Data Geographics:
Traversing New Data Constructs

Stephen E. Arnold
Vice President
Information Access Co.
362 Lakeside Drive
Foster City, California 94404

Abstract: Complex reference questions cannot easily be answer by consulting a single source. Numeric, textual, and directory data must contribute to the "answer." Timesharing companies and database producers have made strides in bringing data from different sources together. Dialog, Mead Data, and Information Access Co. (IAC) have introduced products allowing retrieval in richer, more useful forms. Company Intelligence, a database from IAC, provides directory data and references to current articles about companies. "Hybrid" electronic information products--that is, products blending different data types--mark a watershed because artificial barriers between separate databases disappear and searching becomes easier. Users will demand databases which are intuitive. The longer-term impact of hybrid files is: Create a need for a real-time data presentation in hybrid files; for example, dynamic graphs.

Voyager Corporation (Santa Clara, California) has developed a remarkable optical-disk product for Macintosh computers. For less than $200 a person can buy a performance of Beethoven's Ninth Symphony.

There are several important differences between this disc and the run-of-the-mill audio CD in Tower Records. First, the Voyager product developers include the score of the symphony. A person interacting with this product can select a particular point in the score and hear the corresponding music. A music student, for instance, can work through a score note by note, instrument by instrument. A tone-deaf listener can let the music flow by and doodle with the images and text stored on the disc. In short, Voyager has combined visual and audio data on one 12-inch optical disc, and they use hypertext indexing to link the elements.' Voyager has a catalog filled with nifty multi-media products ranging from a tour of Vincent Van Gogh's paintings to learning discs for young children.

Voyager has never participated in the traditional information industry trade shows. They are content to market their products to the Macfanatics. These are the loyal thousands who take such audio and visual delights for granted. Online and bibliographic CD-ROM products probably seem a bit dull and rather trivial to this band of California innovators.

This organization is typical of a class of innovative companies which poses a threat to the traditional information establishment. Consider this:

- These fellows combine pictures, words, sounds in new and exciting ways
- Locating a particular bit of information depends upon where the listener is in the data and what ideas the information triggers in this Macintosh user
- A mouse is used to navigate from point to point
- Indexing is largely invisible, buried to unseen links which tie the various elements of data together in a way appropriate to the user.

This strikes me as a peculiar three-, maybe four-, dimensional mental task. The person using the Beethoven disc, for example, must keep a mental map of where he is in the experience. The process of mapping and using such a multi-media disc is quite different from the 20-year-old query-response system found on most online and CD-ROM products today.

Voyager wants the user to explore the data. One moves about in a virtual reality. The user needs a different set of mental skills to comprehend words, still and motion pictures, and sound. In a sense, the user of the Voyager products, like the person who plays a computer game in an arcade, explores new territory. The process of building this mental map and navigating through a multi-media data construct is difficult to explain. I would like to suggest the term "data geographics" as a way to describe these new, dynamic, interactive, and multi-media information products.

When databases combine words, pictures, and text in one information product, the user experiences the data construct differently. Information is not so much retrieved as explored. File structure is less important that data relationships. Indexing is not immediately clear, but the links among the various types of data make the user's experience more immediate, somehow

closer to real life. Hybrid databases open new opportunities for database producers and enable a qualitatively different experience for users.

Data Hybrids and Big Gorillas

Remember this quip: "Where does a 200-pound gorilla sit? Anywhere she wants." Optical technology updates the joke: "What can optical discs store? Anything they want." Optical disk may become the universal information storage medium. There is very little by way of information formats--data, text, voice, music, or graphics--that optical disk cannot accept and store in a higher information packing density than any other medium or format.[2]

Information products which contain a variety of different types of media promise to be one of the most exciting areas of innovation in the next three to five years. Several factors are driving this trend:

- The first generation of computer-based information has reached the end of its development cycle. Putting words on a screen has gone about as far as it can. The exploding graphic user interface business and the introduction of products which make text look like typesetting mark the last gasp of character-based display technology.
- Underground online services have introduced still and animated images. These range from cartoons to pornography. Bulletin board operators are devoting more and more disc space to images. The hackers and curious like computer-based pictures. The commercial timesharing services--for example, Vu/Text--have no image capability although the full-text of the newspapers they offer include more and more images with the stories.
- Color displays are the de facto standard for most personal computer owners and users. The high-resolution display is now so commonplace that it takes a 19-inch super-monitor to draw an admiring remark. Even in desktop publishing, gray has yielded to hundreds of simultaneous colors.
- Software allows the developer to combine in a single storage medium a still pictures, line art, full-motion video, sound, text, and facsimiles. The term "hypertext" is loosely used as a way to talk about information retrieval which has more flexibility. Computer magazines carry advertisements for database management systems which allow the developer to build files contain mixed media. PictureWare (Bala Cynwyd, Pennsylvania) provides a comprehensive database, indexing, search and retrieval, and image package.
- Hardware is cheap and powerful enough to make graphics applications operate rapidly.

