

The EWD-P System

Polish Government – Council of the European Union Interoperability Achieved

Grzegorz Bliźniuk^{+#}, Mariusz Momotko^{*}, Bartosz Nowicki^{*}, Jakub Strychowski^{*}

^{+) Military University of Technology, Warsaw, Poland}

^{#) Ministry of Scientific Research and Information Technology, Warsaw, Poland}
gblizniu@mni.gov.pl

^{*) Rodan Systems S.A., Warsaw, Poland}

{Mariusz.Momotko, Bartosz.Nowicki, Jakub.Strychowski}@rodan.pl

Abstract

The Polish government – Council of the European Union interoperability is an important factor of the Europe enlargement success. This article presents how the interoperability has been achieved and further exploited in the European Exchange of Documents – Poland (EWD-P) system – an application responsible for elaboration of the official Polish standpoint concerning adjustment of the newly associated states law to the EU procedures and regulations. The mechanisms and the role of the underlying platform for knowledge management intensive portals development are presented.

1. Introduction

The important ingredient of the European Union enlargement process is mapping and adjustment of the newly associated states law to the EU procedures and regulations. The process will continue even after accession, any time when a national standpoint on legislation issues will be required.

Elaboration of the official Polish standpoint is a complex and recurring process (more than 10 000 process instances a month). It requires involvement of pertinent units of central government (20 ministries), assurance of compliance with relevant documents and previous standpoints and collaborative effort of a group of selected experts from various domains (over a population of more than 12 000 civil servants, see Figure 1 for presentation of involved actors). The results delivery within hard time limits is critical because failure to deliver on time means unconditional acceptance of the original EU proposal.

The primary objective of the European Exchange of Documents – Poland system (EWD-P) is to assure smooth, effective and efficient support for the above presented process. This requires sophisticated classification of documents, their distribution and exhaustive cooperation of dedicated teams. The IT

support turned out to be necessary as previously a significant percent of standpoints could not be elaborated appropriately, hampering the overall integration process.

While the system mission seems to be straightforward, its architecture is challenging. First of all, a conceptual common denominator had to be found with respect to concepts' sets used by both sides (Poland and EU). The vocabularies, not unified yet, varied both in the sense of languages and in the sense of their categorization power. The ontology based approach exploiting the standard representation of topic map [1] turned out to be very effective. The system dynamics is organized around a workflow process which involves a significant number of activities, units, participants, documents and rules that precisely limit the process scope without compromising quality of the final outcome.

The presented solution is a web application and is the first commercial exploitation of the ICONS platform for knowledge management intensive portals development (developed within the ICONS project, IST-2001-32429, completed in April 2004, www.icons.rodan.pl).

2. The EWD-P System Application Domain

The documents from General Secretary of the Council of the European Union (referred further to as UE documents) are received from the U32Mail system (legacy system exploited also by other EU countries). An EU document consists of its envelope and a set of enclosed files of various types (PDF, HTML, Doc97). The envelope comprises of a set of attributes that characterise the nature, source and destination of the document. The files constitute the real content of the document. So far, Poland receives the EU documents in either English or French. Based on EU subject codes and other attributes included in the document envelope as well as in the document content, the EWD-P system classifies it and selects the most corresponding Polish subjects which this document refers to. In the next stages, the selected Polish

subjects are used to determine organisational units and experts that are responsible for processing of a given document.

For the majority of the EU documents a Polish response must be prepared and delivered by a certain date. If the deadline for preparation is not met, it is assumed that there is no objection to the original EU document and that the document is fully accepted. When the Polish response is prepared, it is sent to the Polish Representation Office at Brussels which in turn presents the document to the Council's pertinent teams.

From the information management point of view the 'case' concept plays the crucial role as it secures all information related to a given response (see Figure 2). This comprises the original EU document (triggering the corresponding case; perhaps in versions), response, related cases, external documents etc. The cases are maintained by cooperating experts who attach the consecutive versions of the Polish response as well as other relevant information (e.g. legislative documents). The cases store minutes of meetings on which Polish representatives present the Polish response (this includes remarks and comments).

