Nationaal Lucht- en Ruimtevaartlaboratorium
National Aerospace Laboratory NLR

NLR TP 96789

Overview and discussion of electronic exchange standards for technical information

H. Kuiper and J.C. Donker

DOCUMENT CONTROL SHEET

	ORIGINATOR'S REF.		SECURITY CLASS.
	NLR TP 96789 U		**Unclassified**

ORIGINATOR
National Aerospace Laboratory NLR, Amsterdam, The Netherlands

TITLE
Overview and discussion of electronic exchange standards for technical information

PRESENTED AT
the CALS Europe'96 conference, Paris, May 29-31, under the title "SGML, HTML, the paperless office... what about the forests and the trees", upon invitation of the NL MOD Representative in the conference committee.

AUTHORS	DATE	pp	ref
H. Kuiper and J.C. Donker	**960725**	**36**	**8**

DESCRIPTORS

Computer graphics
Document storage
Document markup languages
Format
Hypertext
Information dissemination
Multimedia
Software tools
Standards
Texts
Word processing

ABSTRACT

Nowadays more and more information is being exchanged electronically. Reasons for this include a higher degree of cooperation between information suppliers and users, an increasing demand for speed (of production and modification, and reduction of time to market), and cost reduction. On the technology side, the advent of the electronic highway enables effective and efficient electronic information exchange.

For reasons of timeliness and life cycle costs, standards and specifications are becoming more important. The aim of this paper is to provide an overview of standards and specifications for electronic exchange of (technical document) information and to discuss the most common ones currently available for text, images, and document exchange. Emerging standards and specifications, such as for audio, video and virtual environments are also briefly discussed. Finally, a brief description is given of a standard for enterprise integration and product data exchange. The availability of tools is described in connection with each standard or specification.

The paper concludes with a brief selection preference for the standards and specifications presented in this paper. To indicate how tools can be integrated for virtual enterprise working environments, a facility for generation of working environments developed at the National Aerospace Laboratory is referenced.

Summary

Nowadays more and more information is being exchanged electronically. Reasons for this include a higher degree of cooperation between information suppliers and users, an increasing demand for speed (of production and modification, and reduction of time to market), and cost reduction. On the technology side, the advent of the electronic highway enables effective and efficient electronic information exchange.

For reasons of timeliness and life cycle costs, standards and specifications are becoming more important. The aim of this paper is to provide an overview of standards and specifications for electronic exchange of (technical document) information and to discuss the most common ones currently available for text, images, and document exchange. Emerging standards and specifications, such as for audio, video and virtual environments are also briefly discussed. Finally, a brief description is given of a standard for enterprise integration and product data exchange. The availability of tools is described in connection with each standard or specification.

The paper concludes with a brief selection preference for the standards and specifications presented in this paper. To indicate how tools can be integrated for virtual enterprise working environments, a facility for generation of working environments developed at the National Aerospace Laboratory is referenced.

Contents

(36 pages total)

Abbreviations

AECMA	Association Européenne des Constructeurs de Materiel Aerospatiale
AFNOR	Association Française de Normalisation
ANSI	American National Standards Insitute
ASCII	American Standard Code for Information Interchange
CALS	Continuous Acquitision and Logistics Support
CCITT	Comité Consultative Internationale du Télegraphie et Télephonie
DoD	Department of Defense
IATA	International Air Transport Association
ITU	International Telecommunications Union
SGML	Stuctured Generalized Markup Language

1 Introduction

Due to the higher degree of cooperation between information suppliers and users, and the information more often being developed and used across company boundaries, the electronic means are being used more and more for information exchange. Also, electronic information has the advantage of being easier to create, easier to check for consistency and easier to maintain in large quantities. An extra benefit is that content and form can be separated from each other.

To achieve quality and efficiency, agreements need to be made on how this information has to be produced, exchanged, presented and supported. Many national and international bodies have emerged to prepare and maintain standards and specifications, such as ISO, IEEE, ANSI and AFNOR; for some areas of industry, application area oriented standardisation bodies have emerged, e.g. AECMA for European aerospace industries, IATA for air transport, and ITU for telecommunications (formerly CCITT).

Other initiatives have emerged which use existing military and/or civil standards. The Continuous Acquisition and Lifecycle Support Initiative (CALS), which started as a US DoD effort, is now being supported by many government and industrial bodies in military and civil application areas.

The CALS initiative was launched in September 1985 as an industry/government strategy to transition from a paper-intensive acquisition and support process to an integrated digitial data exchange environment, with the overall objective of providing the enabling inforation architecture for enterprise integration. Capitalizing on the use of information and communications technology, the US Departments of Defense and Commerce, and a growing number of international governments and industrial consortia are using CALS to implement process improvements, such as an increase in flexibility, responsiveness and quality while reducing costs. An important part of the strategy is based on the use of standard formats for the exchange and sharing of technical data that supports the generation, access, management maintenance, distribution and use of electronic information.

The various application areas of CALS (e.g. acquisition and electronic commerce, design and development, production, integrated logistics support, training, and maintenance) all have a need for information. Each of the areas usually requires a different view on the same central information. As can be seen in figure 1, the information and data bases are produced and managed using a number of integral processes. These processes and application areas generate different views on the information and data, resulting in various electronic documents.

