Databorough

Data Model Extraction
with X-Analysis

» *Recover Application Design*
» *Extract Data Model*

Richard Downey
Stuart Milligan

WHITE PAPER

Preface

Developing tools services for analyzing and re-engineering RPG applications for the last 20 years, has given Databorough a unique view of the very large and complex world of legacy applications running on System i, iSeries and AS/400. This paper focus on the Model extraction (RPG/COBOL/2E) for System i applications using X-Analysis. It discuss new concepts and methods for design recovery of monolithic RPG/COBOL/2E applications into modern application architectures.

Contact info@databorough.com for a copy of the white paper, the Red book and trial software.

Table of Contents

Introduction

An IBM i business application is primarily an application written over a relational database. Therefore, no map of an enterprise application would be complete without the database architecture explicitly specified. Not just the physical specifications and attributes, but the logical or relational constraints too.

With the possible exception of CA:2E systems virtually all RPG or COBOL applications running on IBM i have no explicit relational data model or schema defined. This means that millions of lines of RPG or COBOL code must be read in order to recover an explicit version of the relational model.

The relational model of an enterprise application is an extremely powerful piece of information and potentially valuable asset to the organization. By the term model, we are referring to the foreign key or relational model, not just the physical model of the database. Once explicitly defined, the relational model or architecture of the database can be reused in a number of scenarios including:

- Understanding application architecture
- Data quality referential integrity testing
- Test data extraction
- Test data scrambling and aging
- Building BI applications & Data Warehouses
- Data mapping for system migrations
- Build object relational maps for modernization

X-Analysis has the unique capability of automatically deriving the explicit system data model from a legacy RPG or COBOL application.
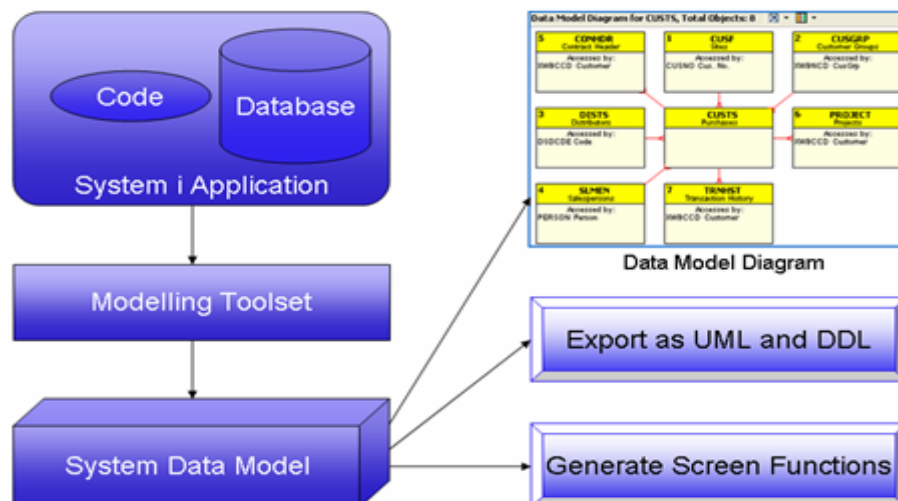

Illustration 1: Overview Modeling Tool Set
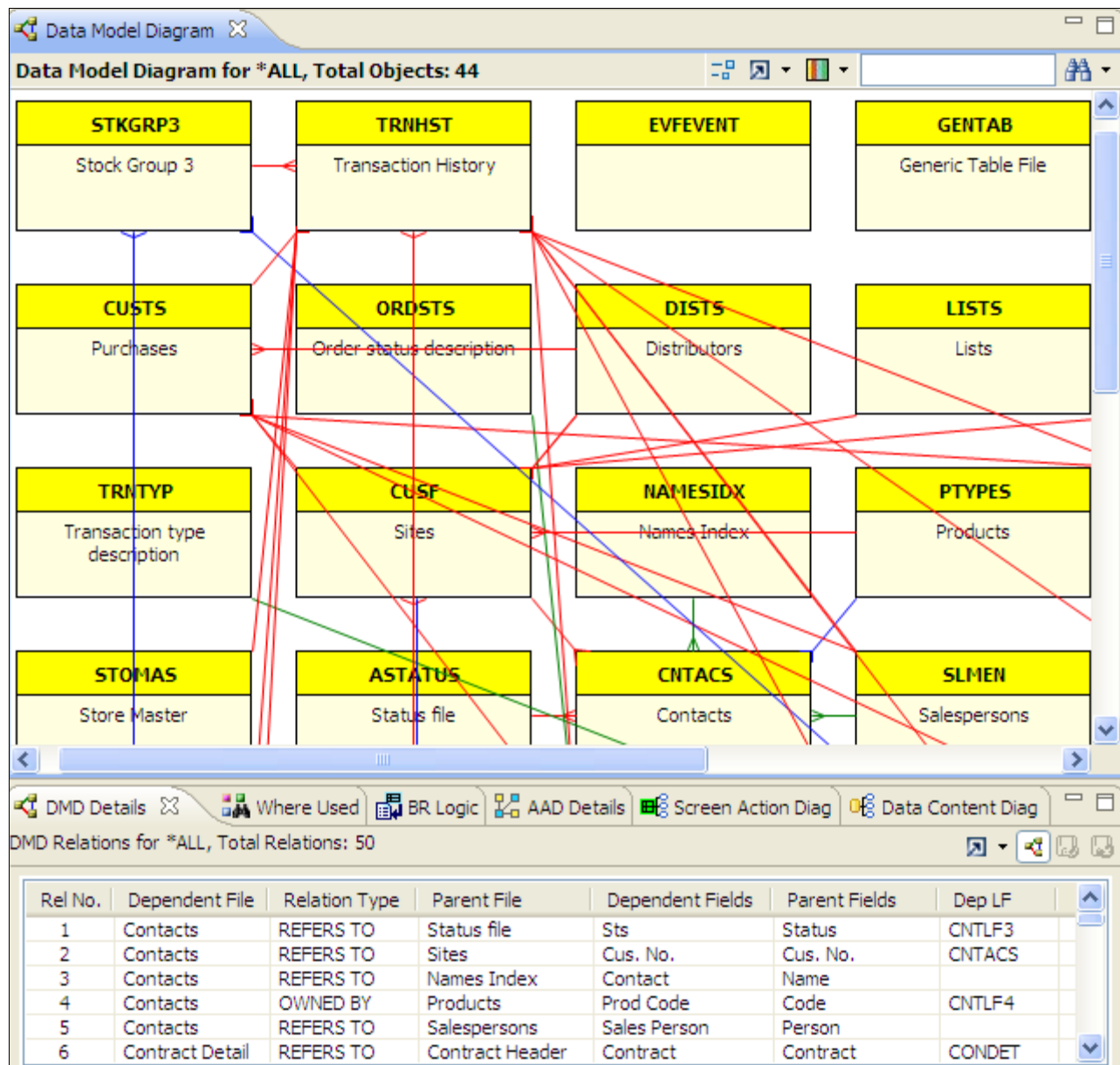
Entity Relationship Data Model Diagrams