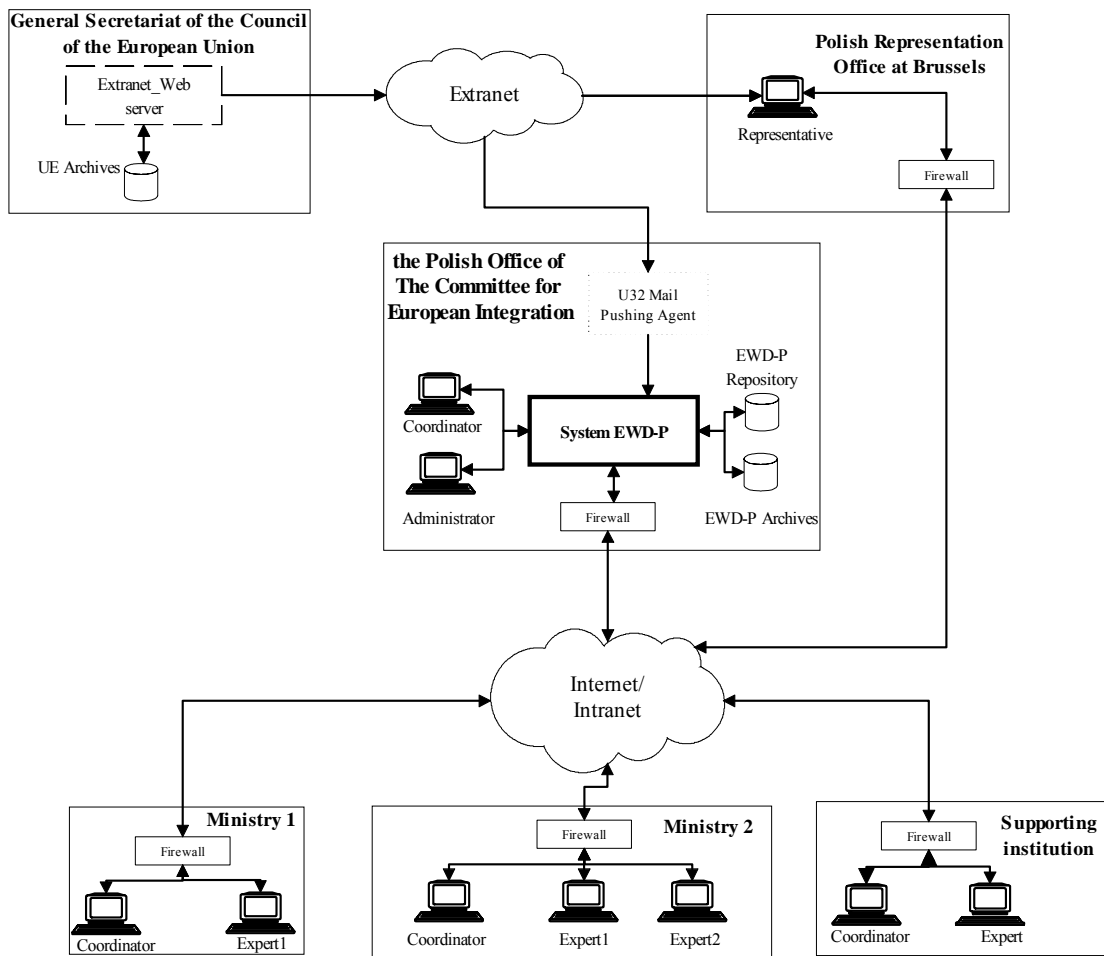


Figure 1. Actors of the EWD-P architecture.

It is critical to ensure consistency with other cases and, to the same extent, preserve the information on previously existing cases that somehow influenced a given response. A net of relations among cases facilitates navigation among related cases and promotes the access to global view on some issues rather than to its individual aspects.

The system dynamics is imposed by processes executed by the workflow engine – one process instance

for each EU document (partial process representation presented on Figure 3). The workflow starts from publication of a original document on the U32Mail gateway, continues through initial dispatching in the Committee for European Integration, precise dispatching on the level of particular ministries' units, establishing collaboration among the selected experts, elaborating of final standpoint and ends up with the final delivery to the Polish Representation Office at Brussels (Figure 1). The

premise is that knowledge management is to the same extent concerned with tacit, as with explicit knowledge, and with people as much as with information.

