New Types of Services and Documents in the Internet, and Their Impact on Information Transfer

Paul Nieuwenhuysen
University Library, Vrije Universiteit

Patrick Vanouplines
University Library, Vrije Universiteit
Document+program hybrids in the Internet, and their impact on information transfer in science and technology

Paul Nieuwenhuysen and Patrick Vanouplines
University Library
Vrije Universiteit Brussel
Pleinlaan 2, B-1050 Brussels - Belgium
Tel: ++32-(0)2-6292436, Fax: ++32-(0)2-6292693

Introduction
A growing range of protocols, tools and techniques is available for the development of interactive services and documents on the Internet, and on the World-Wide Web (WWW) in particular. In the beginning of the WWW, there was (only) the Hypertext Transfer Protocol (HTTP) and the first version of the Hypertext Mark-up Language (HTML). In this contribution, we focus on Internet-based documents, services and in most cases hybrids of these, which contain elements prepared with more recent, advanced methods, tools and techniques.

We look at these phenomena as professional information intermediaries in a university, who wonder about the future impact in the area of online access to information and knowledge. This should update and extend our paper, which was written in August 1996 (Vanouplines and Nieuwenhuysen 1996). Here we focus more on applications in science and technology.

Document+program hybrids in the Internet?
Some relatively new, recent, advanced methods, tools and techniques in the Internet are:

- data collections combined with more or less dedicated programs to view and handle these (excluding of course more classical databases)
- advanced HTML features in WWW,
- additional viewer programs or plug-in programs to be used with a WWW browser,
- the Virtual Reality Modelling Language (=VRML),
- the Common Gateway Interface standard (=CGI) in WWW,
- "intelligent software agents" or "searchbots" or "Webbots",
- Sun's Java programming language and the system built around this,
- Microsoft's ActiveX programming approach,
- program scripts for HTML and WWW browsers,
- WinFrame ICA client-server software by Citrix,
- Internet "push" technology with Web-casting channels,
- trusted systems in the Internet,
- virtual, computer-based chat rooms, social worlds/environments/places in the Internet,

These tools are potentially just as exciting or confusing as the services/documents which they support. They contribute to the evolution:

\[
\begin{align*}
\text{from} & \quad \text{"either document or program"} \\
\text{to} & \quad \text{"document or (?) program"}
\end{align*}
\]

More explicitly formulated, we see an evolution:

\[
\begin{align*}
\text{from} & \quad \text{documents on one side and computer programs at the other side, neatly separated, distinct, apart from each other, without much interaction, so that the static documents can also exist without computers and networks,} \\
\text{to} & \quad \text{document+program hybrids: hybrid systems}
\end{align*}
\]

- where all components are integrated, interwoven, and exist in synergy with each other,
- which can NOT exist without each other,
- which can only work by involving a computer,
- where the classical distinction between the contents and the container is blurred,
- which are newer and often remarkable, surprising, and exciting, because many readers/users are not yet familiar with them,
- which can carry a high added value, they can be more dynamic, more interactive, in comparison with more classical, more static documents, by involving and exploiting the power of computers and networks.

With some of these systems, the user can select information, or -- more generally stated -- express some information need, and obtain the desired information, although they are not classical Internet-based database systems. They can contribute to interesting ways of distributing information, but do not guarantee a high information content. The newest tools and techniques are often used first for test and demonstration purposes, to construct for instance World-Wide Web pages, which are impressive only at first sight, with a regretfully negligible information content.

The methods and techniques and the document+program hybrids mentioned are mainly based on the client-server architecture which is briefly and generally outlined in the table below:
As an illustration we collected some pointers to Internet-based sources (URLs) that offer some scientific or technical information, while using the new tools and techniques mentioned above; they can be found at the WWW server of our university at http://www.vub.ac.be/BIBLIO/. Several examples referred to in the text below can be found there. The exact location may change over time, as well as the name/title given to this collection of pointers. Up to now we have named the document+program hybrids "highly interactive services", but it is hard to find a good name which covers the wide range of Internet documents, services and hybrids that we like to discuss here, because they are remarkable as they:

- use a script / program plus data on the server, or
- address subsequent servers, combine results, eliminate irrelevant, or double information,
  or
- run a script/program on the client computer
- to produce a "result" or "document" with some valuable / interesting information content.