Illustration 2: Data Model Diagram of XAN4CDEM Application

X-Analysis shows the data model as an interactive diagram. By default, all relationships gets displayed but if required, file centric or application area specific filtering of the diagram can be induced by simple point & click mechanisms. The exact nature of the relationships between each entity is colour coded and described in the detail view. A detailed report describing the data model can also be produced, which can be printed if required.

For example here is the data model for the CUSTS file in a program. This shows the relationship between CUSTS and the related objects.
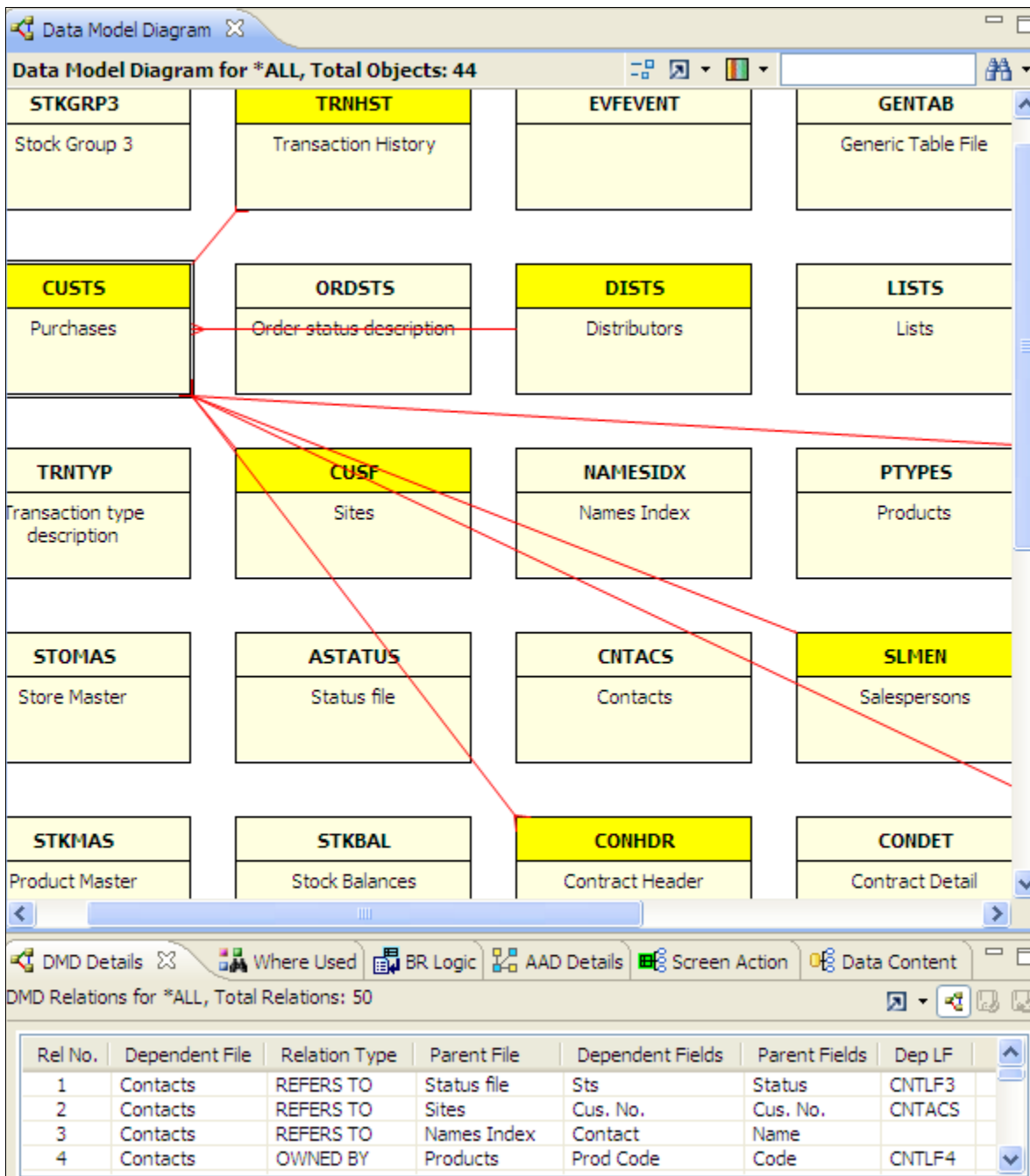

Illustration 3: DMD of XAN4CDEM Application with CUSTS selected

Double-click on 'CUSTS' to better understand the DMD relationships for CUSTS file.

