KNOWLEDGE-DOMAIN INTEROPERABILITY AND AN OPEN HYPERDOCUMENT SYSTEM

Douglas C. Engelbart

Bootstrap Project,
Stanford University
Document #: (AUGMENT,132082,)

Reprinted from *Proceedings of the Conference on Computer-Supported Cooperative Work,* Los Angeles, CA, Oct 7-10, 1990, pp. 143-156. (AUGMENT,132082,). Also republished in *Hypertext/Hypermedia Handbook*, E. Berk & J. Devlin [Ed.]. McGraw-Hill, 1991.

INTRODUCTION 1

This paper anticipates that the tools and methods of computer-supported cooperative work (CSCW) will become harnessed with revolutionary benefit to the ongoing, everyday knowledge work within and between larger organizations. Toward that end, the following concerns about interoperability between knowledge-work domains will have to be met, and something such as the "open hyperdocument system" must become available for widespread use. 1a

As computers become cheaper and we learn more about harnessing them within our cooperative work, they will come to support an increasing number of different domains of knowledge work. Moreover, the sphere of computer-supported activities within each domain will steadily expand as more function and more skill become employed. 1b

It is predictable that increasing functional overlap will occur as these expanding domains begin to overlap. It has become apparent to me that someday all of our basic knowledge-work domains will be integrated within one coherent "organizational knowledge workshop." This leads to thinking about an over-all, integrated architectural approach to the ever larger set of common knowledge work capabilities emerging within a multi-vendor environment. 1c

Much has been accomplished to date in standards and protocols in the highly active field of networked workstations & servers, where "interoperability between hardware and/or software modules" has become a central theme. 1d

This paper considers the "interoperability between knowledge domains." This interoperability theme will be increasingly important for a workable CSCW framework as the scope and degree of CSCW increases. Dramatic increases will predictably create a marked paradigm shift about how to organize and operate cooperative human endeavors. I think that two phenomena will yield changes and a paradigm shift that will make this interoperability of paramount importance: 1e

(1) With a relatively unbounded technological frontier together with immense and growing economic pressure, the speed, size and cost of computers, memory, and digital communications will continue improving by geometric progression; 1e1

(2) Awareness and importance of CSCW is emerging, with a predictable trend toward our doing more and more of our personal and cooperative knowledge-work online. 1e2

Assuming an inevitably gigantic scale for our inter-knit "CSCW world" provides some important guidance for the continuing investment of our business resources and professional time. 1f

For one thing, each year earlier that an effective degree of knowledge-domain interoperability is in place within important organizational or institutional domains could be worth hundreds of millions of dollars -- could mean the difference between vitality and sluggishness. 1f1

And for another, we would prefer to avoid investing our research, product development, or organizational-change resources toward ends that won't be interoperably compatible within that future, radically different paradigm. 1f2

INTEROPERABILITY IN AN INDIVIDUAL'S KNOWLEDGE WORKSHOP 2

To begin with some very basic knowledge-domain interoperability issues, consider your own (future?) "Computer-Supported Personal Work" (CSPW). Assume that you have acquired a fairly comprehensive, online "knowledge workshop," you have found better and better software packages to support the kinds of tasks shown in Figure 1: 2a

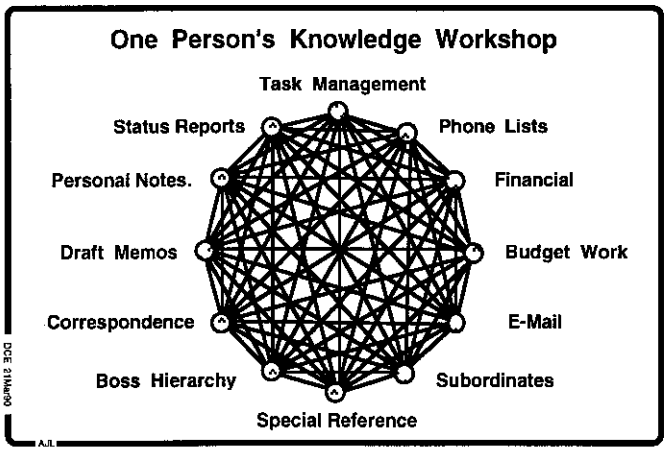

Figure 1. Each functional domain is a candidate for working interchange with all others. 2a1

Consider what you will some day have when your individual workshop inevitably becomes truly integrated. Between the E-Mail and the task management files, or the status reports, or whatever, you really would like to tie these functional domains together with a flexible free-flow of information and linkages. 2b

What kind of interoperability do you have now? I happen to think that the interoperability provided today within most CSPW domains has a great deal of improvement yet to be pursued. But I'd resist any serious arguments about this unless it be approached within the context of a coherent "CSCW interoperability framework" such as outlined below. Let me say in warning, though, that from such a framework I will contend that the marketplace for CSPW will change drastically as CSCW takes hold within our larger organizations and their inter-organizational communities. 2c

