HETEROGENEOUS DISTRIBUTED
QUERY PROCESSING--

THE DAVID SYSTEM

Barry E. Jacobs
Senior Research Scientist
Goddard Space Flight Center
NASA
Greenbelt, Maryland 20771

ABSTRACT

The objective of the Distributed Access View Integrated Database (DAVID) project is the development of an easy-to-use computer system with which NASA scientists, engineers and administrators can uniformly access distributed heterogeneous databases. Basically, DAVID will be a database management system that sits alongside already existing database and file management systems. Its function is to enable users to access the data in the other database and file systems without having to learn the different data manipulation languages.

The scope of this talk will concentrate on the DAVID system and several related issues. First, we describe the problems caused by the diversity of database types and implementations. Second, we discuss two solutions to the problem-- standardization and uniformization. Third, we consider the benefits of the recently developed uniform called "database logic." Fourth, we describe the DAVID system which uses database logic as its framework. Fifth, we outline the status of the development of the DAVID system.

HETEROGENEOUS DISTRIBUTED
QUERY PROCESSING--

THE DAVID SYSTEM

Barry E. Jacobs
Senior Research Scientist
Goddard Space Flight Center
NASA
Greenbelt, Maryland 20771

THE PROBLEM

THERE ARE MANY WAYS FOR A SCIENTIST TO STORE DATA IN A DATABASE:

file approach ——— VMS, DOS, UNIX, ...
relational approach ——— Oracle, Ingres, dBase II, ...
hierarchical approach ——— IMS, RAMIS, ...
network approach ——— DEC—CODASYL, DMS/1100, ...

THIS ABUNDANCE HAS RESULTED IN THE FOLLOWING:

o **A NASA SCIENTIST HAS TO LEARN MANY DIFFERENT ACCESS METHODS IN ORDER TO OBTAIN DATA.**

o **THE DIVERSITY OF TYPES OF DATABASES RESULTS IN AN EXPLOSIVE REPETITION OF DATA AND DATABASES.**

o **THE DATABASE PROFESSIONAL HAS TO LEARN MANY DIFFERENT TERMINOLOGIES IN ORDER TO READ THE LITERATURE.**

OUTLINE OF TALK

1. **THE PROBLEM CAUSED BY THE DIVERSITY OF DATABASE TYPES.**

2. **STANDARDIZATION VERSUS UNIFORMIZATION**

3. **DATABASE LOGIC AS A UNIFORM**

4. **THE DAVID SYSTEM**

5. **DAVID SYSTEM DEVELOPMENT**

STANDARDIZATION VS. UNIFORMIZATION

STANDARDIZATION

o IS AN "A PRIORI" MEANS USED TO ESTABLISH COMMONALITY.

o NEW DATABASES ARE SUBSEQUENTLY BUILT ACCORDING TO THESE "STANDARDS".

o CANNOT HELP IN THE CASE OF ALREADY EXISTING DATABASES NOT ADHEARING TO THE STANDARD.

UNIFORMIZATION

o IS AN "A POSTERIORI" MEANS TO ESTABLISH COMMONALITY.

o PERMITS DATABASE CONSTRUCTION WITHOUT CONSTRAINT ON THE USER.

o CAN HELP IN THE CASE OF ALREADY EXISTING DATABASES.

THE RELATIONAL APPROACH

DATABASES

TAPEINFO				
NOTAPE	TAPETYPE	TITLE1	TITLE2	TITLE3
1003	SDT	NIMBUS6...	BY...	DATA...
1004	SDT	NIMBUS6...	BY...	DATA...

FILEINFO		
PB	FILE	NOTAPE
17400.0	2	1003
17401.0	3	1003
17457.0	2	1004
17459.0	3	1004

RECINFO											
DATE	TIME	LON	LAT	ALT	ZEN	PB	QUALITY	ELECTR	ILLUMIN	CALIB	SCAN
790103	124549	158.57	-1.00	1112.50	153.63	17400.0	0	ON	NIGHT	NO	OFF
790105	145629	25.90	-0.96	1112.30	153.97	17459.0	0	ON	NIGHT	NO	OFF
790117	110204	2.75	10.43	1105.00	34.11	17457.0	0	ON	NIGHT	NO	OFF
790107	105349	4.82	10.84	1103.00	35.59	17401.0	8	OFF	DAY	YES	ON

QUERIES ON DATABASES

```
select (t.tapetype, t.title1, t.title2, t.title3) as result
from   tapeinfo t,
       fileinfo f,
       recinfo r
where t.notape = f.notape
  and f.pb = r.pb
  and r.zen = 153.63
  and r.calib = 'NO'
```