Technologically, the ICONS platform provides a well balanced set of innovative multimedia, process and content technologies implemented into fine grained and encapsulated modules (see Figure 4). Although the modules can be used separately, the highest effectiveness is achieved when modules cooperate in the

seamlessly integrated platform offering, due to their synergy, the outstanding capability. The detailed presentation of the platform exceeds the form of this article (more than 500 pages of technical documentation). In the consecutive sections we succinctly present the platform's modules that played the crucial role in the EWD-P system successful development.

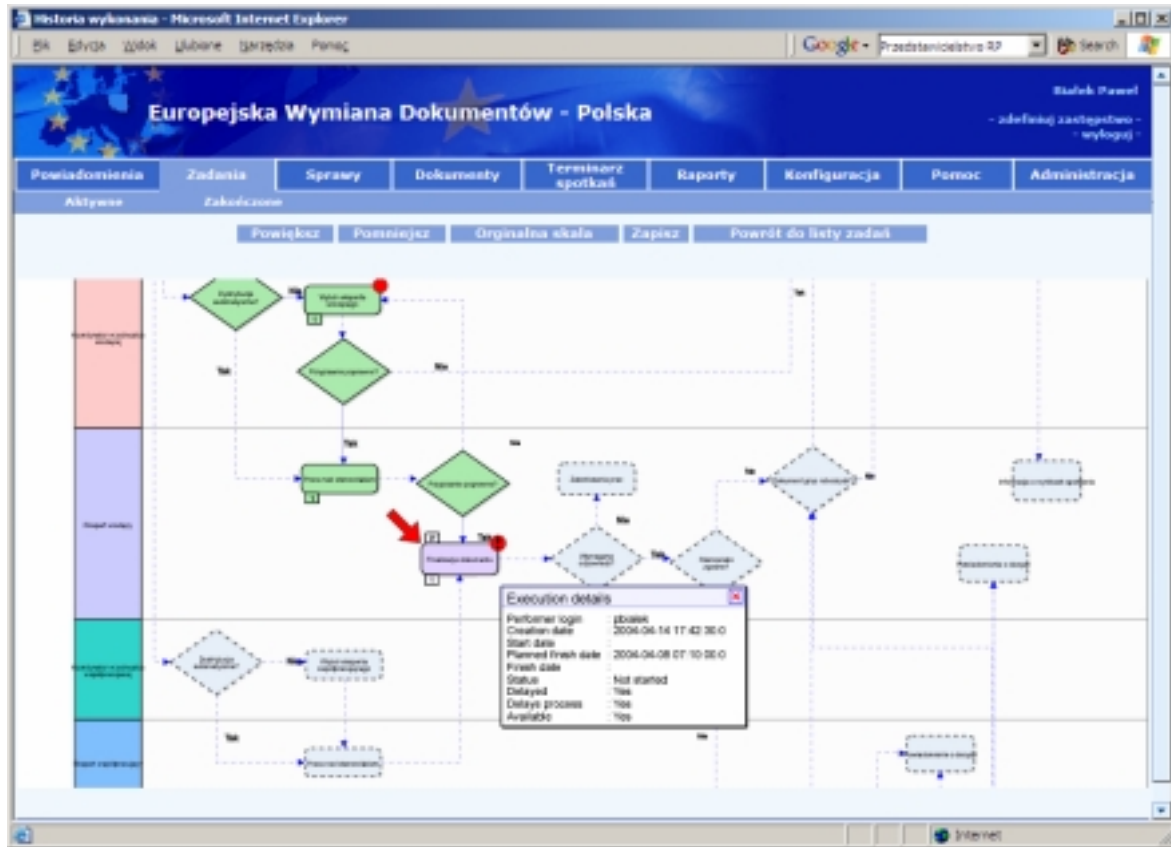


Figure 3. An example of the process execution visualisation (screen from the working system).