Summarised, we focus on
active / dynamic / highly interactive
network- / Internet- / WWW- based
hybrid documents or data collections
+ programs/services
with a high added value
(in comparison with classical, static documents)

Here is not meant the same as so-called "active documents" or "active contents" or "executable documents" or "executable contents", which stands for the more limited and more easily defined concept of documents existing in synergy with a script/program run on the client computer. Nevertheless, many of the document+program hybrids discussed in this contribution belong to this category of course.

A well known document / service hybrid system or document+program hybrid is of course a computer-based database, where selections of records form documents, created when needed. Well known examples are bibliographic databases and all kinds of directories. These are not discussed here, because they are not remarkable.
(anymore). However, some systems allow the user to select information, or -- more generally stated -- to express some information need, and to obtain the desired information, although they are not classical, Internet-based database systems, in the sense that their user-interface is not conventional. See for instance the WWW site offering information about the chemical elements; this service is named WebElements and a snapshot of the first page is shown in the figure below. Selections are made through one of the normal WWW client programs by browsing, pointing and clicking, and multimedia information is offered; more convenient than a printed book.

Another and more recent kind of services in the WWW offer a momentary view (literally) through some instrument, a camera or other sensor. The view can be on a room, a laboratory, a street, the sea coast, an experimental set-up, the contents of a soft drink machine or a coffee pot... However, most offer little interactivity.

As far as we know, there is at present no Internet index with search engine that directly allows to find the "remarkable" information services which we discuss here. Some retrieval tools come close; for instance, AltaVista in advanced mode allows searching by keywords for Java applets. However, one still has to inspect any retrieved item to get an idea of how it works and of its information content and value.

In the following we present a brief overview of the various methods already listed
above, which form the technical basis for the document+program hybrids in the Internet.

**Recent methods, techniques and standards in the Internet, with examples of applications**

**Data collections + dedicated programs**

These are rather popular and valuable in scientific domains like astronomy and molecular biology. For example, astronomical data catalogues are available on Internet through [http://cdsads.u-strasbg.fr/ads_catalogs.html](http://cdsads.u-strasbg.fr/ads_catalogs.html) which gives also links to other such systems.

**Advanced HTML features in WWW**

Advanced aspects and versions of HTML allow for instance:

- forms in pages
- frames on pages
- program scripts in HTML documents
- style sheets to describe outlook of pages
- server push and client pull

The official specifications of HTML are made available online and free of charge by the World Wide Web Consortium (1996).

**Multimedia viewer and plug-in programs for Internet/WWW browsers**

WWW browser programs soon incorporated code to view static picture files (GIF and JPEG format files). Besides these built-in application programs, external programs have been used to interpret downloaded files of types which cannot be directly interpreted by the basic WWW browser program. These should be installed in advance to work in synergy with the browser:

- "viewers" are programs which can be activated by applying the MIME standard, independent of the WWW client program;
- "plug-ins" are programs developed to be incorporated within a specific WWW browser, and to work within the browser window. Many plug-ins are available from The Ultimate Collection of Winsock Software [online] which is mirrored by many servers around the world. Viewers and plug-ins are available for file formats such as those given in the table.
Together with the evolution of viewer software, the multimedia content increases on the WWW. Journals start publishing papers on the Internet which include not only text and pictures but also movies. See for example the new journal New Astronomy at http://www1.elsevier.nl/journals/newast/, with equivalents to paper articles in Portable Document Format, but with online the added value of multimedia elements that do not (cannot) have a counterpart in hard copy.

Virtual Reality Modelling Language (=VRML)

Virtual reality is, computer generated, 3-dimensional, and interactive (user-computer).

Some application areas of virtual reality are:

- modelling of buildings in architecture
- (bio)chemical modelling
- flight simulation
- games
- artificial chatting worlds / environments
- modelling of real cities and other environments for tourists

Virtual Reality Modelling Language (VRML) stands for a few generations of more or less standard computer languages to create interactive simulation. VRML applications on the Internet can be used with a WWW client program that includes a VRML component (built-in or as a plug-in):

- Netscape Navigator extended, professional versions 3 and later WWW browsers from Netscape include a built-in interpreter/viewer for VRML files.
- Microsoft Internet Explorer up to version 3.02 does NOT, and requires one of the available plug-ins.

