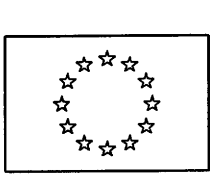


EUROPEAN COMMISSION

INFORMATICS DIRECTORATE
User relations and informatics coherence

INFORMATICS ARCHITECTURE

Implementation

June 1997

Table of Contents

Management Summary

This document is a revised version of the document entitled "Informatics Architecture - Implementation" dated February 1994 (ISBN 92-826-7877-6). It sets out the technical principles which are at a stable level of operation for around three years. The points discussed remain technical, organisational aspects are considered only rarely.

A more detailed level considers the protocols and interfaces used, but also describes the products selected in the framework of "product management". The fact is that for operational implementation, it is workshops or working parties which make the selection of products within the framework of the architecture and develop reference instructions and platforms. Only exceptionally, therefore, does this document refer to products by name. These selections are listed in "product management" documents.

Finally, this document has been compiled for use by the Commission's information technology departments : it does not necessarily represent the Commission's official position. It sets out a reference framework for

- the construction of information systems
- more efficient management of IT products
- the years 1997 - 2000 (with the notion of technological thresholds)

The three major themes at the heart of the new architecture version are:

1. Modernisation of infrastructure

   The technological threshold has been brought up to date to provide appropriate solutions to growing user needs and to keep pace with market developments.

   Definition of this new technological threshold offers users a new degree of stability (not a wild technology chase) whilst at the same time offering a wide range of functions.

2. General simplification

   Five lines of action can be identified:

   - unification within the SNet (Seamless Network) of all types of traffic (computer data, voice, images) and the resulting management benefits;
   - a uniform, multilingual and consistently-performing office information system built around the same family of operating systems for PCs, portables and servers;
   - rationalisation of the information systems support platform, with the decommissioning of the last proprietary platforms and standardisation on a single development tool and a single database system;
   - convergence of interfaces for access to information using the INTERNET paradigm for documentary access and separation of the display component for information systems;
   - component re-usability within information systems

3. Improved integration of systems for

   - administrative document flow, by integration of electronic mail and electronic document recording and management with the mail registry function, and the use of common mark-up and interfaces;
   - coherent data models within information systems;
   - specification of interfaces for exchanges between information systems.

This document is divided into three main chapters: infrastructure, document flow management and information systems.

The chapter on infrastructure is a "bottom-up" update of the former architecture, and covers the following sections:

- the network
- database servers
- the workstation and document server

It will be noted that the notion of infrastructure is evolving upwards and now includes individual office information tools.

Applications servers under Windows NT will be brought into service gradually to support three-tier client-server information systems.

The two following chapters relate to relatively unstructured data (the document) and highly-structured data (information systems).

The chapter on "management of document flows and shared tools" is broken down into

- administrative document flows: administrative correspondence in electronic format, with the associated problem of archiving and exchange formats. The solutions implemented in the office information chain have been drawn from the market, but the products selected are optional in view of their specialised nature. Specific administrative characteristics still require extra developments.

- publication of documents: this is built on the Internet, a universal infrastructure of the IT market .

- work collaboration: a mixed set of technologies permitting groups of individuals to work better together informally. Such technologies are usually known as "groupware".

It may be noted that the **solutions offered** for the management of document flows (described above) and for information systems (described below) **should be perfectly integrated into the underlying infrastructure, i.e. at three levels**.

- **at the workstations** supporting the office systems platform and on document servers

- **at database servers** with the selected database and on applications servers for three-tier client-server

- **on the network.**

The success of the integration of these solutions with the underlying infrastructure is heavily dependent on an efficient policy of development and on appropriate user training for the tools. In particular it is important to avoid the proliferation, as a result of choosing the wrong development tool, of multi-user applications which would be inefficient from the network point of view.

The "information systems" chapter

- describes the general organisation of the information systems. This allows each Directorate-General the tools it needs for management purposes whilst retaining coherence across the institution;

- proposes a form of structure for information, leading to the use of a common reference framework for data across the Commission. Using it guarantees data consistency, facilitates list and report generation and, generally, simplifies the integration of different systems;

- defines an architecture for information systems based on the separation of the search or updating processes and the processes responsible for data display. The separation leads to the development of reusable modules, to the standardisation of the "display" processes and a simplification of the construction and use of such systems;

- describes how information systems use the underlying infrastructure. The database logic and programming language used permit the construction of three-tier client-server systems. This option lightens the processing burden on the PC, facilitates system control and maintenance and may in time lead to the standardisation of the part resident on the PC. This separation of the various modules makes the network less critical to the operation of information systems.

- proposes a rational use of applications packages.

Important note:

A number of technological concepts are omitted from this architecture despite their theoretical attractiveness. Their lack of maturity does not allow the construction of operational solutions within the lifetime of this document. The concepts thus omitted include *inter alia*the "Network Computer", the client-agent-server model for users on the move, and use of the Internet infrastructure for the construction of complex or "mission critical" transactional systems.

CHAPTER 1

INFRASTRUCTURE

SECTION 0 : BREAKDOWN INTO DOMAINS

Organising the informatics (computing) infrastructure of the Commission is not a simple task. In order to overcome the complexity, the architecture is separated into domains which are essentially organisational in conception. The architecture recognises three types of domain :

- the local domain, managed and used by a single Directorate General (or equivalent independent department). This domain is capable of adapting to organisational change within the Directorate General and the institution.

- the common domain, which is the infrastructure common to Directorates General **and includes the Delegations, which are subjects to special considerations**

- the external domain, linking the Commission to its partners : the other institutions, governments of the Member States, public and private-sector organisations within the Member States and individuals.

Each of the three domains corresponds to a different technical and administrative situation and must be considered individually. Integrating them is the job of the communications networks, which form a vital link. Domains retain their own internal individuality, but at their boundaries they respect a series of pragmatic rules which permit interoperability and maintain security. In many cases, interoperability can be ensured only by the choice of agreed products.

The frontier between the external domain and the common domain is specifically the responsibility of the Telecommunications Centre (see Fig. 1).

Within the common domain and each local domain, the use of hardware and software is managed by the domain administrative authority, with due respect to the multi-vendor policy, the rules of security and acquisition via a common Product List. Within the common domain the Computing Centre and the Telecommunications Centre are both the subject of a special definition in view of their pivotal role within the Commission's informatics infrastructure. The Computing Centre is a special domain within the common domain.

Within the framework of the Product List, each Directorate General is autonomous for buying equipment including the common need of acquiring :

- computers capable of running their **information systems** for producing or disseminating information (see database servers)

- computers capable of running their **office automation tools** (document servers and workstations)

The mandate of the Security Office is to define the regulations and the technical requirements concerning security and to monitor their implementation by the managers of local and common domains. These regulations are described in the **document "Information Systems Security Architecture" which links the present document and the C(95)1510, 23rd November 1995 decision of the Commission concerning the protection of information systems.**

For the pragmatic use of standards whose maturity has been demonstrated on the market, the architecture refers in priority to *de jure* standards, otherwise to *de facto* standards with a preference for specifications which are in the public domain.

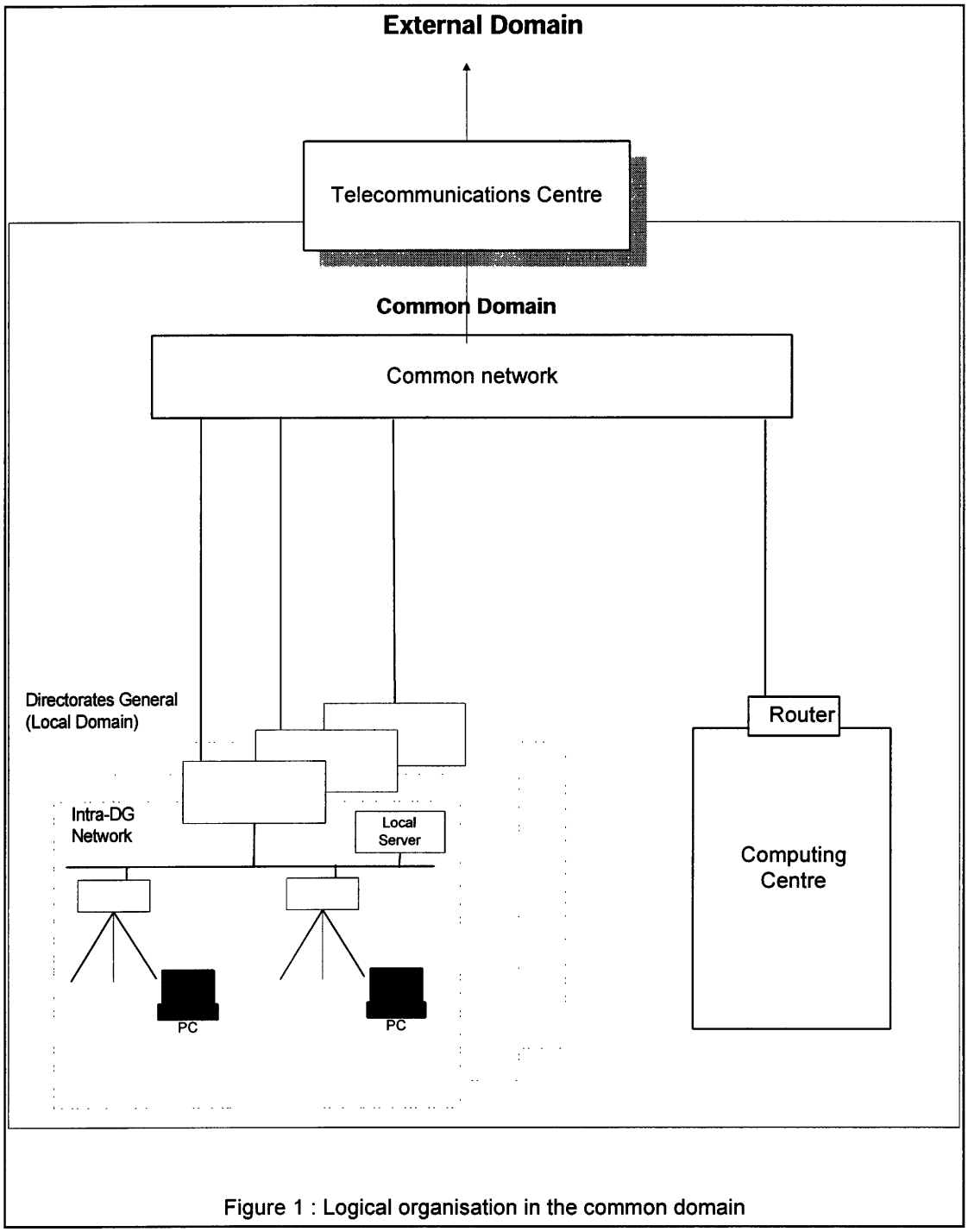


Figure 1 : Logical organisation in the common domain

SECTION 1 : THE NETWORKS

1. THE COMMON DOMAIN NETWORK : IDNET

Since 1992 IDNet is the network connecting the local domains. . It is in effect the communications infrastructure of the common domain, and it possesses gateways (entries and exits) to the outside world via the Telecommunication Centre. IDNet, like the rest of the infrastructure, uses the TCP/IP communication protocol above an Ethernet type network (a migration is planned towards ATM, see paragraph 3 : emergence of the SNet).

A service level agreement details the services provided, the supporting organisation , the quality level warranted by IDNet as well as usage requirements to be followed by users.

The protocols authorised on IDNet are, for batch services

- File transfer (FTP)
- Electronic mail (X400(1988)/ENV41214 and SMTP related protocols)
- remote back-up

and for real-time services

- Interactive access
- Client server access (RPC, SQL, HTTP, ...)

File sharing protocols of the NFS type are required to comply with IDNet usage requirements, since they imply high volumes of data with a short response time and a fine control of security in the absence of passwords and connection controls. Traffic of this nature can be authorised only within a Directorate-General where such interactivity is permissible and in very specific cases where access is closely controlled.[1] This is also the case for X11, TFTP, BOOTP and RIP.

Interactive protocols are being phased out to be replaced by client server type protocols. Their range of functions is too restricted (user-hostility, irrational use of network and CPU power).

Management of IDNet is under the control of a central team responsible for deciding the network installation timetable and monitoring the installation process in cooperation with teams responsible for cabling the buildings. It is responsible for day-to-day operation of the network, ensuring the availability and security, and its optimal operation and management. It is also responsible for any re-location of network hardware.

The Commission has five official TCP/IP addresses. These are five suffixes which allow the creation within the Commission of 1024 sub-networks each with 256

---

[1] In effect, IDNet is not just the inter-DG communications network: when a DG is split over several buildings it forms part of that DG's internal network.

addresses. These sub-networks are made available to Directorates-General according to needs, in such a way to identify each item of hardware unambiguously in a worldwide address structure of type B.

Network components connected to IDNet are managed using tools which rely on the SNMP protocol (Simple Network Management Protocol) to communicate. Associated with the protocol there is a description of objects being managed (hardware, statistics, alarms, etc. ...) and the management tools' access procedures for each of them. The network management protocol is currently being upgraded to SNMP2.

2. NETWORKS WITHIN A LOCAL DOMAIN

For computing purposes each Directorate-General corresponds to the notion of local domain. For purely practical management purposes, a Directorate-General operating on two sites could be given two local domains. Ideally, each local domain is connected to the Commission's private network by a router managed centrally as part of the general management of the IDNet network and in accordance with prevailing security rules. The IDNet router isolates intra-DG from inter-DG traffic. In practice, if a Directorate-General is spread over two or more buildings then there are multiple access points to the IDNet, and thus intra-DG traffic on the IDNet.

Within the local domain, the network is of the LAN-Ethernet type and, like IDNet, based on TCP/IP technology offering the basic data transport functions over the Ethernet. This intra-DG network and any internal routers attached to it are managed with due attention to the rules of addressing and security. To avoid applications proliferation and incompatibility, the methods and tools used for the administration of these networks must be consistent. These tools similarly rely on the SNMP protocol.

The local and capillary networks are segmented in such a way as to optimise traffic flows. In the past buildings were cabled using "fine" Ethernet cable. Current cabling uses twisted pair cables (10 Base T).

Across the intra-DG networks, workstations can receive support from office automation servers for:

- file-sharing;
- printer sharing;
- data back-up and extended storage capacity;
- workstation configuration management;
- local network and access security administration.

The user in front of a PC also has access to

- real-time consultation or updating of databases (client/server or emulation) or hypertext files;
- batch-mode electronic mail and file transfer.

3. EMERGENCE OF THE SNET

The Commission has a number of network problems which are still open.

First, the coexistence on the market of two differing approaches:

- packet-switching networks (Ethernet) established to meet computing needs for reliable data transmission

- circuit-switching networks (e.g. the telephone and ISDN - Integrated Services Digital Network) set up to meet the needs of steady throughput such as voice and motion pictures.

Furthermore the splitting of Directorates-General over two or more buildings, and the consequent removals, raise serious problems of configuration and security, since both are organised on a physical basis. There is a high degree of correlation between the location of an access-point and the services available.

The solution to these problems lies in the SNet (Seamless Network) based on TCP/IP for levels three and four, and on switched technologies for levels one and two. SNet offers

- the services at constant and high speed required for motion pictures (multimedia)

- throughput-sharing services for the optimisation of conventional computer traffic (data)

- and above all the concept of the virtual LAN, permitting the logical definition of a group of users as members of a domain without any constraint of physical location. The fact of moving office then makes no difference as far as software or hardware are concerned: the networks are logical, constructed on a single physical network.

The SNet is implemented at three levels:

- Introduction of the switched Ethernet at 100 Mbps within the Directorates-Generals' local networks, to provide data flow at several tens of Mbps to the servers connected to the intra-DG networks. This migration has already begun in new buildings. The switched Ethernet also permits partial re-use of investments made for the basic Ethernet technology;

- Reduction of the number of PCs on the branches of the 10 Mbps capillary network;

- Gradual introduction of ATM (Asynchronous Transfer Mode - broadband digital technology) on the high-speed lines of the intra-DG and inter-DG network in the framework of modernisation of IDNet.

Two technical problems will nevertheless require solution before ATM technology can be generalised on an SNet:

- either adaptation of network access software from Ethernet technique (i.e. physical segmentation and management by "broadcast") to ATM technique, or

the general implementation of a LAN emulation level on ATM. Such a LAN emulation technique would complete level 2 of ATM;

- ATM is a technology which theoretically covers all networks, irrespective of their distance - short (LAN), medium (IDNet) or long (WAN). In practice, the market implementation of ATM is held up by "bottom-up" improvements to long- and short-distance technologies (e.g. switched Ethernet). For the present ATM is competitive only on medium-distance networks.

See figures 2 and 3 for a technical and functional description of the network at the horizon of the turn of the century.

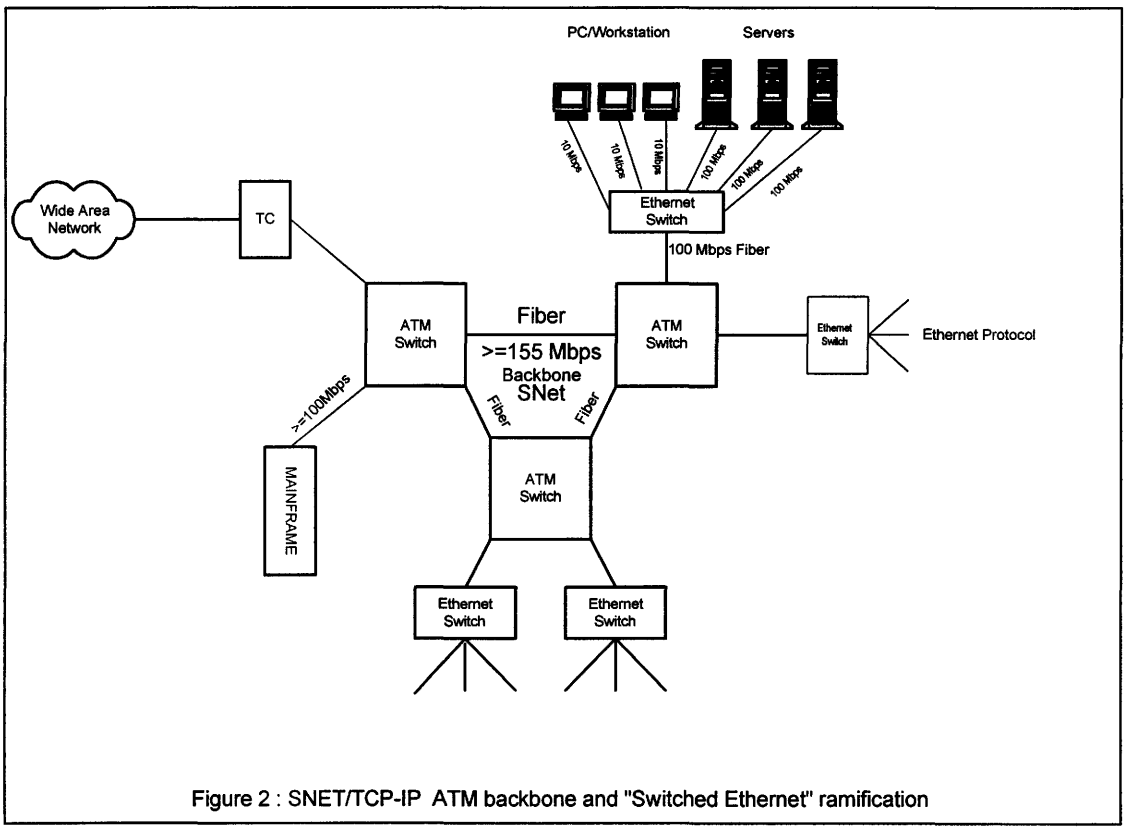


Figure 2 : SNET/TCP-IP ATM backbone and "Switched Ethernet" ramification

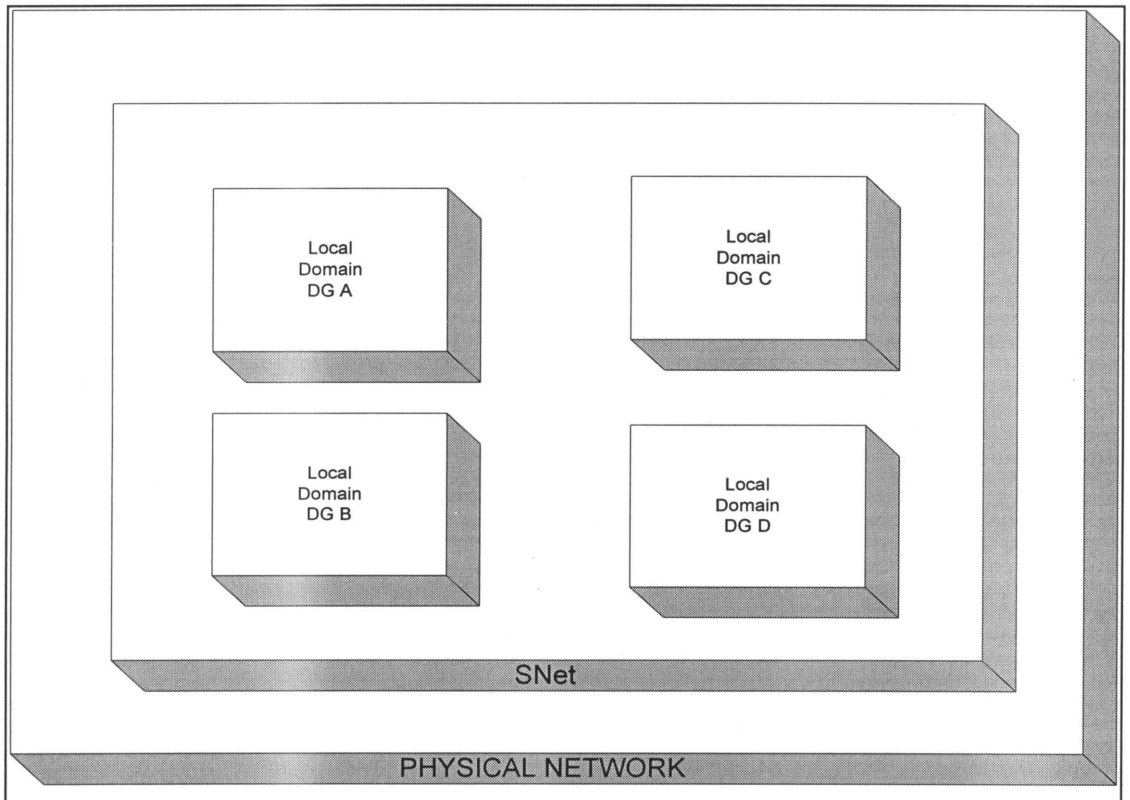


Figure 3 : Functional architecture of the common network

SNet is not merely the successor to IDNet (with the likely introduction of ATM); it is also the successor to the local networks based on an evolution of Ethernet.

