N-STAR: An Automated Analyst Tool for Open Source Data

Offered to the 1992 Symposium:
"National Security and National Competitiveness: Open Source Solutions"
1-3 December 1992

by
Robert K. Driver
Foreign Systems Research Center of Science Applications International Corporation

This paper describes an advanced work station application of computer software for use by intelligence analysts in effectively utilizing open source data The system, called **N-STAR**, uses a state-of-the-art Macintosh application (*SuperCard*) to create, store, synthesize, and display intelligence information in a manipulable and easily displayed format for rapid retrieval of critical information The data system can reside on a stand-alone workstation (including portable), or on workstations which are networked internally or globally.

Management and analysis of open source intelligence data has some inherent difficulties: information arrives in many different formats, its veracity is often questionable and the data must be corroborated; corroboration requires simultaneous access to other types of sources on same topic; varying levels of detail are available; there is a large volume of information to be sorted (which is aggravated by reduced manpower); data typically collected is fragmented and needs to be synthesized with other data; the information changes in incremental/additive ways, it is often analyzed from "corporate knowledge" instead of rigorous methodology; there are inherent language barriers in the indigenous source material, the diverse input data requires many different presentation formats; and there is often an urgent requirement to turn around intelligence analysis to users. N-STAR is a system which, by the nature of its structure and operating methodology, allows the user to solve many of these problems.

Background

Much of the Foreign Systems Research Center's Russian/Soviet knowledge has been captured in the RED STAR interactive database. RED STAR is designed to make knowledge on the former Soviet Union, its military and political decisionmaking procedures, and other relevant information readily available to users. It is analogous to an automated encyclopedia with all relevant entries linked together to assist in cross-referencing and data management. RED STAR takes advantage of an advanced database software package (*HyperCard*) to handle the large amount of information and the free-form data linkages

In order to produce similar data systems and analyst tools for study of non-Soviet countries, the RED STAR approach was expanded. DESERT STAR was produced by the FSRC as a modification of RED STAR for Iraq, using a more advanced application (*SuperCard*). DESERT STAR has similar characteristics to RED STAR, but the improved software application creates more flexibility in information display, manipulation, and updating. N-STAR is be the latest evolution of this concept, taking advantage of the software application and lessons learned from the prototype, DESERT STAR. N-STAR is a generic system which can be made country-specific by data entries and linkages.

System

N-STAR can best be thought of as an automated handbook of information on a broad topic. The breadth of the topic is important to the design of N-STAR because it implies that many different related topics, each describing a different dimension of the overall topic, are required. Thus, a military decisionmaking environment as it might be represented in N-STAR embodies many different aspects of military technology, unit deployments, geography, military organizations, international relations, politics, religion, and so forth. Each such category of information may have associated with it a different type of data to display. For example, military units may be depicted on a map and also have textual information related to the unit's status. Geographic reference material requires maps as well as supporting textual data. Military technology may be represented in tabular form to show the characteristics of weapon systems Organizational data may include textual descriptions of the command elements and perhaps graphical displays of how they are organized. Similarly, intelligence analysis from open sources is best effected by many diverse kinds of data (e g , maps, equipment specifications, unit structures, locations, personnel information) being available simultaneously. N-STAR is configured to make data in many different formats readily available in one system for analyst use

Perhaps the best analogy that can be used to describe the design and organization of N-STAR is that of the traditional "shoe-box" data system. In the case of N-STAR, the entire system has been designed to hold and organize data as if the individual data items were on index cards stored in a large electronic shoe box. However, instead of having to duplicate entries in order to cross-reference them, N-STAR allows a single electronic "index card" to be linked to other cards that are related to it. The N-STAR system configuration allows more efficient gathering of information, which is especially important in a high-volume data environment such as open source.

The N-STAR system has been designed principally to be an on-line, interactive system. Its data display and retrieval methods are optimized for screen use, so an analyst can quickly find the specific information required without searching through file folders, weapons characteristics manuals, or maps with acetate overlays. Controls are provided

on screen to allow the analyst to browse the data and find things of interest, read them on the screen, and then move on to the next item. Associated subjects will be evident on each card, allowing the analyst to jump to those cards as necessary. Mechanisms are built into the N-STAR system to make this browsing function efficient

A key distinction of the N-STAR approach to data management compared to traditional DBMS is that it is not intended to allow you to recover dozens or hundred of records of data which satisfy a set of search criteria. Traditional DBMS deal with large numbers of similar records, each containing the same set of information about a transaction or object. Thus, a warehouse inventory may be represented by a data base of thousands of records, each with an item number, a price, a description, number in stock, etc. The user would search either for selected items or for whole sets of items at a time.

Navigation, as this process is called, is a critical N-STAR capability. With information divided into "bite size" pieces for easy access, it becomes more important to be able to move quickly to the next related data item if the analyst feels the need to explore a topic further. The system provides several alternative pathways for calling up the next desired item. The key to these pathways is the notion of associativity

Associativity essentially describes how individual data items are related to one another. In a traditional, hardbound handbook, the reader will follow the linear flow of text as he reads, use the table of contents to find a desired section, or use an index to find a specific topic of discussion The first is inefficient at best, especially when trying to assimilate large volumes of data The other two are acceptable, but the user will often be leaving the item of current interest to find where the next item is located

N-STAR changes this approach by making more options available from the current data item The equivalent functionality of continuous reading and index search remains, but there are also other mechanisms that permit movement directly to topics which may be related to the data item currently in view. The overall effect is to allow the analyst to browse through a mass of data very easily, and yet find and assemble the information that is important to the analytical problem.