See for instance Virtual Reality Modelling Language [online]. The newsgroup comp.lang.vrml is dedicated to this subject.
A snapshot of an example of biochemical modelling with VRML is shown in the figure below.

CGI, the Common Gateway Interface standard in the WWW

The CGI standard, the Common Gateway Interface standard, is a set of rules to allow more interaction between client and server in the WWW. CGI defines how a WWW server program communicates with another piece of software on that server. That software is called a "CGI program (or script)". In other words, a CGI program handles input and output according to the CGI standard. See for instance Gonzalez 1996. CGI applications allow executing programs on the server, for instance:

- to search a database on the server (a full-text database, a structured text database, or a relational database) to use an expert system on the server
- to start and use an "Internet software agent", which addresses subsequent servers or tools, combines results, eliminates irrelevant, or double information,...
- to convert the data from a WWW form into an e-mail message
- for the remote control of a camera, a scientific instrument,...

"Webbots" or "knowbots" or "searchbots" or "meta-search engines" or "intelligent software agents"

A program belonging to this broad category is started from the client computer or from a remote server computer and:

- seeks information,
- possibly replicates itself,
- possibly interacts with other agents,
- sends reports back to the user.

Applications are news gathering, advanced searching for information, shopping / buying.
Several meta-search engines run on a server computer and are freely accessible for use through WWW. They accept a simple query formulation in the form of one or several keywords. Then the system contacts several Internet search engines, waits for the respective results, integrates these (for example by eliminating doubles) and presents this final result in the form of URL's to the user. Two well-known examples are:

- SavvySearch available at http://www.cs.colostate.edu/~dreiling/smartform.html

Several analogous programs for a client computer are available free of charge, for
instance from The Ultimate Collection of Winsock Software (TUCOWS). More information on this type of software can be found in the article by Jeapes 1996. The first generation of this kind of software is far from perfect but promising (see for instance Tegenbos and Nieuwenhuysen 1997).

Sun's Java

Java is an open programming system, developed mainly by the company Sun. Java application programs for Internet are named "applets". An applet is sent by the server computer, is interpreted / executed by the client computer, and vanishes. Java requires an interpreter built in the WWW browser software. The current leading WWW browsers, from Netscape and from Microsoft, versions 3.xx and 4.0 (beta) versions can deal with Java applets. Not all Java applets are highly interactive: most transfer a simple program to the client computer, perform a simple task, and display a text or picture on the client's screen. Some Java applets on the client can create forms for further interaction with a WWW server computer. See for instance Gonzalez 1995, Dvorak 1996, Flyn and Clarke 1996, McCarthy and Gosling 1996, Notess 1996, Read and Hall 1996, Semich and Fisco 1996, Smith 1996, Van Hof et al. 1996. Discussions about Java are going on in the Usenet newsgroup news://comp.lang.java. A collection of links to Java applets and to information about Java is offered by Gamelan [online] and by Java Applets [online]. The Gamelan collection also includes a list of ActiveX components. Internet search engines can also be useful to find a particular Java applet; for instance AltaVista [online] by Digital, in the advanced mode, allows to search with keywords for Java applets which are freely accessible in the Internet. An example of applets in the area of science and technology is the scientific unit converter. A screen shot is shown in the figure.

Microsoft's ActiveX

This is an open programming system, but only developed for the Windows platform so far, initiated by Microsoft. Using ActiveX software requires a suitable WWW browser, such as Microsoft Internet Explorer. Programs conforming to ActiveX are sent by the server computer, and executed and stored on the client computer. Examples of ActiveX software are plug-ins for Microsoft Internet Explorer to view:

- Adobe Acrobat Portable Document Format files (.PDF)
- Microsoft PowerPoint presentation files (.PPT)
- Virtual Reality Modelling Language files (.WRL)

Of course, Microsoft offers a lot of information about "their" ActiveX as well as ActiveX software, for instance at the MicrosoftÆs ActiveX Resources Area. Collections of ActiveX components are offered also for instance by Gamelan [online], by The Ultimate Collection of Winsock Software (TUCOWS) [online], and by ActiveX.com [online].

