

The EWD-P system. Polish government – European Commission interoperability achieved.

**Mariusz Momotko, Bartosz Nowicki, Jakub Strychowski
Rodan Systems S.A.**

Keywords: Europe enlargement, interoperability, ontology, topic maps, workflow management, text categorization, knowledge management

1. Abstract

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

2. 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 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). 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 management of sophisticated classification, distribution and cooperation patterns. 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 as sets of concepts of both sides (Polish and EU), not unified yet, varied both in the sense of languages and in the sense of the categorization power. The ontology based approach exploiting the standard representation of topic map [ISO/IEC 13250] turned out to be very effective and elastic. The system dynamics is realized by a workflow process which comprises a lot of activities, units, participants, documents and significant number of rules that precisely limit the process scope without compromising the final quality.

The presented solution is a web application and is the first commercial exploitation of the ICONS platform for knowledge management intensive portals development (constructed within the ICONS project, IST-2001-32429, to be completed April 2004).

3. The EWD-P system Application Domain

The EU documents are received from the U32Mail system (legacy system exploited also by other countries). Every received document is registered under a unique number. An EU document consists of the envelope and a set of enclosed files of various types (PDF, HTML, Doc97). The envelope comprises of a set of mandatory and optional 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 by selecting 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.

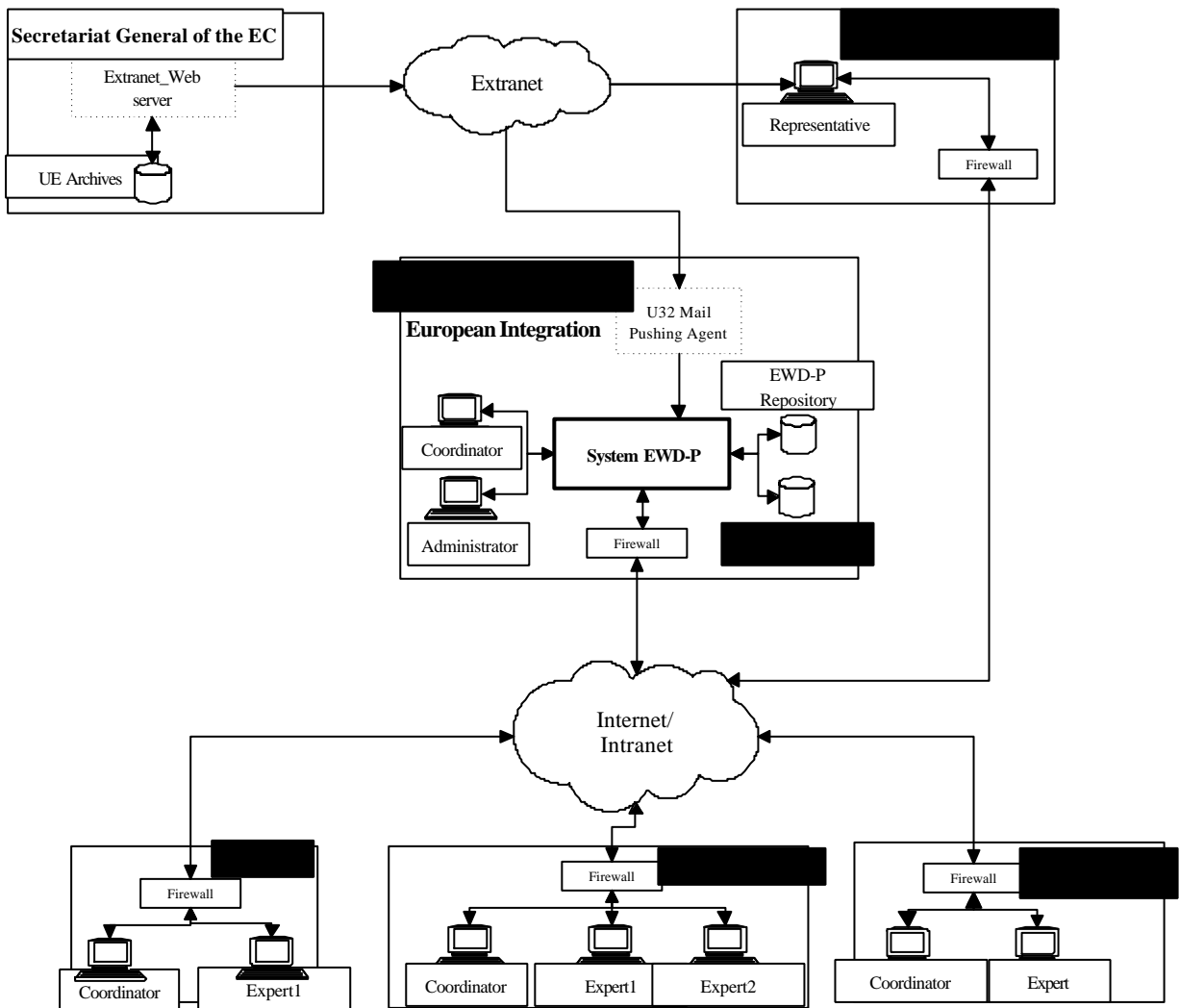


Figure 1. Actors of the EWD-P architecture.

Besides its mission critical functionalities the EWD-P system serves a number of auxiliary services. First of all the users can search over the repository of documents and cases using attribute based search (supported by user friendly search criteria builder) and full text search. This covers also European documents repository, access to which is provided by the U32 gateway. This promotes knowledge reuse based on previous experiences. Secondly, the system provides full accountability by storing the information on contributions of individuals to documents and cases as well as their behaviours in a concrete workflow process instance and discussion forum. Due to this, in the case of any problem the standpoint development history can be examined on the low level of details. The complete information on cases (together with behavioural aspects of contributors) is archived (according to the stringent central government practices and standards [ISO 15489]) – accessible for search and retrieval but closed for any modifications.

