Defining a Person Metadata Model to Improve Data Quality

8th February 2011

Ian Woodrow

Ian.woodrow@atosorigin.com

Atos Origin

CONSULTING > SOLUTIONS > OUTSOURCING

ADVANCE YOUR BUSINESS >>

Make things as simple as possible, but no simpler

Einstein

Essentially, all models are wrong, but some are useful

Box & Draper

If you can't measure it, you can't manage it…

Various known

ADVANCE YOUR BUSINESS ≫

Your Presenter

Role – Service/Project Manager and Information Analyst

Projects

» Data Migration and Cleansing

» Data Standards Implementations

» Sales and Account Development

» Business and Data Analysts

» Accruals Accounting

» Ian.woodrow@atosorigin.com

Employment

» National Audit Office (E&AD)

» Capgemini

» Freelance

» Atos Origin

Recent Training

» TOGAF9 Enterprise Data Architect

» Prince2 Practitioner

» Value Analysis.

ADVANCE YOUR BUSINESS »

Introduction

» Information Management Service

» Corporate Data Model

» Person Metadata Model

» Data Standards

» Data Profile Reporting

» Questions and Queries

» Contact Me.

ADVANCE YOUR BUSINESS ≫

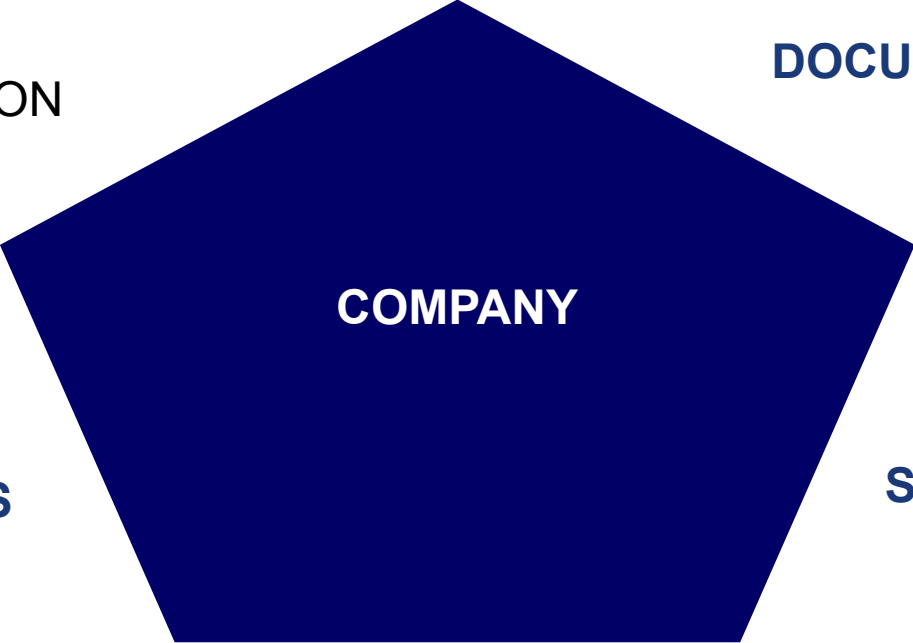# Information Management Service

ADVANCE YOUR BUSINESS »
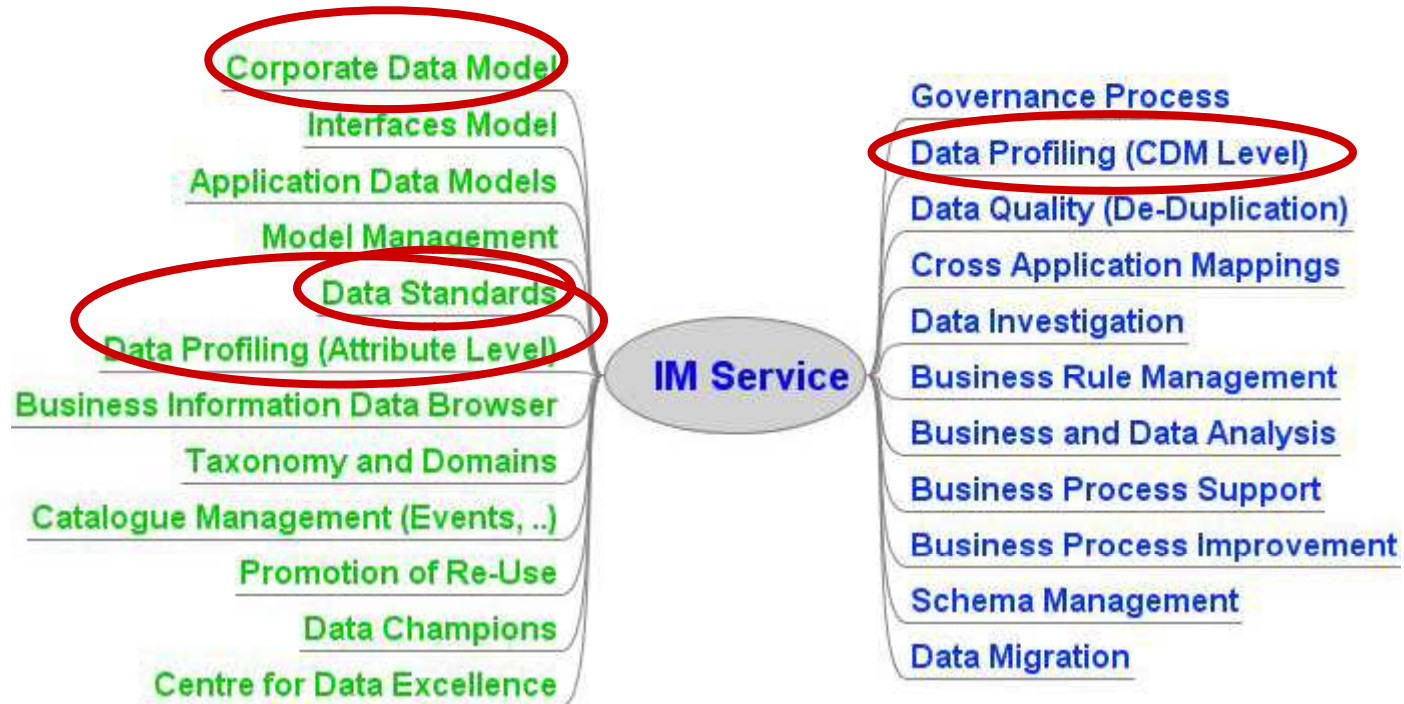
Corporate Data Model