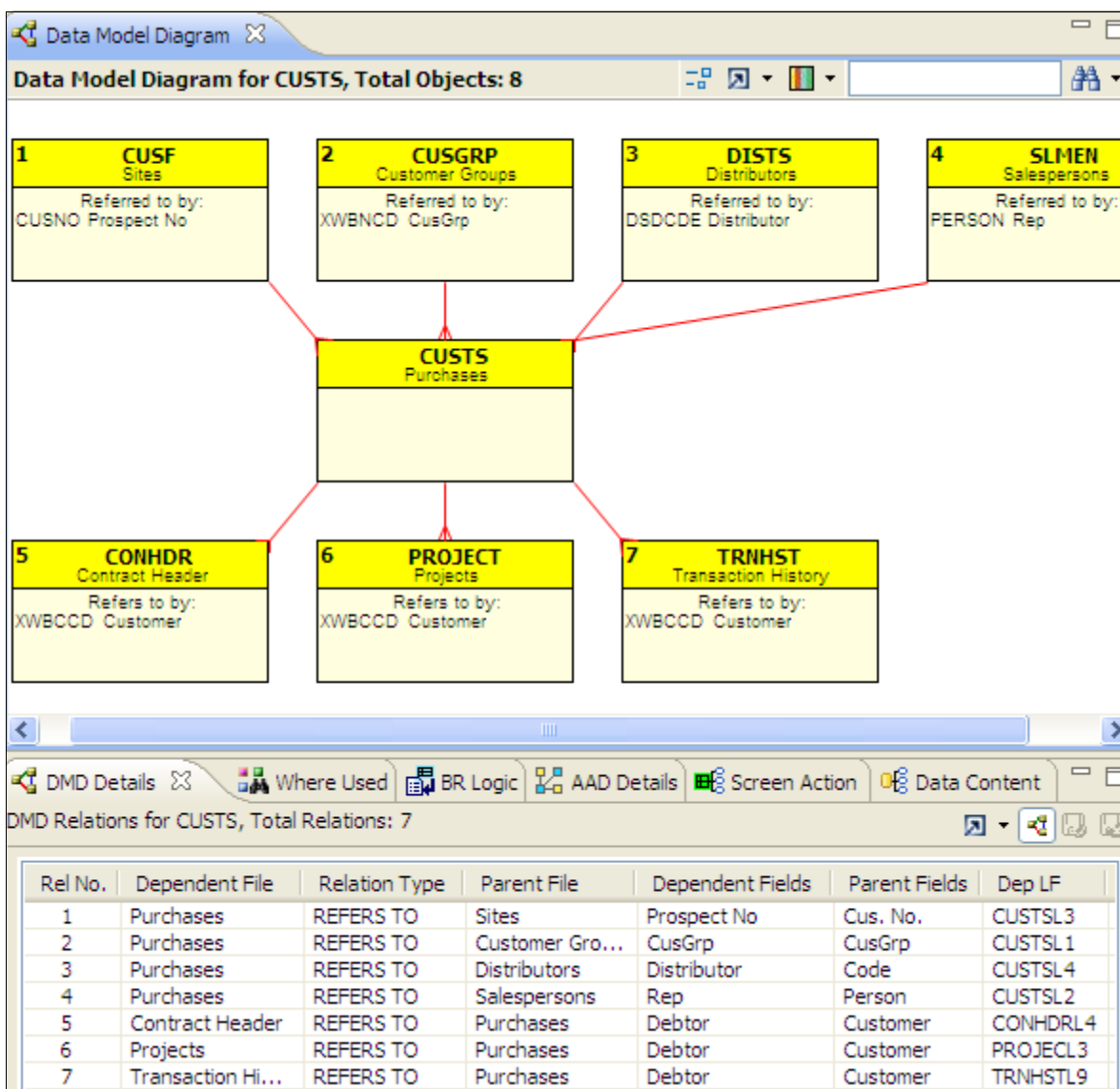| Rel No.	Dependent File	Relation Type	Parent File	Dependent Fields	Parent Fields	Dep LF
1	Purchases	REFERS TO	Sites	Prospect No	Cus. No.	CUSTSL3
2	Purchases	REFERS TO	Customer Gro...	CusGrp	CusGrp	CUSTSL1
3	Purchases	REFERS TO	Distributors	Distributor	Code	CUSTSL4
4	Purchases	REFERS TO	Salespersons	Rep	Person	CUSTSL2
5	Contract Header	REFERS TO	Purchases	Debtor	Customer	CONHDRL4
6	Projects	REFERS TO	Purchases	Debtor	Customer	PROJECL3
7	Transaction Hi...	REFERS TO	Purchases	Debtor	Customer	TRNHSTL9

Illustration 4: Data Model Diagram of CUSTS File

With CUSTS in the middle, we get a much clearer picture of how the objects are interrelated.