INTEROPERABILITY IN A GROUP'S KNOWLEDGE WORKSHOP 3

Suppose that you and a colleague each have a fully integrated CSPW domain, comprised of nicely interoperable sub-domains as in Figure 1. And suppose that you want to work together online. Consider the interoperability between your respective knowledge-work domains, as in Figure 2. 3a

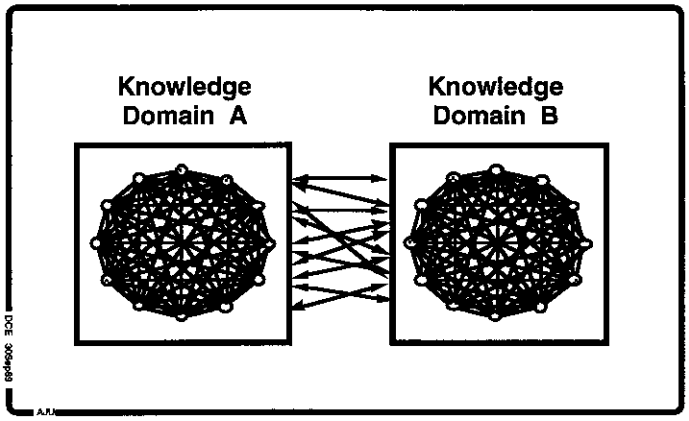

Figure 2. Close cooperation between compound knowledge domains puts new demands on knowledge-work interchange.

3a1

Now you're faced with a new challenge and a new problem. You might set it up so you have a few lines that cross between domains, but why stop there? When do two people in intense cooperative work NOT need total interoperability? In fact they depend on it heavily in the paper world. Why not online? 3b

INTEROPERABILITY ACROSS TIME AND SPACE 4

Yet another example of multiple domains is found in the familiar time-place matrix shown in Figure 3. In many cases, activities in the different quadrants involve the same substantive work content. Is knowledge-work interoperability between the quadrant domains an issue? Very much so. For example, face-to-face meetings need to flexibly utilize anything from the whole organizational knowledge base, and the meeting's records should immediately become an integral part of that same base for later-time work. 4a

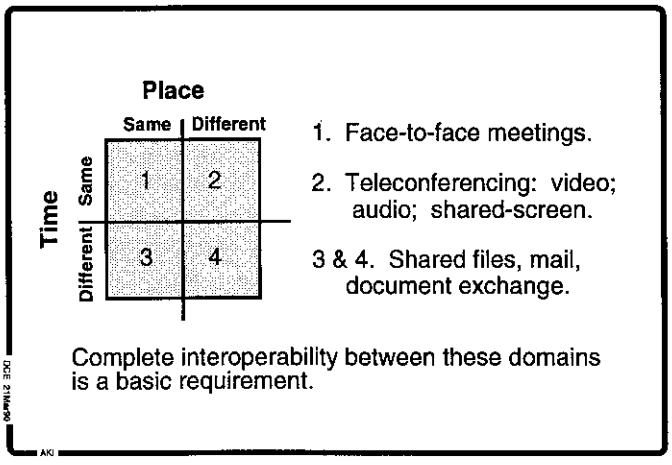

Figure 3. Collaborative processes generally considered 4a1

A Point About Online Group Knowledge Work 5

The matrix in Figure 3 is very neat and ordered. Here in Figure 4 I offer another picture of multi-domain, group knowledge work which isn't so cleanly laid-out. This reflects how I feel about the various knowledge-work domains with which my CSPW domain must interoperate. 5a

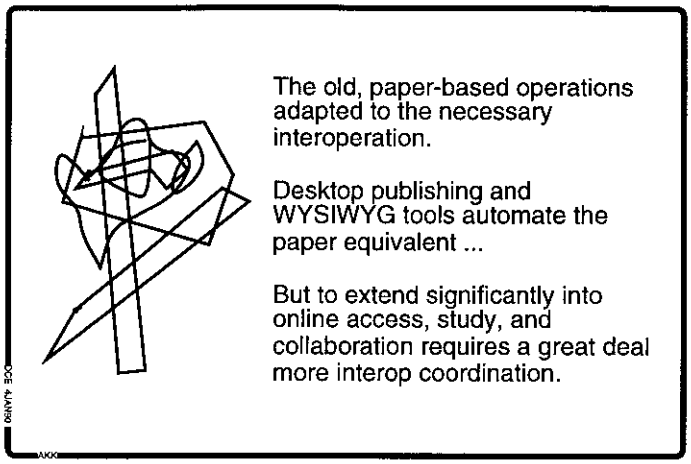

Figure 4. Consider some knowledge domains with which you intersect significantly. 5a1

