



CERN EDMS Study96: Progress Report

Second Progress Report of the Engineering Data Management System Task Force. 31 January 1997

F Dittus, G Faber, C Hauviller (Chairman), J Kuipers, B Nicquevert (Secretary), A Onnela, M Price, W Witzeling /PPE

C Delamare, A-P Hameri, M Mottier, J Nikkola, T Pettersson, J Schinzel, M Tarrant /EST

P Farthouat, P Palazzi, B Rousseau /ECP

M Ferran, N Høimyr, A Osborne, S Santiago /IT

J De Jonghe /AS, P Strubin /LHC, S Oligier /ST.

- ATLAS TECH-No-023
- EST-ISS/97-01
- IT/97/2
- The Publication Numbers of other Divisions and Experiments will be added in due course

ABSTRACT

An Engineering Data Management System (EDMS) is a collection of tools and rules to create and maintain in safety an easily accessible body of reliable information. In 1995 the Task Force specified CERN's EDMS needs and looked for a suitable product on the market. See the first Progress Report [1].

A study of the EDMS market yielded a shortlist of 4 products, from which Matrix [2] was selected to be used in three Pilot Projects. Each Pilot Project built a trial data management framework for a well-defined sector in current CERN activities (LHC, Atlas and CMS). After a three months test the Pilot Project teams concluded that the use of Matrix was clearly better than producing an ad hoc solution. Furthermore Matrix was thought to be sufficiently flexible to be adaptable to different CERN needs and serve as a basis for an EDMS in all parts of the laboratory.

Matrix showed some immaturity in user-friendliness and data security - both key issues - and had no inside links to Euclid and AutoCAD, CERN's principal CAD systems. It includes few engineering functions so these must be provided by CERN, representing expense and commitment. These CERN-specific functions might be, or become, different from mainstream engineering methods. Using a more traditional system would allow, indeed oblige, CERN to use common industrial concepts and functions which are maintained and further developed by the vendors - and by the world at large. To compare directly the pros and cons of tradition with those of modernity - and in the light of better understanding of EDMS realities - CADIM [3], which is the European among the shortlisted products, is being re-examined in a single Pilot Project re-run. (That examination is beyond the scope of this Report.)

The Pilot Projects brought CERN from the realm of theory to the real world of practice - a voyage of discovery. It was encouraging to see that imagined facilities worked well, delivered the data quickly and clearly - and of course difficulties were encountered, sometimes where they had not been foreseen.

CERN will very soon be in a position to go ahead with the CEDAR program. (CEDAR [4] is the name chosen for the implementation - CERN EDMS for Detectors and AcceleratoRs). What must be decided soon is which commercial package we should choose, Matrix or CADIM. The Pilot Projects will continue and serve as growth points in their respective areas. Plans must now be laid for the gradual generalization of EDMS use. Templates, based on present achievements, will help newcomers and serve to accumulate knowhow. Data modelling tools, underpinning those templates, will be set up to ensure the overall coherence of CERN's engineering data. Manpower must be allocated to develop CEDAR in a coordinated way.

Mik Ferran, (editor) 17-March-1997

Table of Contents

- Introduction
- Pre-Selection and Short-List
- Benchmarks
- Reference Visits
- First Choice
- Getting Ready for the Pilot Projects
- The Pilot projects
- Principal Conclusions
- Next Steps

The EDMS Task Force

The Task Force is made up of members of most CERN Divisions and of our most advanced major projects, LHC, Atlas and CMS.

Introduction

CERN's need for an EDMS is dealt with at some length in the first Progress Report where special needs, such as a Web interface, are outlined. The arguments of that report are still felt to be valid and only a few central points are reiterated here.

- More power for each engineer:

Engineers have always written and drawn details of designs on sand, on walls, on paper ... be it for private use in developing their thoughts, for safekeeping or for communication. The quantity, the interdependence and the diversity of origin of design information for large projects can now be so great that both the data and its coherent organization have to be entrusted to computer-based systems. The EDMS must act as an extension of the engineer's mind, enabling him to "know everything" almost as quickly and comprehensibly as from his own memory; effortlessly folding in knowledge from collaborators, however remote.

- Communicate with the past, the future, the far-removed:

From the construction to the decommissioning of LHC, a wide and ever-changing range of engineers must work effectively - often having little personal contact with colleagues. Essential, reliable data must be clearly accessible to these engineers - from the ever-present "extension of the mind".

- Industrial strength with (professional) human kindness:

CERN's EDMS implementation, CEDAR, will be set up using an industrial product which has proved its worth on the open market. This will bring economy of scale and allow CERN to continuously absorb mainstream knowhow from the user community. It is fundamental for the success of CEDAR that it be subservient to the engineer, that it increase his well-being and performance, focusing on his needs and protecting him and his knowledge against all manner of reorganization and interruption. Of course, making CEDAR wonderful requires a disciplined effort from the engineer. It is he who understands his data and must enter it in the overall body of knowledge in the proper way, accepting and following the CEDAR rules - and contributing to the improvement of these rules, for they will become an essential part of his professional life.

