

Business Analytics **Forum**

See The Future Of Decision Making

November 3 – 5, 2010

Gold Coast Convention and Exhibition Centre Queensland, Australia



Information Governance and Business Glossary

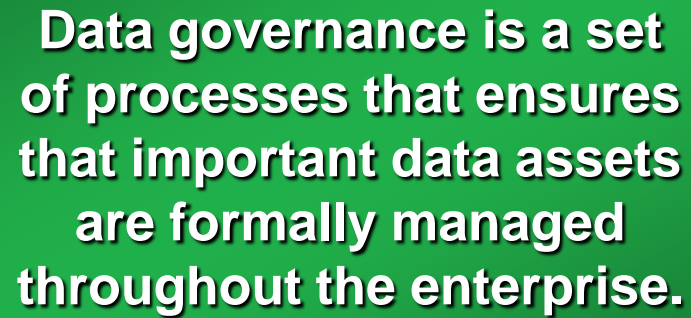
Vincent McBurney
Focus Strategies and NCVER



Welcome



INFORMATION GOVERNANCE WITH BUSINESS GLOSSARY

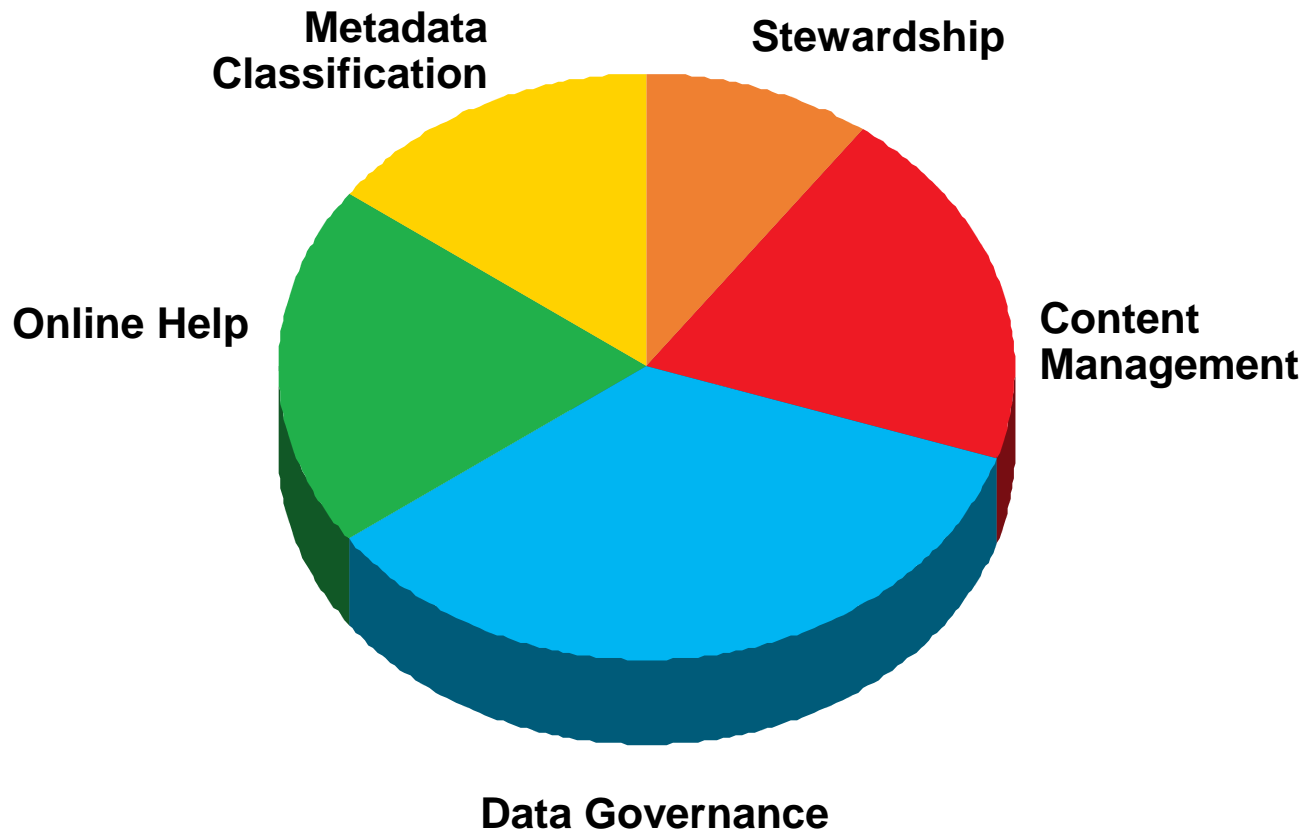
A green 3D rectangular box with a slight perspective, containing white text with a black outline.

Data governance is a set of processes that ensures that important data assets are formally managed throughout the enterprise.

A red 3D rectangular box with a slight perspective, containing white text with a black outline.

Data Governance stops your project from going badly horribly wrong

What is Business Glossary?



Focus and NCVER

National Centre for Vocational Education Research

- Data Collection validation and data quality governance.
- Data Warehouse for analytics and reporting.

HAVE YOU SEEN?



Do managers get in the way of innovation?

[Read more...](#)



The future of VET

[Read more...](#)



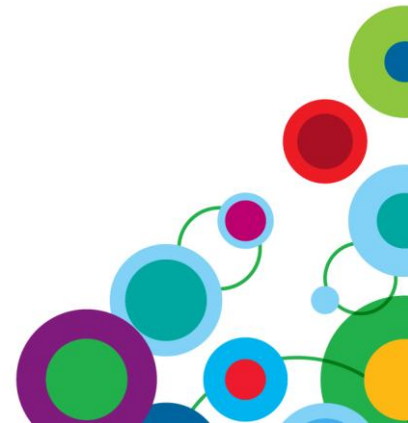
2009 VET financial data now available

[Read more...](#)



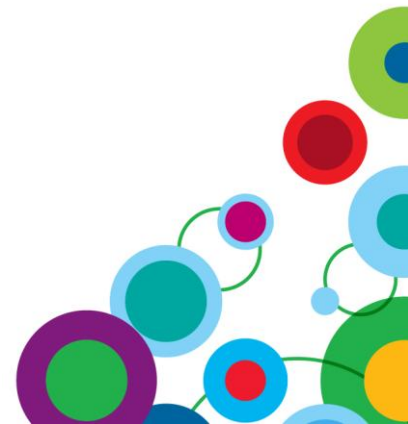
Australian vocational education and training statistics:
pocket guide

[Read more...](#)



