

LEE WHITE'S HISTORY OF THE DIS (as told to R&RS staff, November 1989)

In 1975, DI was established in the Program and Policy Coordination Office. It was actually started before that in the DR Bureau for a couple of months. There was no computerized system at that point. The Development Information System was still being designed as part of a larger development effort called PBAR - planning, budgeting and accounting. PBAR consisted of five applications:

- DIS
- ESDB Economic and Social Databank
- CPDB a forerunner of what is today the Congressional budget system that PB maintains
- PAIS Project Accounting Information System which is still maintained by the Planning and Budgeting Office, and the spin-off PAISHISTORY file, and
- PPTS which was to try and establish project benchmarks for project implementation and performance measurements to see how far along the project cycle had actually gotten.

The target audience of DI was AID project designers and evaluation officers. The AID Reference Center was established with project files that were resident in the State Department Building on the 3rd floor. And, we were an OE-funded operation.

Also in 1975, there was a unit in the Technical Assistance Bureau (TAB), the predecessor of today's Science & Technology Bureau, called the Office of Evaluation Utilization (EU). They designed a project called Information as a Tool for Development (931-0232) which is today the project funding the R&RS and DIHF contracts. The target audience was principally LDCs, and it was program funded.

(Start standard topics 1-10: 1 System 2 Contractors 3 Source Materials 4 Technical Processing 5 Database Processing 6 HW/SW 7 System Access 8 Source Material Access 9 Information Dissemination 10 Information Services.)

In 1976, the DIS became operational. There were two contractors that were responsible for the processing of data: PRC, for about six months, and then LTS. They worked with AID project files in the AID Reference Center - feasibility studies, design reports, evaluation reports, audits, progress reports, end-of-tour reports. They used data coding forms to code bibliographic information, 16-cell logframes, 1000-character abstracts, related project cross reference information, institutional references. All the coding forms were key-punched and the information was processed in a batch mode. The information was all in upper case and there wasn't any cataloging scheme - we designed our own, and it really wasn't a cataloging scheme but more of a bibliographic reference type of information. The databases were two: we had PROJTEXT which printed project information on the primary key for that database (project number), and BIBREF (for which the primary key was BIBID - a long 13-digit

number including the 9-digit project number, bibtype code, and an ID number). For hardware/software, we were running on the AID IBM 360/370 mainframe using the INQUIRE database management software. There was online access - searching by keyword, project number, and BIBID. There was an alphabetical descriptor list, which was a typed list. Source materials were kept in the project files - we only had one copy which was reproduced as required. We also produced long computer listings from the system and tailored information packages on specific topics. The people who did the tailored information packages was a contractor called PCI: Molly Hageboeck was involved, Nena Vreeland and Chris Wolter. That was in 1977 that their contract began. That contract was the original predecessor of what you all are doing today.

So far we've got PROJTEXT and BIBREF, two databases ... on project number. That's about all we had so far.

Also in 1976, we had the project Information as a Tool for Development. Now it's getting funded. And there's a contract with UMI. The AID Technical Reports were processed by this contract. Once again there's data coding of bibliographic reference information and abstracts. PN numbers are assigned to technical reports which are batch processed. All source documents are microfilmed. A sequential tape file is created - that's where all this data is going. The IBM mainframe using COBOL software is used to process the data. There's no system access online. Source documents were available on microfiche, or reproduced on paper using the Xerox 970. Information dissemination was through two products: ARDA, and two volumes of the Catalog for Research Literature in Development. Information service was handled out of a Technical Information Center in Room 105.

So what we have up here (diagram) is a technical report file.

In 1976 we also have a database called ADDS, AID Document Distributio System. It's a mailing list database, managed by direct hire staff. Input is the ARDA subscription forms, who wants to receive ARDA in LDCs. Data is keypunched in with batch updates run on the IBM mainframe, started out as a DSS software database converted into INQUIRE. Access points were countries, dates, organization types, subject interests. It was online. Output was mailing labels, either for ARDA mailing or Catalog mailing. There was really no information service provided. It also had AID/W and Mission addresses.

In 1978, DIU was formed as part of the Development Support Bureau. DIU was the predecessor of CDIE/DI. It merged PFC/DI with the TAB/EU. With the merger the AID Reference Center and the Technical Information Center came under common management. The UMI contract ended and was replaced with a contract with SMA in Virginia Beach, VA. That transferred the source document microfilming operation down to Virginia Beach. What happened with the tape database follows.

In 1978, we created the RANDD system - Research and Development Database - run by direct hire staff. It included the UMI technical reports file tape and any new technical reports. It was an online system that built transaction files for batch processing later on. It was our first application where we started dealing with authority files; we were actually beginning to check to make sure that descriptors were spelled right, that corporate authors matched certain codes, etc. In RANDD the primary access was by PN number. It was on the AID mainframe as another INQUIRE database; it was online with access by PN number and descriptor. The SMA contract handled on demand requirements for paper and fiche copies. Sorted reports were printed from this database, ordered by author, title, subject, organization. The primary information service was the Technical Information Center in Room 105 which was a direct hire operation. Margaret Pope was involved in that.