When integrated into a complete SNet, management of the local domain network will become independent of the geographical location of its components. **With each local network will be associated a logical network, supported by the physical infrastructure of the SNet common network.** These developments will lighten the burden of managing and organising the network, and will allow resources to be concentrated on information systems and office applications solutions.

In conclusion, SNet represents a simplification of the common network and the local networks, permitting all types of traffic to be merged into a single infrastructure and offering greater flexibility to meet user mobility.

SNet is described in more details in a separate document (EC Network Architecture) and like IDNet is subject to a service level agreement.

4. ELECTRONIC MAIL INFRASTRUCTURE

The fundamental aim of an electronic mail service[2] is the natural and reliable exchange of electronic documents using services analogous to those offered by the postal service: mail boxes, recorded delivery, forwarding of mail in the event of absence or removal. Its advantages over the postal service reside in its speed and the richness of its format which (unlike paper) permits computer processing of the documents transmitted.

The Commission has selected a second-generation electronic mail system, INSEM2. It permits the transfer not only of documents but also the general exchange of electronic messages, in structured or unstructured form. INSEM2 operates on the common network and on the intra-DG networks.

INSEM2 is based on a single product implementing the *de jure* X.400 standard in its 1988 version (ENV41214 and ENV 41218 (P7) and 41219 (P3)). It in fact essential to have this single product, since X.400 covers only a part of the functions the Commission requires, and the others are product-specific. The single product forms the skeleton of the Commission's internal electronic mail system. It is resident on robust servers within the Directorates-Generals known, in the specific terminology of X.400, as Message Transfer Agents (MTAs).

INSEM2 is connected to the outside world via a gateway at the Telecommuni-cations centre (Gateway EM in Figure 2,), the MTA boundary which is the only entry point to the Commission for official electronic mail, and a conversion centre towards other protocols including Fax, and also SMTP, X.400 (1984) (ENV 41201 and 41202), telex, and access to message switching via telephone line and PAD (See Figure 4).

INSEM2 offers a wide-base messaging service which can handle not only document transmission (the electronic mail side of its business) but also other messages in widely differing formats such as Edifact (EN29735) which describes message syntax to be used in the framework of EDI (Electronic Data Interchange). This service will become accessible through a PEDI type protocol (ENV 41120) and an API on the server.

INSEM2 manages its user directories in compliance with a *de facto* market standard as a temporary measure until the LDAP (Light Directoty Access Protocol) based on the *de jure* X.500 (1993 version) becomes established on the market. This *de jure* standard covers the same functions, i.e. provides access to an addressee's address from administrative data (institution or organisation), or geographical data (country, city).

Since the introduction of INSEM2, the electronic mail market has evolved considerably. The office automation upgrade (see section 3) and the breakthrough on the market of Internet protocols (at the expense of the X400 standard) has required a reassessment of the electronic mail strategy. A third generation, which should be

---

2 File transfer is a useful adjunct to electronic mail for the sending or collecting of large volumes o data potentially organised as records. File transfer is nevertheless more analogous to a removals or freight service than to the postal service

deployed by 1999, will support in addition to the X400 standard the following Internet protocols (and their natural evolution) :

- SMTP (Simple Mail Transfer Protocol) and MIME (Multimedia Internet Mail Encoding) : covering basic exchanges
- POP3 (Post Office Protocol) or IMAP4 (Internet Mail Access Protocol) : covering mailbox access
- LDAP : for directories lookup

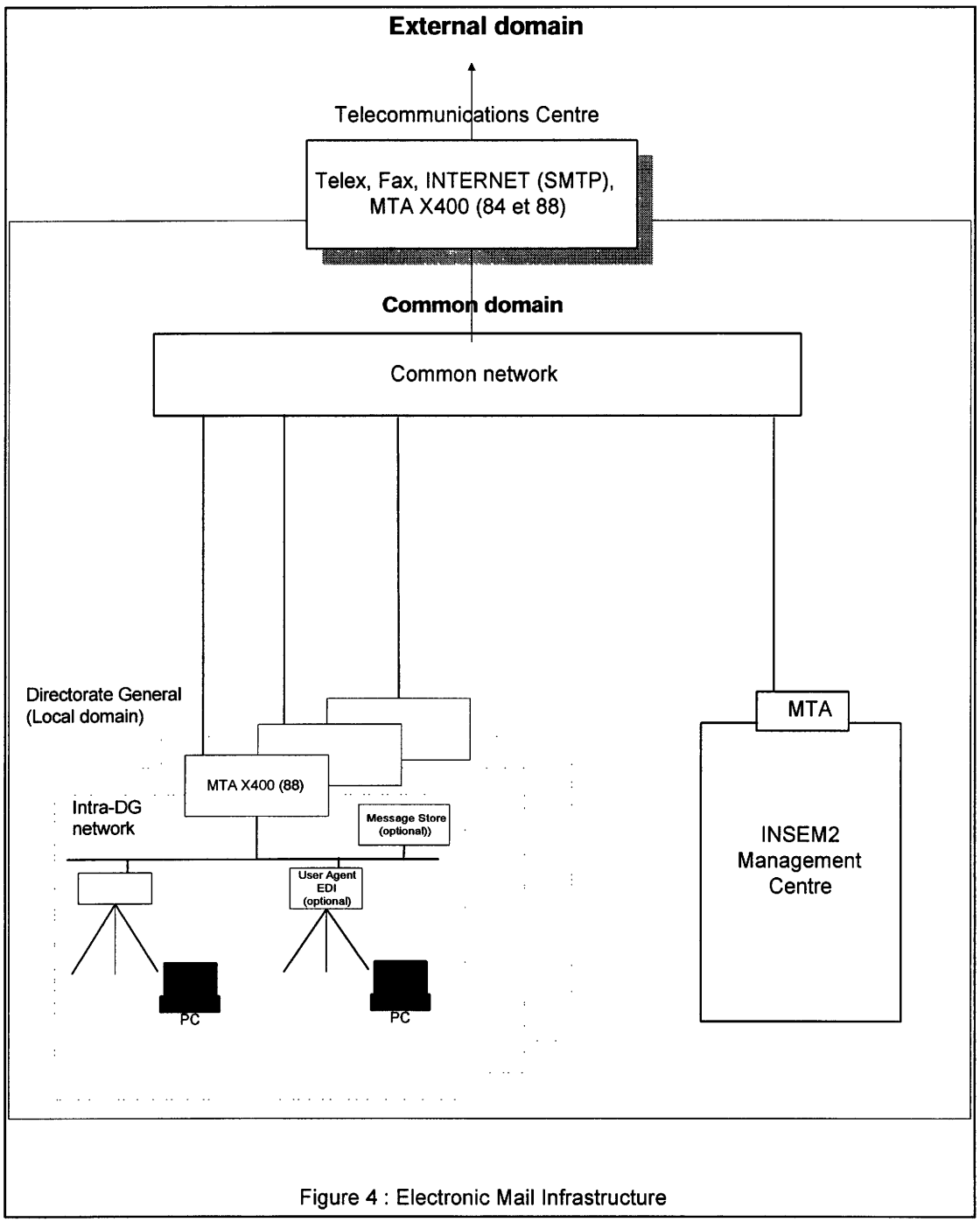


Figure 4 : Electronic Mail Infrastructure

5. THE TELECOMMUNICATIONS CENTRE

The Commission's common domain is separated from the outside world by a Telecommunications centre (see Figure 5). The Centre:

- offers a preliminary entry access **security** check ("firewall") at network level;
- using applications gateways, provides **management services (directories, rights of access, etc.) and protocol (and, where necessary, format) conversion** facilities between the choices made within the common domain and those effected outside;
- offers communications management at the point of exit from the network, particularly as regards **invoicing and statistics.**

The existence of a telecommunications centre managed centrally for all Directorates-General is justified above all in terms of security and by the economics of scale which it generates. These are:

- the concentration of efforts in the quest for technical solutions;
- the aim of a high level of availability (analogous to that of the Computer Centre) under the management of telecommunications specialists;
- rationalisation of connections to external services.

The Telecommunications centre offers batch mode electronic mail and file transfer gateways via

- points of entry and exit for message switching (X.400, fax, telex, Simple Mail Transfer Protocol (SMTP), etc.)
- points of entry and exit for relay mode file transfer;
- points of entry and exit for sundry connections to value-added networks (VANs) whose services are then available to Directorates-General at the best possible price;
- services tailored to the environment of the Commission's delegations, which frequently have no more than the telephone network for their communications.

Real-time[3] access is offered as follows:

- points of entry and exit for interactive access (IA)
- points of entry and exit for client/server access (CS)
- points of entry and exit for Internet access.

---

[3] The Common Access Agent (CAA) offers a protocol conversion and/or management gateway solely for real-time access to the Computer Centre's dissemination databases. The CAA is subject to the Telecommunications Centre's network access controls for incoming communications, but forms part of the Computer Centre

All communications between the exterior and the common domain pass through the telecommunications centre gateways. Apart from E-mail, incoming communications are also systematically screened including for the calling network address. The telecommunications centre E-mail gateway is the sole point of entry to the institution for electronic mail.

The telecommunications centre plays a leading role in inter-institutional exchanges. Priority will be given to optimising these communications and to respecting the EPHOS (European Procurement Handbook for Open Systems) specifications. Connectivity guidelines give further details of the interfaces to de respected and the level of services offered.

It should be noted that the Security Office prohibits the direct connection to an external network of any component of the common domain (e.g. an incoming modem or ISDN connection). Any exception is subject to individual security measures.

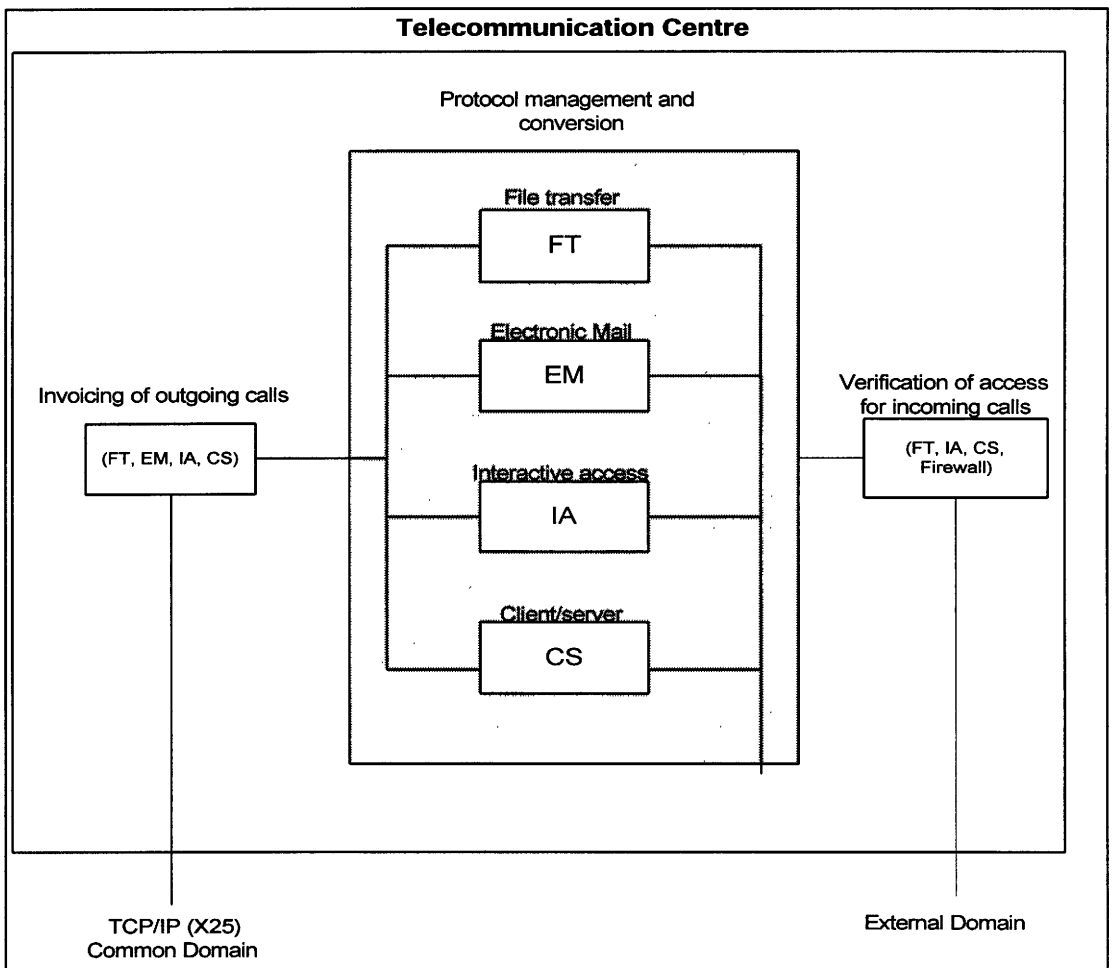


Figure 5 : Architecture of the Telecommunications Centre

SECTION 2 : DATA SERVERS

1. DATABASE SERVERS

For information systems, Directorates-General and the Computing Centre need **database servers**to hold all the data and the procedures giving access to them. The server needs to be few in number in order to concentrate the data (for reasons of administration, security and costs) and provide the means of coping with breakdowns.

The Computing centre strategy is to evolve towards **Unix**[4] in compliance with the X/OPEN specifications. This is the operating system under which **multi-processor machines** may be obtained most easily with the required processing power (scalibility).

Unix allows easy adaptation to **new processors** without the need to re-write applications. Since Unix is also an operating system selected at local level, its adoption in the Computer Centre permits the creation of homogeneous systems within the Commission capable of communicating and collaborating efficiently.

One of the criteria for server acquisition is the machine's compliance with Unix-type specifications based primarily on the Posix public-domain standards EN 29945 and and the specifications of the OPEN GROUP which are in the public domain, together with any other necessary detailed specifications.

The growing availability of servers complying with Unix-type specifications permits the Commission to select them principally on criteria of value for money, whilst allowing a running-in period for any new item. Quality can be evaluated as the machine's robustness (mean time between failures - MTBF) and the efficiency of the support (mean time to repair - MTTR). But this wider availability is still only relative, and there is every reason to limit server heterogeneity by restricting narrowly the choice of suppliers. The penetration of the server market achieved by Windows NTS is being monitored very closely. Unix remains the reference platform for database servers; Windows NTS may be used in accordance with product management procedures and with the relational database systems selected for the institution.

At institution level, a single database applications suite has been selected. This will facilitate management and eliminate all problems of compatibility. Likewise, a single documentary database system has been selected. New developments should adhere to this simplification objective. The selected database must be supported by packages acquired on the market.

---

4 Throughout this document the term "Unix" should be read as indicative of conformity with operating systems complying to the *de jure* ISO standard Posix 1003, together with the OPEN GROUP specifications UNIX 95 ™.

2. APPLICATION SERVERS

With the emergence of the three-tier client-server approach, part of the code (the "business logic") of each information system must reside on a server.

That server could be the Unix machine holding the data. But if the code is to be portable without recompilation and, for reasons of competition, remain independent of hardware suppliers, the server selected must be Windows NTS. For a more detailed explanation of this see Chapter 3, section 4, point 2.

3. THE COMPUTING CENTRE

3.1 Description of the role of the Computing Centre

The Computing Centre has a dual role as a high added-value production unit and as a resource for the dissemination of data.

In its role as a production unit

- The Computing Centre provides for the **management** of data and programs, with a daily back-up, a help desk service **24 hours a day, seven days a week** and the running of a complex tape library and an authorisation and invoicing system capable of handling many thousands of users. All information systems capable of being run more economically by the Computing Centre than elsewhere should thus be found in it. Most frequently these are applications of major importance on account either of the number of users they support, or of the amount of resources (memory, processing power, or both) they require.

- The Computing Centre also has the role of a **back-up service**, in particular for systems which are located within the Directorates-General but require a high level of availability. For systems such as these, the Computing Centre takes over the operations of a temporarily non-operational system, and is capable itself of being backed up by a fall-back system. This is the so-called "white room" concept.

- Data stored at the Computing Centre also include those which require particularly high security (Class B1 or higher), and is becoming reputed as a specialist centre for the handling of confidential data.

- The Computing Centre offers a data management service for Directorates-General which do not wish themselves to manage the infrastructure of their information systems.

In its role as a data dissemination unit

- Data of importance to the entire institution, including data structured for an internal and/or external Web, should also be located within the Computing Centre. For this reason, the Commission's general distribution information systems are held at the Computer Centre.

3.2 Practical organisation

The Computing Centre is required to handle very large quantities of data and their associated access procedures.

For the actual **processing of data**, the Computing Centre relies on very powerful multi-processor Unix database servers **offering very large disk storage capacity and a high rate of data** flow towards the Computing Centre's local network. This is a fibre optic network using FDDI technology (Fibre Distributed Data Interface - ISO 9314) supporting TCP/IP protocols until ATM over optical fibre becomes available using TCP/IP. Only a few machines are used, in order to keep the data together and simplify their management. These machines support both the data and the software for relational and documentary data bases.

In addition to the database application, in order to implement information systems in a three-tier client/server architecture (see Chapter 3) machines are needed which support search and update software (business logic). These can be the same machines as support the database software or other, cheaper systems. The Computing Centre has both, for which the operating systems may be Unix or Windows NTS.

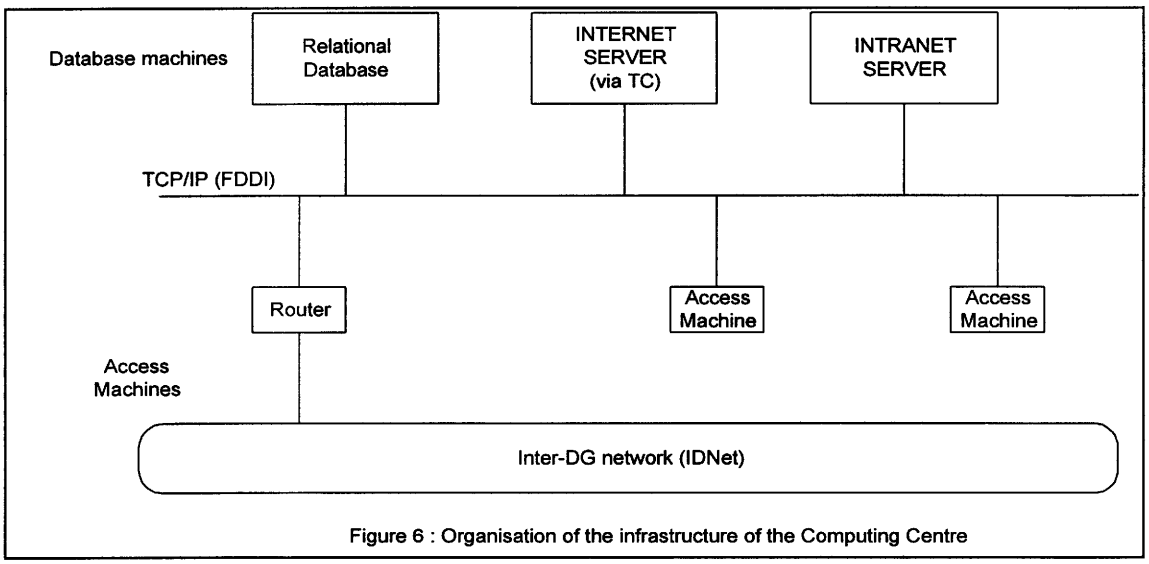


Figure 6 : Organisation of the infrastructure of the Computing Centre

3.3 Migration of the remaining proprietary platforms

The Computing Centre currently still runs a small number of proprietary platforms. Migration of the Commission's information systems assets from these platforms must be analysed jointly with the users, who are anxious to maintain continuity of the services they rely upon. Migration strategy changes with the type of information system and the infrastructure upon which it is built. **The turn of the year 2000 and its impact on date management is likely to accelerate this migration from old systems**

- Documentary databases and menu-access databases will be ported from proprietary environments to the Unix or Windows NTS environment

selected by the Commission. The emergence of Web servers will accelerate this migration;

- Information systems running under SAS and Acumen will be ported to Unix, where those products are available.

- Other information systems will be rewritten in the framework of their natural evolution, with priority going to critical and older systems whose maintenance costs have become particularly high. As regards internal information systems, see Chapter 3.5.

Migration[5] of the last remaining proprietary systems represents a simplification in the sense of unification of the platform supporting the major database servers.

---

5 The target of this migration may change to Windows NT5.X and subsequent evolution, according to the evolution of the large database server market.

SECTION 3 : WORKSTATIONS AND DOCUMENT SERVERS

1. PC

For office systems, DirectoratesGeneral use open market software - the most commonly used on the hardware in question - offering inter alia word-processing, electronic mail and document archiving. This general public software is available on or via all workstations, that is, on the equipment which, by presenting data on screen, permits the user to interact with information technology. Data (text, mailbox) are generally resident on the servers which have the job of organising them, monitoring their integrity and controlling access to them. The data themselves shuttle between the server and the user's workstation where they are processed using office systems software.

The availability of state-of-the-art, integrated office systems software at low cost has hitherto required the workstation to be fitted with the Intel processor (manufactured by the Intel corporation) or a compatible processor, together with the Microsoft Windows 95 or Windows NTW operating systems, themselves the natural development of the Microsoft Windows which dominates the PC market[6] . These PCs are known generically as "Wintel" PC. Windows 95 and Windows NT are 32-bit operating systems which share a strategic array of interfaces known as WIN32 API. Windows 95 was designed primarily to guarantee the compatibility of 16-bit applications; Windows NT offers a higher level of robustness and security, but inferior compatibility and performance.

