



DH01: Accelerate Clinical Development through Management of Data Quality

Dave Smith, SAS UK

Clinical Data Quality

Potential harm caused by poor data quality

Scope of the problem

Solution Components

- Data Governance
- Technology

Benefits of good data quality

Conclusions

Potential harm caused by poor data quality

Warning letters

- Lack of understanding of clinical study assets can lead to studies being omitted from summaries, which can lead to FDA Warning Letters

Time to Integrated Summary

- Time taken to prepare integrated summaries will increase when attempts are made to combine data that does not match

Trust

- Trust in data is reduced, leading to status reports being devalued and ultimately ignored

Scope of the problem

Data

- Data quality is a problem because non-standardised items can disguise the true contents of systems

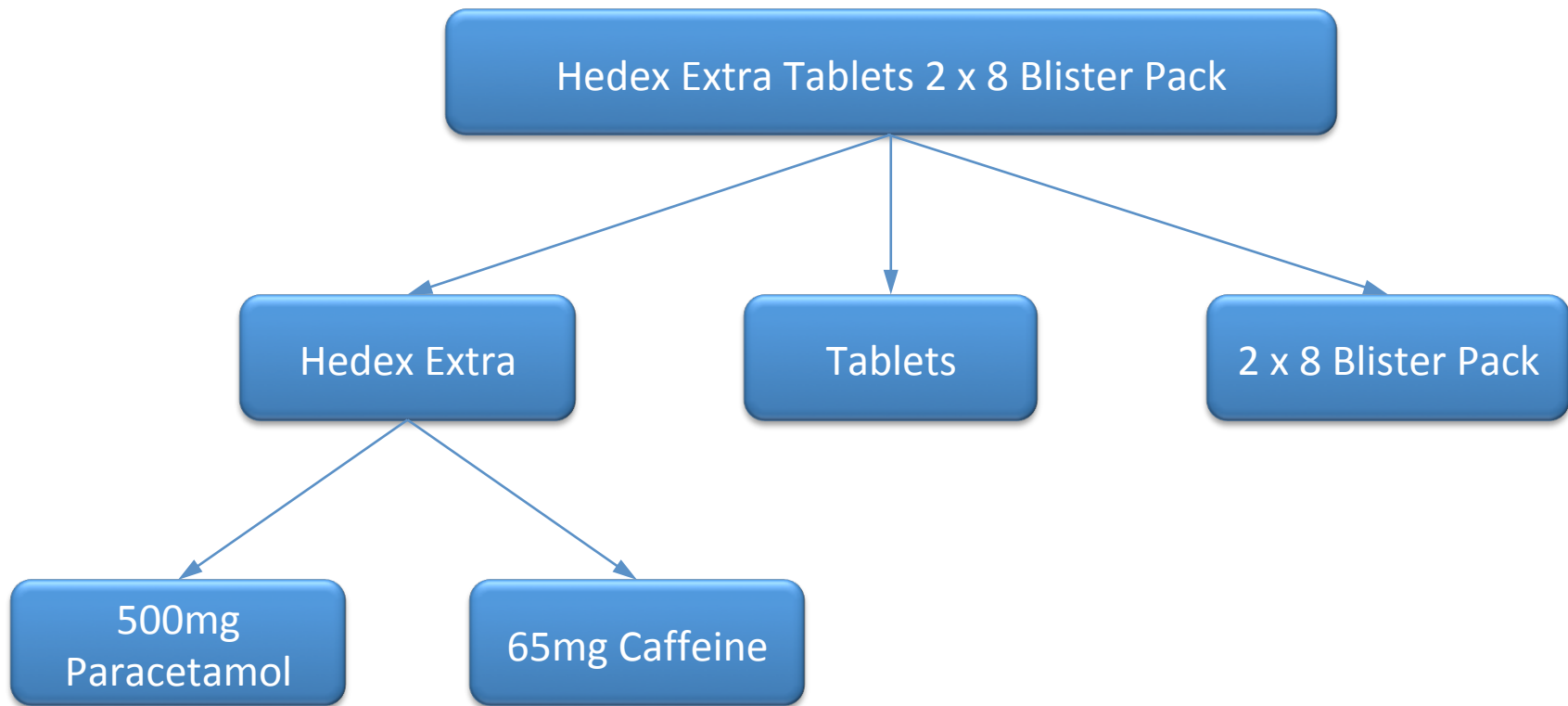
Metadata

- Metadata quality is important because it can lead to failures to match and therefore incomplete or misleading information

