# Data Quality = The Achilles Heel
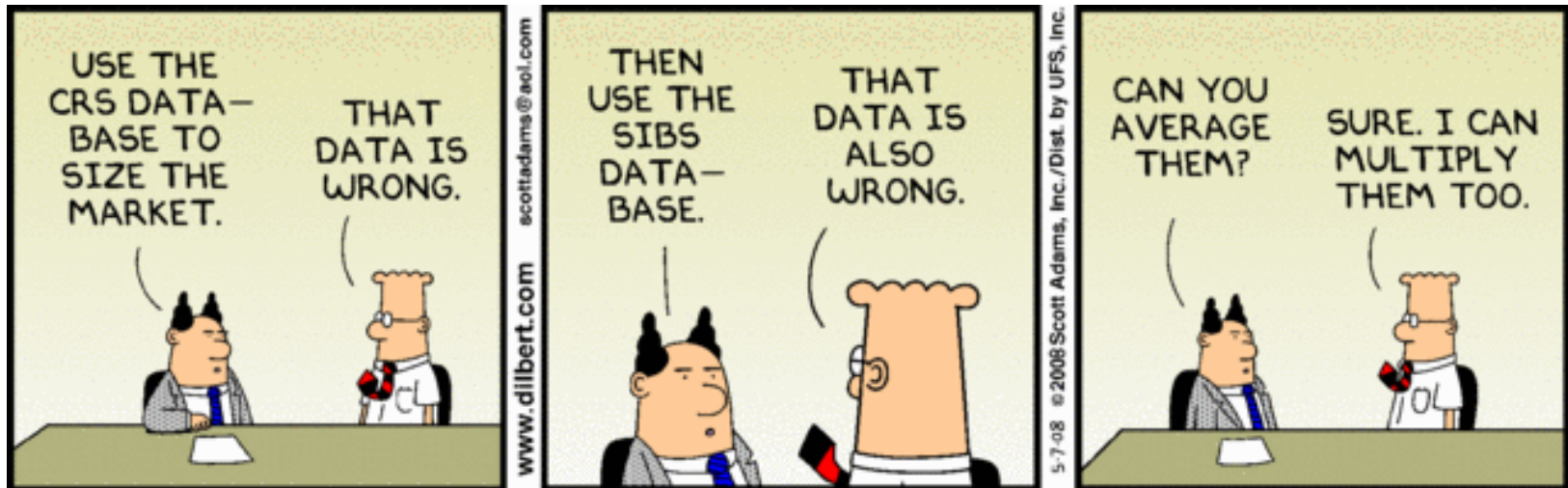## Presented to:
## Oracle User Group NZ

Intech Solutions

Elizabeth Eastwood

# Our customers

# Data Dependency

**Improve The Quality And Utilisation Of Data**

"The greatest enemy of knowledge is not ignorance, it is the illusion of knowledge." Stephen Hawking

Poor data quality could cost companies up to 20% of their operating profits.

(David Howard-Jones of management consultancy Oliver Wyman 2011)

The cost to fix a data error at the time of entry is $1.

The cost to fix it an hour after it's been entered is $10.

And the cost to fix it several months later is $100+.

Integrated Information Quality

# Intech Solutions

**Domains of expertise**

**Intech Solutions**

## Data Matching
- Intelligent Match, Single Customer View, Deduplicate, Search, Master Data Management

## Unstructured Data Collector
- PDF, Text, Email, etc. Input and Parse

## Data Validation
- Email, Phone, ABN, BSB, and others

## Address Validation & Geo-Coding
- Forward & Reverse Geocoding, Boundary Tagging, "type-ahead", bulk, postal

## Data Profiler
- Data discovery, structures, frequency distribution and various data anomalies

## Data Parsing, Cleaning and Standardising
- for Data Integration, ETL, Analytics. Required for Address Validation.

**Integrated Information Quality**

# Business Solutions from Intech

**Delivered by IQ Office**

- Address Validation
- Data Migration
- Geocoding
- Identity Searching
- Intelligence Analytics
- Single Customer View



**Intech Solutions**

IQ Office

Collection Processor
Data Validation & Cleansing
Data Profiling
Data Matching
Data Enhancement
Data Parsing & Standardisation

Integrated Information Quality

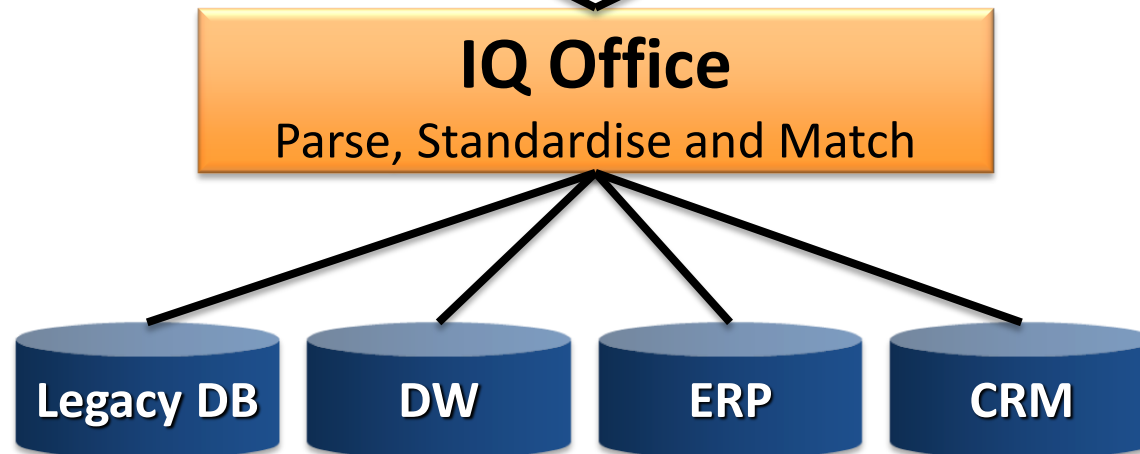# Single Customer View

# How Does It Work?

## Empower Staff, Delight Clients



Internal user identifies correct customer and updates their details

Call centre user identifies existing customer or creates new customer

**IQ Office**
Parse, Standardise and Match

**Legacy DB**  **DW**  **ERP**  **CRM**

Integrated Information Quality

# Jetstar Single Customer View

**Intech Project - On Time, On Budget**

- Single Customer View & Customer Segment Tracking

- Intech Built And Supported

- Data Loaded From Multiple Systems
  - Flight Bookings
  - Call Centre CRM
  - Qantas Frequent Flyer Data
  - Customer Web Portals
  - 60 Million Customer Records and Growing
  - 70% Australian, 30% From Other Countries

- SCV Produces
  - Consolidated View of Customer
  - Enhances Jetstar's Customer Knowledge of Each Customer
  - Enhances Reporting and Analysis Capabilities

Integrated Information Quality

# Jetstar Single Customer View

- Uniquely identify customers from multiple data sources
- Analyse by market segmentation (cohorts)
- Understand customer and segment spend
- Understand customer behaviour (travel destinations)

- Develop strategies focussed on priority market segments
- Enable 'right product, right time' marketing campaigns

- Automated, Integrated, Accurate, Timely

# Intelligence Analytics

## Integrating Information for Enhanced Intelligence

Integrated Information Quality

# Executive Summary

## Challenges addressed by Intech Solutions

- A common failure in data transformation projects is the inability to use information due to bad data quality.

- Implementing new software solutions and processes enable greater use of information, but also require that information to be sufficiently high in quality.

- This is specifically so in intelligence systems where vast amounts of structured and unstructured information is brought together and is required to be used to identify complete and consolidated information.

# Problem Definition

**Intech Solutions**

## Intelligence Functions and Obstacles to Gaining Intelligence

**Functions:**

Hypothesis formulation

Audit, Intelligence & Forensic Support

Enforcement

| Data Acquisition | Interactive Exploration | Evidence Packaging |
|---|---|---|
| (Structured and Unstructured) | Moving from hypothesis to evidence | Satisfying the burden of proof |

**Obstacles:**

Inflexible: difficult to change

Invalid data

Messy data

So much data

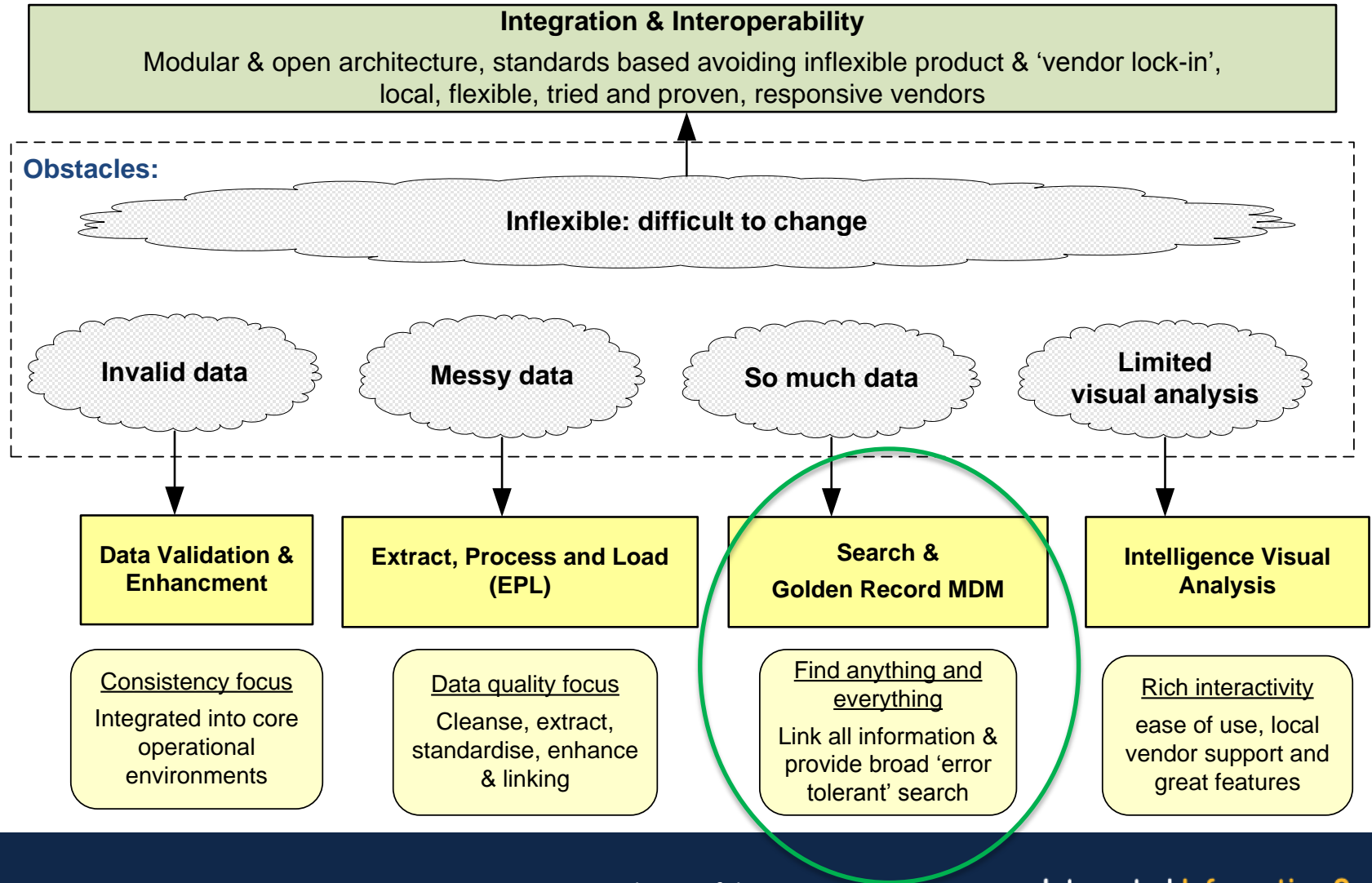Limited Interactive Analysis capability

Integrated Information Quality

# Problem Definition

**Intech Solutions**

## Common Problems

- <u>Invalid data</u> – source systems contain old data, invalid/wrong data, duplicated information which is not linked.