4. The Underlying Platform

Development of any system of the EWD-P size and complexity is always challenging, not to say risky, task. First of all, the end users (on various levels of authority, starting from “business” owners ending up with every-day operators) issues must be coped with. It covers elicitation of valid requirements, process’s reengineering, trainings, deployment and support. Secondly, a bulk of technical details must be designed and implemented within the specified quality and resource constraints. In the EWD-P case the latter source of risk was significantly mitigated as the development was based on generic and powerful services of the ICONS platform for knowledge management intensive portals development [ICONS D21]. One of the crucial points was to map application relevant services to generic services of the platform and fill the gap between what is generic and what is specific. Nevertheless, we hardly believe that the development would be possible without such kind of a workflow / content / information platform and, important to the same extent, experiences and competencies gained during the platform development.

The ICONS platform addresses the most essential needs of knowledge workers in the knowledge intensive organisations. Information redundancy and lack of pertinent information, flood of uncategorised documents, poor integration with external sources of valuable information, lack of uniform, user-friendly and personalised interfaces, huge amount of various stimuli overwhelming direct communication of individuals and groups, low level of exiting knowledge reuse, flat learning curve, etc. that is to mention just few symptoms hampering contemporary knowledge organisations. The primary objective of the ICONS platform based applications is to create conditions, in which “the right person carries out the right task at the right time based on the pertinent and comprehensive information”. However, the focus is not only on content and processes. The ICONS underlying 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 and content technologies implemented into fine grained and encapsulated modules (see Figure 3). 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 with the special stress on modules of the first importance for the EWD-P system.