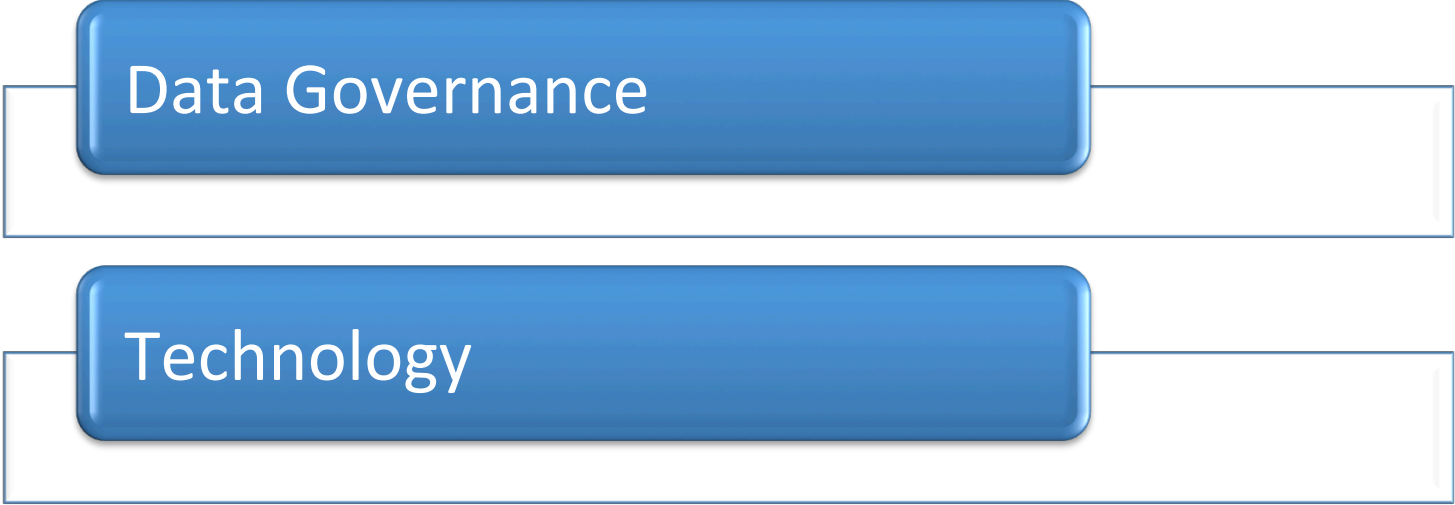
Multiple systems

- The problem is not confined to clinical databases, but to all clinical systems such as CTMS, safety, portfolio management, project management and financial systems

Example



Solution Components



Data Governance

Technology

Data Governance

A business framework to ensure data assets are managed and controlled

Requires People, Processes and Technology

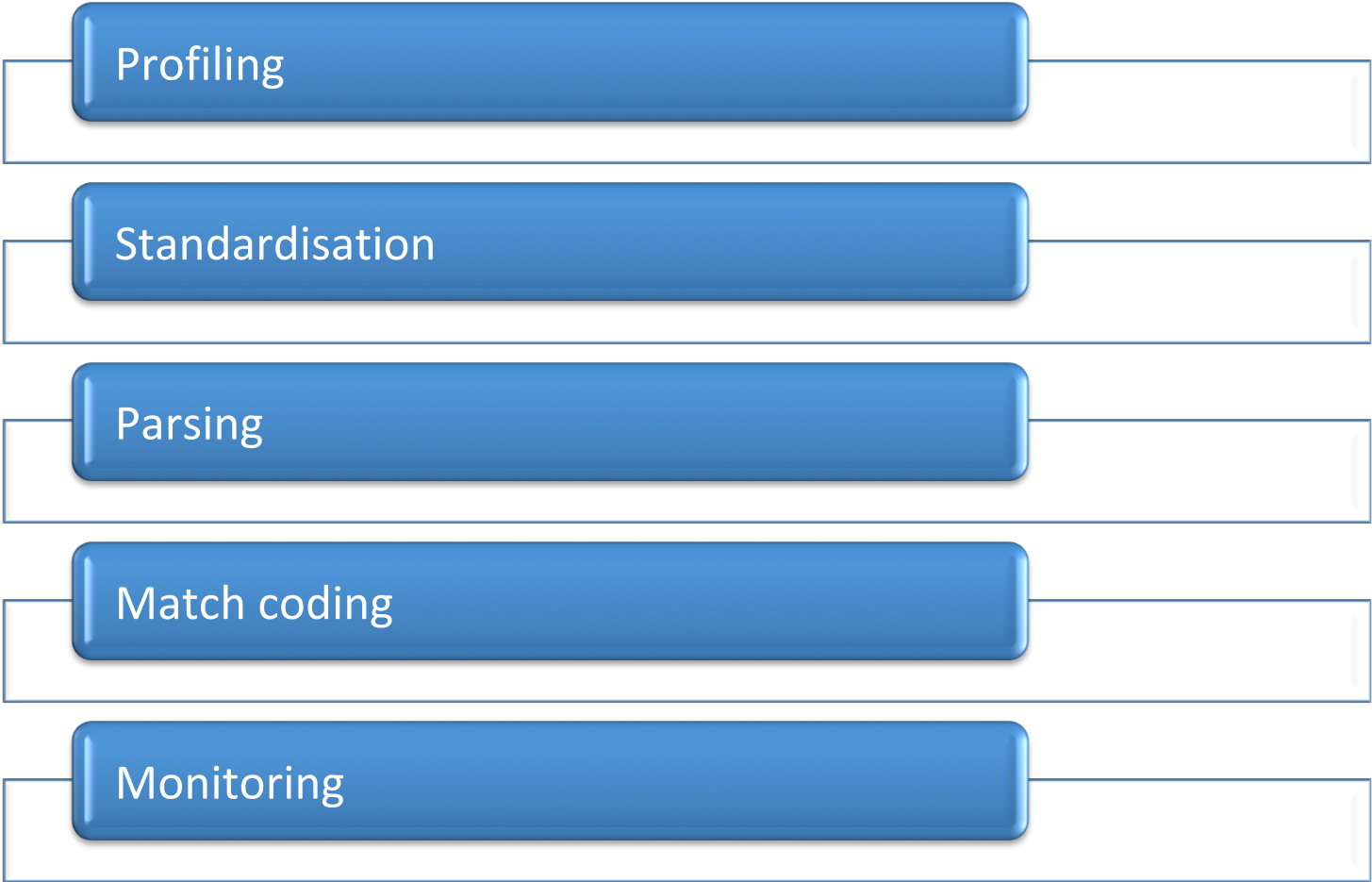
Only effective with high-level sponsorship

Usually involves Data Stewards to manage data assets and develop and enforce procedures

Various Technology Components

- Master Quality Management
- Data Quality Monitoring
- Data Quality Business Rules