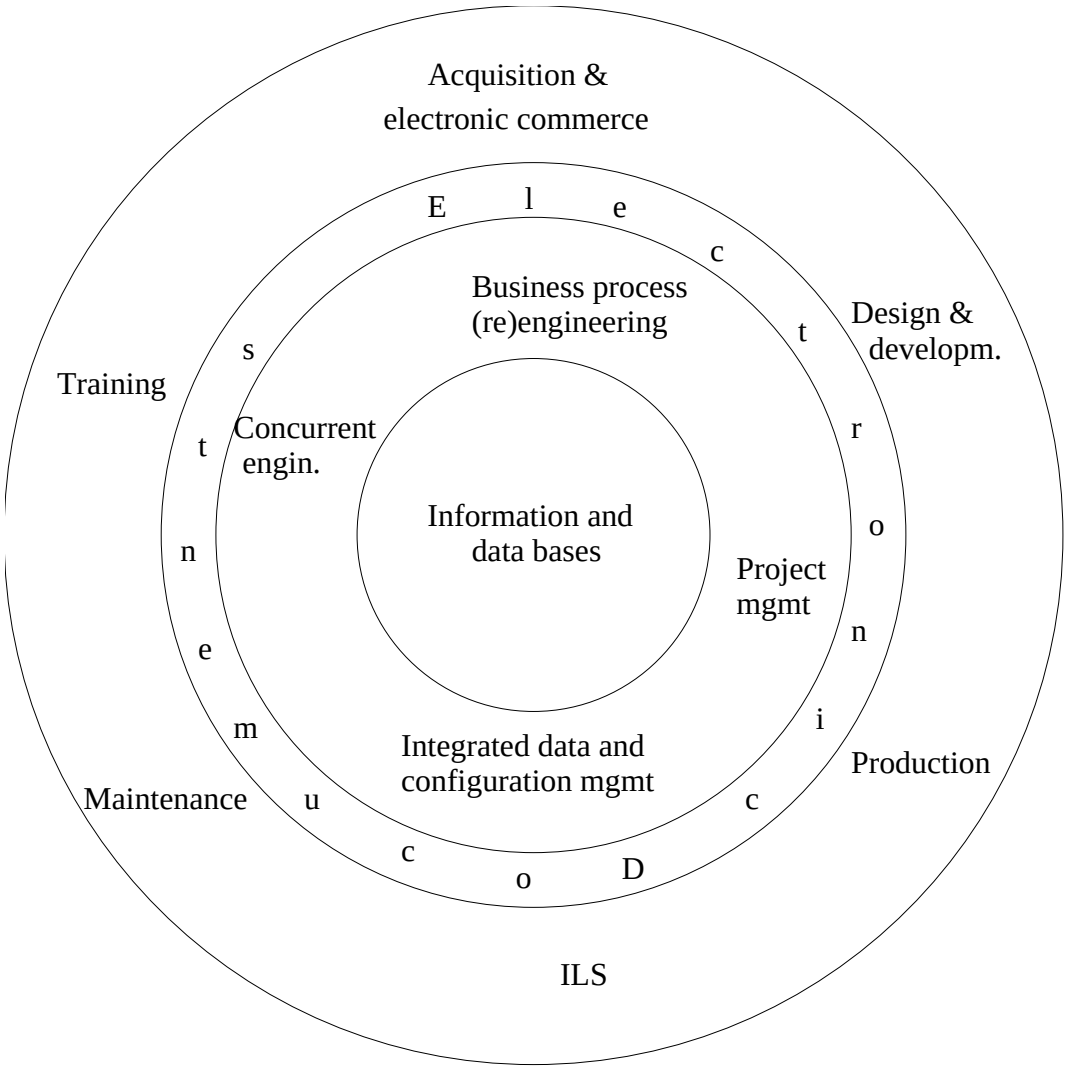

Fig. 1 Electronic documents in CALS

In this paper, we will concentrate on the standards and formats available for the exchange of technical documents. First, some general definitions pertaining to electronic documents are given, including terms in use for electronic document dissemination and viewing on the World Wide Web. Subsequently, the evolving structure of documents is discussed in more detail, before proceeding to identify which classes of documents can be recognized. Using this classification, standards and specifications for electronic text, graphics, multimedia and product data are described, as well as their current tool support. Finally an indication is given of sensible choices, based on openness and current status of the standards and specifications.

2 General definitions[1]

Specifications in this publication are defined as publicly available descriptions for documents, whereas *standards* denote their counterparts defined by formal standards bodies. Where no misunderstanding can emerge, the word standard is used for both the official standard and the vendor- or consortium-dependent specification.

Documents on the World Wide Web are broken down into *nodes*, connected with *links*. Nodes contain information from documents, videotapes or other sources broken down into smaller parts. Each node, whether a paragraph within a text or a scene in a video clip, should convey only one idea or theme. Links, or hyperlinks, connect nodes. The starting point of a link is known as *anchor point*. A user can click on a anchor and the associated link will be traversed, taking the user to the associated node. This process is known as *navigation*.
A *browser* or *viewer* facilitates the perusal of the information thus located. These two functions have become integrated in all modern web clients, and are now commonly called browser. For an overview of navigation and browsing in available Web tools, see ref. 3.
Authoring involves identifying structure for the information that support appropriate accessibility and manipulation. *Authoring environments* offer support for the production and maintenance of text and/or multimedia information. Often, commercial companies offer specific *viewers* for the information produced through their proprietary authoring environments. With these viewers, the information can be displayed and printed, but modification is not possible.

[1] The definitions in this section have been copied from refs. 1 and 2

3 Structure of documents

Paper documents as produced since the advent of book printing are usually of a linear structure. One starts reading at the beginning of the document, and proceeds to the end in a linear fashion. For larger documents, text can be divided into hierarchical components, such as chapters and paragraphs. Exceptions to the linear structure are reference documents, such as dictionaries. With the exception of e.g. maintenance and training manuals, paper documents usually contain more text than graphs and images.

With the advent of computers, the creation of graphs and images has become more widespread, resulting in a higher image content in documents. The next step is not only production, but also electronic dissemination of information and documents (e.g. through the World Wide Web). Ease of search is an important feature of electronic documents. Again, new possibilities are being explored: the addition of sounds and moving images to documents.

At the same time, the primarily linear structure of documents has changed. No longer bound to paper, users gain the possibility to browse through documents in more than one way. Using so-called hyperlinks, related document parts can be connected by the author of such a document, and the user can follow such a link directly from the text, image or other object, instead of going through the table of contents or the index.