PEOPLE/
ORGANISATION

DOCUMENTS

COMPANY

EVENTS

SERVICES

LOCATIONS

ADVANCE YOUR BUSINESS >>

Applications

» Registry

» Membership

» Benefit

» Biometrics

» Resources

» Overseas

ADVANCE YOUR BUSINESS »
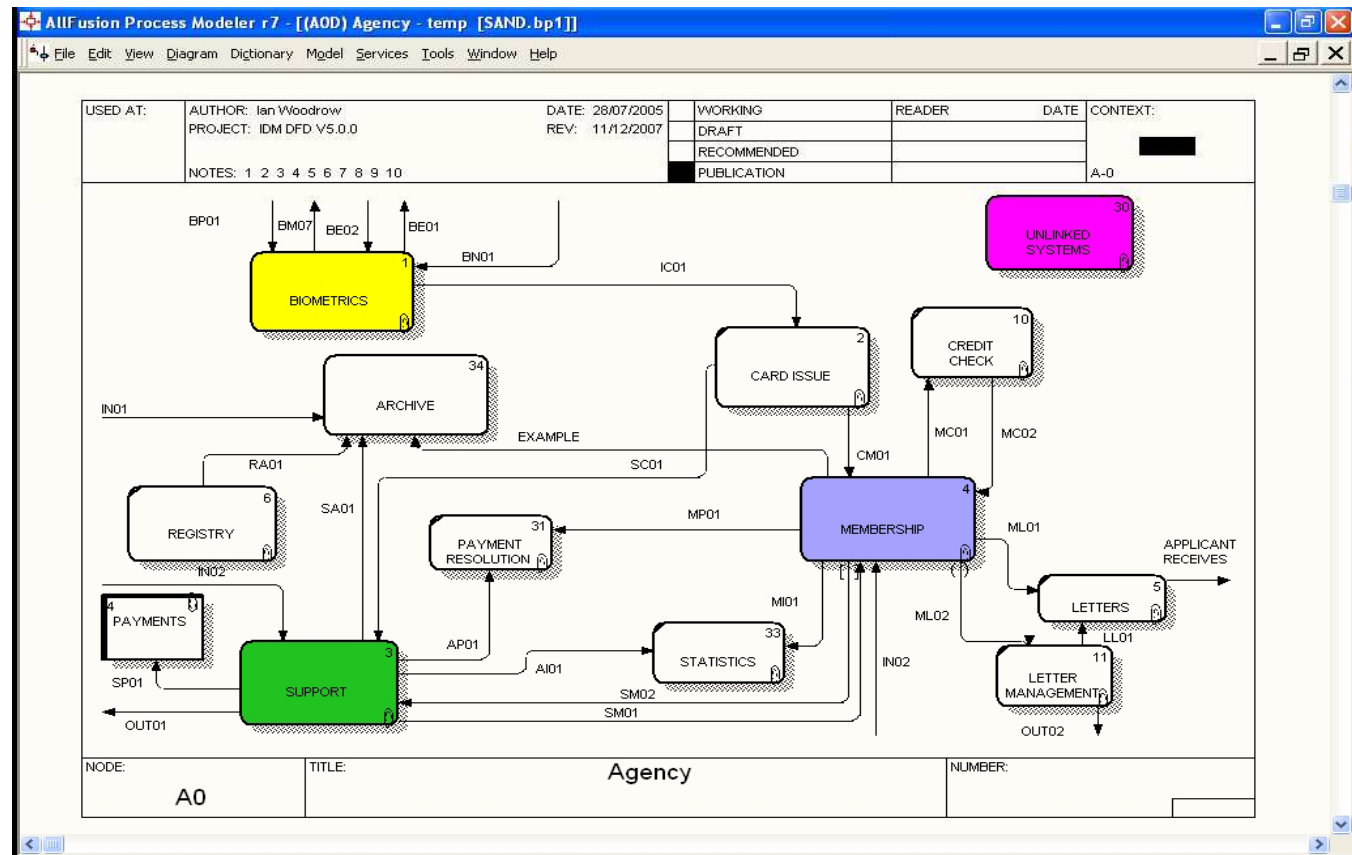
Data Standards

Select Target Attribute

» Identify Local Application Alternatives
» Look for Commonality
» Review Government Standards
» Review UK National Standards
» Review Other Government Standards (e.g. NIST)
» Review International Standards