There is an ever-present need for EDM; clearer management can bring increased productivity, economies in time and staff. Engineering companies have, in recent years, abstracted their production procedures and installed and profitably exploited them in factories in non-industrialized countries, freeing themselves of tradition, of high labour costs etc. Having shown that simple engineering can be managed by explicit procedures, the lure of generalized EDMS becomes irresistible. Clients have, however, difficulty in specifying their needs and choosing a product which suits them. This is entirely reasonable as it entails distilling a clear, coherent description of the engineering process from many separate, specialized activities, some of which may have wilfully been kept secret. There are no guidelines from an established tradition - relying as EDMS does on the recent price-performance and reliability of computers and computer communications. Nevertheless many believe that a solution to their EDM problem can be provided, they accept that no ready-made solution will do everything for them and are willing to invest in a serious creative effort.

Of CERN's two EDMS investments, the bought-in toolkit and the in-house implementation, the latter is by far the greater and the more precious. It will grow to be CERN's corporate heritage in engineering, it will live as long as the laboratory. Thus, it must not rely on particular characteristics of a chosen EDMS. Indeed, as the first Progress Report pointed out, we should require our EDMS data to be perennial and use computer-based semantics to allow it to migrate into the future, or to another system.

Pre-Selection and a Short-List

An invitation to tender was sent to EDMS vendors and included a questionnaire to determine how well they satisfied CERN Users' Requirements. This attracted offers for 11 products. Four offers were eliminated as incomplete or too expensive, leaving 7 contenders, each of which has a serious share of the market. [ADRA, ComputerVision, Eigner & Partner, IBM, Metaphase, Sherpa, Unigraphics].

It was unrealistic to try to carry out thorough tests on more than a few products, so the next step

was to focus on 'several' which were clearly better than the others for CERN. To reduce the number of candidates to a manageable number, products were awarded points according to how well they satisfied users' requirements, whether functions were available immediately or later etc. This study was fairly thoroughly done since each vendor had, as part of the tendering process, filled out an electronic form indicating which of the CERN Users' Requirements he fulfilled, to what extent, when, and so on. Three CERN groups made independent evaluations, to diminish the possibility that a dominant personality with doubtful taste might lead us in the wrong direction. As a further insurance, consultants in the field, notably from CIMdata, an eminent company in EDMS, contributed their outsiders' experience in assessing the offers. While the CERN groups had specialized understanding of the Users' Requirements, the consultants contributed a familiarity with the market jargon and, from previous knowledge of the products on offer, could give some guidance to realistically interpreting the vendors replies. In the end, general opinion, consultants and evaluation groups were all in good agreement on a list of four products. This shortlist was subjected to Benchmark Tests to find the most attractive system on which to implement Pilot Projects in CERN.

Shortlist of Four

1) Metaphase [5] is a market leader. Its native O-O schema is mapped onto relational database systems. (Oracle, Sybase or Informix). While others rested their case on the extent to which they fulfilled CERN's User Requirements, Metaphase warned that requirements were dynamic and that fulfilling them was an ongoing activity rather than an attainable goal. Metaphase offered the widest range of functions.

2) CADIM from Eigner & Partner was the European product (using an Oracle or Sybase database). It has interfaces with Euclid and AutoCAD, the staples of CERN's CAD activity. It is well-established - over ten years on the market - gives high customer satisfaction and is still forward looking - eg to CORBA and STEP.

3) IBM ProductManager [6] is a fairly new offering running on Oracle or DB2 and on non-IBM equipment.

4) Matrix from ADRA is light, flexible and modern - on the leading edge of the technology - but unproved. As the most general solution it did not have the in-built engineering style common to the others. Objectivity, its underlying Object-Oriented DataBase Management System (DBMS), is highly-regarded but is not the leader of this young sector of the DBMS market.

This was as far as we could go with users' requirements and vendors' specifications. The next step was specific tests and hands-on experience.

Benchmarks

Benchmark Test Design

The Benchmark Tests were to allow the vendors to show how their products worked on real CERN data, through both imposed and unrehearsed sequences of operations. The foreseen Pilot Projects, which had already provided scenarios to help in the preliminary assessments, were again used as a source of Benchmark data. The value of arranging tests in realistic scenarios was that each product had the opportunity of illustrating its lucid, consistent behaviour - much more telling than scoring

unrelated points against a check-list.

This Benchmark data was documented and organized in clear conceptual structures such as the Product Breakdown Structure, PBS, and the Work Breakdown Structure, WBS, and stored in readiness on a disk. The tests were carried out at CERN when possible, to allow interested users to see the products first-hand, to see all products in the same context and to see how comfortable each system appeared in the CERN infrastructure.