- <u>Messy data</u> - different formats, structures, and varying accuracy levels, which degrade usability.

- <u>So much data</u> - can't easily find all data related to an entity. Related data present but not linked.

- <u>Limited Visual analysis</u> – 'join the dots' capability is too restrictive.

- <u>Inflexible: difficult to change</u> - inefficient use of new data sources, slow to act on new threats, changed market conditions, and new technologies, and difficult to appropriately collaborate and share information.
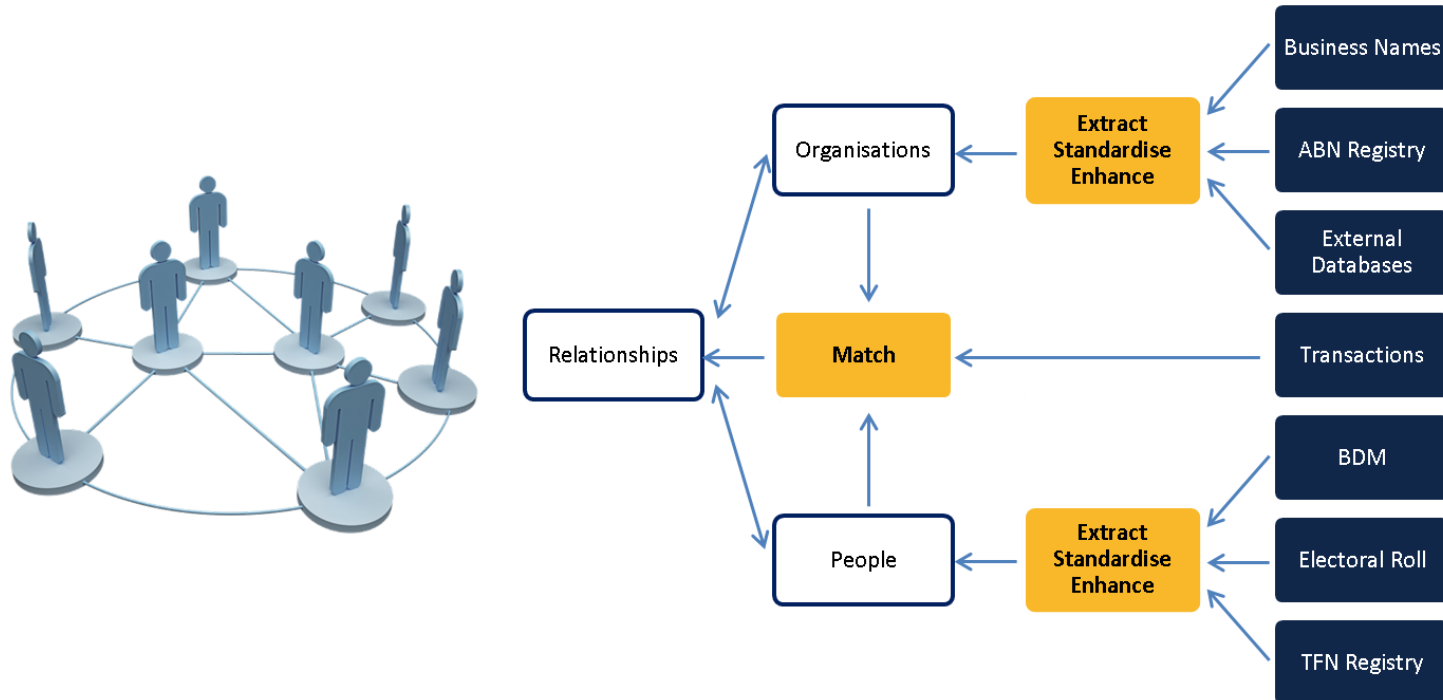
**Integrated Information Quality**

# Problem Definition

## Addressing the Problems

**Integration & Interoperability**

Modular & open architecture, standards based avoiding inflexible product & 'vendor lock-in', local, flexible, tried and proven, responsive vendors

**Obstacles:**

Inflexible: difficult to change

| Invalid data | Messy data | So much data | Limited visual analysis |

| Data Validation & Enhancment | Extract, Process and Load (EPL) | Search & Golden Record MDM | Intelligence Visual Analysis |

Consistency focus
Integrated into core operational environments

Data quality focus
Cleanse, extract, standardise, enhance & linking

Find anything and everything
Link all information & provide broad 'error tolerant' search

Rich interactivity
ease of use, local vendor support and great features

Integrated Information Quality

Intech Solutions

# Solution formation

## A two-fold Solution

1. Deliver an information quality service layer which extracts, standardises and enhances the data across the data collection points

2. Deliver an intelligence information repository that identifies relationships between people, organisations and location records

# Solution Features

## A Holistic and Modular Based Approach
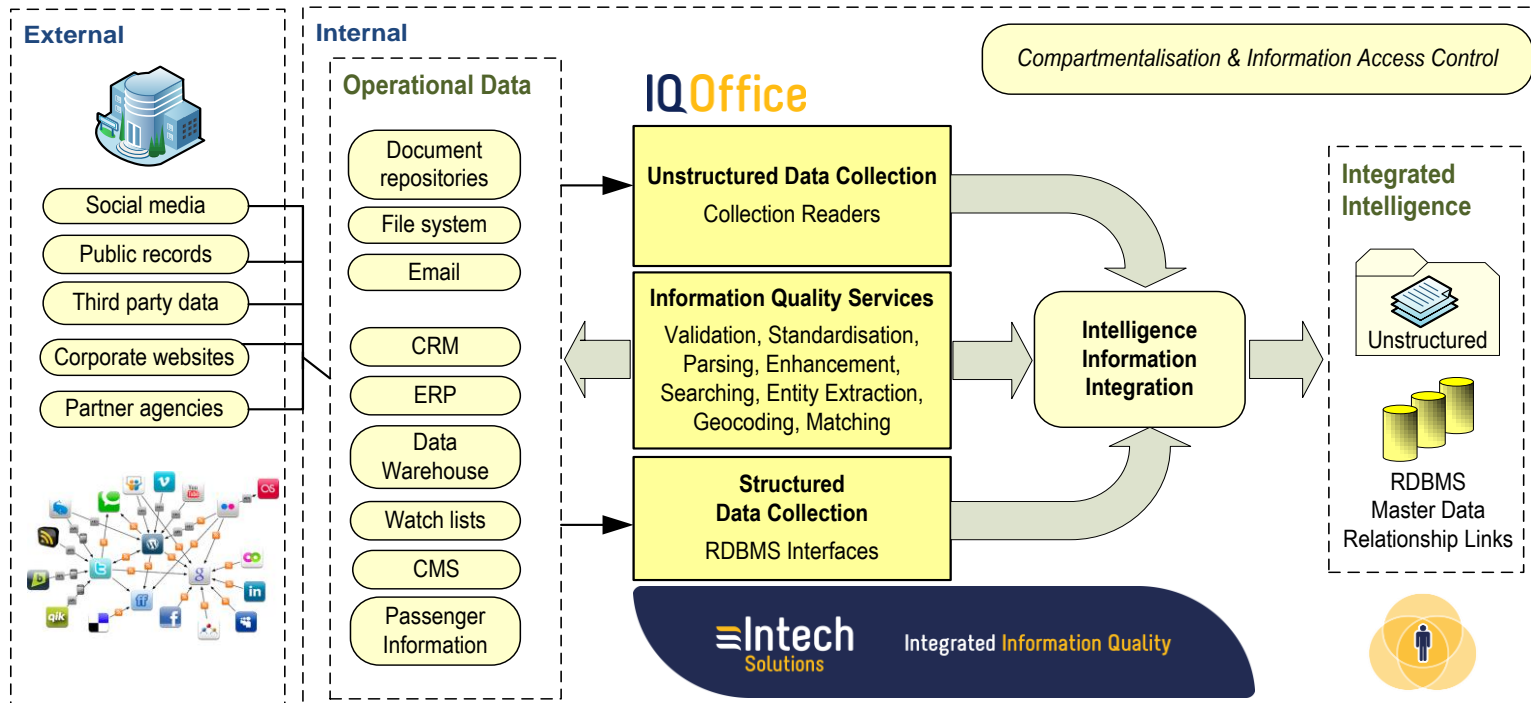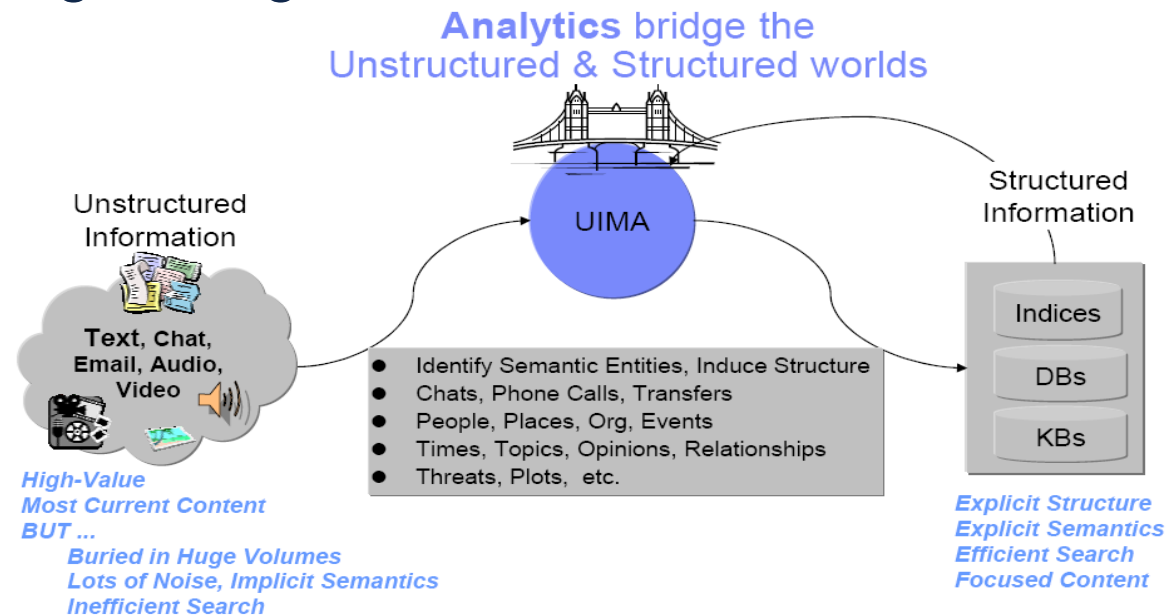


A two pronged approach

- <u>Information Quality Service</u> – validate and enhances data on collection

- <u>Intelligence Information Integration </u>– enhance, integrate, link and make all available information accessible to those who need it.