These explanations, however, dance around the real reason why multi-media information products are so interesting. It all boils down to one point: people think in multi-media. For this reason, the hybrid or blended information file promises to be a product category with considerable commercial promise.

The trigger for multi-media information products is optical technology. The rapid advances in all aspects of high-density optical storage drive image and multi-media products from the fertile minds of developers to the shelves of the local store.

Optical technology allows large amounts of data to be placed in one compact, easily transportable, economical to duplicate, and difficult to damage form. Once developers have room to put images (big memory hogs), sound (modest memory hogs), motion video (mega-memory hogs), and text (memory piglet) in one place, information can be presented in a way inherently more closely matched to the way one gets information in real life.

Text-based CD-ROMs are useful stepping-stone products. They are likely to have value as archival and reference tools. The sound-and-image technologies will storm the marketplace. Let's face it: reading a screen of text is less interesting than warring with the Teenage Mutant Ninja Turtles. Life is a game, and the hybrid information products comes closest to allowing a slice of life to be experienced by the product's user.

Characteristics of the Hybrid Information Product

Let me highlight the attributes of these new multi-media or hyper-information products. When one first meets one of these products, users must build a different mental image is needed of the data. I have never been able to articulate how I see a specific database. Even though I have been involved in building databases for some years, I perceive a database like Compendex Plus as a construct consisting of a number of discrete records.

I can tell you that each record has a particular structure. I can visualize a typical citation for an article in *IEEE Spectrum,* and I can name some of the principal journals which contribute records to the database. My mind then jumps to a sense of the types of questions that Compendex Plus can help me answer. I know that I will be able to get current information about particular engineering disciplines, specific data about testing parameters, and snapshots of products or technologies which will be influencing the world around me.

Whenever I describe a database, then, I fall back upon technical details like fields, tags, and journals lists. If I want to explain the use of the database, I have to provide several applications of the data; namely, brief anecdotes which give a problem and the information that helped me solve the problem.

The multi-media or hybrid information product becomes less difficult to explain. One can sum up the experience by letting the person experience the information: listening, looking at images, reading snippets of text in any order or format comfortable to the user. The data are immediately accessible. Pointing devices, icons, and pictures pull the user into and around in the data. A range of senses is put in play. This makes sense because some people learn by listen-'ing, others by reading, still others by association with kinesthetic actions like moving the mouse to and fro.

The essential differences can be summarized in this way: Hybrid products:

> . Reach out; they are proactive in their presentation of information

- Are interactive, not query based.
- Require different types of mental imaging to explore.
- Have a shape and form which demands that the user get into the information, not retrieve it.

Hybrid products are relatively new. As more appear on the market and users get increasingly familiar with them, this list of attributes can be expanded. Like so many innovations in the information industry, a new paradigm for thinking about these products and a new vocabulary will be needed.

Trends to Watch in the 90s

There are several important trends starting to take shape in the information universe. First, hybrid files--that is, databases which contain information of different types--are beginning to appear with increasing frequency. The grandfather of these databases is Trademark Scan. This database provides the user with an image of a trade and textual data. One can think of this database as the model for many of the new products which are finding their way to market. The Health Reference Center from Information Access Co., for example, follows this pattern.

This product provides an interface which allows the user to retrieve information by placing the cursor on an area of the screen and pressing the enter key. Available on the HRC are medical reference materials, specially-written articles describing common diseases and their symptoms, indices, abstracts, and full text. The product will soon support images..

Another example is the San Jose, California, Multiple Listing Service's image database of Silicon Valley properties. Realtors provide images of properties which are scanned at the MLS office. A database stores the particulars of the property and the image. When an affiliated realtor wants to locate properties with specific characteristics, the realtor queries the database. Images of houses matching the criteria specified can be faxed to the real estate office or to the home of the client.