Essentially the competing vendors agreed on a two-day period in April 1996 for the tests. The agenda for a benchmark session contained three phases; the benchmark 'grid', a series of comprehensive questions and tests; 'scenarios' to show the overall coherence of the product; and 'hands-on time' to allow the testers to see if they really understood how the system worked. Twenty days before the start day of his test, each vendor was given access to his Benchmark Test data, on a disk at CERN. Apart from PBS, WBS and related documents this included CAD files, PostScript, Hewlett Packard Graphic Language (HPGL), text etc. Much of the intent of the Benchmark tests was obvious from the data and explanations but additional CERN-specific exercises were revealed in the course of the tests. All vendors had the same length of time for preparation.

Where possible, the same CERN people, drawn from a pool of six, were present at all Benchmark Tests.

Particular attention was paid to the clarity and control the system offered the user, its openness to the external world and the ease of everyday use, such as to check out an AutoCAD drawing, modify, check in. The grid tests, reflecting User Requirements and consultants' advice, the subjective hands-on sessions, and the creative scenario questions and answers, helped the Benchmarks to cover most aspects of each product. Product experts were present at the Benchmarking to clarify technical issues and avoid confusion and false hopes. Documentation was available for the hands-on exercises - pitched at different levels, from casual user to specialist.

Benchmarks - Some of the Action

(Full details of the benchmark exercise are available in the Benchmark Report [7].)

The IBM team did not find the time to prepare and execute a benchmark test in the time specified but came to CERN for a demonstration. Considering the demonstration, the fact that IBM could not manage a test on the same terms as the others and because the information in their tender was incomplete they were not considered further.

Metaphase was benchmarked at CERN on 18, 19 April. Due to scheduling difficulties in the vendor's team the whole benchmark demonstration was performed by one, exceptionally competent, Metaphase expert. This market leader had good graphical browsing through structures and lifecycles and had comprehensive interfaces to CAD tools. Metaphase looked heavy for a research laboratory and since customizing consisted of stripping down, rather than building up, not all of Metaphase's qualities could be kept.

CADIM was benchmarked at Eigner & Partner headquarters in Germany on 22, 23 April. They had a well-prepared, in-house infrastructure with 4 CADIM persons in attendance. Initial misunderstandings were soon corrected. CADIM has good integration with Euclid, AutoCAD and MS-Office. The tabular browsing which was demonstrated looked ponderous, but some graphic browsing is also available. There was no MacIntosh interface.

Three Matrix persons attended their benchmark at CERN on 25, 26 April. The test was well-prepared, though the scenario implementation was unimpressive. An abstract graphical interface could be mapped to any data structure - but this abstractness could be confusing. On the other hand the structure of the information could be clearly illustrated on the screen. A wide-angled 'indented browser' shows overall hierarchical structures and a close-up 'star browser' gives fuller details of a single - mother-children - generation. No integrations with CAD systems existed (save for CADRA, from the same vendor) though external data vaults, like the CERN Drawing Directory, could be referenced.

Conclusions of Benchmarks

The three remaining products appeared competent and adaptable, but with strengths in different areas. In particular, it was difficult to imagine how our data, managed in one of these systems, would be more accessible, user-friendly, complete, intuitive, helpful and safe than if it were managed in one of the others. That is to say that after the Benchmark Tests it was not possible to choose one product as more suitable than the others, nor was it possible to eliminate one product as less suitable than the others.

Reference Visits

Reference sites for the three candidate systems, installed in engineering environments, were visited to see how the EDMS behaved in real life, what they did for the user and what the user thought of them. The installations visited were arranged by the vendor in each case, giving each the same opportunity for bias. Only in the case of CADIM had the system been in production for some time.

Muller-Weingarten: CADIM User

Muller-Weingarten manufactures presses for the motor car industry. A press has from 40k to 80k parts, half electrical, half mechanical, of which about 30% have to be designed afresh for each new order. The maintenance commitment is long - in their case dating back to 1936. The working metaphor is that all presses are the same, with some differences, so it is relatively easy to be methodical and to understand the processes.

MW wanted their whole delivery/engineering process visible and under the control of the engineers instead of being hidden in masses of paper. An added benefit of making data widely available through an EDMS was the participation of the manufacturing teams early in the engineering process. MW had moved from an in-house Bill Of Material (BOM) expression of the Product Structure to an EDMS which gave Configuration control and Workflow control - and still produced BOMs. This adaptation to a simple existing methodology minimized re-training.