The N-STAR system is designed to grow as it is used. The data should not be considered static. As more pieces of information become available, the analyst can enter them into the system, adding new cards and linking them to other cards as desired. As data is corroborated or refuted, existing cards can be modified. Additional linkages between existing cards can also be added at the discretion of the user in real time. In this way, the system stays robust even in a rapidly changing information environment

Data

N-STAR is designed to provide a broad spectrum of data about a single, underlying topic. In the version represented by the figures in this paper, that topic is the country of Syria. Within that topic, there will be many subordinate topics describing the government of Syria, its ruling political parties, key individuals, its religious institutions, its military organizations and weapons, and other such subjects. Each of these subjects is called a "category," and will be displayed in its own screen window within the N*Syria system. Figure 1 shows the set of categories currently implemented in N-STAR. The analyst can display several category windows at the same time, and move between them with the click of a mouse or through menu and dialog controls.

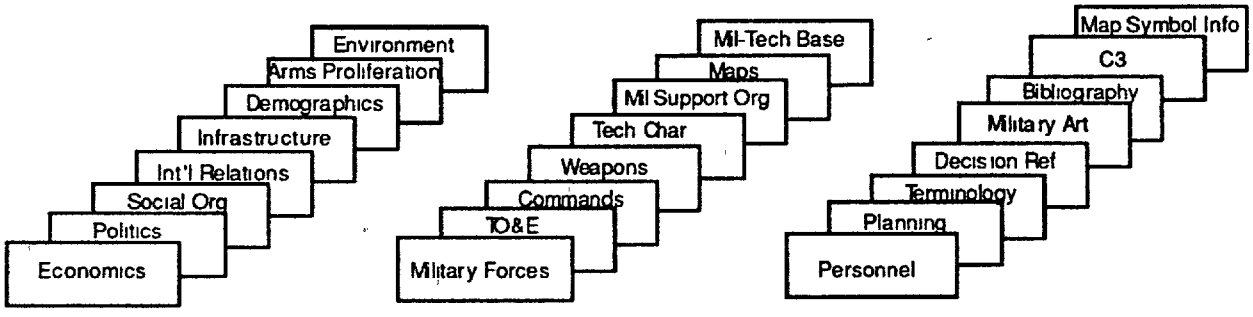

Figure 1

This set of categories provides the only fixed structure to the entire N-STAR system, and even that is not completely rigid. New categories can be added easily through separate configuration controls by an N-STAR system administrator

Data Cards

The basic unit of data storage in N-STAR is the "data card", or just "card." It is analogous to the 3x5 index card in a card catalog, and will contain one of several types of data, as explained below. Each card contains concise data on a given topic in a single format. This compactness lets the user assimilate data quickly, and yet be linked to other data items as required. Each card can be marked with its own classification level and password or other user ID control to maintain multi-level data without risk of exposure Navigation through the system of data follows the free-form nature of the data, allowing the user to browse the data system. Cards from several different categories of data can be visible on the screen simultaneously

Each data card is labeled according to its subject matter; a sample is shown at Figure 2. A unique subject title is placed in the top-most field on the card. This title is used in the card index for each window. The majority of the data card is occupied by the main data area. This will vary in content, depending on the type of card being used On

text cards, this is a scrolling text field. On graphics cards, this is a blank region in which the analyst can draw any figure up to 5 by 7 inches in size. A classification marking can be assigned to each card, making the system adaptable to a security controlled environment. Selection of the classification is done through a pop-up menu. Each category has its own window for display of its data cards. The title bar of that window will display the name of the category, so the user always knows in which section of the data system they are operating.

