Analysing Microsoft Access Projects: Building a Model in a Partially Observable Domain

Santiago Bragagnolo, Nicolas Anquetil, Stéphane Ducasse, Seriai Abderrahmane, Mustapha Derras

Santiago Bragagnolo

@sbragagnolo
santiago.bragagnolo@berger-levrault.com

- 1st year PhD
- Industrial Thesis
- Software Migration

Context: Berger-Levrault

5 Centuries old company

1700 employees

5 Continents

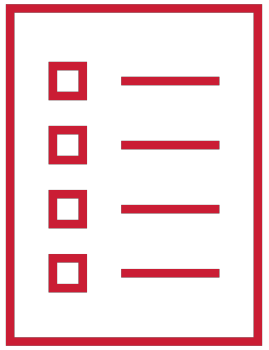
51000 clients that reach millions of users

Berger Levrault

Technology evolves, information systems must evolve too: Migration Happens

Microsoft Access Solutions @ Berger Levrault

> 90 Solutions



900 to 2700 Forms x Solution

Migration Context: Objectives

- Standalone to Multi-layered
- Monolith to Microservice
- Desktop GUI to Web GUI

Migration Context: Requirements

- Selective
- Interactive
- Iterative

Implications

- Up to date Data
- Precise and Detailed Data
- Memory & CPU

Berger Levrault

What does the literature propose as first step on migration?

Related Works

Static Analysis

- AST Extracted from text files
- Focus on a static structure of the software

Related Works

Dynamic Analysis

- Runtime crawling
- Behavior dynamics and real composition

Related Works

Hybrid Analysis

- Static analysis enhanced by dynamic
- Dynamic analysis enhanced by static

Berger Levrault

Microsoft Access: a Partially Observable Domain

Microsoft Access

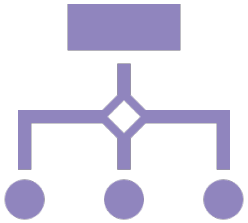
4GL Language

Point and Click Development

Generates GUI related code

Produces Software based on intensive GUI Interaction

Uses proprietary binary format



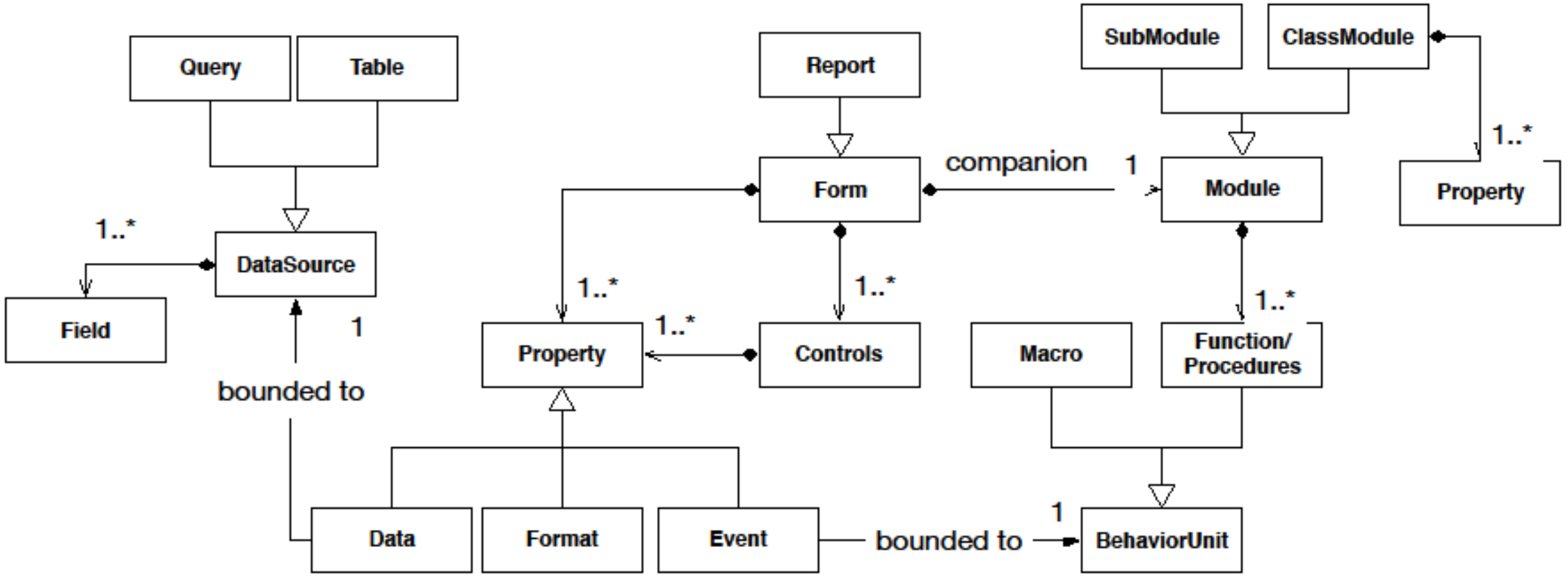
Provides First Class Citizens

GUI: Form and Reports

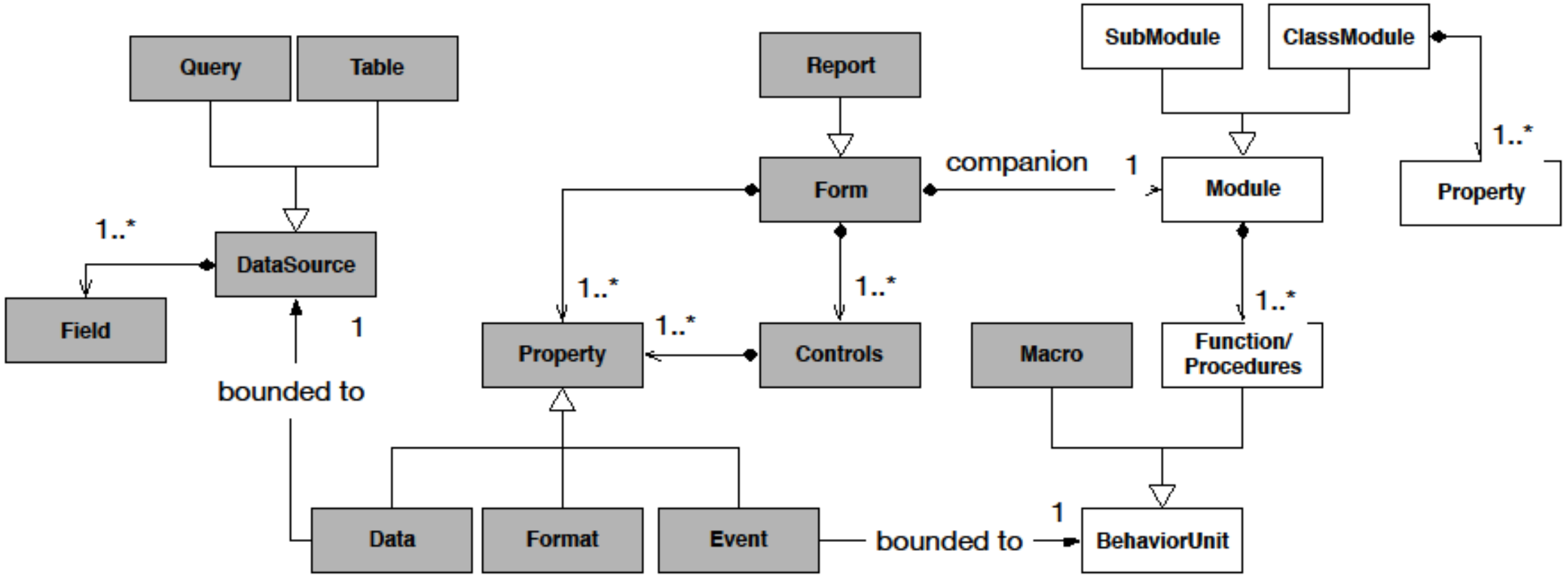
Functional: Class Modules, Modules and Macros

Data: Tables and Queries

Microsoft Access: An Application Overview



Microsoft Access: GUI Exporting Limitations



Microsoft Access: Programmatic Exporting

SaveAsText

- Saves a Microsoft Access first class citizen as a file
- Heterogeneous output
- Non documented

LoadFromText

- Loads an object from a file
- Non documented

Berger Levrault

How to practice a structural analysis on an application if all the data is inside the IDE?