Program scripts for HTML and WWW browsers

Several scripting (programming) systems allow to embed scripts in HTML documents. This can yield more "active", "dynamic", "intelligent" WWW documents.
Such scripts should be interpreted and executed by the WWW client program. In general they allow easier development:

- than when the Java programming language is used
- than when ActiveX components are created.

On the other hand they are by definition more dependent on HTML than the more general programming language Java.

A few script languages are used:

- JavaScript (developed by Netscape) and JScript (implementation by Microsoft), independent of SunÆs Java (but indeed named after it),
- Visual Basic (developed by Microsoft).

Client-server software to run programs on a Windows server
Back to the mainframe?!

Programs for Windows are normally executed on a PC. However, it is possible to execute a Windows program on a server (and to retrieve information for instance), through the Internet, for instance by using:

- WinFrame software by Citrix on the server
- suitable WinFrame client program on the client device
- the Intelligent Console Architecture network protocol (ICA) for client-server communication.

While the program runs on the server, the graphics are displayed by/on the client computer. The WinFrame client software can even work:

- on older, low-end hardware (even PCs with a 80286 microprocessor)
- on "thin" client devices which can be cheaper than a PC

Server and client software is developed and distributed by Citrix, a company which is owned partly by Microsoft.

For a PC, WinFrame client software is available which allows running Windows programs in a WWW browser:

- a plug-in for Netscape Navigator
- an ActiveX control for Microsoft Internet Explorer

For more information, see for instance Quinn 1997 or the WWW site of the software developers at

The technique and required software will probably be integrated directly in future versions of the Microsoft Windows operating systems.

This approach allows of course highly interactive services or document+program hybrids where the program runs on the server.

Internet push technology with Web-casting channels

Channels are web sites that are automatically delivered to the desktop of the user. The system delivers personalised, up-to-date information and/or programs to the client on
the desktop of the user. The method allows self-updating in the background, when bandwidth is available. It is complementary to classical WWW which is not push but pull technology. It makes Internet and WWW more like television. The data can be displayed as a sudden screen flash, or as screen wallpaper, or as screensaver.

You can download push technology client programs, from the Internet, e.g. from http://www.jumbo.com/pages/push/

Some available products are listed in the table below.

<table>
<thead>
<tr>
<th>Company</th>
<th>Service</th>
<th>URL = http://.....</th>
</tr>
</thead>
<tbody>
<tr>
<td>PointCast</td>
<td>PointCast Network</td>
<td><a href="http://www.pointcast.com">www.pointcast.com</a></td>
</tr>
<tr>
<td>Marimba</td>
<td>Castanet</td>
<td><a href="http://www.marimba.com">www.marimba.com</a></td>
</tr>
<tr>
<td>Intermind</td>
<td>Communicator</td>
<td><a href="http://www.intermind.com">www.intermind.com</a></td>
</tr>
<tr>
<td>BackWeb</td>
<td>Polite Agent</td>
<td><a href="http://www.backweb.com">www.backweb.com</a></td>
</tr>
<tr>
<td>Netscape</td>
<td>Netcaster</td>
<td><a href="http://www.netscape.com">www.netscape.com</a></td>
</tr>
<tr>
<td>inCommon</td>
<td>Downtown</td>
<td><a href="http://www.incommon.com">www.incommon.com</a></td>
</tr>
<tr>
<td>Headliner</td>
<td></td>
<td><a href="http://www.lanacom.com">www.lanacom.com</a></td>
</tr>
</tbody>
</table>

For a review of Castanet by Marimba, see Losbin 1997. Also in this area standardisation is far from perfect at this time: most Internet push technology cannot be used simply with all WWW browser software. That makes usage more complicated. Microsoft has proposed a standard to simplify applications:


**PICS document labels in the Internet**

PICS offers a standard for the format of labels (meta-information) about documents on WWW. Software is required to make sense of these labels. See for instance Resnick and Miller 1996, Resnick 1997. This creates improved opportunities to add meta-information and thus value to documents in the Internet in ways impossible and inconceivable in the case of more classical, "unconnected" documents.