4 Document formats and standards

With paper document exchange, the way in which a document was created was only important for the author and the publisher, in case modifications or re-prints were required. Now, with documents being exchanged electronically, and with the number of computer platforms and software packages increasing, the exact format in which the document is stored can be of paramount importance: it determines whether or not the target audience actually can process (e.g. read, modify) the document.

The various document formats and standards can be classified according to a number of attributes. These attributes are:

- *Content description*. This indicates whether the format or standard allows for the actual content of the document to be expressed using the format. If a format does not allow for the content to be specified, it acts upon information expressed using another format which does provide content description.
- *Presentation description.* This indicates whether the format or standard allows for a specification of how the content should be presented to the user. This information includes such things as fonts used, use of italics and bold, etc.
- *Structure capabilities.* This indicates to what extent the format supports specifying structure for the contents. Possible values are *linear* (e.g. ASCII files), *hierarchical* (e.g. SGML) and *networked* (e.g. HyTime and PDF).
- *Semantics support*. This indicates to what extent the format allows for semantics to be added to the content. It ranges from *none* through *(very) limited* to *full*.

The other axis along which formats and standards can be categorized is the sponsoring body of the format:

- *Proprietary* ones are those for which no specification is publicly available, and which are usually supported by one vendor.
- *Open, single vendor* formats are those for which specifications are publicly available, but which are maintained and updated by a single vendor.
- *Open, multiple vendor* formats are similar, but are maintained and updated by a consortium of vendors.
- *Independent* standards are those maintained and updated by an independent standards organization, such as ISO.

These categories are in increasing order of neutrality, thereby increasing the likelihood of independent and continued support. However, the independent standards also often have the slowest ’update’ rate, as it can take a long time to have standards and changes approved. On

the other hand, a longer standard update process does allow more time to upgrade existing software.

Although specifications of the first two types of standards can be changed quickly by a vendor, providing quick adaptability to new requirements or possibilities, there is a real possibility of applications becoming quickly outdated and/or of complete dependency on specific vendors. For large user communities and multiple vendors, conversions are usually well supported.

5 Text standards

5.1 Introduction

For documents, text has been the main information component. Until now, proprietary formats (e.g. Wordperfect, Word, Framemaker, Interleaf, etc.) are being used for production, electronic storage and exchange. A trend is emerging to standardise the formats for exchange of content, structure and presentation. For presentation of information on the World Wide Web, and navigation through that information, use is made of chunks of information (nodes) connected via links. This chapter discusses the standards and specifications for the production, exchange and presentation of text elements, both for paper and electronic documents as end product. Descriptions of individual specifications and standards mentioned in this section can be found in Appendix A.

5.2 Overview

For documents, text is one of the main information components. Text formats and standards fall into two large categories: proprietary (or application-specific) formats and open standards. In the first category, there are the file formats used by packages such as Wordperfect, Word, Interleaf and Framemaker. These formats usually integrate content and presentation description, with usually a hierarchical structure (chapter, paragraphs, etc.). Exchange of content and format between these packages was initially via ASCII, the American Standard Code for Information Interchange. ASCII has the disadvantage that no structure and layout information is preserved. Often, modern document processing packages have a number of intermediate formats and import and export filters which can be used for the preservation of structure and presentation attributes (e.g. RTF, MIF, Wordperfect to Word, etc.). Often, ASCII is used as a carrier for this meta-format. However, as not every format supports all constructs allowed by other packages, during these conversion information will be lost. It is not always guaranteed (and almost never the case) that converting to another format and back will result in the same format or even the same content.

The second category, open standards, ranges from content-only standard ASCII to the Standard Generalized Markup Language SGML. Although SGML has provisions for the inclusion of non-SGML data such as images, it has the strongest position in the representation of structured textual information. Therefore we have included SGML in this section as opposed to the section on document architecture and interchange standards, where it officially belongs. SGML is supported by a number of other standards which can be used to add presentation description (DSSSL, FOSI and SPDL, see Appendices A, C and D) and to add a hypertext structure (HyTime, see Appendix D).

HTML, which is derived from SGML, is an open standard for document formatting used on the World Wide Web. HTML is used in combination with HTTP, the Hypertext Transfer Protocol, a client-server handshaking and communications protocol. In contrast with the other open standards, HTML suffers from vendor-specific extensions, which are widely used, and therefore require a specific editors and viewers to create or browse the documents. More than SGML, HTML also supports the inclusion and manipulation of images. The best-known public-domain package currently in use on the World Wide Web for navigation and browsing is Mosaic. In the past two years, commercial efforts have increased. The best used package to date is Netscape, which is not free for use in commercial environments.

There is a clear tendency to move from application-specific formats to widely supported open (international) standards such as SGML: even vendors of packages such as Wordperfect, Framemaker and Interleaf are now offering SGML-aware versions of their products. The next generation of authoring environments, viewers and browsers will also support SGML.

6 Graphics standards

6.1 Introduction

In this section, an overview will be given of classes of graphics standards. First, an overview will be given of the most common terms in this area, and an overview will be given of the most widely used standards. Finally, the classes of graphics exchange standards will be described, and graphics exchange standards and specifications within each class are listed, with references to more detailed material.

6.2 Definitions[2]

Raster, vector, metafile, page description language and virtual reality modelling language are all terms used to classify the type of data a graphics file contains. *Raster files* (also called bitmapped files) contain graphics information described as pixels. *Vector files* contain data described as mathematical equations and are typically used to store line art and CAD information. *Metafiles* are formats that may contain either raster or vector graphics data. *Page description languages* are used to describe the layout of a printed/printable page of graphics and text. *Animation 3D formats* are usually collections of raster data that are displayed in a sequence. *Multi-dimensional object formats* store graphics data as a collection of objects that may be rendered (displayed) in a variety of perspectives. The *Virtual Reality Modelling Language* (VRML) is a 3D-object-oriented language used for describing "virtual worlds", typically on the World Wide Web. *Multimedia file formats* (see the section on miscellaneous standards) are capable of storing any of the previously mentioned types of data, including sound and video information.