Portable PC are becoming increasingly widely used when mobility is a significant factor in use of the machine. Such PCs are the subject of more particular instructions. Windows 95 is the operating system usually installed on portables, for its reduced power consumption and improved use of extension cards. Extensions to Windows NT nonetheless make this an alternative also for portables.

As a general rule, ink-jet printers are connected to PCs and laser printers are connected to servers or are directly accessible via the network.

To avoid all unnecessary diversity amongst PCs, strict configuration instructions are applied to the setting up of operating system and peripherals. These instructions form part of the reference configuration for office systems platforms and for the client side of information systems access.

The acquisition of Intel-based PC running Microsoft Windows has allowed the Commission to attain a high level of standardisation and workstation support on a technologically advanced base. Improvements have been achieved in particular in memory management and multilingual support. The Commission is capitalising on the maturity of this new level of technology to develop computing solutions and improved user training. The NTP project (New Technological Platform) tackles the

---

6 The dominant operating system on the PC market. Its specifications are the intellectual property of the Microsoft Corporation and, in contrast to those of the OPEN GROUP, not in the public domain

deployment of the 32 bits PC platform, the selection of management tools, the NT domain model implementation and their link with the organisational domains.

2. DOCUMENT SERVERS

Document servers' definition includes office automation servers, Web servers and MTA servers. The Commission's strategy has been centred on Unix servers. This strategy has encountered three problems. First, administration systems of Unix servers are not yet adequately standardised on the market. For the present this diversity is the biggest single barrier to total marketability of Unix; a solution would be the de facto implementation as soon as possible of the ISO/POSIX 1003.7 specifications which make good that shortcoming.

This diversity has been heightened in the administration of services to PC offered by office systems servers, such as sophisticated user management, network printer queue management, and disk space quota management. This diversity is the result of the lack of synergy between PC and server operating systems.

Finally, with the failure of ANDF (Architecture-Neutral Definition Format) Unix has, since it is multi-platform, become a "source" operating system constraining software suppliers to porting, which represents a heavy brake on the availability of Unix software.

In view of the problems suffered by Unix, the Commission's choice of document server operating system has gone to Windows NTS[7] its future line. Its conception relies on the efficient (i.e. consistent) administration of Intel/Windows PC. This choice is identical to that made for applications servers.

The introduction of Windows NTS allows the prospect of a high level of server reliability through "clustering" techniques under which server failures can be recovered provided the software respect a series of interfaces known as "clustering API". This technique makes servers interchangeable, and they all become document servers theoretically capable of specialisation (file sharing, MTA for electronic mail, Web server, or application server) whilst their configuration can change dynamically.

For hardware management, the emergence onto the market of an architecture along the lines prescribed by the DMTF (Desktop Management Task Force) group is awaited eagerly. It should

- permit basic PC components (software and hardware) to identify themselves in a standardised manner;

- include a database containing those data (i.e. the PC's status) capable of being integrated with data describing the status of the network (SNMP database);

---

7 Operating system complying to POSIX for servers. Its specifications are the intellectual property of the Microsoft Corporation and, in contrast to those of the OPEN GROUP, not in the public domain

- offer a management system for that database which, through public and standard APIs would provide a base service ("local agent") to administration systems.

In the absence of such standardisation, PC administration rely on the tools of Windows NT, which is becoming the de facto standard for the markets for office systems servers e-mail servers and Web document dissemination servers

For security another important selection criterion for servers is their compliance with Information Technology Security Evaluation Criteria (ITSEC) F-C2-E2. When needs dictate, these machines can be configured to the next lower level (F-C1-E1). Security rules allow the appropriate level to be determined and any necessary complementary products to be added to the machines.

For further details, see the document "Information systems security architecture".

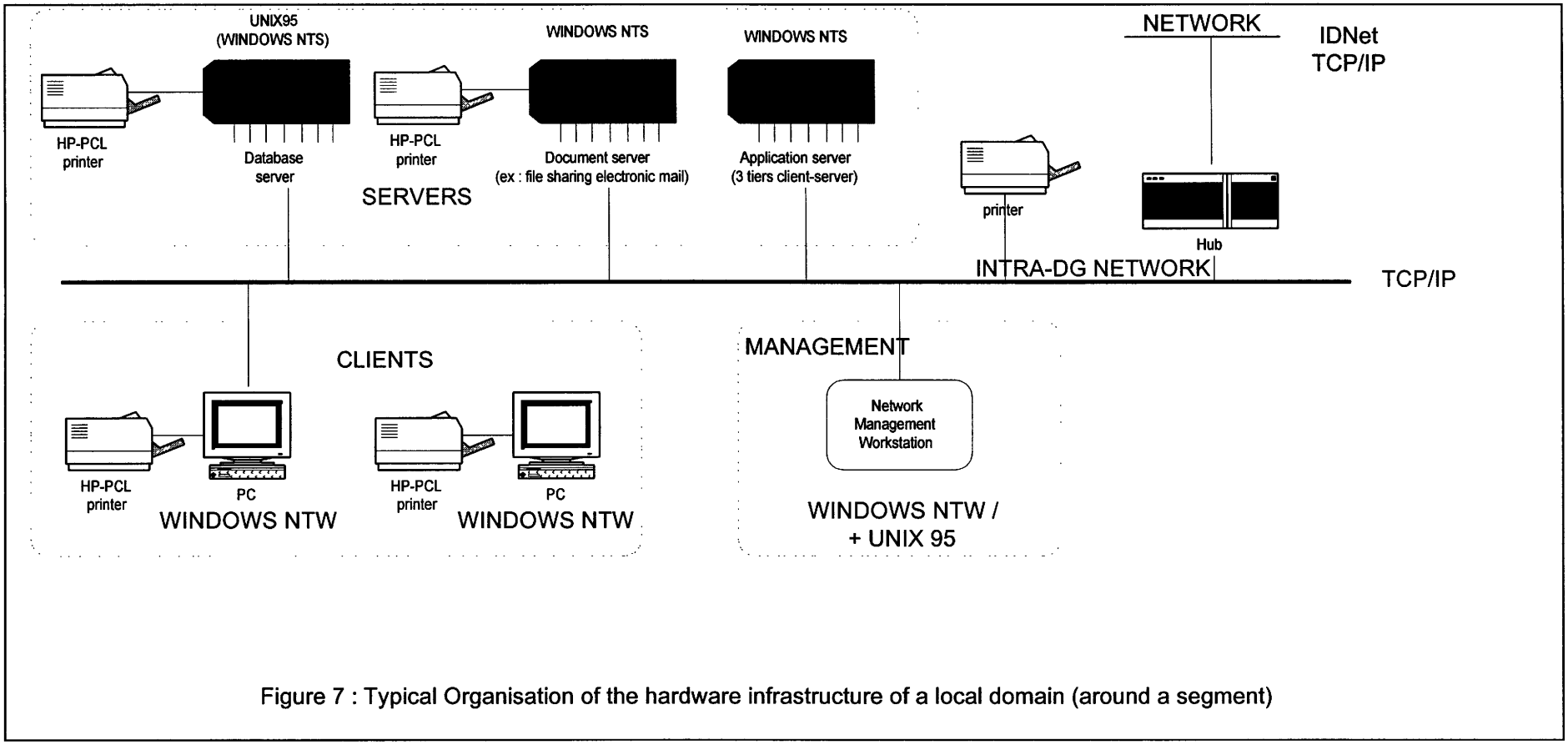


Figure 7 : Typical Organisation of the hardware infrastructure of a local domain (around a segment)

1. 3. BASIC WORKSTATION TOOLS

The users' basic needs are to create, process and print documents consisting of text (correspondence) or numeric data (spreadsheets), or presentations. To this end, they have:

- a suite of software[8] and its logical evolution, comprising a well-integrated word-processor, spreadsheet and presentations application with a consistent look and feel closely matching actual needs. The choice is in line with the majority trend on the market, focused on the purchase of the word-processor with which the spreadsheet and other items are bundled for a modest additional charge. The consistency of the suite ensures simple integration within the mechanisms described in the preceding section. Printing relies on a *de facto* standard.

- Text production should also to conform to Eurolook, a set of general norms aimed at giving a "house" style to documents, and certain specific instructions when the document is likely to be amended as it follows an inter-institutional circuit (DTD SGML structure to be followed;)

- a common programming language for the software in the suite, and for exchanges between them

- help utilities with similar characteristics in each application of the suite, which is a benefit for training

These components complete the office systems reference configuration. A framework agreement generally permits officials to use the same applications at home, which is a benefit in training and mobility.

An important characteristic of this office systems technological threshold is the simplification it offers the similar look and feel of its components. The resulting captivity remains limited and the selections are in line with the majority trend on the market.

---

8 The strategy is to select Microsoft Office 97 and its natural development plus the HP-PCL printer language associated with True-Type fonts. Visual Basic for Applications, a subset of the Visual Basic language is used as a basic integration tool.

1. MULTILINGUALISM

The Commission's working environment being multilingual, simultaneous support is needed for all official Community languages within office systems documents and information systems. The same level of support is not widely available on the market, however, where the offer is generally limited to **restricted multilingualism**, in which the Greek language may not be supported, or support may be limited to Greek and English.

The Commission's earnest wish is to see the generalised adoption of the Unicode 1.1 standard (two-byte encoding under ISO 10646) for operating systems, office systems databases, user interfaces and information systems. This is the only complete solution to the problems of including support for Eastern European languages in a framework of **extended multilingualism**. In an increasingly multi-cultural world the disadvantages of this method of data-coding (the need for more memory) are more than offset by the facilities it offers. Partial solutions are being implemented until this standard becomes available.

The Commission's office systems workstation operates under Microsoft Windows NTW and runs an office automation suite of UNICODE enabled software (see footnote n° 8). This configuration will provide a widely available solution for multilingualism after phase-in.

An associated problem is that of enriching standard keyboards to enable them either to display the full multilingual character set, or at a more reasonable cost to associate with them a range of keyboard simulations from which the user may choose. A multilingual kit is used for keyboard enrichment.

Multilingual data management systems were traditionally located in proprietary environments. Some of these environment were able to support both Roman and Greek alphabets in 8-bit bytes. The migration of these information systems towards Unix requires the definition of a code representing all Community languages. This is Unicode 1.1, which is supported by Windows NTS, the operating system selected for document and applications servers.

Independently of the NT/Unix problem, the combination of Oracle 7.3 and UTF-8 encoding, with SQL-NET 2.3 (or higher) access is the general-issue infrastructure supporting multilingualism under Unicode. This infrastructure will become the rule on all the Commission's information systems.

At client level, the fourth-generation language tool selected must also support Unicode.

Although specific solutions have been developed for the handling of multilingual data, it would be hazardous to introduce a specific multilingual support in directories such as the electronic mail directory. The fact is that any user can receive mail from a country where only the basic ASCII character set (limited to the English language) is

used. The user name König, for example, must be transliterated to Koenig, just as in the telephone directory.

2. INTRODUCTION TO CHAPTERS 2 AND 3

Office systems, whether they are conventional (word-processing, electronic mail) or diffusion-oriented (Web, multimedia), all manipulate more or less standard objects - documents, in which the possible structural type (chapter, paragraph, indent) is largely independent of the content. This means that it is possible to construct standard tool specialised either by type of structure (text, drawing, table, hypertext) or by function (document production, transmission, storage, dissemination). This in turn means that office systems tools are applications sold in high volumes and at low prices.

Shared office systems applications come in three categories, supporting

- administrative document flows (the "paperless" office)
- document publication (Internet paradigm)
- team collaboration (ad hoc shared office automation tools).

Note that it is difficult even to define "shared office applications" - these may be seen as a simple extension of purely personal items (the shared diary) or a general issue infrastructure (electronic mail) or applications similar to those used in information systems (workflow applications).

Information systems seek to represent the specific objects which an organisation such as the Commission manipulates daily. Such objects must be highly structured - it is that structuring which allows them to be manipulated. Such structuring is most visible in cross-referencing between objects. It increases the specificity of objects, and makes them more difficult to standardise. For the purposes of this chapter, "object" means any concept manipulated in a manager's day-to-day work. For a personnel manager, that means that "people" and "jobs" are the objects, for a financial manager, "appropriations" and "payments".

Information systems architecture must answer the following questions:

- How should representation of the structured information be organised?

- How should the systems which create, manipulate and restore the structured information be organised?

- What infrastructure is needed to construct such systems?

This distinction between office systems (poverty of structure but mass distribution) and information systems (enriched structure but specific need) is reflected in the market.[9] This distinction between office products and information systems is far from perfect, but it allows better analysis of the problem and its solutions

---

[9] Note nonetheless that office systems and management systems follow the same path when text is marked up semantically. This is the case when certain types of SGML marking-up are used.

Whether they form part of the shared office systems or of information systems, all the solutions made available must be perfectly integrated into the base infrastructure:

- the workstation supporting the office systems platform and the associated office systems servers
- the database servers with the selected database
- the network.

CHAPTER 2 :

DOCUMENT FLOWS AND SHARED TOOLS

SECTION 1 : ADMINISTRATIVE DOCUMENT FLOWS

1. NEEDS

The arrival of a word-processing application which is open to all and seamlessly integrated into a general office system opens the possibility of significant gains in productivity in the creation of administrative documents. But the Commission's processes also give rise to numerous exchanges of documents within an administrative structure. Making these document exchanges electronic will also improve administrative efficiency, as a result of the gain in transmission speed and improved mail management.

The links in this paperless administrative chain are the electronic mail infrastructure and the electronic document registry and archives, which work in tandem with applications such as a workflow management application and the ADONIS mail registry package. Provision must also be made for a historical archive of documents whose active service is over - see figure 9.

The level of functions on offer and, in some cases, their Commission-specific nature, mean that no general-market solution is yet available commercially. Systems must be constructed module by module from market products to be built into a paperless administrative environment.

These components fit harmoniously onto the basic infrastructure de base of PCs, document servers under Windows NT, the TCP-IP network and database servers under Unix and Oracle. Taken together they amount to a paperless circuit custom-built to meet the Commission's requirements.

2. THE ELECTRONIC DOCUMENT MANAGEMENT SYSTEM

A second generation of electronic document management system (EDMS) complements the basic computing infrastructure described in Chapter 1 with support for images and archiving applications.

In an administration service like the Commission, incoming documents from outside the service are mostly received on paper, and thus prima facie not capable of being incorporated into an electronic document flow.

Such documents can be integrated by means of an EDMS

- by digitising them (using scanners) to produce a file containing an image of the paper (as a facsimile does)
- by addressing storage media matched to the size of these documents and the potential need to retain them (optical disks)

EDMS indexing applications allow electronic documents to be indexed manually or automatically depending on how the document was received, and according to its structure or format. Indices are stored in a relational database.

EDMS storage and index management allow the documents to be presented in the form of files, folders and cabinets whose contents can be searched easily.

Figure 8 illustrates a typical configuration of EDMS hardware components. The implementation inside a local domain of an EDMS infrastructure is optional.

This example includes two scanners: one connected to a workstation for manual data entry, and the other connected to a server for an automatic process which places fewer demands on the network.

The document server is shown separately. This frequently involves digital optical disks; this server can be regarded as a "black box" with its own database system optimised to handle high volumes of data.

The database server, generally running under Unix, contains the base of indices which organise access to the documents.

The third type of server is an applications server, generally running under Windows NT. This server supports applications (workflow, search engines, links to information systems) permitting data to be processed close to the point at which they are stored. For in practice, a substantial proportion of documents are stored in image format, and their file size means they cannot be transported easily around the network without additional cost or loss of time.

Finally, Figure 8 illustrates a document scanning station, a station specialised in encoding stored images, and an example of various research and display workstations, which are top-of-the-range PCs also capable of running under Windows NTW.

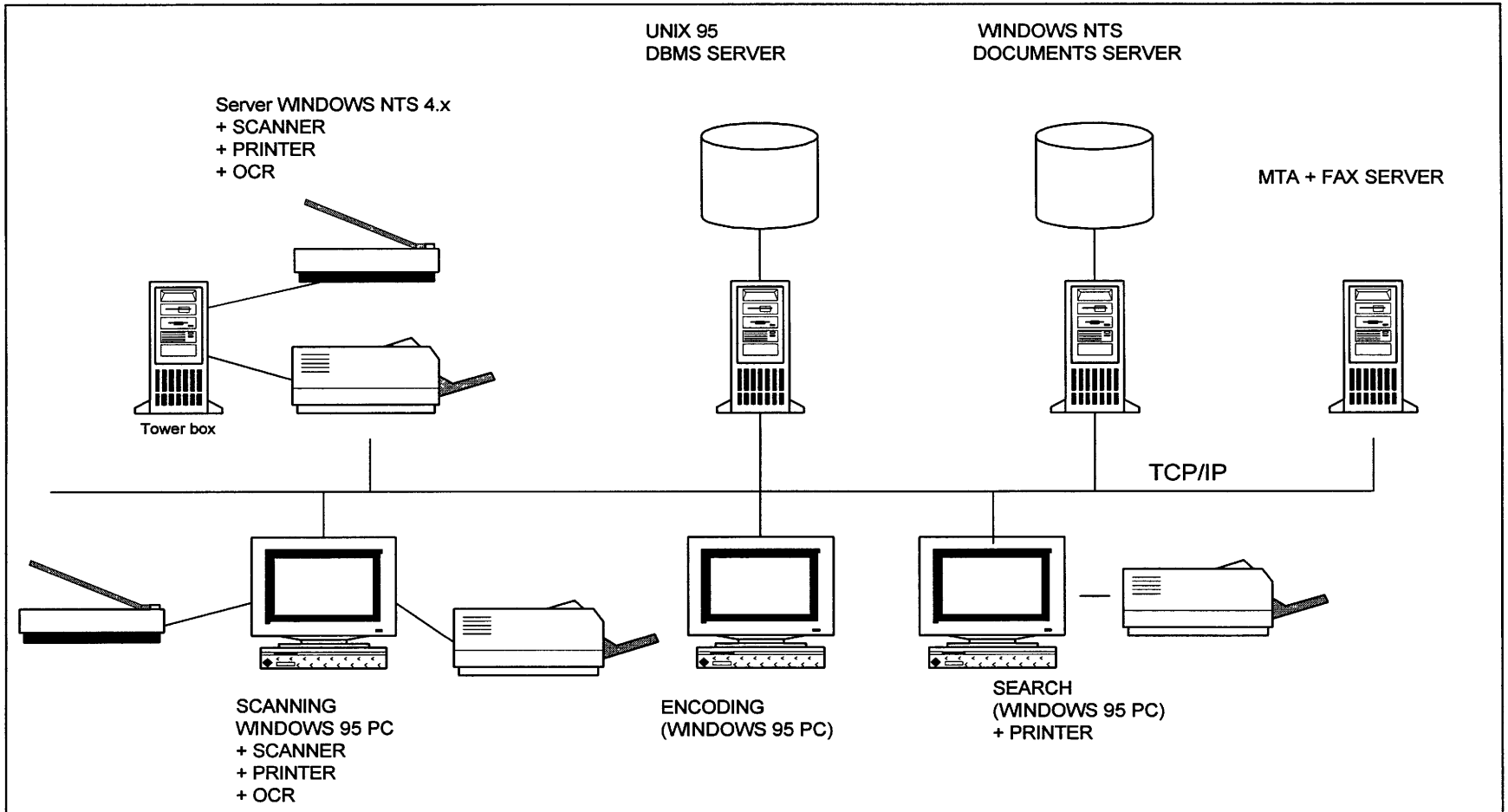


Figure 8 : Typical EDMS hardware components

1. 3. HANDLING OF ADMINISTRATIVE CORRESPONDENCE

The handling of administrative correspondence is built on the infrastructure of electronic mail, directories and electronic document management - see figure 9.

The success of an electronic mail service as a logistic support for administrative correspondence depends on a number of fundamental factors:

- **The user interface must be integrated with the office system**

  This interface runs on the PC and is built on the *de facto* market standard for PCs. It uses a highly efficient protocol for access to mail-boxes (remotely in the case of a user away on mission).

- **Tailoring the application to the organisation,** offering concepts such as

  - "chief/secretary" association for access to mail,
  - the availability of functional, as well as personal, mailboxes
  - association between groups of individuals likely to stand in for each other

- **Efficient document flow management**

  - possibility of filtering unwanted mail ("junk mail" or "all addressees")
  - automatic repeat transmissions making the user independent of network deficiencies
  - accelerated transmission of urgent documents with better-than-average service quality

- **Security,** meaning the absolute guarantee of

  - the origin and confidentiality of an accessible message, by means of digital signature and encryption mechanisms
  - receipt and opening of mail acknowledged by clearly identifiable reports
  - The security measures implemented must be compatible with the smart card selected as a means of identification, and must also comply with legislation on encryption.

- **Integration with document management applications.** This concerns

  - Manual or automated registration of correspondence using **the document structure elements described in SGML**
  - Use of the electronic mail infrastructure by workflow monitoring applications through standard interfaces

- **Multilingual support, i.e.**

  - transmission, reception and display of multilingual documents, including support for Greek
  - multilingualism in the message's envelope, with transliteration of personal names for directory management purposes.

- **support for complete and coherent directories**

For incoming and outgoing mail it is essential to have a **mail registry package** which records incoming and outgoing documents and monitors their circulation and follow-up. EUROLOOK structured documents possess identifiable elements (date, subject, signature, etc.) which can be used by ADONIS, the package used at the Commission.

Archiving software permits storage to be managed - of documents, parts of documents or entire dossiers. The storage may be seen as representing files, and cabinets, which are implemented by attaching data such as key-words and sundry dates to the objects needing to be stored. Without this system, the addressee must print electronic mail and archive it manually.

Mail registry and short- and medium term archiving packages are closely linked. The registry program records the events of the document's life; the archiving package is the dissemination database of registered documents.

Archiving programs have an essential task in management functions affecting documents or part-documents:

- managing access conflicts (multiple writing)
- managing access rights
- administrative support (backup, management of multiple versions, management of language versions)
- ensuring security and confidentiality (encryption)

There are two main types of object to be stored

- documents in revisable format
- documents in image format

Native format is used for "live" documents, i.e. drafts, standard letters, current documents, etc. It is used only for short- and medium-term storage.

