



The value of Metadata for ETL and beyond

Steve Morton

Applied System Knowledge Ltd



What Is Metadata?

- "Data about Data"
- And about Processes
- Describes your Data Warehouse -
Extraction, Transformation, Load,
Published Content and Exploitation
- Your Second-most-important
Information Asset (after your data
warehouse itself)



More specifically

- All information that describes data
 - Structure and format
 - Meaning of data content, how to use it
 - Where it comes from, how it is processed in the d.w. (*Lineage*)
 - Other relevant information, such as
 - When it was created / loaded
 - Quality assessment
 - Ownership, stewardship – original systems
 - And so on



Why use Metadata at all?

- Isn't it just more documentation??
 - Costly 'shelfware', rapidly out of date...
- Value comes from *actively using* information you have gathered anyway
 - In managing the d.w. (technical value)
 - In supporting business use of the data (business value)

Technical metadata

- Basic data description
 - Name & location of tables
 - Structure – column names, data types
 - Relationships (data model)

- Process metadata
 - Alternative to writing code wherever possible – generate from metadata
 - Wrapper for written code where necessary

(c) 2004 Applied System Knowledge Ltd.

5

Tables and Processes in SAS ETL Studio

The screenshot displays the SAS ETL Studio interface. On the left is the 'Repositories' tree showing a hierarchy of data groups and tables. The main workspace shows a 'Process Designer' with a flow diagram involving a 'Splitter: M Split Policyholder a' and several 'Loader' tasks connected to data tables like 'Client_and_agent_details Data Table'. A 'Process Editor' window is open, showing the 'Client_and_agent_details Data Table Properties' dialog. This dialog has a 'Columns' tab with a table listing the table's schema.

#	Name	Description	Length
1	BIRTHPLACE	BIRTHPLACE	16C
2	BIRTH_DATE	BIRTH_DATE	8N
3	BUSINESS_CODE	BUSINESS_CODE	4C
4	CHILDREN	CHILDREN	3C
5	CLIENT_CLASS	CLIENT_CLASS	1C
6	CLIENT_REFERENCE	CLIENT_REFERENCE	10C
7	CLIENT_STATUS	CLIENT_STATUS	1C
8	CONTACT_1	CONTACT_1	40C
9	CONTACT_1_LOC_REF	CONTACT_1_LOC_REF	10C
10	CONTACT_1_FAX	CONTACT_1_FAX	16C
11	CONTACT_1_PHONE_NO	CONTACT_1_PHONE_NO	16C
12	CONTACT_2	CONTACT_2	40C
13	CONTACT_2_LOC_REF	CONTACT_2_LOC_REF	10C
14	CONTACT_2_FAX	CONTACT_2_FAX	16C
15	CONTACT_2_PHONE_NO	CONTACT_2_PHONE_NO	16C

(c) 2004 Applied System Knowledge Ltd.

6



What does it cost?

- Mainly time you have to spend anyway (sunk cost)
 - Information gathering
 - Understanding data and business rules
 - Recording and checking details
- Small addition to ensure it is recorded when building processes
 - Keep together the data and process definitions

(c) 2004 Applied System Knowledge Ltd.

7



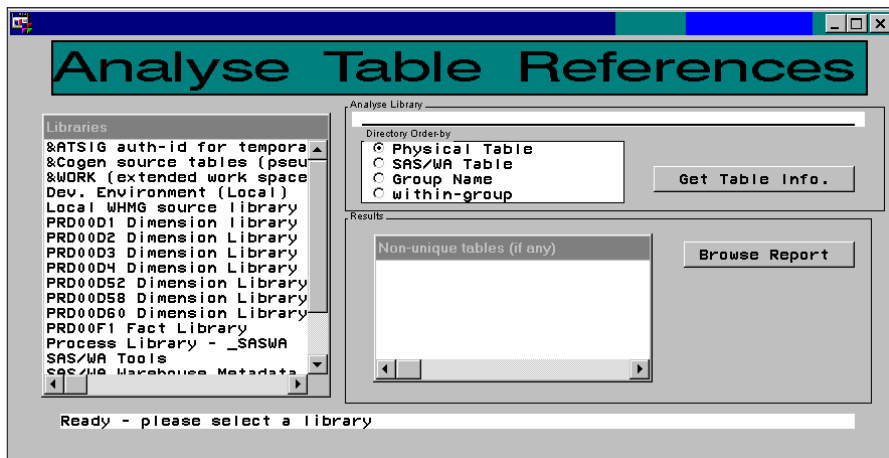
Process metadata: What is it worth?