Individual files could be selected to display relationships pertinent to the selected file only. The detail of each relationship can be displayed, and the underlying data is stored in a structured relational database for easy reporting/analysis. A data model diagram is, by its very nature, extremely complex when produced for an entire application, thus it is necessary in such circumstances to logically sub divide an application into areas.

The diagram can be exported to MS-Word or Visio, and a DDL version of the entire Data Model exported to tools such as CA's - All Fusion, System Architect, Microsoft SQL Server, or Oracle.

Application Area Concept

X-Analysis creates application areas from part of one or multiple systems. It is possible to subdivide programmatically, an application into logical modules or areas. This can be done within the context of a single system e.g. order entry from the operational system, or indeed specific parts from multiple systems e.g. Order Entry from the operational system and sales ledger from the financial system.

The entry of powerful user-defined rules combined with information about the entire application that is stored in the X-Analysis repository are applied by X-Analysis to make this an instant and repeatable task. These application areas can then be used to accurately and confidently put a ring around all objects impacted and required from projects such as re-engineering, re-development, replacement or project management allocation of development responsibility.
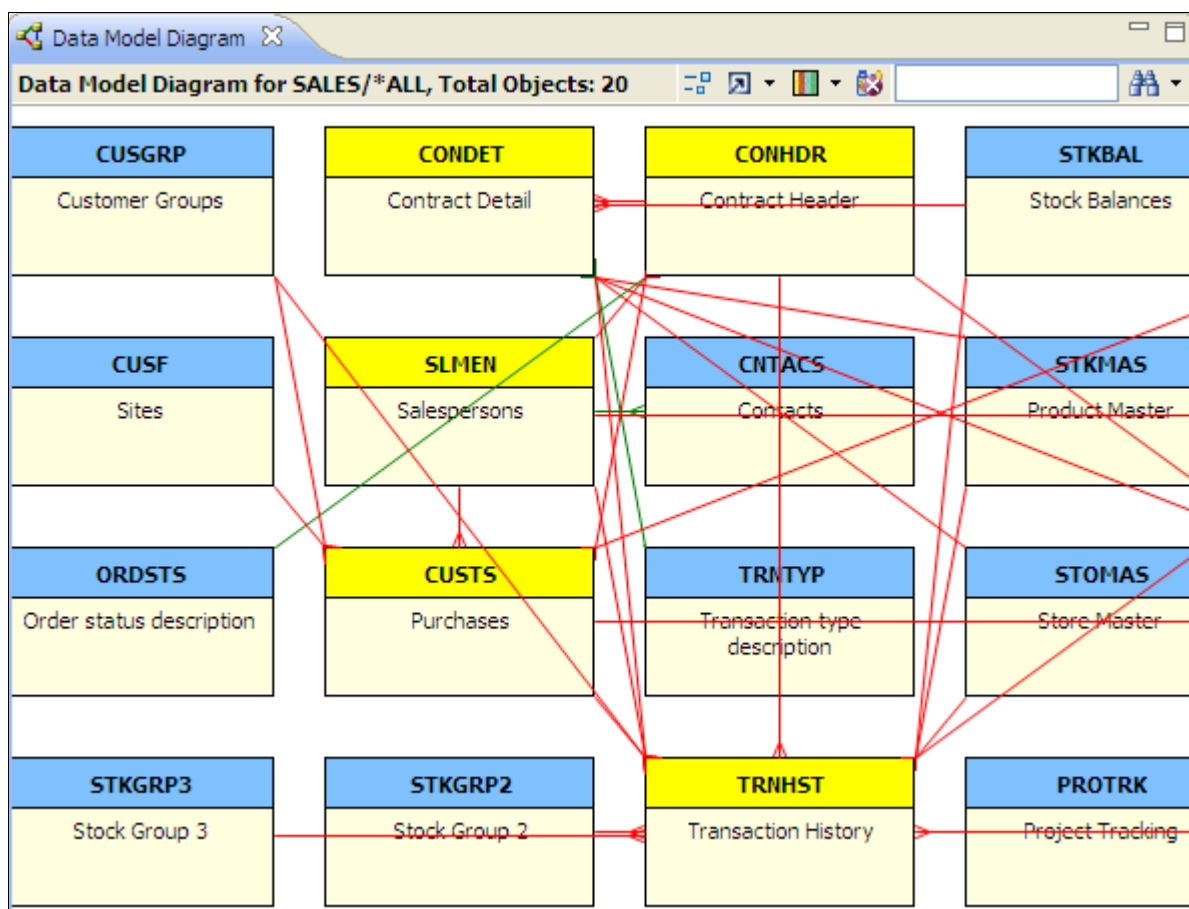

Illustration 5: DMD on SALES application area with external relationships

The DMD shows a few files in yellow and the remaining files in blue. The files in yellow are Internal, i.e. belong to the SALES application area and the files in blue are external.

This implies that the file in yellow, viz., CONHDR is grouped under the application area SALES. All other files of the application are external to the SALES application area.

Thus, we deduce that the Application area facilitates the grouping of an application into different Business Areas, thereby making it easier to understand and maintain the Application.

Data Dictionary

To build a Data Model X-analysis automatically creates a Data Dictionary, which extracts and catalogs detailed information for every field in each file in the application database. Much of this data is the standard meta-data extracted for each file and stored on the XDD file - for instance field and column names, field size and field type.