Integrated Information Quality

# Collection Readers

**Unstructured Information Management Architecture**

- Standard framework
- Vendor neutral
- Interoperate without the need for software programming/coding

**Analytics** bridge the Unstructured & Structured worlds

Unstructured Information

Text, Chat, Email, Audio, Video

High-Value
Most Current Content
BUT ...
Buried in Huge Volumes
Lots of Noise, Implicit Semantics
Inefficient Search

UIMA

- Identify Semantic Entities, Induce Structure
- Chats, Phone Calls, Transfers
- People, Places, Org, Events
- Times, Topics, Opinions, Relationships
- Threats, Plots, etc.

Structured Information

Indices

DBs

KBs

Explicit Structure
Explicit Semantics
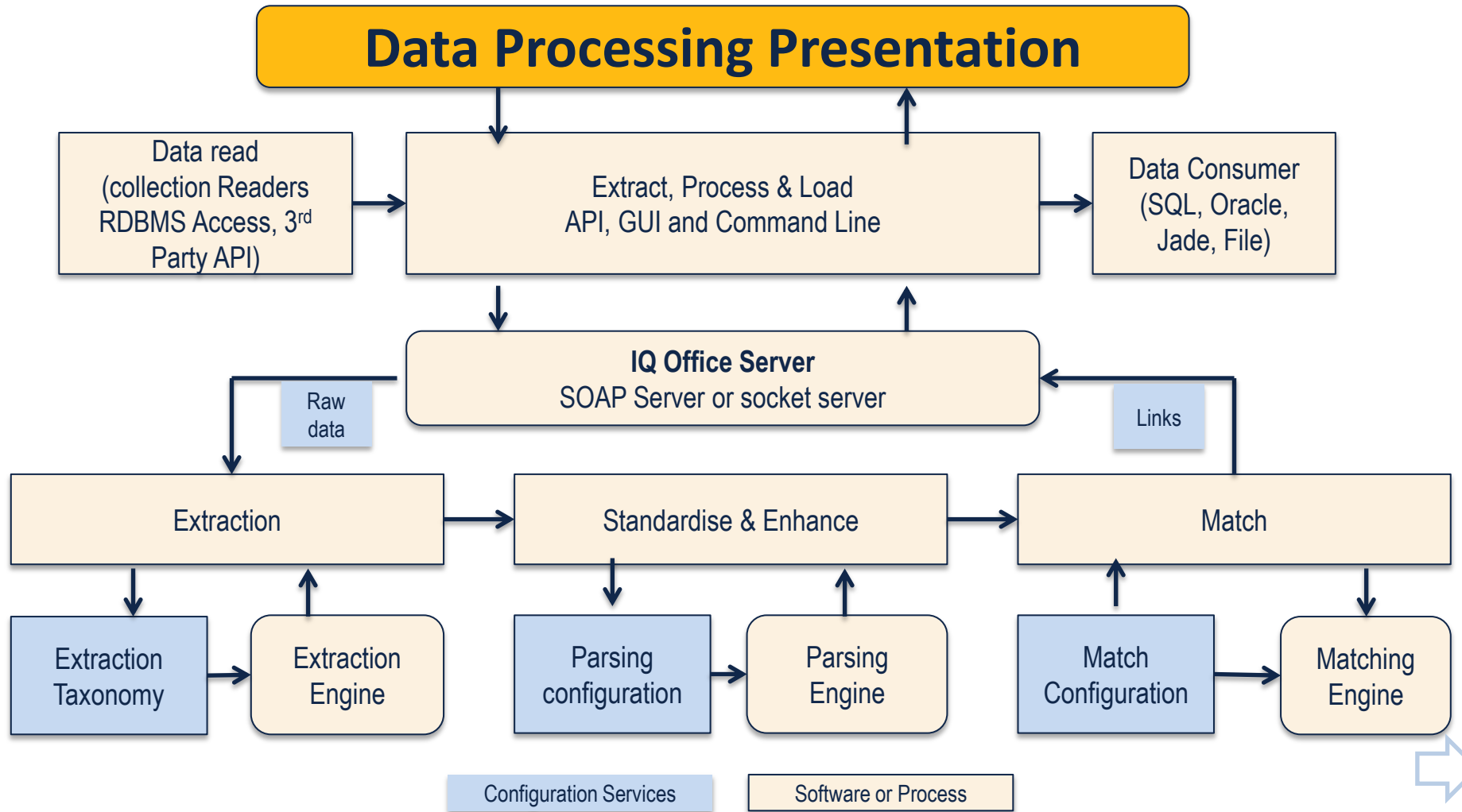Efficient Search
Focused Content

# Data Model

**Bringing Order To Disparate Data Sets**

- The data model is open
- May be set to match your existing data structures
- Several template data models are available
- Or adopt a 'standard' data model
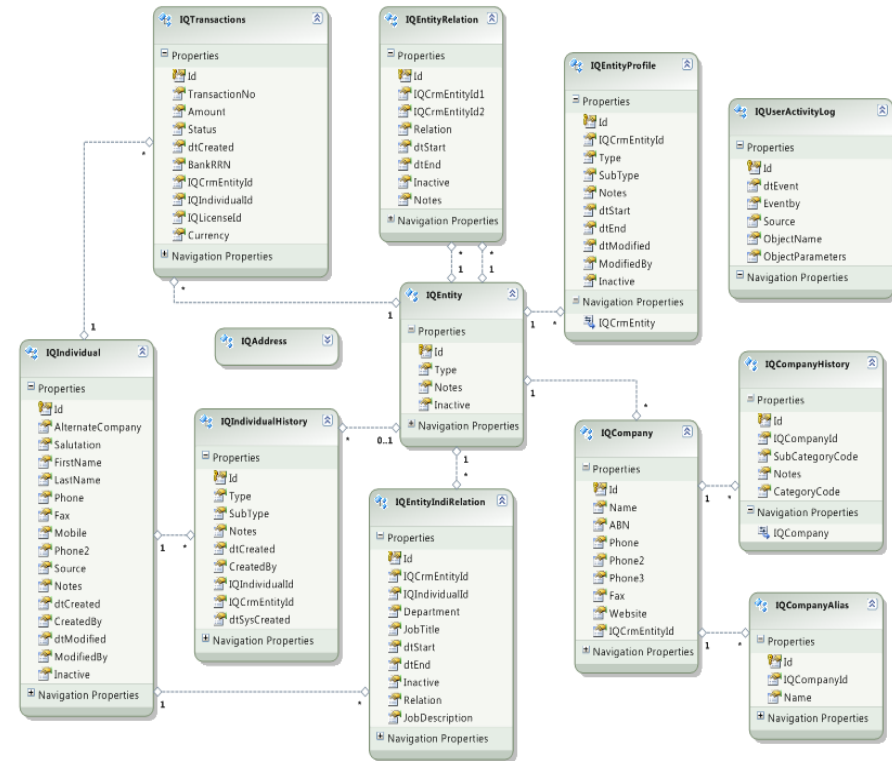
# User Capabilities

**Flexible , Functionality**

≡Intech
**Solutions**

- Search For An Entity

- Update The 'Golden Record' Of An Entity

- Enhance Data

- Find Related Entities

- Submit Complex Queries To The Intelligence Platform

- Read & Write Data

  - Entities

  - Relationships

  - Annotations

- Many Others

**Integrated Information Quality**

# Solution Enablers

**Combination of intellectual property and expertise**

The elements which enable an optimal solution to be delivered are the combination of skills, subject area expertise, and best-in-class technologies, specifically put together to overcome the key obstacles.

- The technologies have been deployed at over 120 sites across Australia and New Zealand including AGD, ABS, DSS, DVA, and Immigration NZ to name a few, and offers a proven solution in demanding business and technical environments.

- The teams which enable this solution to be delivered include experienced project managers and implementation staff.

    - Implementation team have extensive experience at integrating information and resolving data quality issues.

    - Local expertise which is responsive and adaptable to change.

Integrated Information Quality

# Implementation strategy

## Dedicated and specialist expertise



# Information Quality
## Enhanced opportunities, reduced costs

Knowledge, expertise and experience:

- Managed by senior project managers

- Delivered by experts – averaging over 15 years experience
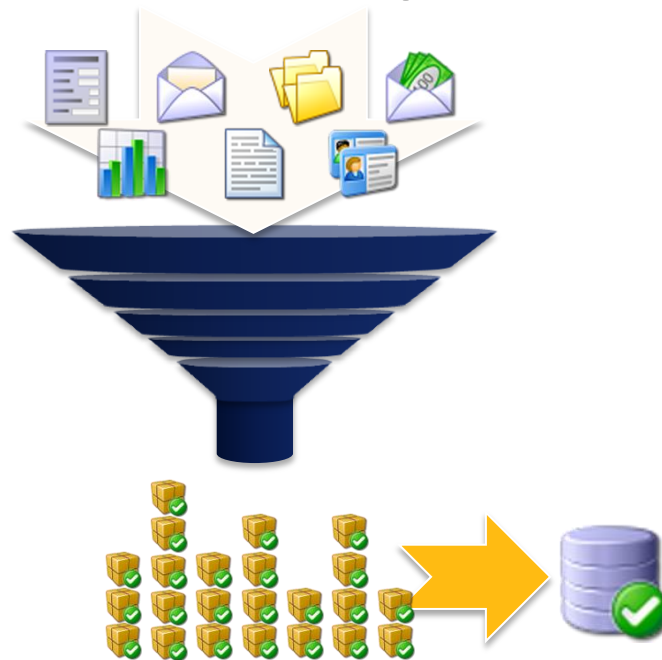
- Staff security cleared to SECRET by AFP