THE DATABASE LOGIC APPROACH

DATABASES

RTAPE										
NOTAPE	TAPETYPE	PLAYBACK								RTITLE
		PB	NOFILE	CATALOG						TITLE
				LON QUALITY	LAT ELECTR	ALT ILLUMIN	ZEN CALIB	TIME1 SCAN		
1010	SDT	e(PLAYBACK) ı1								e(RTITLE) ı1
		81261.4	4	e(CATALOG) ı1						NIMBUS6...
				-133.69 1	-55.62 ON	1123.70 TWILIGHT	98.71 NO	012076 OFF		BY...
										DATA...
				-134.25 1	-56.47 ON	1123.90 TWILIGHT	97.82 NO	012109 OFF		
2010	SDT	e(PLAYBACK) ı2								e(RTITLE) ı2
		81261.0	3	e(CATALOG) ı2						NIMBUS7...
				155.04 0	-80.07 ON	1126.80 DAY	66.65 NO	021349 OFF		BY...
										DATA...

FUNCTIONS

NOTAPE	TITLE1	TITLE2	TITLE3
1010	NIMBUS6...	BY...	DATA...
2010	NIMBUS7...	BY...	DATA...

QUERIES ON DATABASES

```
select (rpc.tapetype, f.title1, f.title2, f.title3) as result
from rtape—playback—catalogue rpc,
     functions f
where rpc.notape = f.notape
  and f.pb = r.pb
  and rpc.zencalib = '153.63'||'NO'
```

DATABASE LOGIC AS A UNIFORM

1. **DATABASE LOGIC IS A THEORETICAL FOUNDATION FOR THE STUDY OF DATABASE ISSUES:**

   First—Order logic is to the Relational Case
   as
   Database logic is to the Heterogeneous Case
   (Relational, Hierarchical and Network)

2. **DATABASE LOGIC HAS BEEN APPLIED TO A NUMBER OF FUNDAMENTAL DATABASE ISSUES:**

   — external—to—conceptual mapping,
   — view update,
   — view integration,
   — database conversion,
   — automatic program conversion, and
   — external axiomatization.

3. **DATABASE LOGIC SERVES AS THE FOUNDATION FOR THE DEVELOPMENT OF A SYSTEM FOR ACCESSING HETEROGENEOUS DISTRIBUTED DATABASES.**

4. **DATABASE LOGIC SERVES AS THE BASIS FOR BUILDING EXPERT SYSTEMS ON TOP OF ALREADY EXISTING DATABASES AND EXPERT SYSTEMS.**