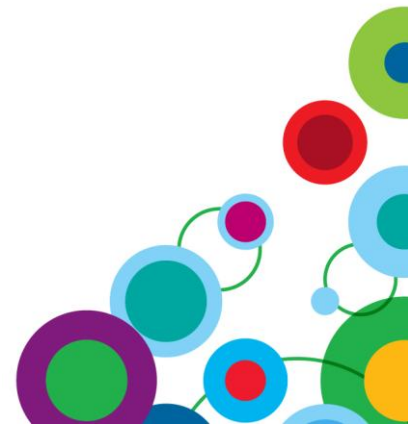
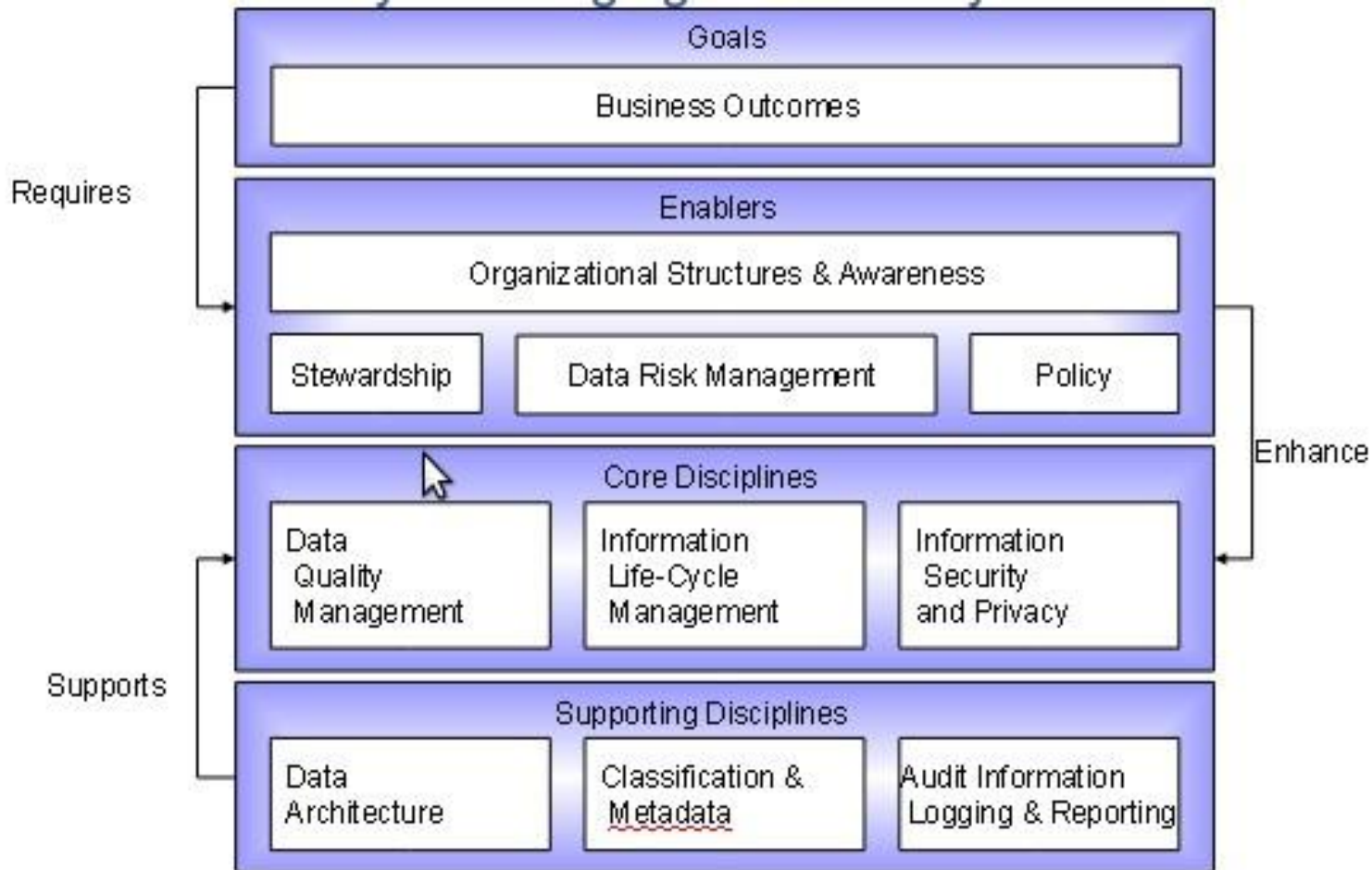
Data Governance Pain Points

- 1) Why are our Projects always late?
- 2) Why do we deliver something that isn't what the business wanted?
- 3) Why does it take so long for someone new on the project to get up to speed?
 - You've got to know the rules.
 - People have been here for years and just know how it works.
 - I don't have long enough to teach them, it's faster to do it myself.
- 4) Why is it so complex?
 - We didn't know it was going to be this hard.
 - We don't know why they did it this way.

