made that the access to the above-mentioned functions is granted not only to www server administrators, but also to the platform's non-technical users. Managing the content published on intranet, extranet and external public websites has become a serious problem in big organizations, so this area should be supported by WCM component.
- **Social content** includes the functionalities of document sharing, collaboration on documents and knowledge sharing within organizational units and project teams. Creation of blogs and wiki knowledge databases as well as support for online instant messaging clients and teleconferencing is another working area required for this component. According to the Gartner's analysts it is the social content – including video files – that is the most dynamically developing category of new content in an enterprise. The name of the component has been changed, as compared to the previously published entry, from "document collaboration" to "social content" in order to include more recipients and new kinds of content.

- **Extended components** comprise of: digital asset management (DAM), document composition, e-forms, search content and analytics, email and information archiving, email management and packaged application integration (integration with office suites, i.e. Microsoft Office or Open Office).

ECM platform technical structure on the example of Alfresco

Alfresco platform is one of the most dynamically developing ECM-class products. Its structure has been presented here to characterize the technical architecture which is present in the products available on the market and used to meet the requirements of the aforementioned above functional components. Alfresco is one of the 22 most popular systems as ranked by the Gartner's analysts in Magic Quadrant for ECM. It is the only open-source product listed in the report. The Alfresco platform has been built since 2005 and is currently used by 1,300 enterprises in 180 countries **(**having important references in Poland**)**. The Alfresco developer (Alfresco Software, Inc.) estimates that their platform is used by 7 million business customers who process 4 billion documents [ALFR14]. The platform's popularity and advantage over its competition are the consequence of its open-source licensing model and the assumption that the main system components are based on open standards, i.e. Java 1.6, Java Content Repository API, Lucene Text Search Engine, JBPM and many more [SHCB09]. Alfresco is available for Windows, Linux and Mac. It can use open-source and commercial relational databases, i.e. MySQL, PostgreSQL, Oracle and MS SQL. The scalable architecture of the product is based on various application servers, such as JBoss Application Server or Apache Tomcat. The product is compatible with all of the most popular web browsers. Its users can also use a mobile client. Figure 2 presents the main technological components of the Alfresco platform. Below will be characterized the four most important ones: Repository, Share, Surf and the newest one – Activity.

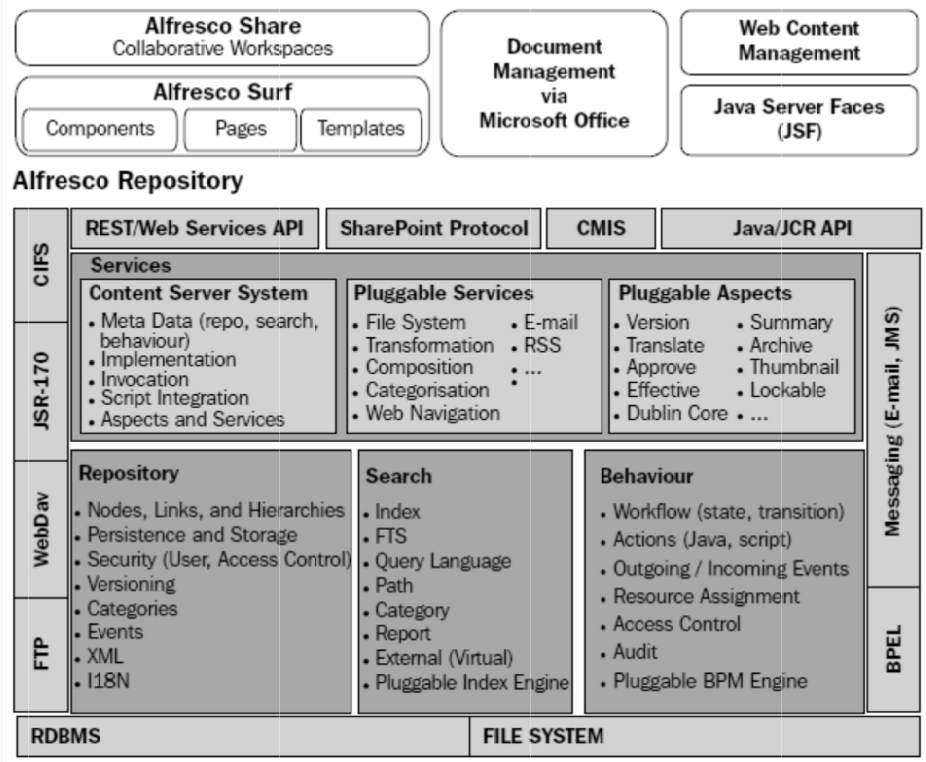

Fig. 2. Alfresco platform architecture

Source: [SHCB09].