Image format is used for

- documents originating in the outside world which arrived in paper format. They must be scanned before being accessible on a workstation;

- the long-term archiving of all documents which have reached the point at which they no longer require amendment, and for which it is preferable to use a format which can reasonably be assumed to be perennial.

The recommended image format is TIFF (16-colour bitmap) compressed in accordance with CCITT Group 4.

Longer-term archiving must offer high-capacity, low-cost support, with an adequate life expectancy. Documents are frequently archived by the dossier rather than individually. Since the archives are in image format, the access infrastructure must be adequate, particularly as regards the network.

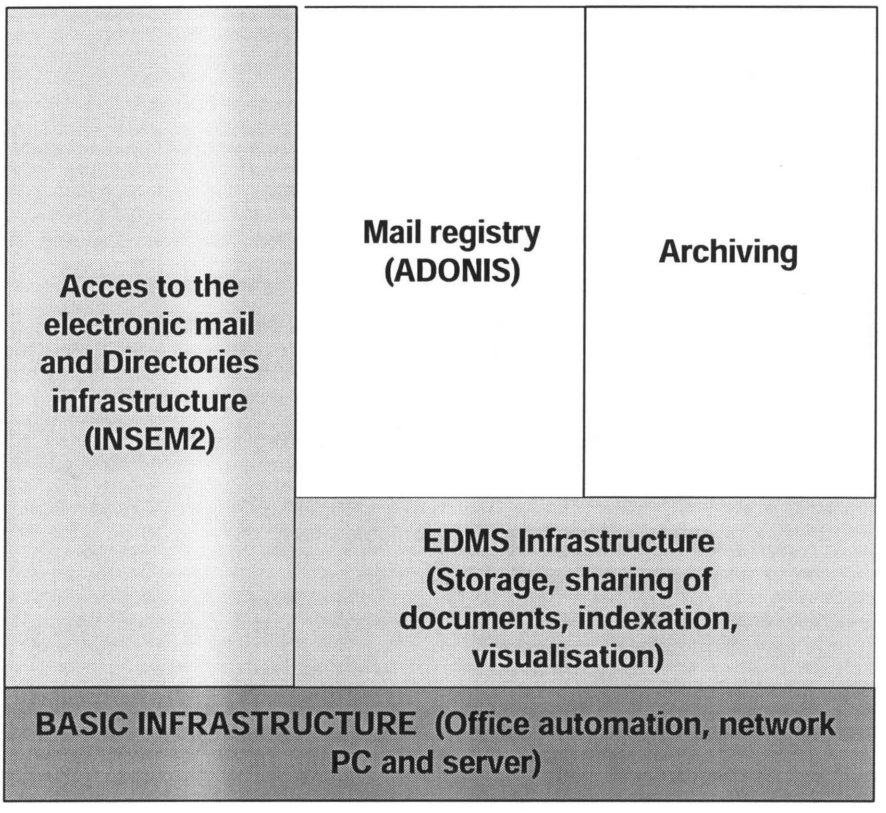


Figure 9 : Software managing the exchange of documents

The Commission's aim is to offer a solution in which registry, archiving and INSEM2 electronic mail work in harness in a modernised environment of electronic document management.

A third generation of integrated electronic mail, directories and EDMS should become widely available on the market. In practice the appearance of Windows NTS as applications server is a catalyst for the development of shared office applications made more robust either by the elimination of application porting to servers, or by the more rational distribution of code between PC and server for performance, security and updating. Evolution in this domain need to be monitored closely for the next technological threshold.

2. 4. MANAGING TRANSMISSION FORMATS

The exchange format should in an ideal world be as rich as possible, leaving the door open to the recipient to manipulate the incoming document in all the ways authorised to him. Document creation software (and native formats, consequently) are highly diverse, however, and this diversity dictates standardisation in a common transmission format. Failing this, the incoming document will be illegible or (at least potentially) deformed by a format conversion. A compromise has to be found between the richness of the chosen format and its universality. (see figure 10).

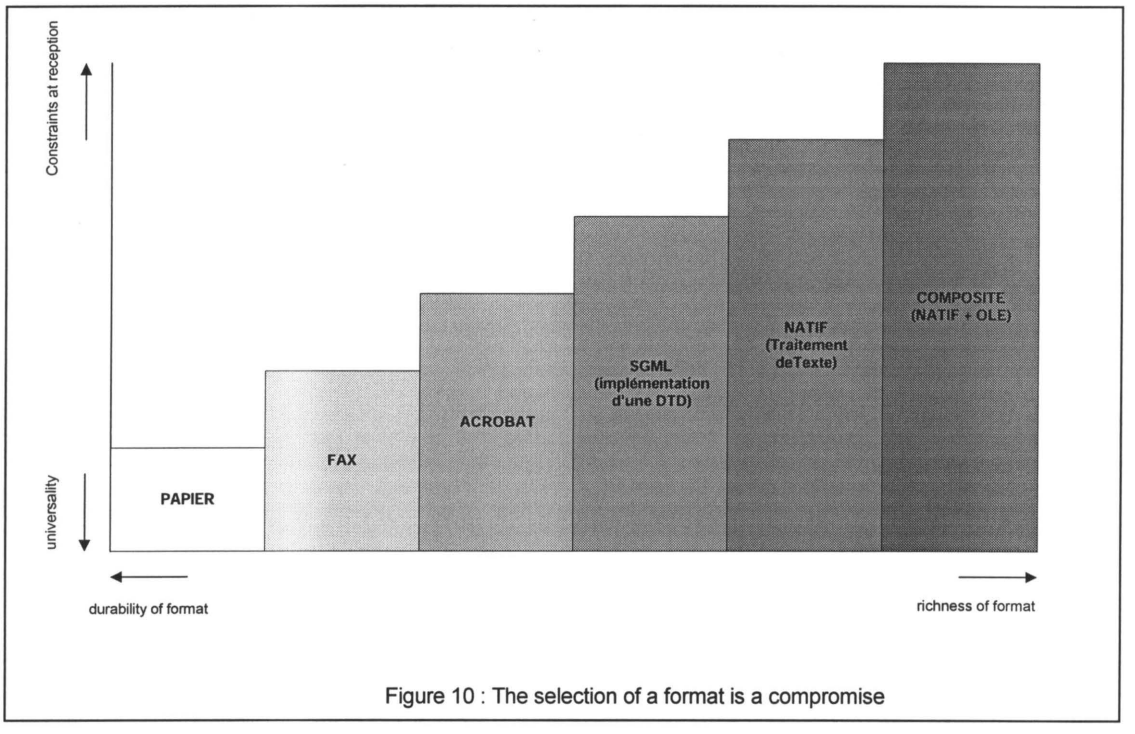


Figure 10 : The selection of a format is a compromise

For the **revisable electronic sending** of**text-type** documents, the reference format selected is Microsoft Word 6 / Word 7 as customised for Eurolook, Word 97 becoming the target format at the end of the office automation migration. For the electronic exchange of more complex documents comprising a text, tabular material and graphics, and with guaranteed revisability of the text, the proposed format will in all probability be based on OLE[10] using Word as reference host format and Microsoft Excel as host format for imported spreadsheet objects. Revisable interinstitutional exchanges are associated with procedures supported by information systems such as SEILEG (*System d'échanges informatisés législatifs*). The format used is reliant on a logical structure specific to the Commission (document table of description: DTD/SGML) to which are associated formats for representation of basic components.

---

10 Object Linking and Embedding: a subset of the DCOM architecture defined by Microsoft Corporation

The document dissemination (see section 2) will create a growing usage of the HTML format based on SGML.

But most electronic document exchanges do not require revisability. The association of electronic mail and a non-revisable format, based on a universally-acknowledged printer language, **PDF** (the de facto standard of the Internet) would be in a position gradually to supplant the facsimile.

For graphics data the transmission formats recommended for use in exchanges with the exterior are CGM (Computer Graphics Metafile), GIF (Graphics Interchange Format) and TIFF (Tagged Image File Format). CGM complies with ISO 8632. GIF is widely used on the Internet and TIFF is the *de facto* standard for bitmapping (16-colour bitmap images) with which compression algorithms (e.g. Fax CCITT Group 4) are associated.

The principal lesson to be learned from the success of fax is the current strength of a format's universality compared with its richness. This observation will probably no longer hold true once the office systems chain is complete: the need to preserve a document's richness from start to finish in order to make it reusable will require mixed solutions: transmitting both revisable formats (native formats readable using free viewers) and leaner formats which are as universal as paper but of better quality, such as Acrobat (PDF).

To these technical considerations relating to format must be added:

- the cost problems arising from the volume of image formats
- the still unclarified legal issues arising from the use of electronic formats

Pending completion of the office systems chain, the current situation relies too heavily on the use of traditional postal services and fax. These two media have a number of disadvantages:

- their basic medium is still paper, which precludes rapid transmission and allows of virtually no further processing of incoming documents;

- they are not easy either to automate or to integrate. For this reason their acquisition should be contemplated with great prudence.

SECTION 2 : DOCUMENT DISSEMINATION

1. INTERNET AND THE WORLD WIDE WEB

Using TCP/IP-based protocols, the Internet network permits world-wide interconnection of a growing number of

- decentralised servers disseminating essentially unregulated information
- workstations with access via navigating tools called browsers.

The information loaded onto the network is organised in the form of files containing hypertext links to other information regardless of their location. These links weave a fabric providing an intelligent link between such information, hence the expression "World Wide Web" or simply "Web".

In addition to the Web type traffic, the Internet network carries file transfer protocols (FTP), electronic mail protocols (SMTP + MIME), and other protocols such as newsgroups (NNTP), etc.

The Web is organised around the following elements:

on servers:

- a data description language (HTML: HyperText Markup Language) based on SGML which describes a simple document structure consisting of pages with hypertext links

- HTTP (HyperText Transfer Protocol), a TCP/IP-based communication protocol optimised for searching out information with hypertext links :URL (Universal Resource Locator), an addressing system enabling a document on the Web to refer to others

- an access interface : CGI (Common Gateway Interface) is an API allowing a server to launch the execution of programs triggered by an HTML page

on clients, a simple navigation tool called a browser which enables the user

either to read HTML documents in page form

- providing searches via HTML links and URL addressing

- formatting documents downloaded by HTTP protocol.

or to trigger processes on servers, for example after filling in simple forms embedded in a HTML page.

2. INTRANETS AND ENTERPRISE-WIDE-WEBS

An intranet can be defined as the implementation of Internet technologies inside a local domain or the common domain. A EWW (Enterprise-Wide-Web) consists of deploying Web technologies over an intranet infrastructure An intranet is distinguished from the Internet by the homogeneity of its user population and its manageable network service quality.

An intranet provides services of three types :

- document publishing via EWW
- information systems support by EWW: awaiting the inclusion of transaction management inside the EWW infrastructure, the supported information systems will be simple or non critical (see section 4 of chapter 3). The market is evolving rapidly to fill this gap.
- team collaboration : built on an electronic mail infrastructure (see section 4 of chapter 1) and EWW. Section 3 of chapter 2 outlines the groupware components used in the Commission.

3. EWW TYPICAL ARCHITECTURE IN THE COMMISSION

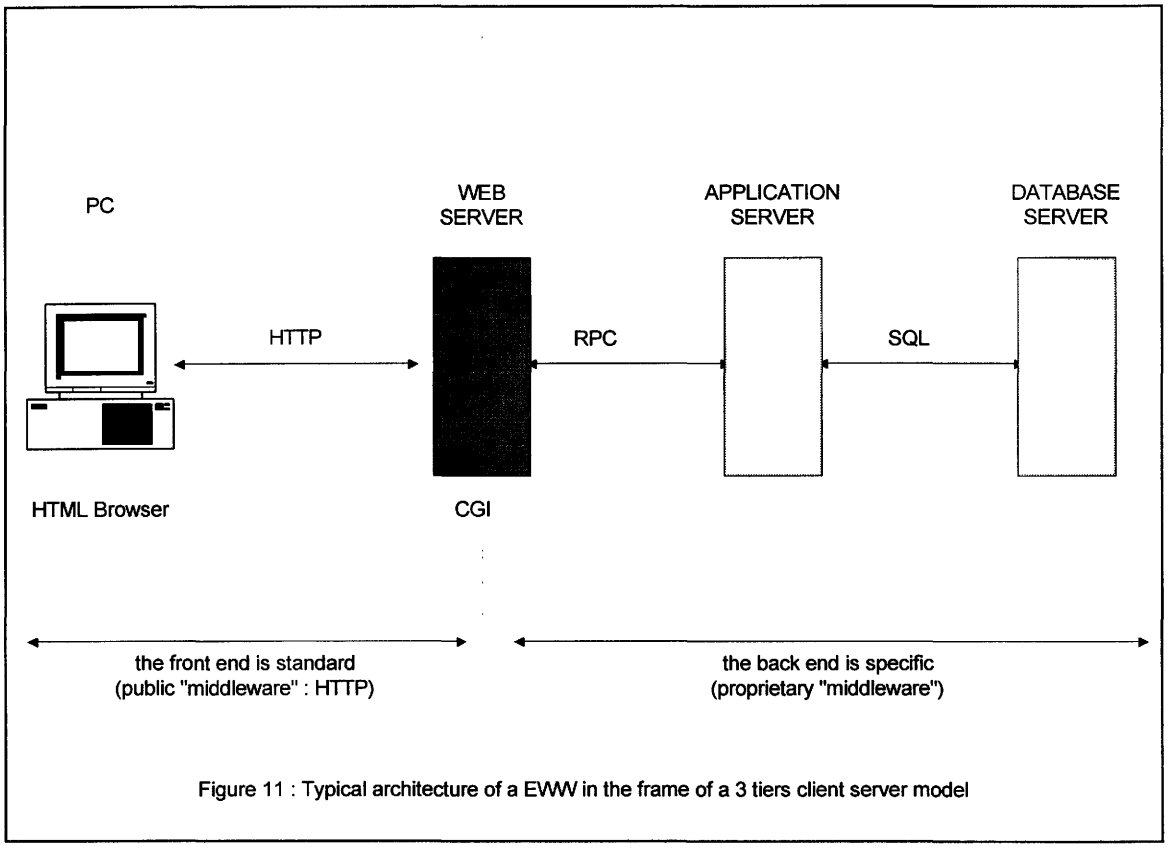


Figure 11 : Typical architecture of a EWW in the frame of a 3 tiers client server model

The architecture of an EWW inside the Commission is a 3 tier client server broken down in two functional parts (see figure 11) :

- the front end is made up of the standard Web elements : an HTML browser and a Web server supporting CGL, linked together via the HTTP protocol

- the back end is made up of applications servers (containing processes in the form of objects or programs) and database servers (containing the RDBMS). The Web server acts as a really between the front end and the back end. The actual code implementating this relay function may also be present on the application server.

The implementation of the 3 tier client server model by an EWW aims at an optimal deployment via the definition of simple and standard client and exchange protocols. The success met on the market is evolving this combination into a universal client server infrastructure.

The EWW client inside the Commission must remain simple in order to fulfil its universal function. Nevertheless, a minimum set of functions are necessary such as the use of forms and an ergonomic look and feel. The EWW client is made up of

- a browser supporting at least HTML 3.2

- a script language JAVASCRIPT

- and browser extensions in the form of JAVA applets for the functions not supported by the script language (format conversions, HTML page animation). The JAVA applets are mini applications written in JAVA language, easy to download, and running on all equipment where a JAVA virtual machine is installed.

This set constitutes the de facto EWW client.

The use of alternatives to JAVASCRIPT and JAVA (such as VBSCRIPT and ACTIVE-X from Microsoft corporation) are possible so long as their deployment is limited to an intranet whose infrastructure is made up of products supporting these functionality alternatives. A higher level would justify a choice diverging from the de facto standard. Implementation guidelines are included in the technical guidelines associated with product selection.

The use of extensions such as plug-ins (applications associated with a browser) are discouraged . Indeed, these plug-ins are too specific to a browser product and their download is too complex.

The EWW server inside the Commission normally runs under WINDOWS NTS and is selected on a price-performance ration. It supports HTTP and CGI as well as a distributed infrastructure based on JAVA in order to ease developments while remaining universal.

---

[11] The architecture of a 3 tier client server is specified in chapter 3, section 4.

Other API (Application Programming Interfaces) are more efficient than CGI, but their acceptance is still only partial on the market. Their use can be foreseen in the particular case of an intranet whose deployment would be limited to an infrastructure supporting these API.

A page set up software is associated with the selection of EWW server.

4. SECURITY

For the purpose of Internet access, the telecommunications centre isolates the common network behind a firewall by filtering IP addresses and carrying out a preliminary check on the protocols used. The firewall is an effective security measure, but it is not foolproof. The firewall depends on a network topology which assumes no direct short-cut access (modem connection). Furthermore, it provides no protection against a user downloading a potentially dangerous application (virus etc.).

It should be recalled that, pursuant to Commission Decision (C (195) 1510 of 23 November 1995), any user making unauthorised modem connections or secretly downloading software faces disciplinary measures.

To make the Web infrastructure secure, the SSL (Secure Socket Layer) located between HTTP and TCP provides a two-way authentication service and guarantees the integrity and confidentiality of messages. For more information, particularly on the use of secure cards and password management, see the Information Systems Security Architecture document issued by the Security Office. The fundamental problem with these solution is bound up with legal constraints on the use of cryptography.

5. EVOLUTION

An open consortium, the W3C (World Wide Web Consortium) was set up to maintain or extend the interoperability of the services available on the Web. The W3C seeks to define standards to specifications which are in the public domain. The domains concerning security and the multilingual aspect of the Web should be underpinned by such specifications. The upgrade of the HTTP protocol in terms of throughput and transaction support is a priority.

The W3C is a useful forum for building a richer client server model definition while preserving its universal applicability. Such a client server model is bound to become an integrator of all the IT services provided to the end-user.

6. EWW IMPLEMENTATION GUIDELINES

In addition to architecture guidelines described here above, the implementation of an EWW in the Commission is subject to :

- technical guidelines concerning product selection (see product list)
- information management guidelines, defining the type of information to publish and the mandates associated with these publications.

1. INTERPERSONAL MESSAGING

Interpersonal messaging is often confused with electronic mail because it can be transported by the electronic-mail infrastructure (see Figure 4). In practice, interpersonal messaging is a sub-set of electronic mail: the messages sent are informal and the format is undemanding; they are electronic "post-its".

SMTP (Simple Mail Transfer Protocol) is the *de facto* standard for interpersonal messaging, and is far more widely used throughout the world than X400. Internet browsers offer easy access to SMTP.

The universality of the SMTP protocol and general access thereto by browsers will make interpersonal messaging a low-cost tool of informal communication. Within the Commission, its use is more organised than on the Internet and it is integrated into the electronic mail infrastructure by shared mailboxes and access to common directories.

2. WORKFLOW

Another use of electronic mail arises from workflow software, for which the market is growing. Via an API, it uses the general mail service or accesses document servers. It offers the option of planning the circulation of electronic documents and monitoring their progress. This enables users to know the status of their file at all times and to respond to any bottlenecks. Workflow software is often associated with the EDMS infrastructure. It is a tool kit which simplifies the planning of dossier circulation and deadlines. Unfortunately, these products are still not standardised (model definition and API), and the benefits they offer are still limited: implementing them imposes constraints on the definition of tasking, of roles and of staff allocation.

There is also procedural workflow, built around mechanisms for storage and database access. This is part of the thinking behind the information systems described in Chapter 3: Common Information System (SIC), for example, provide workflow for managing missions.

The introduction of these new workflow tools has to be negotiated with users, because they always demand some change in current working methods. The organisation for implementing them often resembles that required for information information systems. Authentication and notarisation mechanisms (registration of the exchange by a third party) also have to be set up to locate these new electronic flows within an appropriate legal framework.

3. MEETINGS SUPPORT

The Commission's administrative role is not confined to exchanging or sharing documents; it also has to provide effective management of meetings.

The technologies underlying the products applied to this functional requirement are a broad church.

They can be assigned to two groups:

- shared diaries, logistical tools enabling remote consultation and scheduling
- videoconferencing, currently provided in specialist studios, but which will be technically possible on PCs as the SNet takes shape.

4. DOCUMENT SHARING

A single office automation document (schedule, report etc.) can be shared by software running on local file sharing services. If the user group goes beyond the bounds of a local domain, it will be necessary to use a messaging infrastructure or Web-type access to synchronise processing of the document between the local domains involved.

Document sharing refers to generic tools which are more ergonomic than simple file sharing. As groupware tools, their implementation requires little organisation because they are introduced more as add-ons to existing tools than representing a fundamental rethinking of working methods.

The service offered should not be confused with that of a sophisticated storage system like EDMS (see section 1), which may involve as many as three database systems: one to manage the optical disks, one for the indices and one for full-text searches.

5. DEVELOPMENT TOOLS

Groupware tools include development tools for end users. These are an extension of the software chosen in the Commission[12]. They enable utility programmes to be run for local requirements specific to each Directorate-General.

These tools are less complex than development tools for institutional systems (4GL) and fit easily into the office automation/systems infrastructure. Considering how ill they sit with the network (performance) and the data processing infrastructure (local information is not coherent on a larger scale), their future is clearly in the development of small personal databases.

It should be noted that these tools are a growing presence on the market:

- they can serve clients as small relational databases located on a server
- "file locking" is better managed on a uniform platform.

---

[12] These tools are Access (PC database manager) and Visual Basic for Applications as the local general development language.

6. THE ARCHITECTURAL SHORTCOMING OF THESE COLLECTIVE TOOLS

While workstation-based office systems have been greatly simplified with the appearance of very robust heavy-volume products, groupware is on a rather less stable footing. There are several reasons for this:

- the disparate technologies used;
- the difficulty faced by tools **which want to stay simple** in coping with the diversity of network features: quality of service, interface support or no, cost;
- the need for mass dissemination, which presupposes binary portability. This need is reflected in the transfer of maximum functions to PCs (binary portability) at the expense of UNIX servers (which are only portable at source level). Many of these groupware tools thus pose security and performance problems in widespread use, pending reconstruction for Windows NTS servers.