Predicast's 1990 Promt database contains abstract, tabular, and full-text records. Such a blend was relatively rare only two years ago. Such mixtures of data are becoming increasingly common. The old barriers are beginning to drop as database producers realize that users have different expectations for electronic information. There may be some initial confusion about the new look on the old database, but this will decrease as the customers gain more experience with this new type of file.

One major trend, therefore, is that traditional information products will be enhanced with the addition of less homogeneous data types. A user will no longer be able to predict what type of electronic information a search will generate. The rule of similarity says that each record in a database must be similar in form, structure, and indexing no longer holds. The predictability of ABI/INFORM, for example, is a liability, not an asset in today's market. Databases will include information regardless of type as long as it is germane to the subjects included in the database.

A second trend will be the creation of information products which allow simulations to take place in real time. The first products of this type will be numeric and allow analysts to examine financial trends by gathering real-time data and analyzing that information against the archival data stored on an optical disc. Microsoft's Excel spreadsheet links live data to graphics now. When an optical disc of historical data are available, the live data flow can be mapped against the historical data flow. In short the numeric analysis will be dynamic and much richer than at present.

A third trend will be the gradual replacement of text and primitive graphic interfaces with paradigms defined by the user. The current crop of Motorola-based arcade games take a step in this direction. The graphics and the type of interface project a particular situation; for example, a motorcycle or a racing car. Hybrid databases will provide the user with an interface or interfaces appropriate to the information being explored. Will the Beethoven CD-ROM offer the user a piano keyboard paradigm in five years? Probably.

A fourth trend will be richer, more mature image enhancements. The character-based screens will be lifeless and unimaginative when compared with the next generation optical products. The person wanting information about modem sales will be able to retrieve and capture on floppies segments from "Computer Chronicles" television show, the full text of an interview in the San Jose Mercury News with high-resolution photographs and charts from the original story, and full-text articles displayed on the screen in typeset form. Everything displayed on the computer monitor will be a graphic image, not the text we see today.

A fifth trend will be different companies offering optical products. Few analysts have considered the impact of Sony's aggressive acquisition of television, motion picture, and audio recording properties. To my knowledge, no one has seen a connection between Sony's portable CD-ROM player and the new directions in electronic information products. Is it possible that Sony sees an opportunity to develop or supply data to the new class of products?

A sixth trend will be the use of read-once and read-write optical technology to make image capture a basic, essential part of document management in companies of all sizes. The paper less office does not exist because source documents have to be available to employees. In 1988 only 1% of the information of large corporations has been moved from paper to computers.[3] Scanning source documents and storing them on optical discs saves space at a company and promises easier access to employees throughout the company, not just in the department when the documents are filed. A significant opportunity exists to help companies build upon their source document base by including multi-media data in the process. What multi-media data does an average company have? Here are some representative organizational applications:

- Training information
- Scientific and technical processes which are difficult to document in words
- The organization's photograph archive
- Electronic files created for specific projects
- Interviews with key employees

Obviously several of these media can be mixed and matched. The optical medium gives the managers an opportunity to innovate and experiment.

Outlook

One class of information products will include a range of media. At present, optical products are relatively lifeless and unimaginative. This will change dramatically. The first impact will be in the application of optical technology to games. Shortly thereafter simulation products will find their way to market. These will allows users to explore data which replicates a real-life situation. Once these products have begun their life cycle, increasingly sophisticated blends of various media will flow from database producers. These products will with varying success replicate a real life information construct.

In all of these products, the customer will learn a significantly different way to explore the data. The new products will invite the user into the information. Today's products and users are at arm's length. Building greater intimacy has a price, however. Hybrid information products are more costly because indexing links must be designed into the product, not added as an after thought. Database design will no longer be from the garbage truck school of file building. Data can't be collected and dumped on the customer. Instead, database producers will have to be more disciplined and creative in selecting, presenting, and structuring the new hybrid products. In short, we have a revolution in information products beginning to take place.

Endnotes

[1] Franklin, Carl, "Hypertext Defined and Applied," *Online,* May 1989, Vol 13, No. 3, 37-49.

[2] Frankel, Barry, "Hypertext: A Software Solution," *Inform,* March 1988, Vo12, No. 3, 14-18.

[3] Pemberton, J. Michael, "Optical Disc--Super Media, Super Systems," *ARMA* Records *Management Quarterly*, April 1989, Vo123, No. 2, 64-67.