**Trusted systems in the Internet**

Trusted system rely on additional interactions to yield an additional value to the WWW communication. An important application in the area of libraries is the ability

- to ensure reliable publication,
- to specify terms and conditions for the use of digital works and to control this usage.


**Virtual social worlds/environments in the Internet**

Internet is the basis for several chat rooms, virtual communities, "worlds", or social environments. These are close to older real time communication methods applied in Internet like text chatting and IRC in particular.
Access requires software incorporated in a WWW browser or a separate dedicated client program. For a review of several real systems, see Berger 1997. Client software for this kind of systems is available free of charge, for instance from The Ultimate Collection of Winsock Software (TUCOWS). The virtual world can contain and offer information, for instance as virtual posters on virtual walls. In this way, these can also be seen as a type of document+program hybrid. Another application may be communication in real time with an expert, librarian or information intermediary in and through such a virtual world; that can perhaps serve a need, but it is not yet clear how much added value the real time communication may offer in comparison with other more traditional communication methods like telephone, or even in comparison with not real-time methods like fax and e-mail.

In the following chapters, we comment on the techniques, methods, standards and systems listed above, and on the document+program hybrids which they support.

**Relations between the recent Internet techniques**

The methods, tools, techniques, standards and systems listed above do not belong to distinct, separate, mutually exclusive categories. In other words they are not well modelled as slices of some kind of virtual pie. Instead, they are interwoven and overlapping: there are many relations among each other. This can be confusing and hamper understanding all this. The following examples illustrate this.

- Document-program hybrids can be offered in several forms/versions, to accommodate various client environments. For example: AltaVista LiveTopics is offered in a text, JavaScript and Java version.
- Document-program hybrids can incorporate more than one method and use these more or less at the same time.
- Be careful: One car can hide another! One train can be followed by another! Or in other words, several techniques are based on another one of the techniques listed. For example:
  - Advanced HTML allows program scripts.
  - Many plug-in programs use ActiveX. Viewing VRML items may require an additional plug-in program. VRML will probably be used more and more to build the virtual worlds on the Internet, which are also listed.
  - The Castanet Channels push technology is largely based on Java.
  - Some applications of push technology rely on an independent client program, while others require software built in the specific WWW browser used.
  - Client software for an application of push technology can support also the protocols of other push applications (for instance Netcaster supports also Castanet).
Critical comments on the new Internet techniques

Security?
A subject of dispute is the security of the client computer which uses files coming from the network with applications of:

- Java (in the form of applets),
- ActiveX,
- scripts in documents.


Fast evolution
Technical development is fast. So it is difficult:

- for developers of document+program hybrids to keep pace and to apply the new methods, tools, extensions and standards quickly and usefully.
- for many less experienced users to keep pace, to learn, to adapt, to understand and to apply the novelties usefully; in the first confrontations with strange document+program hybrids, they can be surprised and confused, or even be frightened.

Need for powerful computers
Several of the new methods in the Internet create the need in many applications for a more powerful client computer and network connection for an acceptable execution speed. Thus client computers age fast and become obsolete sooner than the financial director would like. More concretely, we mention here demanding:

- multimedia applications in general,
- movies (with sound) in particular,
- VRML applications in particular,
- Java applets,
- ActiveX components,
- scripts in HTML documents, and
- participation in complex virtual social worlds.

Standardisation and incompatibilities
For several techniques, standardisation is not (yet) perfect, which causes incompatibilities between document+program hybrids (the sources) on the one hand, and the universal, well equipped WWW browser client software needed to interpret and use these on the other hand. Concretely, we mention here:

- the various versions of VRML,
- Java applets,
• JavaScript and JScript.

**Ideal, easy access to information through one browser is threatened**

An ideal for many is uniform access to all kinds of information sources and communication channels by using only one universal WWW browser, grown to a great integrated Internet client software package. The advantages are clear:

• one and only one program to select, install and maintain
• the availability of that one program free of charge or at low cost
• a user friendly interface towards all information sources
• a similar, uniform user interface for information using many formats and communication protocols
• one and the same access method to information on many carriers (local disk, CD-ROM, Local Area Network or intranet, Internet,...)