'Integration' to CAD systems was important in their selection. CAD integration (AutoCAD12, with CATIA for 3D surface models for the dies) uses the information in drawing title blocks and is close to the manufacturer's tradition of BOM evolution. The engineering department manager (together with the IT manager) was in charge of the EDMS project throughout - important in overcoming user resistance. MW did not run a pilot project but their staged (4 months, end '94) implementation served as a pilot. Two months programming by a student gave enough local flavour to CADIM and has lasted two years without change - unaffected by 10 software updates by the vendor. There are 100 daily users - engineers and designers. Two factories share metadata but store their own CAD etc. data. They have 25 document classes and sub-classes, 30 main component classes, and 80 000

different BOM items. One person is responsible for both system administration tasks, like updates and testing, and managing the Entity-Relationship model of the application. Electronics micro-code is managed by CADIM. Conceptual design of the presses has not changed and is done on extremely large sheets of paper.

Eigner & Partner handle problems quickly and the unique system problem (a database lock) was solved without harm. The user is happy. They appointed an EDMS professional to find their system and he chose the one he knew - he had come from Eigner & Partner - for he knew it was adequate.

Foster-Wheeler: Matrix User

Foster-Wheeler Europe in Reading mainly deals with the project planning of process plants, notably for the petrochemical and nuclear industries - Exxon, Shell, BP etc. They were preparing a Matrix installation which had not yet been assigned to a project, but was to control the ordering of materials. They too focused on BOMs, aiming to have accurate information widely and quickly visible to all departments for the ordering of piping, instrumentation and electrical components. Their investment in EDMS was not for engineering data but to free the management of relationships and work-flows from the heavy procedures associated with managing engineering data. They use standard Intergraph CAD database management in Oracle, and home-spun Oracle document management; they did not intend to change those. Their foreseen system, for a total user community of 40, had its (out-sourced) customization specified by their engineer. The system gathers user information in forms written with Visual Basic and conveys data to and from other programs.

Matrix was the second product they saw, and they bought it. They were unhappy about the lack of Matrix knowhow in Europe and the difficulty of attracting the attention - and the best efforts - of American support, but they were happy with MATRIX in general and would choose it again, though they had not yet used it in production.

Metaphase User

Note that all confidential information acquired in this visit has been left out.

SDRC - the CAD vendor has also moved into the EDMS business - sees itself as a partner who installs, customizes and develops a Product Data Management (PDM) system. Just as some Japanese car manufacturers have set up plants in Europe, many European car companies are decentralizing operations worldwide. The company visited to see Metaphase are also dumping their mainframes and buying off-the-shelf software. Their IT department is being hived off as an independent firm - to supply off-the-shelf software, to cut costs and to allow the IT people to compete for more business on the marketplace - all in the modern tradition.

Their Metaphase project, XYZ, will handle geometrical CAD data, manage work-flows and the distribution and exchange of geometry and associated data world-wide. They expect that the very communication which has been forced on them by decentralization will contribute to a more efficient product development cycle. XYZ is a greatly enhanced replacement of their present mainframe system, so users were happy to find the same principles in a friendlier form. The philosophy was straightforward - get a kit to modernize the corporate knowhow and tools; this is a clear goal.

Their shopping list was traditional; flexible; robust; distributed, world-wide capability and support;

custom features. Their procurement was also traditional; User Requirements, selection, benchmarks. (3-week benchmarks, on the site). They were about to go into production at the time of our visit, beginning with MS-Office documents and gradually moving into Metaphase. They had customized interfaces to their corporate users, to the existing mainframe processes, to office documents etc. CAD integrations were additional efforts. That first project would effectively be their Pilot Project - Metaphase philosophy maintains that it is a waste to invest in a trial when you can invest in real life.

The EDMS team all sat in the same room. A HP-server would be used world-wide with PC/Windows or IBM workstation CAD clients. Initially serving 8 sites in several countries the goal is 1800 CAD/CAM stations in 18 locations with 30 000 potential users overall.

XYZ has been budgeted 60 man-years to get to full production (software, tests, training etc) when it will be fully integrated with the current (heritage) environment. The effort-profile so far was 60% specification, 30% testing, 10% coding. Training a CAD-user or engineer will take less than 2 days.

Once the basic system is running they aim to build on workflow, then configuration management, eventually the whole life-cycle.

Conclusions of Reference Visits

Again all products appeared competent and well adapted to the job for which they had been chosen. Of course we had been listening to the converted. The only system we saw in production, CADIM, was doing a very humdrum, predictable job and appeared less attractive than the systems which were not yet actually in use. We had learned more, but not enough to prefer or to exclude.

Choice of a package for Pilot Projects

It had been decided early on to focus on a single package for the Pilot Projects (P_Ps). The fear was that one P_P team, having studied and adapted EDMS_X to its needs would form the opinion that EDMS_Y (which it had not studied) could not adapt to those needs. They would not be convinced by the fact that EDMS_Y had been adapted to another P_P. The fiancailles analogy presented in the first Progress Report will not be re-iterated here. (In passing it is worth noting that the spirit of cooperation in the Task Force has been generally strong and generous - it was not clear at the beginning that this was going to be the case.)