6.3 Overview

In contrast to text exchange, which started with ASCII and now converges towards the use of SGML for the exchange of text-based documents, the situation for image exchange has not in all cases converged towards international standards.

Only for two-dimensional vector graphics such a widely-accepted standard is available in most document processing applications, viz. the Computer Graphics Metafile (CGM) format. For raster graphics, there are a number of widely supported de-facto standards. Currently, many drawing packages support GIF, which is also much-used to incorporate 2D images into documents on the World Wide Web (WWW). Another widely-supported file format is TIFF, of which Adobe owns the copyright. Both GIF and to a certain extent TIFF currently suffer from royalty problems concerning the compression/decompression of GIF and TIFF.

[2] The following definitions have been taken from the Frequently Asked Questions (FAQ) list which can be found on the Internet in the newsgroups comp.graphics.misc, comp.answers, and news.answers.

Therefore, many companies are now moving towards other formats, the most popular of which is the JPEG compression algorithm and JFIF and SPIFF file formats. Many graphics packages still support PCX for static images. PCX was widely used in the PC environment, and is now more and more being abandoned for the MS Windows BMP format, which is native to the Microsoft Windows environment. Finally, a widely used standard for the viewing of images is Postscript, a page description language created and maintained by Adobe. Although it is virtually always possible to view and print images using Postscript, the image is only seen as a bitmap, and therefore it is very impractical to modify such an image.

With the advent of more powerful desktop personal computers and workstations, and the need of integrated document processing containing both text and images, there is a growing need for international standardisation in those graphics areas where these do not yet exist. Moreover, there is a tendency towards more general information exchange in electronic form, which opens the possibility of growth towards interactivity and multimedia capabilities. Many of these capabilities build upon graphics standards. In appendix B, the most common standards are described. An excellent up to date information source can be found on the Info2000 server of the European Community (Open Information Interchange Initiative) (Ref. 1).

In addition to graphics exchange formats, many filters and conversion utilities exist to transform image formats within a certain class to other formats. Some of the most common public domain utilities are mentioned in ref. 4. Recent news can always be found in news groups on the Internet, such as comp.graphics newsgroups.

7 Document specification and exchange standards

7.1 Introduction

As a consequence of increased cooperative production of documents, initiatives in various application areas are now focussing on documents as a whole. Initially, word and document processing systems have defined their own native format. Recently, international standards are gaining more momentum. An overview is given below; more detailed information can be found in appendix C.

7.2 Overview

For exchange of text and images from one word processor to another, the Rich Text Format intermediate format has been and still is in widespread use. RTF is a proprietary specification developed and maintained by Microsoft for the exchange of formatted text and graphics.
For final versions of paper and electronic documents, Adobe's Postscript and, currently to a lesser extent its Portable Document Format, are in widespread use. Officially, SGML and HTML should also be categorised as members of this category. Since these are better recognized for their excellent text markup and exchange capabilities, we have described these under text standards.

A different scope is covered by the Open Document Architecture and Interchange Format (ODA/ODIF) which has been developed in the business and office environment. Classes of Documents are specified through Document Application Profiles; exchange of documents is performed via the ODA Document Language.

The Open Document Specification OpenDoc is another specification for document exchange, supported by a growing number of vendors including IBM, Apple, SunSoft, WordPerfect and Xerox. It supports IBM's System Object Model, a language-independent run-time mechanism for dynamic object linking compliant with the Common Object Request Broker Architecture (CORBA).

Recently, the Object Management Group, a grouping of more than 600 software vendors, developers and end users, has announced the adoption of the Distributed Document Component Facility (DDCF) as part of the Object Management Architecture. DDCF will enable the creation and support of shareable and customizable compound documents across heterogeneous systems.

8 Audio, video and multimedia standards

In addition to text and images, electronic documents can include audio and video components. Again, excellent overviews are contained in ref. 1. The most common standards for video and audio are MPEG-1 and MPEG-2 Audio (Coding of Moving Pictures and Associated audio for Digital Storage Media), and the WAVE format, a proprietary Microsoft and IBM format for Windows 3.1, heavily used as interchange format for audio files on the Internet. Quicktime is a proprietary specification developed by Apple with which video can be made available on desktop computers. Quicktime is now also available on other computers than Apple, and is used for many demonstrations on the Internet. Microsoft has also developed a proprietary standard, Video for Windows/Audio Video Interleave (AVI), which is favoured in personal computer environments.

Specific multimedia standards, which enable the incoporation of multimedia and hypermedia information, are under development. The best-known ones are HyTime for the presentation of audiovisual information into documents and MHEG, for representation of non-revisable final form multimedia and hypermedia. Furthermore, the Standard Multimedia/Hypermedia Scripting Language is under development for embedding scripts in SGML, with which events, actions and behavior of objects can be authored. See appendix D for more information.

9 Product data standards

When thinking of documents as views on product information for a specific purpose (e.g. design documentation, training manual), a closely related area is that of product modelling and exchange. After many years of discussion and harmonisation, the ISO International Standard for the Exchange of Product Model Data has become available. The standard is divided into a series of parts: description methods, integrated resources, application protocols, abstract test suites, implementation methods and conformance testing. STEP uses a formal specification language, EXPRESS, to specify product information. Application Protocols define the representation of product information for one or more applications. The ultimate aim of STEP is to arrive at an integrated product database for the whole life cycle of a product, independent from specific platforms or tools. Essential differences compared to earlier product data exchange standards such as IGES, SET and VDAFS are the description of product data at the semantic level instead of syntactic level, now capturing the meaning of the modelling elements. This is done via EXPRESS and a number of primitive, application independent modelling elements. Application-specific elements are required to be described in terms of the independent primitives.
Major CAD/CAM toolkits for large engineering applications have been quick in implementing the first application protocols, and it is expected that STEP will be adopted by all companies promoting enterprise integration, concurrent engineering and agile manufacturing. STEP is an integral part of CALS.