THE DISTRIBUTED ACCESS
VIEW INTEGRATED DATABASE
(DAVID)
SYSTEM
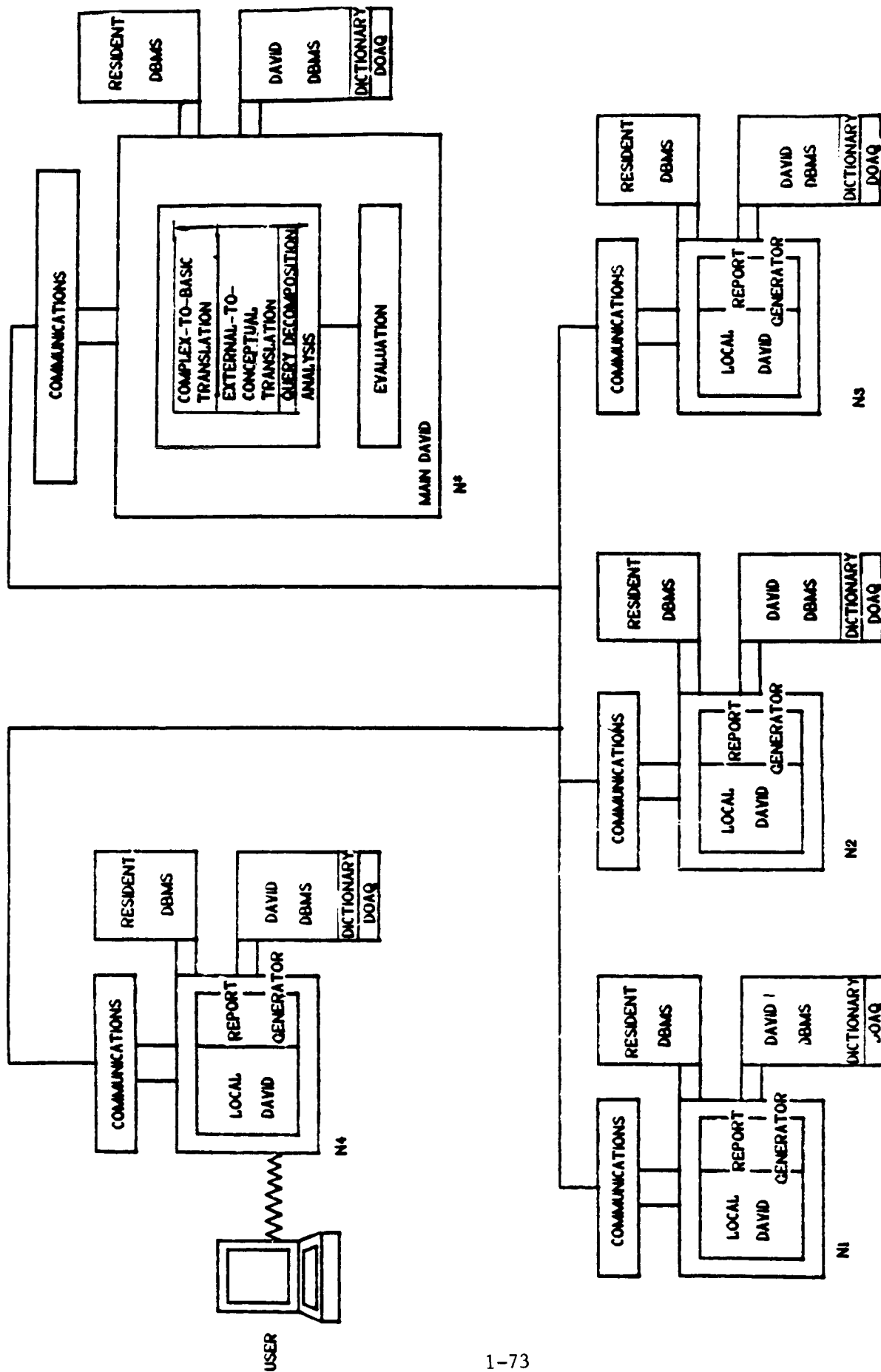
DAVID DEVELOPMENT:

MILESTONES

MILESTONE A. (Summer 1985)

Build a demonstration system using the following environment:

VAX under VMS using an ORACLE DBMS
SUN under UNIX using a DAVID network DBMS
PC under DOS 2.0 using a dBASE II DBMS
MACINTOSH under Mac OS

MILESTONE B. (Summer 1987)

Build a demonstration system using several of NASA's real systems at:

GSFC, ARC, GISS, JPL, NSTL

DAVID DEVELOPMENT:

THE LOCAL DAVID.

Ashok Agrawala, University of Maryland
Michael Anshel, City University of New York
Joseph Aulino, USAF
Kenneth Baum, USAF
Jeanne Behnke, Johns Hopkins University
Nancy Broderick, University of Maryland
Isadore Brodsky, University of Puerto Rico
Dehe Cao, Beijing Polytechnic University
Upen S. Chakravarthy, University of Maryland
Liang Fang, University of Maryland
Adel Gharib, University of Maryland
John Grant, Towson State University
Kim Haynes, TRW Inc.
Hsiao—Fang Hu, University of Maryland
Kyu—Hyun Hwang, SAR Corp.
Barry E. Jacobs, NASA
Thomas E. Jacobs, University of Maryland
Elizabeth Nichols, Digital Analysis Corporation
Joseph Nichols, Digital Analysis Corporation
Surrendra Ray, SAR Corp.
Ira Sack, Stevens Institute of Technology
Michael Shapiro, Bell Labs
Duc Tran, Digital Analysis Corporation
Satish Trapathi, University of Maryland
Cynthia A. Walczak, NIH
Jack Welch, The Catholic University of America

DAVID DEVELOPMENT:

PROPOSED DAVID INTERFACES

CENTER	PROPOSED TESTBED	UNIVERSITY
ARC	PLDS, LAS	C.U.N.Y.
GSFC	PCDMS, PLDS, LAS	U of Maryland Catholic University
GISS	ISCCP	C.U.N.Y.
JPL	PODS, PLDS, PPDS, LAS	U of Southern California
NSTL	PLDS, EOS	Louisiana State University
JSC	TBD	U of Houston Louisiana State University
LeRC	TBD	U of Toledo
MSFC	TBD	U of Alabama
LaRC	TBD	Old Dominion University

SUMMARY