Alfresco Repository is the basic component of the platform, responsible for content storage. Electronic content is stored simultaneously in two locations. The original content is stored in all kinds of multimedia files in a file system specially created for this purpose. Metadata describing the content is stored in a relational database (RDMS). Repositories provide content services for the other platform components, such as content storage or import, content classification, security on content objects, control through content check-in and check-out, and content query services to the content applications. The access control mechanism implemented in Alfresco covers the content, its attributes and modifications, and operates on a single object level as well as hierarchical catalogue structures (this solution combines the traditional approach used in RDMS and various file systems). The security mechanism also provides a full access audit for each of the stored objects. The Repository architects took into consideration the fact that it would be the central location for content storage for the whole en-

terprise, i.e. internal IT systems as well as external customers and partners of the organization. Figure 3 presents the central place taken by the component in the information structure of an organization. As far as content exchange between systems and repositories is concerned, Alfresco uses the latest open JSR-170 Java standard interface. The supplier presented was the first to introduce Microsoft Office Sharepoint Protocol support, which enabled the users to share and simultaneously work on MS Office documents stored in Alfresco Repository.

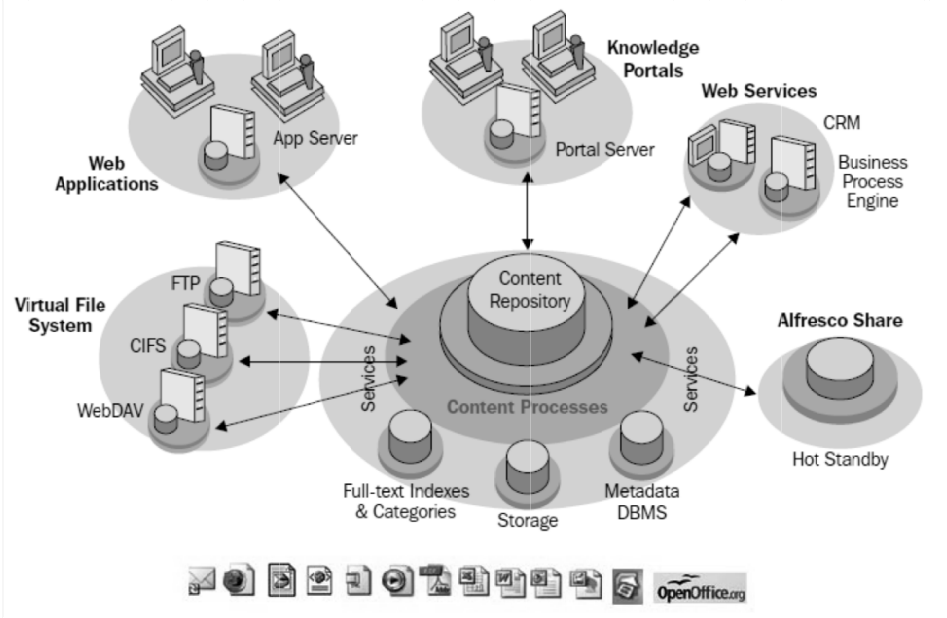

Fig 3. Diagram of Alfresco Repository relations to other systems

Source: [SHCB09].

Repository provides a complete collection of library services such as check-in and check-out, version control, auditing information, and streaking. User can define the library services to be executed automatically, based on business rules. The component is equipped with an efficient mechanism of full text indexing and categorizing based on Lucene 2.1 Text Search Engine technology. Indexing is performed automatically after each upload of any electronic document. The platform has internal OCR mechanism but it can also work with external tools like e.g. ABBYY.

Alfresco Share is a collaborative content management component. Alfresco Share simplifies the capturing, sharing, and retrieval of information across virtual (project or units) teams. The component allows to create project and topical

websites, calendars and to manage team tasks. Team documents are stored in Alfresco Repository. Share is equipped with a built-in Flash viewer, allowing users to view content regardless of the originating application or product version (for example Microsoft Office). The content can be tagged, and access to the tags can be shared. Team discussions can be created on team sites, documents or even topics. Content can also be provided via an RSS feed.

Alfresco Surf allows to create**,** manage and publish content on websites managed by the platform. Surf's interface is based on Web 2.0 AJAX* technology. The content created can be composed of plain text, HTML, XML files, multimedia files and specialized code for user interaction. The latest addition in this area is Alfresco Web Studio, a drag-and-drop visual editor meant to enable non-technical users to create websites. The content presented on the sites managed by Surf comes from Repository. The other parameters, i.e. site navigation, menu and templates are stored in RDMS.

Alfresco Activiti is a component for visual modeling processes, consistent with BPMN 2.0 notation, which are next executed by workflow engine built-in to the Repository. The tool provides modeling and execution of very complex processing of documents and tasks generated in other Alfresco platform components.

The description above presents only the most important components and technologies of Alfresco platform which is just a fraction of this extremely extensive tool. The components listed fully cover the afore-discussed requirement areas as regards ECM tools, mentioned by the Gartner's analysts.