10 Simulation and virtual environments

A virtual environment is a real-time simulation of a real or imaginary world where users navigate and interact with 3D objects within it (Ref. 5). Virtual environments can be modeled using VRML, the Virtual Reality Modelling Language. VRML is a scene description language that standardises how 3D environments are represented on the World Wide Web. Most VRML files are not very big, so bandwidth usually is not much of a problem. VRML files store the most important characteristics of each object, and instructions for the browser. The browser then generates the 3D environment from this information. Consequently, CPU power usually is the limiting factor. Levels of detail are normally used by the browser to first build a rough picture and then fill in more detail.
Tools for production and manipulation of 3D objects include VRML browsers, VRML authoring packages and 3D creation tools (including existing 3D graphics, CAD/CAM and animation packages). Since VRML is very new, no standardisation of tools has taken place yet.

With the advent of VRML, simulation and animation of (virtual) environments has become much easier. This will lead to an increase of cost-effective simulations for training and of remote environments (e.g. 3D modelling and animation of repair procedures, low-cost simulation of flying through simulated environments, 3D simulation of battlefields, teleoperation and telepresence, etc.).

11 Selecting formats, standards and tools

Which (type of) format or standard to use largely depends on the purpose of the documents produced. If document exchange is only necessary during the production phase (i.e. after production only a hardcopy will be used), the format depends on the software packages available. For not too large documents, word processors such as Word or Word Perfect are suitable. For larger amounts of documentation, document processing systems such as Framemaker or Interleaf are more suitable.

If a single software package is not feasible, such as in the case of many companies having different standards, or in case of integration of (partly) existing systems, the best choice would be to choose an open standard as exchange format.

For text, the best choice in the near future will be SGML for structure and content, since it is an international, vendor-independent standard. An additional choice must be made for presentation/viewing. Emerging presentation standards are DSSSL and SPDL, but support is not yet widely available. A practical solution adopted by many distributors of documents is to produce Postscript or PDF files. Postscript files can be presented electronically through public domain viewers such as Ghostview, and printed on virtually any modern printer. PDF files can be read by using a free viewer distributed by Adobe. An alternative would be the conversion from SGML to HTML, which is more or less straightforward. Other possibilities currently supported by many document processing systems include the use of filters to convert from one native format to another, or the use of ASCII. In practice, this means quite some extra work to compensate for shortcomings in conversion, especially for more complicated constructs such as tables and formulas.

For 2-D collaborative production of static images, or images which might need to be modified later on, CGM is the best choice for exchange. Problem with CGM is the use of the LZW compression algorithm in its compression part (see appendix B). If a move to another lossless compression algorithm is not made fast enough, CGM will lose support.

For the production and exchange of product design data, reasoning is much the same. The easiest way is to define the same design approach and product data modelling, using one CAD/CAM tool solution. This is because not only syntax, but also semantics differ between applications. If the use of one standard environment is not possible, which is often the case, product data exchange can best be done via one of the exchange standards adopted by your industry. IGES is the widely-used one, DXF, SET and VDAFS are alternatives in certain industries. In the near future, STEP application protocols will emerge which will enable cooperation and enterprise integration at a higher level. AP 203 is gaining acceptance in aerospace industry.

For audio and video standards, the MPEG standards are most promising. Current applications on the Internet use native formats such as Quicktime and Video for Windows/AVI.
For interactive 3D manipulation of images in electronic documents, VRML is a rapidly emerging standard.

12 Working environments

When producing documents, both for paper and for electronic publication, the variety of tools implementing the standards discussed in this paper is overwhelming. Many of these tools have not been designed to be part of an integrated environment. Additionally, these tools often are available in a hetereogeneous computing environment ranging from PCs and low-end workstations, to database servers, high-end graphics workstations and even supercomputers.

To provide one virtual computing environment to users, NLR has developed a facility, SPINE, that supports the construction, customization, maintenance and operational use of functionally-integrated working environments in computer networks. A SPINE-based working environment provides its end user with access to, and operation of all resources available from the network, as if the resources are located on a single virtual computer: computing, storage, I/O facilities, as well as information in the forms of software, data, and documentation (Ref. 6). The following working environments have already been realised:

- ISNaS for numerical flow simulation
- ISMuS for computer-aided control engineering
- ISRaP for radar data processing
- ISEnS for computer-aided software engineering.

The latter environment currently contains tools for integrating 2D schemes from Computer-Aided Software Engineering tools with document processing tools, document viewers, and tools for version management. Cooperation between tools and conversion facilities are made available to the user via a simple drag-and-drop interface. Extension of this list with an environment for concurrent engineering is foreseen.

13 Conclusions

For the production of paper and electronic documents, a large variety of standards has emerged. In this paper, we have provided an overview of the state of the art in standards and specifications for text, graphics, audio, video, multimedia and product data modelling and exchange.

Tools to support these standards are relatively mature for text processing and 2D and 3D static images. The rest of the tools is quite new and more often than not immature. In contrast to the more mature standards and tools, however, standardisation is now being performed upfront.

The use of a combination of standards and tools is not straightforward. Therefore, there is a need for architectures and working environments to arrive at workable environments for authors and end users. At NLR, a number of working environments have been developed for specific user groups. These enable the use of a variety of tools in a heterogeneous computing environment, via a simple drag-and-drop interface.

