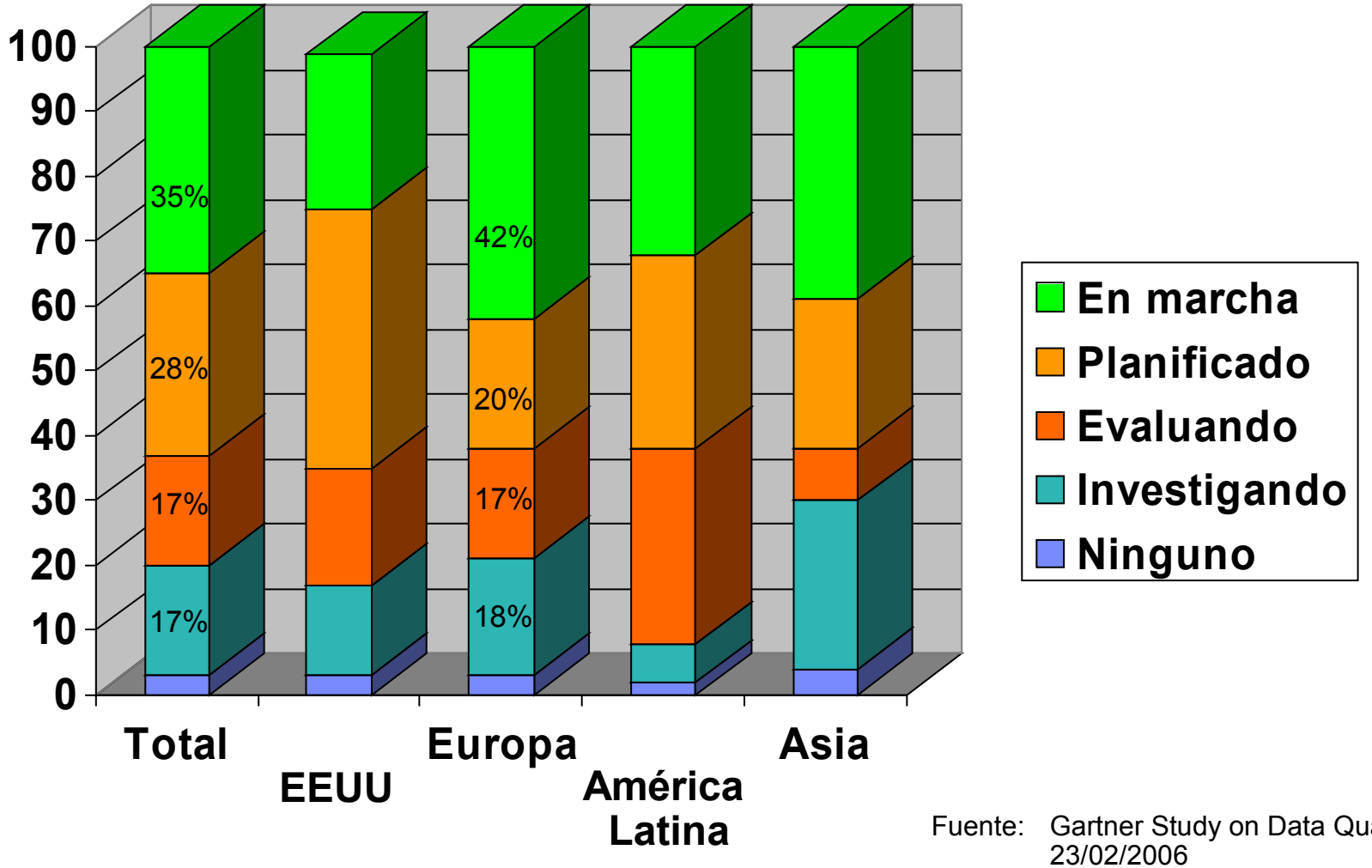
Consiga datos fiables

Philip Little
IBM Information Integration Solutions

Agenda

- Situación actual
- Los procesos para mejorar la calidad de datos

Un problema reconocido

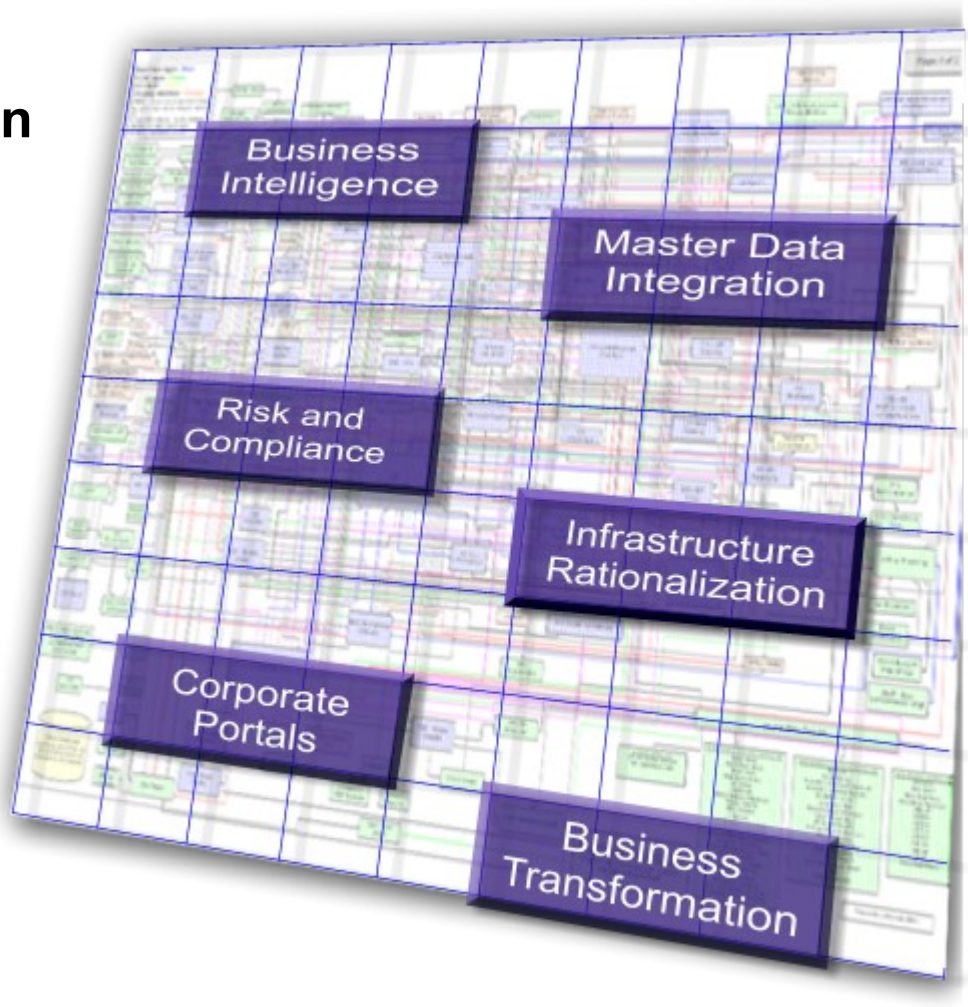


Fuente: Gartner Study on Data Quality 23/02/2006

Iniciativas de negocio críticas que necesiten Información Integrada de Calidad

Todas estas iniciativas necesitan información que sea:

- **Accesible**
- **Fiable / Correcto**
- **Consistente**
- **A Tiempo**
- **Completa**



Calidad de Datos

Los Mitos de la Calidad

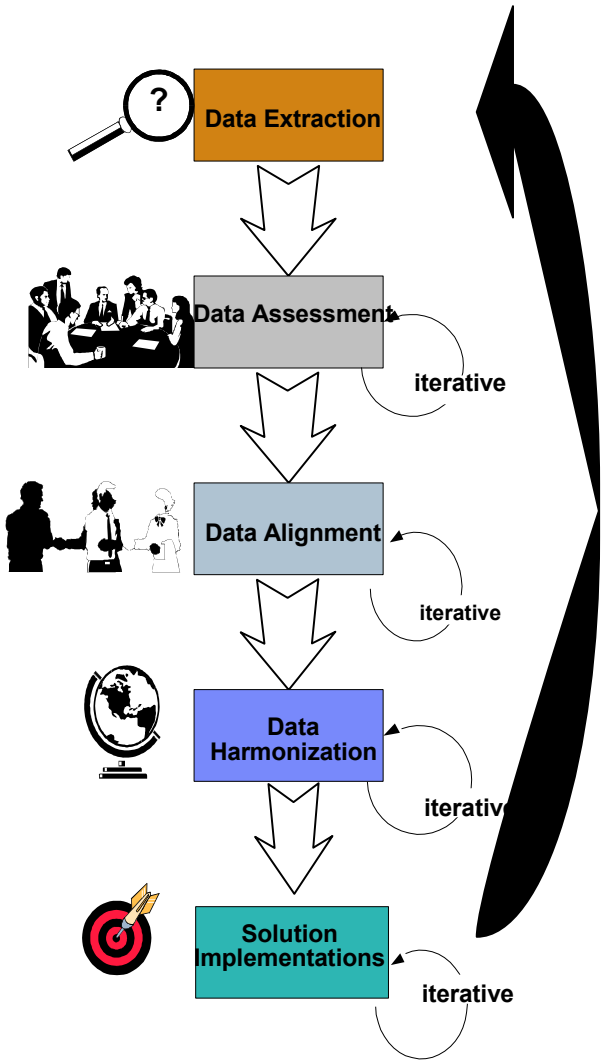
- **La Calidad de Información es para TI**
  - Muy ligada al negocio y sus reglas
- **La Información está bien o mal**
  - La calidad de información depende del uso / necesidad
- **Lo depuramos el año pasado - ¡todo está bien!**
  - La calidad de información es dinámica, afectado por personas, procesos y aplicaciones
- **Conocemos nuestros datos**
  - La documentación existe
  - Los datos corresponden a sus meta datos
- **Las reglas de negocio no han cambiado**

La realidad de los Datos

- La mayoría de organizaciones tienen **distintas aplicaciones** para las áreas de ventas, servicios, marketing, manufacturing, y financiero – **cada una con sus propios "maestros"**
- **No existe un consenso sobre un registro común** de sistema
- Hay que "**re-dirigir**" datos "**antiguos**" a las **nuevas aplicaciones**
- No es sólo un problema de entrada de datos, es un **problema de integración / reconciliación**
- Es **demasiado tarde** y **demasiado caro** arreglar los datos **después de la implantación**

El Control de Calidad necesita

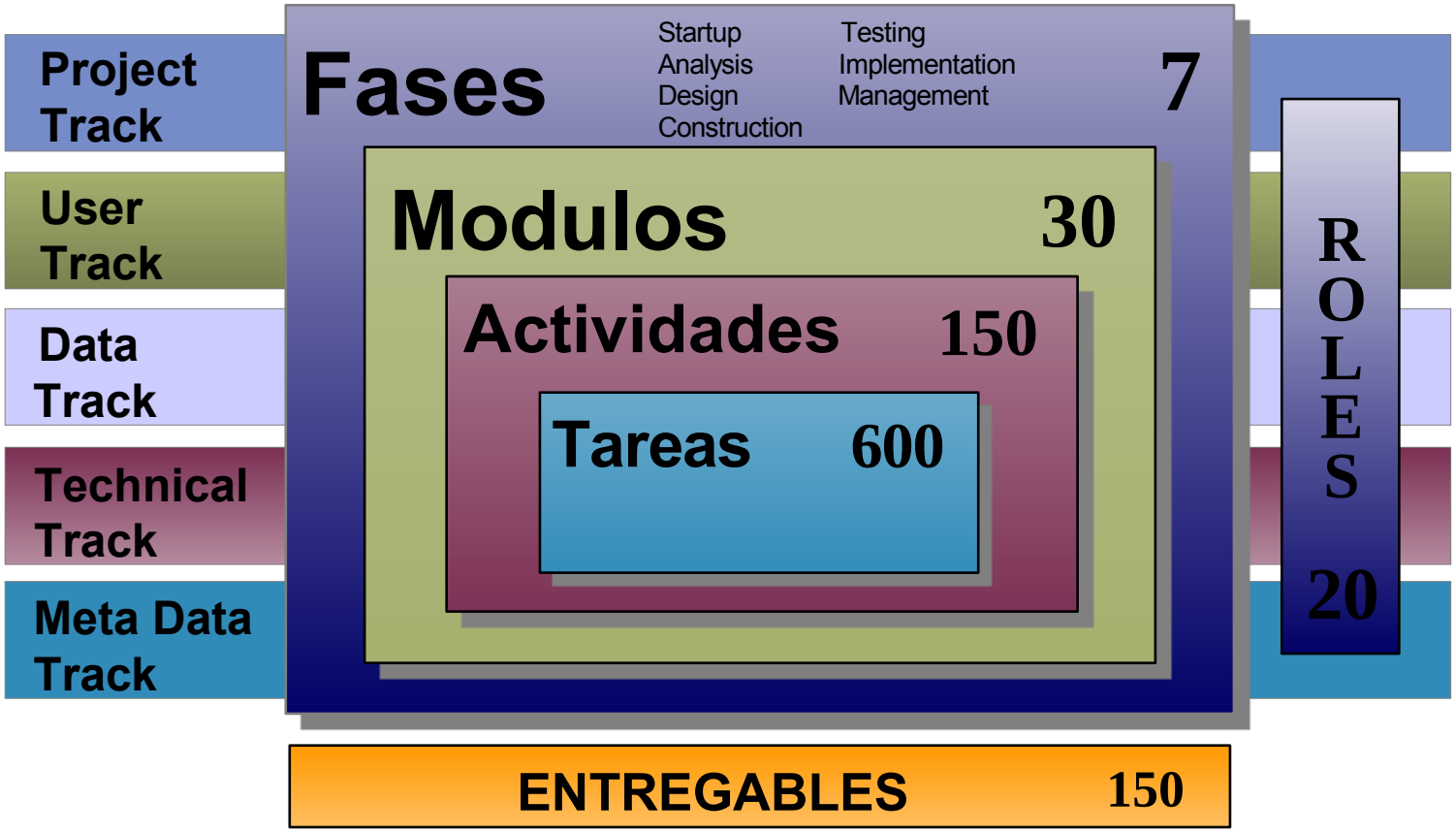
- Estándares para los datos de negocio
- Responsables para los datos de negocio
- Calidad de Información
- Manejo de Meta datos
- Planificación de migración
- Seguridad y Protección de Datos
- Procesos estándares en el uso de datos en el desarrollo de aplicaciones



Productos + Metodología

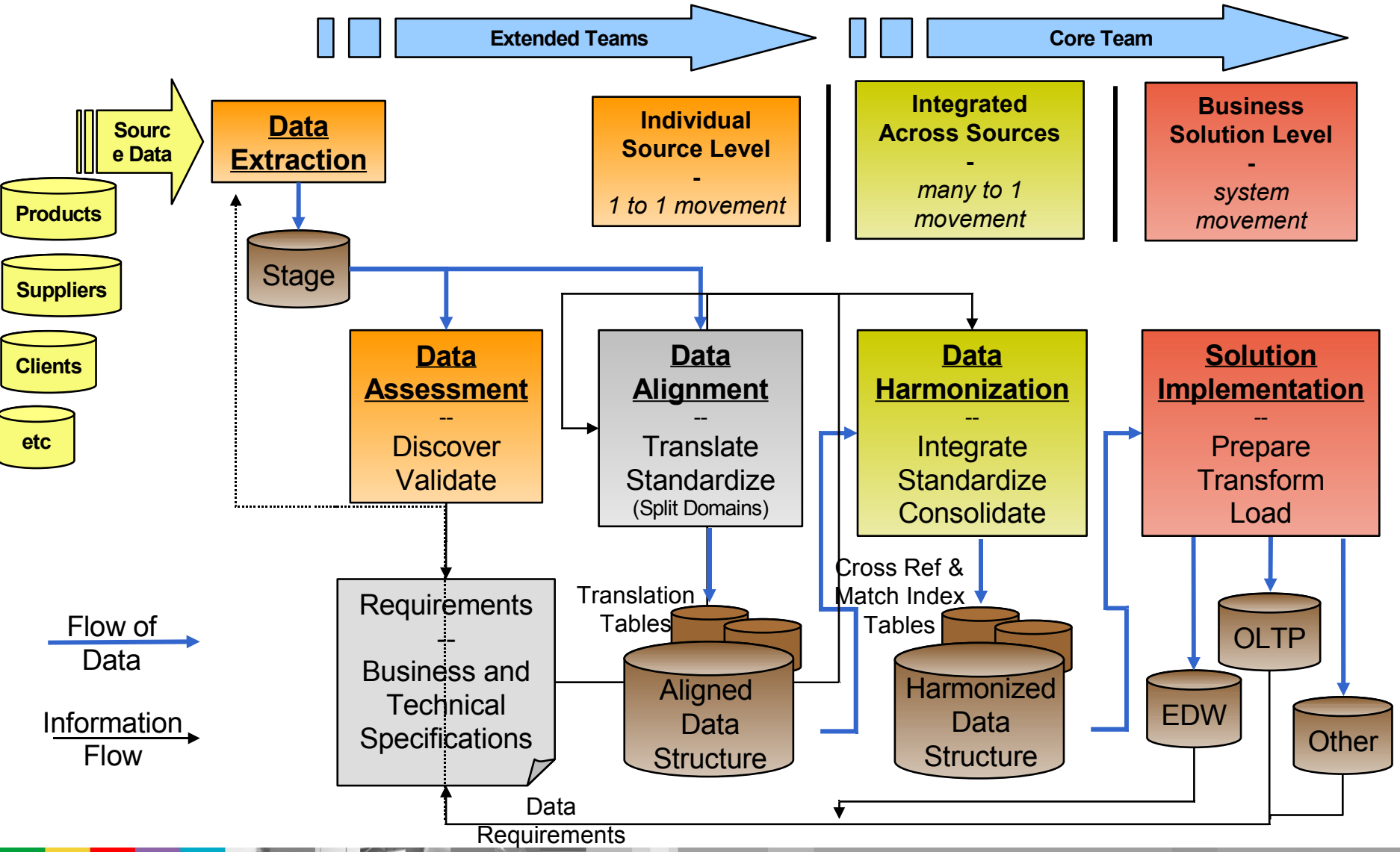
Pre-Project | Strategic Planning | Startup | Analysis & Design | Build | Test & Implement | Monitor & Evaluate

Metodología: Iterations 2



Pre-Project → Strategic Planning → Startup → Analysis & Design → Build → Test & Implement → Monitor & Evaluate

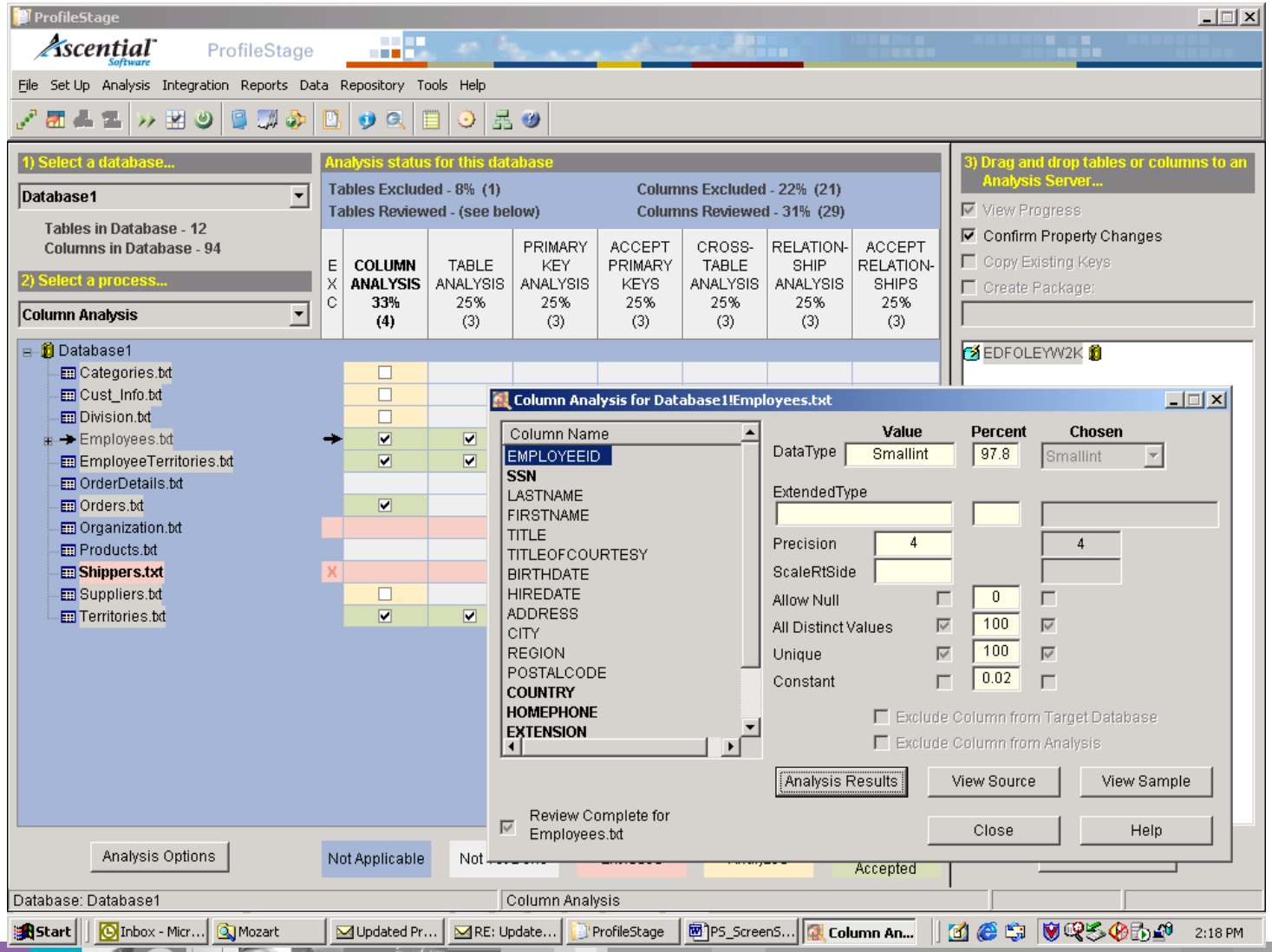
Integración de Datos – el proceso en acción !



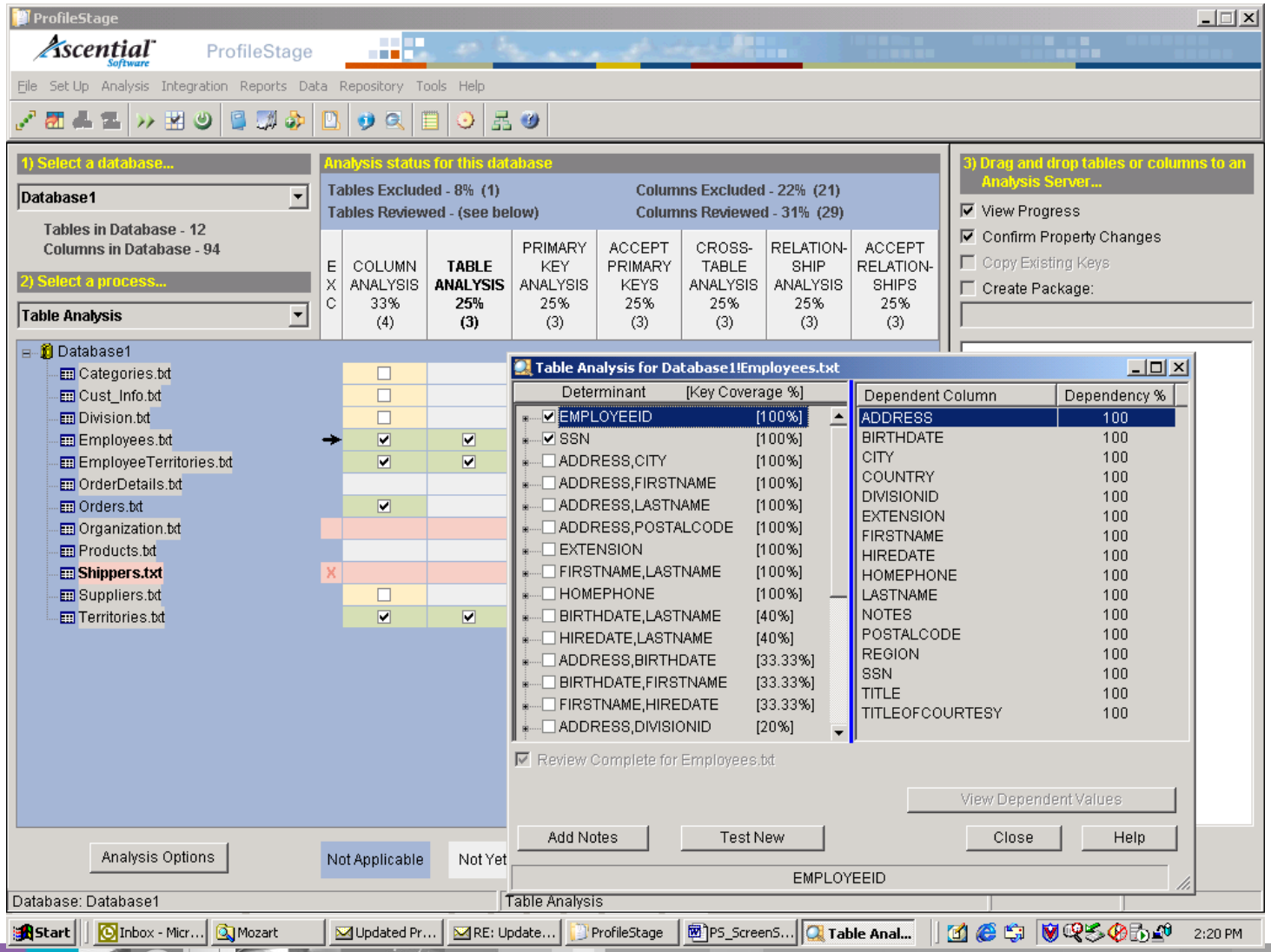
Algunas Tipos de Anomolías

Información Incorrecta	La definición o contenido no está alineada con lo que necesita el proceso de negocio
Campos Inconsistentes	El contenido del campo es inconsistente entre registro y registro. El mismo valor en un campo significa cosas distintas
Sentidos Múltiples	El mismo campo contiene múltiples tipos de información dependiendo del registro
Campos Duplicados	Dos campos distintas con nombres distintas contienen la misma información y sentido
Sentidos Compuestos	El campo tiene códigos múltiples compuestos
Correlación Física Errónea	La características físicas del fuente no corresponden con el destino
Información no Disponible	Falta información obligatoria según las reglas de negocio
Formatos Inconsistentes	El contenido tiene múltiples formatos, dificultando su entendimiento y proceso
Valore Permitidos Inconsistentes	El contenido puede conformar a distintos valores permitidos
Valores Duplicados	No deben existir valores duplicados, pero existen

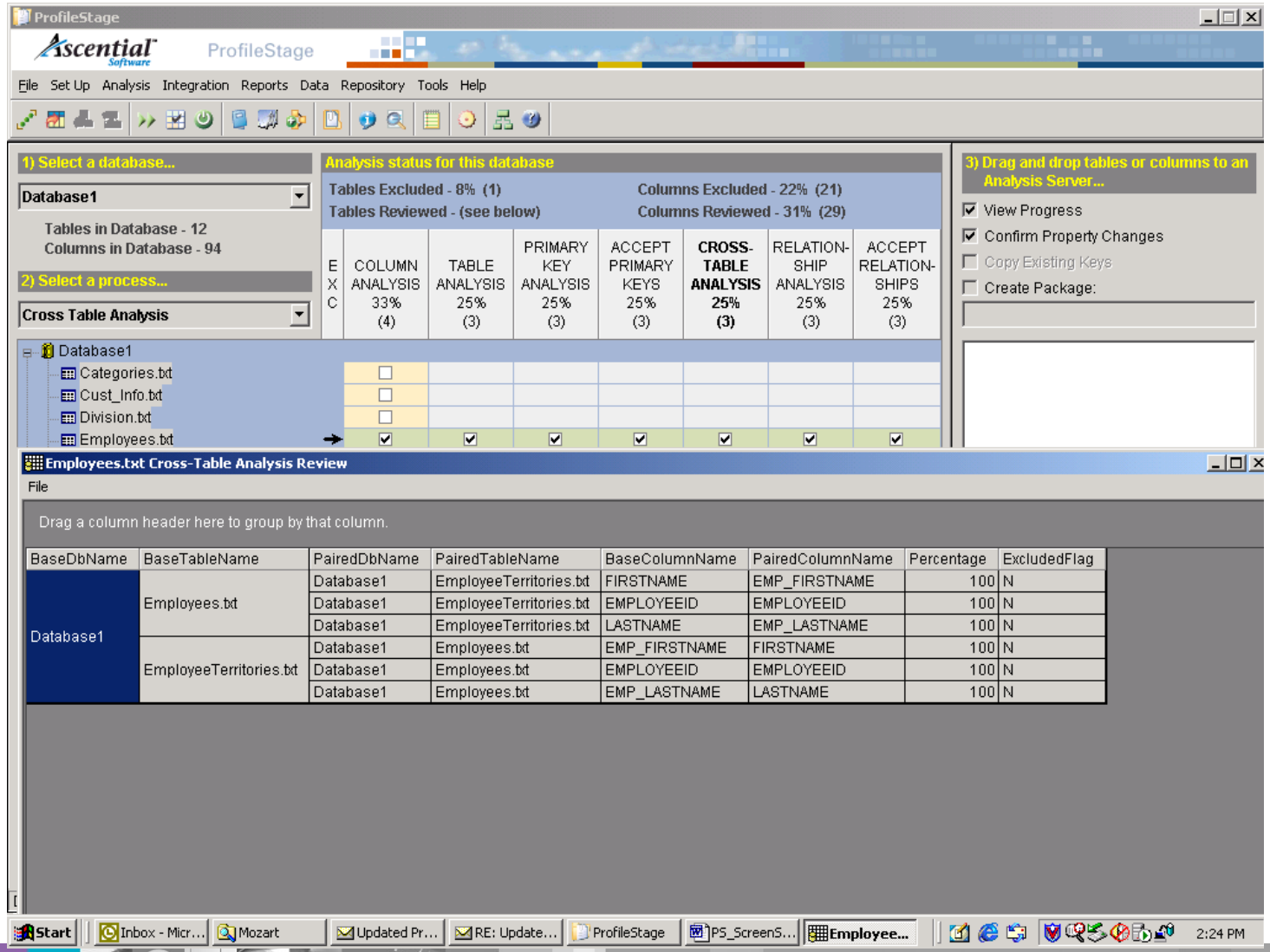
Análisis de Columnas



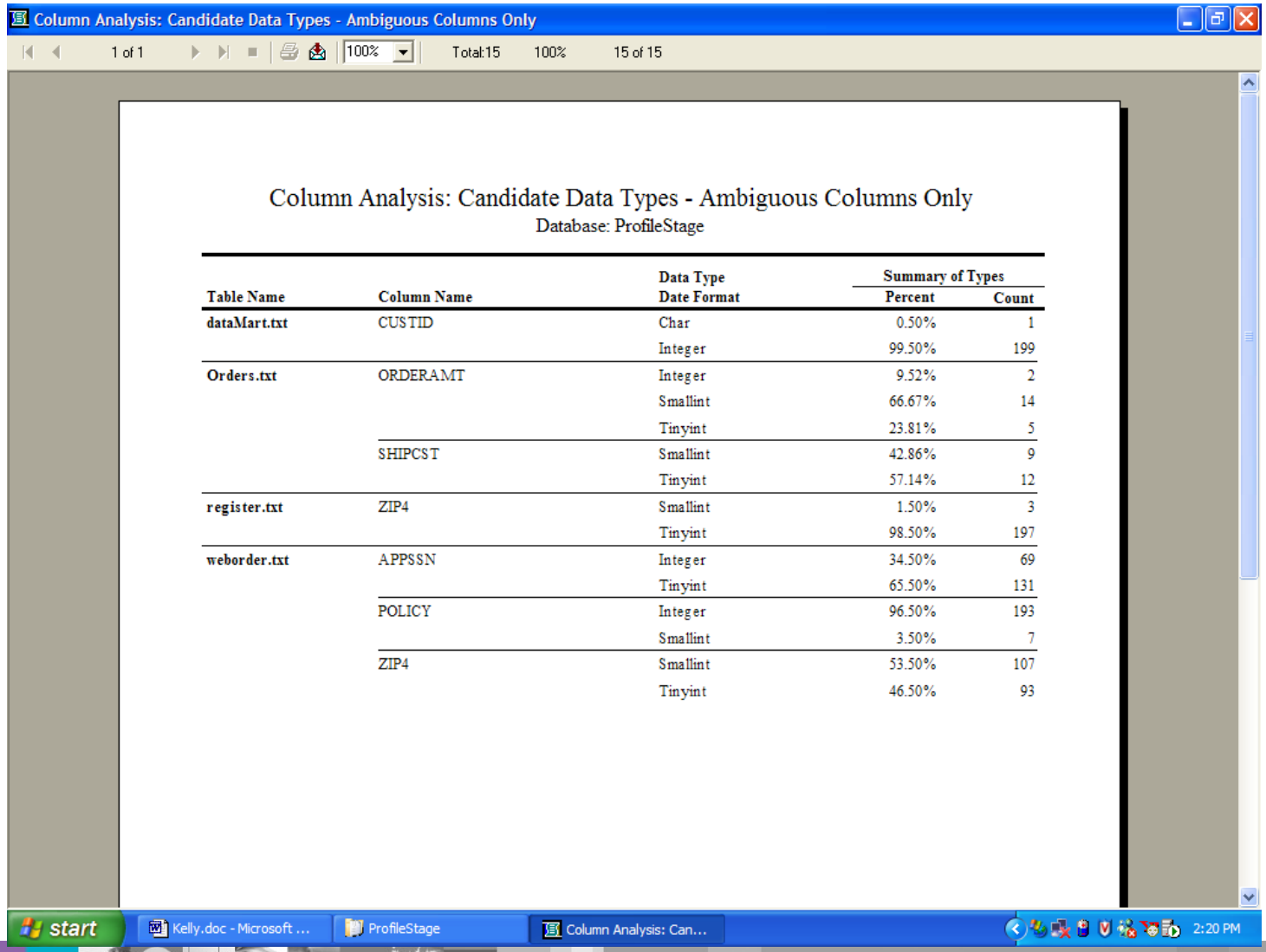
Análisis de Tablas



Análisis de Tablas Cruzadas



Informes



Column Analysis: Candidate Data Types - Ambiguous Columns Only

Database: ProfileStage

Table Name	Column Name	Data Type / Date Format	Summary of Types: Percent	Summary of Types: Count
dataMart.txt	CUSTID	Char	0.50%	1
		Integer	99.50%	199
Orders.txt	ORDERAMT	Integer	9.52%	2
		Smallint	66.67%	14
		Tinyint	23.81%	5
	SHIPCST	Smallint	42.86%	9
		Tinyint	57.14%	12
register.txt	ZIP4	Smallint	1.50%	3
		Tinyint	98.50%	197
weborder.txt	APPSSN	Integer	34.50%	69
		Tinyint	65.50%	131
	POLICY	Integer	96.50%	193
		Smallint	3.50%	7
	ZIP4	Smallint	53.50%	107
		Tinyint	46.50%	93

Tendencias en Calidad

¿Cómo se puede obtener una visión del negocio consolidada y acertada?

1: Investigación - Palabras

123 St. Virginia St.

Parseo:

123 | St. | Virginia | St.

Separación de campos con múltiples valores en piezas individuales

Análisis Léxico:

Number | Street Type | Alpha | Street Type

123 | St. | Virginia | St.

Determinar el significado de negocio de piezas individuales

Sensitivo al contexto:

House Number | Street Name | Street Type

123 | St. Virginia | St.

Identificando estructuras de datos y contenidos variados

"Las instrucciones para manejar los datos son inherentes a los datos mismos"

2: Normalización - Direcciones

Fichero Entrada:

Dirección Línea 1	Dirección Línea 2
639 N MILLS AVENUE	ORLANDO, FLA 32803
306 W MAIN STR, CUMMING, GA 30130	
3142 WEST CENTRAL AV	TOLEDO OH 43606
843 HEARD AVE	AUGUSTA-GA-30904
1139 GREENE ST ACCT #1234	AUGUSTA GEORGIA 30901
4275 OWENS ROAD SUITE 536 EVANS	GA 30809
1775 RUSSELL CIRCLE MILLIS MASSACH	USETTS 02038

Fichero Resultado:

House #	Dir	Str. Name	Type	Unit	No.	NYSIIS	City	SOUNDEX	State	Zip	ACCT#
639	N	MILLS	AVE			MAL	ORLANDO	O645	FL	32803	
306	W	MAIN	ST			MAN	CUMMING	C552	GA	30130	
3142	W	CENTRAL	AVE			CANTRAL	TOLEDO	T430	OH	43606	
843		HEARD	AVE			HAD	AUGUSTA	A223	GA	30904	
1139		GREENE	ST			GRAN	AUGUSTA	A223	GA	30901	1234
4275		OWENS	RD	STE	536	ON	EVANS	E152	GA	30809	
1775		RUSSELL	CIR			RASAL	MILLIS	L260	MA		02038

2: Normalización - Materiales

Fichero de Entrada:

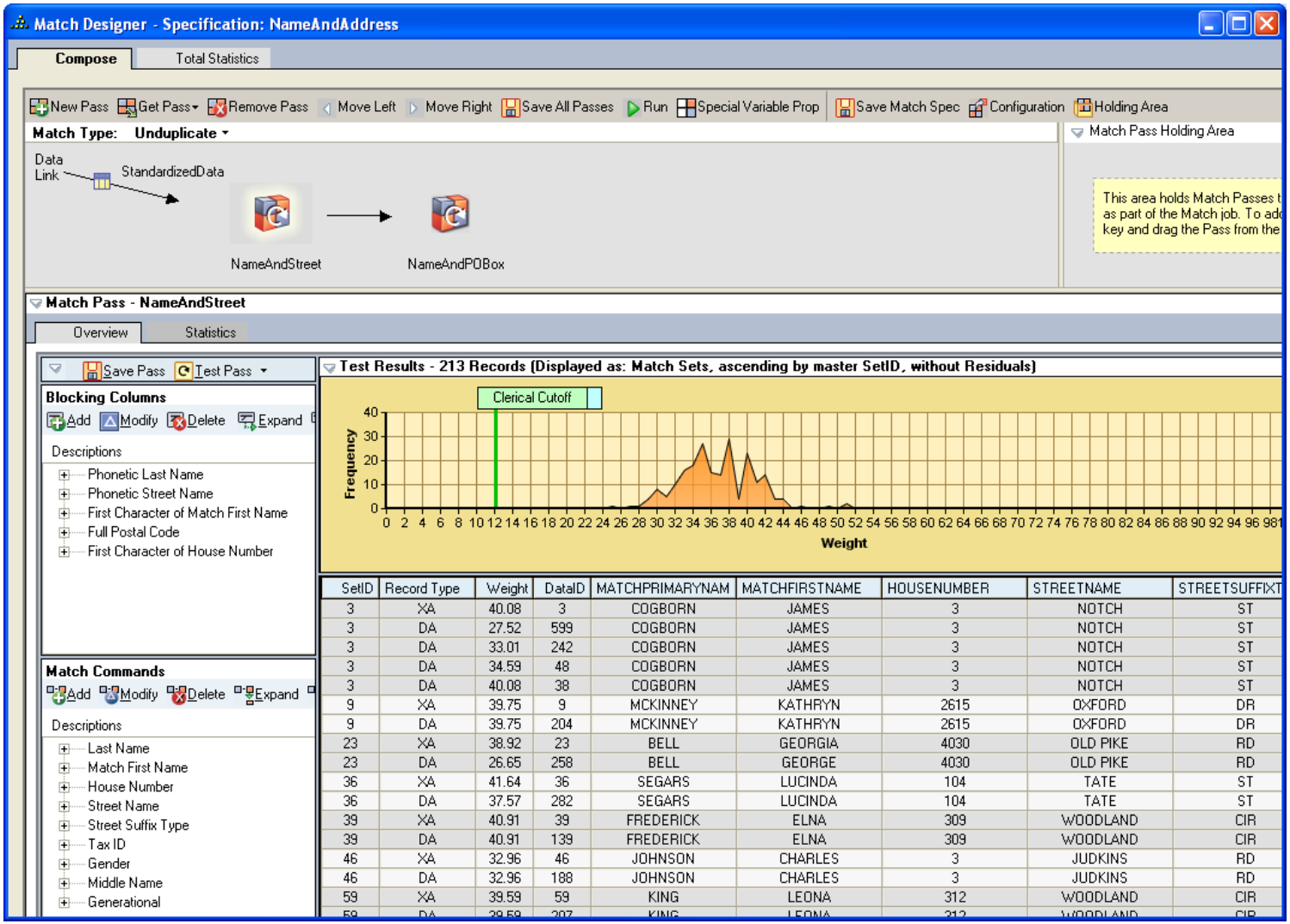
Operation Work Instruction

WING ASSY DRILL 4 HOLE USE 5J868A HEXBOLT 1/4 INCH
WING ASSEMBY, USE 5J868-A HEX BOLT .25" - DRILL FOUR HOLES
USE 4 5J868A BOLTS (HEX .25) - DRILL HOLES FOR EACH ON WING ASSEM
RUDER, TAP 6 WHOLES, SECURE W/KL2301 RIVETS (10 CM)

Fichero resultado:

Assembly	Instruction	QTY	Type	Part	Size	Unit Measure	SKU
WING	DRILL	4	HOLES	HEXBOLT	.25	INCH	5J868A
WING	DRILL	4	HOLES	HEXBOLT	.25	INCH	5J868A
WING	DRILL	4	HOLES	HEXBOLT	.25		5J868A
RUDDER	TAP	6	HOLES	RIVET	10	CM	KL2301

3: Emparejamiento



4: Supervivencia - Piezas

Entrada para supervivencia (Salida del Match)

Group	Record	Operation Work Instruction
1	1	WING ASSY DRILL 4 HOLE USE 5J868A HEXBOLT 1/4 INCH
1	2	WING ASSEMBY, USE 5J868-A HEX BOLT .25" - DRILL FOUR HOLES
1	3	USE 4 5J868A BOLTS (HEX .25) - DRILL HOLES FOR EACH ON WING ASSEM
2	4	RUDER, TAP 6 WHOLES, SECURE W/KL2301 RIVETS (10 CM)

Registros consolidados

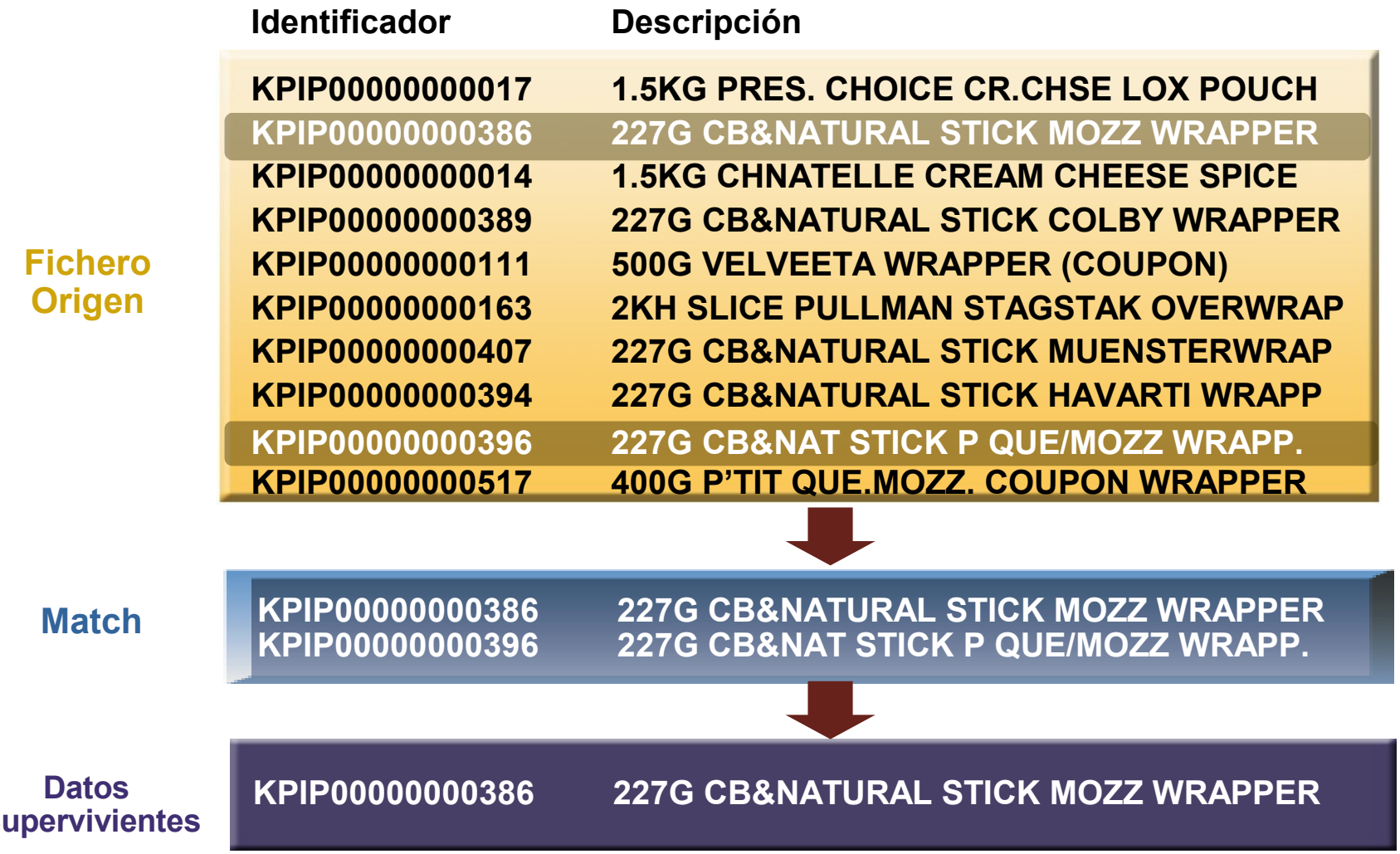
Assembly	Instruction	QTY	Type	Part	Size	Unit Measure	SKU
WING	DRILL	4	HOLES	HEXBOLT	.25	INCH	5J868A
RUDDER	TAP	6	HOLES	RIVET	10	CM	KL2301

Clave de referencia cruzada

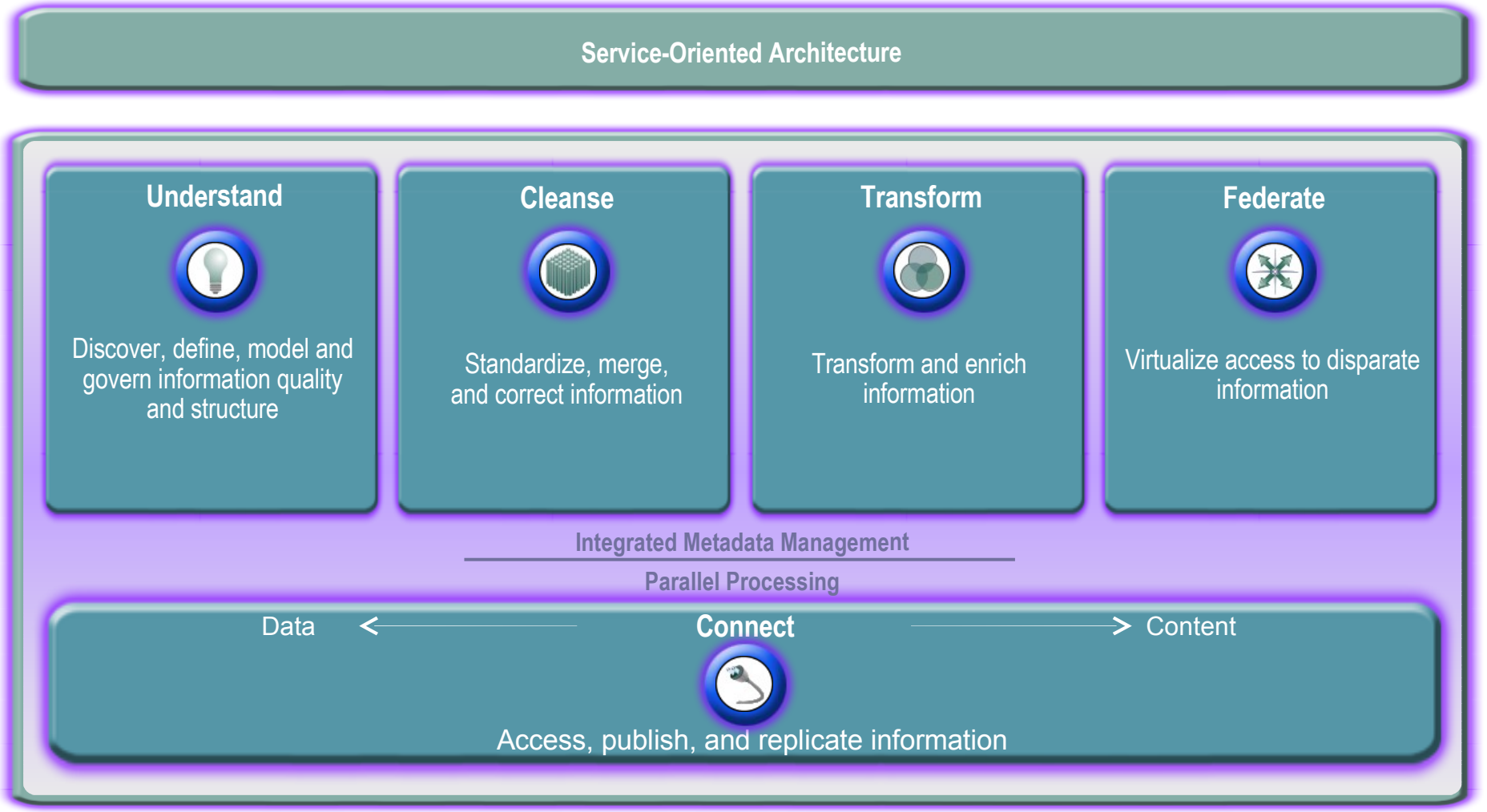
Group	Assembly	Instruction	QTY	Type	Part	Size	Unit Measure	SKU
1	WING	DRILL	4	HOLES	HEXBOLT	.25	INCH	5J868A
2	RUDDER	TAP	6	HOLES	RIVET	10	CM	KL2301

Record	Group
1	1
2	1
3	1
4	2

4: Supervivencia - Datos de Productos



The IBM WebSphere Information Integration Platform




Thank
You