The purpose of interoperability is to avoid having information islands between which information cannot flow effectively. Since we grew up in a paper-based framework, we've given little thought about how much exchange and interoperability support we really do have, and how much we depend upon it. To be interoperable in our CSPW world we could simply print out and hand over the hard copy. With WYSIWYG screens and Desktop Publishing, we're doing that with nicer paper, faster. 5b

So when we inevitably move from computer-supported paper generation and exchange to computer-supported online creation and exchange, we will need the same level of interoperability. And as the number and scale of knowledge domains involved in a given CSCW "web" increases, so does the need for "online interoperability." 5c

INTEROPERABILITY ACROSS KNOWLEDGE DOMAINS 6

To appreciate the extraordinary complexity of heavy industrial knowledge work, and the associated requirements for interoperability, consider the important functional domains within a large manufacturing organization producing a complex product, such as an airplane. It is a serious enough challenge to provide effective interoperability among the knowledge workers within any one of the domains in Figure 5; just consider the inter-domain challenge. And then consider that some of these domains, such as customers and suppliers, exist "outside" the organization, each with its own equally complex multi-domain structure. 6a

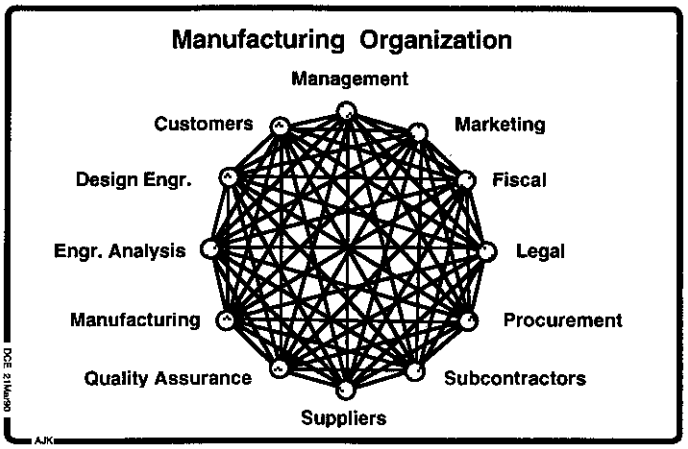

Figure 5. Each functional domain is a candidate for working interchange with all others.

THE LARGE MATRIX ORGANIZATION 7

An interesting example comes from my time at McDonnell Douglas Corporation, where I marvelled at how something as complex as one of their airplanes gets a business plan, and gets designed, manufactured, flown, and supported. Look at any given project or program ("P1" through "Pn" in Figure 6), and the functional support that's required ("F1" through "Fn"), and the exchange that needs to happen within this matrix. 7a

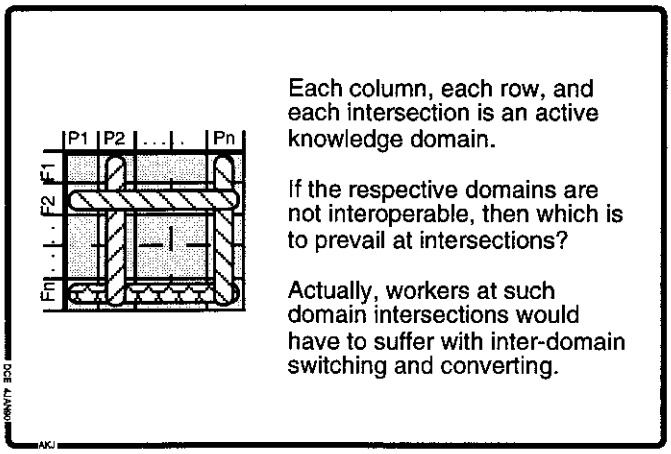

Figure 6. Consider the domains within a matrix organization of projects and function. 7a1

Each function has to share and exchange working information with many programs, and each program has to share and exchange with many functional support areas. Wherever there isn't mutual interoperability, the workers at the domain intersections will have to suffer with inter-domain switching and converting -- which is very expensive. Depending upon this kind of functional program matrix environment will require knowledge-domain interoperability across the whole organization. 7b

THE AEROSPACE INDUSTRY AS A CASE IN POINT 8

To really appreciate the magnitude of this situation, let's look inside one of those aerospace programs. 8a

A Large Aerospace Program. McDonnell Aircraft Company is participating in a bid to build the Advanced Tactical Fighter ("ATF") for the Air Force. It's possibly one of the most technically complex products anyone has ever dealt with. 8b

On top of that, they have an urgent mandate to start practicing "concurrent engineering," where the designers have to work concurrently with the manufacturing engineers. This will require intense back-and-forth cooperation between the two knowledge domains, which no one really knows yet how to do on such a large scale. 8c