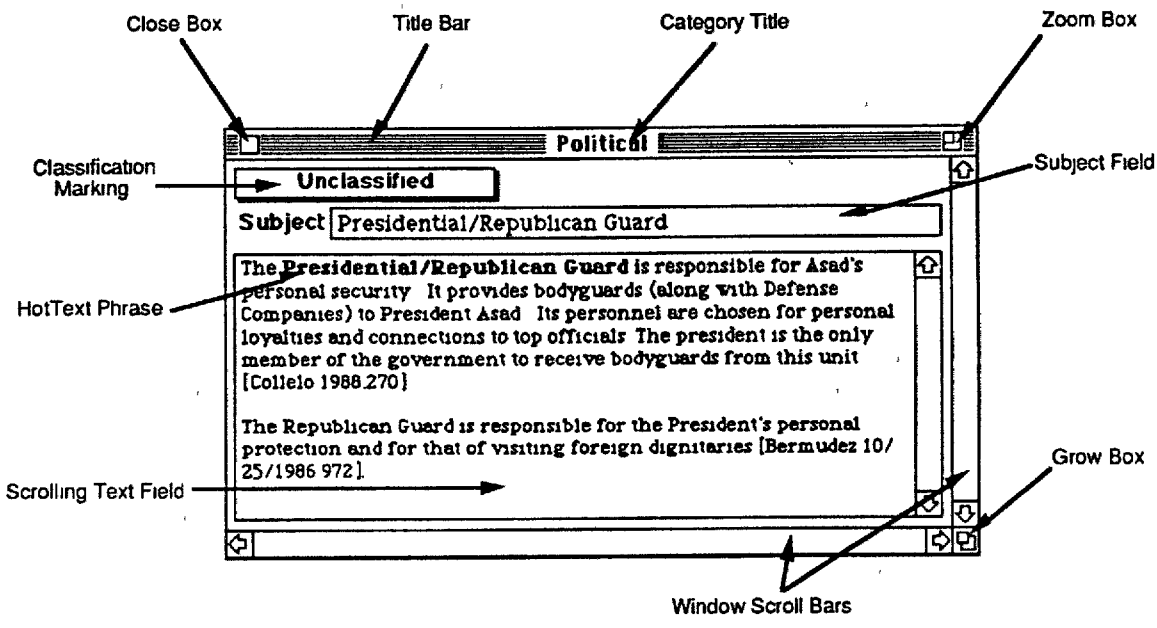

Figure 2

Data Formats

The data cards making up the N-STAR system can contain a variety of different data formats, making the system extremely flexible. For the most part, there are two basic types of data cards: text and graphics. These and other formats are described below:

Each text card (see Figure 3) contains a scrolling text field that can hold up to 32,000 characters of text. This enables a single card to hold a large amount of data on a topic. Any text on this type of card can be made into HotText, which provides a convenient mechanism for connecting two data cards that are related to one another. A user can go to the appropriate related data simply by clicking the mouse on any text designated by bolding as HotText

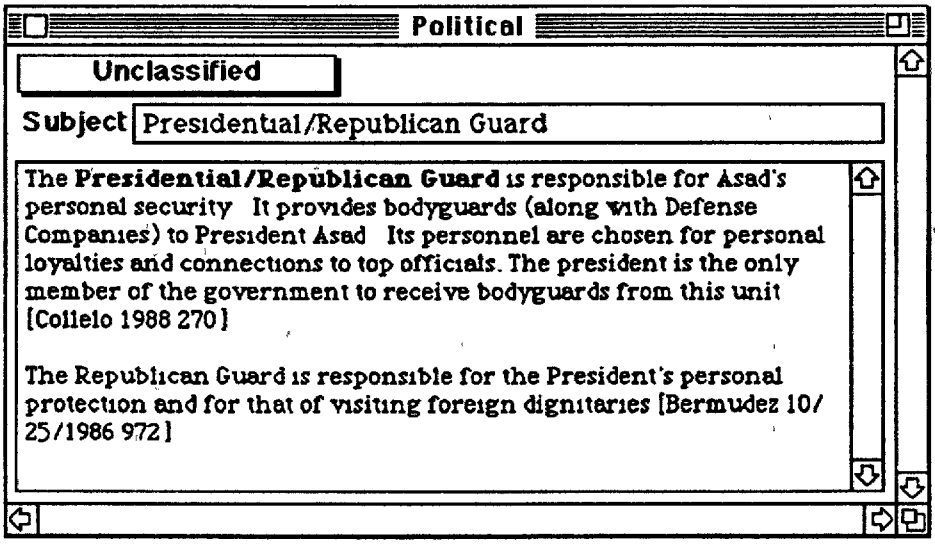

Figure 3

Graphic data cards, such as Figure 4, provide a visual display of complex topics. The graphic tools available within the N-STAR system enable a user to draw simple or complex pictures of organizational charts, warfighting concepts, and so forth. The tools include graphic objects for rectangles, circles, lines, text, polygons, and other object forms. Graphic objects can be linked to associated data in the same way that text can through HotText links These are called HotPix links