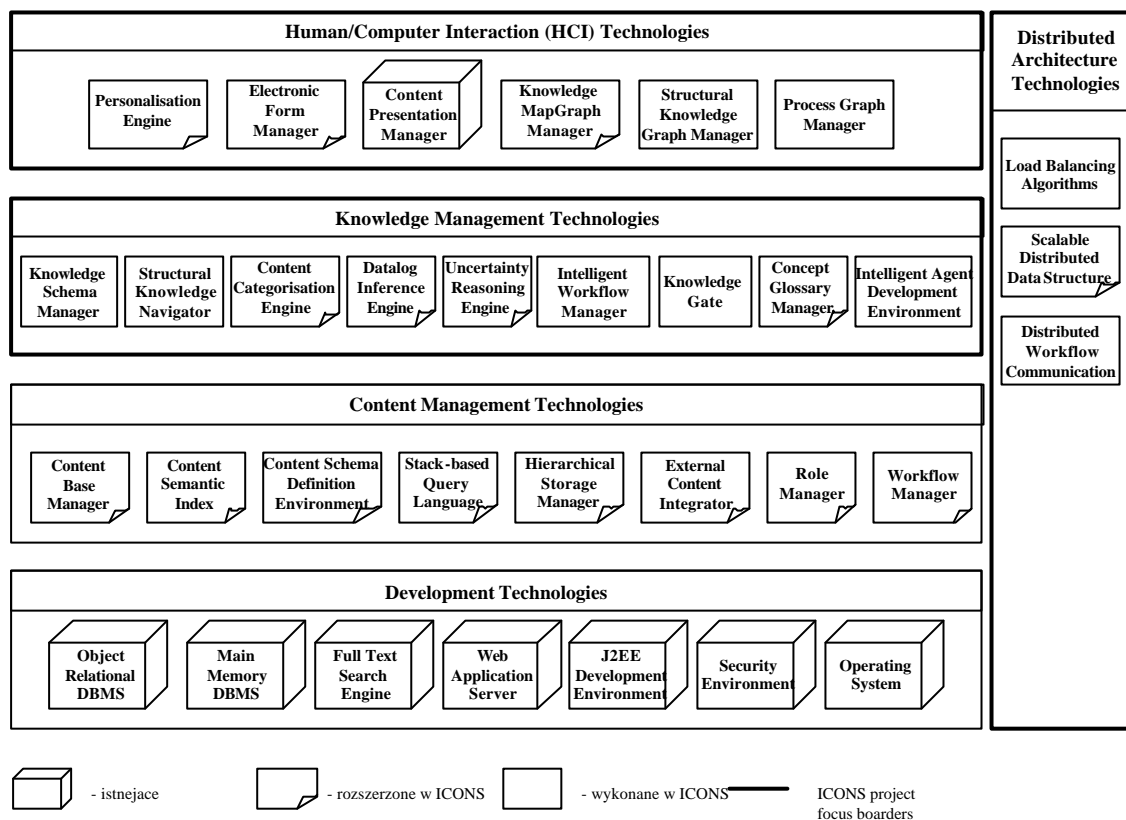


Figure 3. Abstract architecture of the ICONS platform.

5. Ontology – a prerequisite for interoperability

Each incoming EU document is categorized along with the predefined EU classification system. Categories are just multiple values in a given attribute of the document envelope. This set of categories should be used to determine receivers of the document in the Polish administration. The problem is that 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 Topic Map [ISO/IEC 13250]. The EWD-P ontology, constituting the logical heart of the system, provides mapping between inconsistent set of concepts. The mapping is explicitly defined in the system in the form of declarative knowledge.

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 [SOWA 2000]. Ontology promotes systematic exploitation of knowledge representations and services though introducing precision and unambiguity. Elaboration, publication and implementation of well defined vocabulary are of the first

importance for effectiveness of knowledge management systems and, consequently, organizations' productivity. Each important concept is named and defined (on this level concept's semantics comes from interpretation of its name and definition). Additionally the concept meaning is increased by associations with other concepts. Types of these associations differ from general lexicographic (e.g. synonym, homonym) up to application domain specific ones (e.g. compliant with standard). Meaning of associations can be examined in a graphical navigator that presents the ontology in a form of animated graph. The level of ontology abstraction differs and can range from fully abstract entities up to names of concrete instances or dictionary positions.

The module responsible for the ontology management (called Concept Glossary manager) fully implements extremely flexible model presented in the Topic Maps standard [ISO/IEC 13250]. This standardization offers a nice possibility to import and export ontologies defined in the XML Topic Map.

1.1 The Topic Map Model

The Topic Map structure contains a set of topics. Each **topic** describes single subject in the real world. Every topic has set of the 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.

A topic can contain a set of **occurrences** which connect a topic to some information resource related to this topic. E.g. if a topic represents a documents' 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 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 as topics while mapping is imposed by associations of special type among Polish concepts and corresponding EU concepts. EU and Polish categories are stored as topics of types "EU category" and "PL category", respectively. Categories mapping is represented as associations between categories' topics (association type is "categories mapping").

The extraordinary flexibility of the topic map structure (everything is a concept typed by the other concepts, recursively) is definitely too complex to be directly managed even by experienced IT users. Therefore, some constraints were defined and imposed on the ontology in the user friendly interface. In particular, a special set of topics (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 through which ontologists can only introduce knowledge compliant with

the underlying premises. It liberates the users from knowing the ontology schema and guarantees the ontology integrity.

1.2 Knowledge Representation and Topic Maps

Although the main objective of the EWD-P ontology is to manage UE – Poland interoperability by categories mapping, the ontology proved its effectiveness in tasks related to representation of knowledge of various nature [Park2003] e.g. organizational structure of central government units, experts and their competencies, hierarchical dictionaries.

The EWD-P system ontology manages information of each person or governmental organizational unit (represented as topics), which 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 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.