14 References

1. M. Bryan, M-S Li, Open Information Interchange on-line information set. Internet URL: http://www.echo.lu/oii/en/oiistand.html
2. A. Ginige, D.B. Lowe, J. Robertson, Hypermedia Authoring. IEEE Multimedia, Winter 1995, pp. 24-35.
3. H. Berghel, The Client's Side of the World-Wide Web. Communications of the ACM, Vol. 39, No. 1, Jan 1996, pp. 30-40.
4. J.D. Murray (ed.), Frequent Asked Questions list for graphics. Posted to Internet newsgroups comp.graphics.misc, comp.answers and news.answers.
5. Olof Hagsand, Interactive Multiuser VEs in the DIVE system. IEEE Multimedia, Spring 1996, pp. 30-39.
6. E.H. Baalbergen, W. Loeve, SPINE: Software Platform for computer-supported cooperative work in heterogeneous computer and software environments. NLR Technical Publication NLR TP 94359 U.
7. IPO, IGES - Reference documents. Internet URL: htpp://elib.cme.nist.gov/nipde/stds/wh-iges.html.
8. A. Karmouch, J. Emery, A Playback Schedule Model for Multimedia Documents. IEEE Multimedia, pp. 50-61.

Appendices

A Text standards and specifications

In this appendix, a number of the more well known formats and standards are described, using the following format:

- *type of format*, giving the classification of the format or standard using the attributes and categories from the previous section
- *description*, giving a textual description of the format or standard, including some of the features
- *tool support* contains some remarks on available tools for both creating and viewing files with the format in question
- *current status* contains some remarks on the status, extensions being developed, etc.

No application specific formats have been described. These are the formats used mainly by one manufacturer. Examples include the native formats used by word processing packages such as Wordperfect, Microsoft Word and Adobe Framemaker. In the early days the demand for uniform standards was not very high and therefore each software manufacturer "invented" its own file format. As the number of available packages increased and the exchange of document (parts) between users as well, a need arose for conversion between various formats.

American Standard Code for Information Interchange

Type of format Content description. Open standard.
Description The American Standard Code for Information Interchange (ASCII) is a standard which defines how various characters and symbols should be represented.
Tool support Almost all computers can handle ASCII files. Large numbers of software packages have the ability to import and export ASCII files.
Current status Widely spread standard. ASCII is the American version of ISO/IEC 646. However, as these sets do not have provisions for many non-American symbols, extensions to the standard exist, as well as other standards which define more thoroughly existing characters (these standards include ECMA, Unicode, etc.) The more exhaustive standards are not supported to the same extent as ASCII is.

Standard Generalized Mark-up Language

Type of format Content description, hierarchical structure, no formatting, full semantics support. Open, independent standard.

Description The Standard Generalized Mark-up Language (SGML) is a language with which document content and structure can be described by tagging data to allow unambiguous transfer across various systems. The standard defines both the semantics for describing document structure, as well as the syntax for defining Document Type Definitions (DTDs). DTDs describe how a document is (should be) structured. It defines a set of tags, which are used throughout the document to indicate structural entities (such as 'chapter', 'part number', etc.). Content which cannot be described with the SGML syntax (such as images) is stored externally to the document and is referenced from within the document using so-called entity declarations. Document format can be achieved via separate standards (see e.g. DSSSL).
On the World Wide Web, the HTML language is used, which can more or less be described by an SGML DTD.
Tool support Although SGML itself exists since 1986, tool support usually was only available through specialized SGML editors, which where not always user friendly, as they work different from ordinary document processors.
SGML is being used more and more, and because of this, versions of popular document processing packages such as Wordperfect, Interleaf and Framemaker are becoming available which have SGML extensions.
Two additional ISO standards, DSSSL and Hytime, are available to add presentation description and hyper-media facilities to SGML documents. In addition, FOSI is part of the CALS standard.
Current status Actively supported ISO standard (ISO/IEC 8879:1986). Part of the CALS SGML specification (MIL-M28001).

HyperText Mark-up Language

Type of format Content description, (partial) presentation description, networked structure. Open, independent standard with some open, single vendor extensions.
Description The HyperText Mark-up Language (HTML) is the language used to describe documents on the World Wide Web. Although derived from SGML, the DTD is seldomly followed by HTML authors. Content-description tags are often used for presentation purposes. Aside from document content description tags, a number of presentation tags are available, as well as tags especially used for hyperlinking HTML documents and providing options for electronic forms.
Tool support Viewers for HTML documents are widespread. Of the free software packages, Mosaic is the most popular. Overall, Netscape is the most widespread, but is not free for use in commercial environments.
For creating HTML documents, more and more HTML editors for various platforms are becoming available. Companies such as Silicon Graphics and SUN offer complete HTML authoring environments based on their workstations.

Current status The current official version is HTML 2. Work is being done on version 3, which includes facilities for tables, and introduction of more presentation attributes. Extensions to the standard exist and are widespread, most notably those introduced by the Netscape company.

Document Style Semantics and Specification Language

Type of format Presentation description. Open, independent standard.
Description The Document Style Semantics and Specification Language (DSSSL) is a recent ISO standard, providing a language with which for SGML document the presentation of those documents can be described. This description consist of two parts, one with which the structure of the SGML document can be reordered, and one with which the document elements can be formatted for presentation. In fully standardized environment, this could result in SPDL output to drive a printer (see appendix C).
Tool support As the standard is very recent, no tool support is known. Most SGML aware packages have their own way to define the presentation of an SGML document, which is not compatible with other packages. A number of vendors indicated to wait and see how large the request for DSSSL was before implementing it.
Current status An ISO standard since January 1996 (ISO/IEC DIS 10179). A number of members of the SGML Open consortium have defined a subset of DSSSL (the DSSSL online application profile) which will be used by most vendors as a starting point for full DSSSL implementations.