- Confidence that 'metaprocess' fully describes the real processing
 - Reliable maintenance - easy to locate and change the right process
 - True lineage from source systems to the published data warehouse
- Opportunity to 'mine' the metadata
 - Advanced analysis and reporting using API calls

(c) 2004 Applied System Knowledge Ltd.

8

Example: Cross-reference Analysis and Reporting



(c) 2004 Applied System Knowledge Ltd.

9

Applications of metadata

- Cross-reference where any key resource is used
 - Libraries, tables, processing hosts
- Gain overview of process details
 - End-to-end, without needing to read code
- Customised publishing
 - User-oriented, browsable catalogue of available data

(c) 2004 Applied System Knowledge Ltd.

10



Limitations of SAS 8 metadata

- Application based, separate storage
 - ETL(WA), BI (EIS), Data Mining, Portal, ...
- Differing formats
 - No single API – sometimes no API at all
 - Need to “export” it
 - Difficult to share access to it
 - Problem with having a “single version of the metadata truth”



Advances in SAS 9 metadata

- Integrated metadata model
 - No more ‘export to ...’ other tools!
- Easier metadata in ETL Studio
 - Default mappings just happen
 - Better tools and interface
- Better & more flexible XML-based API
 - Call from a choice of languages – Java, Visual Basic, C++, SAS data step
 - Extensions such as ‘search’ calls – no longer need object ID to find an object



Integrated metadata

- Everything really is connected to everything else
- Metadata server – fast, shared service
- Tools and interfaces surface appropriate aspects of the metadata
 - ETL processes (ETL Studio)
 - Environment management (SMC)
 - BI / reporting (Information Maps, Portal)



Easier to define the basics

- 1-to-1 mappings created automatically when table is dropped
 - A big time-saver compared to SAS/WA
- More built-in transformations and loaders, such as Type 2 Dimension load and update
 - Much less need for user-code programming
 - Pre-defined, consistent updates

Improved interface to navigate ETL metadata

The screenshot shows the SAS ETL Studio interface. On the left is a 'Shortcuts' pane with icons for Source Designer, Metadata Importer, Metadata Exporter, Process Designer, Target Designer, and Options. The main window is divided into two panes. The left pane shows an 'Inventory' view of tables, listing various data tables such as 'Agent Dimension Data Table', 'Agent_full_details Data Table', 'Agent_refn Data Table', etc., in alphabetical order. The right pane shows a 'Custom' view of folders, including 'Client ETL Data Group (Tables)', 'Client Extract Data Group (Tables)', 'Client Transform Data Group (Tables)', 'Control tables', 'Jobs', 'Locality ETL Data Group (Tables)', 'ODD's: source groups', 'Policy-ETL Data Group (Tables)', 'Policy-Premium Subject (Tables)', 'Ungrouped', and 'Foundation'. A green callout bubble points to the table list with the text: 'Alphabetic Inventory as well as Custom folders view'. At the bottom left, there is a small 'Inventory' icon and a 'Process Library' icon. At the bottom right, there is a 'Custom' icon.