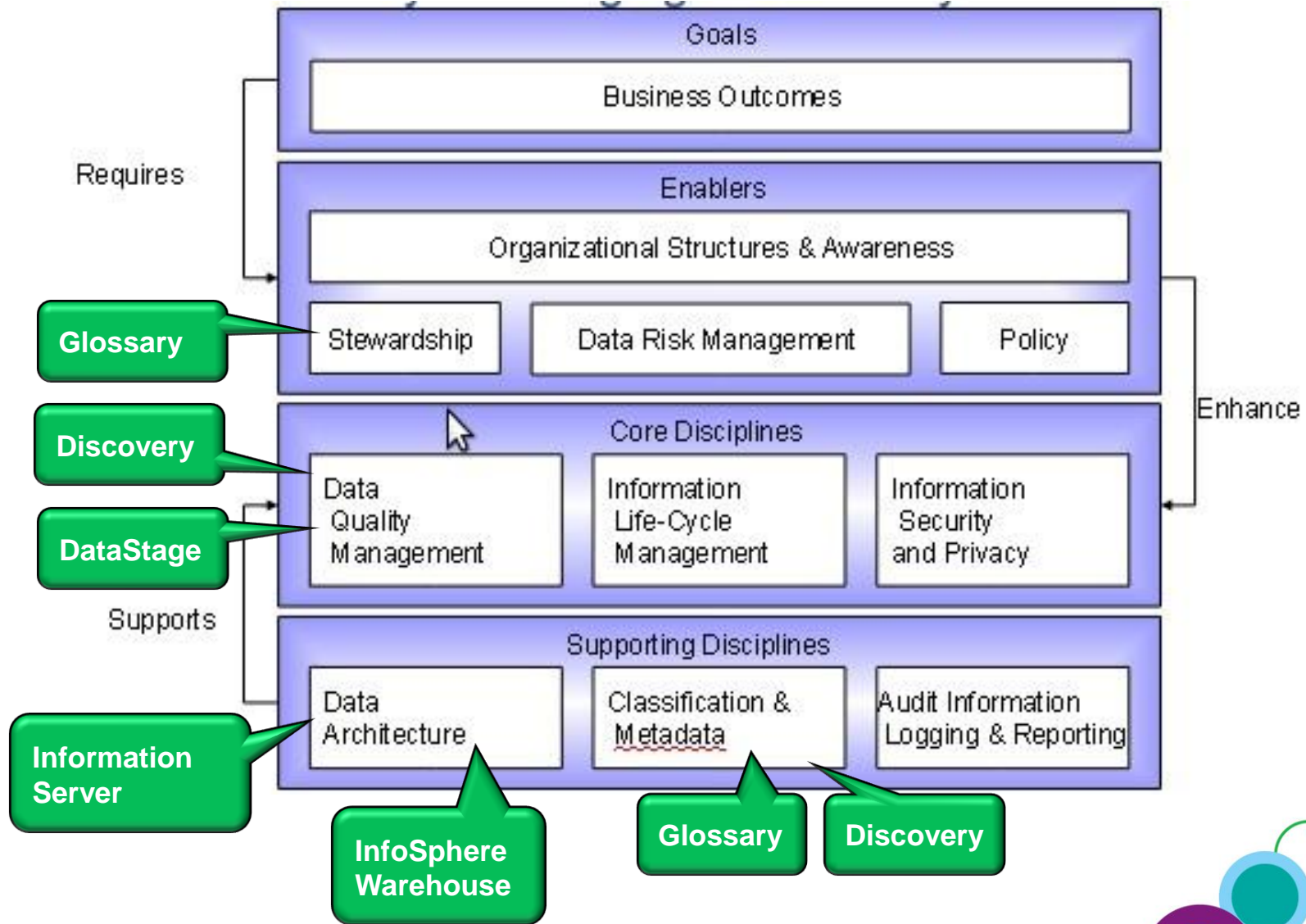


IBM Information Governance Maturity Model

- The elements of effective data governance

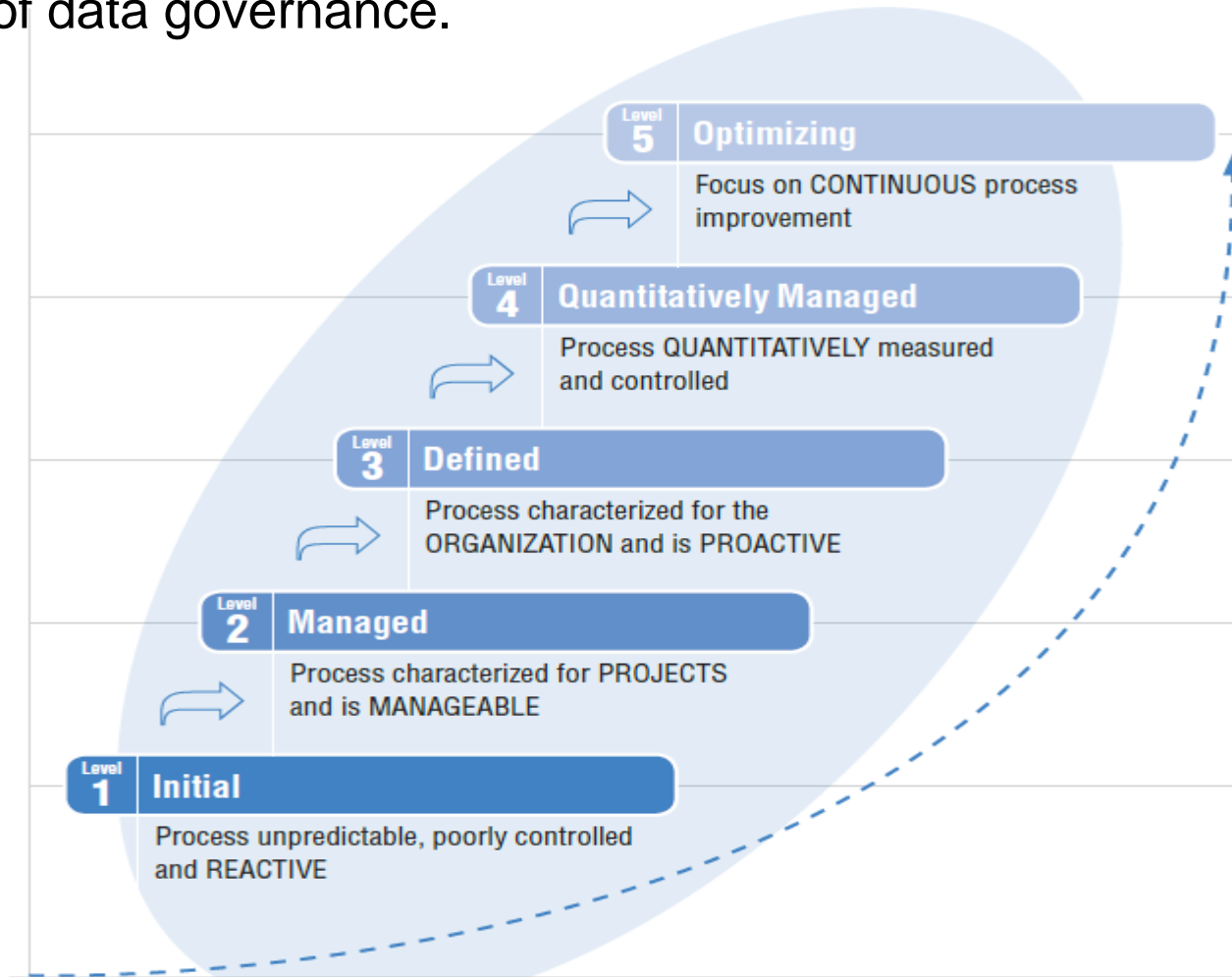


IBM Governance Software Acquisition Since 2004



Capability Maturity Model Integration (CMMI)

- Based on the Capability Maturity Model (CMM) and applied to each category of data governance.



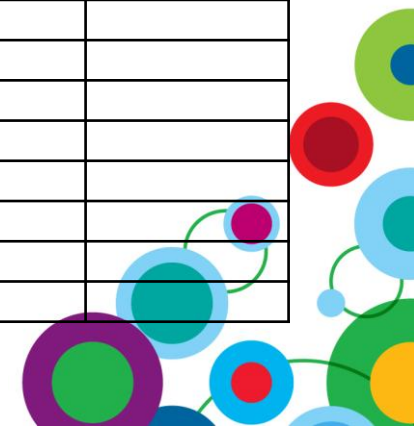
Graphic sourced from Carnegie Mellon Software Engineering Institute

What Maturity do you need?

- Draft recommendations from the last IBM Governance Council meeting

Improve Quality of Reporting and Analytics to Promote Organizational Effectiveness					
	Initial	Repeatable	Defined	Managed	Optimizing
Org Structures and Awareness			X		
Stewardship			X		
Policy		X			
Data Risk Mgmt and Compliance	X				
Info Security and Privacy	X				
Data Architecture			X		
Data Quality Mgmt				X	
Classification and Metadata			X		
ILM	X				
Audit Information Logging and Reporting			X		

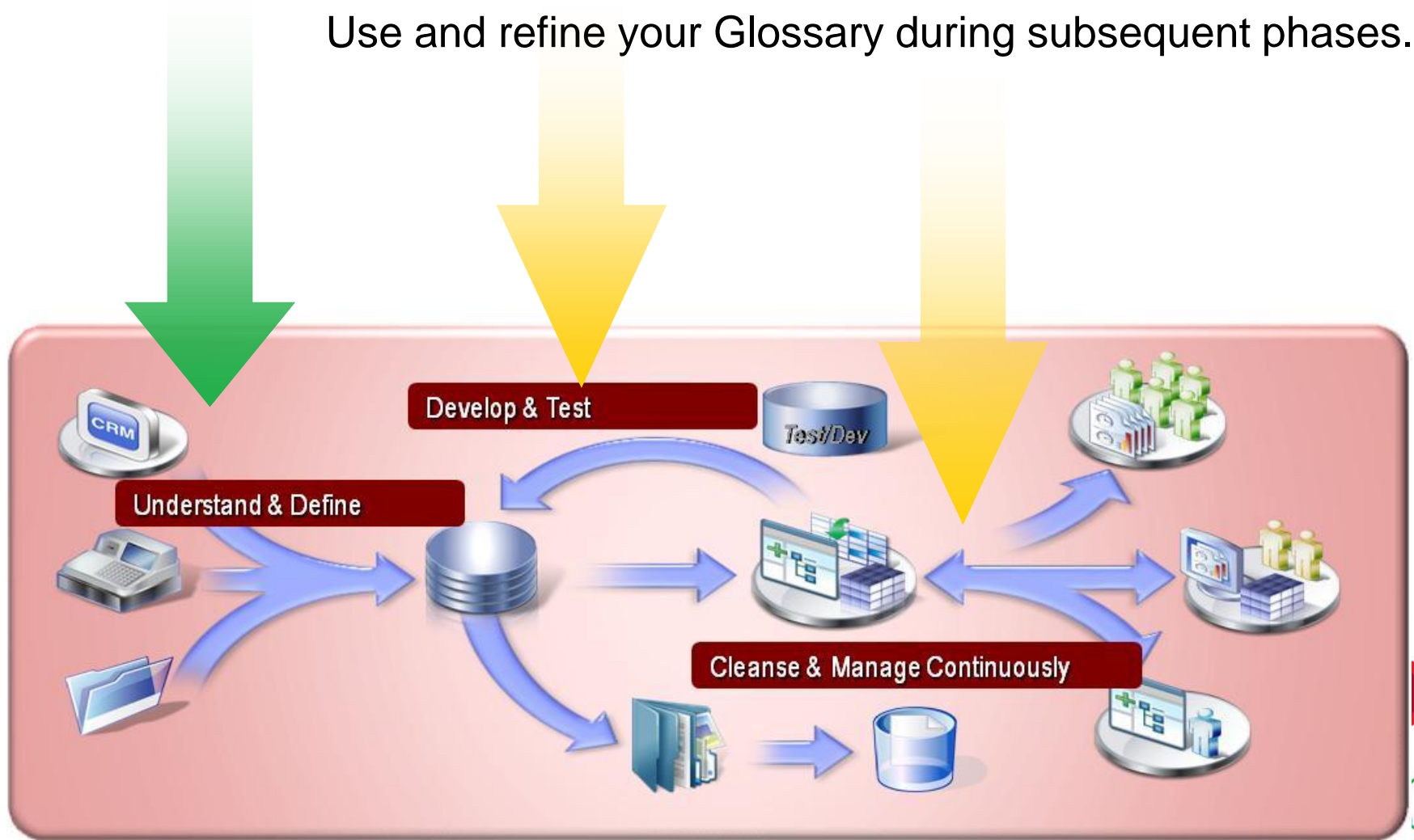
Transform Data into Valued Business Information to Make Smarter, Faster Decisions					
	Initial	Repeatable	Defined	Managed	Optimizing
Org Structures and Awareness	X				
Stewardship		X			
Policy	X				
Data Risk Mgmt and Compliance		X			
Info Security and Privacy		X			
Data Architecture			X		
Data Quality Mgmt			X		
Classification and Metadata			X		
ILM		X			
Audit Information Logging and Reporting	X				