Also, significant design and manufacturing problems are often delegated to the first-tier suppliers shown in Figure 7, so the cooperation with that tier is also close and intense. Then the first tiers hand off to the second tiers, and so on. So, all-in-all, you have something like 6,000 companies cooperating -- each a separate, complex, knowledge-work domain. They are expected to keep track of all business- and technical-exchange records throughout the design and manufacturing process: 8d

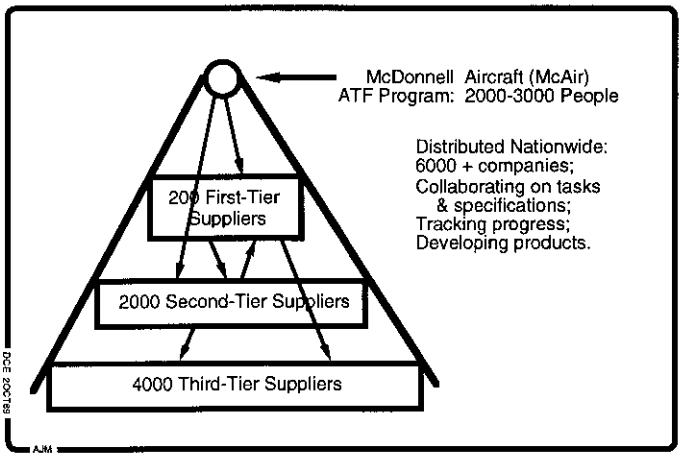

Figure 7. Islands in supplier hierarchy of a major aircraft program would be very costly. 8d1

I should point out here that the arrows in the diagram represent the legal flow of contracts being awarded. The actual exchange of documents would be shown as a two-way flow of continual negotiation and refinement throughout the design and manufacturing process -- developing the specifications, proposals, change orders, testing records, and so on. And for any part within any airplane, the manufacturer must later be able to identify when it was delivered, by whom, and even who was the shop foreman at the time of assembly. 8e

Also, a program of this size in the aerospace world would typically comprise a 10 to 30 year life cycle. So when we talk of Different Time / Same Place, and Different Time / Different Place (Figure 3), the definition of "Time" includes decades, not just hours or days. Even in a short time span and without turnover, it is not unheard of for a project team, in any industry, to occasionally lose sight of some important design decision trails, and consequently waste time and money repeating old discussions or past mistakes. Consider the likelihood, and the cost, of such lost history occurring in this long-term environment. 8f

To comply with the Department of Defense's (DoD's) forthcoming Computer-aided Acquisition and Logistic Support (CALS) mandate, all documents exchanged between the DoD and its contractors must be transmitted, updated, and managed in a standard, computerized form -- a truly gigantic interoperability challenge. 8g

Two Companies Teaming. The situation is even more complex: as with most new, large-system, DoD procurements, the Air Force requires ATF bidders to be joint-venture teams comprised of major aerospace firms. In this case, McDonnell Aircraft is teaming with Northrop Aircraft. Figure 8 shows how Northrop would form its part of the program, with several thousand workers internally, in close collaboration with several tiers of suppliers: 8h

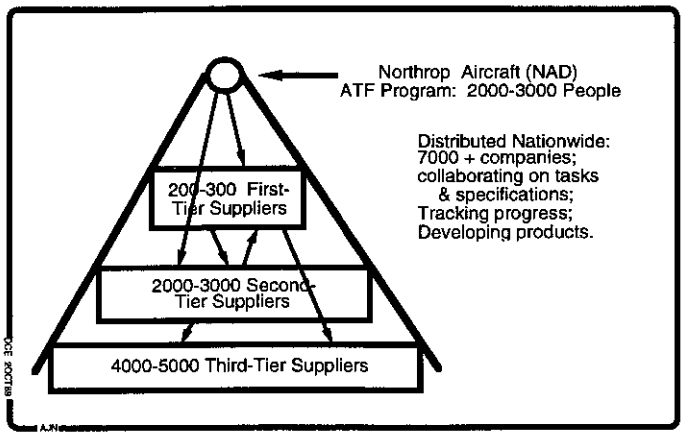

Figure 8. Islands in supplier hierarchy of a major aircraft program would be very costly. 8h1

And then picture the two companies as a team (Figure 9), and consider the intense demands for interoperable recorded document exchange across functional support and project domains within this ATF-contractor team -- within each company, between the two companies, and between them and the DoD (remembering the CALS initiative). 8i