CHAPTER 3 :

INFORMATION SYSTEMS

1. LOCATING INFORMATION

1.1 Links between objects

To represent structured information is to create links between different objects. By contrast with the Web, which also manages links (and thus resembles information systems), it must be possible to create and modify these links very frequently. Thus, to create a "mission" in a mission information system is to create links between the individual who is going on mission, the cities to be visited, the means of transport to be used, a financial allowance to be used and a series of authorisations to be obtained. It must be possible to identify the objects thus linked quickly and uniquely.

To achieve this objective, one solution is to store these objects in one place, a central database containing all the objects with which the institution as a whole deals. The single location guarantees that they are unique and coherent (at least if the base is correctly structured). If the objects are physically close, links can be created quickly and at less cost.

Whereas many organisations such as banks opt for this solution, the Commission usually does not.

1.2 The problems of a centralised approach

First and foremost, in any type of body, concentrating all information at one point poses considerable problems of performance and, to a lesser extent, of resilience and security.

In particular, however, the Commission is highly decentralised. Many Directorates-General process information of concern only to themselves or to a limited number of other directorates. Even when information concerns several Directorates-Generals, they do not use it in the same way and will therefore tend to structure it differently. Even when a Directorate-General is required to give certain data a strictly identical structure, the organisation set up to manage it often differs from one Directorate-General to another, and the use they make of that information can vary greatly.

This is further complicated by problems of physical location. The Commission is organised at several locations. Its Computer Centre is in Luxembourg, while most of its Directorates-General are in Brussels. Locating all its data in Luxembourg would entail considerable extra costs in terms of network capacity.

The architecture adopted by the Commission is therefore based on locating information in the Directorate(s)-General responsible for managing it[13]. This approach enables the Directorates-General to discharge their responsibilities while themselves defining priorities in the construction and implementation of their systems.

1.3 The risks in the decentralised approach chosen

Decentralised architecture does, however, entail the risk of inconsistencies in reference to common objects. It may also lead to waste if several Directorates-General have to manage identical or similar objects and therefore develop systems independently. It also tends to compartmentalise development teams thus preventing the creation of a common approach to structuring data and constructing systems.

Lastly, because a single Directorate-General is rarely exclusively responsible for managing any kind of information, it gives rise to problems in circulating information within the Institution.

2. CLASSIFICATION OF SYSTEMS :

To reduce the risks associated with a decentralised approach, three categories of information systems have been created. Each category poses the problem of decentralised information management slightly differently. They are:

- **administrative systems (SA),** which deal with objects like missions or payments which are common occurrences in most Directorates-General. This document describes the architecture of these systems in greatest detail;

- **operational systems (SO),** which enable a DG to deal internally with objects peculiar to it, such as fishing quotas or export licences. Because these systems are peculiar to one DG they raise relatively few problems of decentralised data management. On the other hand, to facilitate their development and maintenance, and possible integration with certain administrative systems with which they are to be associated (e.g. financial systems), they must comply as best possible with the architecture described in this document;

- **inter-institutional systems,** in which Member States or other institutions are jointly responsible with the Commission for managing certain information (e.g. monitoring goods in transit or the exchange of VAT information). The decentralisation of information is much more difficult to manage in this case because of physical distances, different infrastructures and the lack of any harmonisation in the construction of systems.

---

[13] In some cases, it may be located in the Computer Centre, which then provides data management (backup, fall-back machines, SIC area, etc.). The responsibility for implementing and operating management systems nevertheless remains with the Directorate-General which has to manage the data.

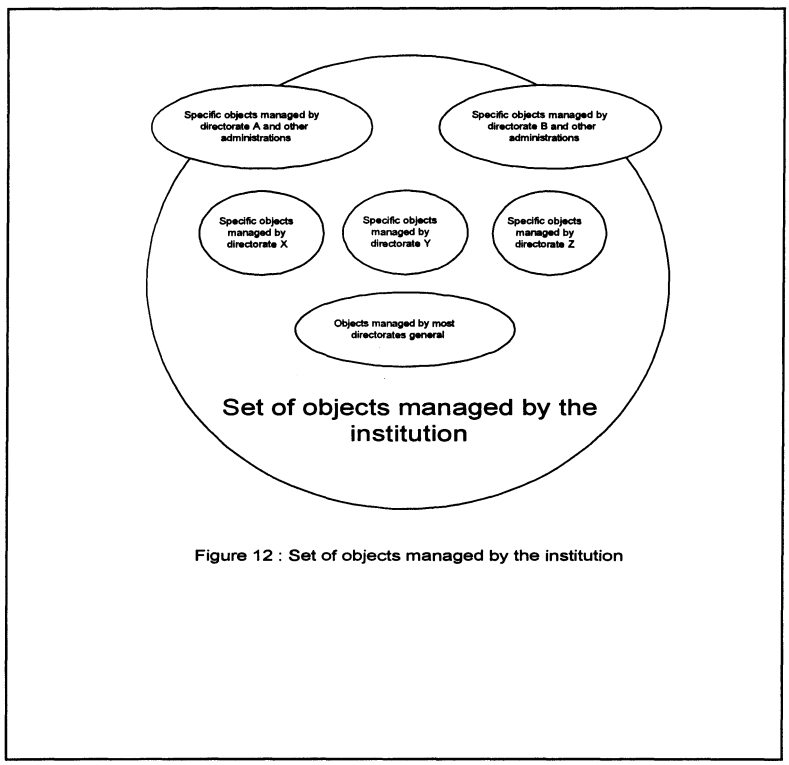


Figure 12 : Set of objects managed by the institution

Classification depends on responsibility for managing different objects:

- objects managed by most Directorates-General;
- objects managed by each Directorate-General;
- objects managed by the Commission and other administrations.

Needless to say, there are no rigid boundaries between these three categories.

3. SAI AND SIC

For administrative systems for manipulating objects which occur in most Directorates-General, there is usually a "leader" Directorate which is responsible for guaranteeing consistent management throughout the Institution. It is usually a so-called 'horizontal' directorate, like DG IX (Personnel et Biens), DG XIX (Finance) or the Secretariat-General (documents). This Directorate-General defines the management rules to be observed. Directorates-General are free to organise their local management autonomously provided that they observe these rules.

3.1 SAI

Each "leader" directorate has an Interinstitutional Administrative System (SAI) for its management. This system lays down the reference standard for the objects it manages. In principle, the SAI is implemented directly by the Directorate-General responsible (occasionally with other directorates which are closely involved, as in the case of the management of commitments and payments, where DG XIX - budgets - and DG XX - financial control - both have access to the SAI).

An SAI always provides an external interface enabling information to be exchanged with other management systems. The structure of this interface is described in subsequent chapters. In addition, a SAI may also have a user interface affording other Directorates-General direct access to SAI. This interface enables them directly to create and display objects in the central system.

3.2 SIC

Directorate-Generals which are not "leaders" use Common Information Systems (SIC). These are centrally-developed local systems which are implemented in the Directorate-Generals and run by them on decentralised terms. SIC are designed to be adapted to Directorate-Generals' specific requirements. They usually run on a single local database in the DG and have a similar interface regardless of the objects managed. Because they work on a single database, they have a coherent picture of the information. (Occasionally, for reasons of performance, however, a Directorate-General has to use several replica databases instead of one central database.)

SIC and SAI communicate via interfaces defined at SAI level. An interface committee is responsible for changes in the co-ordinates of these interfaces. One Directorate-General's SIC never communicates directly with another Directorate-General's SIC. This restriction enables access security to be guaranteed.

SIC-SAI architecture has two extremes:

- on the one hand, a DG may have no interest in specifically managing certain objects, but may be responsible for providing information. This is so with the management of the furniture or the financial management of some small DG, for example, where there is no local system and the DG interacts directly with the central system.

- On the other hand, management of certain types of objects may be completely decentralised. In this case, a central mailbox mechanism may be needed to enable different SIC to exchange information.(e.g. when an official moves from one DG to another and records must be transferred from his former DG to his new one.)

3.3 Workflow

In both SIC and SAI, information has to circulate from one individual to another for approval (controls) or simply because different parties contribute information concerning an object.

- Within a SIC (or SAI) workflow provides access to information located on a single database. This access flow is transmitted from user to user. Generic sub-systems permit such access flows to be easily integrated into any information system.

- Between the SIC and the SAI the "workflow" is an actual data exchange using SIC-SAI interfaces.

This easily implemented structured information workflow often needs to be associated with a document, text or image flow. Thus, for example, requests for payment have to be matched to invoice copies. This documentary workflow eats up more resources and requires heavier investment in terms of network and storage capacity and a scanner. Instead of routing these documents through the network, it is often preferable to store them near the point where they come into the system by scanning them and making them accessible on request to users who need them for an approval procedure.

4. SO (OPERATIONAL SYSTEMS)

Operational systems involve fewer data to be shared within an institution. They come under the direct responsibility of DG, which determine their structure as a function of essentially historic constraints. However, new systems observe the architecture defined in this document. Such architecture:

- makes for easier integration with administrative systems (SIC and SAI) when this is necessary (in particular, many SO end up generating financial objects such as commitments and payments);

- enables generic blocks common to several systems to be compiled (like systems for managing structural funds or for managing quotas);

- simplifies the construction of systems by re-using standard components (security, for example).

Moreover, when a new SO is defined, the data model is defined so as to observe the structure used in administrative systems and thus facilitate subsequent global consolidation of input for reporting tools.

5. INTER-ADMINISTRATIVE SYSTEMS

The architecture of inter-administrative systems is much more complex because the systems are different. The differences emerge at all technical levels (networks, software, operating systems, databases) and in the data modelling.

The architecture of administrative systems is set out in a separate document produced by the IDA program. Essentially, this architecture aims to standardise and manage the interfaces between systems on the basis of the few standards in current use (X400, EDI etc.). This architecture is burgeoning currently with the emergence of widely used new *de facto* standards and the decline of proprietary systems.

1. ACCESS TO AN OBJECT TYPE

Every information system manipulates certain types of objects. Thus, for example, a financial information system manipulates commitments and payments; a personnel information system manipulates individuals and jobs. These are the objects managed by the system, and it can create new objects, create links with other objects and, very rarely, destroy the objects under its management.

The objects which a information system does not manage but which it needs to establish links are "external" to the information system.

In principle, an object is managed by one single information system only, which is the sole means of managing the object. To guarantee this exclusiveness, every object knows which information system is authorised to modify it and will prevent any other system from updating it.

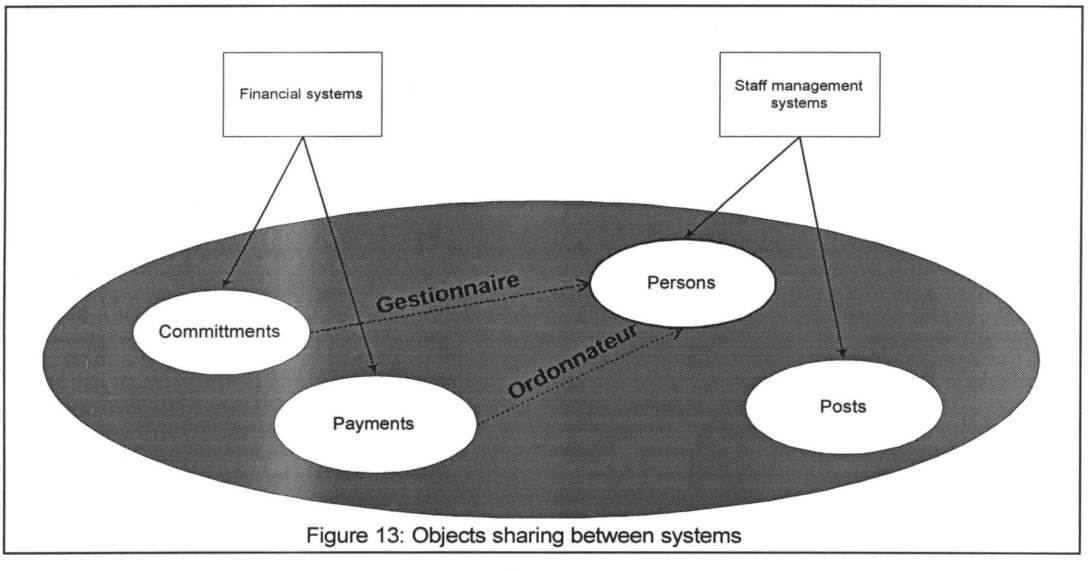


Figure 13: Objects sharing between systems

Because an object can be modified by only one single system, there is no risk of inconsistencies arising in updating or problems emerging if the object undergoes slight changes.

If a system other than the information system needs to create an object which is managed by the information system, it must approach it via an API (Application Programming Interface), the structure of which is described in a subsequent chapter. Thus, a system for managing financial interventions which is called on to produce a commitment will approach the financial information system to produce the commitment.

In such decentralised architecture, it is important for the objects managed by the different systems to be unambiguously identifiable.

Because SICs and SAIs can attribute each identification to an object, both systems must be able to exchange their mutual identifiers at interface level. One of the two systems must retain the association between the two identifiers (unless one of the systems adopts the same system of identification as the other).

Direct read access to external objects via a information system is entirely admissible. Such access is usually via filter views which enable the data structure to be modified if necessary without changing the image via the view and which also enable the information to be filtered so as to display only that information to which the user is entitled (security).

2. REPLICATION OF INFORMATION

Because a given type of object is managed by only one system, it is possible to reproduce the information if necessary for management purposes. Thus, if personnel is managed in one building and finance in another, it is possible to create two databases. One will be the "master" base for personnel and will reproduce its data on the other, while the reverse process will apply for financial data. It is also possible for two different databases to manage a single object, such as missions for DG located in both Brussels and Luxembourg. The principle will then be to partition all the identifiers by location and to observe the rule that an object initially managed by one site is always managed by the same site.

3. EXTENSIONS (CLASSES AND SUB-CLASSES)

The structuring of objects also aims to observe the concept of classes and sub-classes of objects. In relational implementation this means that there is at least one relational table per class or sub-class. Thus, every individual in an information system will be represented by a line in a relational table (containing his identifier). If the individual is also an official, the data concerning his status as such (e.g. the link with his work location) will be found in another table.

Both these tables (joined by a 1:1 link) will be linked by the identifier.

This architecture offers acceptable performance levels and enables extensions to pre-existing objects to be created easily. Thus, if DG XVII manages the work of nuclear inspectors, officials specialising in nuclear inspections, it is sufficient to create an additional table containing the information specific to these inspectors. Extension mechanisms at the level of the standard personnel information system will enable this to be given parameters such that supplementary data can be managed without upsetting its structure.

The data structure set out in Figure 14 aims to avoid redundant and therefore incoherent data representation insofar as possible. Thus, a city, country, region or language will never be referred to via a chain of characters which are supposed to identify them (Brussels, France, Sicily, Dutch). The reference will always be to one object in a table identified by its single, arbitrary identifier (a number). The same is true for the beneficiary of a commitment, who will never be identified as free text, but by reference to an organisation.

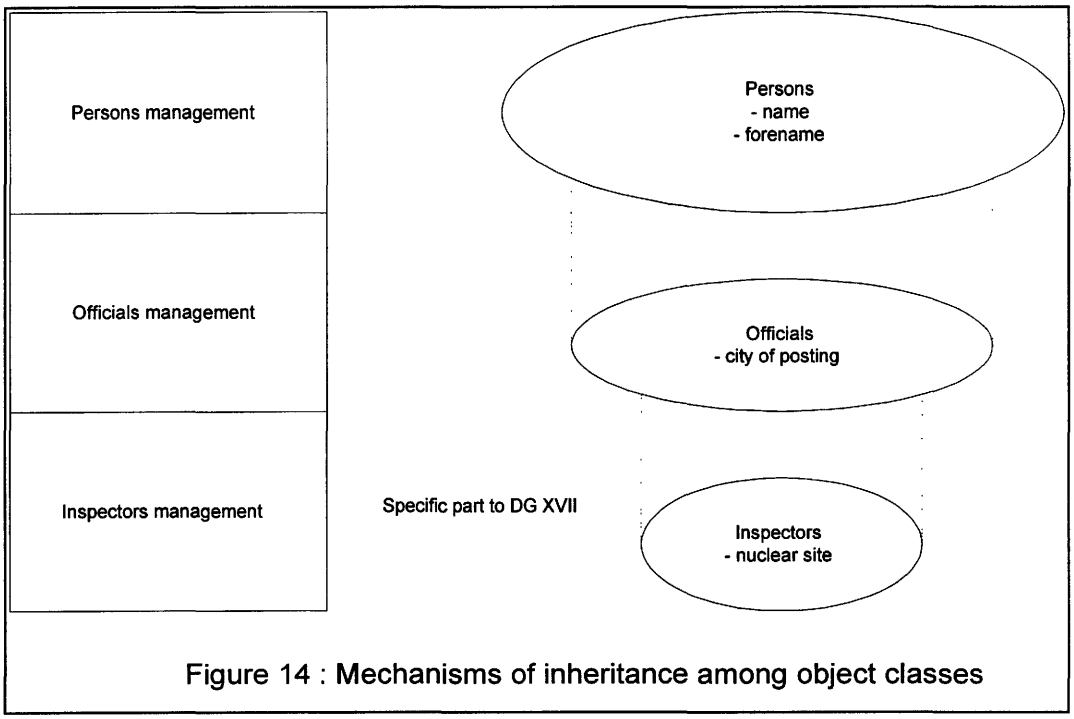


Figure 14 : Mechanisms of inheritance among object classes

Such an approach assumes that someone is responsible for updating these data. It occasionally implies certain amendments (e.g. for managing missions, in certain cases, the exact location of the mission will be given as text with a reference to the nearest city shown in the table). It is indispensable, however, if the aim is to construct systems capable of exchanging information. It is essential if the user is to be given the option of easily recovering the information by navigating through the base.

4. OBJECT REPOSITORY

To represent the structuring of information, the Institution has chosen a single tool which is thoroughly integrated in the software used for systems development. This tool, CASE (Computer Aided Systems Engineering), makes it easy to design a database and can associate the necessary integrity co-ordinates to the data. In time, it can be used to create a data dictionary (metadata), describing the data and making for easier system maintenance. **A well crafted 'meta' model can also serve to create generic systems which no longer have to be specifically programmed for each type of object they have to manage.**

5. COMMON DATA

Many systems (and SIC in particular) now use the same data structuring (the same relational tables) to represent a certain type of object. This set of common data provides for single structuring throughout the Institution of country codes, city codes and currencies, to say nothing of budget headings, individuals (whether or not they are officials) and organisations (the Commission directory, outside firms). **Such single representation facilitates data exchange, speeds up systems development and generally simplifies data management**. It can only produce real results, however, if it is backed up by strict, timely management of changes in data structuring.

SECTION 3 : THE STRUCTURING OF INFORMATION SYSTEMS

An information system is usually built around a relational database system which can be accessed via a more or less standardised variant of Structured Query Language (SQL). A program translates the behaviour expected of the information system into a sequence for accessing or updating the database.

The program which manages the objects can be divided into two parts:

- one part permits the **ongoing management of the objects** managed by the system (updating)
- one part permits **data searches and extraction (research)**.

1. ONGOING OBJECT INFORMATION SYSTEMS

The part which permits ongoing management divides into two sub-parts:

- **the part which deploys the logic associated with the coherent updating of the object (business logic) ;**
- **the part which interacts with the user.**

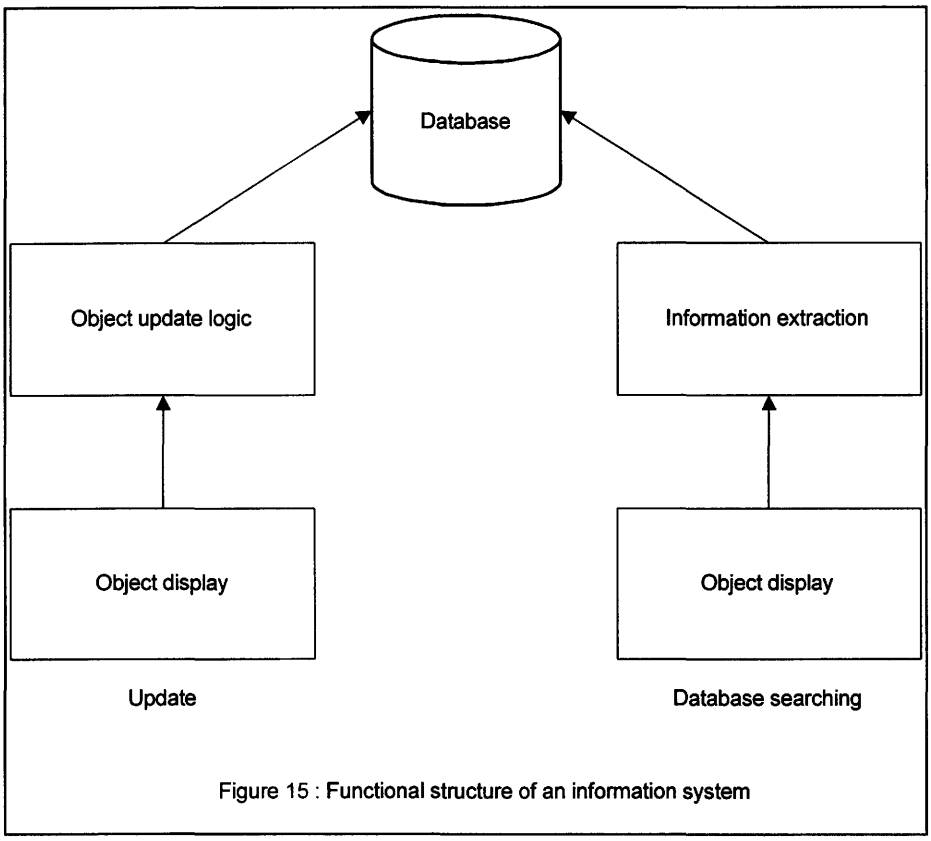


Figure 15 : Functional structure of an information system

1.1 Business logic

The part which deploys the logic associated with the coherent updating of the object (generally called "business logic") offers functions which enable the object to be updated while retaining control over it. Thus, for credit management, one function offered will be a transfer from one line to another. This function will check the transfer authorisations before effecting the transfer while ensuring that the amount debited from one account matches the sum credited to the other account.