We had to choose one of three competent products; each earns a good living, each has qualities and faults and was seen in a 'real' environment. We found no feature to distinguish The Product For CERN from the others, nor even a silver bullet to eliminate one of the contenders. After specific examinations through the eyes of different users, balancing ease, clarity and flexibility against control and security, estimating - again - how the user requirements were satisfied, there was no clear winner.

This well-established equality, however, vanished when cost was considered. The catalogue prices of the three products were comparable but the Matrix offer to CERN was clearly lower than the others.

The inescapable conclusion that we should buy Matrix for the Pilot Projects occasioned some heart-searching. Matrix was now alone in a sharper focus and its pros and cons became much more important.

The wariness Matrix invoked in some was balanced by the enthusiasm it awoke in others. The freedom in using Matrix, of doing things "our way", meant CERN implementing functions, and functions which have to be implemented come out more expensive than those bought off the shelf - and they generate a maintenance commitment. It was clear that if CERN engineering data methods could be accommodated by existing functions in CADIM or Metaphase,

1. CERN would avoid some of the cost of development and long-term maintenance.
2. Some degree of harmony could spread through the laboratory.

But we knew also that a major element in an EDMS implementation would be to analyze and express the organization of CERN's information in a formal, computer-friendly manner. That phase was obligatory, regardless of the system chosen.

There was a lot to find out and cogitation alone was sterile, indecisive. We had to move forward. We had not eliminated Matrix by direct confrontation with its competitors so now that it presented a clear financial advantage the uncertainty was over. Matrix could do the job so - conscious of our fallibility - it was chosen for the Pilot Projects. The level of confidence at the meeting which took this decision was recorded in the minutes: "The Task Force might come back on this decision after the Pilot Projects phase".

We were looking on the bright side and leaning as far into the future as was reasonable. Matrix is modern, adaptable and free of the rigid formalities which seemed to weigh down other products. It is not tightly bound to engineering - FermiLab bought it to help in document management for the computing division but also plans a Matrix exercise in tracking aspects of the SVXII detector upgrade [8]. A new product avoids some old problems, and if it is immature, with bugs, instabilities, missing features, etc. it will probably still be alive - and mature - when today's mature products are dead. Matrix's underlying database is the Object-Oriented DBMS, Objectivity, which is outside CERN's Oracle tradition, though the RD 45 project plans to use Objectivity for LHC physics data management. People have long been dreaming of an infinitely adaptable O-O DBMS to manage their large, complex engineering enterprises. Matrix might make the O-O dream come true.

Getting Ready for the Pilot Projects

Matrix was installed mid-July and training started in the summer. Mapping the realities of the P_Ps into a Matrix representation was the first creative exercise. Difficulties were expected in translating cultural procedures into a computer-based system, however it turned out that there was frequently no procedure to be translated, many situations were dealt with ad hoc. P_P teams had to discern such lack of procedure, find out what should be done and articulate this in Matrix with the consent and cooperation of the users. This would have been easier if CERN had a corporate procedural style to guide the creation of solutions - or some validating mechanism. Finding and fixing flaws in procedure was at times felt to be a tiresome, uphill beginning to the P_Ps, but it is the very essence of an Engineering Data Management. (... by the sweat of thy brow ...)

The three P_P teams had frequent meetings [9] and discussions to pool knowledge and techniques. Their databases were federated to favour the use of common schemata and tools. There was a common will to focus on existing Matrix facilities, resisting the temptation to write new code or address non-EDMS issues.

[Another important tendency has been the decision - not yet in all quarters - to adopt the

ATLAS/LHC Quality Assurance Plan [10], as a cohesive cultural institution. This Plan, a charter to underpin a "CERN corporate style", will provide a porthole for taking on board useful elements of advanced industrial practice, such as may be found in the STEP standard, keeping CERN in tune with the world - in a position to use off-the-shelf products.]

There was much to learn and do, but before the train started to roll into the real world two important issues had to be dealt with. First, the Web is an essential element of our context, it had to be present in the P_Ps and the Matrix@Web interface was not yet ready. Second, especially with the unfettered Matrix, different people working on different problems would be more likely to produce Babel than a coherent corporate dataset. We needed help to develop good practices and to share them, to make innovation fruitful but avoid the negative aspects of anarchy.

A Stand-in Web Interface

ADRA had not produced its Matrix@Web interface, so TuoviWDM -Web Data Manager- [11], produced by a Finnish team working at CERN, was adopted. TuoviWDM allowed people who felt more at home with a Web browser than with the unfamiliar Matrix interface access to a useful subset of the EDMS functions - and it brought these functions to remote sites. At present TuoviWDM is simple and rugged. It relies partly on frequent snapshots of available data - arranged in disk file directories parallel to the Product Breakdown Structure - independent of Matrix. And it can communicate with Matrix via command files for some services. Access to the full power of Matrix should certainly be restricted to the Matrix@Web facility (the Matrix Query Language, MQL, server callable from C/C++ and Java is now available in preliminary form). However TuoviWDM proved effective in the P_Ps and it, or its descendants, may well play a valuable role for years to come.