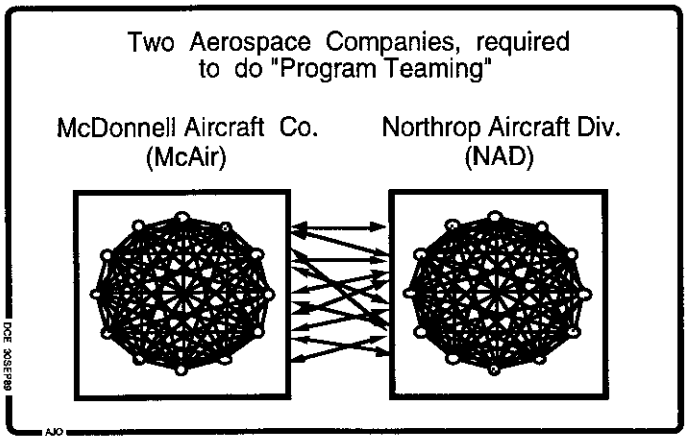

Figure 9. Close cooperation between large organizations puts new demands on knowledge-work interchange. 8i1

And then consider Figure 10 and all of the recorded interchange between these two companies and their supplier hierarchies, throughout the multi-decade life cycle of the program. 8j

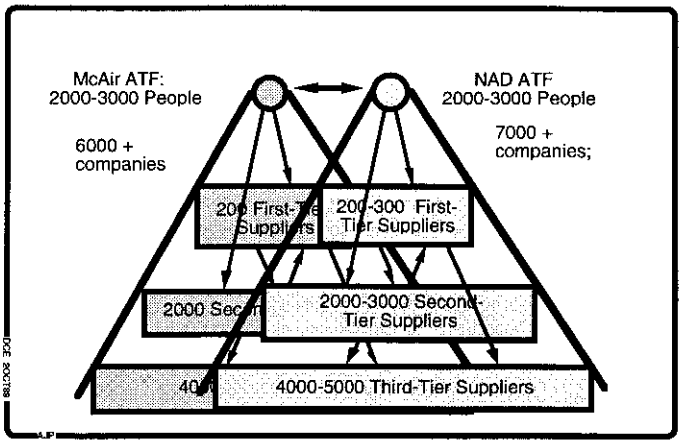

Figure 10. Teamed aerospace program -- immense demand for knowledge-work exchange. 8j1

The Web Of Aerospace Relationships. Now consider all the other large-program webs of aerospace contractors, suppliers, and customers represented by the small sub-set shown in Figure 11. A great many of these suppliers and customers will work with many of the same contractors. The complexity becomes staggering. Within such an inter-knit web of cooperative knowledge domains, there is no practical solution for effective interoperability other than industry-wide standards -- adhered to by contractors, customers, and suppliers. 8k

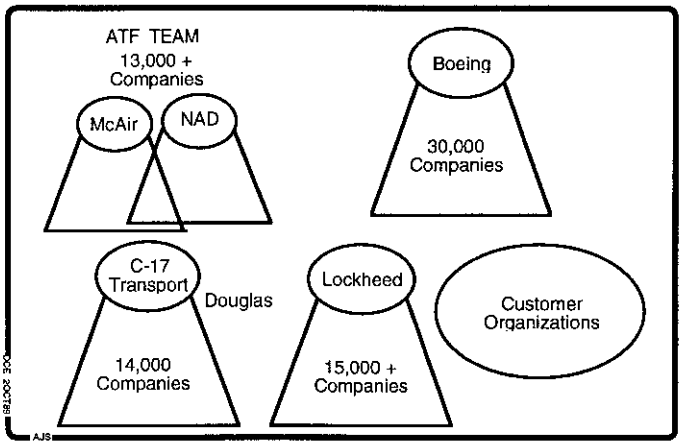

Figure 11. With common customers and suppliers, an aerospace industry can't afford islands. 8k1

And every other large industrial sector must also achieve CSCW interoperability. And those sectors must themselves interact effectively. The CSCW-interoperable web will cover the world, as has clearly been or will be done for transportation and communications (e.g. telegraph, telephone, radio, or TV). I think a strong case can be made that the cost of NOT having total knowledge-domain interoperability would far exceed the cost of achieving this interoperability. 8l

So how will this urgent need be satisfied -- for intense, computer-supported cooperation across the knowledge domains of our rapidly approaching future world? It would seem that our "CSCW future" must include something like the solution characterized below as "an open hyperdocument system." And if so, then all of our research, product development and application exploration should align with and properly affect the concepts and principles by which the future state is pursued. 8m

Towards An Open Hyperdocument System 9

Several years ago at McDonnell Douglas Corporation we coined the term "Open Hyperdocument System" (OHS) and began to define the associated functional and interoperability requirements for the kind of wide-area online cooperative knowledge work described above. This followed several years of careful study, and some pilot trials -- one of which involved thousands of knowledge-workers using a prototype system containing many of the required capabilities. 9a

Note: McDonnell Douglas is poised to move forward with requirements such as below as the basis for functional specifications and a workable procurement process. 9a1

In the following, I assume a need to provide basic capabilities so generic as to satisfy both the CSPW and CSCW application requirements over a broad spectrum of knowledge domains within a wide variety of organizations -- including for instance universities, standards groups, and the U.S. Congress. 9b