Prospects for the ECM market and conclusion

The growing importance of ECM tools on the IT market can be indicated by two elements: the market's major players as well as its value and expected growth rate. Referring to the Gartner analyses [GART12], the ECM market's major players are companies like Microsoft (with SharePoint), Oracle, IBM or Xerox which only proves they see this segment's enormous potential. The dynamics of the market shows in the results of analyses carried out by a market research firm The Radicati Group, and presented in the report "Enterprise Content Management Market, 2013-2017" [RADI13]. According to the data presented there, in 2013 the global market for ECM systems was worth $5.1 billion. Analyses of historical data and sales forecasts of companies participating in the

* AJAX (Asynchronous JavaScript and XML) – Web development technique where the client − server interaction takes place without reloading the whole document, asynchronously.

survey show that in 2017 the market value will reach as much as $9.3 billion. These results allow to calculate that the ECM market will grow at an average annual rate of 16%. When compared to other areas of the IT market, this very high growth rate confirms the thesis that the ECM platforms will be a dominant technology implemented by the enterprises in the coming years.

References

[ADAM07] Adamczewski T.: Systemy zarządzania treścią. In: Strategie i modele gospodarki elektronicznej. C. Olszak, E. Ziemba (eds.). WN PWN, Warszawa 2007.

[AIIM14] AIIM: What Is Enterprise Content Management (ECM)? AIIM, http://www.aiim.org/what-is-ecm-enterprise-content-management, 2014 [access: 05.05.2014].

[ALFR14] Alfresco Software, Inc.: About Alfresco, www.alfresco.com, 2014 [access: 15.05.2014].

[ĆWIK03] Ćwiklicki M.: Koncepcja zarządzania elektronicznymi zasobami informacyjnymi w przedsiębiorstwie. Zeszyty Naukowe Akademii Ekonomicznej (obecnie Uniwersytetu Ekonomicznego) w Krakowie nr 616, Kraków 2003.

[GART12] Gartner Analysts: M. Gilbert, K. Shegda, K. Chin, G. Tay, H. Koehler Kruener: Magic Quadrant for Enterprise Content Management 2012, Gartner Inc., http://www.project-consult.de/files/Oracle_Gartner-Magic-Quadrant-ECM_20121018.pdf, 2012 [access: 05.05.2014].

[KAHA10] Kaplan M., Haenlein M.: Users of the World, Unite! The Challenges and Opportunities of Social Media. „Business Horizons" 2010, 53 (1), pp. 59-68.

[KAMP13] Kampffmeyer U.: EIM Enterprise Information Management. PROJECT CONSULT Unternehmensberatung GmbH, http://www.project-consult.de/files/EIM-KAMPFFMEYER-EN.pdf, Hamburg 2013 [access: 05.05.2014].

[MOLL05] Moller C.: ERP II a Conceptual Framework for Next-generation Enterprise Systems? „Journal of Enterprise Information Management" 2005, 18 (4), pp. 483-497.

[NOTA00] Nonaka I., Takeuchi H.: Kreowanie wiedzy w organizacji. Jak spółki japońskie dynamizują procesy innowacyjne. POLTEXT, Warszawa 2000.

[RADI13] Radicati S., Yamasaki T.: Enterprise Content Management Market 2013-2017. http://www.radicati.com/wp/wp-content/uploads/2013/05/Enterprise-Content-Management-Market-2013-2017-Executive-Summary.pdf [access: 09.05.2014].

[ROCK03] Rockley A., Kostur P., Manning S.: Managing Enterprise Content: A Unified Content Strategy. New Riders Publishing, 2003.

[SHCB09] Shariff M., Choudhary V., Bhandari A., Majmudar P.: Alfresco 3 Enterprise Content Management Implementation. Packt Publishing, Birmingham 2009.

[TRAB13] Trąbka J.: Specific Analytical Perspectives in the Modelling of Workflow Systems. Zeszyty Naukowe Uniwersytetu Ekonomicznego nr 153, Katowice 2013.

[WIKI14] Wikipedia, Enterprise content management: http://en.wikipedia.org/wiki/Enterprise_content_management [access: 10.05.2014].

[ZIEM07a] Ziemba E.: Technologie informatyczne w zarządzaniu wiedzą niejawną. In: Strategie i modele gospodarki elektronicznej. C. Olszak, E. Ziemba (eds.). WN PWN, Warszawa 2007.

[ZIEM07b] Ziemba E.: Portale korporacyjne. In: Strategie i modele gospodarki elektronicznej. C. Olszak, E. Ziemba (eds.). WN PWN, Warszawa 2007.

PLATFORMY ENTERPRISE CONTENT MANAGEMENT − AKTUALIZACJA KONCEPCJI, ROLA W ORGANIZACJI I GŁÓWNE TECHNOLOGIE

Streszczenie

Koncepcja zarządzania zasobami informacyjnymi przedsiębiorstwa (Enterprise Content Management, ECM), a także technologie z nią związane zmieniły się w okresie ostatnich 20 lat. Ta ewolucja i aktualna rola zarządzania zasobami informacyjnymi w przedsiębiorstwach jest przedstawiona w dwóch pierwszych rozdziałach artykułu. Trzeci rozdział jest poświęcony głównym komponentom i technologiom stosowanym dla tworzenia platform ECM. W artykule umieszczono krótką analizę światowego rynku systemów ECM i wskazano tendencje przyszłego rozwoju. Jest on wprowadzeniem do dyskusji na temat nowych narzędzi dla analizy systemów i projektowania, które mogą być użyte dla wdrażania platform ECM.