» Assess Applicability
» Decide Solution
» Confirm Solution
» Publish Solution

Target Attribute Metadata

» Datatype
» Length
» Values Lists

» Description
» Constraints
» Defaults

ADVANCE YOUR BUSINESS ≫

Tooling: Best of Breed Solution

PRODUCTS

Erwin Data Modeller
Model Manager
Process Modeller (BPWin)

Discovery

CHANNELS

INTRANET

TECH DOC LIBRARY

ADVANCE YOUR BUSINESS »

ADVANCE YOUR BUSINESS »
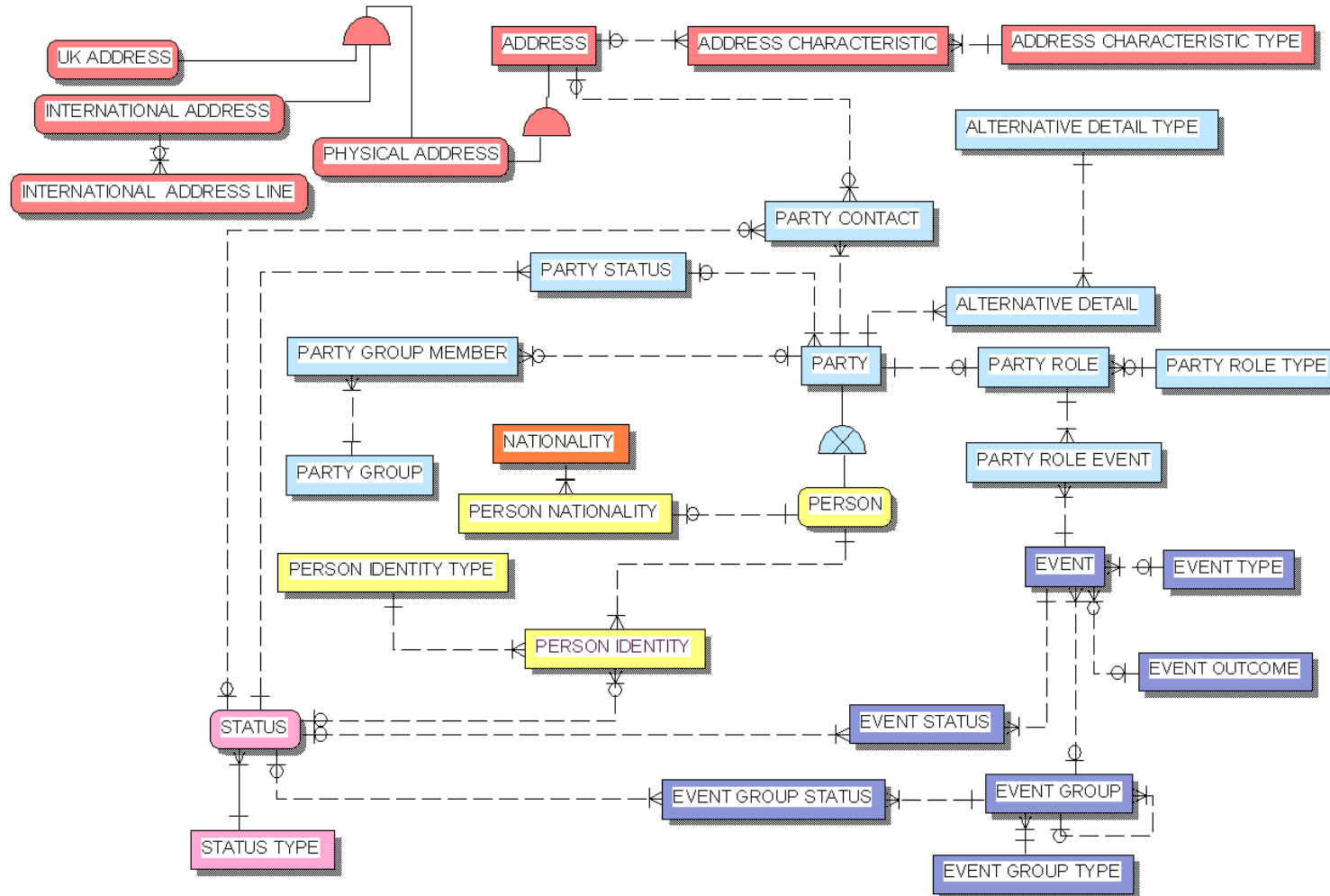
Cross-reference CDM to Applications