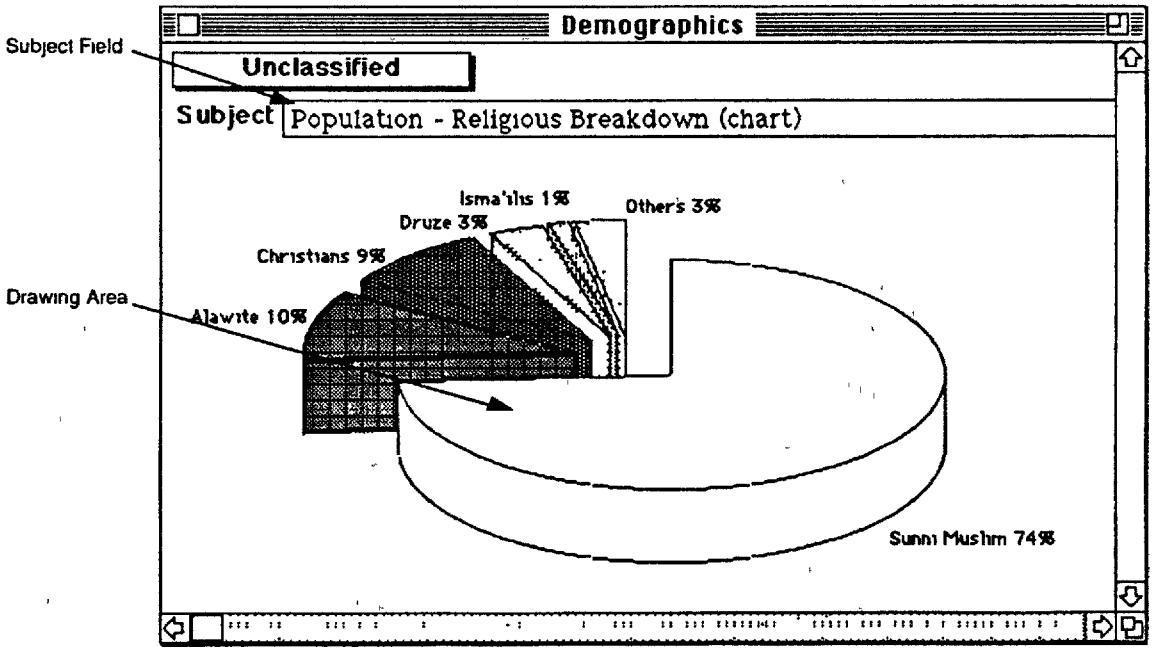

Figure 4

A type of graphic data card is the imagery card. It uses the graphic card format, but the graphic is not a picture drawn with tools. Rather, it is imported from other programs and pasted onto the data card (see Figure 5). Examples of imagery items might be: a photograph of weapon; satellite imagery; or the bomb damage assessment photos after a mission. Once placed on the data card, it is possible to add other graphic objects to annotate a picture.

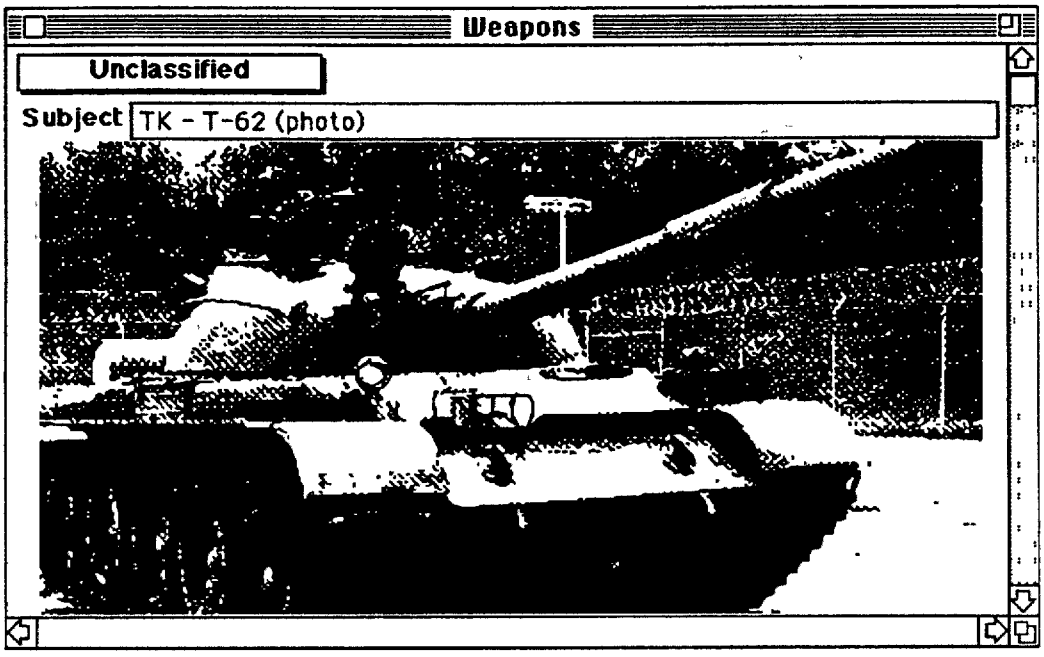

Figure 5

A very important format of data in N-STAR is map data, which is used to display the nation or area covered by the N-STAR system, its geography, and its military and industrial facilities. A sample of a map card is shown in Figure 6, depicting a portion of the Syrian-Lebanese border. Each base map can accommodate hundreds of individual overlays representing such map information as Military Units, Air Defense Sites, Airfields, Bases, Railways, Roadways, Rivers, Communications Links, and anything that could be conceivably drawn on a map. The analyst can quickly create graphic decision aids by judiciously combining overlays on the display. Each symbol or icon on each overlay can be linked to other data cards anywhere within the system, which could describe the facility or entity represented on the map in more detail. Additionally, each map symbol representing a dynamic object such as troops or units can be moved on the map in real time to correspond with actual relocation or detection. The data cards in the system linked to that symbol remain linked as the symbol is moved to a new spot on the map. Clicking on the icon for a facility will cause a data card of that installation to be displayed.