Glossary in a Project

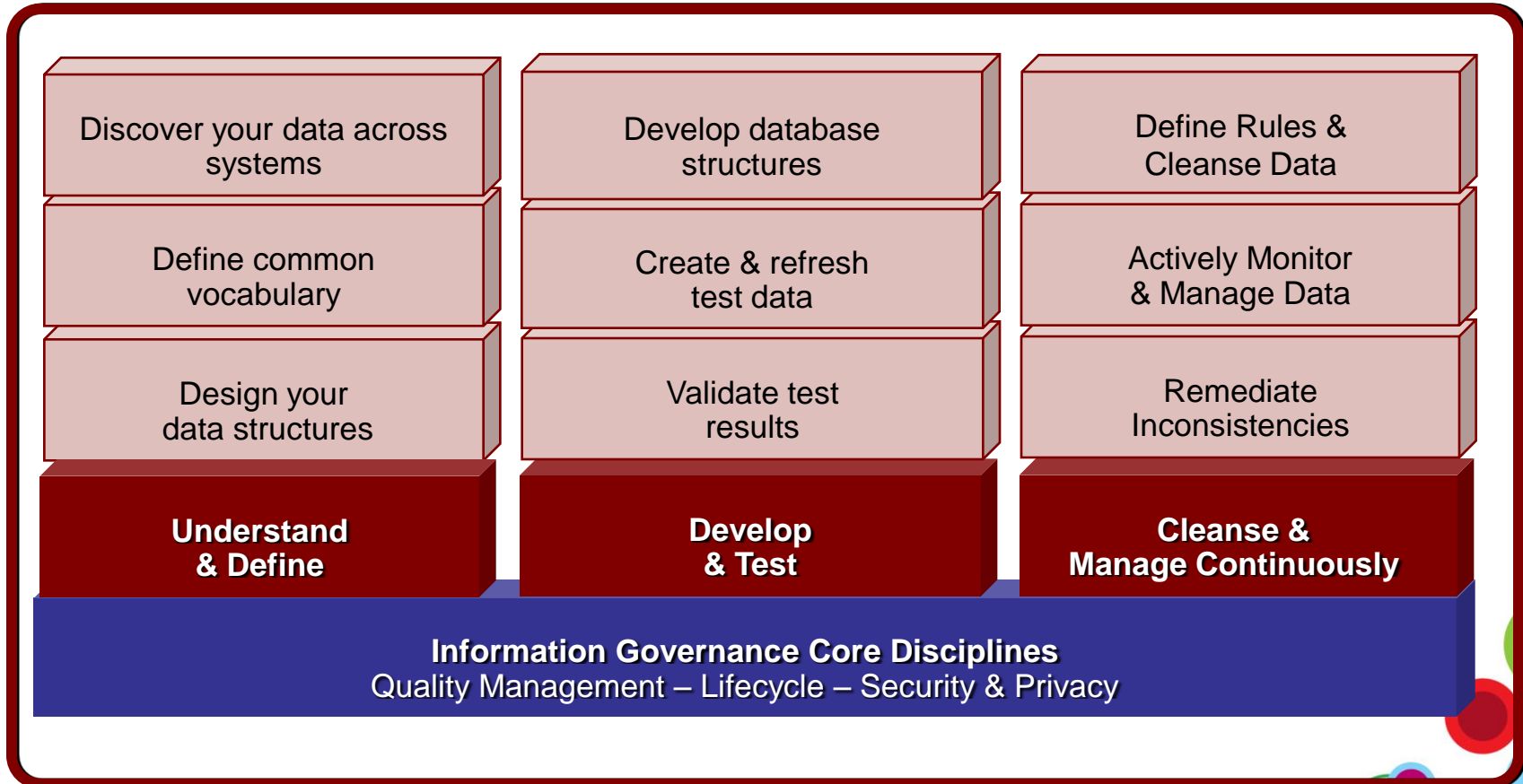
Create your Glossary during the Understand and Define stage.

Use and refine your Glossary during subsequent phases.



Understand and Define

Data Discovery and Glossary accelerate the first phase and lower the cost, risk and duration of subsequent phases.



INFORMATION GOVERNANCE PROJECT VICTIMS UNIT

- Why does a simple enhancement request take so long?
- Why are our estimates always wrong?
- Why does everyone take so long to do things?



The Victim Statements – the Business

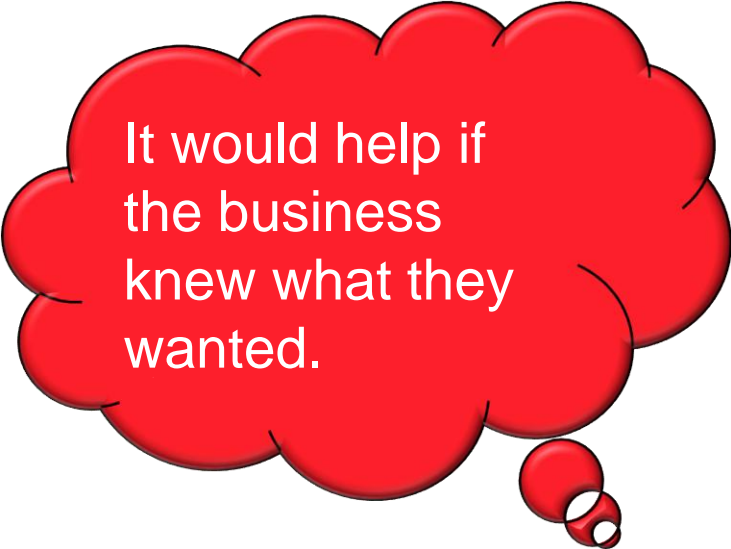
We spend all this money on IT and what do we get for it?

They just spent so long on meetings and documentation and didn't build anything!

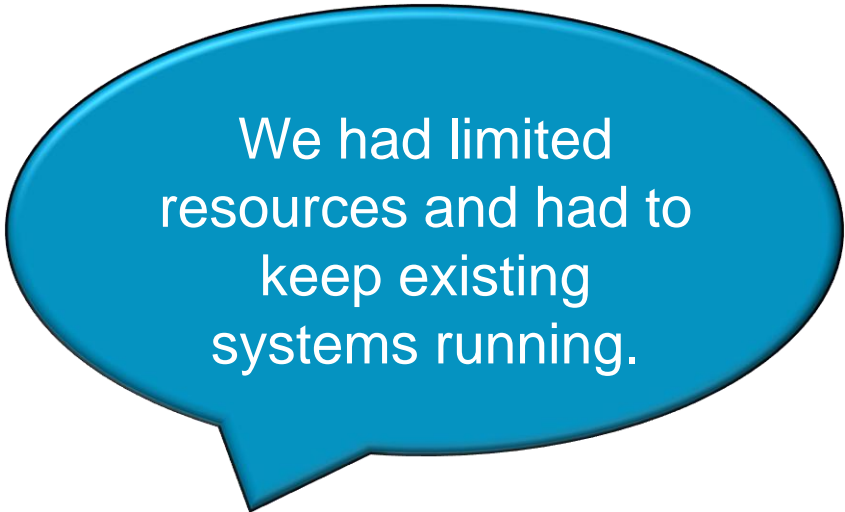
When we got to UAT Testing there were bugs and it had to be fixed over and over again.

We could have built this faster ourselves


Initial Suspect – Internal IT



It would help if the business knew what they wanted.



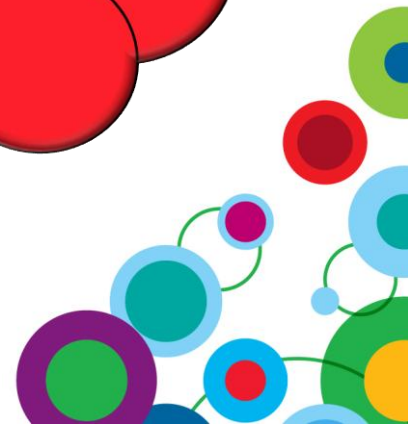
We had limited resources and had to keep existing systems running.



We brought in consultants and they took a long time on deliverables.



These guys earn 10 times what I do!



Obvious Suspect – the Consultants

We had some unexpected change requests and scope creep.

Requirements and rules kept changing right up through testing.

No one told us there were three different definitions for client status.

You thought that was bad, wait until you see phase 2.

Obvious Scapegoat – the New Guy

I didn't do a proper handover as the guy I replaced was always out to lunch.

I don't know where the application documentation is.

I don't even know who was managing the project.