Since this part checks the integrity of the update, it is important that it should not be possible to evade it or amend it without supervision. In infrastructure terms, this part should be located close to the server.

This part must offer its services via an API which is accessible to the other local or remote information systems. This API should be easily used by the part which interacts with the user. The API should have the same structure from one information system to the other and should, in particular, be extendible to support the division of the objects managed into classes and sub-classes (it should be possible to extend the API for managing individuals to manage officials and, in the case of DG XVII, inspectors if required).

This makes the case for implementing this API in the form of a "business object" (a management object) with the benefit of the inheritance approaches of object-oriented systems.

Each managed object will be associated with a business object inherited directly or indirectly (via the definition of other business objects) from a general object which defines the general structure of the API.

To update the information on a managed object, the information system creates (or reuses) an instance of the business object associated with the managed object. On this instance, it executes the following operations via the standard functions:

- re-initialisation of the object;
- choice of function to be executed (create, update, delete);
- provision of the necessary data (by successive calls on the same function);
- request for update.

Such a system supports a hierarchy of classes and sub-classes (and therefore an extension mechanism) for the managed objects. If one type of object (e.g. an official) is a sub-class of another type of object (e.g. an individual), the business object which implements the updating of an official will be a sub-class of the business object which implements the update of the individual.

It will use the functions available for updating the data on the individual and will supplement them by functions updating information specific to officials. This incremental approach avoids the creation of redundant codes.

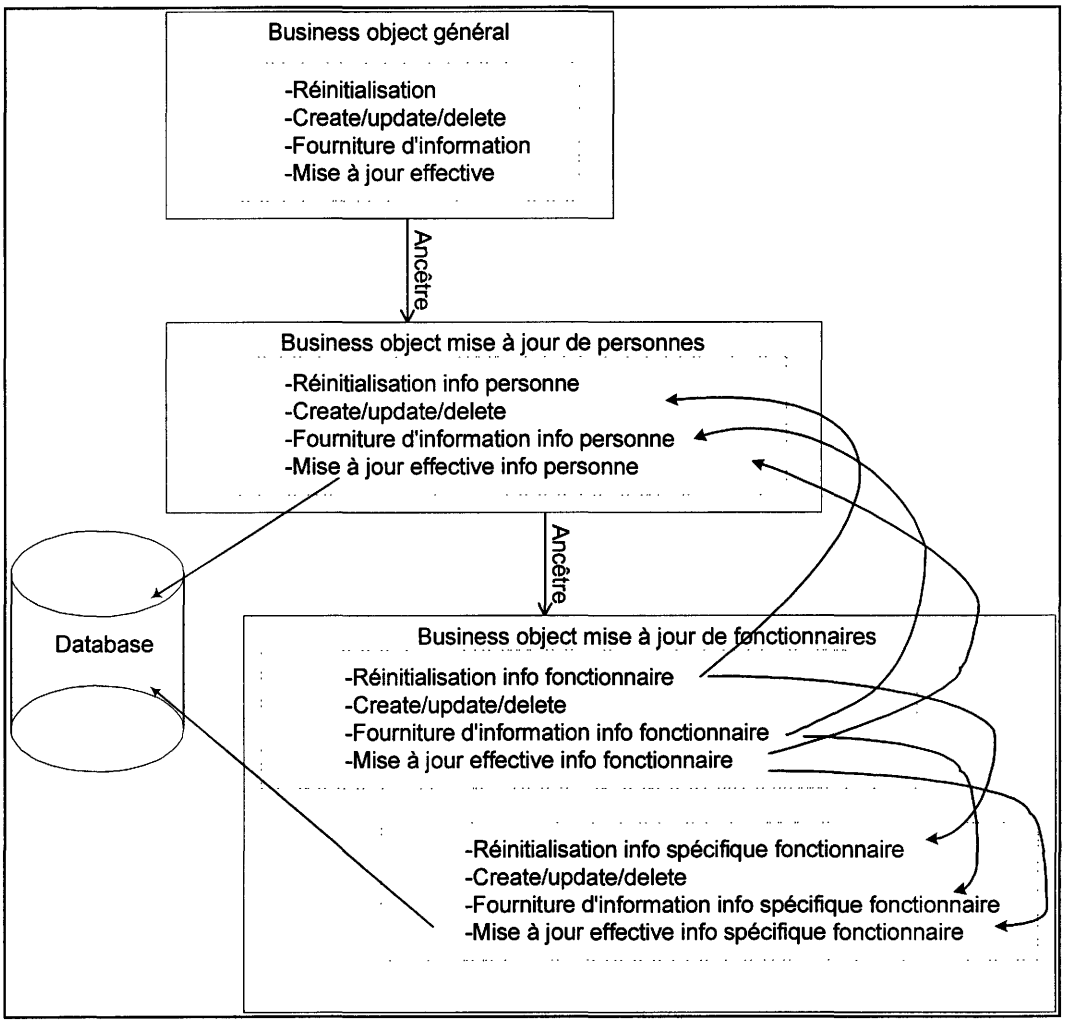


Figure 16 : object-management inheritance mechanism

The API definition should take care to permit optimum integration with the "display" part without pointless format changes.

Because a business object API is the sole location which contains the object logic, using it avoids numerous systems repetitively programming the same update procedures. **This simplifies the construction of the system and makes for an approach based on reusable components.**

1.2 Display

The display part of the object should provide a standard user interface regardless of the object managed and the data system used for management. Within the Institution, the usual model is one in which each type of object has a corresponding MDI[14] page showing a list of objects of this type at the top and the details of one of the objects managed (as chosen by the user) in the bottom part of the window. The detail is presented via a tab mechanism in which each tab contains a set of data on the chosen object.

[14] An MDI (Multiple Document Interface) page is a data display "window" which can be on screen in several copies at any time.

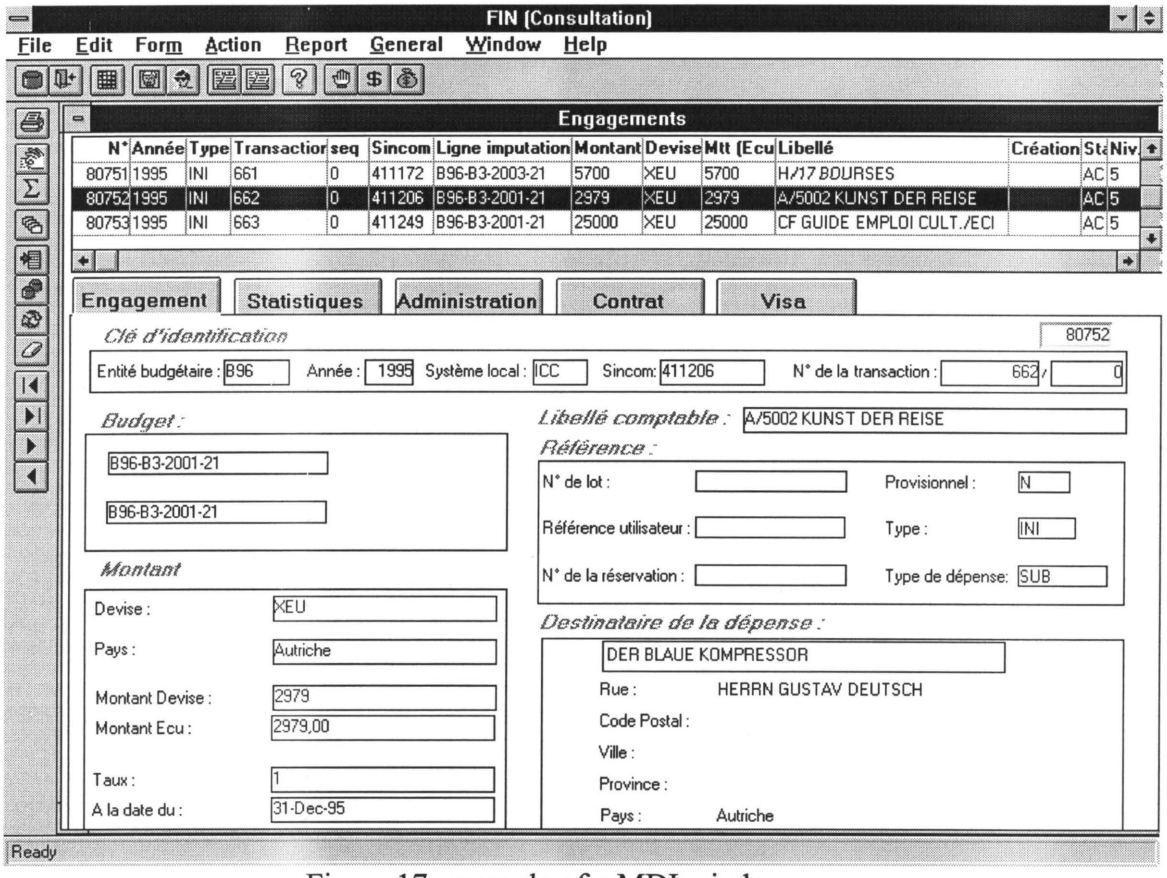


Figure 17: example of a MDI window

On the active MDI page, i.e. the one on top, actions may be performed on the chosen object (modify, delete), on all the objects on the list (sort, list), or simply on the type of object on the list (add a new object of this type, search for objects of this type according to certain criteria).

Standardising the user interface and separating off the display and business logic parts enables the display part to be further standardised and makes it increasingly independent of the information system.

Besides this "Windows-oriented" interface, there interfaces which are "Browser-oriented". They emphasise the links between objects rather then the intrinsic object properties.

1.3 Vertical and horizontal systems

This standardisation also enables sufficient sets of windows to be defined to manage a type of object. A set of windows for managing a type of object enables a user to display, search and update data on this type of object. It may be placed in dynamic libraries accessible to a set of information systems. It is thus possible to construct "vertical" and "horizontal" systems using not only the same business logic but the same display windows.

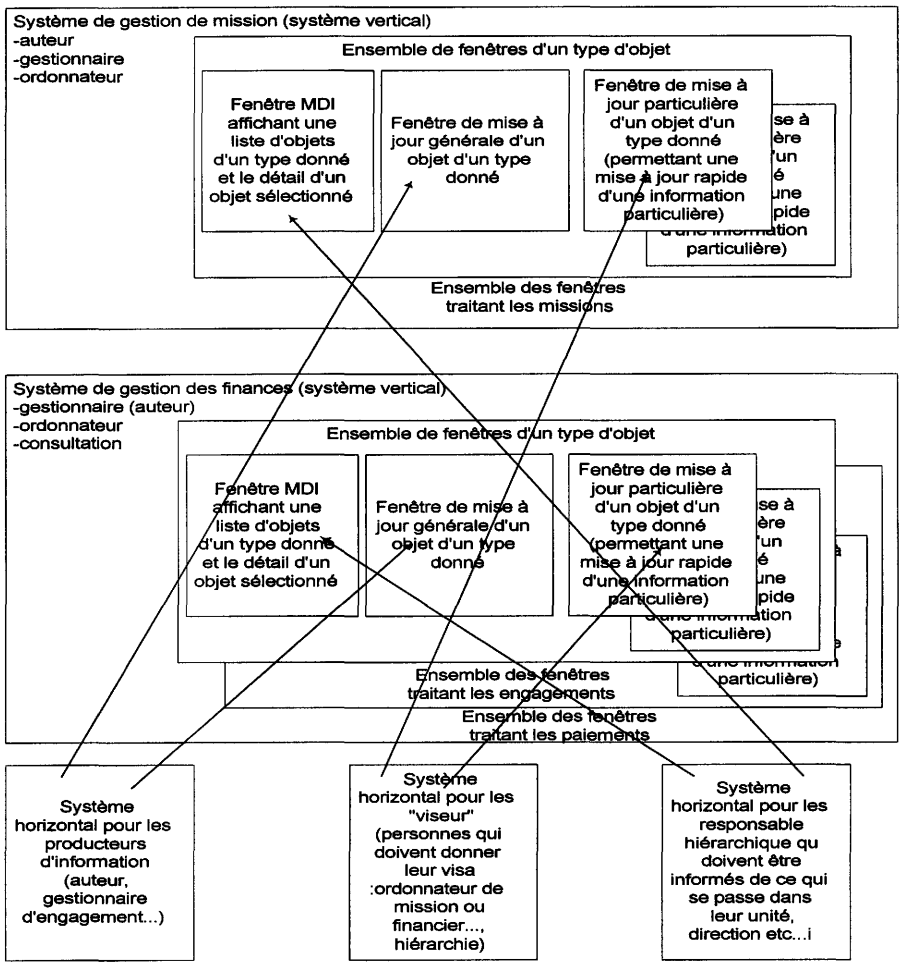


Figure 18 : Sharing MDI windows

Vertical systems, for example, will be used to manage missions, finance and leave. All these systems have screens enabling the "author" to create an object and circulate it for approval ("countersigning") before finally creating the object. Lastly, all these systems provide a hierarchical superior with a view at a glance of a unit or directorate's missions, finances or leave.

Alongside these vertical information systems, it is possible to create a horizontal system to permit an "author" to create data (on missions, or commitments, payments or applications for leave). Another horizontal system will enable the individual responsible for countersigning requests for missions, commitments or payments or applications for leave to do so without having to switch from one system to another. A last horizontal system will give a hierarchical superior direct access to all these dashboards.

To construct these horizontal systems (defined by their function rather than by the managed object), it is sufficient to create a framework to call up the different windows for managing each type of object. Such horizontal systems will even enable simpler systems than information systems to be integrated, such as systems for replacing "typewriter" standard forms with "electronic" forms.

2. SEARCH AND EXTRACTION

Although the management part for updating data always requires some specific programming which may at times be complex, the search and extract part can be handled by more general tools. The principle here is to separate the search for the object and the extraction part from the characteristics of the chosen objects.

2.1 Lists (reports)

To create lists or reports on a type of object, the "search" part of a information system will present the user with different levels of complexity :

- it may display static reports whose value is operational (such as all missions expenses to commit)
- it may display static reports specificly designed for this user
- it may allow the user to build his own research

In the latter case, the "search" part will present the user with a series of fields from which he may choose values from those in the base.

He may combine choices in the different fields by "and" and "or" and perhaps save this search. The result of such a search will be a selection of objects of the chosen type. On the basis of this selection, he will also be able to navigate to other types of objects by following the links between objects. Once the objects have been chosen, they can be applied on a print 'template' which selects the relevant information to print on paper. Alternatively, the data may be assigned to a clipboard and transferred to a spreadsheet or word-processing system or be produced in HTML format for dissemination by the Intranet infrastructure.

The search part of an information system is well fitted for the use of an universal browser (Web browser).

2.2 Data warehouses

An ongoing information system only retains the current state of information. In numerous cases, it is also necessary to record the background, i.e. links between objects which existed previously but no longer do. **When it is important to record such background and to be able to extract data from it, separate electronic storage systems - "data warehouses" - are constructed.** A data warehouse permits a user to take certain objects and "dig up" their links with other objects. It is particularly specialised in managing the time link. It also enables data from different systems which are hard to access or integrate to be consolidated in real time. The data model of a data warehouse is often different from that of the corresponding management database. Because it only holds "dead data", a data warehouse poses fewer problems of coherent updating, but it should make for easier searching.

The Commission use packages tackling the analysis of data warehouses and is carefully watching the rapid evolution of this market.

As has been said, the infrastructure supporting data storage is a relational database. At institution level, a single database software package[15] has been chosen. This facilitates management and avoids compatibility problems. Developers are asked, however, to use standard SQL for any interaction with the database so as to avoid excessive dependence on this software.

The Commission also uses a single fourth-generation[16] tool for developing information systems. It fits well into the Windows environment, is stable and is a market leader. It is essential for such a tool to be timeless because its infrastructure has a direct bearing on development investment.

The architecture obtained from the combination of the development tool and the database software is a Client/Server architecture on two or three levels.

1. TWO-TIER CLIENT/SERVER ARCHITECTURE

The first systems developed with this infrastructure were based on two-tier architecture in which the client, on the PC, contained both the update software (the business logic) and the display software. The data exchanged with the server is therefore SQL. An approach like this is acceptable for simple systems used within a single DG.

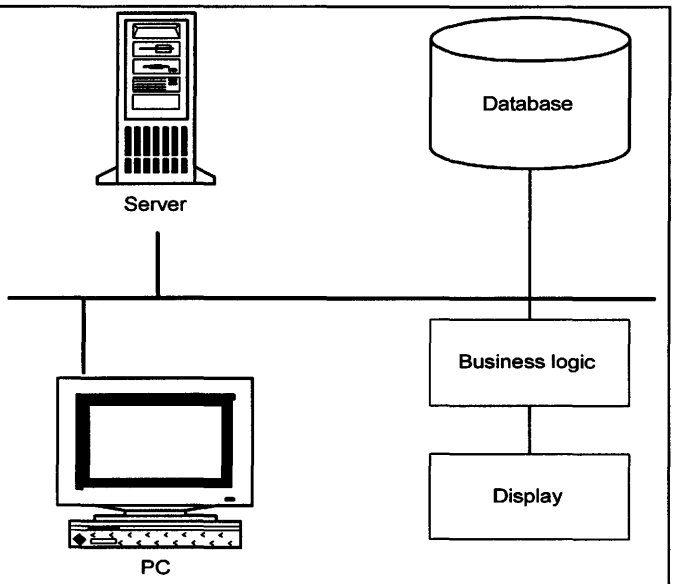


Figure 19 : 2-tiers client-server

This approach has the advantage of providing readily portable systems.

---

[15] ORACLE ™

[16] PowerBuilder ™

The server operating system is transparent, and the server can therefore be a variant of Unix or NT. The client infrastructure is standardised on the PC and any discrepancies at PC platform level are overcome by the 4GL software.

Another advantage of this architecture is that it lends itself readily to prototype approaches (RAD: Rapid Application Development) where the display and business logic parts are not strictly separated.

This advantage quickly becomes a disadvantage, however, because the lack of separation in a prototype risks persisting into the final system. The system therefore has no API for access to other systems. Furthermore, the information system coding becomes intractable. Lastly, the mixing of the display and the business logic parts means that each client of each information system has to be specific to that system.

This **two-tier approach** presents several other disadvantages

- because the update logic is on the PC, an unprotected tool, it is necessary to provide complex security mechanisms to prevent a user from being able to create and use a "ghost" application for updating without observing the integrity constraints;

- because the update logic is reproduced on all clients, as versions change, strict upward compatibility has to be guaranteed because some PC are still using an old version while others are using a new one. Furthermore, a bug in the business logic part can only really be corrected if the correction code is installed on all the clients.

To avoid these problems, in certain cases the client code is downloaded from the server every time the user asks for access to the information system. Such a solution is a heavy user of network bandwidth, however.

2. THREE-TIER CLIENT/SERVER ARCHITECTURE

The three-tier approach consists of moving the business logic part to the server supporting the database or to one or more additional servers.

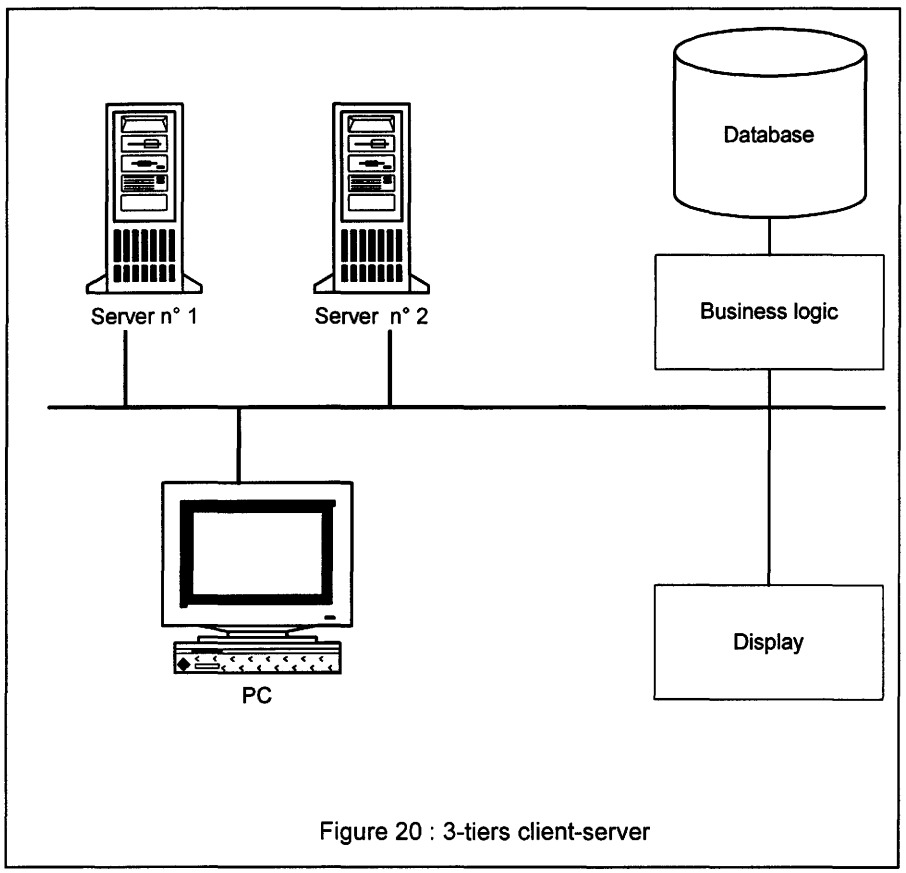


Figure 20 : 3-tiers client-server

This approach solves the problems associated with the two-tier approach, and has the following specific advantages:

- if the break between the display and the business logic parts is properly achieved, there are fewer transactions circulating on the network and therefore fewer instances of network, or even server, overloading;

- switching the business object to the server frees up the PC and does away with the need for frequent PC upgrades (CPU, memory, disk).

- Although tempting in theory, until recently this approach was problematic in practice :many information systems mixed business logic and display, and it would have been too expensive to rearrange their architecture.

- The API business logic structure was not standardised. This API, which has to circulate on the network in three-tier architecture, was not supported by widely-accepted standard protocols.

- The business logic coding on the server either dictated low-level language coding which was portable after recompilation (C type), or programming in a language supported at database level. A low-level language made the code too difficult to develop and the source portability involved considerable work on each new type of machine. The languages linked to the database were relatively limited and did not have adequate support tools. Furthermore, both approaches sacrificed the advantage offered by 4GL used to develop the client on the PC.