DataFlux Technology Components



Profiling

Standardisation

Parsing

Match coding

Monitoring

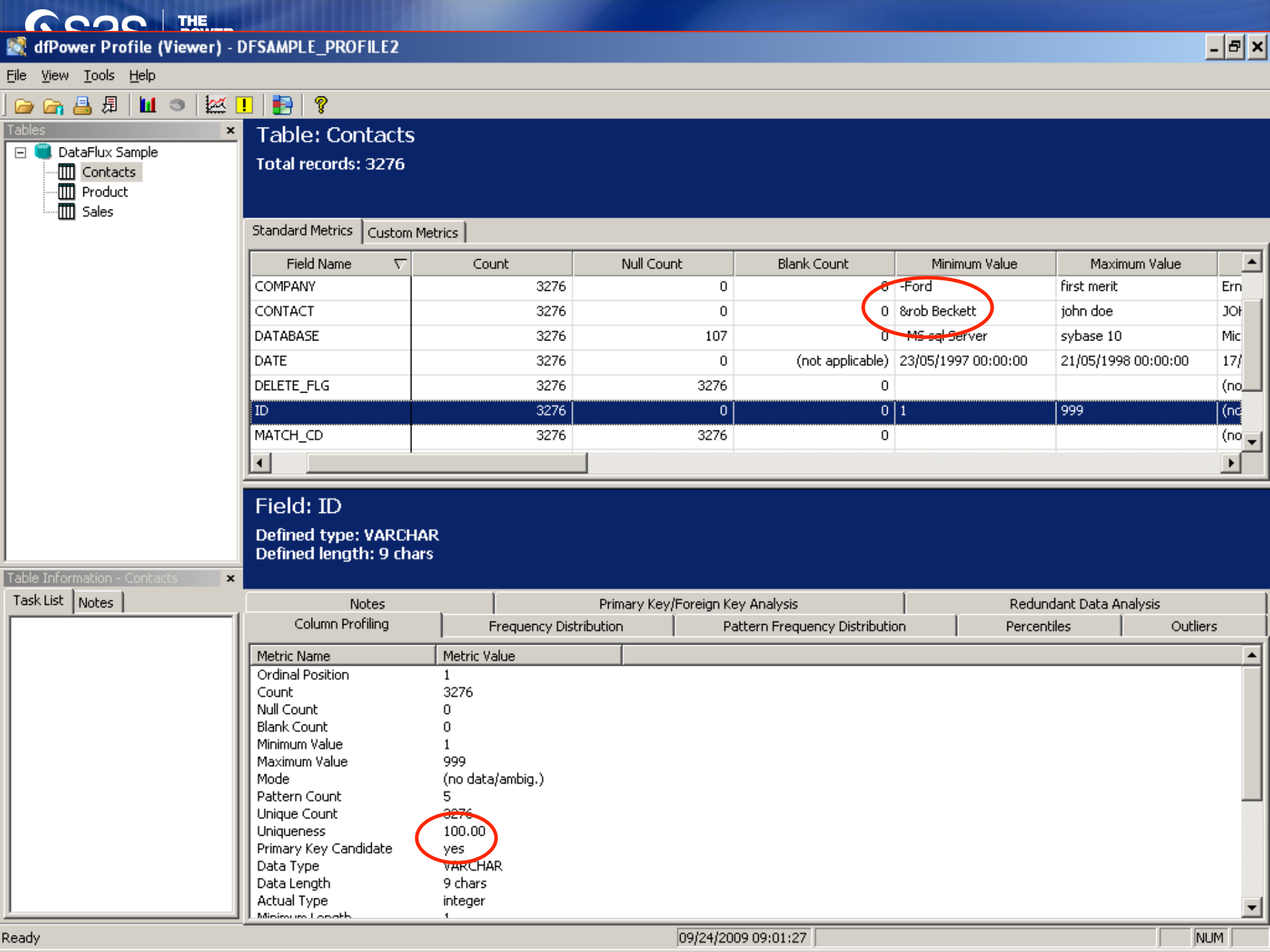
Profiling

Assess the structure and content of data

Quickly identify issues

- Missing data
- Non-unique data
- Failure to meet required patterns

Key analysis



Standardisation

Create a scheme to map raw data to standard terms and spellings

Can be manual, can use grammars and dictionaries

Can generate SQL code to apply the scheme

File Edit View Report Tools Help



Report	Total count: 58
--------	-----------------

Group	Value	Count
1	Acyclovir	1
2	Alopurinol	1
2	Allopurinol	1
3	Amlodipine Besylate	1
4	Aciclovir	1
5	Azithromycin	1
6	Atenolol	1
6	Atenolol	1
7	Atorvastatin	1
8	Bisacodyl	1
9	Diclofenac	1
10	Domperidone	1
11	Domperignon	1
12	Enalapril Maleate	1
13	Erythromycin Estolate	1
14	Ethamsylate	1
15	Fluconazole	1
16	Fexofenadine	1
17	Glibenclamide	1
18	Gliclazide	1
19	Glipizide	1
20	Gatifloxacin	1
21	Hyosine Butyl Bromide	1
22	Isosorbide Mononitrate	1
23	Chloroquine	1
24	Ketoconazole	1
25	Loratidine	1
26	Lisinopril	1
27	Lisinopril (Anhydrous)	1
28	Losartan Potassium	1
29	Levofloxacin	1

Standard: Acyclovir

Add To Scheme

Scheme	Entries: 7
--------	------------

[illegible]

Add...

Edit...

Delete

Options...