Thus, record meta-data is readily available for use by other applications.

The main Data Dictionary sub-menu options are:

- Entities
- Access Path
- Fields
- Relationships
- Relationship Details
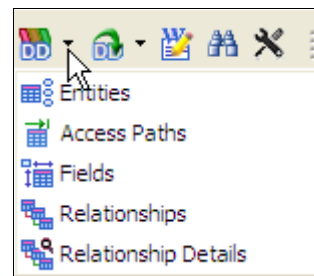- Business Rule Summary
- Business Rule Programs

Illustration 6: Data dictionary Menu Options

Entities (XPIDS)

The Primary Identifiers view is displayed. This is also the default view when the Data Dictionary is invoked from X-Analysis.

The identification of the correct primary identifier is crucial to the building of an accurate data model. The primary identifier is determined by an examination of all the access paths for the file and is verified against the data in the file. All the primary identifiers are written to the XPIDS file.

PF Name	PID File	Key1	Key2	Key3	Key4	Key5	Key6	Key7	Ke
ASTATUS	ASTATUS	STATUS							
CNTACS	CNTLF4	PRPCDE	STATUS	CUSNO					
CONDET	CONDET	XWORDN	XWABCD						
CONHDR	CONHDR	XWORDN							
CUSF	CUSFL3	CUSNO							
CUSGRP	CUSGRP	XWBNCD							
CUSTS	CUSTS	XWBCCD							
DELIVA	DELIVA	XWBDCD							
DISTS	DISTS	DSDCDE							
EVFEVENT	EVFEVENT	EVFEVENT							
GENTAB	GENTAB	FLDNAM	CODVAL						
LISTS	LISTS	LSLCDE							
NAMESIDX	NAMESIDX	IXNAME							
ORDSTS	ORDSTS	XWSTAT							

Illustration 7: Data Dictionary - Entities

Access path (XKEYMAP)

This file records all of the access paths available to each physical file. There is one XKEYMAP record for each access path.
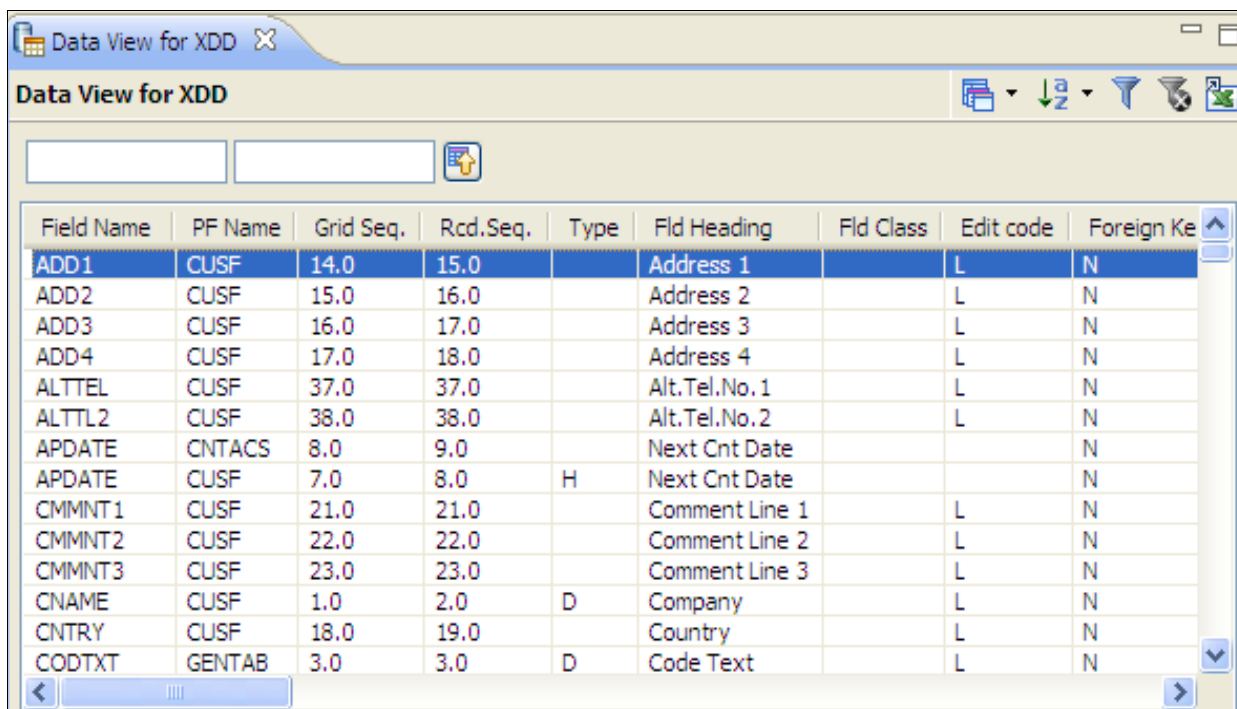

Illustration 8: Data Dictionary – Access paths

Up to 10 key fields can be associated with each field.

Fields (XDD)

This view gives information for every field in each file in the application database. The data dictionary file holds the field information extracted by the Data Modeling process. Each field from each physical file in XPIDS is listed. The meta-data extracted can be interpreted and utilized by applications to read, write and format the fields correctly.