Building a CERN DataModel

Matrix can control the evolution of its complex dataset but offers no help in the conceptual design of the dataset or in the evolution of this design. The Object Modelling Technique (OMT) graphical notation and a CASE Tool supporting it, OMTTool, was adopted to help make and manage data models. OMT diagrams clearly picture data structures and relationships - and the pictures clarify thinking, help in trying out ideas and in communicating.

[Behind the scenes these diagrams have equivalent formal, computer-friendly expressions which may be subjected to all sorts of general rules and constraints. Programs, monitoring the use of OMTTool and aware of all the other models which already exist, might warn a user that he has broken a rule or suggest actions to comply with existing designs. In the long run CERN's corporate data model will be so large that such computer monitoring of its development will be essential, but for the pilot projects this was not necessary.]

OMTTool is easy to use and above all its diagrams constitute both a clear documentation and a source from which Matrix MQL scripts can be generated to define the Matrix equivalents (objects, types, relationships.. ..). For a Matrix user, particularly a newcomer, the OMTTool makes pre-existing models available - with structures, conventions and notation - for information or to use as templates. It is also, as a Computer Aided Software Engineering (CASE) tool, available around the clock, around the world.

Experience convinced us that the OMTTool approach is right and it appears that it will be necessary for any EDMS. Apart from the practical benefits of bringing clarity to the design and

documentation, and the possibility of ensuring formal correctness and coherence on a grand scale, the OMTool provides an independent description of the CERN dataset. As it is not tied to the methods of our chosen Engineering Data Management System, this "pure" description would help us map our dataset onto a new EDMS - should we ever be obliged to swap horses. Even more important, it will help to model and control the many inevitable interactions with other Data Management Systems: Euclid will have its own appropriate data storage style, as will the Human Resources of the laboratory, maintenance or other contractual arrangements, and so on.

The adaptation of OMTool to the P_Ps could be improved, for instance, its structures do not map precisely to the full Matrix capability, it does not aim to do a global validation of the corporate datamodel. This is a rich field for development: like the EDMS it is generic and it is not CERN's core business. As in the case of an EDMS, an "OMTool" is necessary and we must find an adequate industrial product for the job.

The Pilot projects

- CERN context for Matrix

Three quite different Pilot Projects allowed Atlas, CMS and the LHC to use an EDMS in a familiar domain doing real work. These also allowed Matrix to show its adaptability.

The Projects lasted from September through November. Each was distinct and independent though regular meetings helped to share experience and ideas, make common decisions and learn more quickly. Each P_P was deliberately limited in scope so that clear questions and answers could be formulated at the end. Other projects realized that an EDMS could benefit them but they had to be discouraged from joining the Matrix activity, Matrix had not yet been mastered and free-lance development would have led to chaos. Above all, conventional wisdom dictates that an initial EDMS implementation should be severely limited in scope to avoid sinking the whole initiative beneath an overload.

Support view of Matrix [12]

The CERN infrastructure was Matrix-friendly, AFS distributing to UNIX and Novell to PCs. The existing Euclid_to_Net channel was extended to transmit Euclid-made files to Matrix. Backups and database administration were fairly straightforward. But, of course, some things went wrong. Matrix deliveries were late with incomplete documentation. ADRA, the Matrix vendor, did not seem to listen to our needs for support - or even to hear them. There were bugs, like the MacIntosh interface did not work (due to Objectivity problems) and had to be banned as dangerous. There were mysteries, like occasional periods of sloth. Matrix is young and far from foolproof so the novice users in the Pilot Projects experienced crashes from time to time. Some early work was lost and had to be repeated.

- User view of Matrix

Many users found Matrix was not intuitive. But the fact that those who devoted time to it managed to learn quickly indicates that it should be possible to present Matrix more clearly. Users also felt that operations could be needlessly awkward. And some realities sank in. Matrix is not particularly Object-Oriented and even hides Objectivity's O-O world from users. The official interface delivered with Matrix was not popular with everyone but the TuoviWDM Web interface, which illustrates the users' work context, was adequate for data

access in the P_Ps.

Certainly, any brand-new system lacks many of the features which have been built into, or bolted onto, older systems during their years on the road. Matrix users managed to work around their problems. After all, the main thrust of the P_Ps was to discover what an EDMS could do, not what it could not do.

- Usefulness of Matrix

Each P_P wanted a safe, well-organized data vault which would be easy to understand, to navigate and to maintain. All types of data would be dealt with - or pointed to - texts, photos, magnetic fields, etc. All of these things were achieved, in as far as they could be tested in the P_P context.