Parsing

Break a string of text into its constituent parts

Need to understand the data being analysed

Create a vocabulary of known words (can be imported)

Create a grammar of structures

Manage delimiters, Stop words

Probabilistic approach – may be more than one solution

dfPower Architect - Parse Pharma Products

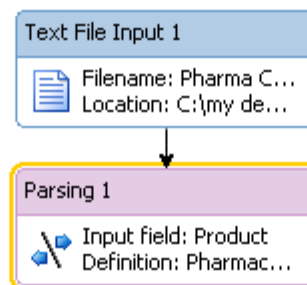
File Edit View Insert Job Tools Help

Toolbox Details 100% Insert Node

Toolbox

Nodes Accelerators

- Utilities
- Profiling
- Quality
 - Gender Analysis
 - Gender Analysis (Parsed)
 - Identification Analysis
 - Parsing
 - Standardization
 - Standardization (Parsed)
 - Change Case
 - Locale Guessing

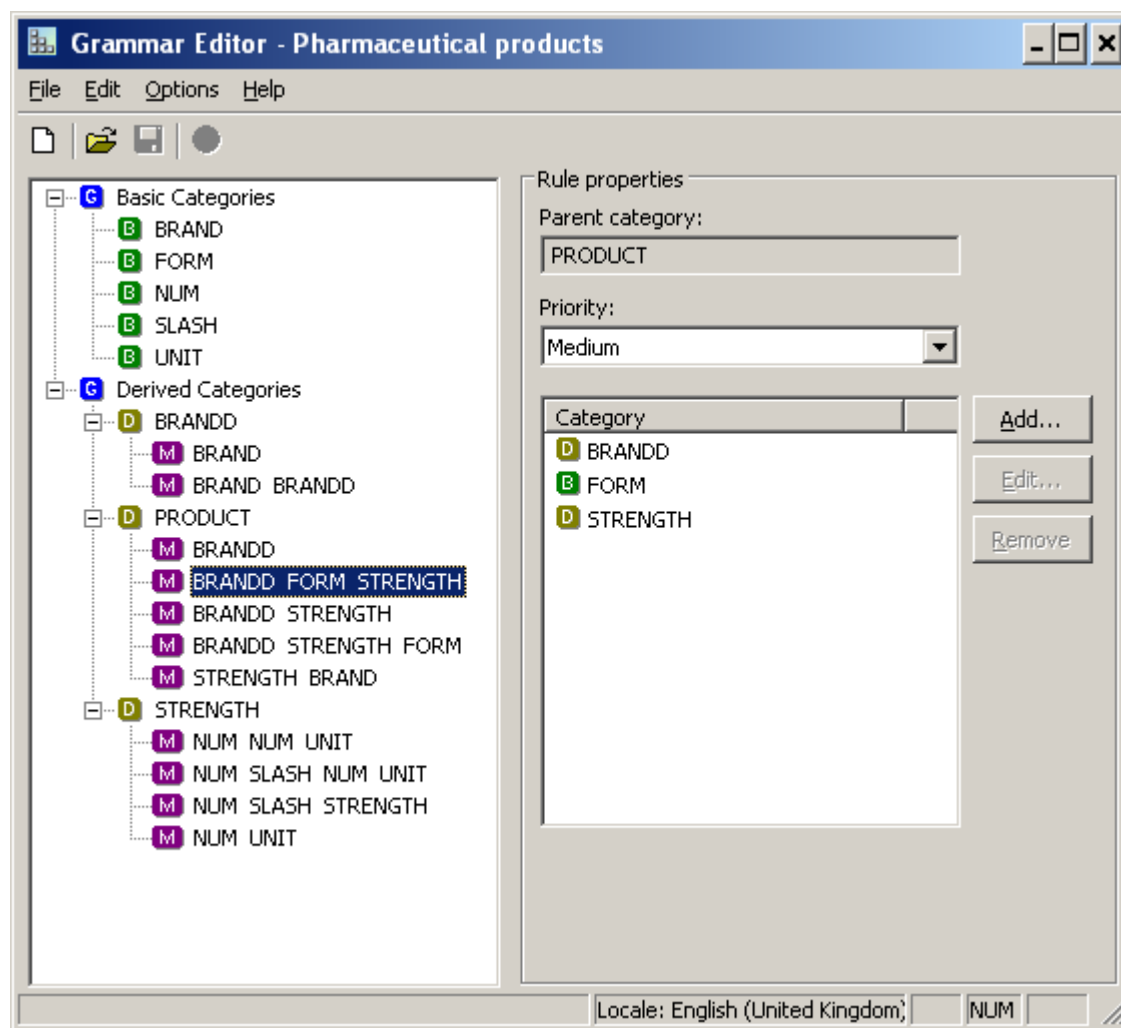


1 - Page 1

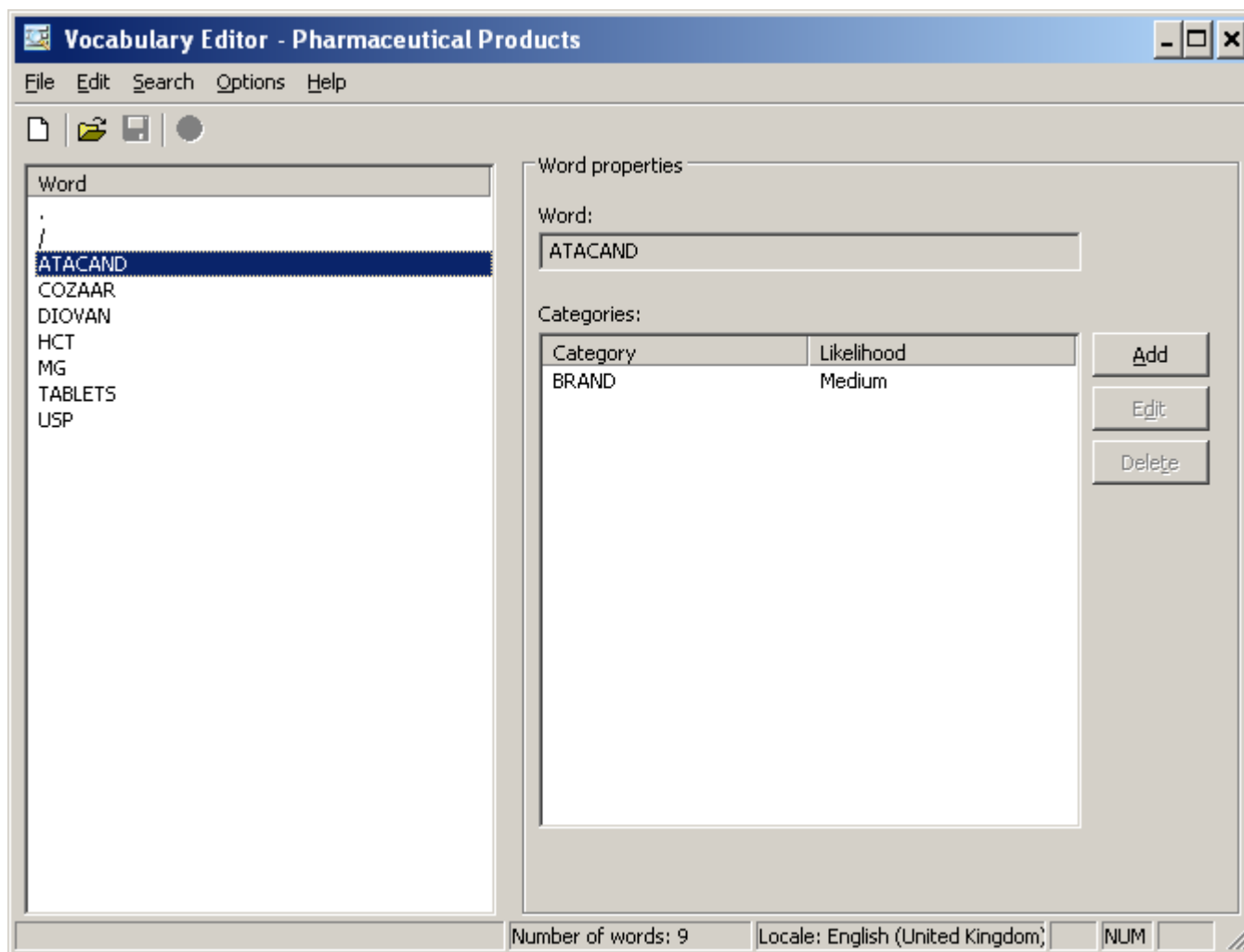
Information Preview Log

Product	Brand	Form	Strength						
25 mg COZAAR	COZAAR		25 mg						
ATACAND tablets 4 mg	ATACAND	tablets	4 mg						
ATACAND tablets 8 mg	ATACAND	tablets	8 mg						
ATACAND 32 mg	ATACAND		32 mg						
ATACAND tablets 16 mg	ATACAND	tablets	16 mg						
ATACAND tablets	ATACAND	tablets							
Diovan HCT USP 80/12.5 mg	Diovan		80 /12 .5 mg						
Diovan HCT USP 160/12.5 mg	Diovan		160 /12 .5 mg						

Parsing: Grammar



Parsing: Vocabulary



Parsing: Test

Parse Test - Pharmaceutical Product

Input text:

Atacand Tablets 8 mg

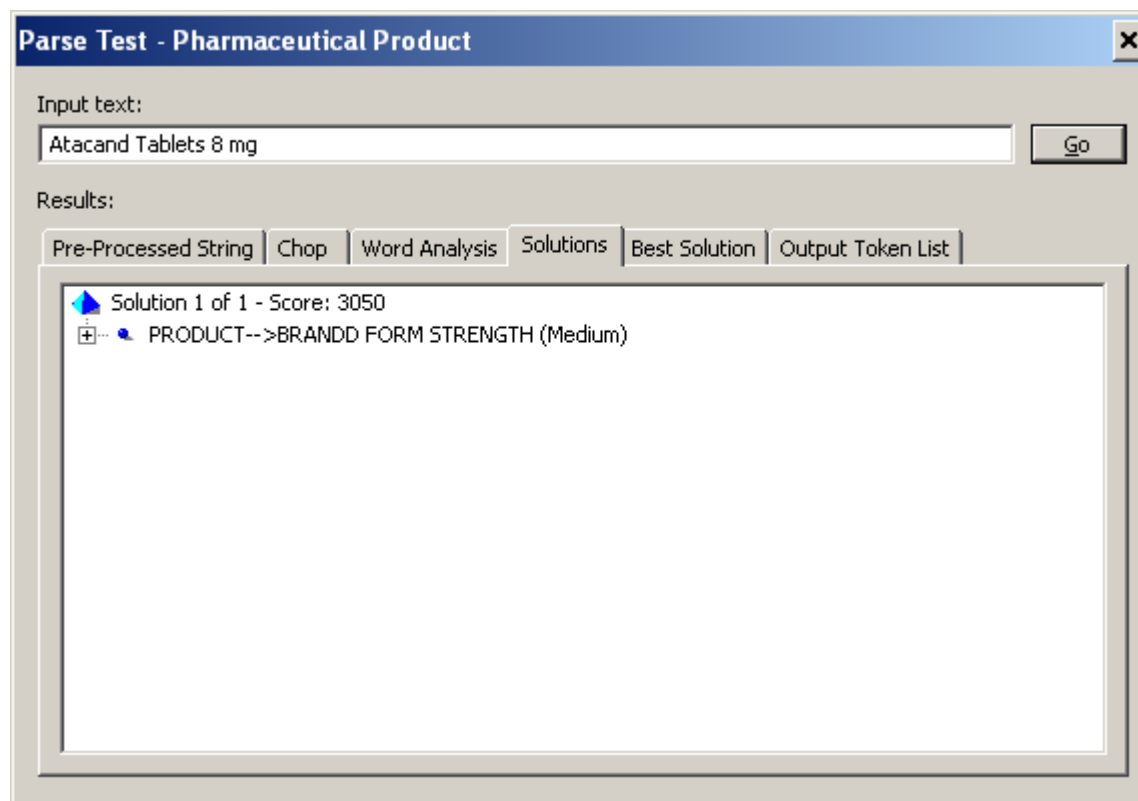
Go

Results:

Pre-Processed String Chop Word Analysis Solutions Best Solution Output Token List

Word	Normalized Word	Category	Likelihood
Atacand	ATACAND	BRAND	Medium
Tablets	TABLETS	FORM	Medium
8	8	NUM	Medium
mg	MG	UNIT	Medium

Parsing: Test



Parsing: Test

Parse Test - Pharmaceutical Product [X]

Input text:

Atacand Tablets 8 mg [Go]

Results:

Pre-Processed String | Chop | Word Analysis | Solutions | Best Solution | Output Token List

Token	Value
Compound	Atacand
Form	Tablets
Strength	8 mg

Match coding

Create an intelligent code to contain the essence of a text string

Use a sensitivity to determine how much information is passed to the match code

Enables fuzzy matching of data

dfPower Architect - Parse Pharma Products

File Edit View Insert Job Tools Help

Nodes

Accelerators

Data Inputs

Data Outputs

Utilities

Profiling

Quality

Integration

Match Codes

Match Codes (Parsed)

Filename: Pharma C...

Location: C:\my de...

Parsing 1

Input field: Product

Definition: Pharmac...

Match Codes 1

Sensitivity: 75

1 - Page 1

Information Preview Log

Product	Product_MatchCode							
COZAAR 25 mg tablets	34YH5BFY~M\$\$\$\$\$							
COZAAR 50 mg	34Y50BFY\$\$\$\$\$\$							
COZAAR tablets 100 mg	34Y~MW~Z00\$\$\$\$\$							
25 mg COZAAR	H5BFY34Y\$\$\$\$\$\$							
ATACAND tablets 4 mg	8~3P8~MW~S\$\$\$\$\$							
ATACAND tablets 8 mg	8~3P8~MW~D\$\$\$\$\$							
ATACAND 32 mg	8~3P8KHBFY\$\$\$\$\$							
ATACAND tablets 16 mg	8~3P8~MW~Z\$\$\$\$\$							
ATACAND tablets	8~3P8~MW~\$\$\$\$\$\$							
Diovan HCT USP 80/12.5 mg	8VP3~4ND0Z\$\$\$\$\$							
Diovan HCT USP 160/12.5 mg	8VP3~4NZ60\$\$\$\$\$	←						
Diovan HCT USP 160/25 mg tablets	8VP3~4NZ60\$\$\$\$\$	←						
Diovan HCT USP 320/12.5 mg tablets	8VP3~4NKH0\$\$\$\$\$							

Ready Selected: 1 node NUM

dfPower Architect - Parse Pharma Products

File Edit View Insert Job Tools Help

Toolbox

Nodes Accelerators

Data Inputs
Data Outputs
Utilities
Profiling
Quality
Integration
Match Codes
Match Codes (Parsed)

Filename: Pharma C...
Location: C:\my de...

Parsing 1

Input field: Product
Definition: Pharmac...

Match Codes 1

Sensitivity: 50

1 - Page 1

Information Preview Log

Product	Product_MatchCode								
COZAAR 25 mg tablets	34YH5B\$\$\$\$\$\$\$\$								
COZAAR 50 mg	34Y50B\$\$\$\$\$\$\$\$								
COZAAR tablets 100 mg	34Y~MW\$\$\$\$\$\$\$\$								
25 mg COZAAR	H5BFY3\$\$\$\$\$\$\$\$								
ATACAND tablets 4 mg	8~3P8~\$\$\$\$\$\$\$\$	←							
ATACAND tablets 8 mg	8~3P8~\$\$\$\$\$\$\$\$	←							
ATACAND 32 mg	8~3P8K\$\$\$\$\$\$\$\$								
ATACAND tablets 16 mg	8~3P8~\$\$\$\$\$\$\$\$	←							
ATACAND tablets	8~3P8~\$\$\$\$\$\$\$\$	←							
Diovan HCT USP 80/12.5 mg	8VP3~4\$\$\$\$\$\$\$\$	←							
Diovan HCT USP 160/12.5 mg	8VP3~4\$\$\$\$\$\$\$\$	←							
Diovan HCT USP 160/25 mg tablets	8VP3~4\$\$\$\$\$\$\$\$	←							
Diovan HCT USP 320/12.5 mg tablets	8VP3~4\$\$\$\$\$\$\$\$	←							