# IQ Collection Processor
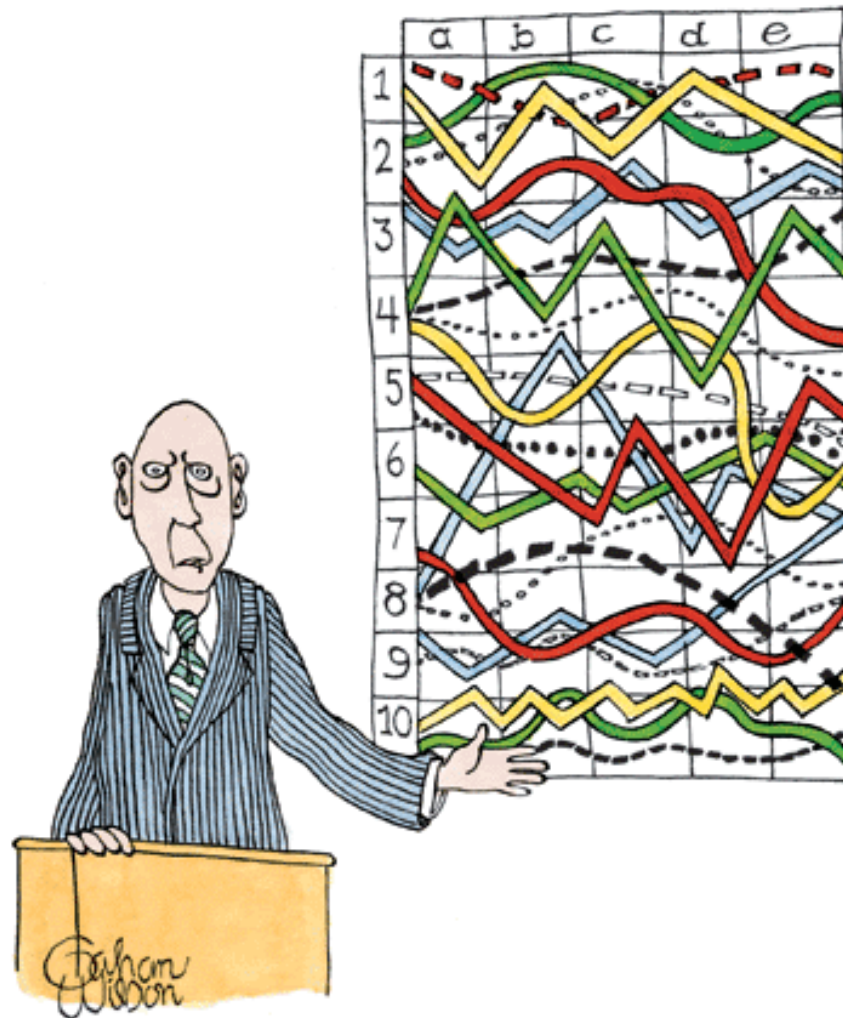## Unstructured Data Import & Cleanse

**Integrated Information Quality**
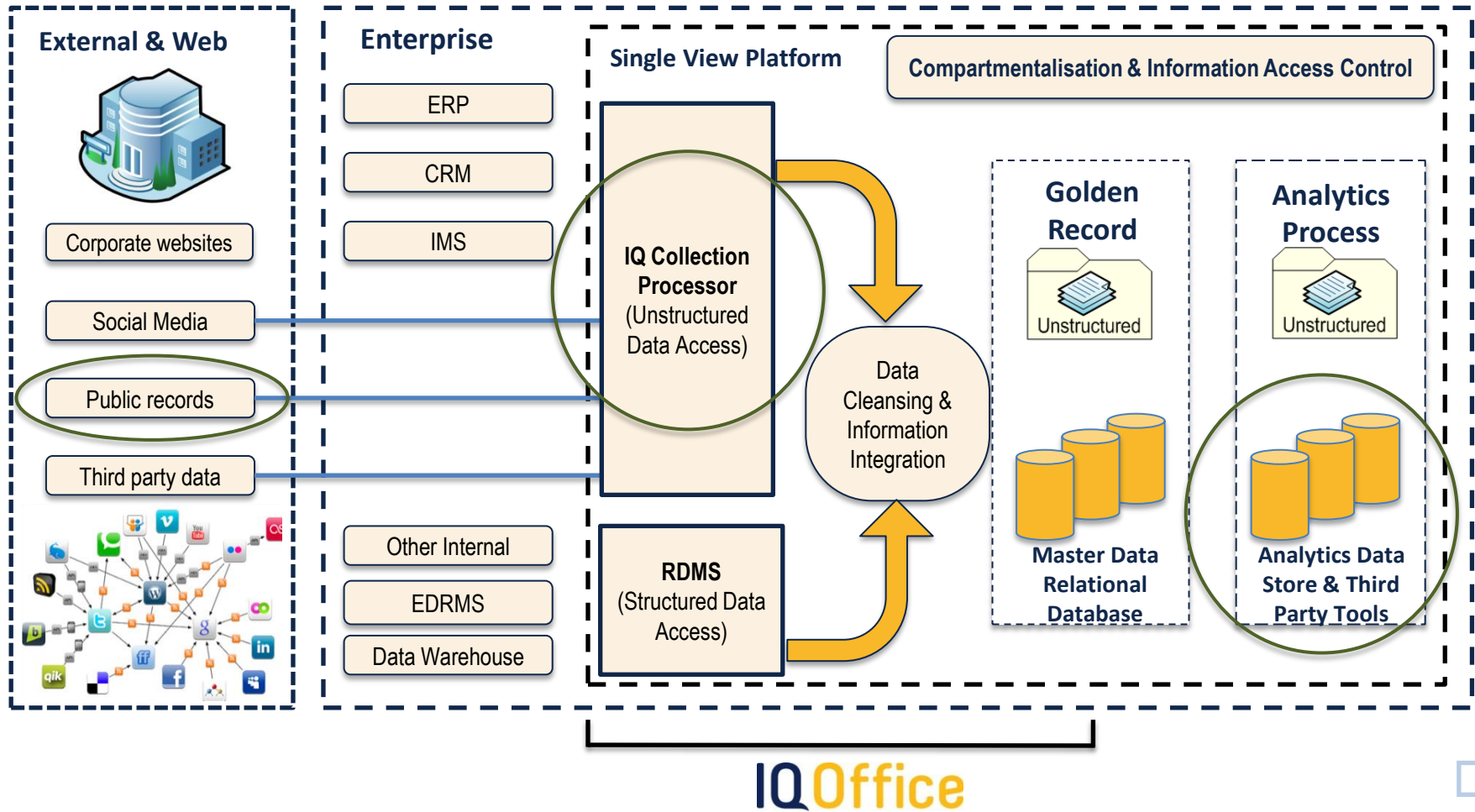
"I'll pause for a moment so you can let this information sink in."

# Data Integration & Importing

## IQ Collection Processor



**External & Web**
- Corporate websites
- Social Media
- Public records
- Third party data

**Enterprise**
- ERP
- CRM
- IMS
- Other Internal
- EDRMS
- Data Warehouse

**Single View Platform**

**Compartmentalisation & Information Access Control**

IQ Collection Processor (Unstructured Data Access)

RDMS (Structured Data Access)

Data Cleansing & Information Integration

**Golden Record**
Unstructured

Master Data Relational Database

**Analytics Process**
Unstructured

Analytics Data Store & Third Party Tools

**IQ Office**

Intech Solutions

Integrated Information Quality

# The challenge

**Automating Import Of Unstructured Data**

- ASIC Extracts
  - Individual
  - Company

**GAFFWICK PTY. LTD. 010 584 522**

ASIC - Current & Historical Extract - GAFFWICK PTY. LTD. ACN: 010 584 522

t contains information derived from the Australian
and Investment Commission's (ASIC) database under
74A of the Corporations Act 2001.
rise ASIC of any error or omission which you may identify.

**IDENTIFICATION**

ACN: 010 584 522
ABN:
Current Company Name: GAFFWICK PTY. LTD.
Registered in: Queensland
Place of Registration:
Registration Date: 26/06/1985
Previous State Number: 85B03744X
Governance Type:
Review Date: 07/08/2013

**CURRENT COMPANY DETAILS**

Name: GAFFWICK PTY. LTD.
Period from: 08/07/2003
Name Start: 26/06/1985
Status: REGISTERED
Type: AUSTRALIAN PROPRIETARY COMPANY
Class: LIMITED BY SHARES
Subclass: PROPRIETARY COMPANY
Disclosing Entity: N