Also in 1978, the PCI analysts contract was phased out. This was an AID management decision; they wanted to cut down on the program budget and rely increasingly more on direct hire staff. The information service was provided by a direct hire librarian/technical information specialist. That meant we had less research analyst type work, less in-depth tailored information packages and more long computer listings that were sent out for project designers to look at, more quick reference type projects, more short-term information specialist projects. Also in 1978, the DIS descriptor vocabulary was brought under control using Avocon, an INQUIRE product that lets you automate the management of vocabulary control.

In 1979, we had an AID Thesaurus concept paper developed. LTS, which had been doing technical processing for the DIS system, now had come up with an approach using the Wang DIS to establish a front-end for all the word processing associated with the abstracts. We were all using coding forms up until 1979; now we were able to word process all the abstracts and logframe information and build a front-end transaction file to do authority file checking. We were also able to shift to upper and lower case in our abstracts. 1979 was also when a TDY was made to IDRC to look at Minisis. And, the DIS project file documents were being microfilmed by SMA, the same contractor handling the technical material.

1980 was when we developed the first Mission application. We developed a Wang VS version of the RANDD database and put it up in Cairo so that people could search by subject, by author, etc. The AID Thesaurus approach was refined; we started looking at how we would deal with merging two separate lists: the DIS descriptor list, the RANDD descriptor list, and Library of Congress modified Subject Headings.

1981 was when the DIHF RFP was originally developed. What it did was merge the LTS contract with the SMA contract, the ARDA printer contract and other small contracts. What we got was a contract with CDSI (Computer Data Systems Inc.), including sub-contracts with CSB (Capital System Group) and LTS. September 1981 was when the DIHF contract began. What we did systemwise was merge

BIBREF (PD's) and RANDD (FN's), two INQUIRE databases, into DOCUMENT. Major access points were DOCID, PROJECT NO. We converted the PROJTEXT file into the PROJECTS database, with Project Number as a major access point. Abstracts was a common database, transparent to the user. In 1981-82 the AID Thesaurus was developed, using OECD Macrothesaurus terms as preferred terms and took up parts of the Unesco Thesaurus structure. The ADDS database was converted to REQUESTR.

In 1982, we basically had the new DIS system in place. The contractors were CDSI, CSG and LTS. We were processing project documents, technical reports and ARDA mailing list forms. We were using modified AACR2 cataloging. We were doing all of these functions online - acquisitions, cataloging...we had never done all of this online before, where updating was simultaneous, etc. Abstracting was quasi-online, using HPWord to enter all the abstracts and some COBAL programs for processing the abstracts and add them. All the source documents were microfilmed. We had Document, Projects, two views or relationships or sub-models with the databases Prjdoc and Prjdoc2, the Thesaurus database and the Requester database. We were using the Minisis software; access was through an online dial-up. You know all the fields you can search on. Source documents were supplied in paper or microfiche. ARDA as a publication was now photocomposed directly from the database, which had never been done before. And we started producing some special bibliographies from the database. Around the end of 1982 was when the USDA Graduate School RASA was started for the Research and Reference contract - that was our information service arm.

1982 was when APDMS was started - the AID Project Document Micrographics System. The input was the official AID project files - online indexing of source documents. Databases were created, one for the Africa Bureau, originally two for the Asia Bureau, one for contracts (Prjsrd), loans and Sntrd. Processing was through Minisis, with online access via project number, title, contract number, year and country. Documents were made available in microfiche or paper copy. Primary dissemination was through computer lists and microfiche sets, and the principal information service is the records management service done through the bureaus.

1984 was when the contract with the USDA graduate school was cancelled and the AED contract was started. 1984 was probably the earliest that the APDMS system was searched for DIS acquisitions.

1985 was when the Acquisitions Lists were begun for Technical Reports. It was also when we first started developing our MicroDIS software product.

Other databases: Inv, Classified database (MicroDIS application), close to 90 MicroDIS installations (projects, PVD's, 7 active Missions, 12 at various stages of development), Bankrd. We started Inventory because we were getting so many materials in to process - we didn't want to spend a lot of time and effort, today we're doing a better job of cataloging them. WIDrd. We're relying

more and more on the bureaus for identifying design documents. Catalog database for non-AID materials in the AID Library. Serials, in the AID Library. Orderrd, used to track orders. Avrd/Avds is a new one for audio-visual materials; there is a possibility of Office of Public Affairs (OPA) materials, and eventually Mission materials.