4. Ontology – a prerequisite for interoperability

Each incoming EU document is categorized along the predefined EU classification system. Categories are just multiple values in a given attribute of the document envelope. This set of categories determines receivers of the document in the Polish administration. The problem is that the Polish and EU classification systems are not unified yet and vary both in the sense of languages and in the sense of the categorization power. One EU category can be mapped to several Polish categories. Similarly, one Polish category can be mapped to several

EU categories. The approach applied in the EWD-P system is based on ontology implemented as the Topic Map [1]. The EWD-P ontology, constituting the logical heart of the system, provides mapping between inconsistent set of concepts.

An ontology is a catalogue of the types of things that are assumed to exist in a domain of interest D from the perspective of a person who uses a language L for the purpose of talking about D [4]. Ontology promotes systematic exploitation of knowledge representations and services through introducing precise and unambiguous meaning. Elaboration, publication and implementation of well defined vocabulary are of the first importance for effectiveness of any knowledge management systems and, consequently, organizations'

productivity. Each important concept is named and provided with a definition (on this level concept's semantics comes from interpretation of its name and definition). Additionally the concept meaning is refined by associations with other concepts. Types of these associations differ from general lexicographic (e.g. synonym, homonym) up to the application domain specific ones (e.g. compliant with a standard). Meaning of associations can be examined in a graphical navigator that presents the ontology in a form of an animated

graph. The level of ontology abstraction differs and can range from fully abstract entities up to names of concrete instances or typical dictionary positions.

The module responsible for the ontology management (called Concept Glossary Manager) fully implements the extremely flexible model of the Topic Maps standard [1]. This standardization offers a nice possibility to import and export any ontology defined in the XML Topic Map format.

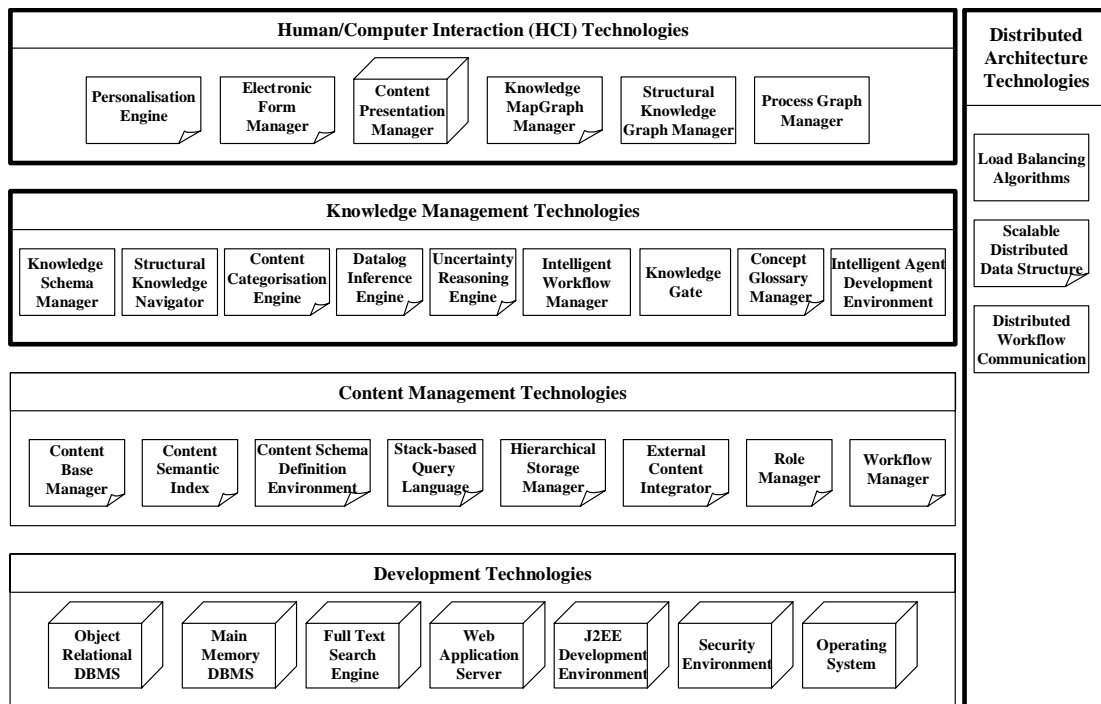


Figure 4. Abstract architecture of the ICONS platform.