ATTRIBUTE	CDM	REGISTRY	MEMBERSHIP
Full Name			
Family Name	PERSON.FAMILY NAME	REGISTER.PRINCIPAL NAME	PEOPLE.SURNAME
Given Names			
Date of Birth			
Gender			
Nationality(s)			
Language (s)			
Alternative Details			
Organisation Start Date			
Organisation Reference			
International ID (Passport)			
NI Number			
First Line Address			
Postcode			
Telephone Number(s)			

ADVANCE YOUR BUSINESS »

Publishing the Results

Defining a Person Metadata Model to Improve Data Quality

ADVANCE YOUR BUSINESS >>

Person Metadata Model – Vision

» To Create a Whole Customer View
  » Person Identity
  » Key Events
  » Sufficient for Identity Resolution

» Organisation Facts
  » Siloed Data Sources
  » Several Suppliers
  » Considering:
    » Master Data Management
    » Service Oriented Architecture

» Technical Facts
  » Different Databases
    -SQL Server
    -Oracle
    -Access
  » No ETL Tool on server

ADVANCE YOUR BUSINESS »

Benefits of Person Metadata Model Approach

» Address data shortfall in existing applications (enrichment)

» Form basis for new application data models

» Bench mark internal developments and suppliers offerings

» Baseline for Profiling
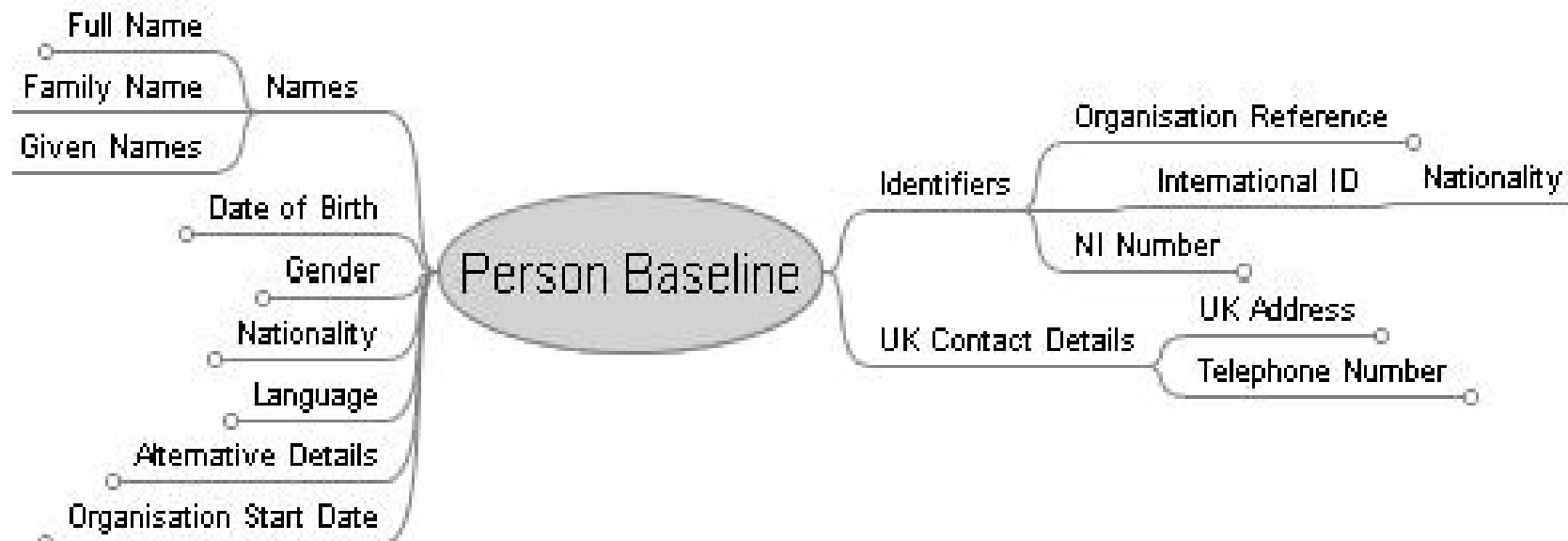
ADVANCE YOUR BUSINESS ≫