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

6. Text categorization

Despite the possibility to categorize documents that can potentially come without the envelope, text categorization is applied in the process of the standpoint elaboration leader selection. As it was stated previously, each Polish category is assigned to experts and organizational units that will 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 responsibility for both the standpoint validity as well as for the overall elaboration process (leading expert and leading unit).

The full text categorisation engine is employed to solve 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 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 preprocessing (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 Neighbors based Model, K-Nearest Neighbors, Rocchio) of various effectiveness for different type of texts

[Sebastiani2002]. 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 it strongly depends 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 subprocess 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), a coordinator (special user) reconstructs the training set using friendly user interface. New classifier is provided to the categorisation engine and the quality is improved.

7. Advances in Workflow Management

The EWD-P dynamics is fully based on the workflow management approach. The workflow process (in the current version - comprising of 45 activities and involving 8 roles of executors) explicitly defined in XML Process Definition Language (XPDL) [WfMC-TC-1025] is responsible for classification and distribution of 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 system 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 participants based on pertinent information in appropriate order at due time”. A process instance is perceived by the workflow participant through a task list in which all commitments of a given user are sorted along with some priorities. The key feature that made the workflow engine successful within the EWD-P system are presented below.

Assignment of workflow participants is dynamic and based on the selected coordinators, experts etc. This selection is based on assignment of experts and organisational units to documents' categories according to the competence they have. For instance, the Polish Ministry of Infrastructure is responsible for the 'aviation' category. Lets further assume that in this ministry there is one coordinator (responsible on the ministry level) and two candidates for the leading expert. The system suggests to the coordinator these two candidates and allows selecting of the most suitable (tacit knowledge application). In fact, the coordinator can even select a person not present on the short list generated by the system.

The expression and valuation of the above described rules for workflow participant assignment is possible thanks to the Workflow Participant Assignment Language (WPAL) [Momotko2002]. WPAL is a functional language that enables workflow designers to define dynamic workflow participant assignments. It is also possible to declare whether one or all participants are supposed to perform a given activity (quantity) and if such assignment has to be done manually or automatically. Owing to WPAL it was possible to express complex workflow participant assignment straight in the process definition.

Communication between individual process participants is also dynamic. Usually, the coordinator monitors elaboration of the Polish standpoint and, if necessary, gives some hints and feedback for the experts involved to improve the quality. This communication pattern may be hardly expressed in workflow processes, since there is no algorithm defining when this communication occur and in which way it will be carried out. To complement strict rules of the traditional workflow processes, discussion forums are 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 nature are grouped in thematic threads. Integration between workflow process and discussion forum assumes 2 dimensions: time and participants. The discussion forum is active from the process commencing up to its finalisation (then read only mode is available). As participants are identified dynamically during process enactment on the base of WPAL rules valuation, the participants are joined to the forum 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.

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 objections. 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 the overall process level. To meet this requirement the ICONS workflow engine implements time management algorithm proposed in [Eder2001]. This algorithm checks deadline and duration constraints for the whole process and 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 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).

As the EWD-P workflow is really complex (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 4). This graphical representation of the process instance is an extension of Business Process Modelling Notation [BPMN2003]. Such 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) [Momotko2003]. As the work context is definitely increased it positively impacts participants' knowledge and consequently productivity and quality.

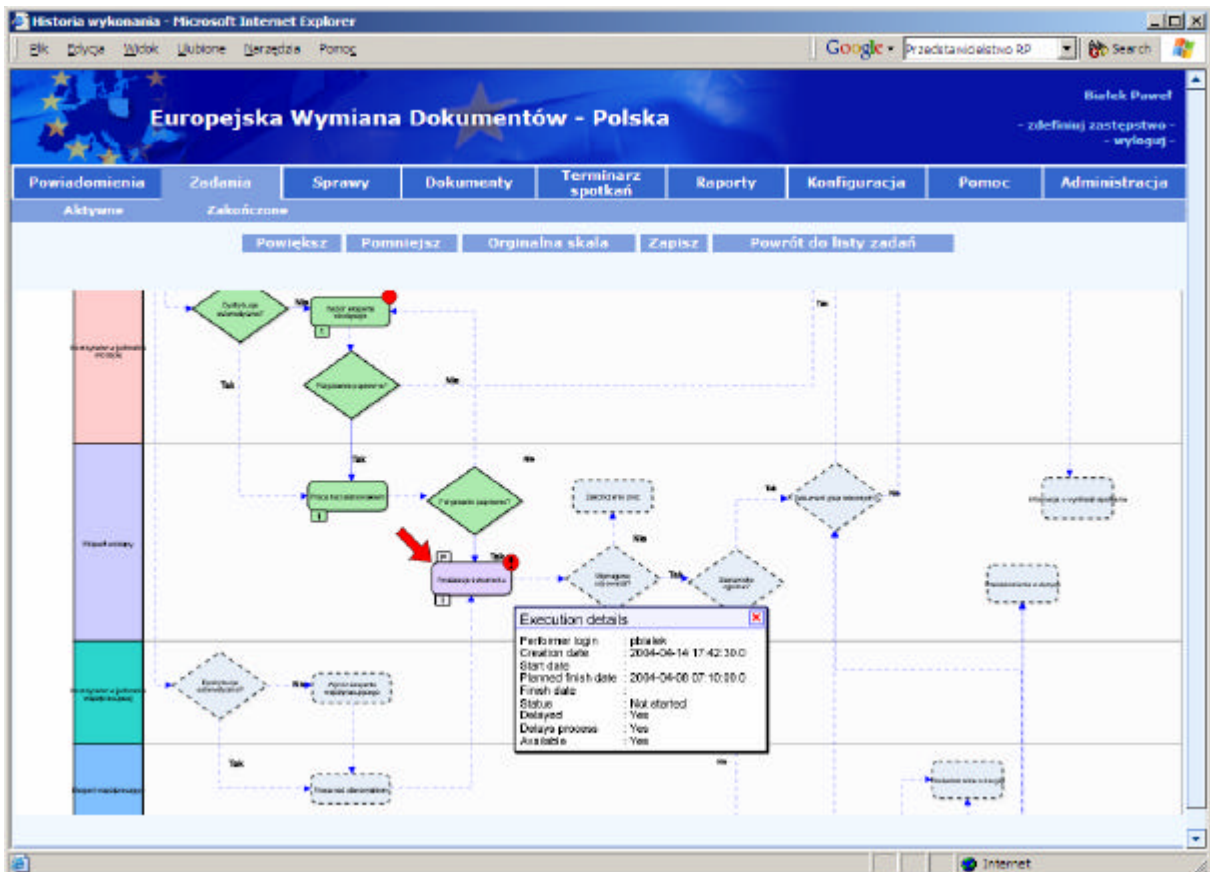


Figure 4. A concrete process execution visualisation (screen for the working system).

8. Summary

At the time being (April 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 environment ensures that the processes execution times vary from 2 to 14 days while all pertinent information 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. Thus we hope to present more exploitation lessons learned during the conference presentation.

The 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, to be completed 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 systems.

9. Bibliography

- BPMN2003 Business Process Management Initiative, Business Process Modeling Notation, working draft, version 1.0, August 2003.
- Eder2001 Eder, J., Paganos, E., Managing Time in Workflow Systems, in Workflow Handbook 2001, Layna Fischer (Ed.), Future Strategies Inc., 2001, USA.
- ICONS D21 The IST-2001-32429 ICONS Consortium, The ICONS Software Technical Design Manual, www.icons.rodan.pl, August 2003
- ISO 15489 ISO 15489, Information and documentation “Records Management”, ISO 15489-1:2001(E)
- ISO/IEC 13250 ISO/IEC, Topic Maps, December, 1999
- Momotko2002 M.Momotko, K.Subieta. [Dynamic change of Workflow Participant Assignment](#), ADBIS’2002, Bratislava, Slovakia.
- Momotko2003 Momotko M., Nowicki, B., Visualisation of (Distributed) Process Execution based on BPMN, submitted to 14th Conference on Web Based Collaboration, Prague, Czech, 2003.
- Park2003 Park J., Hunting S. (eds), XML Topic Maps: creating and using topic maps for the Web, Pearson Education Inc. 2003
- Sebastiani2002 Sebastiani Fabrizio, Machine learning in automated text categorization, 2002, ACM Computing Surveys, Vol. 34, No. 1, March 2002.
- Sowa2000 Sowa, J.F., Knowledge Representation: Logical, Philosophical, and Computational Foundations, Brooks/Cole Publishing Co., Pacific Grove, CA, 2000.
- TC-1003 WorkFlow Management Coalition, The Workflow Reference Model, TC-1003, issue 1.1, January 1995.
- TC-1025 WorkFlow Management Coalition, Workflow Process Definition Interface -- XML Process Definition Language (XPDL), issue 1.1, October 2002.