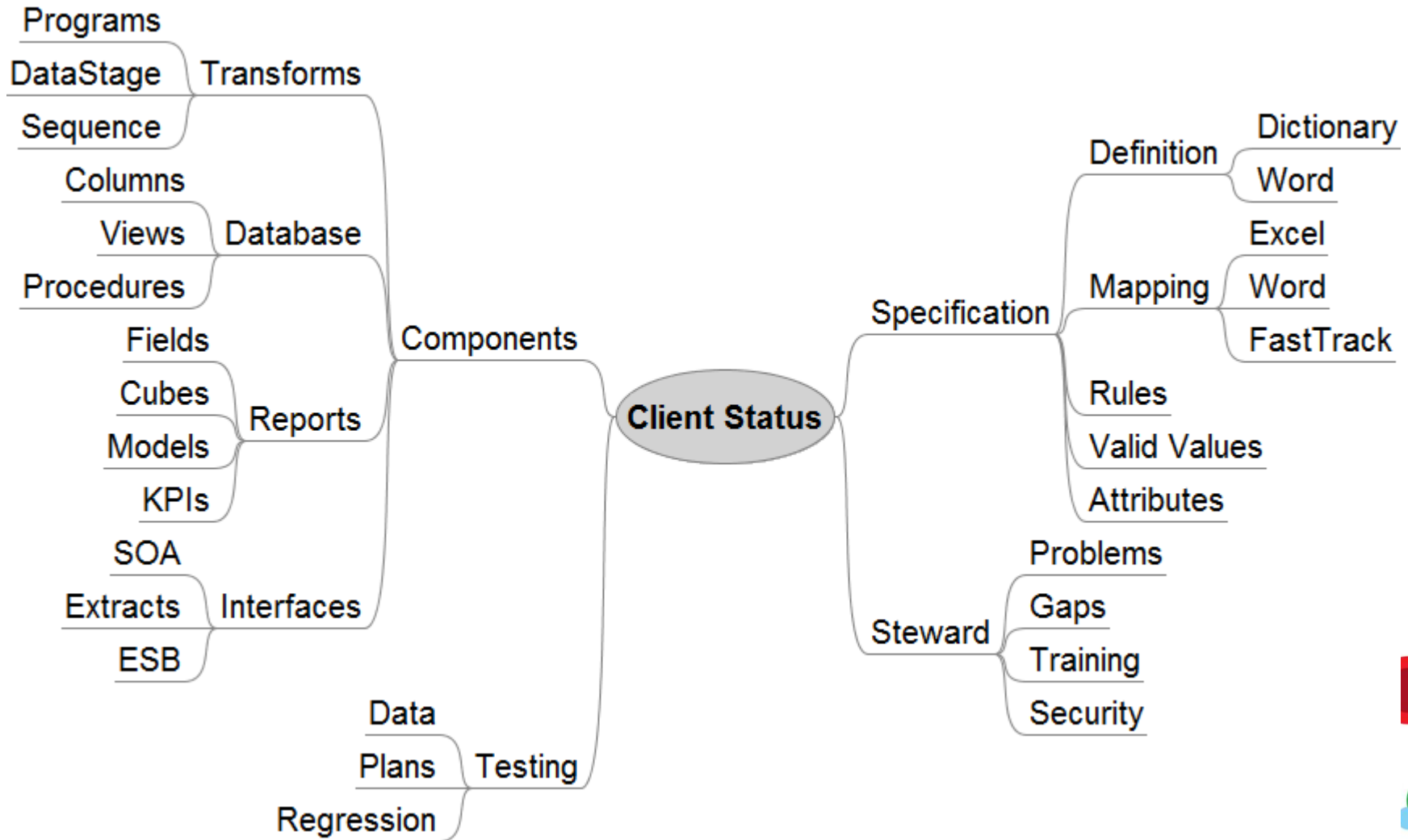
Turned out the Functional Spec I was using was out of date by two years.

There are three different definitions for client status? Wait, which client status are we talking about?

I need to update my resume.

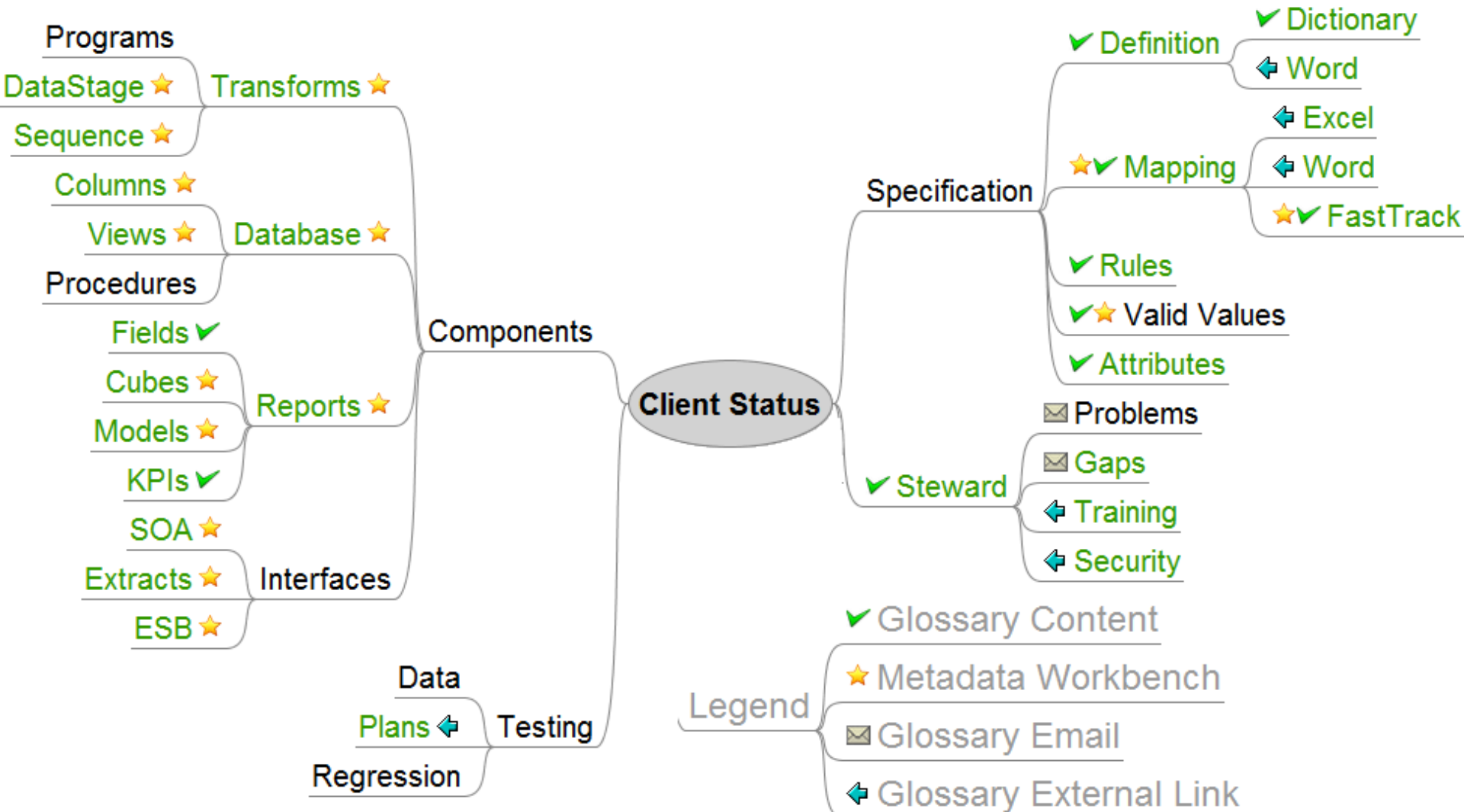
The Coroners Report

- The team did not have the information and the context for the change.