SOME GENERAL ASSUMPTIONS 10

In an open hyperdocument system, basic standards for document architecture are of course important. But beyond that, facilities for creating, transporting, storing, accessing and manipulating the hyperdocuments are embedded within an open, interoperable information-system environment, and the combined functionality is available within the knowledge-work domains of every class of worker (working from any vendor's terminal/ workstation of suitable capability). Under these conditions, the role and value of hyperdocuments within groups, and between groups, offers very significant improvements in productive knowledge work. 10a

Two unique issues differentiate this new environment from document-support systems to date: (1) interlinkage between objects arbitrarily located within a large, multi-topic and extended-history document & data collection; and (2) extensive, concurrent, online utilization for creating, studying, organizing and linking within and between the many overlapping and nested knowledge domains. 10b

These differences introduce paradigm shifts that produce different system requirements from those that have been evolving in the predominantly CSPW marketplace. For instance, WYSIWYG will give way to WYSIWYN -- "what you see is what you need (at the moment)" -- providing different options for how you'd view selected portions of the document space in your windows. The WYSIWYG view would be but one option (and likely to be utilized with decreasing frequency). Other expected shifts are implicit in some of the following suggested OHS requirements. 10c

Besides special, "document-system architecture" features, full achievement of large-domain CSCW gains awaits two things: 10d

(1) widespread implementation of integrated, open-system architectures; and 10d1

(2) widespread adoption of new knowledge-work processes (or, "knowledge processes"). 10d2

To me, these new knowledge processes are especially relevant. They will involve new systems of skills, conventions, roles, procedures, methods and even organizational structures. I believe that they will provide a much more effective

matching of basic human capabilities to the heavy knowledge-work and collaborative tasks within the functional human groupings that we call "organizations," and within the mission-specific groupings that we call "projects." 10e

In my experience, truly effective new knowledge processes will emerge only via a co-evolutionary process -- new knowledge processes and the new tools evolving together in real working environments. Explicit evolutionary pursuit with numerous, well-run pilot groups, seems called for. 10e1

From this is derived the position that a really good set of requirements and functional specifications for an OHS can only emerge from solid prototypical experience, in which advanced knowledge processes were developed and exercised along with advanced tools. 10e2

Note that the following list was derived from extensive experience with the evolution of the AUGMENT System (an OHS prototype owned by McDonnell Douglas) and its concurrent application within numerous real-work pilots. 10e3

ESSENTIAL ELEMENTS OF AN OHS 11

Mixed Object Documents -- to provide for an arbitrary mix of text, diagrams, equations, tables, raster-scan images (single frames, or even live video), spread sheets, recorded sound, etc -- all bundled within a common "envelope" to be stored, transmitted, read (played) and printed as a coherent entity called a "document." 11a

Explicitly Structured Documents -- where the objects comprising a document are arranged in an explicit hierarchical structure, and compound-object substructures may be explicitly addressed for access or manipulation of the structural relationships. 11b

View Control of Objects' Form, Sequence, and Content -- where a structured, mixed-object document may be displayed in a window according to a flexible choice of viewing options -- especially by selective level clipping (outline for viewing), but also by filtering on content, by truncation or some algorithmic view that provides a more useful view of structure and/or object content (including new sequences or groupings of objects that actually reside in other documents). Editing on structure or object content from such special views would be allowed whenever appropriate. 11c

The Basic "Hyperdocument" -- where embedded objects called "links" can point to any arbitrary object within the document, or within another document in a specified domain of documents -- and the link can be actuated by a user or an automatic process to "go see what is at the other end," or "bring the other-end object to this location," or "execute the process identified at the other end." (These executable processes may control peripheral devices such as CD ROM, video-disk players, etc.) 11d

Hyperdocument "Back-Link" Capability -- when reading a hyperdocument online, a worker can utilize information about links from other objects within this or other hyperdocuments that point to this hyperdocument -- or to designated objects or passages of interest in this hyperdocument. 11e

The Hyperdocument "Library System" -- where hyperdocuments can be submitted to a library-like service that catalogs them and guarantees access when referenced by its catalog number, or "jumped to" with an appropriate link. Links within newly

submitted hyperdocuments can cite any passages within any of the prior documents, and the back-link service lets the online reader of a document detect and "go examine" any passage of a subsequent document that has a link citing that passage. 11f

Hyperdocument Mail -- where an integrated, general-purpose mail service enables a hyperdocument of any size to be mailed. Any embedded links are also faithfully transmitted -- and any recipient can then follow those links to their designated targets in other mail items, in common-access files, or in "library" items. 11g

Personal Signature Encryption -- where a user can affix his personal signature to a document, or a specified segment within the document, using a private signature key. Users can verify that the signature is authentic and that no bit of the signed document or document segment has been altered since it was signed. 11h