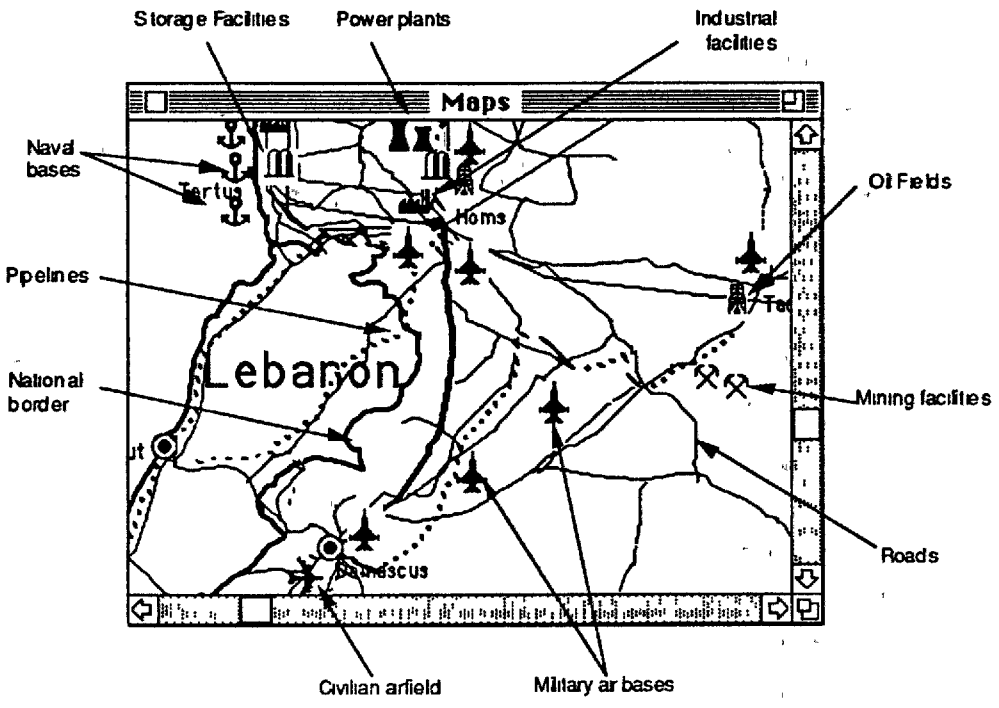

Figure 6

Another data type which is available in N-STAR that is not normally found in data systems is the use of actual mathematical models. It is possible to build in a function on an individual data card which performs desired computations, such as loss rates, speeds of advance, or more complicated modeling problems. These provide the user with a readily accessible method for evaluating the impact of changes and "what if" types of questions

Special Card Formats

In certain categories, there is a need for specialized formats, other than text or graphics, that identify discrete types of data that should apply to the unique category of information or fit the type of data better.

Tabular data (Figure 7) is a special form of text data, in which the data card is organized like a multi-column table or matrix. Text and numeric values can be placed in the matrix, giving a better organized display than simple text alone. This allows an analyst to quickly get the information necessary without wading through large amounts of text. An example of a tabular data card might be a table of organization and equipment of a ground force unit. Descriptions of each item would be placed in the left hand column, and the number of items of that type would be entered in the right hand column

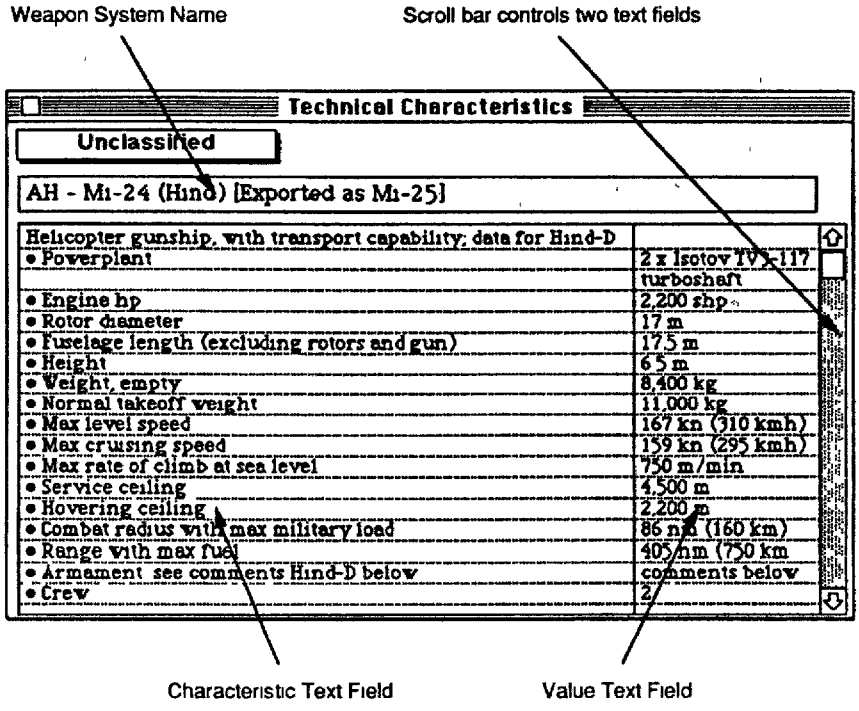

Figure 7

Biographic data cards, such as Figure 8, provide information about personal histories of individuals. This is used to summarize the background and positions of persons in government, the military, or other national institutions.