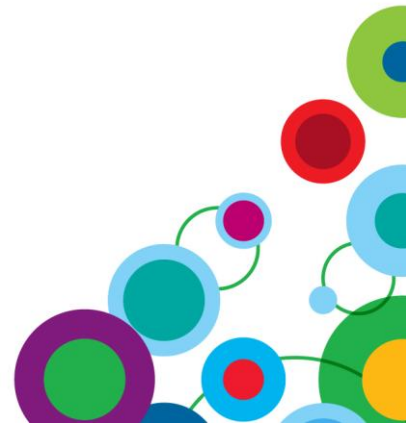
The Information Server Approach

- Metadata Workbench and Business Glossary provide context

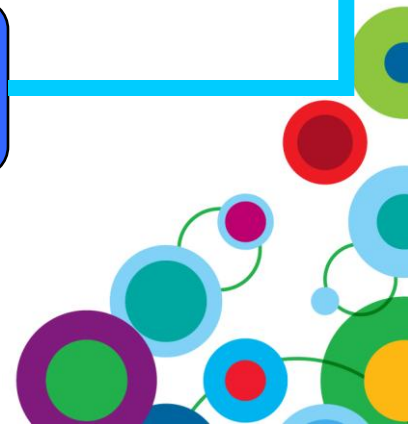
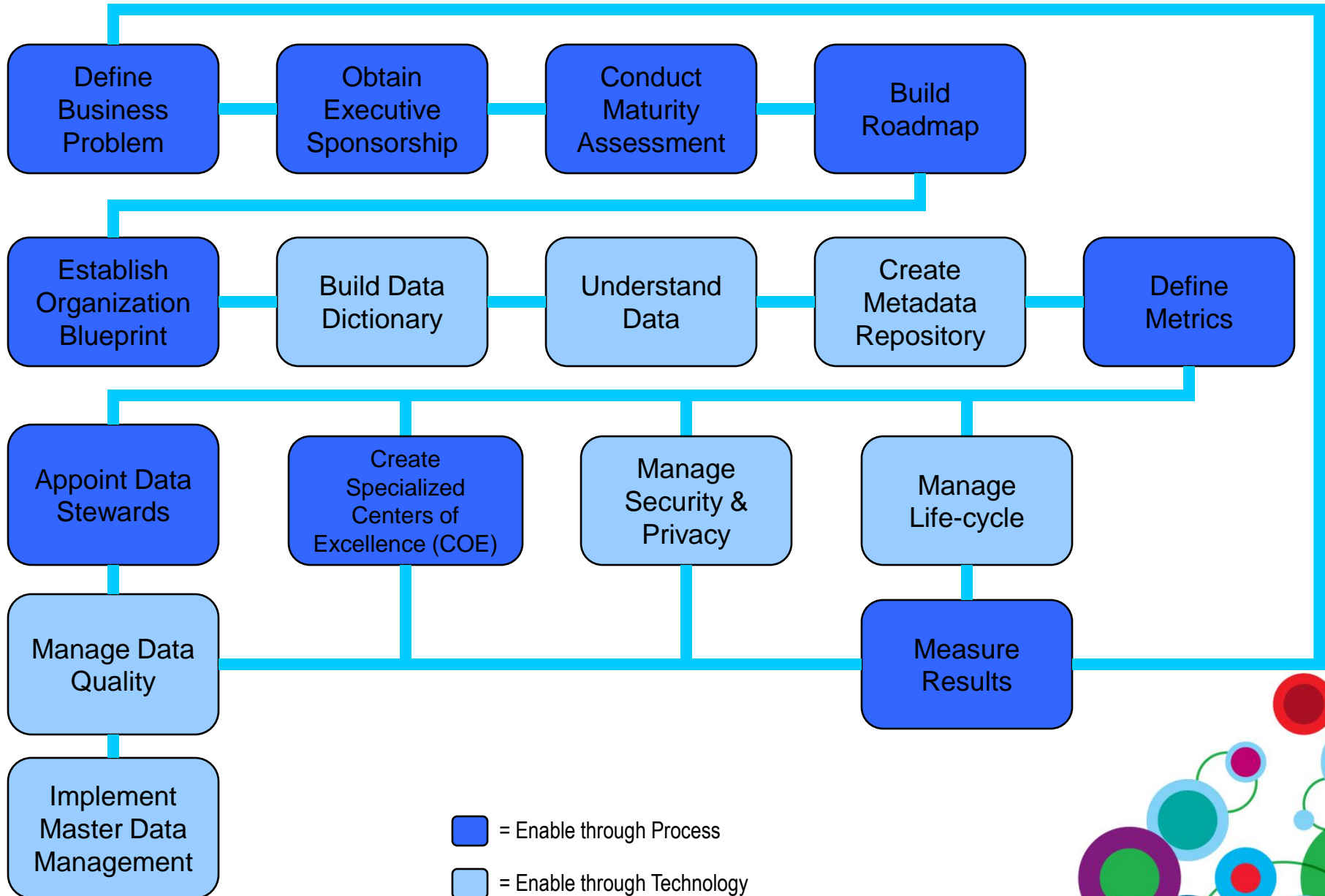


BUSINESS GLOSSARY GETTING STARTED

- The steps to a successful Business Glossary.



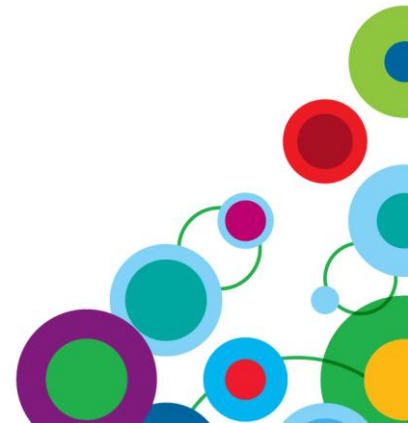
Define Business Glossary in the Unified Process



Defining the Scope and Structure

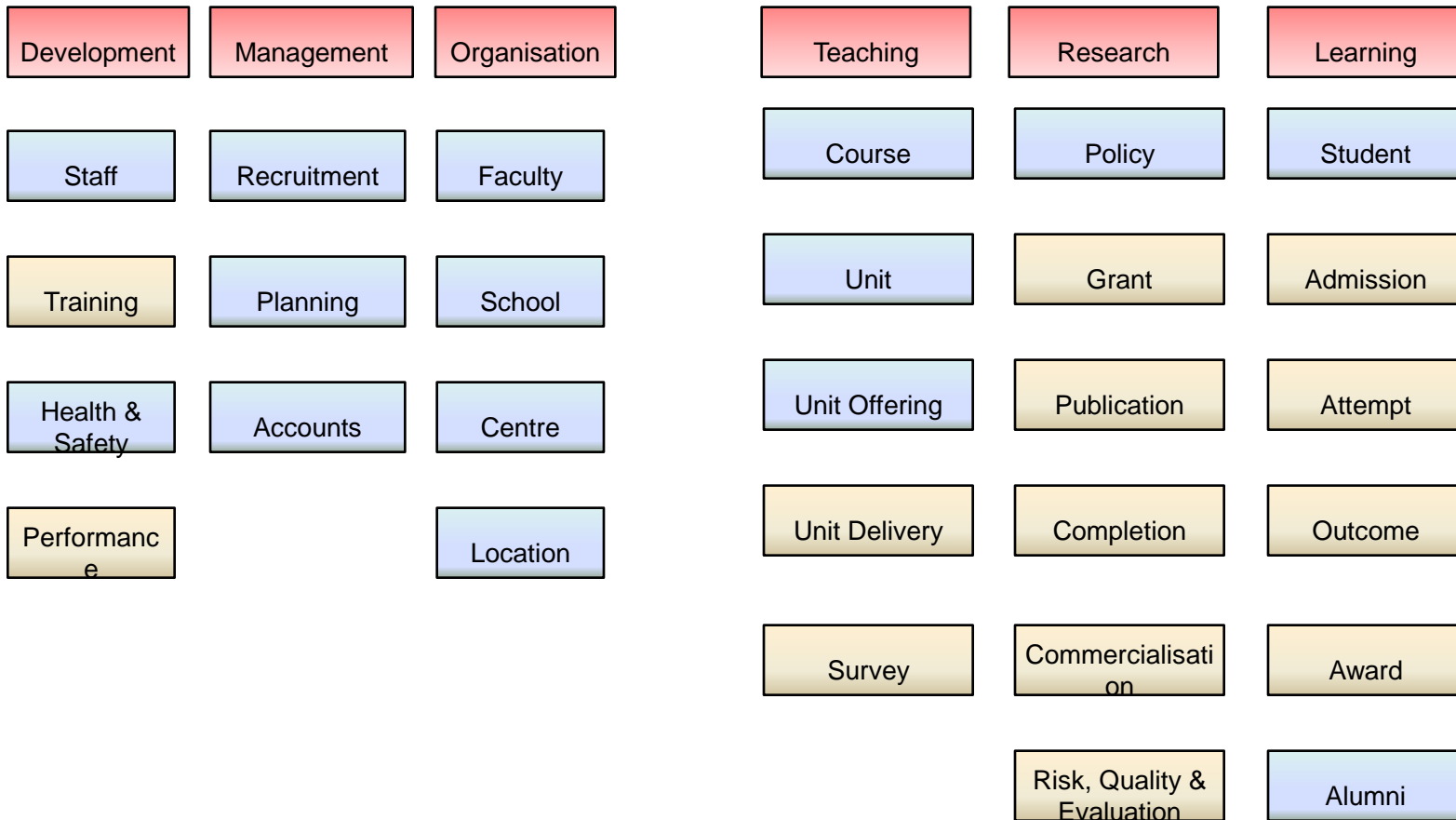
Work out what terms are included and how they are organised.

- Choose your Taxonomy – can be different across subject areas.
 - Flat Taxonomy
 - Try to keep less than 30 categories for scrolling/scanning
 - Only about 25 terms fit on a standard Glossary browser page
 - Hierarchical Taxonomy
 - Wide and Shallow is less work to navigate
 - Do not branch for a single category
 - Sufficient content in each category to warrant existence
 - Network Taxonomy
 - Define once and use multiple times
 - Terms can reference each other
 - Categories can **hold** or **reference** a term



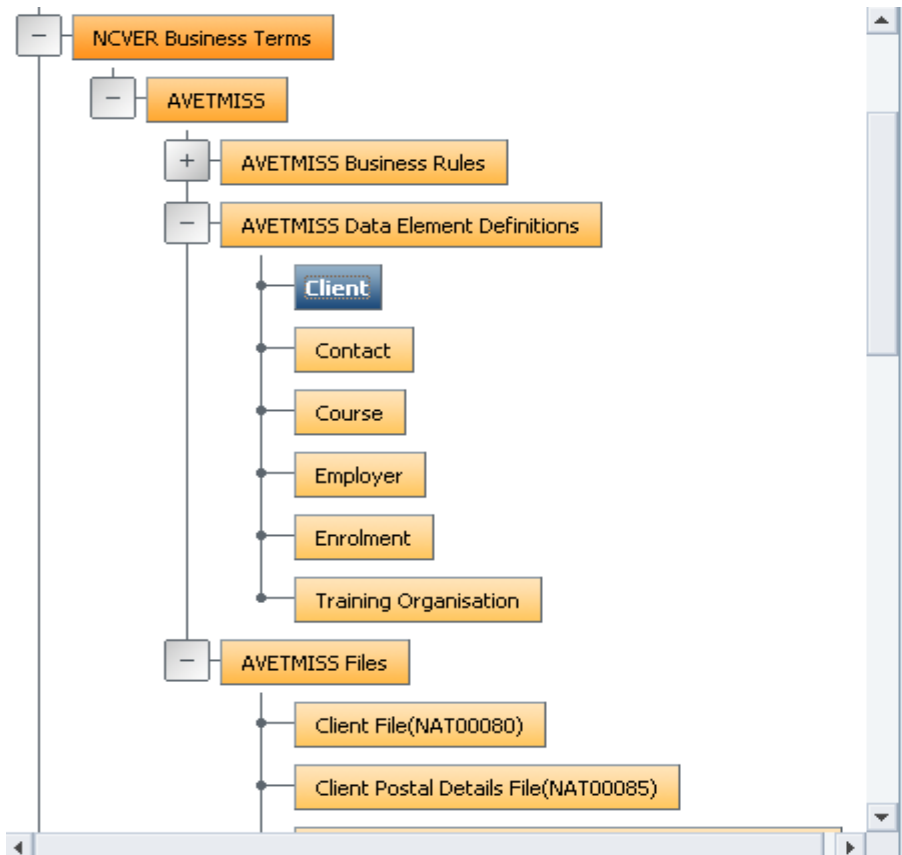
Identify Subject Areas

- If you are using a Business Glossary to support a Data Warehouse then start with the high level conceptual data model.



Start with a Formal Vocabulary

- Focus helped create a Glossary for a Data Collection at NCVET.
- Clearly defined Data Dictionary with elements and rules.



Category:
Client

[Click to View Full Details](#)

Description:
Concepts that describe a Client of training providers.

Steward:
Undefined

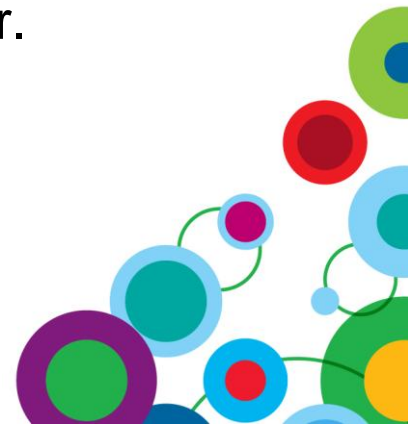
- Terms:
- [At School Flag](#)
 - [Client First Given Name](#)
 - [Client Identifier](#)
 - [Client Identifier - Apprenticeships](#)
 - [Client Identifier - TYIMS](#)
 - [Client Last Name](#)
 - [Client Title](#)
 - [Date of Birth](#)
 - [Disability Flag](#)
 - [Disability Type Identifier](#)
 - [Highest School Level Completed Identifier](#)
 - [Indigenous Status Identifier](#)
 - [Labour Force Status Identifier](#)
 - [Main Language other than English Spoken at Home Identifier](#)
 - [Name for Encryption](#)



Define your Steward Structure and Policies

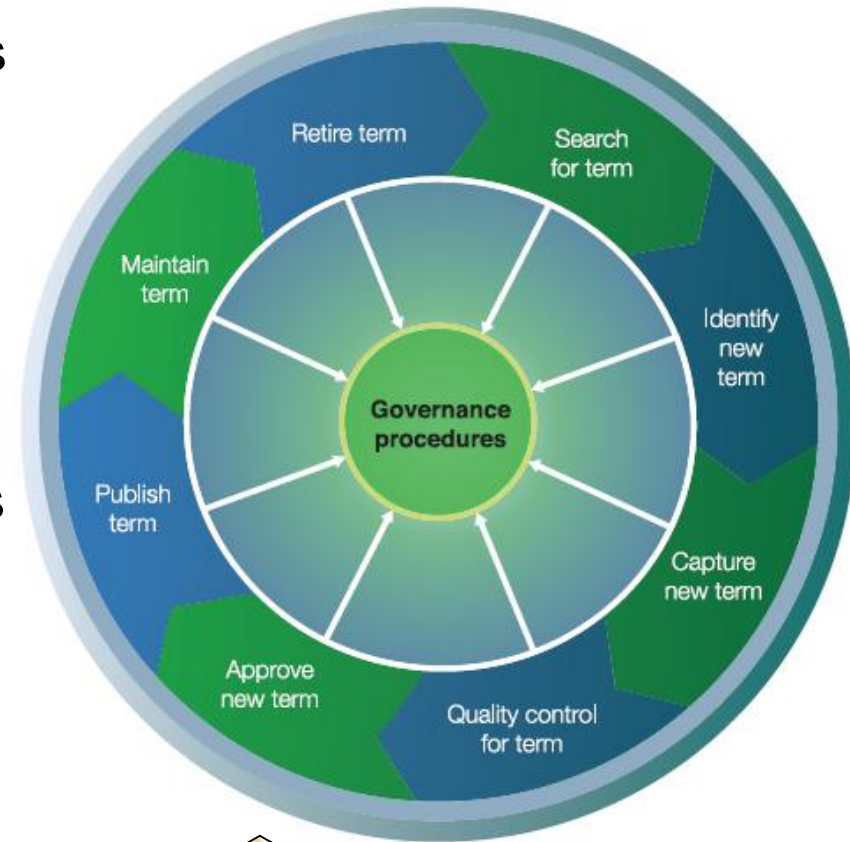
Are Terms and Categories owned by People or Groups?

- In Business Glossary stewardship refers to the person who maintains the content and answers questions.
 - Terms can only be **owned by** a single person or group and passed between people easily.
 - Terms can be **related to** multiple people but it can be painful to maintain.
 - Glossary users can quickly and easily find the term owner.
- Business Glossary introduces a simple stewardship structure.
 - Business Rules stewardship is in Information Analyzer.
 - Data Exception stewardship is in Exception Manager.

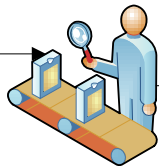


Define the Lifecycle of Terms

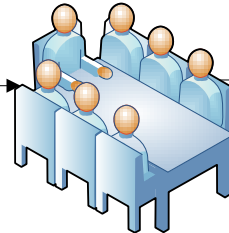
- Work out the Data Stewardship Policies
 - How to use the Term status
 - Identify review groups
 - Collaborate via email
 - Track changes over time
 - Report to track progress of reviews



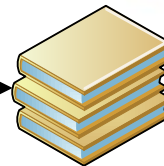
Term Added – “Candidate”



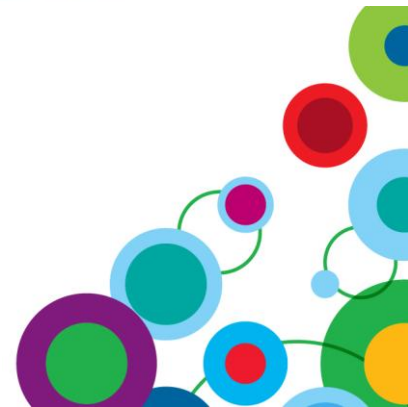
Small Term View – “Accepted”



Large Team Review – “Standard”

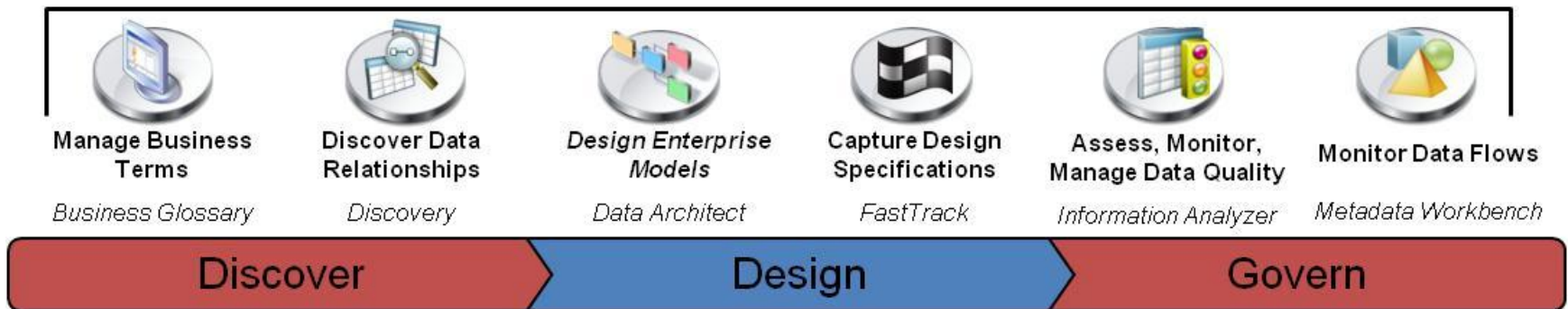


Enterprise Term



Tell Everyone where it fits

- Business Glossary content needs to become part of the Software Development Lifecycle.
- The more Information Server products used the better the chances to maintain good quality Business Glossary information.
 - Discovery helps classify physical data.
 - FastTrack helps map physical columns to Glossary terms.
 - Metadata Workbench can be a faster way to associate IT assets to terms.