Formatting Output Specification Instance

Type of format Presentation description. Open, independent standard. Part of SGML CALS specification (MIL-M-28001, Appendix B).
Description While a DTD (see SGML) describes content and structure for a specific application, a specific FOSI interprets the style and formatting requirements of a specification. The current standard provides formatting capability for paper SGML documents; it is being expanded to include electronic presentation as well.
Tool support Limited vendor support. Most SGML aware packages have their own way to define the presentation of an SGML document, which is not compatible with other packages. See also DSSSL, which was developed later.
Current status Used in paper-based military environments.

B Graphics standards

B.1 Vector Graphics

There are a considerable number of standards in this category, which stem from various organisations varying from single vendors to international standardisation organisations. For a comprehensive overview, see ref. 1. Main specifications and standards of relevance to the CALS electronic document processing community are CGM (ISO and MIL-STD) and IGES (CAD/CAM drawing-oriented exchange standard now adopted by ANSI and CALS). In this section, CGM and IGES are worked out in more detail.

Additional standards include the Graphical Kernel System (GKS) (2D) and GKS-3D, DXF (AutoCad 3-D Exchange Standard supported by many CAD and other vector drawing packages), PHIGS and PHIGS-plus (ISO standard mainly in use for manipulation of 3D objects on high-end graphical workstations), SET (AFNOR standard for the storage and exchange of geometric engineering data, mainly European aerospace projects), and VDAFS (DIN for the interchange of engineering data, mainly in German automobile industry).

Within ISO Subcommittee 24, a Computer Graphics Reference Model has been developed as an ISO standard (ISO/IEC 11072: 1992), which provides a framework for future characterisation and standardisation of 2D and 3D computer graphics data formats and processing.

Computer Graphics Metafile

Type of format Content description, presentation description for two-dimensional vector graphics. Open, independent standard. Contained in CALS as MIL-D-28003.
Description CGM is the most widely used standard for two-dimensional monochrome, grayscale and colour graphics. It is a widely-adopted standard in both the civil and military communities. It is a metafile standard, in which both vector and raster-type graphics can be used. However, for raster-type graphics alone, other more applicable standards are available. CGM is a static standard: it can not provide for dynamic effects on partially defined pictures. CGM is included in the Office Document Architecture developed in the office automation sector, and in CALS, for the incorporation of images in electronic documents.
Tool support CGM has been widely supported by word and document processing systems since the late 1980s. In 1992, a new version has become available, which offers facilities previously only existing in page description languages such as Postscript.
Current status Actively supported ANSI/ISO standard (ANSI/ISO 8632 - 1992, two amendment in 1994 and 1995). One major disadvantage of current CGM implementations at this moment is the wide-spread use of CGM in combination with the compression algorithm LZW (named after their inventors Lempel, Ziv and Welch) which in 1995 has successfully

been patented. Many suppliers of filter and conversion packages are currently transferring to other compression algorithms.

Initial Graphics Exchange Specification

Type of format Content description, presentation description. Open, independent standard. Part of CALS (MIL-STD-28000).

Description The IGES standard was developed to exchange engineering information between CAD/CAM systems. The focus is on graphics (drawings, three-dimensional wireframe models and surface models), although annotations are also part of the standard. IGES information at the sending end is produced by CAD/CAM systems; on the receiving end the information is intended to be read by humans. IGES is the most common format for exchange of this type of information. Since IGES supports all entities of widely varying CAD/CAM systems, mappings of representations is quite challenging. Therefore, Applications Protocols are being developed for certain application areas. These form the basis for the implementation of STEP, the ISO 10303 Standard for the Exchange of Product Model Data (see under miscellaneous standards).

Tool support All major CAD/CAM systems support IGES. In addition, some of the best-known CAD/CAM systems in the aerospace industry (CATIA, CADAM and AES) share a common core of IGES interface routines enabling faster translation by partly eliminating the production of IGES text files. In addition, the IGES standard contains a reference to tools available from the NIST IPO Office. An overview of these tools (public domain viewers, shareware browser, and commercial tools) can be found on NIST's WWW IGES pages (Ref. 7).

Current status The standard is maintained by the IGES/PDES Organisation as part of the US Product Data Association. The current version IGES 5.2 was approved by ANSI and published as ANSI US PRO/IPO-100-1993. IGES is also adopted as Federal Information Processing Standard 177. In the years to come, IGES will remain supported for engineering applications. Since 1994, the initial release of the STEP standard provides product modelling not only at a syntactic level, as in IGES, but also at a semantic level (see under "Other standards"). Once enough Application Protocols and tools will become available, industry will probably upgrade to this standard.

B.2 Raster graphics

Raster files, also called bitmapped files, contain graphics data in the form of pixels. Normally, data compression is applied to reduce the file size of the images. Below the more common formats are described.

Fax Groups 3 & 4

Type of format Content description. Open, independent standard.
Description The Fax Groups 3 & 4 format is the image encoding scheme used for fax transmission of images. The encoding scheme is also used for other formats, including TIFF.
Tool support Various implementations of the encoding algorithm exist.
Current status Actively supported standard by the Internation Telecommunications Union (ITU/CCITT)

.

Graphics Interchange Format

Type of format Content description. Open standard, patented by IBM and Unisys Corporation.
Description The GIF format allows for multiple images in one file. Images are compressed using the LZW algorithm.
Tool support Very widely supported format. Default image format on the World Wide Web.
Current status Because of the decision of Unisys Corporation to enforce their patent on the LZW algorithm, developers are moving towards royalty-free formats, such as JPEG. For the World Wide Web, a new format is being developed, the Portable Network Graphics format (PNG)

.

Joint Photographic Experts Group Format and Still Picture Interchange File Format

Type of format Content description. Open, independent standard.
Description The JPEG and SPIFF formats can be used to store all types of still image raster graphics. For the compression algorithm, there is a choice between lossy or lossless (with lossy algorithms, the result after decompression is only an approximation of the original image).
Tool support JPEG is widely supported, SPIFF less. Both hardware and software implementations exist, including free software implementations.
Current status Actively supported jointly by ITU and ISO (ISO/IEC 10918).