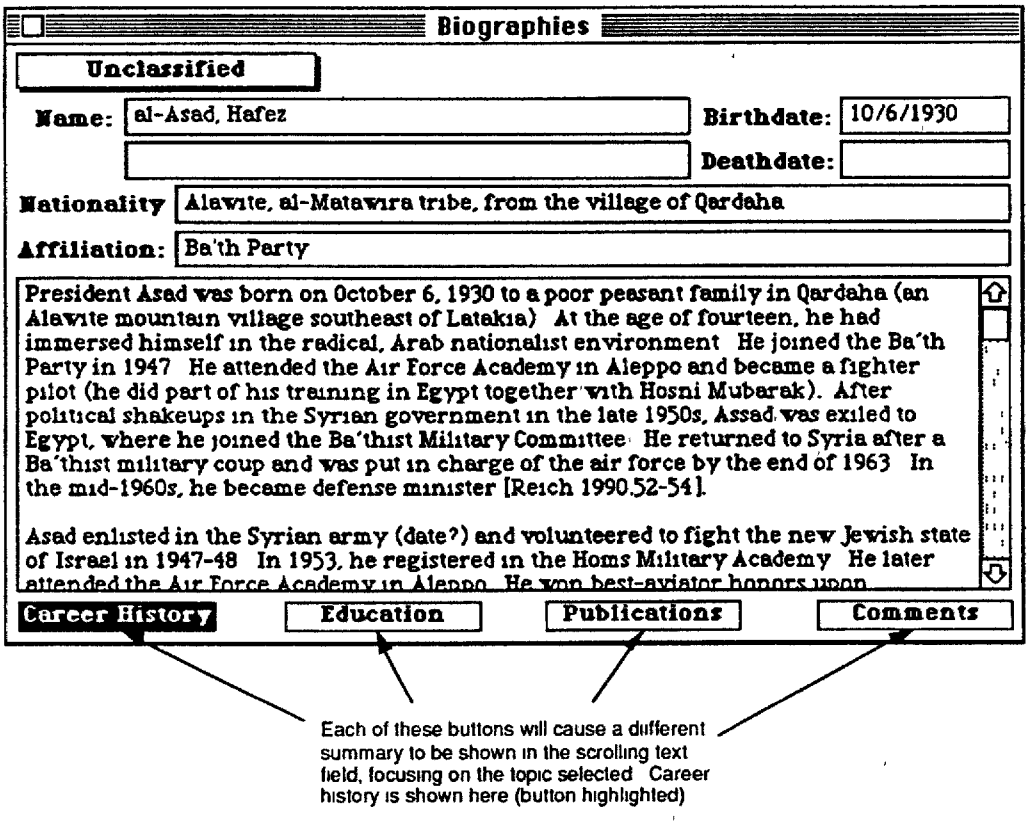

Figure 8

The Map Symbol card is another special format of data card, containing information about a facility or location that will be found on a map overlay. Map symbol cards provide a formatted means of making information available on any icon that is drawn on a map overlay Figure 9 shows an example of such a card. Categories of Map Symbol information cards (and Overlays) include

Cities
Industrial Centers
Major Roadways
Railways
Waterways and Coastal SLOCs
Power Stations
Natural Resources
Airbases
Naval Bases
Army Bases
Air Defense Sites
Radar Sites
Other Troop Concentrations
Communications Nodes
Communications Links
Major Storage Facilities