Study path

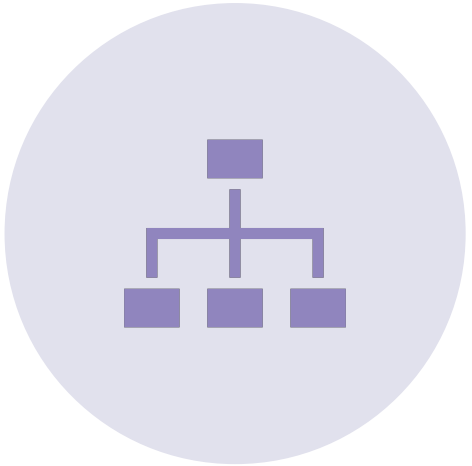
RQ#1	Can we obtain an application representation by querying the IDE runtime?
RQ#2	Are we able to reengineer the meta-data sources into a model?
RQ#3	Is the obtained model suitable for migration?

Berger Levrault

Can we obtain an application representation by querying the IDE runtime?

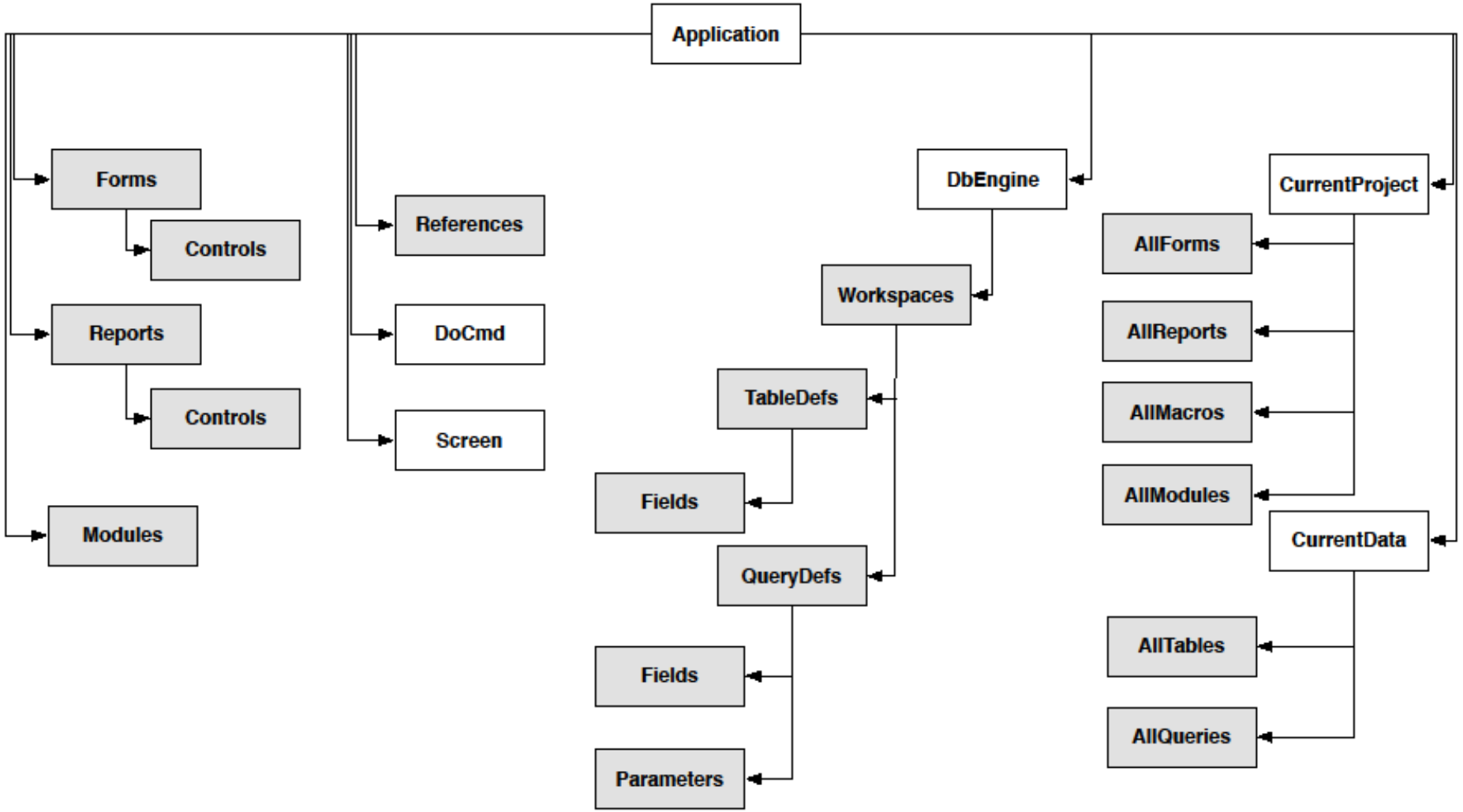
Microsoft COM Technological Overview

STANDARD BINARY INTERFACE



COMPONENT OBJECT MODEL

Microsoft Access Object Model & COM



Object Model Overview

COM Handle	**Pointer to a remote object**
Application: Represents a running Microsoft Access Application	**DoCmd**: Command object. Reifies the available operations
	References: Collection of static dependencies
	CurrentProject: Project currently opened by the running application. Contains Forms, Reports, Modules and Macros descriptions
	CurrentData: Database currently opened by the running application. Contains TableDef and QueryDef
	Forms Reports and Modules: Collections containing the opened Forms / Reports / Modules. Each object is bounded to the real "thing"

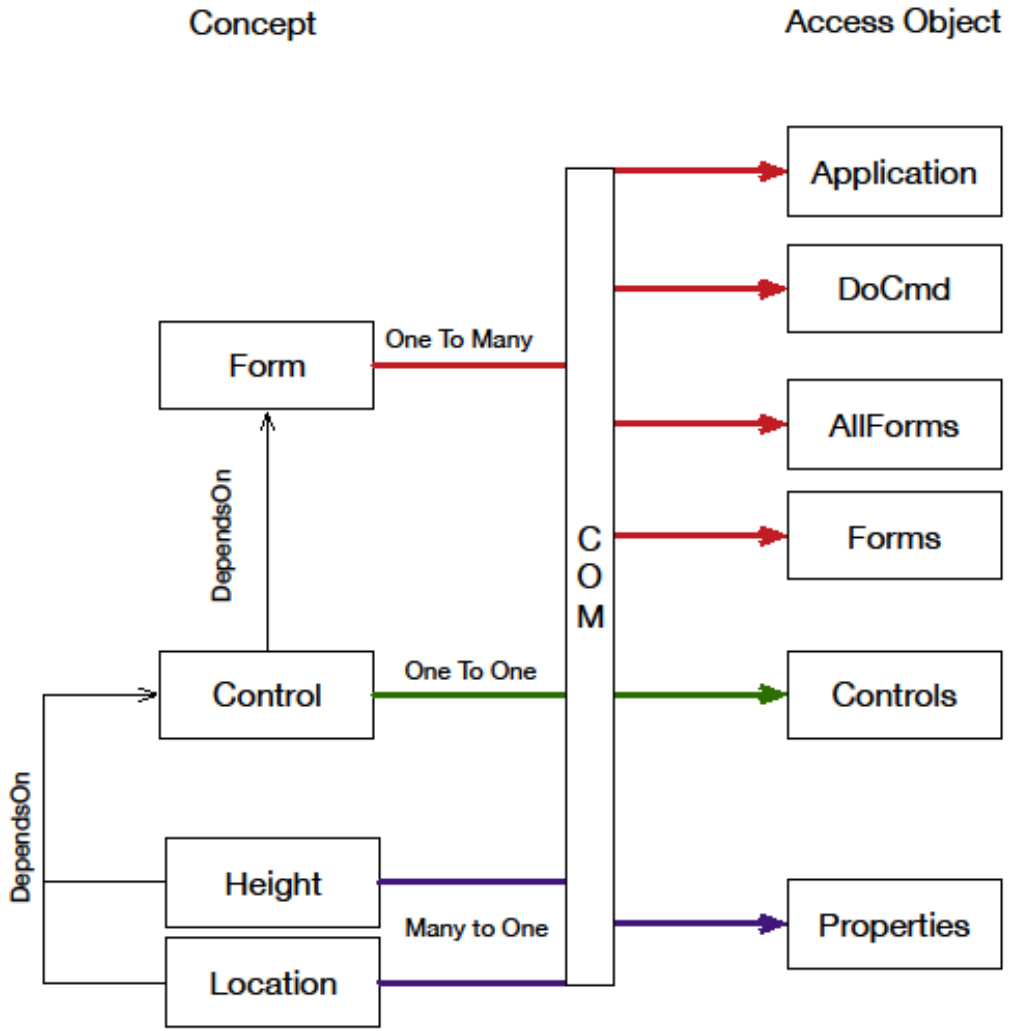
Re-Engineering Challenges

Reading a property of a COM Entity may give us an other COM Entity

A model entity type may be mapped to more than one entity type

First class citizen objects must be dynamically loaded to be analyzed

Different model Entity types may be mapped to one COM entity type



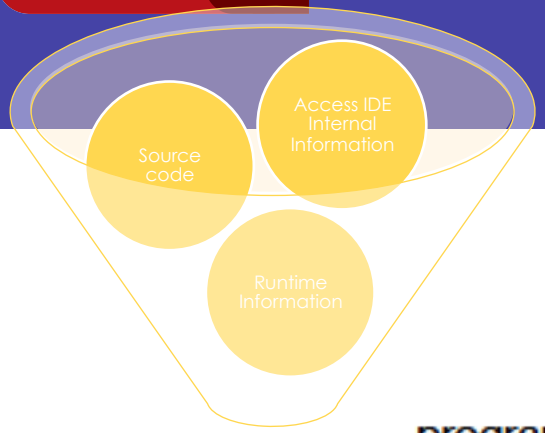
Berger Levrault

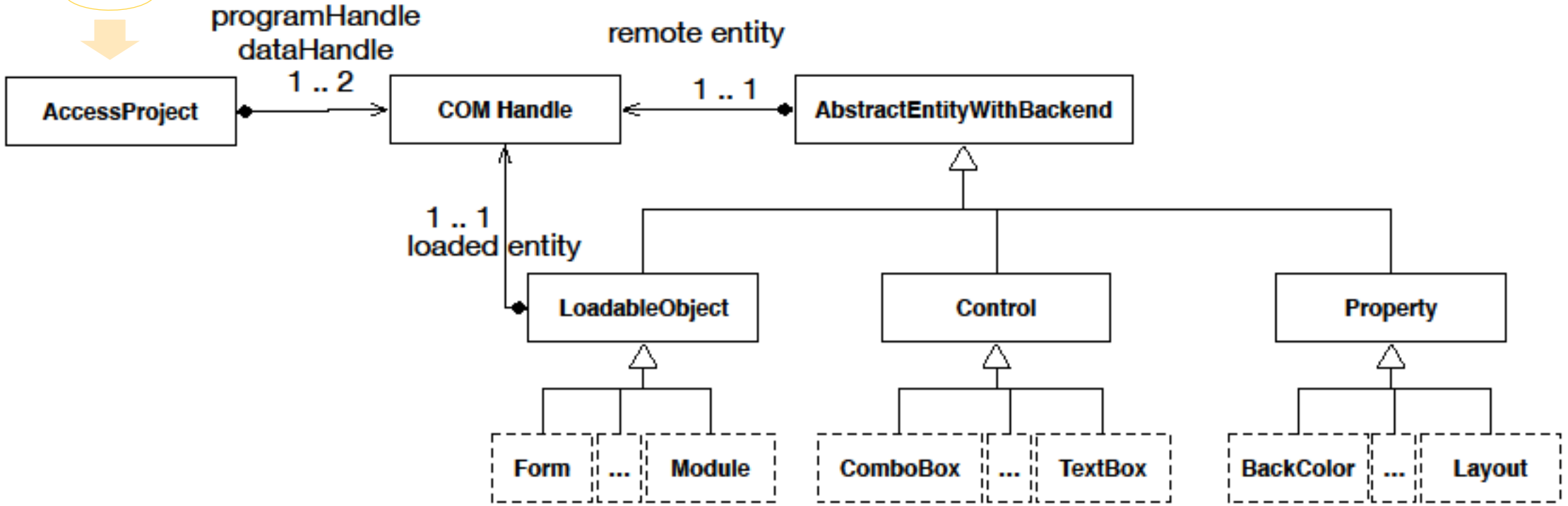
Are we able to reengineer the meta-data sources into a model?

Mixing Static and Internal MS Access information

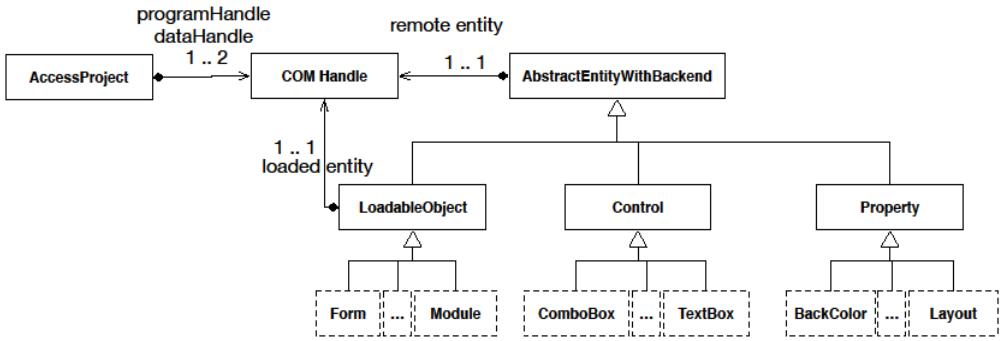
Microsoft Access Analysis

An Architecture Implementation for an Online Model





Addressing challenges



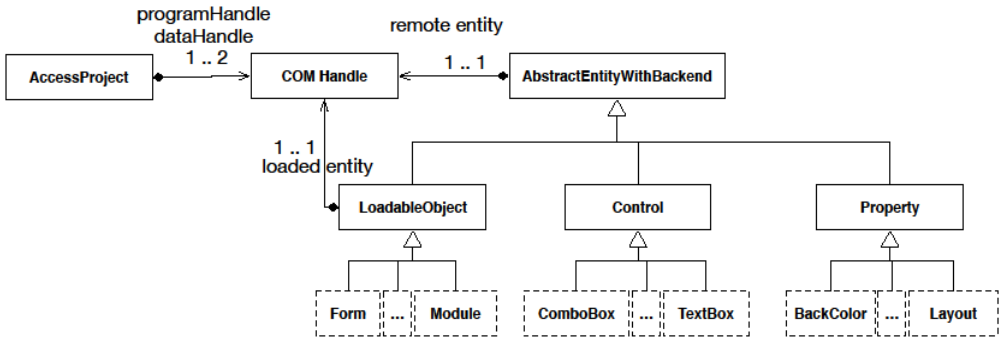
Each model type must know which readings return COM entities. A factory is used for mapping by COM type

Only Loadable objects are mapped to more than one entity. This is managed individually

Managed At the level of LoadableObject Class

When reading properties a Factory that is used for mapping types by property value

Addressing Restrictions

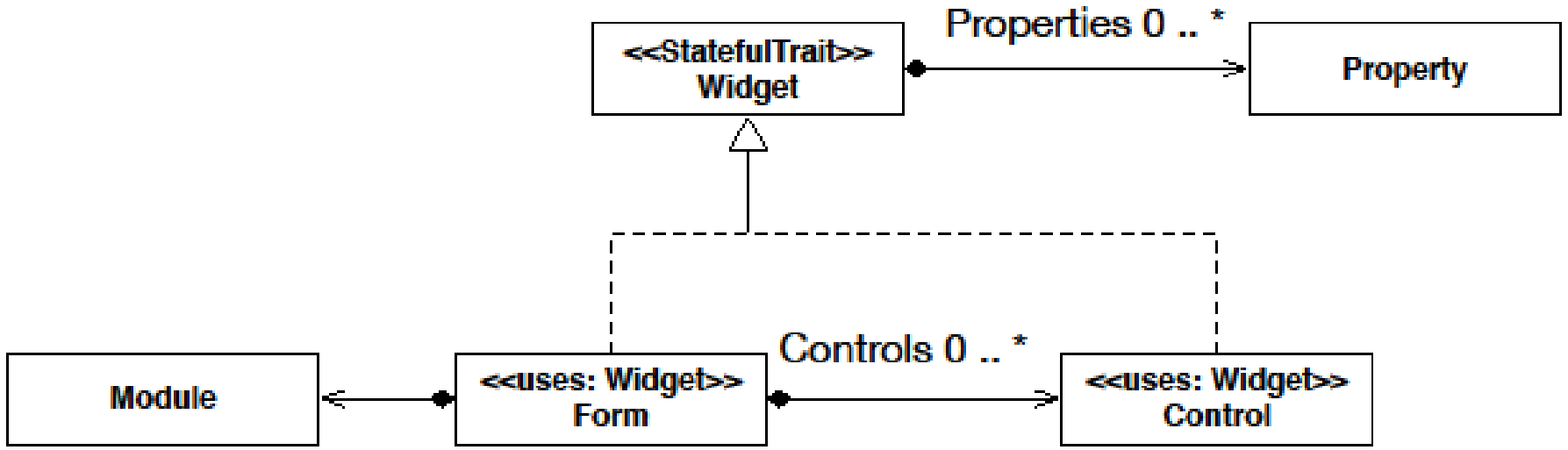


The online nature of the implementation ensures up to date data

Our solution can access what ever information reachable by COM

Lazy load + caches

A Metamodel Implementation for an Online Model



Berger Levrault

Is the obtained model suitable for migration?

Is the obtained model suitable for migration? Validation Required!

What is to migrate?

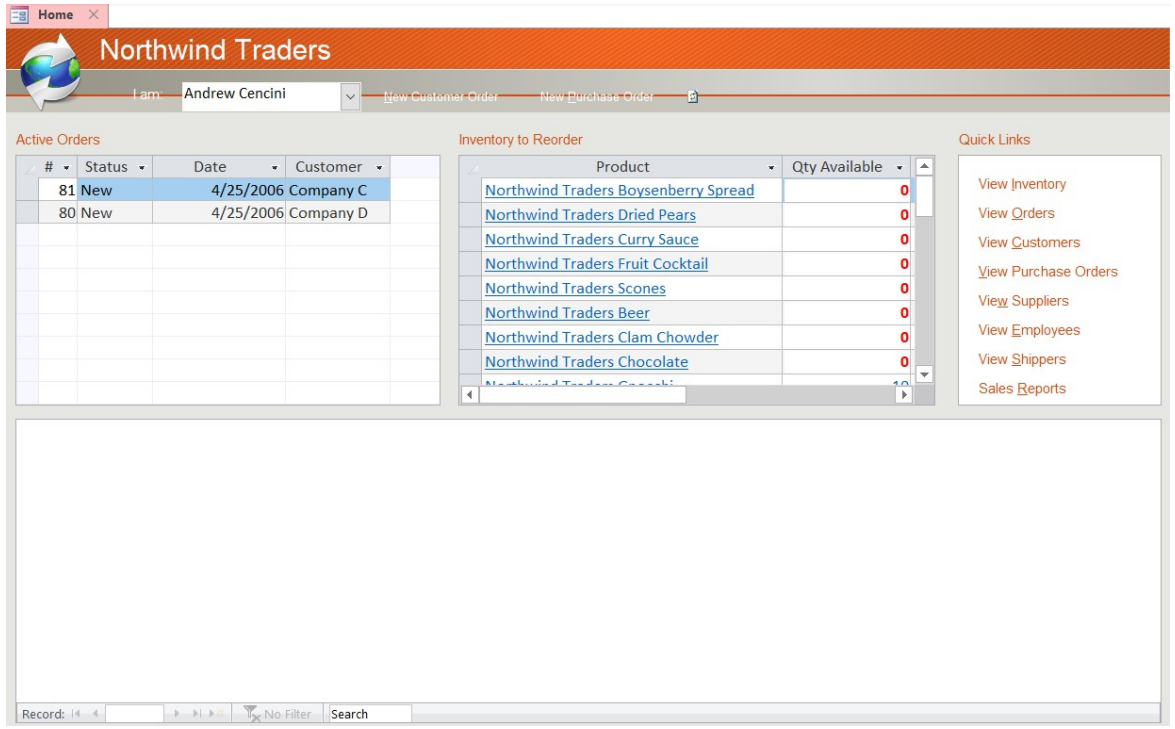
To reproduce a software defined in an origin environment into an other environment

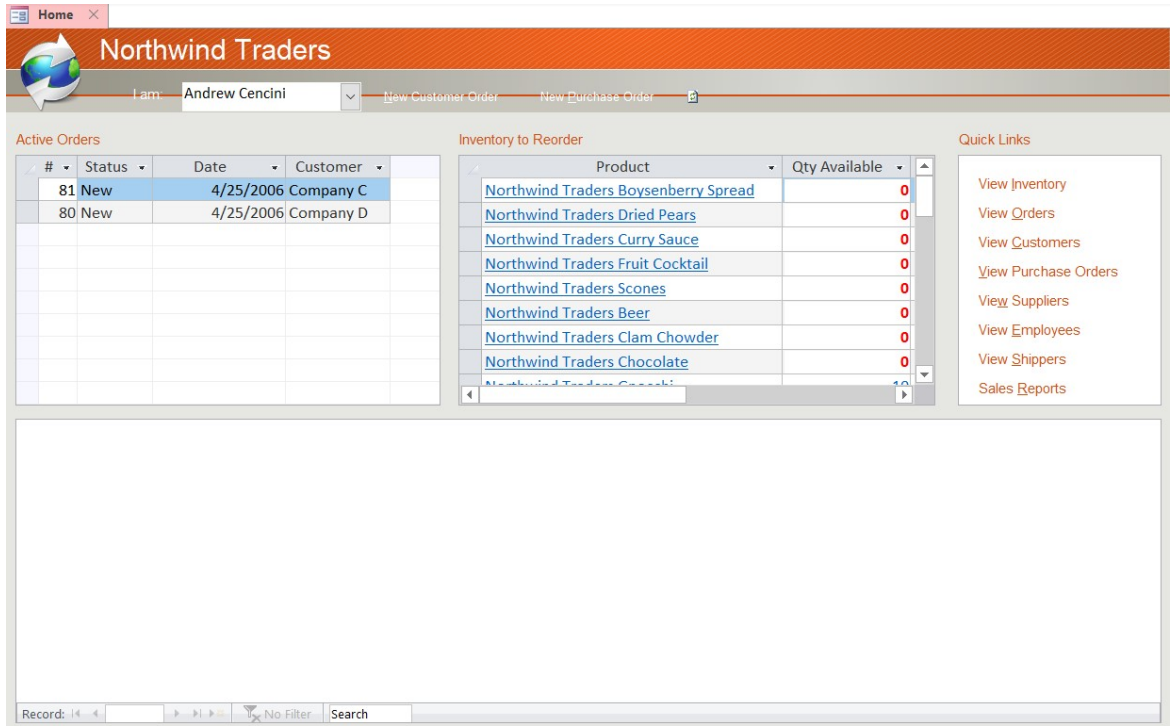
Is the obtained model suitable for migration? Validation Required!

What is the simplest migration?

To reproduce the same software in the same environment: To Replicate

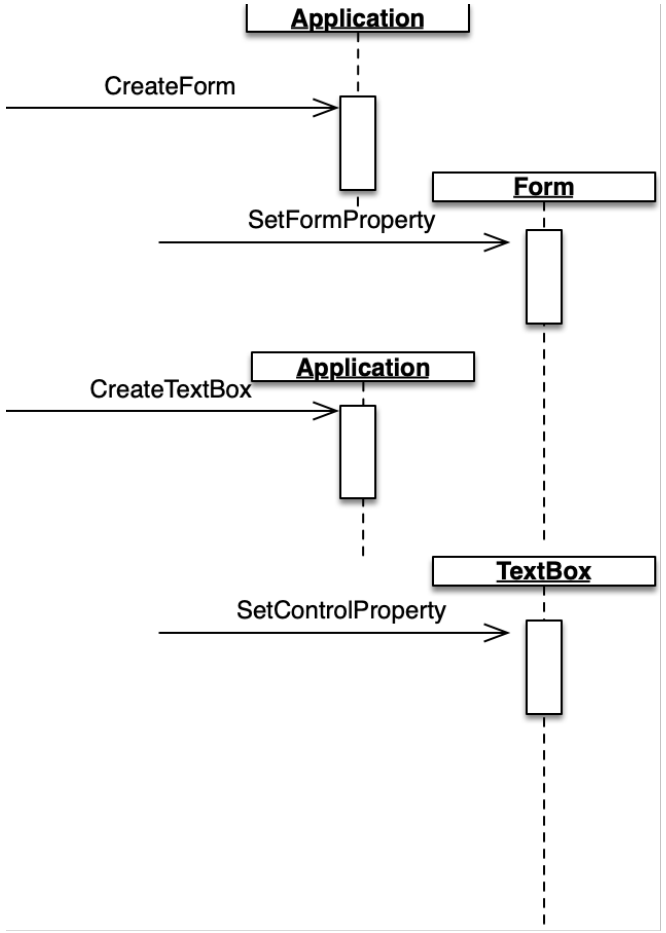
How should a replica look like?





How do we replicate a project?

1. Access allows programmatic creation
   - Projects
   - First Class Citizen
   - Controls / Columns
   - Properties



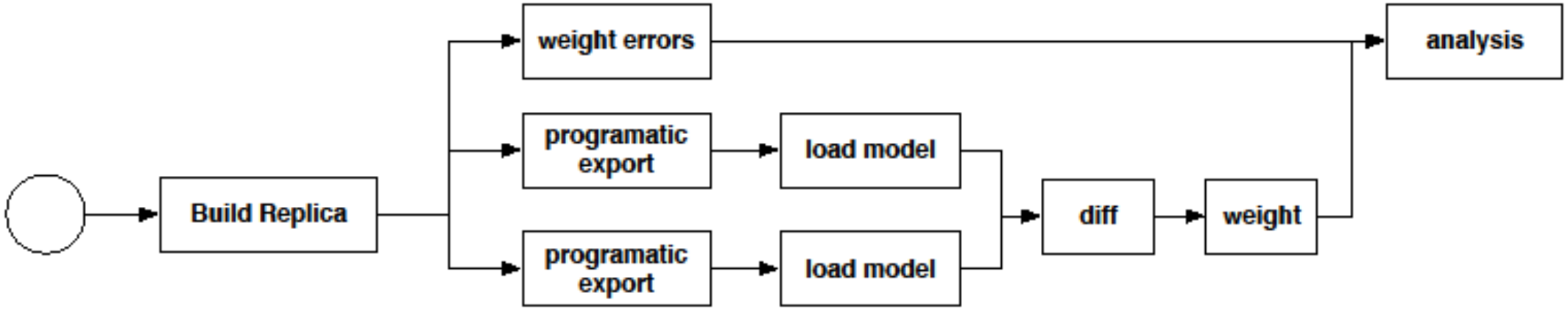
General demographics

Project	Remote Table	Table	Remote Query	Query	Module	Classes	Report	Forms
Northwind	0	20	0	27	6	2	15	34
CUTLCOMP	0	1	0	0	3	0	0	0
CUTL	7	3	0	1	26	62	0	8
CRIR	5	4	0	16	6	0	2	3
CPDI	0	1	0	0	2	0	0	0
CHABIL	11	2	0	27	8	1	1	10
CDDE	0	1	0	0	2	0	0	0
CAUNIT	0	1	0	0	4	15	0	1
ACCUEIL	25	7	0	13	6	67	5	33
Access Examples	0	10	2	14	11	1	8	13

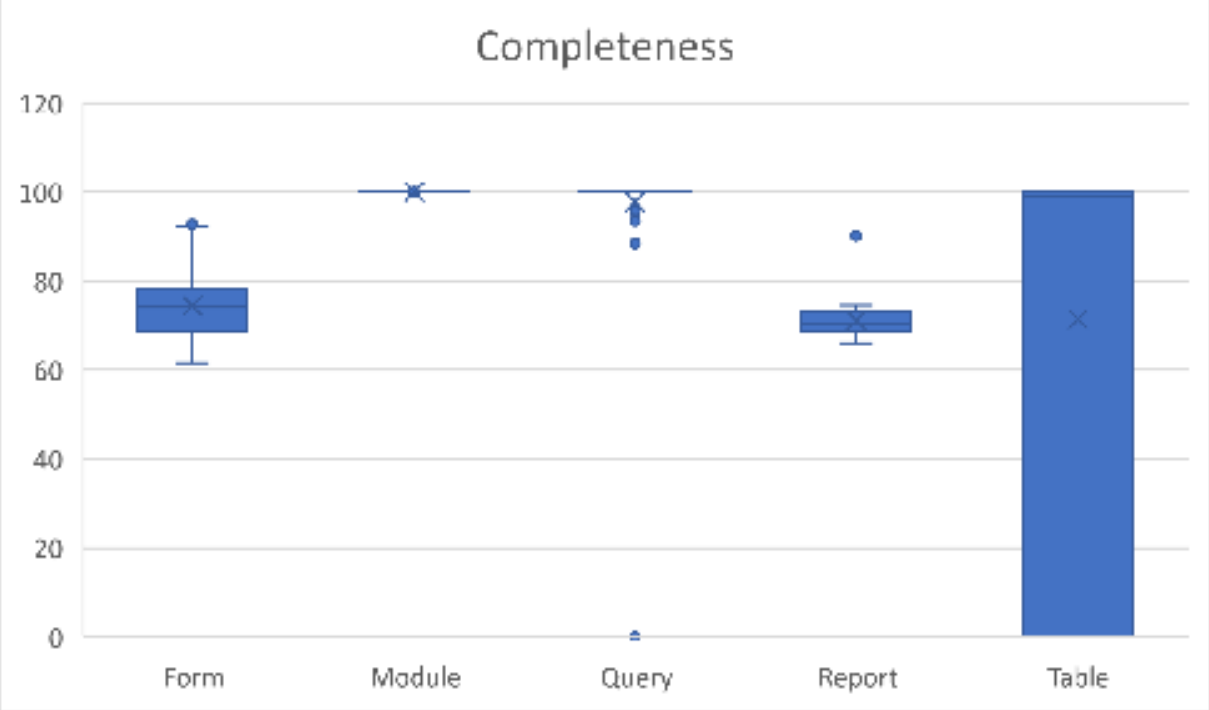
Results Overview

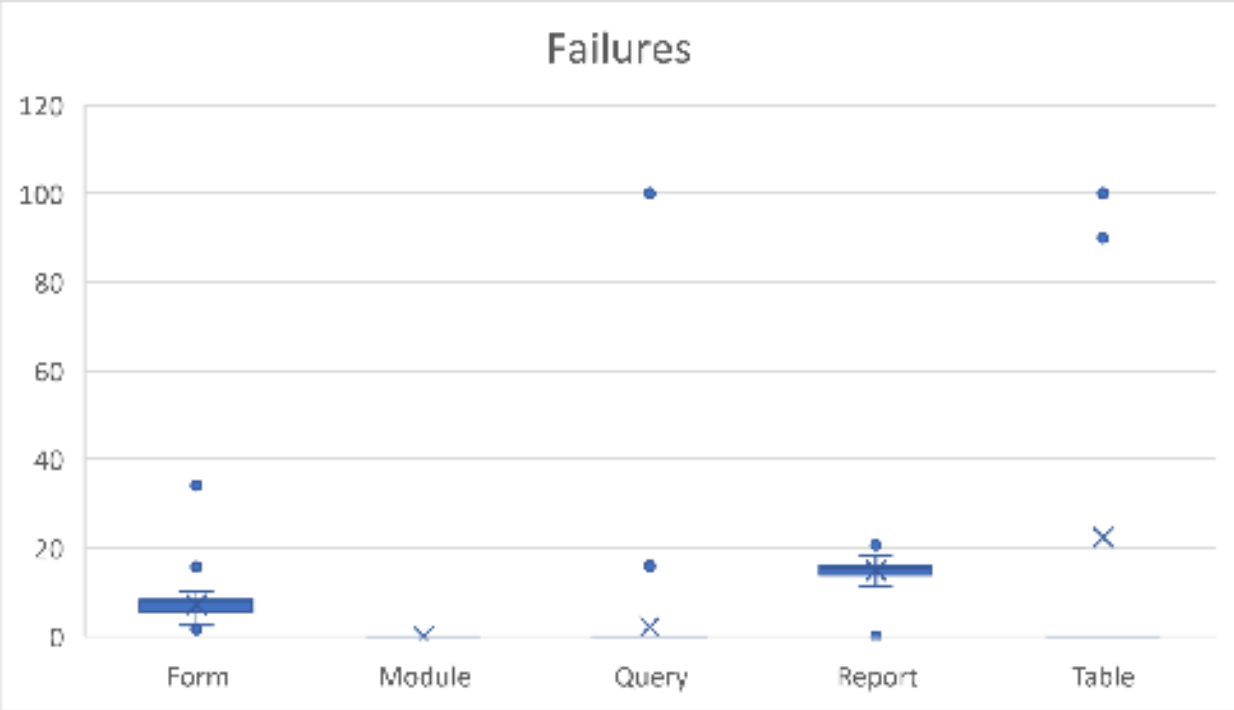
	Reference	Table	Query	Module	Report	Forms	Total
#Elements	70	98	100	222	31	102	623
#Replicated	69	50	98	222	31	102	572
#Failures	1	48	2	0	0	0	51

Validation Methodology

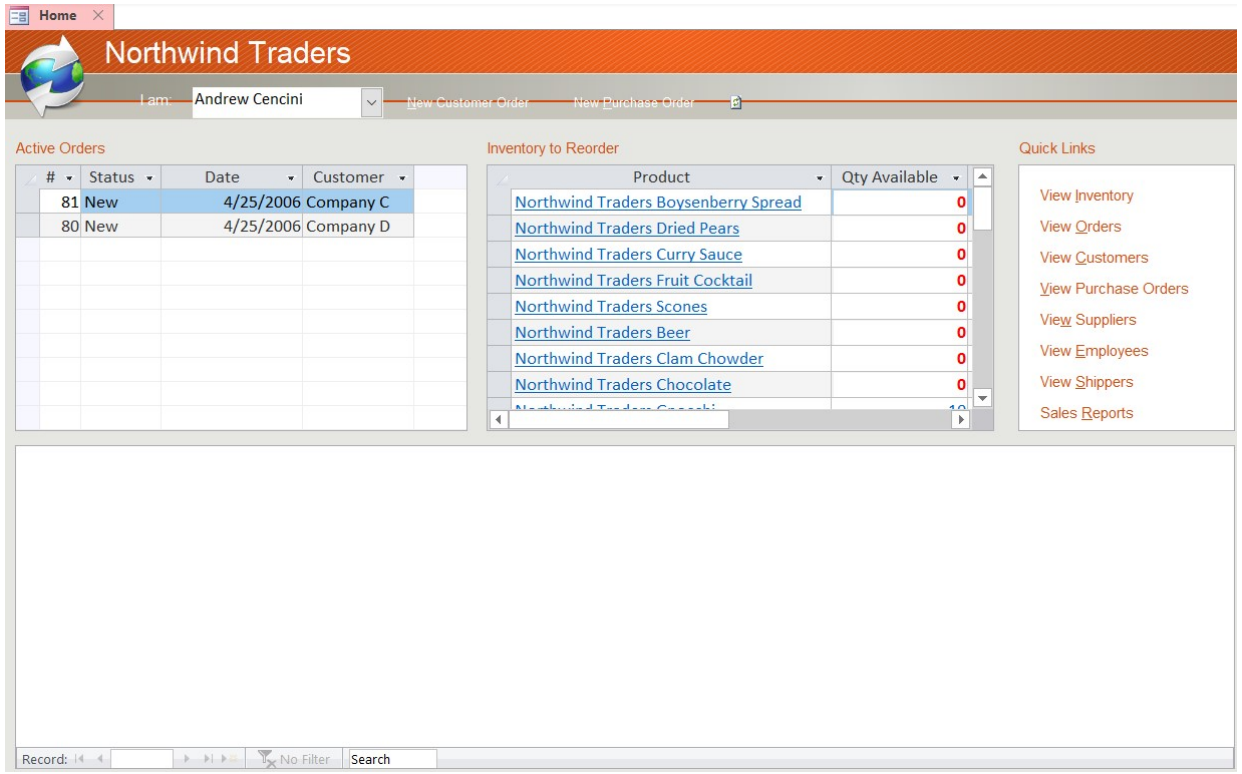


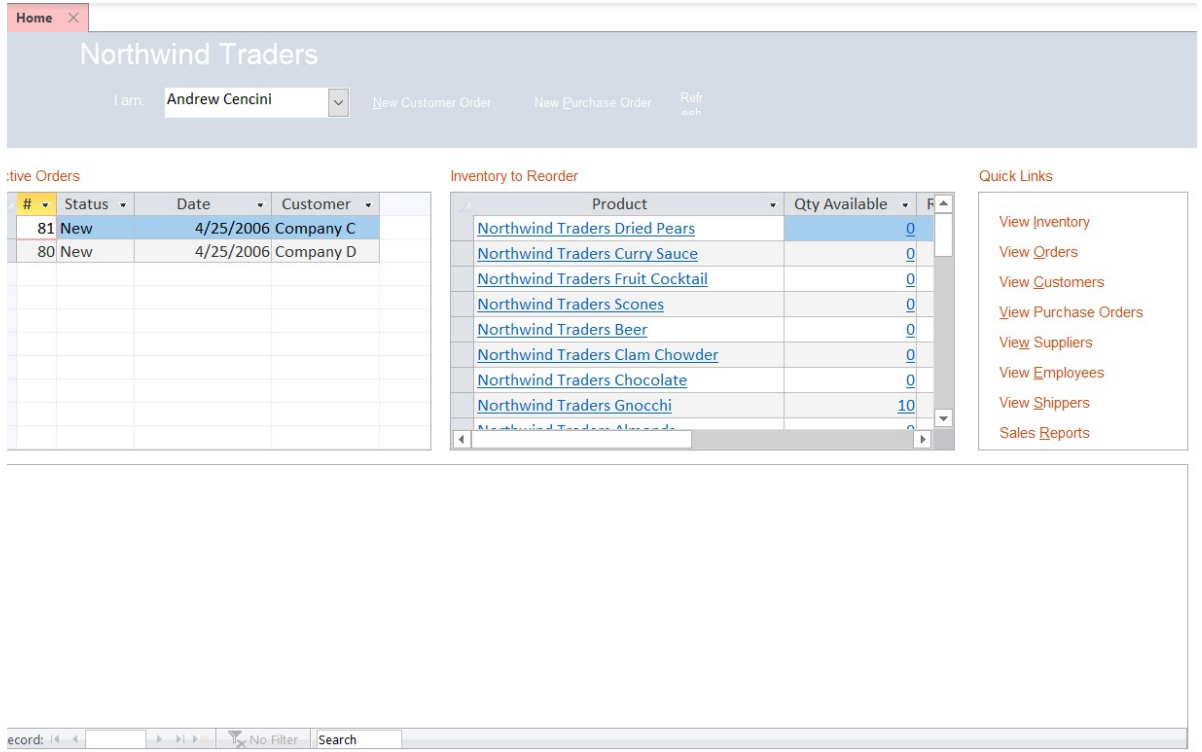
Completeness and Failures Overview





Human Insight & Opinion





Threats to Validity

Empirical Study

- 10 Projects
- 623 Components
- Many kinds of projects may not be represented

Multiple Version

- Different versions between origin and destination may suppose a problem

Undocumented Features Used

- Approach based on SaveAsText

Discussion

1. Source version control-based solution
   - Undocumented features
2. Context and performance
   - Selectiveness
3. What our validation does not validates
   - Text representation is incomplete

Contributions

- 1 Novel approach on modeling
- 4 Challenges – 3 Real industry requirements
- 1 Implementation
- 1 Validation approach

Do you have some **Microsoft Access or Visual Basic 6** Based Migration?

Let's work together! ☺

S. Bragagnolo

@sbragagnolo
santiago.bragagnolo@berger-levrault.com
santiago.bragagnolo@inria.fr

Contributions

- 1 Novel approach on modeling
- 4 Challenges – 3 Real industry requirements
- 1 Implementation
- 1 Validation approach

Do you have some **Microsoft Access or Visual Basic 6** Based Migration?

Let's work together! ☺

S. Bragagnolo

@sbragagnolo
santiago.bragagnolo@berger-levrault.com
santiago.bragagnolo@inria.fr

Results: Forms

Forms								
Projects	**Completeness**				**Failures**			
	Max	Min	Median	SDev	Max	Min	Median	SDev
Northwind	86.24	61.55	69	4.72	8.68	5.32	9	0.73
CUTLCOMP	–	–	–	–	–	–	–	–
CUTL	86.73	63.46	82	8.45	7.62	1.66	5	1.79
CRIR	74.73	73.06	74	0.96	8.31	8.18	9	0.064
CPDI	–	–	–	–	–	–	–	–
CHABIL	78.3	65.89	70.5	4.14	8.95	5.44	6	1.134
CDDE	100	98	100	0.83	0	0	0	0
CAUNIT	79.96	79.96	79.96	0	5.27	5.27	5.27	5.27
ACCUEIL	92.71	70.01	78	6.22	15.71	2.49	6	2.7
Access Examples	88.52	61.91	76	8.35	34.05	6.6	9	7.25

Results: Reports

Reports								
Projects	**Completeness**				**Failures**			
	Max	Min	Median	SDev	Max	Min	Median	SDev
Northwind	71.55	65.97	70	1.5	16.57	11.33	16	1.3
CUTLCOMP	–	–	–	–	–	–	–	–
CUTL	–	–	–	–	–	–	–	–
CRIR	73.13	71	73	1.5	20.64	15.64	18.5	3.54
CPDI	–	–	–	–	–	–	–	–
CHABIL	71.21	71.21	71.21	0	13.87	13.87	13.87	0
CDDE	–	–	–	–	–	–	–	–
CAUNIT	–	–	–	–	–	–	–	–
ACCUEIL	74.57	73.34	74	0.53	14.16	13.87	14	0.13
Access Examples	90.14	66.1	72.5	7.48	18.15	13.87	16	1.51

Results: Tables

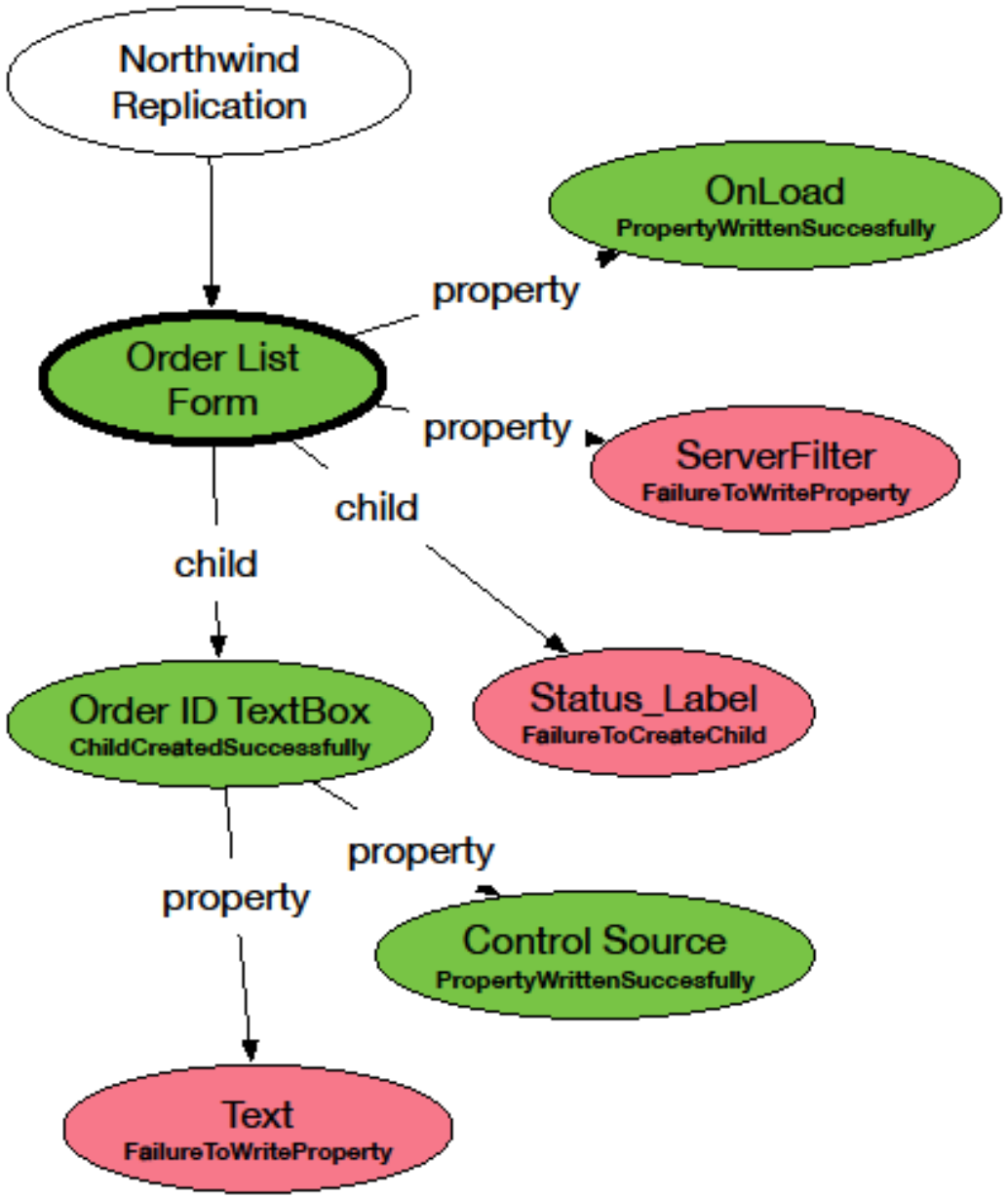
Tables								
Projects	**Completeness**				**Failures**			
	Max	Min	Median	SDev	Max	Min	Median	SDev
Northwind	100	0	100	23.72	22	0	0	5.14
CUTLCOMP	100	0	100	27.62	0	0	0	0
CUTL	100	0	98	5041	10	0	0	46.43
CRIR	100	0	99	46.04	100	0	0	44.42
CPDI	100	0	100	27.62	0	0	0	0
CHABIL	100	0	0	50.65	100	0	0	50.38
CDDE	–	–	–	–	–	–	–	–
CAUNIT	100	98	100	0.75	0	0	0	0
ACCUEIL	100	0	0	49.6	100	0	100	50.63
Access Examples	100	0	99.5	32.34	90	0	0	17.65