Tag Image File Format

Type of format Content description. Open, multiple vendor standard by Adobe Systems Inc and Microsoft Corp.

Description The file format uses so-called tags to describe 2D raster data. The image itself can stored using various compression techniques, including LZW, Group 3 & 4 Fax and JPEG.

Tool support The TIFF format is widely supported with desktop publishing software and image scanners. However, not all application supporting TIFF implement the same tag-set, creating a situation where not every package can read all TIFF files.

Current status ANSI is currently standardizing a TIFF subset for image technology applications, in order to reduce the current incompatibilities. The new standard, TIFF/IT is currently an ANSI draft.

C Document interchange standards

Rich Text Format

Type of format Content description, presentation description, hierarchical structure. Proprietary format by Microsoft Corporation.
Description The Rich Text Format (RTF) is based on ASCII and contains information on fonts, layout etc. Also, images are converted to ASCII, thereby creating a file that can be transferred across networks easily. The primitives used in RTF correspond to word processing functions. Information that cannot be represented using these primitives is lost.
Tool support The RTF is format is gaining support as the primary interchange format between word processors, and therefore is supported by most packages.
Current status Each time Microsoft releases a new version of their word processing software, the standard is updated. There is no guarantee that word processing packages use the same RTF version.

Postscript

Type of format Content description, presentation description. Linear structure. Open, single vendor standard (Adobe Systems Inc).
Description Postscript is the most common printer language available. Text and images are described using a stack-based language. Although it can be used to exchange documents, editing Postscript documents is practically impossible.
Tool support Almost every software package that has hardcopy output capabilities can create Postscript files.
Current status The de-facto standard for printers, currently at level 2. The new Standard Page Description Language extends on Postscript.

Portable Document Format

Type of format Content description, presentation description, networked structure, no semantics. Proprietary format from Adobe Systems Inc.
Description PDF is a page description language based on Postscript, another Adobe format. PDF files have the possibility for hyperlinks to be included in the document (hot links), thumbnails (for fast browsing), page annotations (which allow a user to add personal comments to a document) and chapter outlines (containing meta-information such as summaries and indexing information). PDF files are meant to be independent of the application software, hardware, and operating system used to create it.

One of the strong points of PDF files is that the appearance of the documents remains the same, independent of the viewing platform.

Tool support For a large number of platforms, Adobe has released their Acrobat viewer for free. Development tools can be purchased from Adobe, and include the PDF Writer and Acrobat Distiller packages.

Current status The Portable Document Format (PDF) has recently begun gaining support, since Adobe is offering PDF viewers for free for version 2 files. It is still not clear whether the PDF will become more popular then the HTML format used on the WWW (an area also targeted by Adobe). Adobe states that PDF is complementary to HTML and provides "helper" applications for HTML viewers, so these can also display PDF files.

Standard Page Description Language

Type of format Presentation description. Open, independent standard.

Description The Standard Page Description Language (SPDL) is the standard which can be used in the final step of the publication process: it describes how the required presentation of a document (through DSSSL) can be realized on various output devices.

SPDL provides an international reference version of the features found in Postscript level 2.

Tool support Unknown.

Current status An ISO standard since 1995 (ISO/IEC 10180:1995)

Open Document Specification

Type of format Content description, presentation description. Open, multiple vendor dependent standard.

Description The Open Document Specification (OpenDoc) is a CORBA compliant object-oriented method for specifying contents and structure of documents. It is being developed by CILabs, a consortium of vendors including IBM, Apple, Novell, Oracle, SunSoft, WordPerfect and Xerox.

The method uses support functions such as overlapping text and images and nested text objects. Applications need to register the set of functions supported, and in order for documents to be successfully exchanged, the same set of functions need to be available at both sides.

Tool support Unknown.

Current status The final specification was published at the end of 1994. Sofar, few products based on the specification have become available.

Open Document Architecture and Interchange Format

Type of format Content description, presentation description, hierarchical structure. Open, independent standard.
Description The Open Document Architecture and Interchange Format (ODA) is mainly used for the interchange of business documents, such as reports, letters, forms, invoices and memoranda. The document model defines a treelike structure of a document, with each node in the tree having a number of attributes describing the node. Content nodes can contain both text and images. Using the Office Document Interchange Format (ODIF), documents are interchanged between systems. Continuous media such as audio and video currently cannot be incorporated into ODA-compliant documents.
Tool support Unknown.
Current status ODA is made up of a number of ISO standards. The primary standard is ISO/IEC 8613, containing of 14 parts. Some of these parts are currently being developed to extend ODA with among others audio content and networked structures (HyperODA).

Distributed Document Component Facility

Type of format Content description, presentation description, hierarchical structure. Open, multi-vendor specification.
Description The Distributed Document Component Facility is based on a revision of OpenDoc technology to include support for CORBA services and the OMG Definition Language, which is the interface for CORBA.
Tool support OpenDoc is an implementation of DDCF. Tool support is currently unknown.
Current status DDCF has just been introduced by the OMG. Many of the OMG standards are quite succesfull. For heterogeneous client-server environments, DDCF stand a good chance of becoming an important specification.

D Multimedia standards

Hypermedia/Time-based Structuring Language

Type of standard Presentation description, networked structure. Open, independent standard.

Description The Hypermedia/Time-based Structuring Language (HyTime) is an SGML application that provides methods for the specification of hypertext links and time scheduling of a document. The standard has been developed to allow creation of multimedia documents.The standard provides extensive support for addressing, hyperlinking and event scheduling. It does not include multimedia authoring tools and support (Ref. 8).

Tool support A number of HyTime "engines" exist, including HyMinder, HyOctane and SoftQuad Explorer.

Current status An ISO standard since 1992 (ISO/IEC 10744:1992). Not all HyTime engines are concentrating on all parts of the standard, which could result in different HyTime engines being developed for different market segments.