Some sophisticated options had to be excluded or only partly tested to be sure of coming to clear conclusions. For instance, Matrix enables the implementor to create different views of the same dataset according to the user's needs so that the electrician, say, does not see the same objects as the plumber when he browses the detector, though some elements may be common to both paths. Similarly the role and relations which a given element has in the Product Breakdown Structure (the favourite P_P view) will be different from those it has in the Work Breakdown Structure. While Matrix caters for such different intersecting views they were not studied in depth.

Naming schemes, of which there is already a traditional wealth at CERN, were used to classify documents and, for those who knew the names, were a powerful basis for queries. Attribute values could also be assigned to objects and they too could be queried.

- CMS integration of engineering and services [13]

Essentially the CMS Project loaded the CMS integration data into Matrix to make the existing integration information available to the people involved in the CMS sub detector projects both through workstation/Mac/PC interfaces and the Web. They also had in mind an approval procedure for integration. The team is satisfied that Matrix is capable of doing all that is required. A separate evaluation done in the context of the CMS CRISTAL [14] project confirmed that Matrix is usable in this project too. In particular, the matter of tracing each individual part of an assembly, giving its date and place of manufacture, component origins, results of tests and calibrations, history of assembly etc. was not in the scope of the original P_Ps. However these particular facilities are very important for certain production, operation and maintenance operations in the CMS detector. The CRISTAL team verified that all of these functions worked to their satisfaction.

- Parameters of layout and integration in LHC (Short Straight Section, SSS) [15]

The LHC team wanted to see to what extent they could adapt Matrix to the existing SSS working methods, rather than bend the engineers to suit the system. Their aim was to transcribe the PBS to Matrix, populate it with drawings and carry out some functions such as generating an Engineering Change Request (ECR), or a call for tender on the basis of the entered information. It was essential to make the information available to their collaborators of the CEA and CNRS in Paris.

All of the functions were implemented successfully and heritage Oracle data from the CERN

Drawing Directory was made available, if somewhat awkwardly (copying data related to Matrix from CDD to Matrix overnight with a pointer back to CDD which gave direct access from Matrix). By the end of the P_P the users were quite enthusiastic about using an EDMS for most of the targetted functions.

- Design and integration of ATLAS TRT (Transition Radiation Tracker) [16]

About ten users were exposed to this P_P and they found the system satisfactory in general. A heartfelt conclusion in the TRT team, who spent much time in data modelling, was that there was a need for simplicity in data structures. These would have to be simple and adaptable since the EDMS and its data model must be general enough to apply to countless, as yet unheard of objects. Objects' peculiarities should be expressed mainly in names and in attributes rather than structures. Above all they found that planning their work with the OMTTool was essential.

Matrix could be made to do more or less anything the P_P team wanted to do, with some thought and work. This must be distinguished from a product like MS-Project which, with limited thought or work, can be made to do almost anything it wants to do.

- Summary [17] and Conclusion of the Pilot Projects use of Matrix

The Pilot Projects showed that Matrix was a satisfactory EDMS. The main difficulties encountered seemed to be due to the fact that Matrix is a young product. There are worries about stability and response times and LHC project engineers did not find the interface easy to use. But at the end of the Pilot Projects all participants believed that CERN needs an EDMS and that Matrix would be capable of providing the functions required.

The main investment in introducing an EDMS into CERN will be in the customization of the product. Much of this customization work is in the analysis and expression, in formal terms, of all of the relationships and dependencies and procedures, and is the stuff and essence of project management. This work must be carried out in one form or another, even if there is to be no EDMS.

However, in the case of Matrix, where functions have to be supplied by CERN, considerable resources must be committed to setting up and keeping up our corporate datastructure. In particular the understanding and use of an OMT Tool becomes an immediate obligation. This seems resource-intensive, innovative or going out on a limb, compared to adopting a traditional EDMS solution where users would be dealing with the same - if rather blunt - concepts as their peers in the engineering industries. Opinion drifted towards the position of not trying to go too far in one step. The thought of introducing a novel EDMS into the widely varied CERN environment and at the same time learning to deal with an almost infinitely adaptable data model seemed, well, adventurous; with a real chance that it would not converge in the short term.

In other words some people started to get cold feet and to argue that Matrix had not been perfect and that in the light of our real hands-on experience we should review the shortlist before making the final decision. It was agreed that we should re-do an accelerated Pilot Project (Atlas-TRT) on one of the other, more traditional products and that product would be CADIM, the European contender. (Nationalist arguments should always be suspected). Reaching out towards generalized corporate data management from a traditional position could be postponed into the far future when the engineering issues were comfortably settled.

And CERN's evolution towards a corporate data solution would benefit in that phase from the experience of others.

- A footnote on the rate of progress

Despite the success of the P_Ps and the firm realization that an EDMS must be brought into CERN, there is disappointment in some quarters that the technology has not matured more quickly. But good management has always been highly prized (and highly paid). Management is simply a difficult task, even in particular cases. So it will take quite a long time to set up a general management system - even though we only want to manage obedient, docile engineering data.