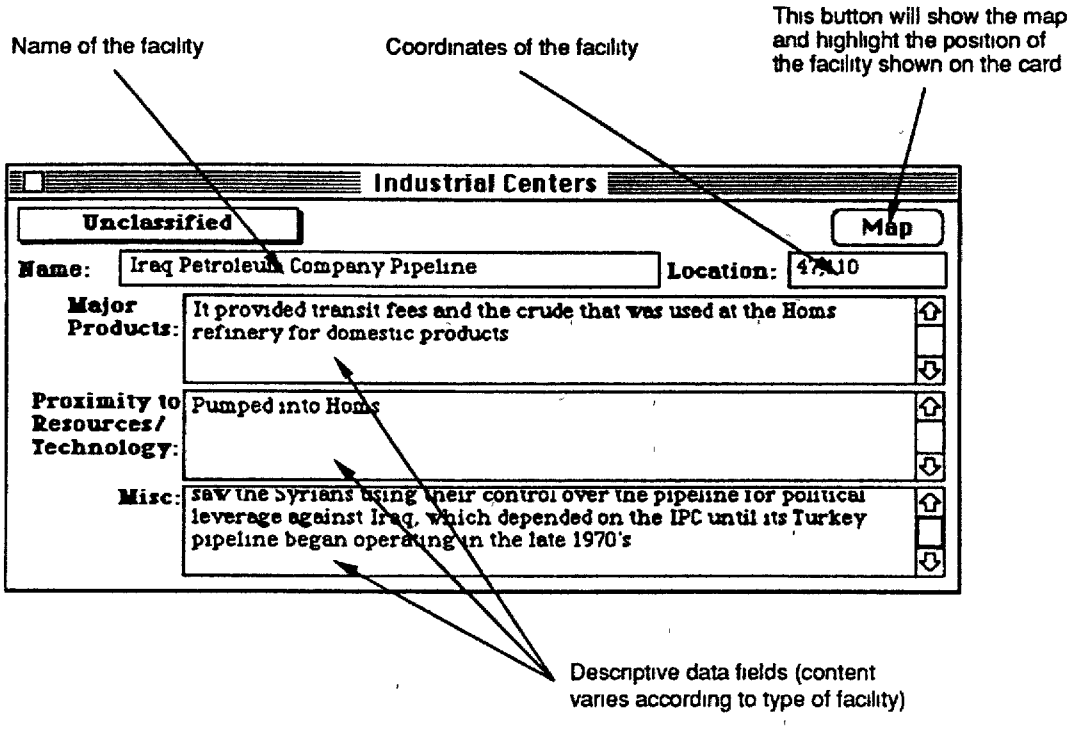

Figure 9

Data Relationships

A key feature of the N-STAR system is its ability to provide easy navigation through the many data entries, in a way that suits the needs of the individual user This feature is made possible by the lack of an overall rigid structure to the data. Traditional DBMS will impose a fixed set of record types and relations between records, forcing retrieval by the pre-defined set of fields in that schema. To move from one set of records, found by a given set of criteria, it is necessary to back out of the current match, and start the search process all over again. N-STAR takes the opposite approach, minimizing the organizational structure of the data, and maximizing the alternatives for finding the next item of interest (Figure 10). Finding data by search criteria is only one method possible in N-STAR. Other means of navigation include simple paging through the collection of cards, using indexes to find an entry by its subject, refining that search by keyword, HotText links between cards based on key phrases in the text, and several others.

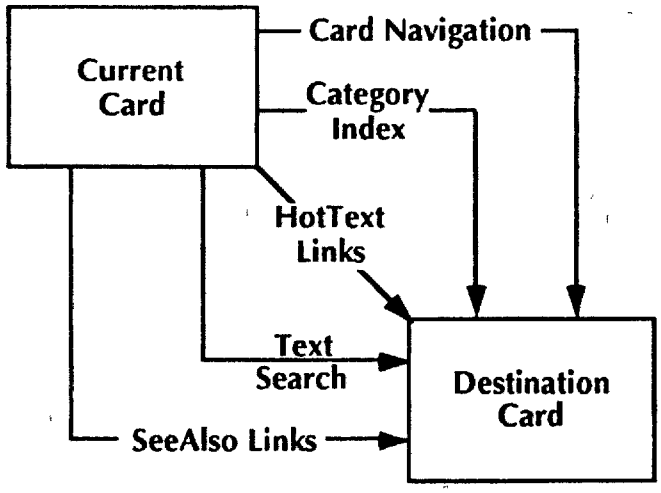

Figure 10

The N-STAR approach enables a user to move through the data system in a dynamic way, led by their requirements, interests, and curiosities, rather than by a fixed set of search commands It also means that no two users will employ N-STAR in precisely the same way. Figure 11 illustrates how two users, examining the same set of data cards, would do so in different orders, and possibly with different objectives. In both cases, the users start at the same data card. However, User #1 sees something on the first card that lead him, in succession, to cards 2, 3, and 5 Meanwhile, User #2 sees something entirely different on the first card He then follows a data path from cards 1 to 4 to 5 to 3. Thus, the structure of N-STAR is sufficiently flexible to define multiple pathways among the data cards

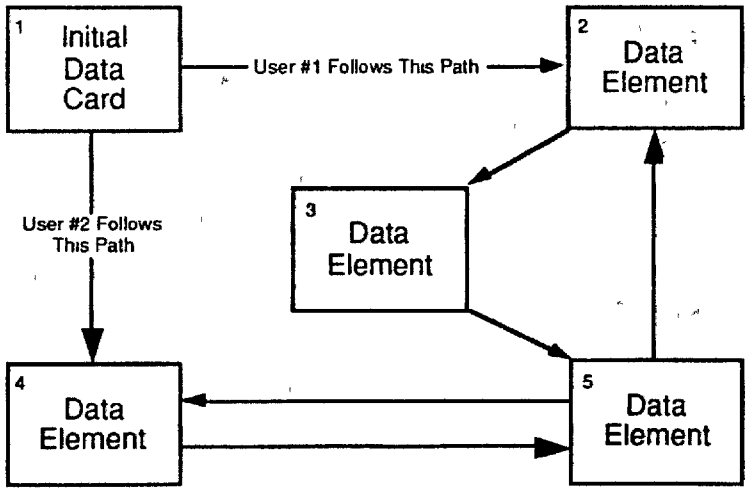

Figure 11

Keywords

While there are many different categories of data represented in the N-STAR system, these categories can only provide the broadest demarcation of the data base. Within a given category, there are data of many different varieties Looking at the subject index for a category window, therefore, may not provide enough discrimination of the variations in topics. The user can always read through the index list, but this may not be the most efficient way of finding the data items desired. Nor will the subject title of each card always capture the entire content of the data on that card

To alleviate this problem, N-STAR provides a means of describing the contents of a given card using one or more subject keywords. These can be used as a quick subject summary of the card data, much the way that a library card catalog has subject cards for its books. In this case, however, there is only the one card for the data, and the system will mark the subject keywords electronically for the analyst. Each category window is associated with its own list of keywords, so that different keywords can be used for each broad topical area

Multiple Distributed N* Workstations

It is also possible to operate the system across a wide area network, with updates to data items distributed automatically, even allowing for discontinuous communications between sites When linked together, multiple users can transmit changes to other workstations to ensure each analyst is working with current data Each user can change data on the spot as the information evolves There are a variety of system configurations available to handle data interchange, depending on the means of communication available. Figure 12 shows a conceptual diagram of the overall data system, including the likely computer network configuration at central and remote sites.