The situation is changing thanks to:

- better structuring of systems and advances in development tools, which can now also function on the server whether or not they are integrated into the database software;

- standardisation at server level, whether obtained via Unix-type binary APIs, unification at server level with the emergence of Windows NT (it is proposed in particular to take a three-tier approach with two servers; the application server being Windows NT and the database server Unix) or using a Java-type portable interpreted language.

3. STANDARD CLIENT (WEB BROWSER)

A well-standardised three-level architecture should eventually lead to a standard-display client PC. This promise was first held out by the X-Windows approach, but the excessively low-level interface was an unacceptable burden on the network and both server and client machines.

For search and display systems, this becomes possible again with the emergence of the Web and a client standard capable of activating a "search server" located near the database.

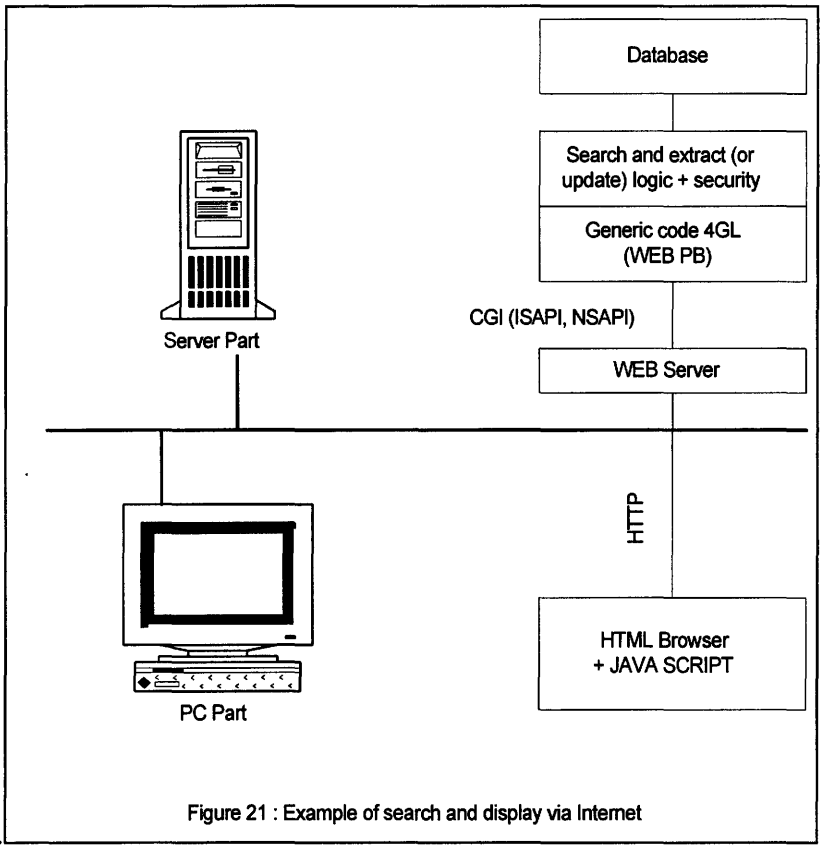


Figure 21 : Example of search and display via Internet

The search and extraction logic is still implemented in a program which can be fairly general but which has nevertheless to be given parameters for each information system. (This program would usually be written in the language used to program the information system or its data warehouse.) Instead of interfacing with Windows, this programme simply runs a generic code implementing the relay function between the Web server (the standard part of the Web server in figure 11) and the business logic programmed in 4GL (business logic of figure 20). Usually, it is also able to decode parameters given for the CGI interface (or ISAPI or NSAPI) and to call 4GL functions. Local validations are performed by the browser with script language such as JAVASCRIPT. A full description is given in figure 21, with the option of locating the relay code on the application server.

Although this approach has many advantages in terms of PC infrastructure management, it assumes that HTML page format can be sufficiently rich to meet users' requirements. It is necessary for a HTML page dynamically to fill drop down list boxes to enable the user to fill in fields, for page size to be definable dynamically, and several pages to be present simultaneously, and it must be possible to retain the table structure so as to permit "cut and paste" operations, a set of conditions which have only recently obtained.

Yet the major problem with using a Web browser arises from the non session-oriented nature of the HTTP protocol and thus from the 'stateless' approach of the server. This approach means that sessions have to be re-established on every client request, and this solution is unacceptable in performance terms for sophisticated or numerous extractions. With this approach, security controls must be repeated for each transaction.

For the present, of course, this approach is inadequate for updating data, due to the same problem of server 'statelessness'. Only in very simple systems with little data to transmit is it possible to send data to the server and to "COMMIT" in the same transaction. In more complex systems, the user must be able to create a series of objects in the database and to "COMMIT" only after a complex set of interactions. In that case, with the HTTP protocol, it is difficult to complete correctly the transaction in the event of a network error.

In the medium term, however, it is possible that HTTP-type but session-oriented protocols (HTTP-NG: New Generation) will come onto the market. While the appearance of such protocols will undoubtedly radically simplify the problems of managing a pool of PCs, it will have scant bearing on development tools and techniques. These will simply be reworked to produce their output to the standard protocol. At this point, what is needed is to ensure that the chosen 4GL meshes well with the protocols and Web. This integration already enables the Web infrastructure to be used to disseminate lists or reports: the infrastructures and the 4GL tool do not clash; they complement each other.

An important issue in the proposed Web architecture is the key role of the 4GL as an interface between the database and the Web browser. In the near future, specialised managers (for mission, leaves, ...) will continue to use specialised non-Web clients. Using the Web infrastructure should not lead to a regression in terms of programming techniques by unnecessarily forcing programmers to write in low level languages (CGI scripts) or general purpose languages (JAVA, C++,...). The 4GL tool should ensure that an application correctly written with it leads automatically (or at a very low cost) to a "classical Windows-oriented" release or a "Web-oriented" release. The tool selected by the Commission does not yet fulfil this promise but it is evolving in the right direction.

4. THE COMPONENT-ORIENTED APPROACH

The architecture described in this document tends to permit the creation of reusable components at both business object level (using the APIs they offer) and at display level (via the sets of windows for managing each type of object). This "component" approach is further reinforced by the use of standard sub-modules for managing security, countersignature circuits or the allocation of finance (appropriations). It also features in the use of skeleton common systems to speed up systems development and standardise their ergonomics.

Such a component approach is only possible if it is technically possible dynamically to link the business objects and the display objects (windows etc.). In three-level logic, the business objects should also be accessible from the network.

While operating system solutions (like ActiveX, DCOM etc.) are beginning to emerge, the tool used to construct data systems offers its own mechanisms which satisfy needs in an integrated environment. In particular, object instantiation mechanisms, which permit a dynamic choice of sub-class to be instantiated as a function of information extracted from the database, are the key to the construction of standardised systems which can nevertheless be extended to cater for specific requirements. Only this flexibility provides a way out of the "incomplete system" -

"excessively complex system" dilemma, the main reason why resources are wasted in duplicating efforts.

5. COMMUNICATION BETWEEN SYSTEMS

Different information systems are often required to communicate. They normally do so via the API in the business logic. This API permits a system to interact with another system in real time.

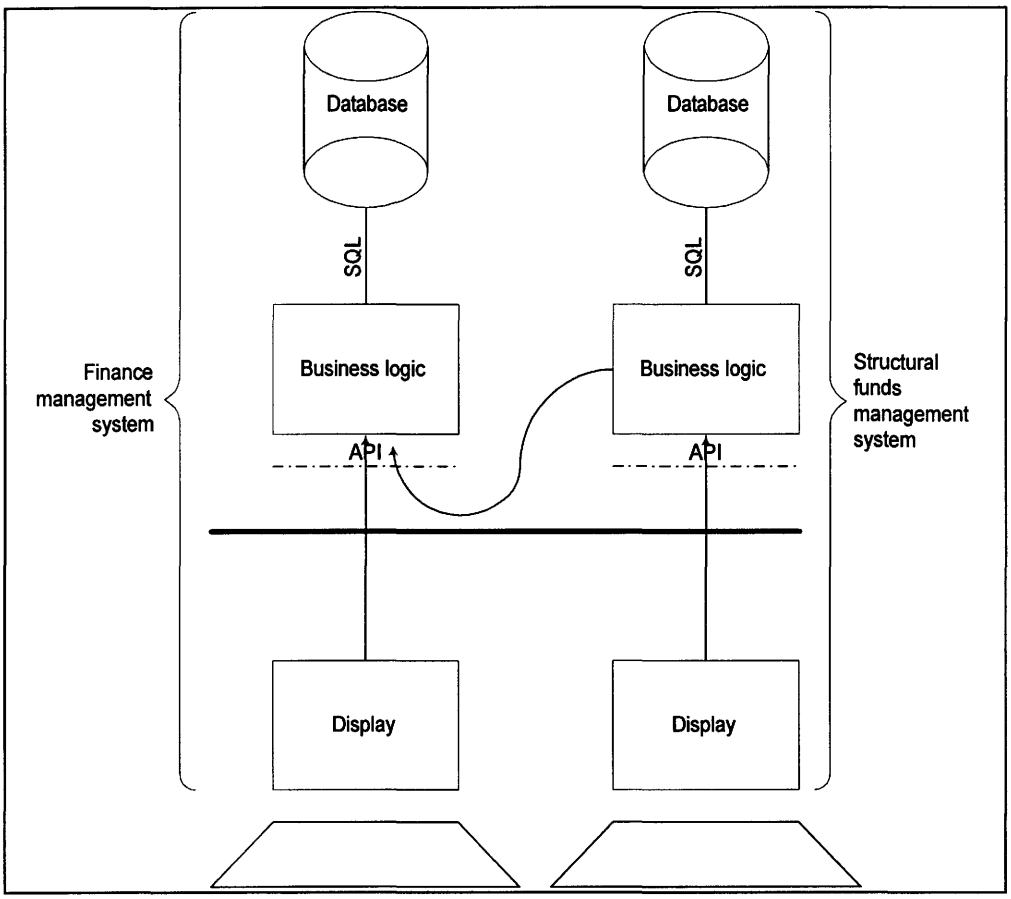


Figure 22 : Typical communication between systems

In certain cases, a information system provides a conventional API, with a series of functions. If this API is accessible via a network, it can also be used for communications between systems. One way to have such a conventional API is to use "stored procedures" associated with the database software. It is also possible to encapsulate the API provided by the fourth-generation languages to have this API act as for example methods of an object OLE2 (OCX or ActiveX) which can be called up via the DCOM distributed architecture.

When real-time exchange is not required, the exchange procedure automatically becomes more complex, because an asynchronous error code return mechanism is needed. A simple solution is to use a "mailbox" database. One system places the data to be transmitted in a structured database. The other system then searches for the data in that base. Exchanges are made by inserting, updating and deleting data in the mailbox.

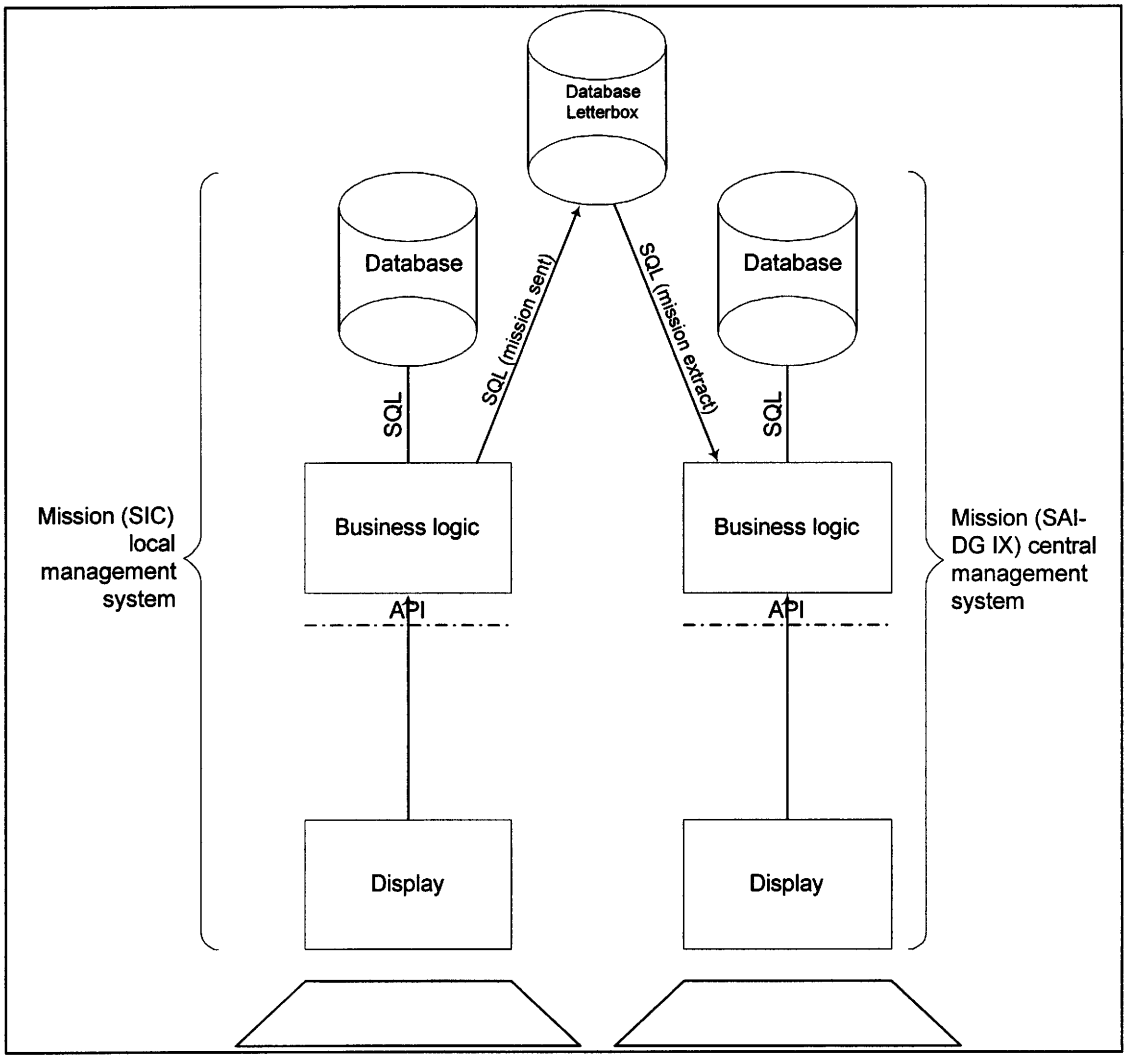


Figure 23 : Exchanges between systems via a mailbox

This approach has practically nothing but advantages provided a common protocol permits SQL on the network:

- The communicating systems do not have to interface with specific APIs.
- The SQL transport protocol provides reliable data encryption/decryption.
- The information structure may be complex.
- "COMMIT" enables the atomicity of data updating to be closely monitored.
- Integrity constraints and triggers, etc., may be used to guarantee maximum syntactical consistency of data on mailing (immediate error message).
- The system receiving the data can use complex strategies to access the data to be processed as a priority.

Another popular approach to data exchange is to use EDI messages or EDI structures via electronic mail. In this case, it is important that an API be available at client or server level. The electronic mail manages the queues for applications. It is often useful to separate data reception queues from those for receiving error messages.

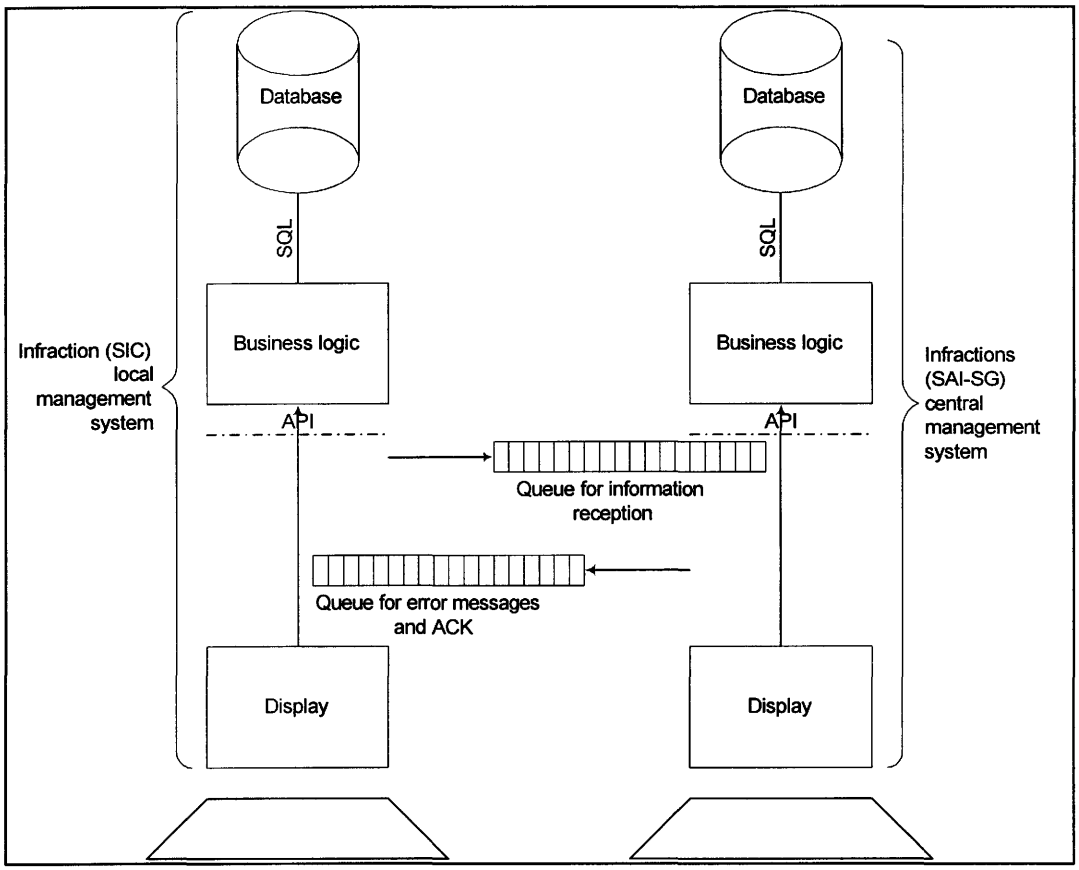


Figure 24 :Exchanges between systems via EDI messages

One problem with the electronic mail solution is the potential limitation on the size of the information which can be placed in a single message (a few megabytes).

Lastly, on very old systems, the only possible approach may be to send the file. This solution should be avoided insofar as possible because of the security problems it poses. It is also necessary to create a mechanism for alerting the destination that the transfer is complete when it has been effected.

SECTION 5 : RATIONAL USE OF APPLICATONS PACKAGES

1. THE FORCES INVOLVED

Formerly, every information system was constructed specifically for the organisation which needed it. Nowadays, different forces are pushing this approach in conflicting directions.

On the one hand, one set of forces tends to favour ever-wider use of application packages:

- Users are accustomed to using standard office software. They do not understand that, in information systems, they need expensive specific software which takes time to develop.

- Many information systems are old and impossible to maintain. They have to be given a new architecture, or simply be rewritten. Managers who know how many man-years went into developing these systems shrink from the idea of having to overhaul their entire systems and therefore naturally think that a single product could solve their problem.

- The market offers an ever greater wealth of systems for day-to day management (personnel, finance etc.). Looking at the titles of these systems and the general description of their functions, managers tend to think they should suit them. Because they are used to different office systems working together (using OLE-type techniques), they expect the same flexibility in applications packages.

On the other hand, other forces seek to perpetuate the specific development approach:

- The emergence of powerful, robust 4GLs greatly facilitates development work and nurtures hopes of halving the times needed for traditional development.

- As an understanding of IT becomes more widespread in firms, most managers, even outside the IT area, realise that the complexity of a information system is at least proportional to the complexity of the management rules to be automated. **Renewal of a system is therefore often accompanied by a procedural simplification which makes it easier to implement.**

2. THE ARCHITECTURAL CHARACTERISTICS OF THE PACKAGE APPROACH

These two opposing forces face development managers with difficult choices. Several elements have to be taken into consideration:

- A package only makes sense if the managed object is standard to different types of organisation, which is why pay or accounting packages (often country-specific) are easy to find, while the same cannot be said of quota management packages. A package can, of course, be given particular parameters to some extent but if these new parameters are at all substantial this is tantamount to programming in a specialist language.

- An application package comes with its own data structure over which the client who buys the product has no influence. Furthermore, there is no guarantee that this model will remain stable through subsequent releases, a property which makes it hard to integrate with other software.

- A package intended to satisfy a broad range of requirements is necessarily more complex to manage than software geared to a specific requirement. This complexity is reflected in the number of tables needed to implement it and the complexity of the functions provided for access. This complexity often dictates calling on specialist consultants to implement the package, and it is hard to monitor these consultants' work.

- Most of these systems come with their own architecture in terms of infrastructure. While they generally run on several database programme suites, they bring their own fourth-generation languages into the company, along with their own client software for installation on every PC. As a rule, they do not observe the systems architecture described in this document.

3. THE STRATEGY

To conclude, there is no miracle solution, and a balance needs to be struck which minimises disadvantages. The Commission uses packages to implement some complex information systems. It observes the following principles:

- Packages are used for their business logic functions in particular. The user interface is not used, or is used only for a limited user population.

- Packages are only used in "leader" DGs managing a type of object. The other DGs using a SIC have a system developed in-house and based on the standard architecture described in this document.

- Packages should offer points of entry and exit in the form of APIs. This enables the access points to be isolated and "wrapped" so that they are activated as business logic objects or via triggers from a mailbox-type database.

- Packages are associated with a data warehouse which allows data extraction with a known structure. Alternatively, they should have stable views providing a picture of the base.

Figure 25 illustrates the SINCOM 2 (financial information system) architecture.