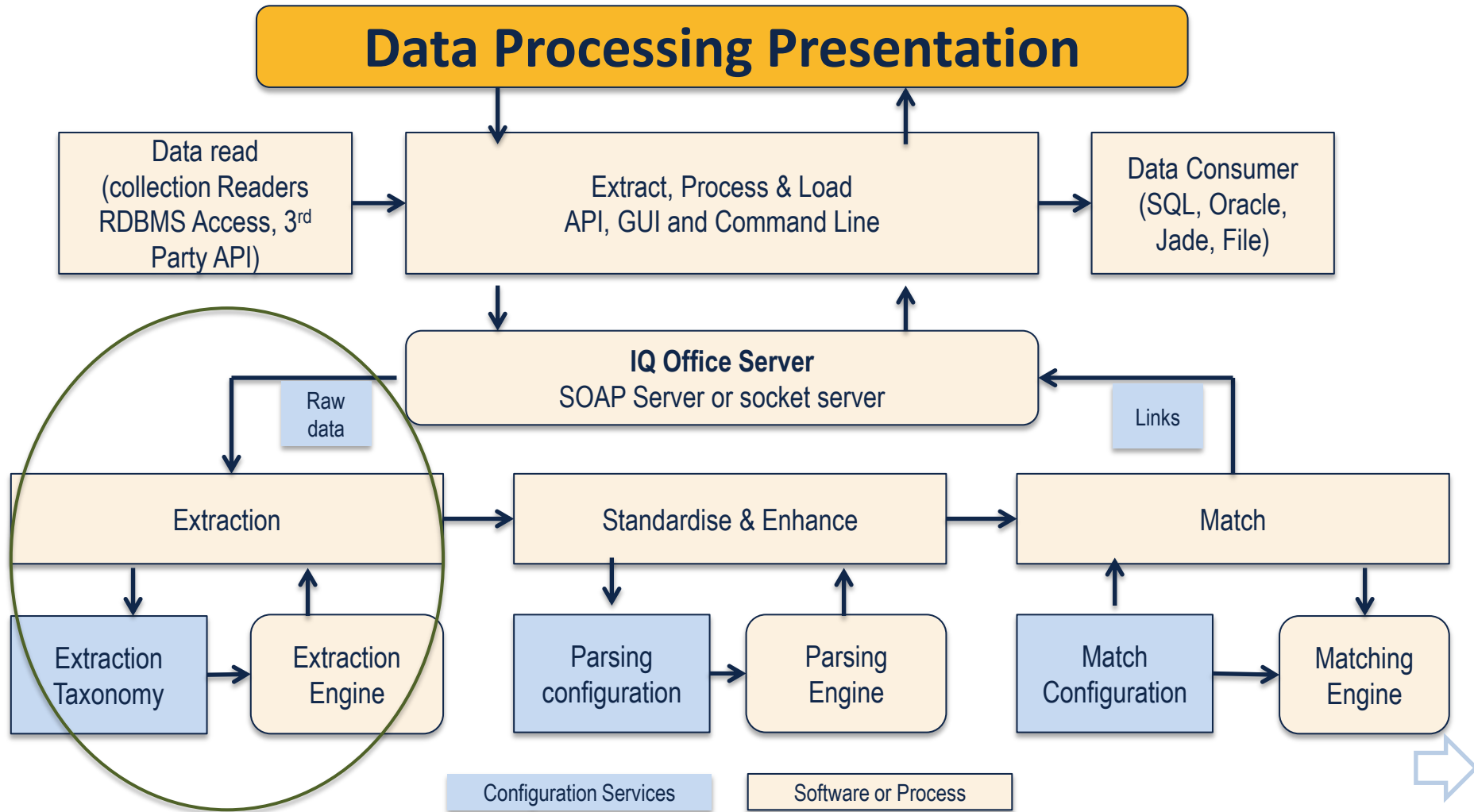**CURRENT COMPANY ADDRESS**

Address Type: Registered Office
Address: 56 HOPETOUN AVENUE
MOSMAN NSW 2088

- Task → process and import to SQL

ASIC = Australian Securities & Investment Commission

Integrated Information Quality

# IQ Office

## Raw Data To Analytics - Ready Today

**Intech Solutions**

**Data Processing Presentation**

| | | |
|---|---|---|
| Data read (collection Readers RDBMS Access, 3rd Party API) | Extract, Process & Load API, GUI and Command Line | Data Consumer (SQL, Oracle, Jade, File) |

**IQ Office Server**
SOAP Server or socket server

Raw data

Links

| Extraction | Standardise & Enhance | Match |
|---|---|---|

| Extraction Taxonomy | Extraction Engine | Parsing configuration | Parsing Engine | Match Configuration | Matching Engine |
|---|---|---|---|---|---|

Configuration Services

Software or Process

Integrated Information Quality

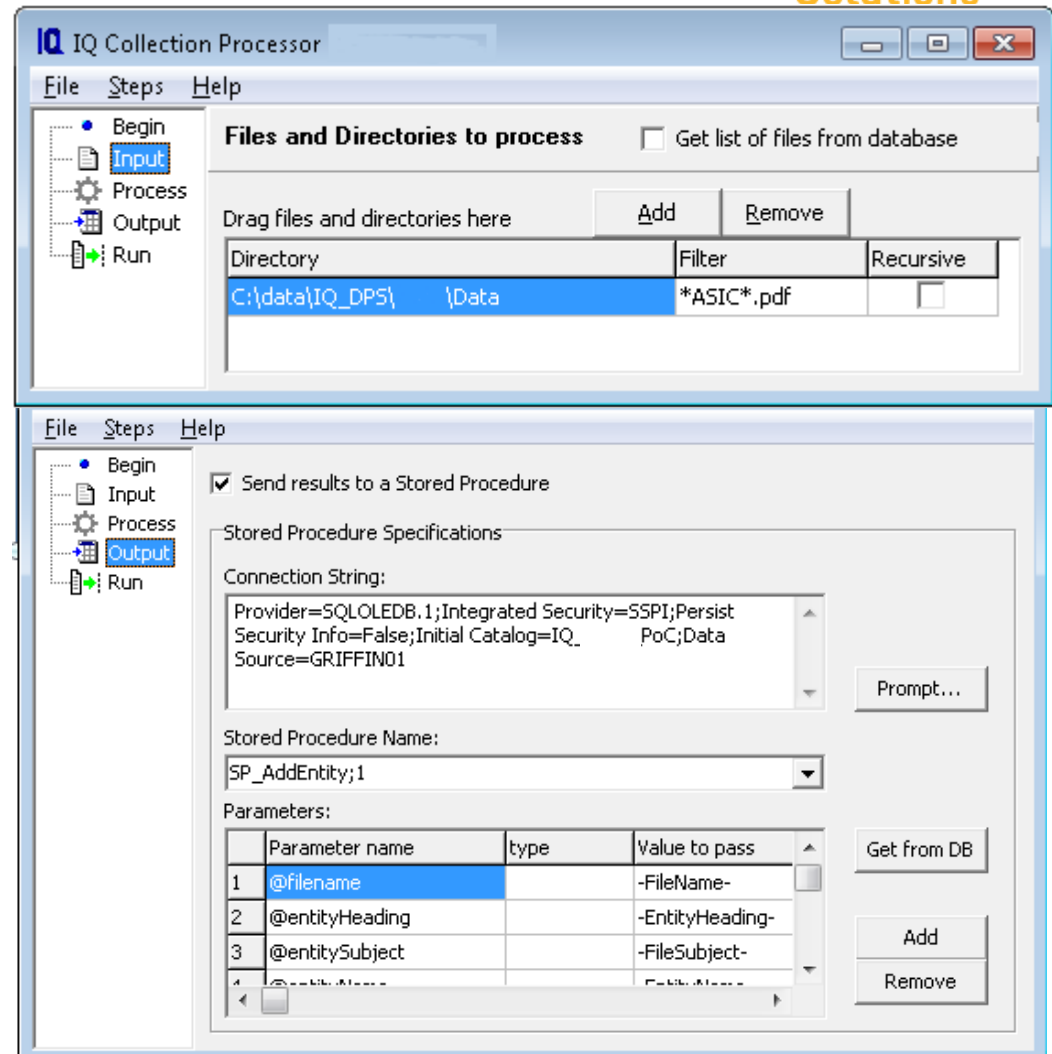# Results

## IQ Collection Processor

- Configurations stored for re-use
- User selects file type
  - ASIC extracts
- Process can be integrated and executed from other applications
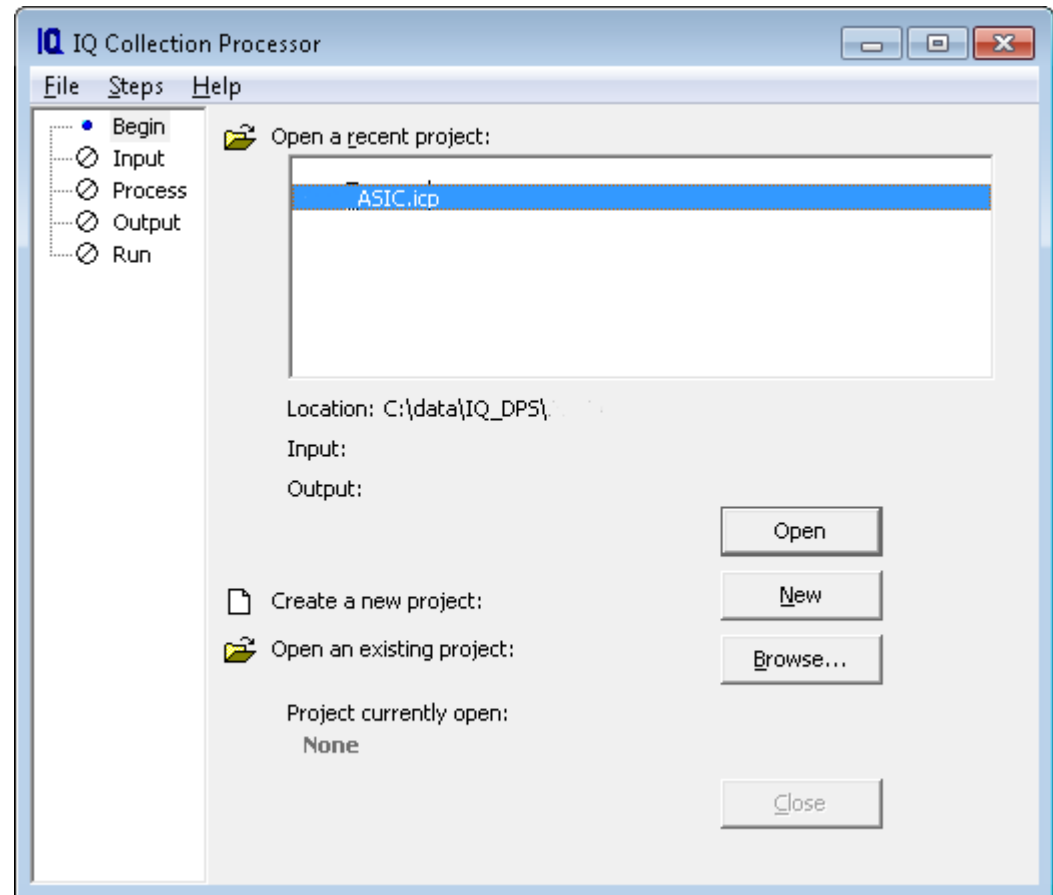
# Results

**IQ Collection Processor**

- Input configured to
- read data from PDF
- Output goes to SQL Server
  - Via stored procedure
  - To tables
- Stored procedure can send data directly to your systems

# Results

## IQ Collection Processor

- Process executed
- Fields extracted and written to SQL Server



**IDENTIFICATION**

ACN: 010 584 522
ABN:
Current Company Name: GAFFWICK PTY. LTD.
Registered in: Queensland
Place of Registration:
Registration Date: 26/06/1985
Previous State Number: 85B03744X
Governance Type:
Review Date: 07/08/2013

**CURRENT COMPANY DETAILS**

Name: GAFFWICK PTY. LTD.
Period from: 08/07/2003
Name Start: 26/06/1985
Status: REGISTERED
Type: AUSTRALIAN PROPRIETARY COMPANY
Class: LIMITED BY SHARES
Subclass: PROPRIETARY COMPANY
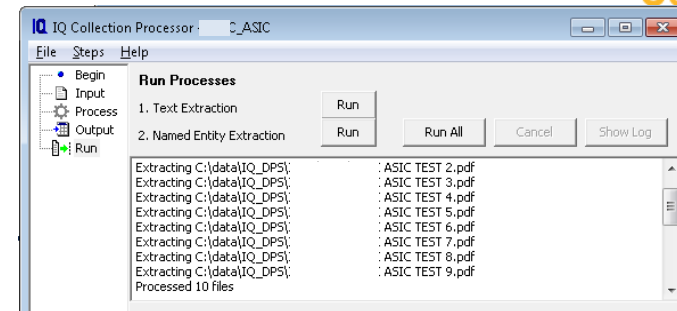Disclosing Entity: N

**CURRENT COMPANY ADDRESS**

Address Type: Registered Office
Address: 56 HOPETOUN AVENUE
MOSMAN NSW 2088
Period from: 21/08/2007

Address Type: Principal Place of Business
Address: 56 HOPETOUN AVENUE
BALMORAL NSW 2088

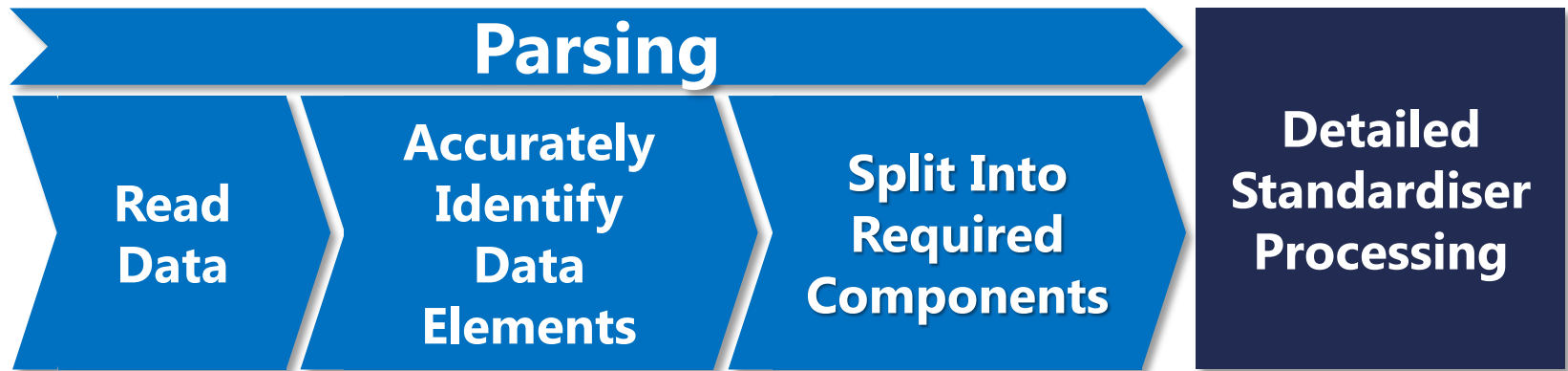| | entityHeading | entityName | entityType | entityValue |
|---|---|---|---|---|
| ASIC TEST 4.pdf | IDENTIFICATION | ACN | ACN | 010 584 522 |
| ASIC TEST 4.pdf | IDENTIFICATION | Current Company Name | Company | GAFFWICK PTY. LTD |
| ASIC TEST 4.pdf | IDENTIFICATION | Registered in | | Queensland |
| ASIC TEST 4.pdf | IDENTIFICATION | Registration Date | Date | 26/06/1985 |
| ASIC TEST 4.pdf | IDENTIFICATION | Previous State Number | | 85B03744X |
| ASIC TEST 4.pdf | IDENTIFICATION | Review Date | Date | 07/08/2013 |
| ASIC TEST 4.pdf | CURRENT COMPANY DETAILS | | Company | GAFFWICK PTY. LTD |
| ASIC TEST 4.pdf | CURRENT COMPANY DETAILS | | ACN | 019 349 710 |
| ASIC TEST 4.pdf | CURRENT COMPANY DETAILS | Period from | Date | 08/07/2003 |
| ASIC TEST 4.pdf | CURRENT COMPANY DETAILS | Name Start | Date | 26/06/1985 |
| ASIC TEST 4.pdf | CURRENT COMPANY DETAILS | Status | | REGISTERED |
| ASIC TEST 4.pdf | CURRENT COMPANY DETAILS | Type | | AUSTRALIAN PROPRIETARY COMPANY |
| ASIC TEST 4.pdf | CURRENT COMPANY DETAILS | Class | | LIMITED BY SHARES |
| ASIC TEST 4.pdf | CURRENT COMPANY DETAILS | Subclass | | PROPRIETARY COMPANY |
| ASIC TEST 4.pdf | CURRENT COMPANY DETAILS | Disclosing Entity | | N |
| ASIC TEST 4.pdf | CURRENT COMPANY ADDRESS | Address Type | | Registered Office Doc# 1E3541350 |
| ASIC TEST 4.pdf | CURRENT COMPANY ADDRESS | Address | Address | 56 HOPETOUN AVENUE |
| ASIC TEST 4.pdf | CURRENT COMPANY ADDRESS | Address | Locality | MOSMAN NSW 2088 |
| ASIC TEST 4.pdf | CURRENT COMPANY ADDRESS | Period from | Date | 21/08/2007 |
| ASIC TEST 4.pdf | CURRENT COMPANY ADDRESS | Address | Address | 56 HOPETOUN AVENUE |
| ASIC TEST 4.pdf | CURRENT COMPANY ADDRESS | Address | Locality | BALMORAL NSW 2088 |
| ASIC TEST 4.pdf | CURRENT COMPANY ADDRESS | Period from | Date | 15/10/1999 |

Integrated Information Quality
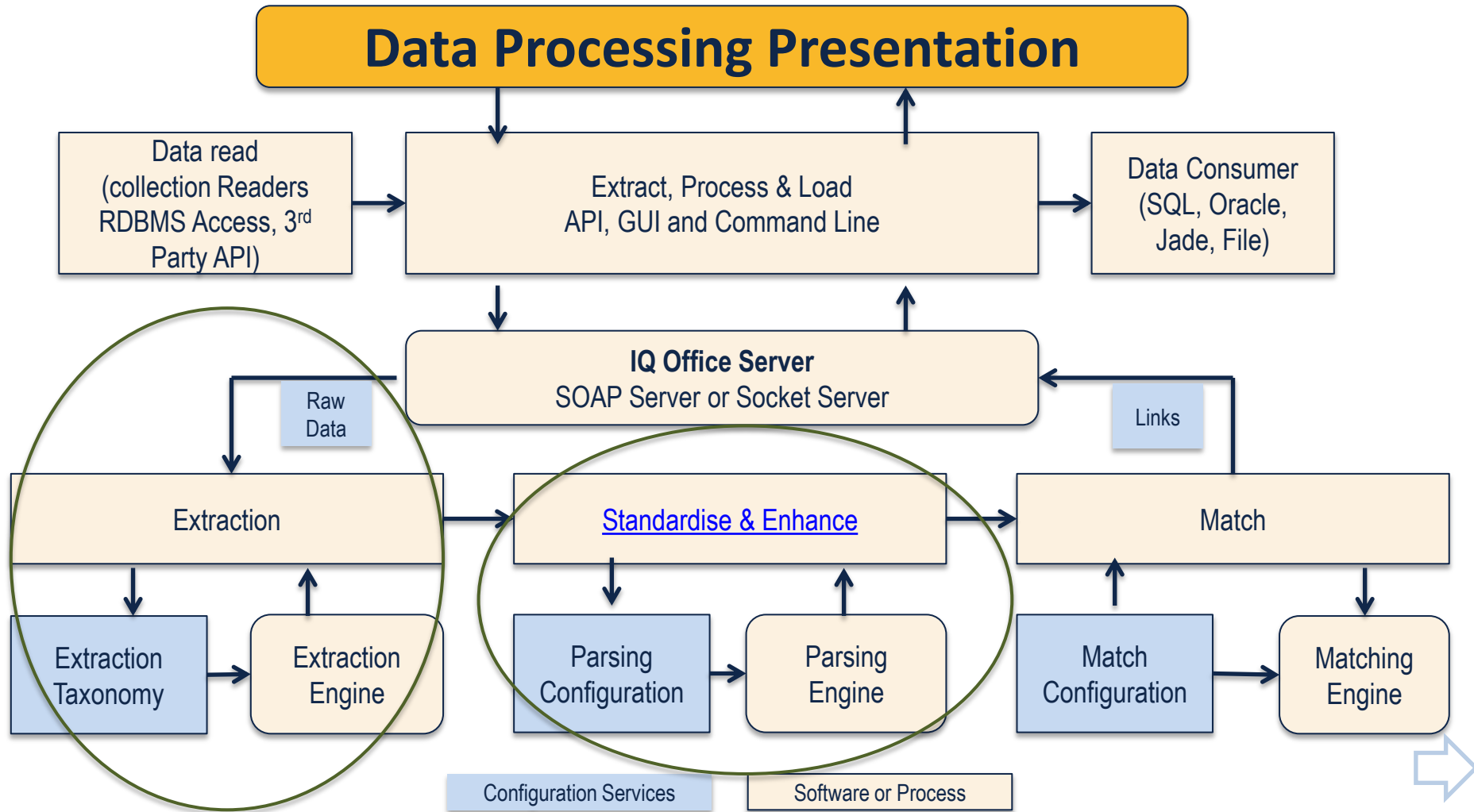
# Data Parsing & Standardisation
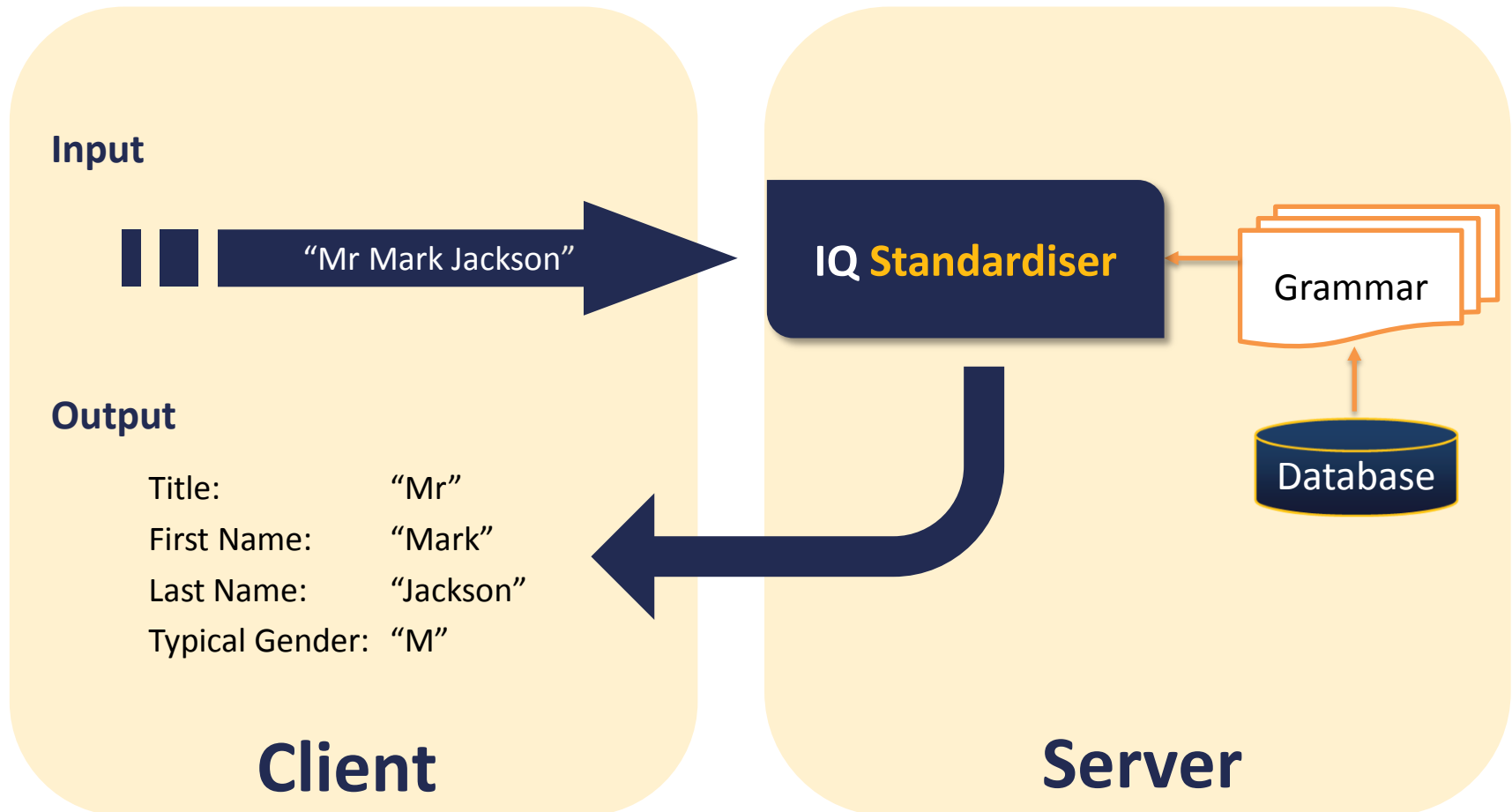
# Parsing & Standardisation