However some of the methods and techniques discussed here threaten this ideal. Concretely we mention here:

• the problems with standardisation and thus with incompatibilities, which make it (still) necessary that a particular client is best used for particular document+program hybrids;
• the need to install additional viewers and plug-in programs;
• the requirement to use separate client programs besides a WWW browser for some applications of push technology and virtual worlds;
• the fast evolution of the big and growing WWW browser software packages, so that the many previous versions (older generations, releases between generations with additions and bug fixes, and beta releases of browsers), lack features required for the most ôhotô or ôcoolô applications of today and certainly of tomorrow.

**New information and communication media**

Is the subject of this contribution not comparable to discussions of other communication and information media at the time that they were or are "new"? All new media need good contents and also the medium (the container) must be well suited to deliver that contents, to be accepted by the public, the users, in other words for a successful breakthrough.

Often new media, methods, tools and techniques are used first for test and demonstration purposes. This seems also the case here. For instance many World-Wide Web pages are created and offered which incorporate one or several of the new methods, which may be impressive, but only at first sight, regretfully with a negligible information content or value in general. Nevertheless, it is clear that the methods discussed here are potentially useful. They are not yet widely implemented to publish useful academic, scientific and technological information, but the evolution is fast.

The critical comments on the new document+program hybrids make it clear that
many users (will) for example still prefer solid, old-fashioned, hard copy documents, accompanied by a hand calculator: cheaper and more robust. And they will have their preference for a long time to come.

Anyway, the document+program hybrids certainly contribute to the increased importance of computers and networks as essential, required tools in information distribution and usage. In this way they force information end-users as well as information intermediaries or librarians once again and perhaps more than ever to adjust to the permanently changing information landscape.

**Implications for information intermediaries**

Some tasks and roles of librarians and more generally for information intermediaries should receive more attention:

- to create guides to search tools and sources;
- to engage more than ever in education and training of each other and of end-users at all levels;
- to co-operate with each other and other experts in this kind of activities, because these are big tasks;
- to ensure access to Internet workstations for end-users;
- to ensure the quality and suitability of these workstations for the "work" they should support;
- to work on archiving, conservation and preservation of document+program hybrids, and on ensuring access also in the future, in spite of the fast evolution of hardware and software, which complicates already archiving of and providing access to "old" computer-based information items even when their format is more classical;
- at least in those institutes where the structures are still rather traditional with high and thick walls between a library on the one hand and computer and communication or information technology centre on the other hand: to work together more and more, informally or even formally, thereby changing historically grown structures and entities for the benefit of all end users of information and communication services.

**The future**

Computer and Internet specialists should and will co-operate more and more with information specialists and information content providers to create attractive and useful document+program hybrids. The prime focus will shift from developing and demonstrating frameworks in which information could possibly fit, to offering useful information content with a high added value, based on the new techniques.

To close the paper we repeat its main theme:

We observe the steady evolution:

from well separated, easily understandable *or* document, *or* computer program,

to more complex hybrid entities

*document or (?) program.*
Finally we can wonder about an analogous evolution:

from well separated, easily understandable or human, or computers,
to even more complex, hybrid entities:

*human or (?) computer.*

However, the authors wish to leave that last theme to science-fiction writers and to scientists in other areas.

**References**


Gonzalez, Sean. Java & Hot Java: waking up the Web: the end of browser add-ons and the beginning of Web interaction. PC Magazine (USA), October 25, 1995: pp. 265-266.


Notess, Greg R. Java creates new potential for Internet information. Online,
March/April 1996: pp. 82-86.


Resnick, Paul. Filtering information on the Internet. (Look for the labels to decide if unknown software and World-Wide Web sites are safe and interesting.) Scientific American, March 1997: pp. 54-56.


Smith, Tony. Push and pull. ("Push" technology is the net's latest killer app, likely to pull in big bucks and eventually make the web as powerful a medium as TV.) [PointCast, Marimba, BackWeb, inCommon] Personal Computer World (England), June 1997: pp. 198-204.


Virtual Reality Modelling Language [online] Available from: