



Xerox: REDEFINING THAT X-VERB

Who better to provide the new world of multilingual, transnational Total Documentation than a multilingual transnational with Total Documentation needs of its own?

Andrew Joscelyne

Like hoover and scotch tape, Xerox is one of the privileged few trademarks to become a commonly-used verb found in your average dictionary. And whether Xerox Corporation likes it or not, when most folk hear this Greek-rooted word meaning *dry* – first used to denote the product developed from Chester Carlson's 1938 invention of the xerographic image – they inevitably think of office-based photocopies.

Yet by the time Xerox had produced their two millionth copier in 1988, almost 40 years after their first Model A, the whole environment of office document handling was radically changing. This was largely due to the results of long-term R&D projects funded by Xerox itself.

Today Xerox spans the globe, with major operations in over 60 countries. It is ideally placed to refocus its technology strategy on the changing information economy of the 90s. As a far-flung multinational producing multifarious documentation for its product range, Xerox understands the needs of users who circulate corporate and technical data globally.

Xerox has set itself the task of offering a holistic solution to handling documents – from their creation and organization, to their printing and mailing – in more languages than most Berlitz schools can offer courses in.

At the heart of its new corporate communication nexus is what Xerox calls the “compound document” – a potential combination of text, structured CAD/CAM technical graphics, corporate graphics like pie charts and tables, data scanned via optical character recognition or hoicked out of a remote database and inserted in other documents, and soon, perhaps, digitized speech.

NICE PROJECT, SHAME ABOUT THE NAME

According to David Goodstein, president of Massachusetts-based Interconsult Inc. and long-time consultant to Xerox, the 90s should be a very strong period for Xerox, especially in Europe. He sees their organizational strength, superior field service and highly consistent way of providing technical support over a wide range of countries making Xerox a market leader in its field.

It's at Rank Xerox's European HQ (don't ask what Rank's re-

lationship with Xerox is – you know what multinational organization is like. They're uh, *close*), a superb new greenfield site at Marlow in the Thames Valley, where under the leadership of Lyndon Haddon, senior products development managers are master- and mistressminding the prosaically named Integrated Systems Operations (yes, *another* ISO).

This major IT project grew out of the Xerox view that the rapid growth of transborder corporate activities, combined with a customer-environment where individual vendors increasingly have to act cooperatively with competitors to solve technical problems, means that relevant information is no longer the kind traditionally handled by data processing departments in large-scale companies.

Products Marketing Chief David Jones spelled out the bottom line: “There's no more ‘it-saves-money’ blood to be squeezed from the IT stone. Everything is shifting away from the 70s idea of cutting costs via automating routine activities. Instead, the market is driving towards providing users with the tools for gaining competitive advantage. And this means increasingly doing what Xerox has always been good at – handling documents.”

Jones likes to use the pharmaceuticals industry as a concrete example of the potential of document handling to swell the corporate coffers: if the genetic engineers who've been devouring money like the Bermuda Triangle of venture capital finally discover a new drug with possible downstream commercial applications, that drug can be patented for a total of only 17 years before High Street druggists start producing an analogue. After at least five years of testing before going to market, the drug will have just 12 years of protected life. If you get the two zillion pages of documentation necessary for the Federal Drug Administration trials ready a wee bit quicker than in the past, you might lengthen the actual time your product is covered by patent protection, and hence increase your profits.

“If that document preparation period can be reduced by just 2%,” explains Jones, “you've gained another five weeks of extra patent protection – meaning anything up to US\$25 million in revenue. Only when an office system (in the widest sense of the

Foto: Dyana Van Campen



Products Marketing
Chief David Jones:
"We are probably
alone at the high end
of integrated,
multilingual
document manage-
ment systems."

Mike Garrido, Xerox's

transborder man:

"Technology
should not
be a barrier to
the use of any
language."



Photo: Dyana Van Campen

term) focuses on the document as the major information support can you start to see quantifiable cost savings."

The "compound document" may sound like Xeroxspeak for shifting files around in a multimedia environment, but according to Jones it's a way of drawing attention both to the technology and to the material support of corporate documents.

"We all predicted the paperless office four or five years ago," he admits, "But paper is proving to be a very user-friendly medium. The trouble is there has been no technology to help people manage docu-

ments that emerge as paper."

WE (STILL) ♥ PAPER

Xerox sees documents gradually becoming *objects* with an electronic half-life for storage and communication – yet still requiring paper support as a medium of use.

What you might like to have then is a total document management station that lets the compound document play out its life fully in corporate communication, now as electronic file, now as demand published text. It needs to be powerful enough to interface with the big old main-

frame database server so as to pull in and handle large chunks of data, yet friendly enough for non-specialist users to create and edit material in their own language and script without lengthy interface learning.

A tall order, yet Xerox feels it has most of the answers in its ViewPoint software. This runs on a proprietary 6085 Xerox Professional workstation, supporting networking facilities, multitasking and the integration of other software, plus printing in up to 40 languages.

"If you ask for the documentation manager in a company today, the answer

Multilingual Activities At Rank Xerox

Xerox is not just a multinational but also a multilingual corporation.

MULTIPLE SCRIPTS

"Technology should not be a barrier to the use of any language," insists Mike Garrido, Rank Xerox's Marketing Support Manager with special responsibility for multilingual activities.

Garrido is the very personification of multilingualism, having had the unlikely honor of being born in a corner of Portuguese-speaking Brazil that became a corner of Spanish-speaking Colombia soon afterwards. He's a transborder man, particularly suited for heading the team which coordinates work on what Xerox calls its "multinational language" capabilities.

With Xerox's ViewPoint software, language is a parameter of a document. Being able to shift easily between documents and scripts is the most natural thing in the world. In fact, Xerox pioneered many of the coding standards which give their software this flexibility to support multiple scripts in an intelligent way.

A standard ViewPoint package supports over 20 languages, with all the diacritic signs, accents and typical document symbols of the major "Western" languages, including Greek, Turkish, Swedish and Catalan.

A regular Babel of optional language packs covers the Arabic scripts (plus Persian variants), Hebrew, Chinese, Korean, Japanese, and many more. In Singapore, script capabilities for Devanagari languages and Bengali are under development. Altogether, a total of 40 different "strategic" (i.e. commercial) languages, in a rich variety of scripts, may co-exist in a single document, both onscreen and as printer output.

ViewPoint uses a "virtual" keyboard in the lower left hand corner of the screen, which transforms a QWERTY or AZERTY keyboard into the standard keyboard for the language in question. A Polish typist can key in text by calling up the virtual Polish keyboard, without having to change systems. With a full WYSIWYG screen and a set of virtual keyboards, all hooked up to a high-quality printer, documents can typically be created in any combination of those 40-plus languages.

The keyboard facility also allows customization: you can design your own virtual keyboard, adding keys relevant to your particular language needs. Under ViewPoint you can also enter Japanese via a Latin alphabet keyboard; the software automatically converts the

input to Kanji script. This utility has become a standard feature of Japanese language technology, but Xerox are extending it to other inter-script situations – with Latin-Cyrillic and Latin-Arabic currently under development.

INTEGRATION EXPERTISE

Xerox is also working on more standard language smarts: spell-checkers (only UK and US English are available, so far), autohyphenation rigs, specialized glossaries for legal and medical terms, dictionaries, and – eventually – thesauri in the major languages. Using Xerox PARC-originated compression algorithms (favored by the likes of Proximity and Microlytics), Xerox is tooling up its docware with large wordlists. These will become error-checking utilities for OCR software.

The company is also designing interfaces to bridge the gaps between different sources of tools within a multilingual environment, among them such future standards as Electronic Data Interchange (EDI). Screen and error messages for ViewPoint software are now edited in a special routine that allows expansion for foreign language versions – as is currently the case with many developers.

In addition, Xerox is planning interfaces to allow workstations to access online terminology banks at the double click of a mouse, giving rapid access to the massive EUROCAUTOM and other Infoterm sources. To make sure that they cover all the technological options for information accessing, they are also considering integrating CD-ROM-driven multilingual dictionary materials into the desktop environment.

With this integration expertise growing in a variety of multilingual applications, Mike Garrido would like to see Xerox position its internal resources in a more commercial environment.

"So far there hasn't been much of a commercial offer of services in computational linguistic tools," he comments. "I would like to see the company spin-out some of Xerox's development expertise into joint-ventures or service outfits to develop a full range of more powerful language tools to supply the market.

"Ironically, I feel we need a company actually outside Xerox in order to develop more dynamic ways of doing business in this area." A first attempt in this area would appear to be Rank Xerox's collaboration with technical communication company Phoenix for translation management.

If Mike Garrido his gets way, there may well be more.

will be the same as asking for the DTP manager six years ago: 'Who?'" quips David Jones. "There are currently a large number of DTP systems and wordprocessors for low-end, small-scale needs. Yet we are probably alone at the high end of integrated multilingual document management systems."

THE BEEF

In many ways, ISO builds on the by-now familiar WIMPS (windows, icons, mouse, etc) interface technology originally developed in the 1970s by Xerox researchers at PARC in Palo Alto, Califor-

nia, and then made famous in "other environments."

It also exploits another Xerox R&D product – Ethernet, the well-known local area network standard. This now forms the bottom layer of Xerox Network Services (XNS), a powerful software and network architecture.

After a vigorous, but clearly difficult effort to get its own proprietary 6085 hardware platform accepted, Xerox will concentrate future development of the compound document management system on opening it up to industry standard platforms. By year's end, ViewPoint software

will be running on Sun SPARCstations under Unix, and users will be able to integrate DEC or IBM PC applications into their Xerox desktop work environment.

"In the long run," explains Jones, "This means that Xerox will concentrate on software development. The current 6085 will eventually be phased out, as Sun workstations and Unix become cost-effective alternative hardware platforms."

Just as the compound document itself is composed of elements from a variety of structurally different sources, so Xerox's potential customers have made investments over the years in a variety of differ-

Translation Technology At Rank Xerox

As a multinational company, Xerox has a sizeable translation load, and an organized management structure (created in 1975) to cope with it. In meeting its inhouse needs, the company has found the product ideas and philosophy to launch itself into the MT arena.

1992: TRANSLATION EXPLOSION

Xerox's US translation center started using the Systran MT system in the late 70s. Today, the Systran software is installed at a site in Webster, New York. It's used for the translation of photocopier service and maintenance manuals from English into French, German, Spanish and Italian. To speed up the translation process, Xerox astutely introduced a special simplified writing and vocabulary style. Called Multinational Customized English (MCE), it is now the required style for Xerox documentation authors.

In 1985, a translation unit was formed at Rank Xerox in the UK. Based just north of London in Welwyn Garden City, the unit has since expanded from eight to 40 employees: 22 systems and translation support staff, nine translators, and nine editors, who coordinate the work of freelance translators.

The UK translation unit accesses the Webster-based Systran system online. For converting English into languages not supported by Systran – Portuguese, Dutch, Danish, Finnish, Norwegian and Swedish – they use ALPS translation software.

The material translated here is software and user documentation for copiers and typewriters. This kind of text has a wider variety of syntactic forms than the maintenance manuals translated in the USA, so it needs more post-editing. The interactive nature of ALPS software is particularly suited to this process.

The translation load of the UK unit has increased from around 1M words in 1985 to 5.5M words last year. It is expected to double, reaching 11M words (or 44,000 pages) in 1990. Consequently, Rank Xerox have been obliged to sub-contract a part of the load to documentation and translation service company Phoenix. The company, which has a 13-strong unit dedicated to Xerox's needs, also uses the ALPS system.

But the translation problem has not been licked yet. Xerox forecasts an annual load of 25 M words (that's 100,000 pages) by 1992, as new language markets combined with new products – such as documentation-intensive network systems, like the XC80 – generate a monstrous crunching of lexemes.

VIEWPOINT ON MT

To anticipate some of these needs in the context of Xerox's genuine commitment to MT, plans are underway for a range of interfaces allowing direct communication between ViewPoint (Xerox's state-of-the-art document processing software) and Systran and ALPS.

Xerox has already developed an interface connecting ViewPoint and Systran, and is working on one to link ViewPoint and ALPS. The present interface is used inhouse, for translations from English only. Documentation is created on a Xerox workstation, stripped of its publishing/printing formatting, transferred to the MT server, retrieved in the target language, and finally reintegrated into the original document format, complete with fonts, paragraph attributes and other features.

True to their global systems integration philosophy, Xerox eventually intends to commercialize this front-end to Systran, providing the perfect integrated translation solution for clients such as British Steel. Xerox licensees will be able to send text from ViewPoint, IBM software, or the ViewPoint MCE facility, via a Remote Batch Server on XNS to their company mainframe, receiving translated documentation that can be displayed or printed side-by-side with the original on a 6085. Productivity gains from minimal document post-editing promise to be significant.

In Europe, however, according to European Commission statistics, translation from English source text represents only 40% of user demand. Consequently, Xerox plans to develop the same kind of dynamic interface facility for non-English source language input to Systran and ALPS.

In the even longer term, Mike Garrido would like Xerox and Rank Xerox to offer ViewPoint front-ends for the *whole range* of current second generation MT systems – Logos, Metal (they are particularly interested in the market possibilities of this Siemens rig), Tovna, GETA (the Grenoble engine) and, with reckless optimism, Eurotra ("We think they'll have to come up with something very soon," smiles Garrido).

Obscure Russian and Hungarian rigs will also be examined, and over in Japan the Fuji Xerox subsidiary is looking at interfaces for Fujitsu's Atlas and Bravice's Japanese remake of Wiedner software. At this rate a fully integrated Xerox photocopier of the future might translate your document as it copies it.

And maybe the dictionary will one day read, *Xerox v. i) to copy; ii) to translate.*

ent computer technologies.

With the accent in the 90s on solving users' management problems rather than on lobbying geewhizz boxes of tricks at them, being able to offer soft- and hardware conforming to open standards will be crucial. Xerox believes this will add value to their integrated systems and allow them to be more easily integrated into current customer configurations.

"Look at the development of standards in the US military," says Jones. "Documents accompanying bids for equipment and weapons systems will have to be CALS (Computer-aided Acquisition and

Logistics Support) compliant or you won't get the contract. This isn't just a problem for a few big contractors, but for all the subcontractors that those prime contractors use.

"We started looking into CALS compliant systems about 18 months ago, both in the USA and now in Europe, and our ViewPoint document software will incorporate this standard, along with SGML (Standard General Markup Language) as a substandard."

If schedules are kept, Xerox might also be the first document manager to bring out an ODA (the emerging European

standard Office Document Architecture) interface by the end of the year. While SGML is aimed at the producer of highly-structured documents on a large scale, ODA is geared to *ad hoc*, unstructured docs around the office environment. "Despite what has been suggested, there's no real conflict between SGML and ODA," reckons Jones. "Rank Xerox are backing both horses in any case."

The Xerox strategy of opening its software to industry standards is also at work down in the very language parameters of their software. When ViewPoint software was originally developed in the mid-70s with a multilingual capability, Xerox developed a character set with a then-revolutionary 16-bit code offering sufficient space to handle a vast potential range of languages, including a range of Chinese and Indian languages.

Xerox consulted the ANSI, ASCII, ISO, JIS and ASMO international standard codes so that there would be full support for these codes in their own character space.

DOC AROUND THE CLOCK

If the 19-inch bitmapped screen of the current 4085 Xerox workstation offers a very smooth environment for creating and editing documents, it will usually constitute a single node in a whole network of stations.

The French company Arianespace, for example, uses Xerox software to handle its massive space program technical documentation. With an HQ in France and a test site on the South American coast at Kourourou (Guyana), it needs a fully communicative approach to the management of its documentation, allowing staff to mail files to-and-fro without spending all day punching keys.

So does Boeing, which uses Xerox for creating single documents from different sources scattered around the USA. And so does Voice of America, to prepare its bulletins in various languages around the world.

The compound document they all use is essentially a document of multiples – multiple authorship from multiple locations using multiple sources of data from multiple applications – in a multiplicity of languages. To ensure there are links to support this underlying proliferation of components, Xerox needs the powerful network resource of XNS "the backbone of the document management system," as Jones puts it.

Xerox should be well-placed to integrate XNS with industry networking

standards, since its architect, Jim White, is also chairing the international standards committee for X.400 mail, fast becoming a worldwide standard. Not surprisingly, X.400 already bears a strong resemblance to XNS.

E-mail and networking are not just technical nice-to-haves touted by the marketing boys to dubious customers; they play an essential competitive role in the distributed corporate system itself. Xerox itself is wired together around the globe in a powerful illustration of communicative networking called *Internet* (not to be confused with its oft-infiltrated US homonym).

At Marlow alone, some 800 staff with access to 500 workstations regularly communicate via E-mail by double clicking an icon to activate the facility. Mail is delivered not to a workstation but to a server close to it, so that wherever users might be on the globe, they can call up their own mail server and have their messages sent to the workstation where they happen to be logged on.

Beyond the local network, dedicated lines fuse the whole corporation together worldwide, allowing some 35,000 Xerox users to send messages to each other, globalizing awareness and expediting the corporate business. As David Goodstein says: "Xerox is its own biggest customer when it comes to networking products."

A typical use of this kind of facility, allowing both globalization of product development and a net gain in time and cost can be found in Rank Xerox's translation department (see *Translation Technology* sidebar), from where bits of doubtful text are regularly zapped-off to relevant local country operations for the OK on sensitive terms.

I'VE SEEN THE FUTURE, AND IT...

Other doc-management technologies being actively pursued by Xerox include scanning by Optical Character Recognizers (OCRs) and, more recently Intelligent Character Recognizers. These use rule-driven pattern recognition, rather than traditional pattern matching techniques.

Developed through their Kurzweil subsidiary in the USA, the aim is to provide scanning systems which will allow text in any font in any language to be input to a document. Xerox is also developing a series of interfaces to relational databases to allow users to pull structured information straight into their application with the minimum of manipulation. Again, it's a case of integrating facilitated

access into the basic document handling environment.

In a return to its roots, Xerox is in the process of revolutionizing our old friend the light-lens photocopier. The high concept is that by splitting the functions between a scanner and a laser-printer, tomorrow's "photocopier" will be a total interface for documents entering an organization. As a data-gateway, it will allow information to be created, manipulated, filed and printed, blurring the distinction between documents from outside and inside a company.

Dubbed "system reprographics," it will effectively allow an entire integrated information management and publishing system to be orchestrated from a plain vanilla PC.

In a final shift to industry standards, still under wraps but poised to elicit a further collective "Wow" from low-end marketland, Xerox are planning to port ViewPoint software with all its networking and document processing smarts to the PC environment. With a product portfolio like that, saying "I'll Xerox that" in a few years time could have a whole new meaning in the dictionary.

Andrew Joscelyne is EW's Paris-based peripatetic, polysyllabic, senior contributing editor.



"There's no more 'it-saves-money' blood to be squeezed from the IT stone. Everything is shifting away from the 70s idea of cutting costs via automating routine activities and driving towards providing users with the tools for gaining competitive advantage." – David Jones