Ready

Selected: 1 node

NUM

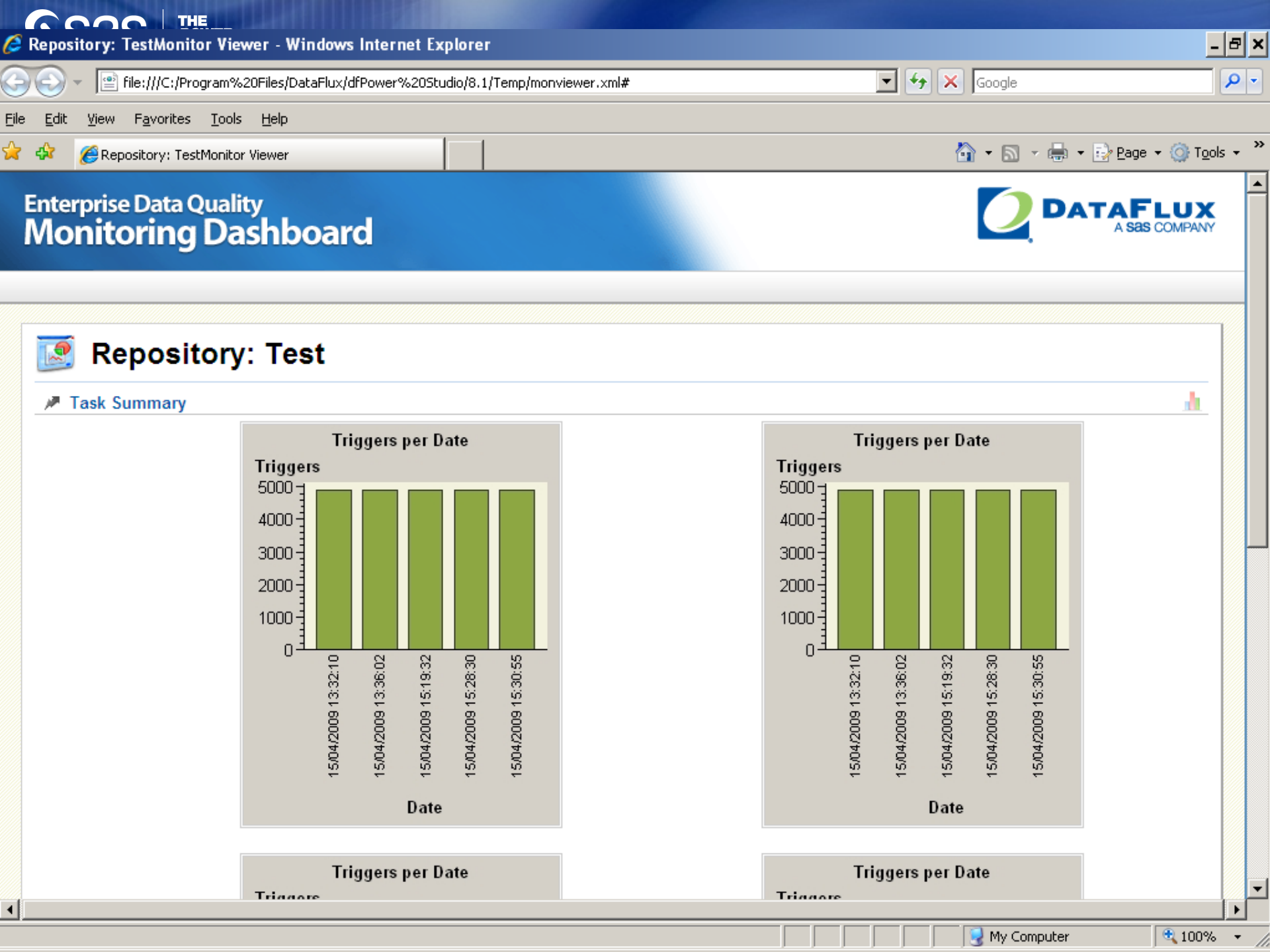
Monitoring

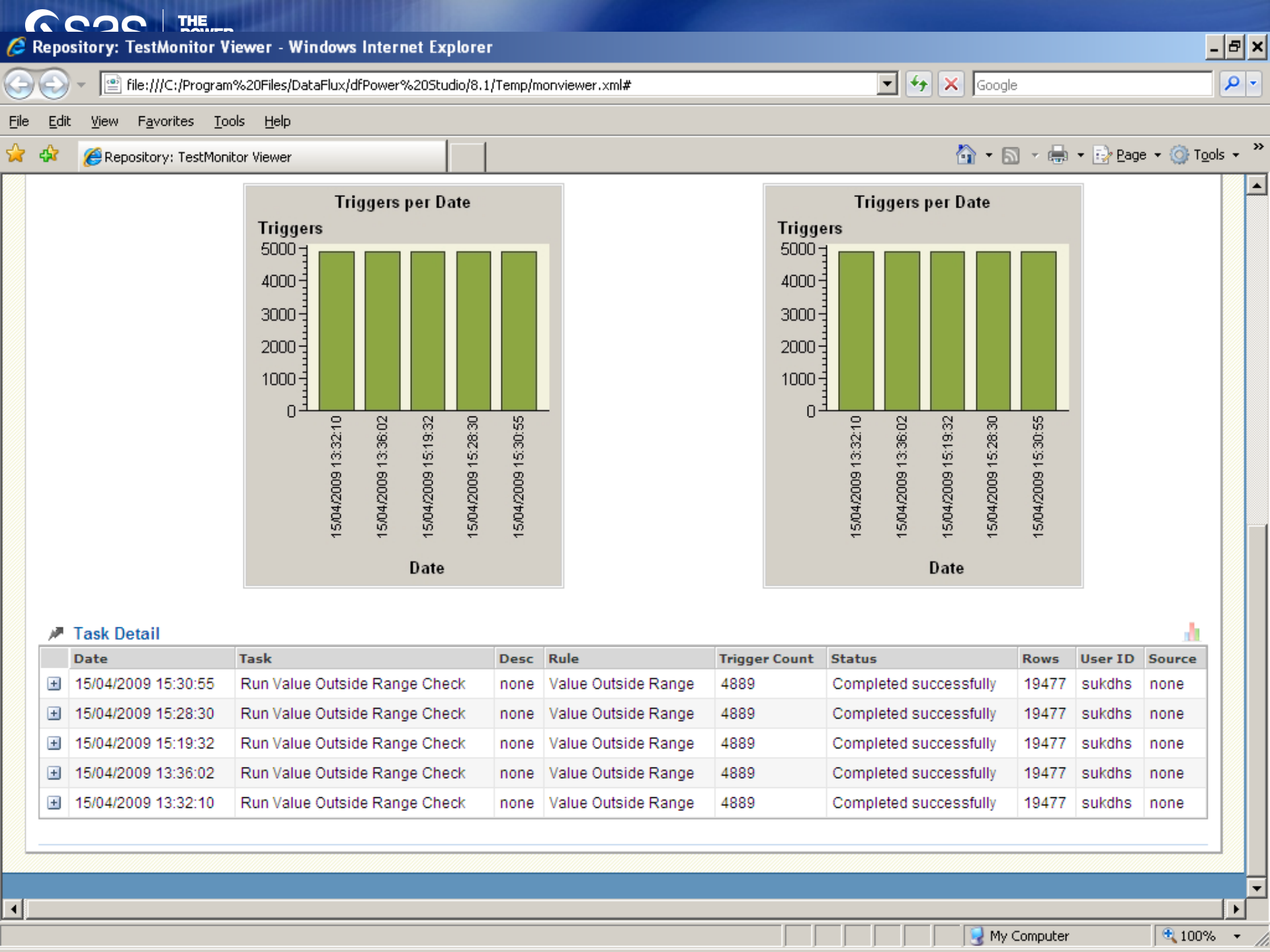
```
graph TD; A[Track data quality indicators over time] --> B[Apply business rules to create alerts if data quality drops below certain levels]; B --> C[Create data quality scorecards];
```