**Customisable and Standard Grammars**

- High performance parsing
- Data transformation and scrubbing engine
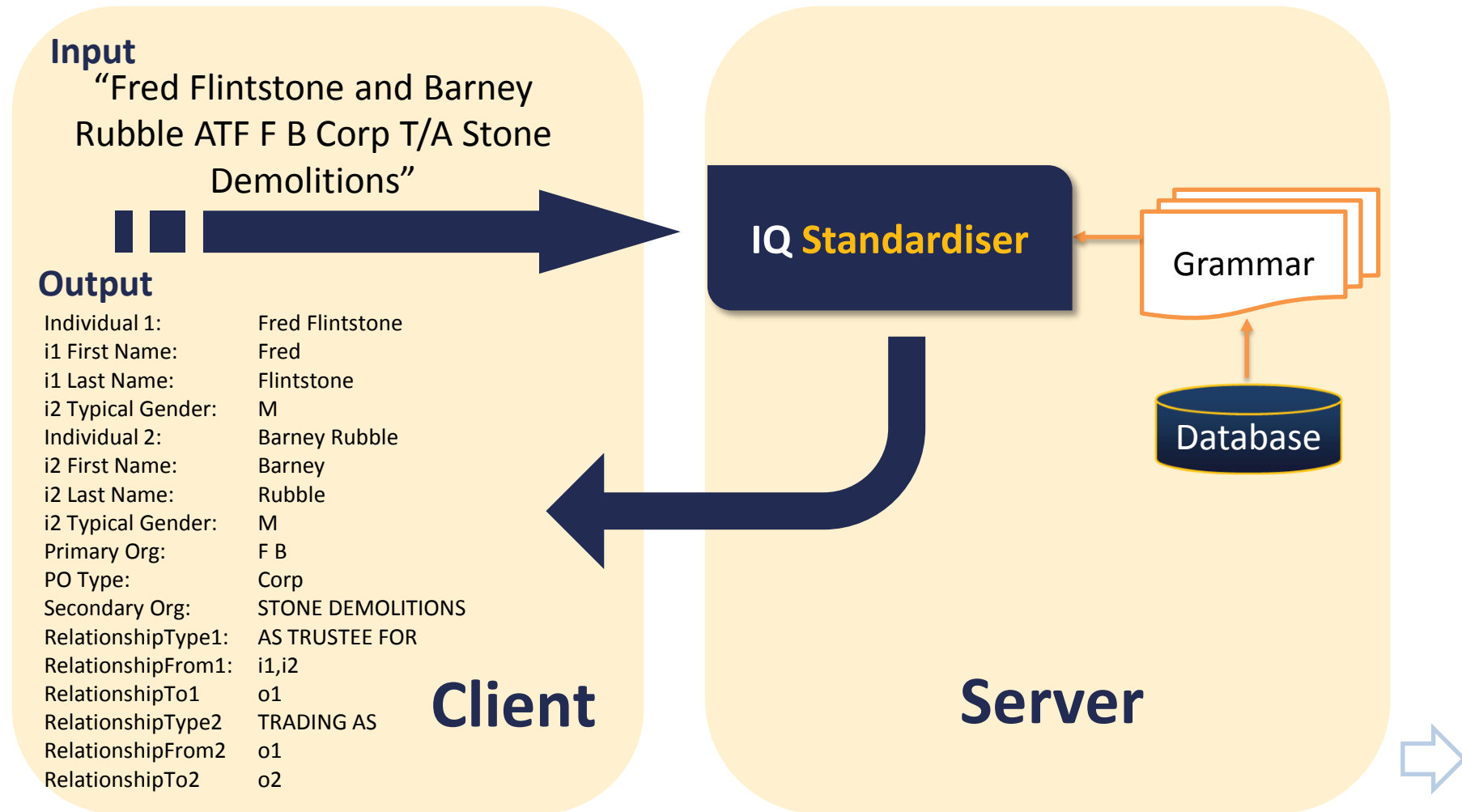- Transforms data of any type to custom defined standards.

**Parsing**

| Read Data | Accurately Identify Data Elements | Split Into Required Components | Detailed Standardiser Processing |

# IQ Standardiser

## Names Process

**Input**

"Mr Mark Jackson" → **IQ Standardiser**

Grammar

Database

**Output**

Title: "Mr"
First Name: "Mark"
Last Name: "Jackson"
Typical Gender: "M"

# Client

# Server

Integrated Information Quality

# IQ Standardiser

## Account Names Process

**Input**

"Fred Flintstone and Barney Rubble ATF F B Corp T/A Stone Demolitions"

**Output**

| | |
|---|---|
| Individual 1: | Fred Flintstone |
| i1 First Name: | Fred |
| i1 Last Name: | Flintstone |
| i2 Typical Gender: | M |
| Individual 2: | Barney Rubble |
| i2 First Name: | Barney |
| i2 Last Name: | Rubble |
| i2 Typical Gender: | M |
| Primary Org: | F B |
| PO Type: | Corp |
| Secondary Org: | STONE DEMOLITIONS |
| RelationshipType1: | AS TRUSTEE FOR |
| RelationshipFrom1: | i1,i2 |
| RelationshipTo1 | o1 |
| RelationshipType2 | TRADING AS |
| RelationshipFrom2 | o1 |
| RelationshipTo2 | o2 |

**Client**

**Server**

**IQ Standardiser**

Grammar

Database

**Integrated Information Quality**

# Identity Search

"I'm Spartacus!"

# Identity Search

**Auto-resolution of entity's across multiple sources**

- Execute a probabilistic matching algorithm
- Use all reference data to create matches
- Enables identification of unique identities, and relationships
- Stop duplicate records being introduced

| Name: | richard | | Bosley | DOB: | |
|---|---|---|---|---|---|
| Locality: | BONDI BEACH | | | | Match |

| Score | Id | First Name | Surname | Address |
|---|---|---|---|---|
| 29.64 | 999265 | Dick | Beasley | 8 Campbell Pde BONDI BEACH NSW 2026 |
| 27.03 | 993789 | Richard | Beasley | 42 DAYSDALE WAY THURGOONA 2640 |
| 25.32 | 242759 | Richard | Bosse | 20 NOLAN PL BALGOWLAH HEIGHTS 2093 |
| 24.07 | 242760 | Richard | Bosse | WIDGELLI 2680 |
| 22.03 | 993788 | Richard | Beasley | 13 GANGES PL BEECHBORO 6063 |

Integrated Information Quality

# Advanced Search

**Keywords, defined terms, single entity and multiple entities**

- Probabilistic / "Fuzzy"
- Quickly and accurately locate entities
- Highly scalable, flexible
- Can be configured to be used to search within your data Platform
- Can be configured to search other databases

# Federated Search

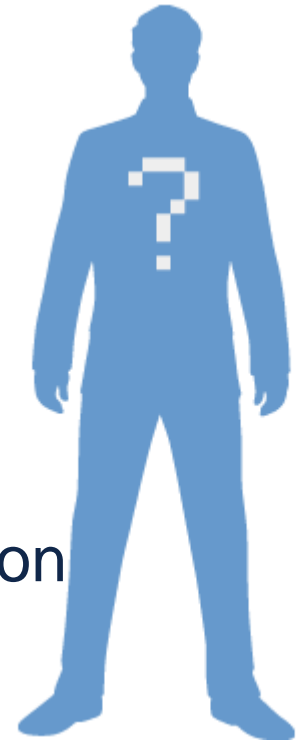**Use within custom build applications or standard CRM and ERP's**

- Advanced matching makes it easy to locate customer records.
- Multi-lingual and foreign name support
- Intelligent sorting ensures most relevant record is presented first.
- Extensive alias name support
- Ability to fine tune to any dataset and business rules

# Enhanced Identity Searching

**Benefits**

- Improved customer service
- Decreased data duplication
- Better fraud detection
- Higher cross selling opportunities
- Fosters operational efficiencies
- Improved analytical capabilities
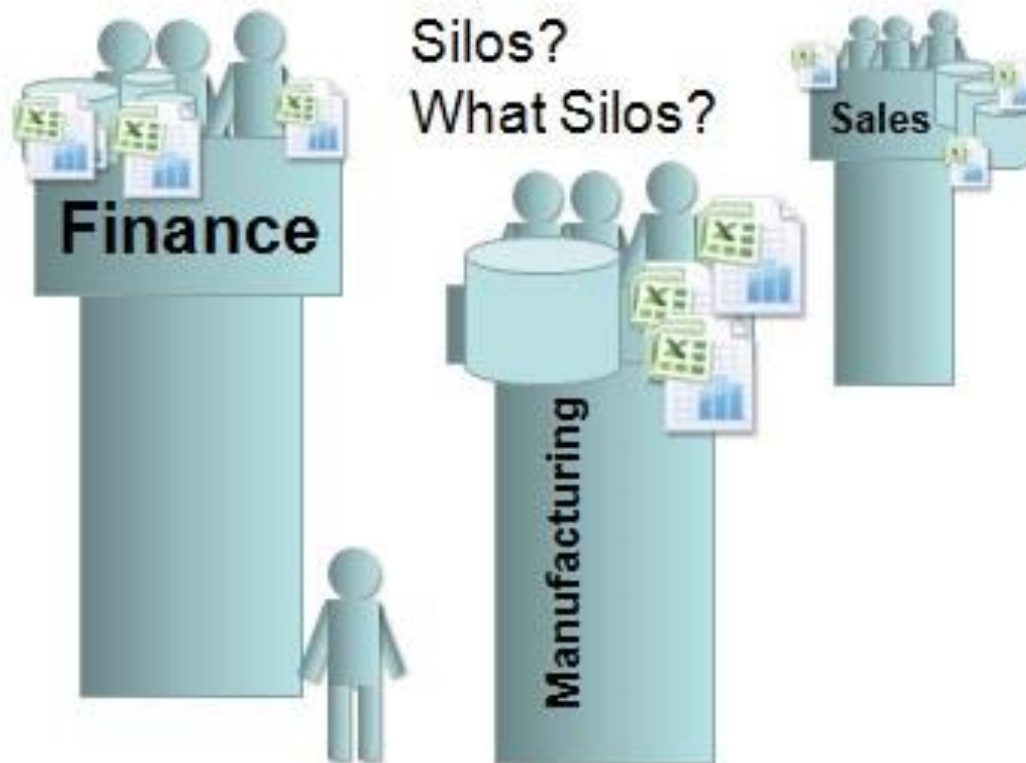- Increased customer acquisition and retention

