

DOCUMENT MANAGEMENT SYSTEM (DMS) FOR THE TECHNICAL REPORTS OF THE GEOLOGIC INSTITUTE OF CATALONIA/CARTOGRAPHIC INSTITUTE OF CATALONIA

Jorge Fleta; Carles Serra; Joan August Van Eeckhout; Anna Lleopart; Eulàlia Pujal; Antoni Roca and Enric Palma

Institut Cartogràfic de Catalunya (ICC). Parc de Montjuïc, s/n 08038 Barcelona. jfleta@icc.es

KEY WORDS: *document management system, technical reports, knowledge base, Catalonia.*

FOREWORD

The Geology unit of the ICC, previously Geological Survey of Catalonia, has generated a great number of technical information and documentation throughout its more than 25 years of history, the public services that it provides, the projects of investigation and development, in its administrative and organizational management, among others. This set of information, kept mainly stored in book format of written reports of different formats.

As of the year 2000 it has begun to generate a certain number of information in digital format Adobe PDF. In the year 2004 was decided to digitize the technical report of the Geology unit with the objective to guarantee the preservation of the information and to facilitate the search and recovery. So, in this way the geologic file would be the test pilot in the introduction of the document management system (DMS) of the ICC.

ANTECEDENTS

From year 1979 to the year the 2003, the inventory of data was located listing only in a local node of the network, the paper copies were stored in the file. Thus the metadata were only compiled.

Between the years the 2003 and 2005 inventory of information is made by means of Microsoft Access application InfoGeol (López-Arenas and Fleta, 2003). This application is located in a server node and they have certain functionalities (extraction of tables, georeferentiation, reports). It is begun to produce digital files the habitual way and the copies in analogical support are stored in the archives. This way one has separately the metadata and data information.

As of year 2005 the ICC DMS project with the Library of the technical reports of the archive of the Geology unit (DOCSGEOL) starts. This project includes the scanning process of about 2400 reports and the implementation of the DMS. In absence of the digital certification of the files, the documents in original format paper previous to

year 2005 with signature are conserved in safekeeping boxes on restricted access. So, the production of files passes to be digital totally, in this way, for the first time the metadata and the data are together in the same interface.

The equipment characteristics in the scanning stage includes a network server with Oracle database and a node for indexing with printer for the bar codes generation and three scanner stations with different resolutions and different dimensions (colour DinA0 of 400-800dpi, black/white DinA3 and colour DinA3).

DMS PROJECT

The implemented technology for the ICC DMS is Hummingbird Enterprise DM (Webtop and Windows Extensions modules) that delivers a consistent user interface and secure, scalable client/server architecture (Fig. 1).

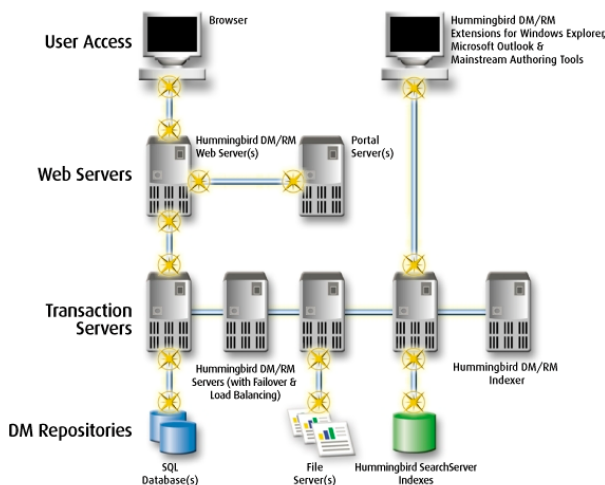


Figure 1 – Scalable architecture of Hummingbird Enterprise DM.

The following checklist enumerates some of the many features found in the DMS (Fig. 2):

- Drag and drop documents with a specific profile (metadata).
- Full-text indexes enabling custom search options against both profile metadata and document contents.
- Version information and control access.
- Document lifecycle (retention and availability).

- Configurable permissions for role-based activities and security policies.
- Ability to preview document via web in native format.
- Workflow management.



Figure 2 – Hummingbird DM concept.

ICC DMS requirements:

- Format files and information structures must be open, standard and well-documented (now only the PDF format is perdurable for 30 years).
- Hardware and software are open, standard and its fitted to the ICC DMS.
- DOCSGEOL library requirements:
- Possibility of recovering analogical documents until having certification of files.
- End-repository of documents, reports, maps and related databases.
- Control of document versions is not required.
- Documents always will be active.

Normally, the reports performed inside the projects have 3 steps well-distinguished, raw data, elaborated data and final report and associated databases. It is decided to implement Step 2 and 3 in the DOCSGEOL library project (Table 1).

Step 1	Raw data	high cost	high volume data
Step 2	Elaborated data	middle cost	intermediate volume data
Step 3	Report and associated databases	low cost	little volume data

Table 1 - Performing process report.

The ICC DMS includes an Oracle database, the drag and drop forms and the searching full text query syntax forms (Fig. 3). Also we have a set of tools to scan and generate the title pages and bar codes, management of scanned lots by resolution

and dimensions as well as the searching location in its safekeeping boxes.

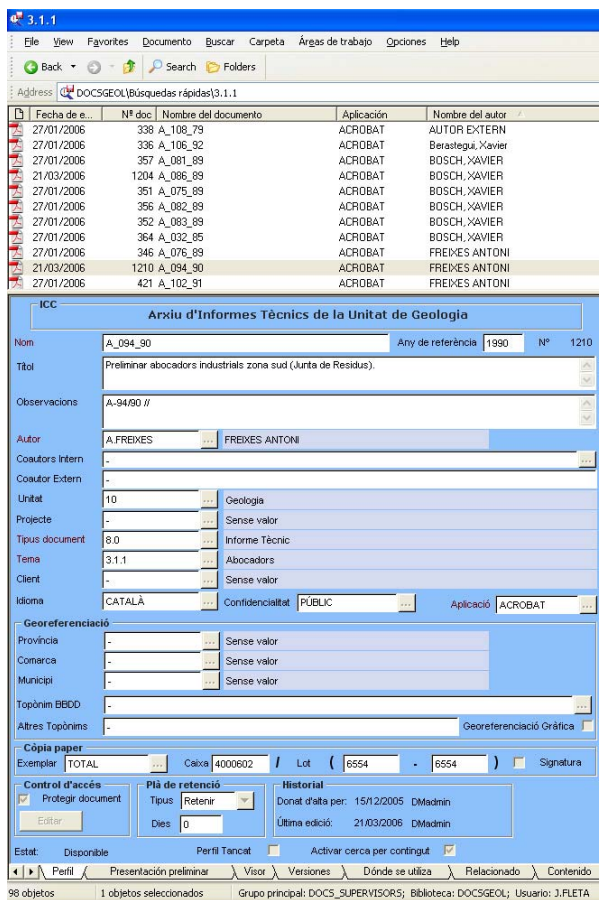


Figure 3 – Standard profile searching screen that combine simple query syntax (ex.98 identified reports searching for theme 3.1.1. Waste dumps). You can see at the bottom part of figure the coordinates of scanned pack (Caixa / Lot).

The server functionalities include the storage of documents and the metadata outlined, and the security policy as well as the lifecycle service of documents. The security policy is based on different restricted user groups: administrators, members of the Geologic users group, select members of other units and ICC users.

STATE-OF-THE ART

As of now two massive document loads have been made, 2400 reports. Very soon, we will be inserting about 1100 technical reports dwells. Part of the document repository includes in addition files and associated data bases. Thus most of the file has OCR capacity except for the maps and documents of bigger format.

DMS server contains the following numbers:

- Scanning process previous related to information previous to year 2005: 2392 scanned reports, 92446 scanned pages, 368 scanned maps of bigger format,

29918 coloured pages, 60700 b/w pages, 2196 scanned bigger format and 5959 scanned lots.

- Inventory later than 2005: 1059 technical reports.
- Number in the repository is of 3451 technical reports, March 2006.
- Mainly, it corresponds a three different typologies: Adobe PDF (technical report), compressed file (technical report and associated databases) and compressed file (interactive CD-Rom).

The geologic repository can be considered as a first step to a geologic knowledge base with different thematic contents (Table 2): geologic cartography information, geology of urban areas, geotechnic data, geophysics data, geologic hazard, geoheritage, hydrogeology, geology and health.

Thematic type	03 06	%
Garbage dumps	108	3,1
Flash floods	20	0,6
Snow avalanches	65	1,9
Cemeteries	230	6,7
Geology dissemination	7	0,2
Applied geophysics	231	6,7
Geology	247	7,2
Geotechnical	142	4,1
Geothermal energy	8	0,2
Hydrogeology	163	4,7
Interest for Geological Survey of Catalonia	10	0,3
Quarries	114	3,3
Risk and engineering geology	1272	36,9
Seismology	203	5,9
Geologic heritage	324	9,4
Geologic maps	158	4,6
Geologic cartography	112	3,2
Investigation licences	37	1,1
Total	3451	100,0

Table 2 – Number of technical reports by thematic contents.

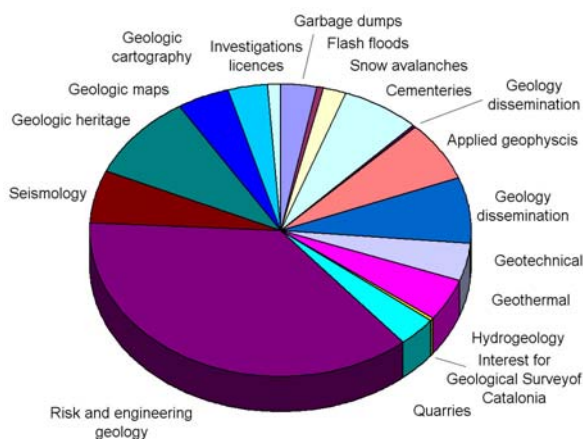


Figure 4 – Geologic repository distributed by thematic contents.

Another work-line in implementation phase is the spatial analysis within the DMS. From the georeferentiated explicit and graphical informations, within clients based in technology of geographic information systems, will be able to be

made searches outposts within the system (Fig. 5).

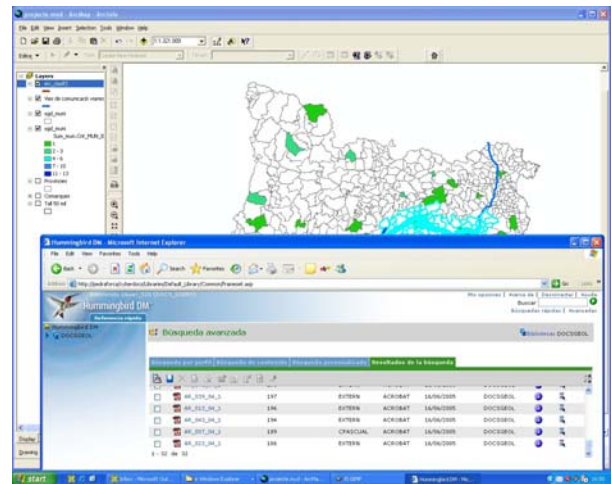


Figure 5 – Spatial query analysis linked to DMS output screen.

CONCLUSIONS

The phase of implementation of a document management system of the technical information of the geologic file has been completed that constitutes the embryo of the corporative ICC DMS. Also others sources and workflow tools are being implemented. The ICC DMS project is designed as an open system, also to formats. It's suggested a deeper standardization of protocols of information delivery and associated data bases we attempted, till we obtain the digital certification of the files. Finally, it is necessary to emphasize that the geologic document repository of ICC DMS must be considered like the beginning of the Center of Documentation and Geologic File of the recently created Geologic Institute of Catalunya.

ACKNOWLEDGEMENTS

We thank to all the people, organizations and institutions that throughout the 25 years of history of the Servei Geològic de Catalunya have collaborated and collaborating with us. We thank ADHOC Synectic Systems and GINFO for the DMS implementation.

REFERENCES

- HUMMINGBIRD DM, 2004. Hummingbird Document Management Feature Comparison. Differences between DOCSFusion 4.0 and DM 5.1.05 [Internet: http://mimage.hummingbird.com/alt_content/binary/pdf/collateral/br/migration_docsfusion.pdf 2006.04.10 9:00].
- LÓPEZ-ARENAS, A., FLETA, J., 2003. Aplicació InfoGeol: Informes tècnics de la unitat de geologia. Manual de l'usuari. Informes del Servei Geològic de Catalunya, D-006/2003, 51 pp. Generalitat de Catalunya.