Person Metadata Model

ADVANCE YOUR BUSINESS >>

Person Baseline Analysis

Defining a Person Metadata Model to Improve Data Quality

ADVANCE YOUR BUSINESS >>

CDM Attribute Scoring

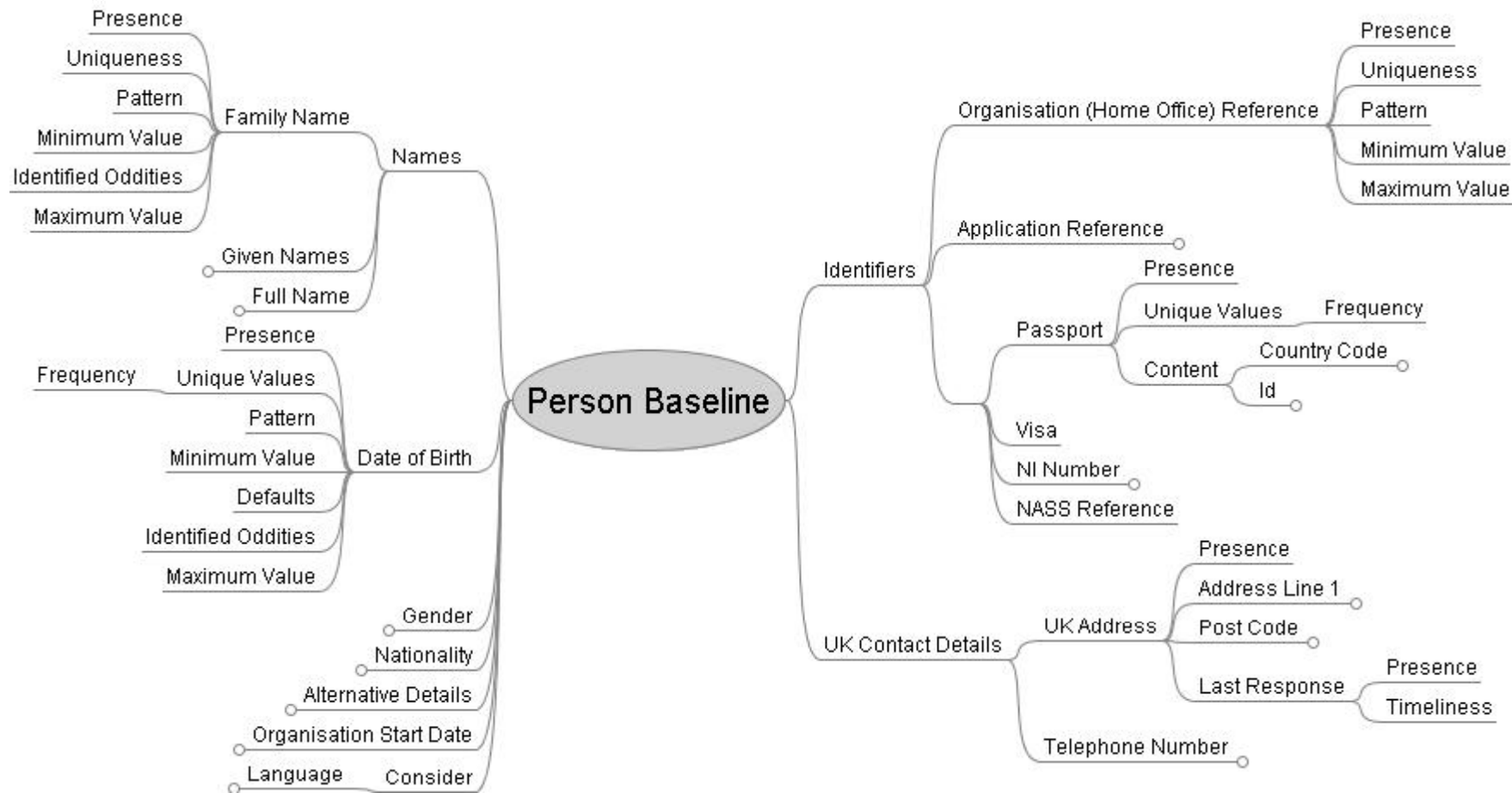|	MODEL COMPARISON ATTRIBUTE	CDM Maximum	Options	
1	MODEL COMPARISON ATTRIBUTE	CDM Maximum	Options	
2	Full Name		2	
3	Family Name	6		
4	Given Names	6		
5	Date of Birth	6		2 for each part
6	Gender	3		
7	Nationality(s)	6		3 + 3 for dual Nat
8	Language (s)	4		2 + 2 for multiple Lang
9	Alternative Details		4	2 for each name, only 2 if held at a composite part (maximum of 2)
10	Organisation Start Date	2		
11	Internal Reference Available externally	2		
12	Home Office Reference	4		
13	International ID (Passport)	2		
14	Country Issuing International ID	2		
15	International ID Details Linked	2		
16	NI Number	2		
17	First Line Address	3		
18	Postcode	3		
19	Telephone Number(s)	3		
20		56	6	