Access Control -- Hyperdocuments in personal, group, and library files can have access restrictions down to the object level. 11i

Link Addresses That Are Readable and Interpretable by Humans -- one of the "viewing options" for displaying/printing a link object should provide a human-readable description of the "address path" leading to the cited object; AND, that the human must be able to read the path description, interpret it, and follow it (find the destination "by hand" so to speak). 11j

Every Object Addressable -- in principle, every object that someone might validly want/need to cite should have an unambiguous address (capable of being portrayed in a manner as to be human readable and interpretable). (E.g., not acceptable to be unable to link to an object within a "frame" or "card.") 11k

Hard-Copy Print Options to Show Addresses of Objects and Address Specification of Links -- so that, besides online workers being able to follow a link-citation path (manually, or via an automatic link jump), people working with associated hard copy can read and interpret the link-citation, and follow the indicated path to the cited object in the designated hard-copy document. 11l

Also, suppose that a hard-copy worker wants to have a link to a given object established in the online file. By visual inspection of the hard copy, he should be able to determine a valid address path to that object and for instance hand-write an appropriate link specification for later online entry, or dictate it over a phone to a colleague. 11l1

HYPERDOCUMENTS IN A GENERAL INTEGRATED ARCHITECTURE 12

Besides the aforementioned Hyperdocument Mail and Hyperdocument Library features, there are other important CSCW features that are dependent upon an "integrated system". 12a

Shared-Window Teleconferencing -- where remote distributed workers can each execute a related support service that provides the "viewing" workers with a complete dynamic image of the "showing" worker's window(s). Used in conjunction with a phone call (or conference call), the parties can work as if they are sitting side-by-side, to review, draft, or modify a document, provide coaching or consulting, and so on. Control of the application program (residing in the "showing" worker's environment) can be passed around freely among the participants. 12b

Inter-Linkage Between Hyperdocuments and Other Data Systems -- for instance, a CAD system's data base can have links from annotations/comments associated with a design object that point to relevant specifications, requirements, arguments, etc. of relevance in a hyperdocument data base -- and the back-link service would show hyperdocument readers which passages were cited from the CAD data base (or specified parts thereof). 12c

Similarly, links in the hyperdocuments may point to objects within the CAD bases. And, during later study of some object within the CAD model, the back-link service could inform the CAD worker as to which hyperdocument passages cited that object. 12c1

External-Document Control (XDOC) -- Same "catalog system" as for hyperdocument libraries -- with back-link service to indicate links from hyperdocument (and other) data bases, for any relevant material that resides offline or otherwise external to the OHS. 12d

THE INTEROPERABLE OHS ENVIRONMENT 13

Here's what the share-and-exchange domain within an open hyperdocument system might look like: 13a

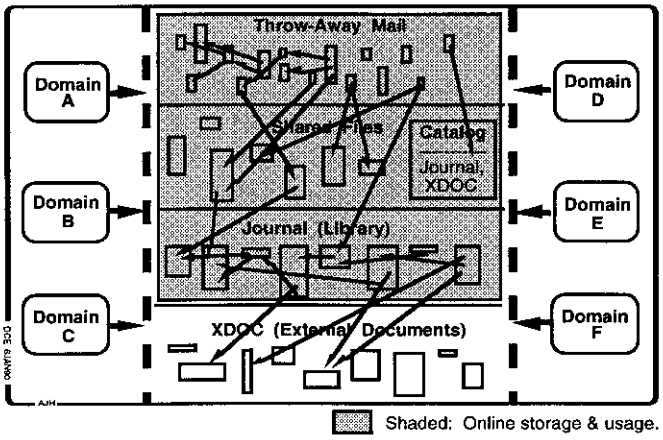

Figure 12. Knowledge-domain interoperability is greatly enhanced by hypertext linkage capability. 13a1

The requirements outlined above form a basic support platform for any group knowledge work effort, with interoperability across time and space (including all quadrants of the Time / Place matrix), across knowledge domains, and across organizational domains. 13b

THE INTEROPERABILITY INVESTMENT 14

It could take a lot of effort and expense to get such knowledge-work interoperability. You might say, "Why don't I just do the part that's important?", as in Figure 13, Choice A. Someone else's idea of what's important to share and exchange may look like Choice B: 14a