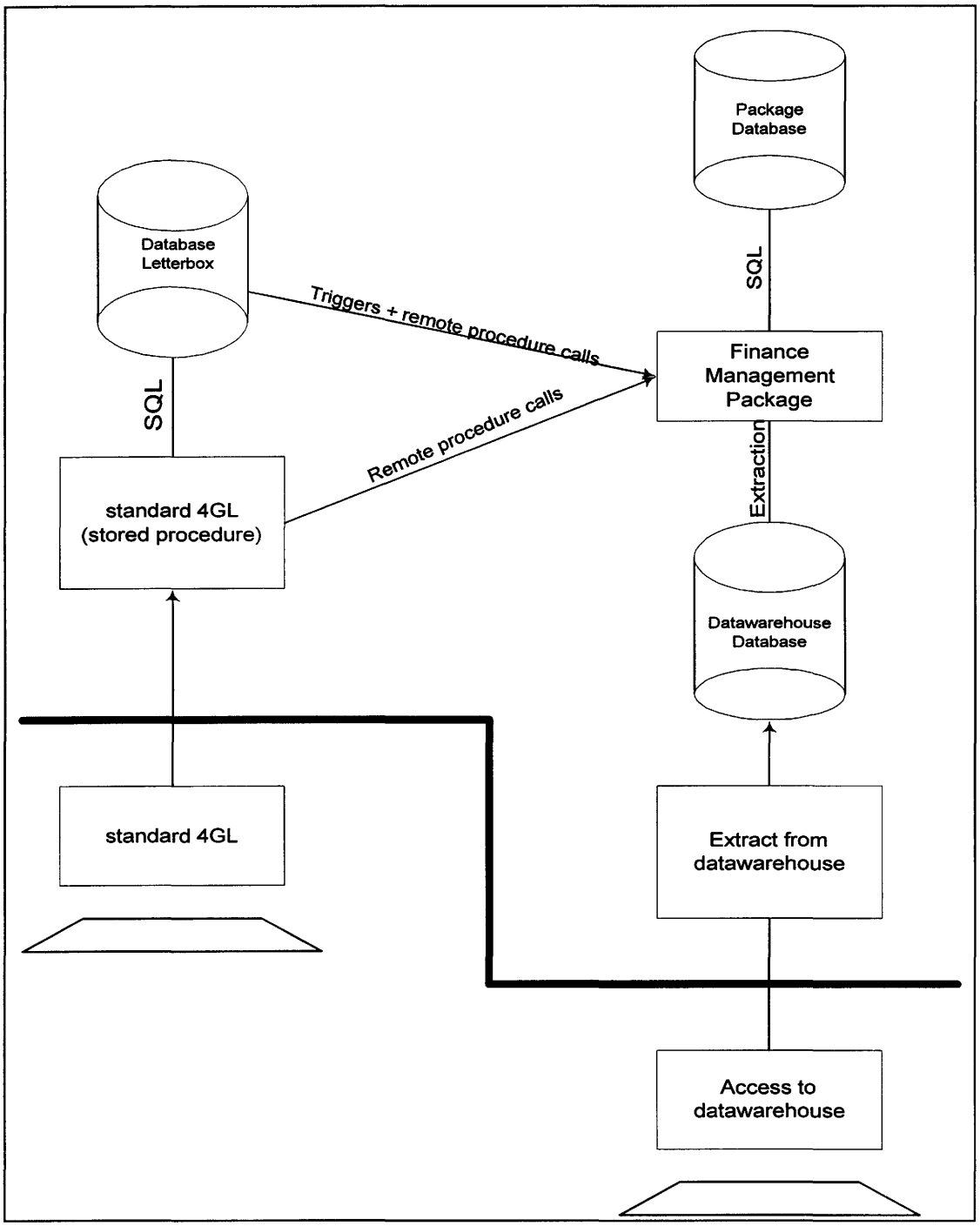
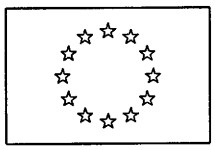


Figure 25 : SINCOM 2 Architecture



EUROPEAN COMMISSION

INFORMATICS DIRECTORATE
Data transmission service

CONNECTIVITY GUIDELINES FOR ELECTRONIC DATA TRANSMISSION WITH THE EUROPEAN COMMISSION

Addendum to the informatics architecture of the European Commission

Version 1

December 1997

Editor: G. SCHÄFER

1. OBJECTIVE ANDRATIONALE

The present document is addressed to program managers, project leaders and other personnel responsible for informatics within the European Commission in charge of community-wide information systems between Administrations.

If Directorate Generals of the European Commission (DG) are part of such information and communication systems they have to exchange data between the European Commission and external partners, such as other European Institutions and member state administrations, across the boundaries of the Commission.

It can currently be observed that every project has a tendency to implement its own infrastructure although common services of similar characteristics have already been implemented for all DGs of the Commission. This situation leads to duplication of efforts, higher development costs and higher follow-up costs during the course of a project.

This document provides guidelines as well as information on procedures, technical facilities and interfaces defined for electronic data transmission with the European Commission and external partners in the framework of inter-institutional and trans-european projects.

The guidelines should facilitate the usage of common data transmission infrastructure to take advantage of :

- Cost efficient services with the required service quality level for data transmission;
- Identification of procedures, interfaces and access points;
- Unique and coherent management of addresses and directories.

The guidelines should be seen as covering the whole range from best practice to compulsory rule. Unfortunately it is not always possible to define guidelines which cover all possible cases. The European Commission services and "trans-european" projects are simply too heterogeneous in their objectives and conditions in order to establish fully binding guidelines for every possible occasion.

In practice, a project should have good reasons which would justify an exception to one of these guidelines. Such exceptions should be specifically argued in documents, like feasibility studies or system design specifications. In the case of an exception, it is important for a project to define already the evolution towards an implementation in conformance with the guidelines.

The guidelines have been designed to be fully in line with the IDA Architecture which foresees the concepts of EURODOMAIN and EUROGATE. They describe the terms of the concrete implementation of these concepts for the European Commission. Furthermore, the guidelines of this document respect the guidelines for security as defined by the Security Office and the Informatics Architecture approved by the IRMB as valid for all aspects of informatics at the Commission.

2. SCOPE OF THE GUIDELINES

The most frequently used services are described, in particular :

- Electronic mail covering :
    - Interpersonal Electronic Mail (personal messaging and document exchange)
    - Electronic Mail Application Programming Interfaces (API)
    - EDI, Electronic Data Interchange
- File Transfer
- Internet (WWW, World Wide Web) access to information
- Interactive access, Transaction processing and client-server access

Other facilities, like general Directory Services are under study. Respective guidelines will be added to future versions of this document.

3. GENERAL PRINCIPLES

Role of the Telecommunications Centre

The Commission's Telecommunication Centre (TC) is the unique and exclusive access and exit point for electronic data transmission to and from the European Commission.

The TC architecture offers a preliminary security check at network level for incoming requests for access. It offers communications management at the point of exit from the network, particularly as regards to accounting and invoicing. The TC uses application gateways, provides management services (directories, rights of access, etc..) and protocol conversion facilities between the choices made within the Commission domain and those effected outside. In cases where technically adequate the TC also covers format conversions, e.g. back end conversion to Fax format (G3) in addition to PC based front end conversion. Details of service are subject to the evolving service definition of the TC.

The existence of the TC centrally for all DGs is justified, above all, in terms of security and by the economies of scale which it generates. These are the concentration of efforts in the development of technical solutions, the aim of a high level of availability under the management of telecommunications specialists, and the rationalisation of connections to external services.

The TC is a single logical concept despite the fact that there exist two physical instances of the TC in Luxembourg and Brussels. Both installations complement each other in the fulfilment of its coherent single purpose.

TC Quality of Service

As a unique and exclusive entry point of the European Commission telecommunications the TC is obliged to offer to the Commission services and trans-european projects the quality of service they require in order to fulfil their objectives. For that purpose the TC has to develop Service Level Agreements (SLA) for its various services which can then be used by the employing services and projects. (see the section below on the usage of the SLA). Parameters of importance in this context are availability, security, delivery times, intervention times in case of problems, disaster recovery, etc. An SLA exists for E-Mail and IDNET and is in preparation for the various services of the TC.

Exclusivity of TC access

Any telecommunication access to and from facilities within the boundaries of the Commission in Luxembourg and Brussels which would bypass the TC is to be strictly avoided. Existing facilities of this kind are to be considered only as intercept solutions on a short term basis. Any legacy systems should be phased out progressively. Any new installation is only permissible with a distinct strategy and concrete plans towards a TC supported architecture.

Scope of the Commission domain

The Commission decides on which organisational entities, called domains in the architecture document, are to be considered within its boundaries and have thus to be accessed from outside via the TC. They include the Luxembourg and Brussels sites of the Commission. Some parts of the Commission, although legally belonging to it, are located

remotely and therefore access the main parts of the Commission via the TC. These remote parts of the Commission are, for example, representations in the member states, delegations in third countries and different types of antennas of individual DGs.

Relay function

In general, the Commission does not provide relay functions for third parties of any kind. A relay function in this context is a communication between two external parties without direct implication of the European Commission. For example, messages transmitted from one member state administration to another should not use the routing facilities of the Commission TCs. Only remote sites which are legally part of the Commission may do so in a manner tightly controlled by the Commission.

There may be exceptions to that rule if the European Commission is legally obliged to provide such a relay function either on a permanent or a temporary basis.

Informatics Architecture guidelines

Projects implementing services which use data transmission functions must observe the Commission guidelines for the Informatics Architecture at least for those parts of a service which will be implemented within the boundaries of the Commission. A document describing the Commission Informatics Architecture is regularly updated. It is available from the Commission Informatics Directorate.

Preference for official services

Projects implementing services which use data transmission functions must use the official services of the Commission, at least for those parts of the communication which happens either within the Commission domain or between the Commission and external partners. For example, no project specific electronic mail (E-Mail) or file transfer installations should be allowed. This is, of course, under the hypothesis that the official service can deliver the service quality required by the employing service.

Usage of Service Level Agreement

It is imperative that a project using services of the Commission should carefully check whether the service quality parameters as expressed in the respective Service Level Agreement (SLA) of the official service are sufficient for its purpose.

If there are requirements defined in the feasibility study of a project which go beyond the service characteristics of this service, it should be checked whether those requirements can be implemented as part of the project on top of the official service. For example, for electronic mail this is possible for security features by using measures or tools in addition to those provided by the official service. Such additional measures may be less evident for other quality criteria, such as delivery times. In general, it should be avoided wherever possible to implement project specific architectures.

Project specific Service Level Agreement

An explicit SLA concerning the usage of the Commission services by the employing service should be established already in the planning phase of a project prior to any physical implementation in order to avoid expensive adjustments later on. This specific

service level agreement should concentrate on the essential quality aspect of the service important for the project.

External connection decisions

In order to provide its service, the TC establishes external connections. As a basic rule for an employing service, the usage of the connections generally provided as part of the Commission's regular service should be privileged. Total cost considerations, however, may lead to different connections. For example, E-Mail should as a default use ADMD services but in case of heavy traffic it is more cost effective to use direct PRMD-PRMD connections and functional requirements may speak for file transfer instead of E-Mail which is the preferred default service. Equally, in certain cases, a Value Added Service (VAN) may be preferable to the use of public networks.

The decision for a concrete connection have to be done in co-operation between the TC and the employing service. The TC takes a consulting role but it remains the final responsibility of the service to decide which connection will provide the required quality for its purposes.

Product management

Projects implementing services which require software or hardware in addition to what is available in the concerned DGs of the Commission has, whenever possible, choose amongst the products in the Commission's informatics product list. Exceptions have to be agreed in co-operation with the responsible Commission service, i.e. the product manager of the product group concerned. These exceptions must be treated according to the Commission's informatics product management procedures.

Usage conventions of services

A project that is implementing services within the boundaries of the Commission must adhere to agreed usage conventions, such as addressing or naming conventions. Exceptions to this rule can only be allowed if there is an absolute incompatibility with the service convention.

Application Programming Interfaces

The Application Programming Interfaces (API) recommended by the Commission should be used within projects for the interface between the application and the transmission services, at least for the parts of the service implementations to be installed within the Commission. Wherever possible project specific "middle ware" developments should be avoided. Repeated experience shows that also attempts to develop some kinds of horizontal multi-project interfaces did not succeed to get the acceptance their authors had hoped for. The required adaptation of such middle ware to the evolving technical and market context as well as the support and documentation efforts involved make it prohibitively expensive in the long run.

Exceptions to this rule are allowed only in cases in which a intra-project harmonisation takes absolute precedence over the harmonisation between Commission services and where the usage of the Commission recommended API(s) is not feasible.

Transmission costs

The transmission costs are to be paid by the calling party as a general rule. Normally this is the party which sends the data. Infrastructure costs, such as initial costs required for setting up the service, as well as fixed base costs, like leased line charges, should be shared by the participating administrations on the basis of traffic volumes unless taken in charge by IDA or other specific budget lines. A service arrangement is to be established in the SLA for the project.

The Commission may, in exceptional cases, be legally obliged to take over the charges for a certain project in violation with this principle. The legal basis for such an exception has to be well indicated in the project feasibility study.

Interworking tests

A project with parts within as well as outside the Commission is itself responsible for the inter-working tests via the TC. These tests are done in co-operation with the Informatics Directorate.

Real-Time versus Store-and-Forward

A project has to carefully examine the required quality in terms of the certainty that the communication has to be completed within a certain time limit. It should be noted that, in principle, any store-and-forward based technology, such as E-Mail, cannot guarantee a definite maximum delivery time if the connection is established through servers in addition to the origin and destination server. Only a statistical probability can be given. A definite maximum delay of transmission demands a session oriented service, such as file transfer.

4. TECHNICAL FACILITIES

The Commission has established connections to a number of public network services, such as X.25, INTERNET, ISDN, Dial-up, etc. It also has connections to Value Added Network (VAN) services offering basic communications as well as higher level services. The Commission has also leased-line connections with X.25 or TCP/IP communication to other European Institutions as well as other bodies.

As the Commissions internal network (IDNET) is TCP/IP based, the preferred way to communicate with the Commission is via an IP network. TCP/IP can also be used over an ISDN or an X.25 connection. VAN (Value Added Network) suppliers have been selected in a public tender by the DI and IDA (TESTA). Alternatively, public networks can be used.

4.1. *Electronic Mail*

E-Mail is a key service for Commission internal and trans-european projects. Its role in the projects depends on the way it is used. This paper distinguishes between:

- Interpersonal Electronic Mail
- Application Programming Interfaces to E-Mail
- Electronic Data Interchange via E-Mail

The E-Mail service quality criteria, the service limitations and organisational responsibilities for the service are defined by the Service Level Agreement (SLA). In using the E-Mail service a project recognises the conditions outlined by the SLA. Of particular importance in this context are the guaranteed availability and delivery time criteria.

Interpersonal Electronic Mail

The Commission's E-Mail service defines usage conventions in co-operation with the DGs in the following areas:

- addressing and naming,
- user registration and change procedures,
- system configuration,
- systems operation,
- service quality monitoring and reporting,
- security and fall-back conventions and procedures.

These conventions should be followed wherever possible.

The preferred E-Mail technology is X.400 (1988 specifications) in conformance to the currently valid EPHOS specifications (EPHOS 2). The Commission's E-Mail service selects only products strictly conforming to this specification.

INTERNET access (Simple Mail Transfer Protocol, SMTP) from and to the Commission is provided via gateways adhering to RFC 1327 and RFC 1495. The Commission strategy is oriented towards a full MIME support by the gateway making the connection to INTERNET as transparent as possible for the end user. This service is accessible by a Commission user via the X.400 User Agents generally implemented as part of the E-Mail

service. Separate specific INTERNET client software which would allow an E-Mail communication detached from the Commission E-Mail service should be avoided.

FAX and Telex access for outgoing communication is provided as part of the E-Mail service . A Commission user accesses these service through the standard E-Mail User Agent.

The E-Mail service includes the possibility of accessing the user mailbox in the DG via a Remote User Agent (RUA) from outside the boundaries of the Commission, e.g. for missions. The network connection is passing via the TC which also handles the security aspects.

The TC also provides an External Mail Service (EMS) with RUAs for those users who are located outside the Commission boundaries. This service has been created for the connection with external partners which do not have an own E-Mail service. It should not be used for the communication among the external partners themselves (avoidance of the relay function).

The E-Mail service provides a centralised quality monitoring and quality analysis facility with regular reporting to the participating DG.

The Commission has concentrated its efforts on the development of an E-Mail infrastructure based on personal, nominative mailboxes. Almost every official with access to informatics resources has by now a personal mailbox. Functional mailboxes representing departments, projects, and various services have also been introduced and are about to become a second principle for organising the E-Mail addressing infrastructure. Communication with functional mailboxes is nevertheless considered interpersonal E-Mail unless an application is attached to the mailbox.

An important element of the interpersonal electronic mail is the document format, in particular, the word processing format. For a transition period, the WordPerfect 5.2 and Winword 6/7 formats are allowed with Winword 6/7 becoming the future single exchange format for revisable documents. The evolution of Winword may in the future require the adaptation to future formats of Winword.

Non revisable documents may be exchanged in the PDF format. SGML is also used for specific document exchange requirements between the European Institutions.

For non word processing files, such as spreadsheets, graphics, etc., the E-Mail system of the Commission incorporates support for frequently used formats, e.g. Excel, Visio, HTML, G3, Postscript, etc. which allows easy handling for the end users. The exchange of other file formats requires a mutual agreement between the sender and the recipient.

Please note that the E-Mail policy of the Commission is currently under review in a project concerned with the future E-Mail system of the Commission.

Application Programming Interfaces to E-Mail

An important part of the E-Mail strategy of the Commission is the effective support of applications exchanging messages with other applications or end users. This requires Application Programming Interfaces (API). The two important platforms for API access to E-Mail are NT and UNIX servers as well as PC based application running one of the Windows versions.

Within the boundaries of the Commission the APIs are generally implemented in order not to affect the dedicated E-Mail servers. The API is residing together with the application on a platform different from the E-Mail server. This allows to select the hardware platform of the application independently from the E-Mail server and to keep the E-Mail server dedicated to E-Mail.

The Application Programming Interface (API) to the Commission's E-Mail is via a Common Messaging Call (CMC) client on a UNIX application host or via Messaging API (MAPI) or CMC on NT servers and PCs. Other APIs should be avoided. An X/OPEN MT API may be used in co-ordination with the E-Mail service manager at the Commission.

The choice of hardware and operating system to be used for those parts of a project related application should be discussed with the E-Mail service provider in order to avoid unnecessary porting of the CMC client software. A range of different platforms is already supported.

The quality monitoring of application related E-Mail traffic is done by the E-Mail service.

Electronic Data Interchange (EDI)

The preferred communication technology for EDI in the Commission as well as in trans-european projects is E-Mail conforming to X.400 (specifications of 1988 or later).

The EDI technology should conform to the X.435 specifications of 1988 or later. Many products implementing this architecture are available so that special Commission developments should be avoided.

As soon as products are known to be mature and are used in important projects, EDI converters should be subject to the normal product management procedures of the Commission. In particular, the requirements for a product, the choice of a product and the relationship with the supplier should be done in co-ordination with the DI. The product should also be classified for a regular follow up within the relevant product group.

Where possible, data formats specified by EDIFACT or other international standards bodies, such as the X.12 standards should be used. In areas specifically designated by EDIFACT as areas of interest, e.g. customs, and statistics data conformance to the EDIFACT standard should be considered a must.

If possible, the normal Commission E-Mail service should also be used for EDI purposes. Specific hardware servers and MTAs may, nevertheless, be installed. They will then be integrated into the INSEM E-Mail architecture, e.g. in terms of addressing and the various services regularly provided to E-Mail servers.

Particular attention is to be given to the required E-Mail service quality criteria. If the INSEM criteria are considered sufficient the EDI servers could be treated by the regular INSEM quality surveillance and follow-up. If, however, they are considered as not

sufficient additional quality measures and control can be established in the framework of the specific project. They are then within the responsibility of that project.

4.2. *File Transfer*

The generally used product for file transfer within the Commission is FTP.

Contrary to the E-mail facility, file transfer requires that both sender and receiver are running at the same time and a password exchange is required to set up the connection.

Additional facilities to E-mail are provided as part of the file transfer service. These facilities cover:

- restart of interrupted transfers,
- remote execution
- efficient support of very large files
- file fetch

File transfer across the boundaries of the Commission has exclusively to pass via the File Transfer Gateway (FTRG) in the TC. The FTRG supports the following file transfer protocols:

- FTP
- FTAM
- Kermit
- UUCP

The FTRG Service Level Agreement defines the guaranteed quality level.

File transfer via FTP is of course possible within the framework of World Wide Web access as it is an integral function of the browser software in the market. Its usage is less a technical issue than one of user guidance.

4.3. World Wide Web (WWW) Access to Information

The European Commission makes certain data available to a greater public via online access via the Internet according to the World-Wide-Web standards on a server known as "EUROPA". Other servers supplied by the Commission services have been created for specific purposes outside the Commission domain, e.g. ISPO.

For specific areas and trans-european information systems, data can be made available through the same technology, but with controlled and restricted access rights for well defined closed user groups ("Virtual Private Network") using the Internet with advanced security features (e.g. SSL - Secure Socket Layer) or using restricted access data networks.

The HTTP protocol and HTML language based on HTML 2.0 (RFC 1866) is used. A short term evolution to HTML 3.2 can be expected.

4.4. Interactive Access, Transaction and Client-Server Access

Some projects use interactive access to servers at the Commission. The servers have to be installed outside the Commission network in order not to allow uncontrolled access to its

internal network. Such a solution is generally easy to implement but has serious deficiencies, for example in the user interface, communication costs, etc. They should be seen as preliminary solutions, for example, to create an initial pilot and be replaced by more advanced technical architectures.

Trans-european projects may need transaction oriented processing in which a complete sequence of operations has to be fully completed before a commit to the new state can be done. A partial execution of the sequence of operations has to result in a roll-back to the previous defined state. Guidelines for the communication in such projects have to be developed.

Certain projects start to use the facilities offered by data base products with client-server connections, e.g. SQL-NET. Guidelines for those projects have equally to be developed.

5. PROCEDURES

A project planning to use the above named communication services should include in the project plan an explicit co-operation with the DI outlining the Commission services and their respective role in the project. The responsible service within the DI for the co-ordination is 'Support des Systèmes d' Information', SSI (Mr. Deasy).