4.1. The Topic Map Model

The Topic Map structure assumes a set of topics. Each topic describes single subject of the real world. Every topic has a set of its names. A topic name is a text which visibility and functionality can be defined by a topic map scope. Each topic has also its subject identity and a set of the subject indicators. The subject represented by a topic can be identified either by a reference to the resource that the topic represents (subject identity) or by reference to a resource that in some way describes the subject. Each topic can be typed by an arbitrary topic (this provides for very flexible knowledge representation model).

A topic can contain a set of occurrences which connect a topic to some information resource related to this topic. For instance, if a topic represents a document's category then that topic can have an occurrence that points

to a document belonging to this category. The inline occurrence stores information about topic in an explicit form of a short description. The inline occurrence is used as a concept definition or a type of meta-data about the concept. Every occurrence has its occurrence type. Occurrence type is specified by a selected topic.

A topic map structure is build out of associations. The association construct may include additional typing information which specifies the nature of the relationship among topics and also specifies what role each topic plays in the relationship.

In the EWD-P system document categories are represented by topics while mapping is imposed by associations of special type, linking the corresponding Polish and EU concepts. The EU and Polish categories are stored as topics of types "EU category" and "PL category", respectively. The categories' mapping is represented as associations between categories' topics (of the type "categories mapping").

The extraordinary flexibility of the topic map structure (every topic construct is a concept typed by other topics, recursively) is definitely too complex to be directly managed even by the experienced IT users. Therefore, some constraints were defined and imposed on the ontology in the user friendly interface. In particular, a special set of core topics (the ontology schema) was defined to represent types further denoting other concepts, associations and roles. The ontology editor reads the ontology schema and dynamically generates user interfaces (e.g. dynamically generated menus comprising all registered types) through which ontologists can only introduce knowledge compliant with the underlying schema. It frees the users from knowing the ontology schema and guarantees the ontology integrity.

4.2. Knowledge Representation and Topic Maps

Although the main objective of the EWD-P ontology is to manage the UE – Poland interoperability within the area of categories mapping, the ontology proved its effectiveness in tasks related to representation of knowledge of various nature [5] e.g. central government units organizational structure, experts and their competencies, hierarchical dictionaries.

The EWD-P system ontology manages information on each civil servant or government organizational unit (represented as topics) that are involved in the standpoint elaboration process. Complex relationships among experts, units and categories cover e.g. person and organizational unit (is employed in), person and category (is an expert in), organizational unit and category (is responsible for). The relationships, modelled as associations and exploited by the workflow engine, allow for precise selection of the most suitable (knowledgeable) experts and the responsible unit.

Finally, the EWD-P ontology, although serving various purposes (out of which interoperability is the most important), is internally consistent and play the vital role in the end users' education and communication.

5. Text Categorization

Despite the possibility to categorize documents that can appear without the envelope, text categorization is applied in the process of selection of the standpoint elaboration leader. As it was stated previously, each Polish category is assigned to the experts and organizational units that can be involved in the elaboration process. Although the workflow system "knows" every person that have to be involved in the process to achieve a complete view on a given problem, it is not decidable who bears the ultimate responsibility for both the standpoint validity as well as for the overall

elaboration process (i.e. who is the leading expert and leading unit).

The full text categorisation engine is employed to solve the problem of leaders' selection. An incoming EU document is categorised against the Polish classification system. As a result a set of Polish relevant categories (subjects) together with the associated trust level is delivered. The trust level represents the intensiveness of a given category in a particular document. The category having the highest trust level determines the leader.

The applied Text Categorization Engine involves an inductive process that automatically builds a classifier by learning from a set of previously classified documents and then uses the classifier to assign a predefined category to each new document. The quality of the categorization is guaranteed by the comprehensive document pre-processing (native format to pure text conversion, tokenization, recognition of national languages, stemming, removal of stops words, development of the document vector, information gain feature selection) in combination with well known approach to the classification process. The module implements a number of classification algorithms (Support Vector Machines, K-Nearest Neighbours based Model, K-Nearest Neighbours, Rocchio) of various effectiveness for different types of texts [6]. The selection of the optimal (with respect to the predefined metrics) algorithm is realized automatically during the learning phase.