(c) 2004 Applied System Knowledge Ltd. 15

The screenshot shows the SAS ETL Studio interface with the 'Reverse Impact Analysis' window open. The left pane shows the same 'Inventory' view of tables as in the previous slide. The main window is divided into two panes. The left pane shows a tree list of the 'Reverse Impact Analysis' for the 'Coverage Dimension Data Table'. It includes a 'Loader: Coverage Dimension Data Table', an 'SQL Join: U: Update Coverage dimension', and several 'Branch' nodes. The right pane shows a graphical view of the same analysis, with nodes connected by arrows. A green callout bubble points to the graphical view with the text: 'Impact Analysis with tree list as well as graphical view (same for Reverse Impact Analysis)'. At the bottom left, there is a 'Report' icon and a 'Graph' icon.

(c) 2004 Applied System Knowledge Ltd. 16

Reverse Impact Analysis - Coverage Dimension Data Table

Coverage Dimension Data Table

- 5. Load/Update Star Schema
- Loader: Coverage Dimension Data Table
- SQL Join: U: Update Coverage dimension
 - Branch 1
 - Coverage_nk Data Table
 - Branch 2
 - Coverages Data Table
- 4. Transformations
 - Loader: Coverages Data Table
 - Extract: M: Coverage columns
 - Coverage_details Data Table
 - 1. Latest Policy changes and assoc
 - Loader: Coverage_details Data Table
 - SQL Join: J: Select coverage de
 - Branch 1
 - POL_DET_C ODD table
 - Branch 2
 - New_pol_trans Data Tabl

Coverage Dimension Da...

Update

5. Load/Update Star Schema.S...

Coverage_details Data Table Properties

General Columns Indexes Keys Physical Storage Notes Extended Attributes Advanced

#	Name	Description	Length	Type	S
1	ANN_BASE_RATE_PREM	ANN_BASE_RATE_PREM	8	Numeric	(N)
2	ANN_COMMISSION	ANN_COMMISSION	8	Numeric	(N)
3	ANN_GROSS_PREM	ANN_GROSS_PREM	8	Numeric	(N)
4	ANN_NETT_PREM	ANN_NETT_PREM	8	Numeric	(N)
5	CANC_LAPSE_REV_IND	CANC_LAPSE_REV_IND	1	Character	(N)
6	CAN_LAP_REV_EFF_DT	CAN_LAP_REV_EFF_DT	8	Numeric	(N)
7	CAN_LAP_REV_TRANS	CAN_LAP_REV_TRANS	10	Character	(N)
8	COMM_RATE	COMM_RATE	8	Numeric	(N)
9	COVERAGE_CODE	COVERAGE_CODE	5	Character	(N)
10	COVERAGE_NUMBER	COVERAGE_NUMBER	5	Character	(N)

New Import Columns... Delete

Report

OK Cancel Apply Help

(c) 2004 Applied System Knowledge Ltd. 17

Reverse Impact Analysis : Coverage Dimension Data Table

Coverage Dimension Data Table

- 5. Load/Update Star Schema
- Loader: Coverage Dimension Data Table
- SQL Join: U: Update Coverage dimension
 - Branch 1
 - Coverage_nk Data Table
 - Branch 2
 - Coverages Data Table
- 4. Tra
 - Lo
 - Properties
 - View Job
 - Views the job's process flow
 - Print
 - SQL Join: J: Select coverage de
 - Branch 1
 - POL_DET_C ODD table
 - Branch 2
 - New_pol_trans Data Tabl
 - Loader: New_pol_trans D
 - SAS Splitter
 - Extract: Subset latest tra
 - POL_TRAN ODD table

Coverage Dimension Da...

5. Load/Update Star Schema.S...

Coverages Data Table

View / Edit any impacted Job directly from impact analysis

(c) 2004 Applied System Knowledge Ltd. 18



New flexible API

- XML-based data structures
- Call from Java, Visual Basic, C++
- 'PROC METADATA' in SAS