Another application that is probably going to come up this fiscal year is going to be the Development Education database, for lack of a better term. That will have access to a lot of audio-visual materials developed under AID grants. They want to use this database for online access, but also to create a directory to disseminate. It is a small collection. Several months ago they sent out a survey form to identify other materials produced by Biden-Pell grants.

The Request database was started under USDA.

As you know the DIHF contract and your contract ends this fall. One of the things we want to do is a feasibility study with two principle partners - SER/MO and our office. We're also going to have IRM involved. The primary purpose is to look at alternative technologies to microform. Right now that's how we capture, store and access all the source documents. SER/MO (Service Bureau, Management Operations) is using microform for records management and to send to the Archives. Everything that they filmed after project files were closed the microfilm will go to the Archives, and the standard that Archives accepts after the Master is the silver (second generation microfilm) - that meets their requirements that National Archives has set up for retiring official agency records. There's an interesting, problematic situation that has to be addressed before this occurs. They haven't sent anything to Archives yet although they're ready to. You notice that in the DIS system there are a bunch of documents that have a number like xxxxxxxx -- all those come from the SER/MO system. We catalog those documents, but don't refile them. We have been preparing ourselves for the day when they actually retire their records. We don't have our own copy, so what we're going to have to do for retrospective records that we've selected is use the silver to make a negative. What we're going to have to do from now on is to somehow get earlier in the loop in an automatic process. Our interest is in the institutional memory - it's those records that really have to be retained indefinitely for the agency to learn ... incorporate that into program planning, project design, research. The institutional memory really is a subset of the project files. What we're going to do is look at different optical technologies as an alternative medium for storing source documents. Right now we're using microform; at some point in the future optical technologies may become suitable for long-term record storage and may become a standard established by the Archives. So what we want to do is be more... in terms of what optical technologies are close to become a standard, which technologies should we be experimenting with over the next five years 1990-1995 that we may want to incorporate into the next DIHF contract. Out of this feasibility study, we basically want a statement of work to spell out in more detail what we want to do in terms of alternative technologies to microform.

Optical technologies take a lot of forms; there's WORM - Write Once, Read Many. Probably going to be done by an IQC contractor, not just technologies, but also cost-benefit type issues, looking into what the costs are of microfilm, trying to get some idea of what alternative costs are, looking at management issues, staffing issues.

Another key question is going to be, with optical technology, the question becomes how are we going to get the document with that optical technology. Scanners, two types: digital scanners store images (which can be displayed/manipulated), OCR scanners store text (which can be searched). This raises the question of how much are we going to store digitally and how much in text form - what is important. It's probably going to be a mix.

What we're doing now is using a piece of equipment that the DIHF recently bought, the TDC ImageScanner, to convert microform via a laser printer onto paper (that replaces the Xerox 970 - it broke). But the other thing it does, is it converts the microform into an optical storage medium. One of the things we want to test this year is see how well this works - see if it can handle any of our retrospective conversion needs.

Another thing we're investigating with scanners... we now have something called Truescan, a board that's put in a PC that works with an HP scanner to store the document (?). The other thing we're investing in is called Dmipage, it also works with an HP scanner. The difference between the two is that Dmipage is software based, Truescan is a hardware circuit board. We want to look at these for the purposes of remote scanning of documents on paper.

So we're looking into scanning. There are two purposes of scanning: (1) one is to store a document electronically for later access by someone (instead of microfiche); (2) the second purpose of storing it optically is for use by an end-user, connecting a PC to the HP. In the latter case, you have to look at what kind of a PC you have to have; it may be similar to what Jeanne has for the desk-top publisher, a special display (high resolution) and laser printer.

The key issue is how good the scanning quality is (in image scanning), because what is most important for us initially is how we're going to replace the microfiche - and microfiche is an image capturing medium. A secondary issue is which of these documents we're going to OCR scan - I can't believe that everything that is processed into the system needs to be scanned word for word to search full text on. But there probably is some small critical mass of documents that is probably worthwhile. What we have to do is identify which ones they are.

One big issue still facing us that we haven't resolved in terms of the DIS is that we don't have good access to program information. DIS is still a project level system. One possibility is to OCR scan (full text) this information into a whole new database (from the CDSS's, ABS's, CD's, Action Plans).

(Alternatively, to electronically capturing the information) IRM's strategy is to develop an AID/W area network. Using the mainframe (in Beltsville), they plan to buy a software package called Softswitch - a data transfer or document transfer package. There will also be an e-mail package. IRM would then link all PC's to the mainframe. Softswitch provides the directory services, so that whenever there's a change in the directory, that change is automatically made on the IBM mainframe so Softswitch knows where to send things. What this means for us is the possibility of document transfer (and capture in full text) via electronic mail - with the assistance of IRM. There are problems: for example, how do we know when something is the final copy.