The categorization recall and precision are at the level of 70% and they strongly depend on the classifier (classification model) quality which in turn depends on the training set quality. In the search of better quality the full text categorisation improvement sub-process is put into operation. Basically, it assumes the classifier reconstruction based on the improved training set (comprising e.g. cases that especially violate the common sense classification). If automatic classification fails (e.g. a leader refuses to take responsibility for a given standpoint claiming that it is not his / her area of interest), a coordinator (special user) manually selects the appropriate category and reconstructs the training set. New classifier is provided to the categorisation engine and the quality is improved.

6. Advances in Workflow Management

The implementation of the EWD-P dynamics is fully based on the workflow management technology. The workflow process (in the current version - comprising of 45 activities and involving 8 roles of executors) explicitly defined in the form of Conceptual Business Process Model [7] and XML Process Definition Language (XPDL) [8] is responsible for classification and distribution of the EU official documents as well as for

preparation of the Polish response. The process of elaboration of the Polish standpoint is instantiated for every incoming EU document and then executed by the workflow management system.

The major challenge addressed by the workflow engine in the EWD-P system was to optimize workflow processes executions in order to “to assure that appropriate activities will be performed by the right (knowledgeable) participants based on pertinent information in due time”. A process instance is perceived by the workflow participant through a task list in which all commitments of a given user (possibly involved in a numerous cases) are sorted along with some priorities. The key features that made the workflow engine successful within the EWD-P system are presented below.

Assignment of workflow participants is dynamic.

The workflow engine uses the rules of assignment of experts and organisational units to documents' categories according to the competence they have (carefully specified in the EWD-P ontology) or based on the role they already played in previous activities of a given process instance (e.g. let B will be carried out by the same person that carried out A). For instance, the Polish Ministry of Infrastructure is responsible for the ‘aviation’ category. Let’s further assume that there is one coordinator (responsible on the ministry level) and two candidate leading experts in this ministry. The system suggests to the coordinator these two candidates and allows selection of the most suitable one (tacit knowledge application).

The expression and valuation of the above described rules for workflow participant assignment is possible thanks to the Workflow Participant Assignment Language (WPAL) [9] being the performers-oriented specialization of OfficeObjects® Business Process Query Language (BPQL) [10]. WPAL is a functional language that enables workflow designers to define dynamic workflow participant assignments (just before activity commencement rather than static specified on the process definition level). The language enables to declare whether one, N or all participants selected by the WPAL rule are supposed to perform a given activity and if such assignment has to be done manually or automatically. Owing to WPAL it was possible to express complex workflow participant assignments rules straight in the process definition rather than in the application code, what makes the application more maintainable.

Communication among process participants is flexible. Usually, the coordinator monitors progress of elaboration of the Polish standpoint and, if necessary, gives some hints and feedback to the experts involved to improve the quality. This communication pattern is difficult to express in the form of the traditional workflow process, since there is no algorithm defining when this communication occur and in which way it will be carried

out. Thus, to complement strict rules of the traditional workflow processes, a mechanism of team collaboration management (TCM) has been introduced. This is important as especially teams creating new knowledge need more elastic forms of cooperation with more space left for innovation, creativity and spontaneity. Messages of various natures are grouped in thematic threads and moderated by the coordinator. Integration between workflow process and TCM assumes 2 dimensions: time and participants. The discussion forum is active from the elaboration process commencement up to the process termination (read only mode is available then). As participants are identified dynamically during the process enactment on the base of WPAL rules valuation, they are joined to the TCM as soon as they are selected by the workflow system (starting from the process owner during the process initialization). Since this moment he / she can read messages, start new threads and reply to messages of others. The discussion-forum-like facility fosters unconstrained collaboration of the process participants, what is really indispensable while complex, interdisciplinary and multidimensional problems are addressed.