- Also a SAS Function interface
 - Data step
 - SCL



XML-based API

Utility in Management Console to interact with API

XML output and input

```

<Objects>
<TransformationStep Id="ASTUVM04.A6000001" Name="Loader: Changed_pol_base Data Table"/>
<TransformationStep Id="ASTUVM04.A6000003" Name="SQL Join: 3: Extract pol_base records"/>
<TransformationStep Id="ASTUVM04.A6000004" Name="Loader: New_pol_trans Data Table"/>
<TransformationStep Id="ASTUVM04.A6000006" Name="Loader: Canc_details Data Table"/>
<TransformationStep Id="ASTUVM04.A6000007" Name="SQL Join: 3: Details of cancellation"/>
<TransformationStep Id="ASTUVM04.A6000008" Name="Loader: Canc_trans Data Table"/>
<TransformationStep Id="ASTUVM04.A6000009" Name="Loader: Coverage_details Data Table"/>

```

Explore using the SAS Metadata Browser

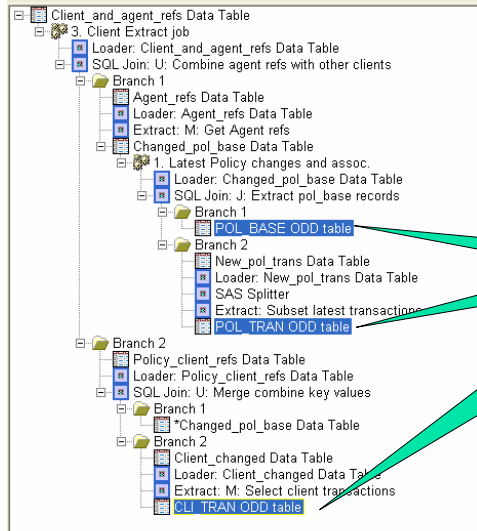


Name	Value
StoredText	DATE TIME0
Id	ASDQ1PFG.A900003U
Desc	Contains the source code that will be used upon a record insert the v.
MetadataCreated	22Sep2004:18:01:49
MetadataUpdated	22Sep2004:18:01:49
Name	SCD2 Beginning Date Expression
AssociatedTransformation	Association

Understand structure and relationships before writing API-based tools

Data Lineage from Reverse Impacts

Reverse Impact Analysis : Client_and_agent_refs Data Table



Identify source tables that provide this data



Value of Technical Metadata

Mainly derived from:

- Management of ETL – reduced coding
- Process maintenance
- Data lineage and dependency
- ... etc.



The Business Value of Metadata

Mainly derived from:

- Saving business users' time
 - Quickly locate the right data, report or analysis
- Reducing or eliminating errors and misunderstandings
 - Describe data at the point of use
 - Identify who to call



The Business Value of Metadata

- Where is the right data?
 - Information catalogue or channel
- What does it mean?
 - In business terms
- Where does data come from?
 - Operational sources (lineage again)
- How and when is it processed?
 - High-level, not the programming details
- Who is responsible for it?

(c) 2004 Applied System Knowledge Ltd.

25



Delivering business metadata

- Some is carried forward from technical metadata for data structures
 - Names, Descriptions – but may need more
 - Navigation metadata for OLAP
- More is derivable from ETL
 - Lineage, timing, quality/profiling results
- Final delivery tools allow augmentation
 - Portal channels, pre-defined calculations, more descriptive detail etc

(c) 2004 Applied System Knowledge Ltd.

26



Differing audiences for business metadata

- Report browsing and OLAP needs simple navigation and documentary
- Report creation needs more detail
- Data mining and ad-hoc analysis may need the most detail
 - Depending on business user expertise and knowledge
- Provide the relevant metadata for different data consumers

(c) 2004 Applied System Knowledge Ltd.

27



Looking forward

- Personal aim to make more use of SAS 9 delivery routes for metadata
 - Portal, Information Maps, connections from ETL to applications such as data mining etc, ETL exception alerts
- Also aim to use the new API
 - For analysis of metadata
 - To automate some repetitive tasks

(c) 2004 Applied System Knowledge Ltd.

28



Thank you!

Steve Morton
Applied System Knowledge Ltd
Steve.Morton@AppliedSystem.co.uk

© 2004 Applied System Knowledge Ltd