ADVANCE YOUR BUSINESS »

Application Data Model to CDM Analysis

	A	B	C	D	E
1	MODEL COMPARISON ATTRIBUTE	CDM Maximum	Options	Registry	Membership
2	Full Name		2		2
3	Family Name	6		6	6
4	Given Names	6		6	
5	Date of Birth	6		6	6
6	Gender	3			3
7	Nationality(s)	6			3
8	Language (s)	4			2
9	Alternative Details		4		2
10	Organisation Start Date	2		2	2
11	Internal Reference Available externally	2		2	
12	Home Office Reference	4		4	4
13	International ID (Passport)	2			
14	Country Issuing International ID	2			
15	International ID Details Linked	2			
16	NI Number	2			2
17	First Line Address	3			3
18	Postcode	3			3
19	Telephone Number(s)	3			3
20		56	6	26	41
21					
22					
23					

ADVANCE YOUR BUSINESS »

Data Standards - Approach

Select Target Attribute

- » Identify Local Application Alternatives
- » Look for Commonality
- » Review Government Standards
- » Review UK National Standards
- » Review Other Government Standards (e.g. NIST)
- » Review International Standards
- » Assess Applicability
- » Decide Solution
- » Confirm Solution
- » Publish Solution

ADVANCE YOUR BUSINESS »

Data Standards – Observations and Issues

» Local Standards
  » Name
    - Two Fields (Given Names and Family Name)
    - Single Composite Fields
  » Problems
    - Which is the Family Name (single field option)
    - Cultures with no Family Name
    - Length, Which Character set to Use
    - Using National Standards to International Situation
  » Gender
    - Single Character pretty universally (but also text)
    - Which standard to apply (M/F or H/D, not known, not disclosed, not specified)

ADVANCE YOUR BUSINESS »

Data Standards: Given Names Resolution

» eGIF

» Originally CDM mandated to be compliant with eGIF
  » Too short

» Application Standard
  » Acknowledged as almost long enough!

» New Standard for length 100 characters prevails

» Promoted the publication on CDM website/Release Note

ADVANCE YOUR BUSINESS »

Data Standards: Gender Resolution

» Application examples
  » Values: M/F, M/F/U, M/F/D

» eGIF:
  » 4 values 0,1(Male), 2(Female), 9

» Use of ISO/IEC 5218:2004 standard, enables international interchange

» Standard for 4 UK values with mappings

» Promoted the publication on CDM website/Release Note

ADVANCE YOUR BUSINESS »

Data Profile Reporting

» Standard Data Profiling
- Field Type Distribution
- Field Uniqueness
- Relationship Integrity

» CDM Profiling
- Person Baseline
- Event Baseline

» Issues
- Root Cause Analysis
- People/Process/Technology Approach
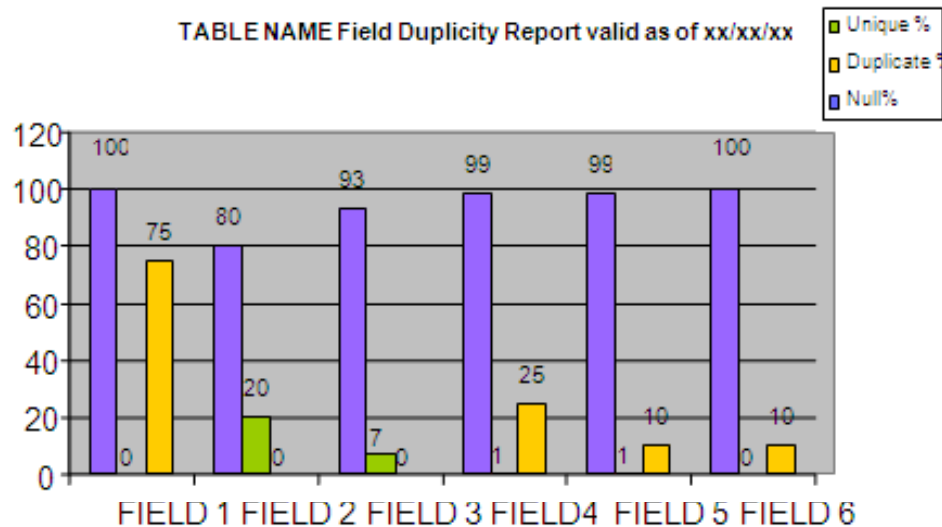
ADVANCE YOUR BUSINESS »