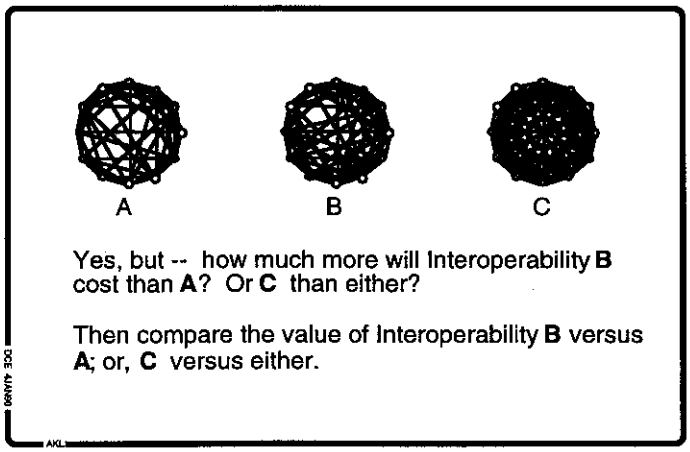

Figure 13. Providing for extensive interoperability will be expensive. 14a1

As more and more of the knowledge work in each domain is done online, then the benefits of a comprehensive degree of CSCW interoperability will rapidly increase. 14b

How do you decide how far to go? You'd compare the value of A vs. B, or B vs. C. And you'd say, "Well, let's see, with each successive choice I'd save more money, wouldn't I?" So how much more? We don't know how to quantify it yet. But, once you start finding a way to make some of the major sub-domains interoperable, by the time you've picked these selective lines in Choice A or B, what would be the incremental cost in dollars and effort to get Choice C? 14c

But the real question is, what does it cost in dollars and effort NOT to have the interoperability. 14d

THE OHS MOVEMENT 15

I asked people familiar with complex aerospace projects, "Well all right, let's make a guess -- if the kind of hyperdocument interoperability we are talking about here were installed for instance under the whole design & manufacturing operation of this ATF program, what might the yearly dollar benefit be?" They look back and forth at each other for a while ... So I offer: "$300,000,000 a year?" And they look at it and say, "At least." 15a

User organizations must realize that they can't just sit back and wait for the standards groups and computer vendors to deliver this, because there hasn't yet been enough orientation or application experience in this area. It seems necessary for the larger user organizations to take responsibility, to become pro-active -- e.g., with exploratory pilots, active development of associated knowledge processes, and cooperative requirements definition -- and then show the vendors that there is a sizable market for this. 15b

But they must also realize that it isn't just a matter of specifying, procuring, and installing the resulting system -- they have to learn how to employ it effectively in this extremely complex environment. And they must realize that they have to cooperate more intensively than before: The stakes are extremely large; there is too much to learn and events are moving too rapidly; the resources and degree of stakeholder coordination involved are both very high. 15c

To find this effort emerging from within the aerospace industry seems likely enough to me: it is the most complex work environment I know of, and a most urgent candidate for harnessing the benefits of wide-area CSCW and effective knowledge-domain interoperability. But other large organizations also have pressing needs for exactly this same capability -- for example, car manufacturers, computer vendors, government agencies, consulting firms, universities, consortia, and standards groups. 15d

To me there is a real need for a cooperative movement -- among large organizations that are heavily dependent on group knowledge work -- to coordinate planning and operation of advanced, cost-effective pilot explorations in this area and to share the experiences and results. This relates to what I am currently doing at Stanford University with the Bootstrap Project: exploring with a number of larger organizations how a "cooperative, CSCW community" could be set up and run to provide both valuable pilot-application experience and substantive knowledge products. 15e

One of the first projects of this community would be to collaborate on the requirements for an open hyperdocument system, and on a procurement approach. The community would employ a prototype OHS platform (initially AUGMENT from McDonnell Douglas) to collaborate on this and other related projects. This hands-on experience will be an important part of the exercise, and should provide valuable insight into how to employ these capabilities effectively. Similar pilot trials will be launched within member organizations. 15f

ACKNOWLEDGEMENTS: 16

This work was sponsored by grants from the Kapor Family Foundation, Sun Microsystems, Inc., and Apple Computer, Inc. 16a

For more background on the source experience from which these proposed OHS requirements grew: 17

Engelbart, Douglas C., "Toward High-Performance Knowledge Workers," OAC '82 Digest, *Procedings of the AFIPS Office Automation Conference*, San Francisco, April 5-7, 1982, p. 279-290 (AUGMENT,81010,). Re-published in "Computer Supported Cooperative Work: A Book of Readings," ed. Irene Greif, Morgan Kaufmann Publishers, Inc., San Mateo, CA, 1988, pp 67-78. 17a

Engelbart, Douglas C., "Authorship Provisions in AUGMENT," COMPCON '84 Digest, *Procedings of the 1984 COMPCON Conference*, San Francisco, CA, Feb 27-Mar 1, pp. 465-472 (OAD,2250,). Re-published in "Computer Supported Cooperative Work: A Book of Readings," ed. Irene Greif, Morgan Kaufmann Publishers, Inc., San Mateo, CA, 1988, pp 67-78. 17b

Engelbart, Douglas C., and Lehtman, Harvey G., "Working Together," BYTE Magazine, December 1988, pp. 245-252. 17c