Principal Conclusions in January 1997

The EDMS Task Force started out with the hopeful belief that the computing power which has become available and the datastores and networks which have proved themselves in other spheres, can all be made to help the engineer to face the many challenges of the LHC era. This belief persists and is strengthened by the experience of all of the Pilot Projects. CERN information has been adequately and usefully organized through the EDMS more quickly, clearly and economically than could have been done with an in-house system. However the fact that Matrix was got to work in the Pilot Projects did not mean that it was the best suited to CERN's needs. A choice was about to be made and that choice would stick for many years. It was felt that, based on newly acquired understanding, another one of the short-listed EDMS, CADIM, should be re-examined in a well-known Pilot Project.

So, - Product Specific:

- Matrix is capable of supporting a CERN EDMS, if its Web interface is successful.
- Matrix was successful in the Pilot Projects. Its adaptability to areas other than engineering would allow its use in a CERN corporate database.
- CERN functions in Matrix need CERN manpower, defeating somewhat the purpose of buying an EDMS.
- CERN's needs might be fulfilled by using pre-existing functions in a traditional EDMS. As a test the Atlas P_P will be re-run in the long-established CADIM.

- General:

- CERN should adopt an EDMS.
- The principal EDMS interface must be the Web.
- The chosen EDMS must guarantee a capable Web interface but this need not be operational on the first day as the TuoviWDM has proved to be a very capable stand-in.
- A modern design tool (CASE tool) must be used for CERN's data model. (This is an urgent requirement - and particularly severe if Matrix is chosen). This tool will enable a clear understanding of the data abstraction issues and ensure that the data may be migrated.
- To evaluate what we have learned, and our reservations, it would be politic to come up for air before making the final choice of product.
- The Task Force will choose its recommended EDMS system in March 1997.
- Meanwhile we can plan development of the P_Ps and new growth points.

Next Steps

Choose an EDMS system for CERN for the next N years

This will be Matrix or CADIM.

- The CADIM Pilot Project.

The Atlas TRT Pilot Project is being re-run using the CADIM system. Since this P_P has already been analysed and implemented in Matrix, a six week frame has been defined for the CADIM version, end January to mid March. The main aim is direct comparison between Matrix and CADIM using the same data on the same computing infrastructure.

The CADIM data model will be exploited as much as possible to find the extent to which CERN data can comfortably conform to an existing scheme and avoid innovative data modelling. The exercises already done with Matrix will be repeated - implement an Engineering Change Request, interface to the TuoviWDM Web interface, communicate with CERN applications...

From mid-March users should be able to work with the same data through Matrix and CADIM and decide which they prefer - the customer is always right. Meanwhile the other qualities of the two systems will have been discovered and assessed by the P_P team. In the light of all this new knowledge a choice will be made.

Elaborate a Final Report

The mandate of the Task Force was to find out if an EDMS would help CERN with the LHC work, to choose a product and to propose a methodology to introduce it into the laboratory. We now know much more about EDMS than we did at the time that task was defined and we must now stand back and consider how best to continue. The next step of the Task Force should be to recommend to the CERN management that it purchase a certain EDMS for CERN, that this EDMS be introduced in a certain manner. In particular the implementation of the EDMS will extend from the details and general culture of specialized activities and services through to a central, CERNwide coherency. This is not a major difficulty - conceptual, social or other - for those who have been studying EDMS in the framework of the Task Force. But an appropriate administrative structure and resources must be defined.

References

1. <http://cadd.cern.ch/cedar/documents/progprep95/progress.html>
2. <http://www.adra.com/>
3. <http://www.ep-ka.de/products/cadimedb.html>
4. <http://cadd.cern.ch/cedar/>
5. <http://www.pdmic.com/sdrc/sdmetaph.html>
6. <http://www.clearlake.ibm.com/MFG/engineering/prodmgr.html>
7. The Benchmark Report has not yet been released
8. <http://www-dbi.fnal.gov/dbi/>
9. <http://cadd.cern.ch/cedar/support/minutes/>
10. <http://atlasinfo.cern.ch/atlas-bin/list?dir=QA>
11. <http://tuovi.cern.ch/TuoviWDM/>
12. http://cadd.cern.ch/cedar/pilot_reports/matrix_sw_summ.fm5.ps

13. http://wwwcn.cern.ch/~kuipers/matrix/note97_001.ps
 14. <http://hpcord02.cern.ch/cristal/main.html>
 15. http://cadd.cern.ch/cedar/pilot_reports/sss_pilot.ps
 16. http://cadd.cern.ch/cedar/pilot_reports/atlas_trt_pilot.ps
 17. The Pilot Project Summary has not yet been released
- Mik Ferran, (editor) 17-March-1997*

Sorry, the following references have not yet been released for publication, this will be done as soon as possible.

- The Benchmark Report
- Pilot Project Summary