# Questions & Discussion

**Intech Solutions**

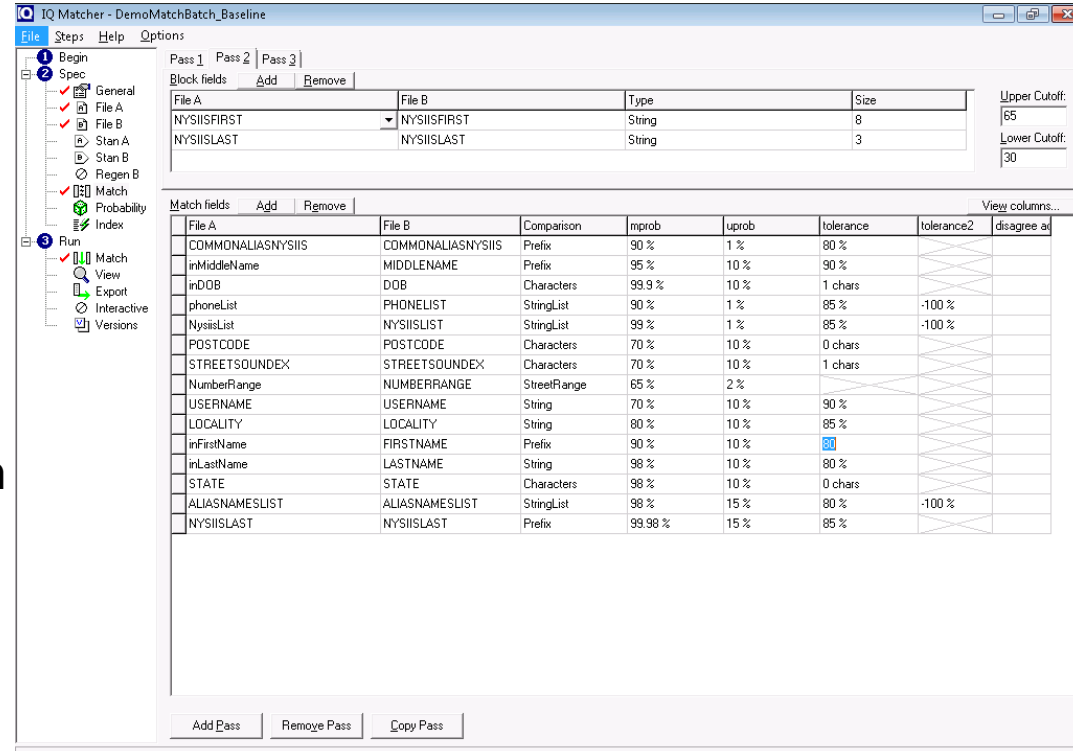Information Quality
Supporting
Identity Search

**Identity Search**

IQ

Integrated Information Quality

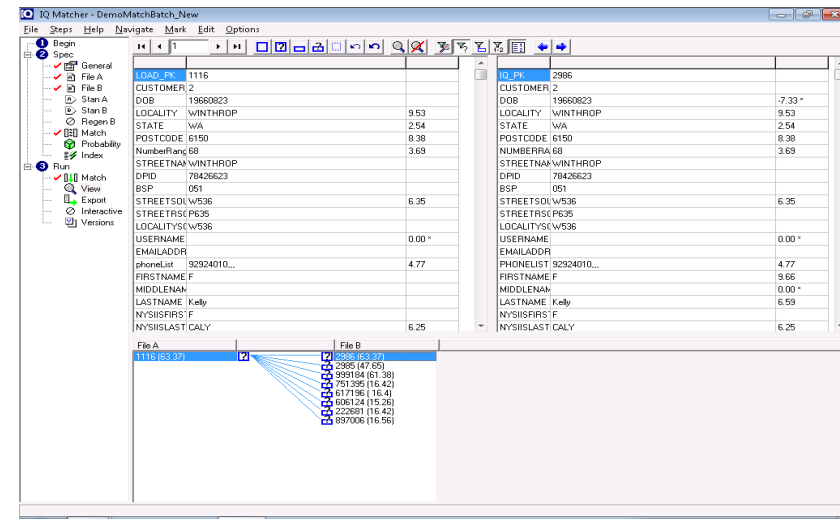# Data Matching

# Data Matching

## Various Types

- Loose/Fuzzy Match

- Measures Strength of Match

- Real Time, Across Many Data Sources

- Standardise, Match

- Cleanse And Integrate Data From Multiple Sources

- Detect And Correct Duplicates
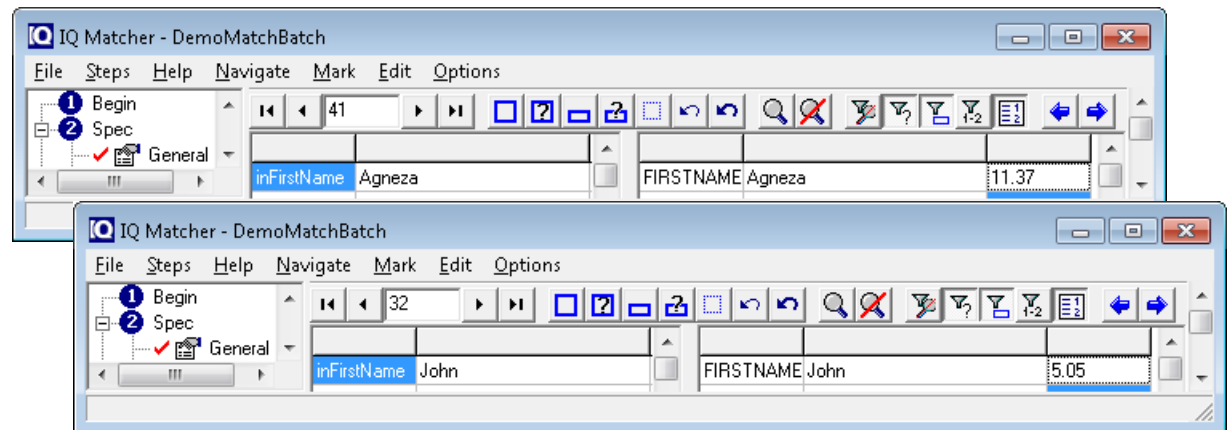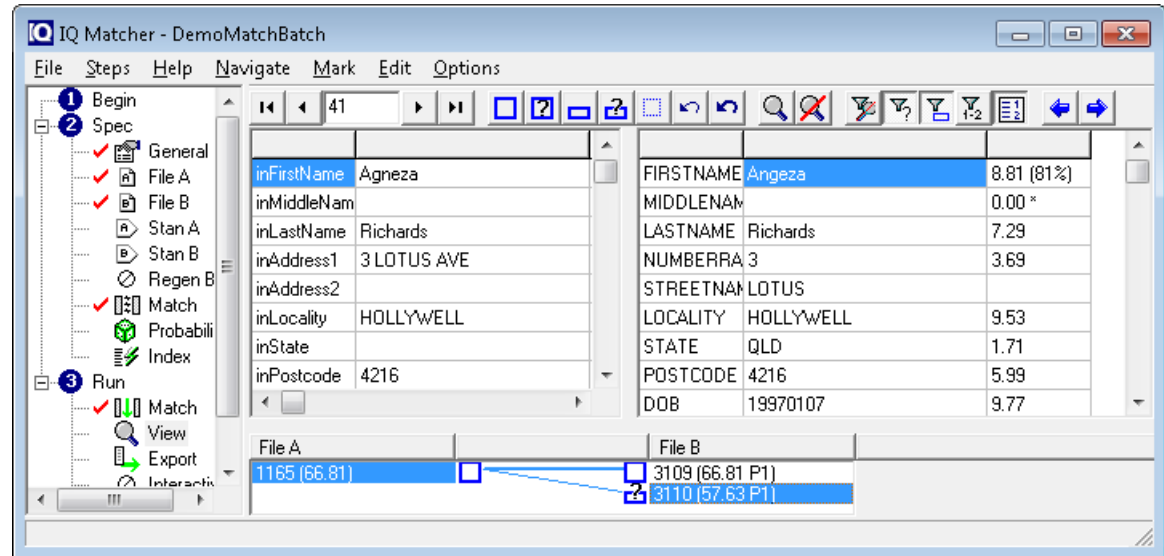
# Data Matching

## IQ Matcher

- High Performance Record Matching
- Designed Specifically For Large-scale Databases
- Enterprise-wide Record Matching
- Real Time, Across Many Data Sources
- Varying Type Of Record Fields And/Or Disparate Information
- Misspellings, Address Variations And Other Inconsistencies
- Advanced Probabilistic Record Linking Statistical Algorithm

# Data Matching

## IQ Matcher – GUI Example

- Match File A to File B
- Match is sensitive to your data
- John is more common than Agneza, so lower match score for John

Integrated Information Quality

# Questions & Discussion

Information Quality
Supporting
Data Matching

**Data Matching**

**Integrated Information Quality**

# Conclusion

# The Intech Difference

- The Details Matter, We Get It Right

- Australian & New Zealand Specialisation
  – Local Support
  – Local Development
  – Local Data Expertise

- Validate Against Any Reference Data Source
  – Eg Address, Agency specific regions/boundaries

- Open Architecture Designed for Configuration and Enterprise Integration

Integrated Information Quality

# Questions

Intech Solutions

**Intech Solutions**

Prime Property House
2 Woodward Street
Wellington 6142
intechsolutions.co.nz

☎ +64 4 499 5414

## Elizabeth Eastwood

☎ +64 4 499 5414
🕽 +64 21 404 035
✉ elizabethe@intechiq.com