Field Name	PF Name	Grid Seq.	Rcd.Seq.	Type	Fld Heading	Fld Class	Edit code	Foreign Ke
ADD1	CUSF	14.0	15.0		Address 1		L	N
ADD2	CUSF	15.0	16.0		Address 2		L	N
ADD3	CUSF	16.0	17.0		Address 3		L	N
ADD4	CUSF	17.0	18.0		Address 4		L	N
ALTTEL	CUSF	37.0	37.0		Alt.Tel.No.1		L	N
ALTTL2	CUSF	38.0	38.0		Alt.Tel.No.2		L	N
APDATE	CNTACS	8.0	9.0		Next Cnt Date			N
APDATE	CUSF	7.0	8.0	H	Next Cnt Date			N
CMMNT1	CUSF	21.0	21.0		Comment Line 1		L	N
CMMNT2	CUSF	22.0	22.0		Comment Line 2		L	N
CMMNT3	CUSF	23.0	23.0		Comment Line 3		L	N
CNAME	CUSF	1.0	2.0	D	Company		L	N
CNTRY	CUSF	18.0	19.0		Country		L	N
CODTXT	GENTAB	3.0	3.0	D	Code Text		L	N

Illustration 9: Data Dictionary – File Fields

Relationships (XRELS)

This file records the relationships between all the physical files that form the data model.

Illustration 10: Data Dictionary – File Relationships

There are three types of relationship that can be identified:

- Owns - PID to PID relationship
- Accesses - Access Path to Access Path relationship
- Refers to - foreign key to PID relationship.

Relationship Details (XSHKEYS)

This file is the counterpart to XRELS. It describes in further detail the composition of every 'File-to-File' relationship recorded in XRELS. XSHKEYS describes each of the 'Field-to-Field' relationships that make up these XRELS relationships.

Data Browsing

Another important feature of the Modeling is Data Browsing. File data may be browsed using X-Analysis. X-Analysis provides an 'View Data' option which displays the records of *FILE type objects (PF).

Illustration 11: Browse Data

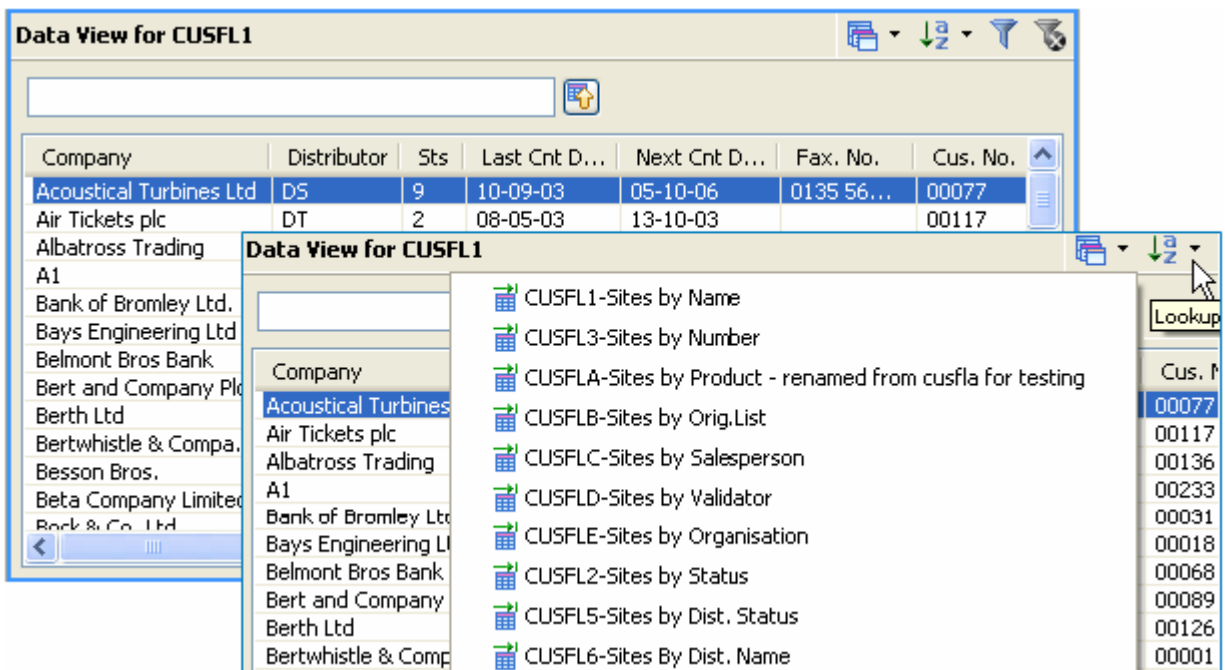X-Analysis provides the facility of subdividing an application area into groups of objects that meet the user defined selection criteria. The Application Area diagram displays an overview for the application. It is an interactive diagram and by clicking on different parts of the system you can see the relationships between either all parts or just the area you've clicked on and the areas it relates to.

Verifying Data Relationships

Verify Data Relationships uses the data model information to automatically, verify that all application data satisfies the relationships deduced by the data model. The product examines each data record in turn to see if its referential integrity relationships are satisfied. Each integrity relationship that is breached is separately reported on in a comprehensive audit print produced by the product.

Illustration 12: Verify Data Relations Dialog

Parameters
The parameters passed internally in the verification are:

Function Library

The library which contains the data model and application specific files created by the command XDMODEL. This is the name of the library that was entered as the Function Library on the XDMODEL command. Its passed the X-Analysis X-Reference library.

Database Library Name

The library name which contains the application database files. The command is submitted with the value *LOADLIB, implying, each database file should be located in the same library used when XDMODEL was run.