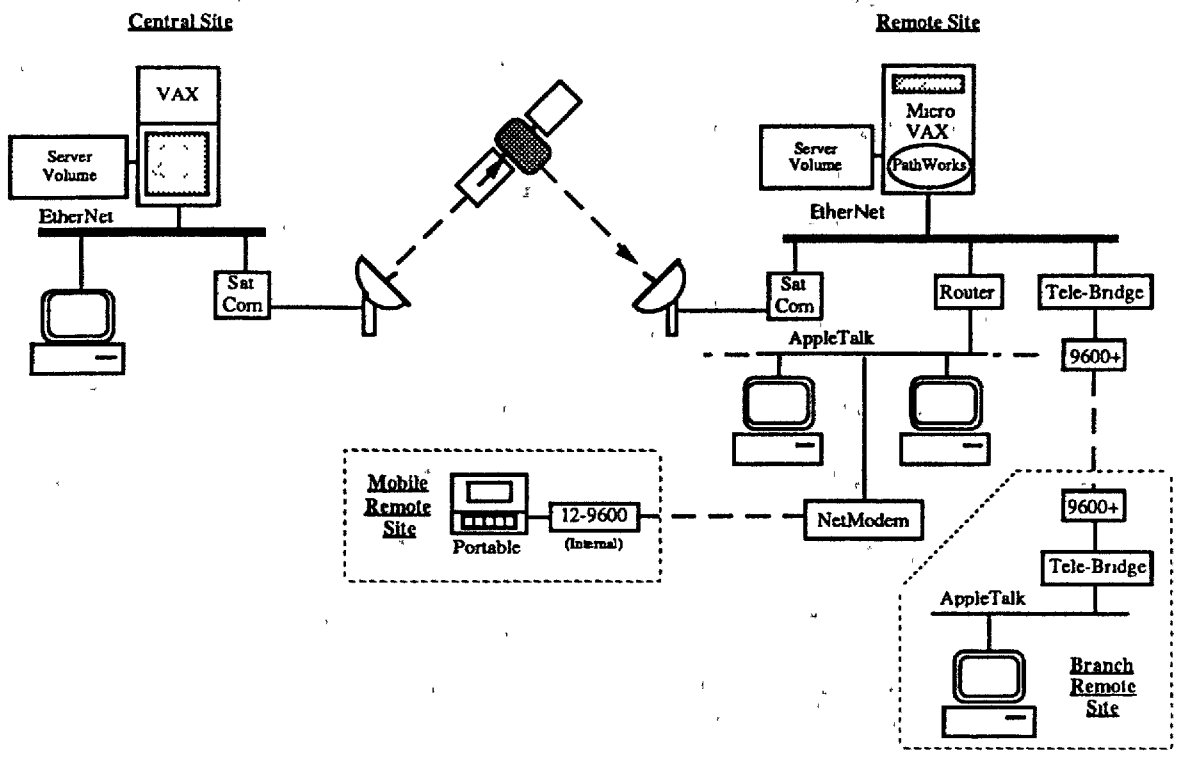

Figure 12

A configuration such as this example would allow distribution of secure data in many formats to widely dispersed parts of the world. A central location will serve as the focal point of the system, assembling data from centrally located institutions, processing it, and then sending that data to remote sites as they require In addition, the central location will accept new data from the remote sites, which will then be assimilated into the central data system. The central location must maintain a complete version of the data base. Portions of the data bases relevant to each separate remote location would be distributed to those sites All sites (central and remote) will operate in a server-client mode. Users at remote sites will normally be connected to their server at all times. Server computers, in turn, will normally be continuously connected to the central site using EtherNet protocols over high speed (56K, T1, or equivalent) data links, including satellite communications.

Summary

N-STAR is an interactive analysis system that provides a flexible and dynamic way of collecting, displaying, and assessing open source information Written in an advanced Hypertext application, it does not look like a traditional data base and provides exceptional versatility in information retrieval and display. Information is recorded in a variety of forms images, graphics, text, tables, etc , with an extensive system of cross-references. Since N-STAR is a free-

form data base (i.e., the form of the data is not specified by the data base structure), it can incorporate and relate diverse types of data in a single system. In addition, the entire system can be tailored to an individual mission by adding additional categories of information ("windows") or focusing on areas of the world in detail or in general The system can thereby be modified for specific uses as diverse as drug interdiction, arms proliferation monitoring, mission/target planning, or strategic assessment.

FIRST INTERNATIONAL SYMPOSIUM: NATIONAL SECURITY & NATIONAL COMPETITIVENESS: OPEN SOURCE SOLUTIONS Proceedings, Volume II - Link Page

Previous	THE FUTURE FEDERAL INFORMATION INFRASTRUCTURE:
Next	Communication to Mike Nelson, Estimated Costs of Networking Nation's Public Schools

Return to Electronic Index Page