As Polish standpoints have to be prepared before some deadline, **time management** is of particular importance. If a standpoint is not prepared on time, the delivered EU document is accepted with no objection. Therefore, it is crucial to detect and signalise any delay in the process as soon as it occurs both on individual activity level as well as on the overall process level. To meet this requirement the ICONS workflow engine extends and implements time management algorithm proposed in [11]. This algorithm checks the deadline and duration constraints for the whole process and for its individual activities. It also determines the best and worst cases for these constraints taking into consideration different possible scenarios (paths) of the process execution. The predefined workflow participants (especially coordinators) are informed on delays. Again, to select the optimal group of notified participants, WPAL is used to fully explore the execution context (e.g. notify process owner i.e. a person from the delayed activity department who carried out the first activity belonging to the process).

To mitigate the problems following from the standpoints elaboration process complexity (a lot of activities, a significant number of spread geographically organizational units and thousands of potential participants) a facility of **process execution visualisation** in a human-readable form is used (see Figure 3). This graphical representation of the process instance (on the contrary to the visualisation of the process definition) is an extension of the Business Process Modelling Notation [12]. This visualisation allows performers to better understand the process history (what was done before, by whom, what were the recommendations, what were the

time constraints), presence (what its current state is, what are the requirements for the current activity) and future (who will continue the process, what are potential consequences of current decisions) [13]. As the work items (picked up from individuals' tasks list) are put into definitely broader context, it positively impacts participants' knowledge and consequently improves productivity and quality.

7. Summary

At the time being (June 2004) the EWD-P system has been successfully deployed in the Polish Office of the Committee for European Integration and 3 ministries (almost 1000 of registered users). The highly productive workflow-based environment ensures that the processes execution times vary from 2 to 14 days while all outcome data are secured. As this pilot application proved its viability and has been thoroughly accepted by the end users, the remaining ministries (altogether 20 ministries and almost 12000 users) will be covered incrementally, to complete the whole process by the end of this year.

The EWD-P system implementation and deployment in such strict time regime (the first increment December 2003 – March 2004) was possible only due to the fact that it did not start from the scratch but rather built upon generic and powerful services provided by the ICONS platform for knowledge management intensive portals development (constructed within the ICONS project, IST-2001-32429, completed in April 2004). At the moment the platform is under industrialization, will be incorporated into the OfficeObjects® family of products and will become a core of even more sophisticated and challenging applications.

8. References

- [1] ISO/IEC, Topic Maps, December, 1999
- [2] ISO 15489, Information and documentation "Records Management", ISO 15489-1:2001(E)
- [3] The IST-2001-32429 ICONS Consortium, The ICONS Software Technical Design Manual, www.icons.rodan.pl, August 2003
- [4] Sowa, J.F., Knowledge Representation: Logical, Philosophical, and Computational Foundations, Brooks/Cole Publishing Co., Pacific Grove, CA, 2000.
- [5] Park J., Hunting S. (eds), XML Topic Maps: creating and using topic maps for the Web, Pearson Education Inc. 2003
- [6] Fabrizio S., Machine learning in automated text categorization, 2002, ACM Computing Surveys, Vol. 34, No. 1, March 2002.
- [7] iGrafx, Process 2000 User's Guide, 2000
- [8] Workflow Management Coalition, Workflow Process Definition Interface -- XML Process Definition Language (XPDL), issue 1.1, October 2002.
- [9] Momotko, M., Subieta, K., Dynamic change of Workflow Participant Assignment, ADBIS'2002, Bratislava, Slovakia.
- [10] Momotko, M., Subieta, K., Business Process Query Language – a Way to Make Workflow Processes More Flexible, accepted to Advances in Database Information Systems, ADBIS'2004, Budapest, 2004.
- [11] Eder, J., Paganos, E., Managing Time in Workflow Systems, in Workflow Handbook 2001, Layna Fischer (Ed.), Future Strategies Inc., 2001, USA.
- [12] Business Process Management Initiative, Business Process Modeling Notation (BPMN), version 1.0, May 2004.
- [13] Momotko M., Nowicki, B., Visualisation of (Distributed) Process Execution based on BPMN, submitted to 14th Conference on Web Based Collaboration, Prague, Czech, 2003.