Results: Queries

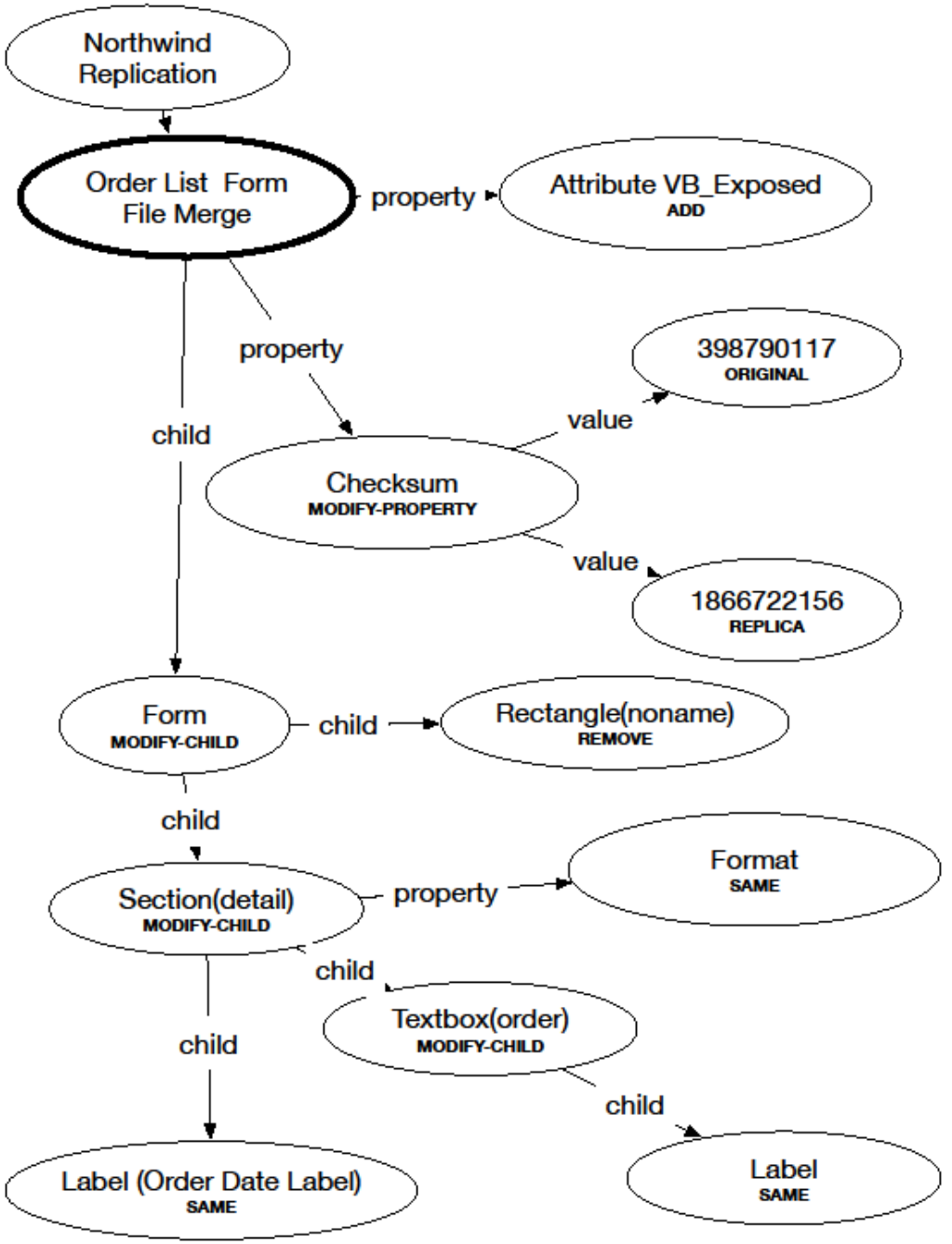
Queries								
Projects	**Completeness**				**Failures**			
	Max	Min	Median	SDev	Max	Min	Median	SDev
Northwind	100	100	100	0	15	0	0	3.05
CUTLCOMP	–	–	–	–	–	–	–	–
CUTL	100	100	100	0	0	0	0	0
CRIR	100	88.46	100	2.88	0	0	0	0
CPDI	–	–	–	–	–	–	–	–
CHABIL	100	93.49	100	1.25	0	0	0	0
CDDE	–	–	–	–	–	–	–	–
CAUNIT	–	–	–	–	–	–	–	–
ACCUEIL	100	95.76	100	1.17	0	0	0	0
Access Examples	100	0	100	25.81	100	0	0	34.15

Weighting Errors

$$F(o) = \begin{cases} 1 & o \in \{Failure\} \\ 0 & o \in \{Success\} \wedge |c_o| = 0 \wedge |p_o| = 0 \\ (0.9\dfrac{\sum_{i=1}^{|c_o|} F(c_o i)}{|c_o|} + 0.1\dfrac{\sum_{i=1}^{|p_o|} F(p_o i)}{|p_o|})0.5 & o \in \{Success\} \wedge |c_o| > 0 \\ \dfrac{\sum_{i=1}^{|p_o|} F(p_o i)}{|p_o|} & o \in \{Success\} \wedge |c_o| = 0 \end{cases}$$



Weighting Differences

$$Completeness(u) = (1 - M(u)) * 100)$$

$$M(u) = \begin{cases} 1 & u \in \{Add, Remove\} \\ 0 & u \in \{Same\} \\ (0.9\dfrac{\sum_{i=1}^{|c_u|} M(c_u i)}{|c_u|} + 0.1\dfrac{\sum_{i=1}^{|p_|} M(p_u i)}{|p_u|})0.5 & u \in \{ChildModif\} \wedge |c_u| > 0 \\ \dfrac{\sum_{i=1}^{|p_u|} M(p_u i)}{|p_u|} & u \in \{ChildModif\} \wedge |c_u| = 0 \\ u_r - u_o & u \in \{PropertyModif\} \wedge u_r, u_o \\ & \in \{Native\ type\} \end{cases}$$