Field Type Distribution Report

Presents the percentage of rows within a particular field that fall into to any of the following field type categories:

» Null

» Integer

» String

» Decimal

» Space

ADVANCE YOUR BUSINESS »
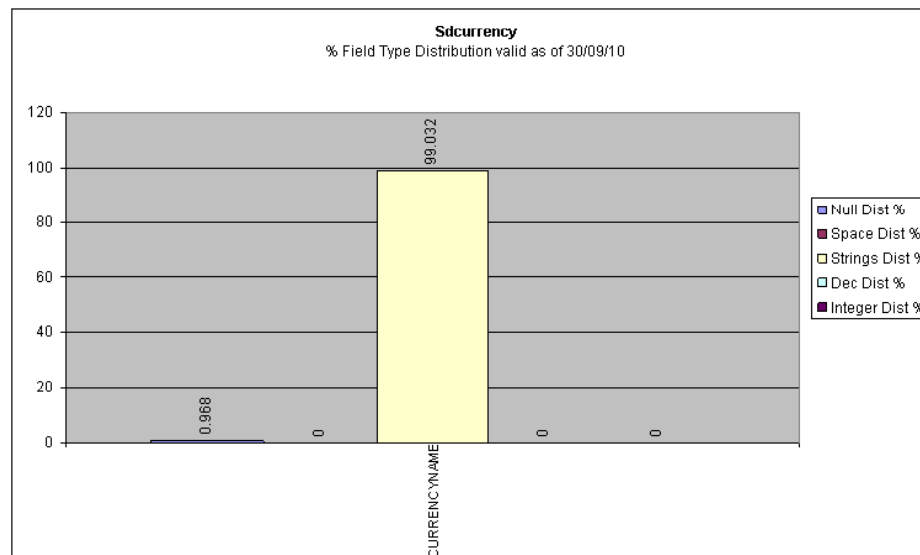
Field Uniqueness Report

Presents the percentage uniqueness of a column

TABLE NAME Field Duplicity Report valid as of xx/xx/xx

Legend: Unique %, Duplicate, Null%
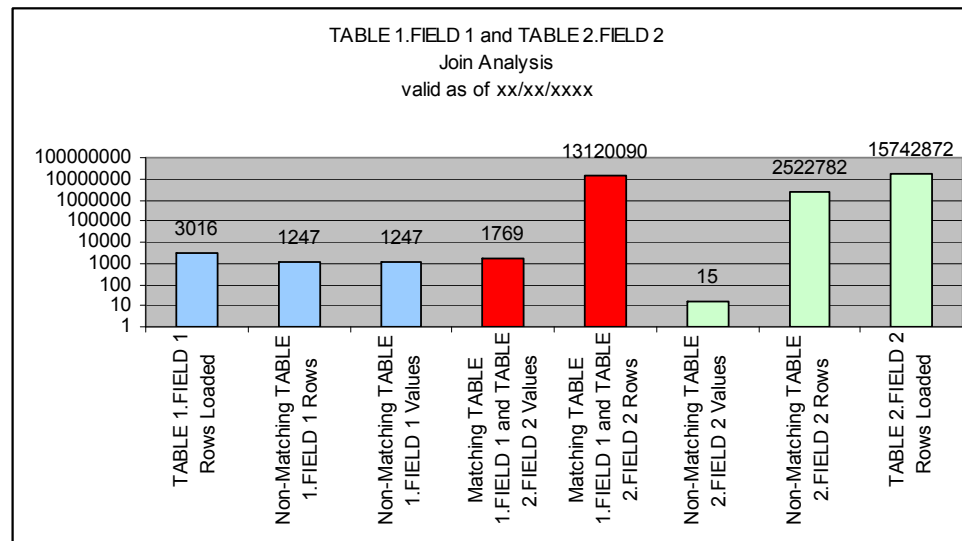
» Null Percentage

» All Distinct e.g. ID fields

» Reference Data (few distinct values)

» Form a view of adequate uniqueness

ADVANCE YOUR BUSINESS »

# Relationship Integrity Report

Presents the widows and orphans by number/percentage

TABLE 1.FIELD 1 and TABLE 2.FIELD 2
Join Analysis
valid as of xx/xx/xxxx

ADVANCE YOUR BUSINESS »

# Profiling Examples: Given Names

» Presence – 101 Null

» Unique Values – 99,999 N/A

» Patterns – 4,000
  » Leading Spaces – 100
  » Contains Digits – 50
  » Initials Only – 25
  » Trailing Spaces – 50
  » Leading Punctuation – 100
  » Non-printable Characters – 10

» Minimum Value – 0

» Maximum Value – Â…

ADVANCE YOUR BUSINESS »
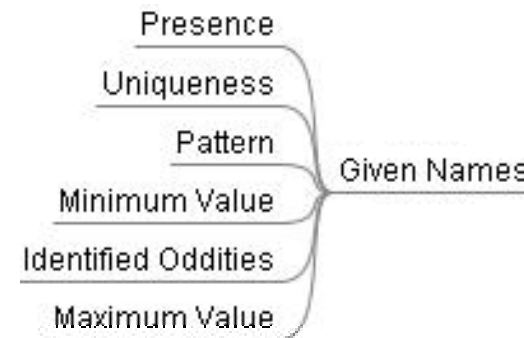
Profiling Examples: Gender

» Presence – 25,000 Null

» Unique Values – 25 values

» Patterns – 8
  » Leading Spaces – 1
  » Contains Digits – N/A
  » Initials Only – 2
  » Trailing Spaces – 3
  » Leading Punctuation – 4
  » Non-printable Characters – 5

» Minimum Value – male

» Maximum Value – vnm

ADVANCE YOUR BUSINESS »
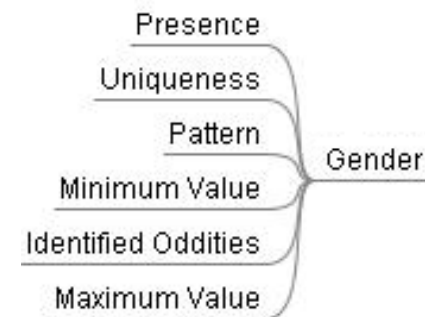
Gender – Unique Values

Value	Frequency	Dist %	Length	Soundex	Metaphone	Pattern	Mask
Male	357227	75.375	4	M400	ML	a4	AAAA
Female	83865	17.696	6	F540	FML	a6	AAA...
Unknown	6523	1.376	7	U525	UNKNN	a7	AAA...
male	5	0.001	4	M400	ML	a4	AAAA
M	2	0.001	1	M000	M	a	A
male	1	0.001	5	M400	ML	_a4	AAAA
,M	1	0.001	2	M000	M	pa	,A
`m	1	0.001	2	M000	M	pa	`A
f	1	0.001	1	F000	F	a	A
FEMALE	1	0.001	6	F540	FML	a6	AAA...
female	1	0.001	6	F540	FML	a6	AAA...
GHA	1	0.001	3	G000	F	a3	AAA
ind	1	0.001	3	I530	INT	a3	AAA
ma	1	0.001	2	M000	M	a2	AA
make	1	0.001	4	M200	MK	a4	AAAA
MALE	1	0.001	4	M400	ML	a4	AAAA
vnm	1	0.001	3	V500	FNM	a3	AAA

ADVANCE YOUR BUSINESS »

CDM Profiling – Report and Presentation

» Report
  » Executive Summary
  » Observations
    - People
    - Process
    - Technology
  » Profile Scope
  » How to read the Document
  » Business Drivers
  » Profile Objective
  » Profile Approach
  » Assumptions
  » Observation Details
  » Next Steps

» Presentation

  » Introduction

  » Process Overview

  » Observations

  » Data Facts
    » Standard Reports
    » Person Baseline Analysis

  » Findings

  » Recommendations

  » Next Steps

ADVANCE YOUR BUSINESS »

Addressing the Issues

» **Root Cause Analysis**
  » Fact Based
  » To preclude problem recurring

» **Examine**
  Code   Specifications
  UI and UI Standards   Use of dropdowns
  Field Validation   Talk with Users

» **General BA Tools.**

» **Reporting Approach**
  » People
  » Process
  » Technology

ADVANCE YOUR BUSINESS »

Thank You

Defining a Person Metadata Model to Improve Data Quality

ADVANCE YOUR BUSINESS >>

Atos Origin ™

CONSULTING > SOLUTIONS > OUTSOURCING

For more information please contact:

Ian Woodrow

m   +44 (0)7974 674045
ian.woodrow@atosorigin.com

Atos Origin (UK)
4 Triton Square
NW1 3HG, London
www.atosorigin.com

ADVANCE YOUR BUSINESS »