X-Analysis Application Area

The command uses the application area, on which this option is invoked, to control which files in the data model should be verified. Only relationships between files in the specified application area will be verified. Since an application area has been specified, the X-Analysis X-Reference library too is passed as parameter to the command.

Run Mode

Run mode parameter is passed as *REPORT, implying, generate a report showing the primary keys and the foreign keys of each record that breaches a given relationship up to a maximum number for each relationship as specified by the Maximum records parameter.

Maximum Records

If you have selected a run mode of *REPORT then XVERIFY will print a report showing the primary keys and the foreign keys of each record that breaches a given relationship. This parameter allows you to specify the maximum number of records printed for each relationship.
If a dependent file contains less than the maximum number of records specified (but more than zero records) and all the records breach the relationship then that relationship will also be updated.

Furthermore, X-analysis can show a report showing the primary keys and the foreign keys of each record that breaches a given relationship.

Illustration 13: Orphaned Records Report for SALES application area

Subset Data

The Subset Data option creates a subset of all the files in an application area specified in X-Analysis. It assumes the XDMODEL (the new model database) database is in the same library as the X-Analysis database. It also assumes the name of the group filter to be used is the same as the name of the application area.

In the example in the Subset Data dialog, a Test library(CUSCARES1) containing the files falling under the CUSCARE application area. The records on the file obey the Subset Filter criteria specified on the files under CUSCARE application area.
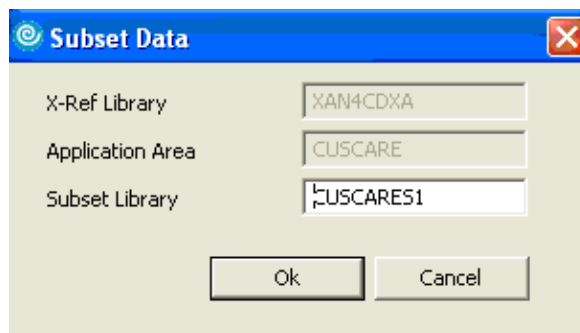
Illustration 14: Subset Data Dialog

Parameters
The parameters passed internally in subset data are:

X-Analysis Application Area

The command uses the application area, on which this option is invoked, containing the files to be subset.

X- Analysis library

The name of the X-Analysis cross-reference library being used, containing the application area definition.

Test Library Name

The name for the test library specified, which this command will create.

Subset Mode

The subset mode is passed as *LIVE, implying, records are copied to the test library if they meet the filter criteria.

Include owners

This parameter is passed as *YES, i.e. copy records in owning files which own records, which have been copied to the subset.

Rebuild triggers/journals data

This is passed as *NO, i.e. do not rebuild the data. (It the data has already been built then any triggers, journals or constraints will be removed.).

After a subset of data has been created its data can be viewed in the Data View window.

Customer	Name	Statement Account	CusGrp	Rep	Distributor	Prospect No	Related Account	Ta
ACC1	Bertwhistle & Company Ltd	ACC23	PV	STU		1	ACC43	ta
ACC2	Bock & Co. Ltd		GV	JKL		14		
ACC3	Besson Bros.		GV	NWD		15		
ACC4	Media Enterprises Ltd		AG	MTT		17		A
ACC5	Bays Engineering Ltd		PV	STU		18		ta
ACC6	Gough Research plc		RT	JKL		6		A
ACC7	Karst plc		RT	JKL		50		
ACC8	Bays Engineering Ltd	ACC1	AG	MTT		18		
ACC9	Newt Foods UK Ltd.		GV	MTT		9		
ACC99	Prime Computer Systems	ACC1	GN	STU		56		

Illustration 15: CUSTS grid

Note the records displayed are for Customer ACC1 and ACC2 only.

In addition, a report for the subset CUSCARES1 can be created, showing the primary keys and the foreign keys of each record that breaches a given relationship.

File	Owner	Primary Key	Foreign Key
		Orphaned Records Report For SubSet Data Library CUSCARES1	
CONDET	CONHDR	XWORDN\XWABCD:125486\000003	XWORDN:125486
CONDET	CONHDR	XWORDN\XWABCD:125487\000003	XWORDN:125487
CONDET	CONHDR	XWORDN\XWABCD:125489\000004	XWORDN:125489
CONDET	STKBAL	XWORDN\XWABCD:000005\000031	XWABCD\XWAACS:000031\UK
CONDET	STKBAL	XWORDN\XWABCD:000012\000077	XWABCD\XWAACS:000077\USA
CONDET	STKBAL	XWORDN\XWABCD:125478\XA	XWABCD\XWAACS:XA\UK
CONDET	STKBAL	XWORDN\XWABCD:125478\000005	XWABCD\XWAACS:000005\UK
CONDET	STKMAS	Relationship verified.	
CONDET	STOMAS	Relationship verified.	
CONHDR	CUSTS	Relationship verified.	
CUSTS	CUSF	Relationship verified.	
STKBAL	STKMAS	Relationship verified.	
STKBAL	STOMAS	Relationship verified.	
TRNHST	CONDET	XWE4NB\XWDLDT\XWC8DT:200201\2002-01-14\00000000009	XWORDN\XWABCD:000001\000001
TRNHST	CONDET	XWE4NB\XWDLDT\XWC8DT:200201\2002-01-18\00000000018	XWORDN\XWABCD:000001\000001

Illustration 16: Orphaned Records Report for Subset Data Library CUSCARES1

Archive Data

A data model can be used to set up complex archive requests for automatic running. A full GUI over the database allows both specifications of archive selection criteria & also for reinstatement of any records incorrectly archived.

In the example to the right the Archive library, CUSCAREA1, containing the files falling under the CUSCARE application area is produced. The records on the file obey the Archive criteria, which can be specified, on the files under CUSCARE application area.

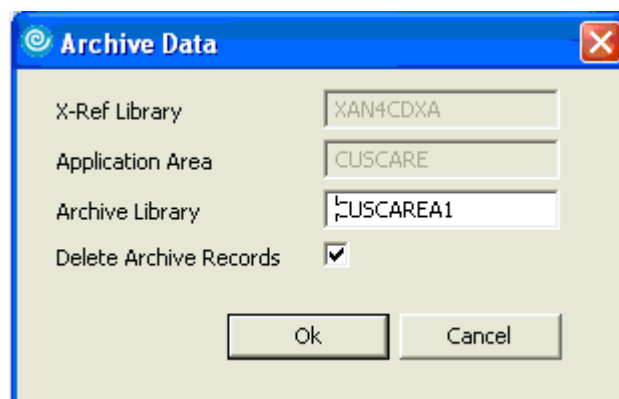This adds the archive library, CUSCAREA1in this case, as a sub-item to the CUSCARE application area.

Illustration 17: Archive Data Dialog

The Archive Data option works by submitting a XARCAREA command in batch. This command archives all the files in an application area specified in X-Analysis. It assumes the XDMODEL database is in the same library as the X-Analysis database. It also assumes the name of the group filter to be used is the same as the name of the application area.

Parameters
The parameters passed internally to the XARCAREA command are:

X-Analysis Application Area

The command uses the application area, on which this option is invoked, containing the files to be subset.

X-Analysis library

The name of the X-Analysis cross-reference library being used, containing the application area definition.

Archive Library

The name for the Archive library specified, which this command will create.

Archive Mode

The archive mode is passed as *RETAIN, i.e. records are created in the archive library if they meet the archive criteria but they are not deleted from the original.

Include owners

This parameter is passes as *YES, i.e. archive records in owning files which own archived records.

Rebuild triggers/journals data

This is passed as *NO, i.e. do not rebuild the data. (It the data has already been built then any triggers, journals or constraints will be removed.).

After a data has been archived it can be viewed in a data view window.

For example, to see the records in CUSCAREA1library do the following:

Customer	Name	Statement Account	CusGrp	Rep	Distributor	Prospect No	Related Account	Ta
ACC1	Bertwhistle & Company Ltd	ACC23	PV	STU		1	ACC43	ta
ACC2	Bock & Co. Ltd		GV	JKL		14		
ACC3	Besson Bros.		GV	NWD		15		
ACC4	Media Enterprises Ltd		AG	MTT		17		A
ACC5	Bays Engineering Ltd		PV	STU		18		ta
ACC6	Gough Research plc		RT	JKL		6		A
ACC7	Karst plc		RT	JKL		50		
ACC8	Bays Engineering Ltd	ACC1	AG	MTT		18		
ACC9	Newt Foods UK Ltd.		GV	MTT		9		
ACC99	Prime Computer Systems	ACC1	GN	STU		56		

Illustration 18: CUSTS grid

Note the records displayed are for Customer ACC1 and ACC2 only.

X-analysis can also reinstate archived records. Reinstate on a selected Archive library under the application area brings up a dialog showing the selected referenced Library, Application area and Archive library. Further it prompts the user to reinstate all records or tagged records in the specified archive library. If the request is successful, the message, "Reinstate" job submitted successfully, appears on the status bar.
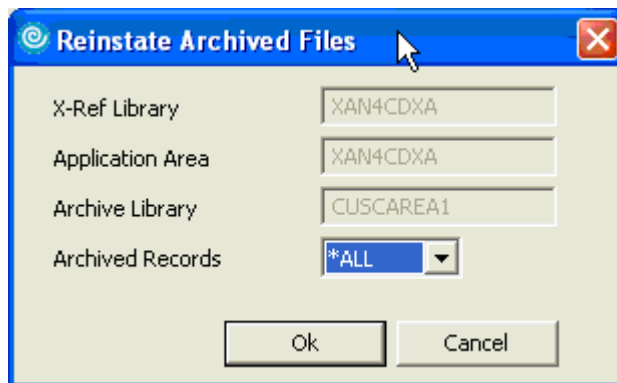

Illustration 19: Reinstate Archived Files

The *ALL option simply reinstates all the records in the archive library back into the application data library.

The *TAGGED option reinstates only those archived records which have been marked for reinstatement. The next section explains how individual or selected records can be reinstated.

Summary

A complete data model accurately describing all possible relationships between each file is essential for productive maintenance and development work.

The X-Analysis data modelling extraction process interrogates a host application and database and outputs the Data Model files which describe the internal structure and comparison of the system. The data-model thus generated is used for:

Visualization

Data-model can be viewed and analyzed in X-Analysis

Export to other CASE Tools

The Data Model can be exported to UML, DDL and other CASE tools such as COOL:Biz

Generating Screen Functions

The generation of Screen Functions is incorporated into the XDMODEL command

This data model information describes all possible relationships between each file and is essential for productive maintenance and development work. Such a model also provides the foundation upon which critical data administration tasks can be automated.

Richard Downey and Stuart Milligan
© Databorough