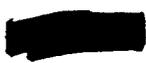


22 SEPT 93



Progress report on the creation of an integrated machine file at the Air Force Intelligence Center for all raw data of all types.

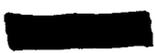
AN IDIOT SYSTEM FOR INTELLIGENCE
Harry Hopkinson

The Air Force Inspector General, in February of 1962, remarked that "a serious weakness in our system permits individuals, without proper training and experience, to exercise judgment beyond their capabilities." He was talking about the Air Force, not intelligence, but do we in intelligence occasionally see a glimmer of this truth breaking up through our own silver-lined cloud layer like a forming thunderhead? Armed forces world wide all have a bit of the idiot system—rules prescribed by geniuses for idiots to follow—built into their normal operating procedures, for good reason; and it may be that something not quite the same but analogous to it has a place in the collation of raw intelligence data.

Intelligence analysts are not likely to be idiots, but many are apt to be a little parochial, a little myopic beyond their particular specialties and resources. This is an occupational infirmity generally inflicted on them by a workload that forces them to confine most of their efforts to data files readily accessible from their desks, files limited by type or source or security classification of the data. Let us examine some case histories that show the infertility of such partial, separate, uncohabitant files.

The MK Numbers

Many years ago, for a number of apparent reasons, the Russians put the clamps on the individual enterprise then exercised in the procurement of foreign books, periodicals, and other publications and set up Mezhdunarodnaya Kniga (hereinafter referred to as MK because it is so hard to spell) to act as the official procurement agency for Soviet organizations. In a bureaucratically efficient manner, they devised a form and a numbering control system, assigning an individual number to each major organization in the country, which



then assigned a sub-number to its own organizational elements, who in turn assigned a third, sequential number to each order as it was prepared. MK also assigned a number to each foreign country and a second number to its procurement agency there. In the United States this agency is the Four Continent Book Store at 822 Broadway in New York City, with a miserable staff plagued on the one hand by internal bickerings and union troubles and badgered on the other by MK and the Soviet Embassy for doing a poor job. It is a registered Communist organization, headed at the time of interest to us by a chap named Allen Markoff, unhappily making a living the hard way.

Book orders, typed at the local level in the Soviet Union, wend their way up to MK, where they are dispatched in triplicate to the appropriate foreign collection agency. Those for the United States were all sent to Four Continent, who took action by sending two copies of the form (plus a couple of its own dealing with billing and payments) to the appropriate publisher or subscription agency, the Government Printing Office, or other source for procurement.

In 1957 an Air Force Intelligence Center analyst, looking for new sources, started collecting these forms through the GPO and the FBI. The data, including the identity of the Soviet organizations that originally placed the orders, were put on IBM punched cards. Many of these identities were somewhat obscured by the use of Post Box addresses without any organizational name, but their assigned organizational numbers required by bureaucracy were helpful not only to MK but also to our analyst. Periodic machine runs of various sorts were made to determine the primary interest, as expressed in the orders for source material, of the various organizations.

At about this time the Ministry of the Aviation Industry (MAP) underwent a reorganization, changing its name to the Committee of Aviation Technology and acquiring a new street address, but not changing its assigned MK book order number. If the latter had not been enough to identify the new organization as a replacement of MAP, an analysis of the typing by the FBI laboratory in Washington established that the same typewriter had been used to prepare some of the

book orders under the old name and some from the new outfit. Russian reorganizations are assumed to be similar to our own: no matter how the organizational boxes are shuffled, the same functions generally continue to be performed by the same people with the same old equipment.

The edge was taken off the sweet taste of accomplishment in this identification by a Soviet public announcement of the organizational change in question. Not announced, however, was the fact that after the change the Committee acquired a marked interest in missiles, as evidenced by a small flood of book orders for such items as "The Mechanical Aspects of the Vanguard Flight Control System," "Re-entry of Spherical Bodies into the Atmosphere at Very High Speeds," "Redstone Arsenal Ballistic Ramp," "Effect of Air Drag on Elliptic Satellite Orbits," and hundreds more in the same vein.

So far so good. But in 1958 this enterprising analyst was transferred, and the effort ceased. It was picked up later elsewhere as the maintenance of a distinct category of data, the CIA Book Order File; but who has used this or will use it as a separate file? The identification of those mysterious Post Box numbers can be accomplished only by systematically playing this unclassified, open-source material against other files of a completely different nature and classification which might appear to contain altogether unrelated data.

Coolers, Arsenal, and Elephants

Some years ago a trainload of air coolers was shipped to Ashkhabad on the Russian-Iranian border. An intriguing fact. A carload wouldn't have bothered; but a trainload? What could the Turkmen slope-heads do with this amount, since they do not afford themselves the luxury of individual air conditioning? So far as I know, no one ever found the answer; and the reason no one did is that there was no clue to tell which files of the multiplicity maintained in the intelligence community one might search next for any possible associated activity. And if there had been, no one would probably have had time to undertake it anyway. Nor would it generally be possible for a machine-type custodian of 25 or 35 varied automated files to suggest a particular two or three to be searched to shed light on this anomaly. As a matter

of fact, the file that produced the anomaly itself is one associated neither with Ashkhabad nor with air coolers, but one concerning a different subject entirely and an unlikely place to find such an item at such a time.