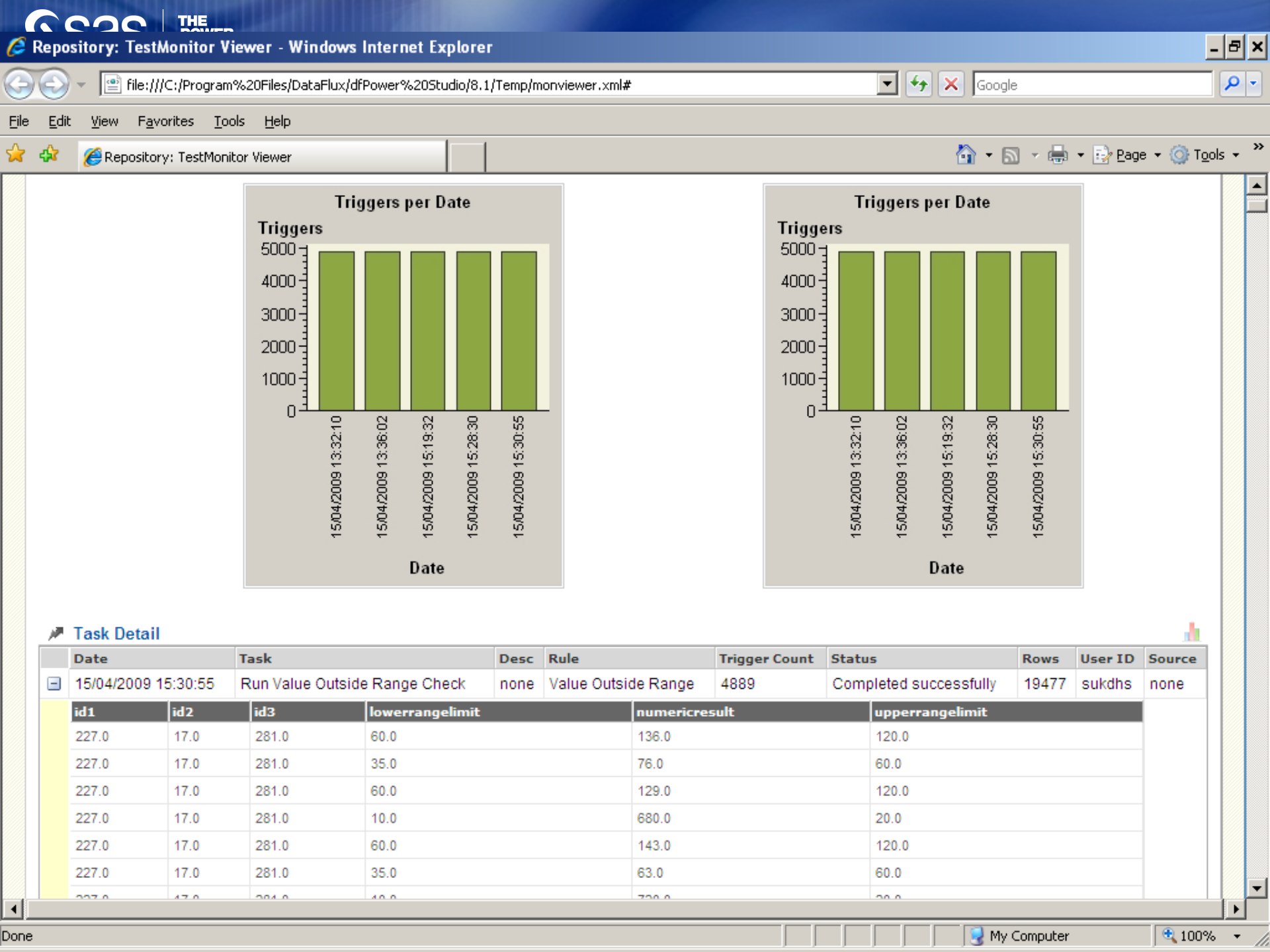
Track data quality indicators over time

Apply business rules to create alerts if data quality drops below certain levels

Create data quality scorecards







Benefits of good data quality

Lower risk of an FDA Warning Letter

Faster reporting process, especially in ISS/
ISE and SAE reconciliation

Better ability to answer regulatory
questions

Better overview of project portfolio

Conclusions

Data quality is an imperative

Direct costs

- Warning Letters
- Legal Action

Opportunity costs

- Failure to identify safety issues early enough
- Failure to identify resource issues early enough



Contact

Dave Smith
SAS UK
Wittington House
Henley Road
Medmenham
Marlow
Bucks
SL7 2EB

Work Phone: 01628 404379

Email: david.smith@suk.sas.com