The inexplicable installation at Mozhaysk, a word to produce tarantism in some people even today, was an unusual example of complete frustration. We had excellent photo coverage of the thing—all that could be asked for, one would think—yet never in the history of intelligence has so much labor been expended on a single question as on the search of *all* files, manual and machine alike, in the endeavor to discover what this was. Every existing information file was milked not once but many times by different searchers; but one wonders whether, if a single homogeneous group had been able to conduct an integrated search of all of these files, scrutinizing each fact in relation to all other available data, it would have taken some three years to reach an agreed solution.

It would perhaps be an exaggeration to say that we handle raw intelligence information today in the same way it was handled by the Union Army during the Civil War, but in a sense this is not far from the truth. Indeed, we might go back even as far as Hannibal and reflect on our resemblance to the Roman Scipio, trying to establish the operating characteristics of the African elephant from scattered and unrelated reports. And speaking of elephants, Aesop's fable of the seven blind men is a classic example of compartmented intelligence—and recommended reading for all.

The Unifile

There is no disagreement that the increasing masses of data, the excessive demands on available manpower, the need for increased speed in reaction and analysis, and the exploding technology of weapons systems all add up to a pressing need for automated data handling. To stop short with the application of this forward step to separate files, to be operated or queried individually, would be a grave error. The habits, training, operating constraints, and environment of today make the establishment of an idiot-proof, single-query, single-search type of raw data file the minimum requirement

for automated data manipulation, and an even more critical necessity for the high-speed intelligence analysis and synthesis of the future.

In 1960, following years of contractual probings aimed at designing a complete intelligence data handling system, with no tangible results except a very liberal education for certain personnel of AFIC, an ambitious concept was born, and an important decision was made. The concept was an integrated, complete raw data system using in combination the IBM 7090, the AN/GSQ 11A (Minicard) document retrieval system, and the photo-handling equipment of the now defunct subsystem I of a proposed satellite program. The important decision was to perform the systems design in-house, utilizing contractor personnel only for detailed design and programming under the strict direction of AFIC personnel. After years of pie in the sky, always associated with a requirement for literally hundreds more people, this was an attempt to get on with the job while keeping at least one foot on the ground.

The modus operandi was the establishment of an ad hoc committee known as the Information Exploitation Group, composed of four full-time people selected because of their ingeniousness and backgrounds of experience in both intelligence analysis and electronic data processing. With the help of in-house systems designers, programmers, and indians and chiefs of assorted experience, they were ready at the end of ten months (under some pressure) to start building what they called the Duplex Unifile System. This system was to provide not only for processing new information on an all-source basis, but also for storing and handling information previously indexed and machined by other organizations, so that a sort of instant library in depth might be achieved rapidly. It had to have growth potential, be sufficiently flexible to shift with the advent of new equipment, and be redirectable to handle presently unanticipated problems.

A babel of machine codes has developed in the intelligence community, and not even transliteration of the Russian language is standard. It was necessary to design a computer record format of universal character, one that could accept any existing machine-readable file. Conversion programs would be prepared to put into this Unifomat all the various

acquired files and subsequent material up-dating them. The consolidation would create an electronic file unprecedented in scope and depth, an instrument for correlating all pertinent available data regardless of its class, source, or security classification.

The universals of Unifomat are the three basic elements which are repeated in all files regardless of format—*Identifiers* (people, places, and things), at a *Location*, in *Time*. With an added provision for the expression of interrelationships among these three, all information can be adequately represented by them, and the system is intrinsically suitable for machine manipulation, document retrieval, and photographic records. In connection with the latter, a Unifile P.I. Report was designed and is currently in operational use in the Air Force, serving as the photographic intelligence input to both the 7090 and the Minicard system.

During the development and conversion stage, duplicate files have been supplied to us by our very cooperative fellow-members of the intelligence community, including the photographic intelligence files in NPIC, several CIA registers, special intelligence files from NSA, some selected sensitive files from FTD, ACIC photo data files, and several 438L-developed files from SAC.¹ By the end of the year 25 files now scattered geographically, organizationally, and by security barriers, comprising two and a half to three million records, will have been merged into Unifile.

Computer programming for data correlation—the end objective of the development—has already been initiated. An analyst will be able to query all these manifold sources at a single stroke. But the computer will not wait idly on the analyst's initiative: under a "trigger" program being developed it will keep searching the Unifile to detect the convergence of indicators of target activities—long-range missile sites, nuclear weapon storage facilities, etc. When the sum of evidence of such an activity crosses a threshold value, output will automatically be triggered, giving a hypothesis of the

¹ The last three abbreviations stand for the Foreign Technology Division of the Air Force Systems Command (successor to Air Technical Intelligence Center), the Aeronautical Chart and Information Center (at St. Louis), and the Strategic Air Command.

Idiot System

existence of the activity and a statement of the evidence supporting the hypothesis. Stringent tests of the programming are planned, and before winter's end we will have completed our universal research instrument—or at least will be within revision distance of it.