Get a Fast Start with Imports

- The Quality of Business Glossary imports has a major impact on the success of the implementation.
 - We use Excel to CSV import first for content.
 - We use Excel to XML import for relationships.
 - We use Excel to XML import for External Website links.
 - We use extra Excel to XML imports to add custom attributes.
- CSV imports can only handle terms and attributes but not relationships.
- XML imports can handle everything – including relationships between term and assets.
- Excel is a good way to grab over a large chunk of content that would be too slow to data enter straight into Glossary.



Excel for Glossary XML

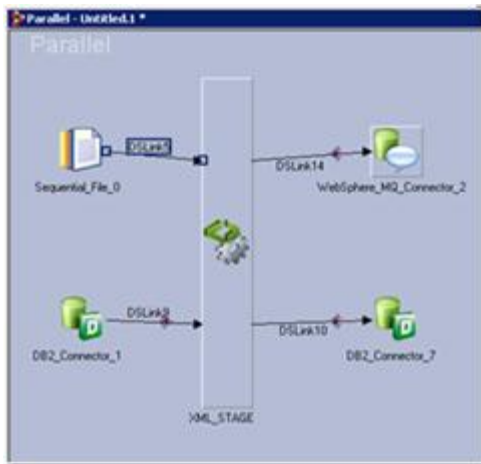
- We can collect Synonyms, Related Item links, additional Custom Attributes, External links etc in normal Excel tables.
- Paste or link the data into our Excel to XML import sheet.
- Create an XML text file for import into Glossary.

<terms>						
<term name="	Activity dust diseases lev	"><parentC ategory identity="	Business Terms::Workers Compensation::Policy	"></parentCategory><customAttributes><cu stomAttributeValue customAttribute="PowerDesigner Unique Identifier" value="	DR000008	"/></customAttributes><relatedTerms ><termRef identity="Business Rules::
<term name="	WIC premium adjustment flag	"><parentC ategory identity="	Business Terms::Workers Compensation::Policy	"></parentCategory><customAttributes><cu stomAttributeValue customAttribute="PowerDesigner Unique Identifier" value="	DR000250	"/></customAttributes><relatedTerms ><termRef identity="Business Rules::
<term name="	Wages for determined classes dust disease lev	"><parentC ategory identity="	Business Terms::Workers Compensation::Policy	"></parentCategory><customAttributes><cu stomAttributeValue customAttribute="PowerDesigner Unique Identifier" value="	DD000246	"/></customAttributes><relatedTerms ><termRef identity="Business Rules::
<term name="	Determined classes dust diseases lev	"><parentC ategory identity="	Business Terms::Workers Compensation::Policy	"></parentCategory><customAttributes><cu stomAttributeValue customAttribute="PowerDesigner Unique Identifier" value="	DR000087	"/></customAttributes><relatedTerms ><termRef identity="Business Rules::
<term name="	Amount of wages paid W1 (last period)	"><parentC ategory identity="	Business Terms::Workers Compensation::Policy	"></parentCategory><customAttributes><cu stomAttributeValue customAttribute="PowerDesigner Unique Identifier" value="	MD000019	"/></customAttributes><relatedTerms ><termRef identity="Business Rules::



New Import Method – DataStage 8.5

- Import, store and browse the Glossary XSD schema file.
- Read lists of glossary terms and convert them to Glossary XML.
- Filter out bad data before it breaks the glossary import – remove rich text and funny characters.



The screenshot shows the 'Assembly Editor for Stage XML_0'. The 'XML_Composer Step' is selected in the 'Assembly Outline'. The 'Configuration' tab is active, showing the 'Mappings' section. A table lists the source and target mappings for the XML structure.

Source	Result	Target
top	+	document_collection
top/DSL_Parser/result/ns/catalog/book	+	book
top/DSL_Parser/result/ns/catalog/book/author	+	author
top/DSL_Parser/result/ns/catalog/book/title	+	title
top/DSL_Parser/result/ns/catalog/book/genre	+	genre
top/DSL_Parser/result/ns/catalog/book/price	+	price
top/DSL_Parser/result/ns/catalog/book/publish_date	+	publish_date
top/DSL_Parser/result/ns/catalog/book/description	+	description
top/DSL_Parser/result/ns/catalog/book/price	+	@id

The 'Details Inspector' window is open, showing the details for the 'price' attribute:

- price**
- Attributes:** Name: price, Type: decimal, Type Name: decimal, Type: http://www.w3.org/2001/XMLSchema:URI, Namespace: 2001/XMLSchema:URI, Facets: Whitespace: COLLAPSE

Basic Glossary Entry

- Using Glossary just for Definitions

[NCVER Business Terms](#) » [AVETMISS](#) » [AVETMISS Data Element Definitions](#) » [Client](#) »

Client First Given Name

Long Description: Client First Given Name contains the client preferred first name.

Status: Candidate

General Information

Usage — Client First Given Name is used for addressing correspondence.

Classification — VALUE : DESCRIPTION - CLIENT FIRST GIVEN NAME
text : Client's preferred first name

Guidelines — Not applicable

Related Terms — Not applicable

Rules — Client First Given Name must not be the client family name even if it is their first name.

Created By — isadmin

Created On — 2010-10-07 15:04:59

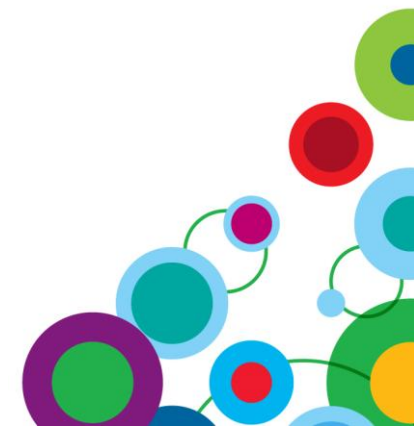
Modified By — isadmin

Modified On — 2010-10-08 11:06:54

▶ Assigned Assets

▶ Notes

▶ History



Synonyms and Related Terms
















- Synonyms track different names for the term across the Enterprise.
- Related Terms are used to define validation business rules.

▼ General Information

Synonyms —

1. [403 - Student given name - First](#) ( [University](#) »  [DEEWR](#) »  [Commonwealth Assisted Students](#))

Related Terms —

1. [Client First Given Name\(NAT00085\) Rule1](#) ( [NCVER Business Terms](#) »  [AVETMISS](#) »  [AVETMISS Business Rules](#) »  [Client Postal Details File Validation\(NAT00085\)](#) »  [Client Details\(NAT00085\)](#))
2. [Client First Given Name\(NAT00085\) Rule2](#) ( [NCVER Business Terms](#) »  [AVETMISS](#) »  [AVETMISS Business Rules](#) »  [Client Postal Details File Validation\(NAT00085\)](#) »  [Client Details\(NAT00085\)](#))
3. [Client First Given Name\(NAT00085\) Rule3](#) ( [NCVER Business Terms](#) »  [AVETMISS](#) »  [AVETMISS Business Rules](#) »  [Client Postal Details File Validation\(NAT00085\)](#) »  [Client Details\(NAT00085\)](#))

Usage — Client First Given Name is used for addressing correspondence.

Classification — VALUE : DESCRIPTION - CLIENT FIRST GIVEN NAME
text : Client's preferred first name

Guidelines — Not applicable

Related Terms — Not applicable

Rules — Client First Given Name must not be the client family name even if it is their first name.

Created By — isadmin

Created On — 2010-10-07 15:04:59

Modified By — isadmin

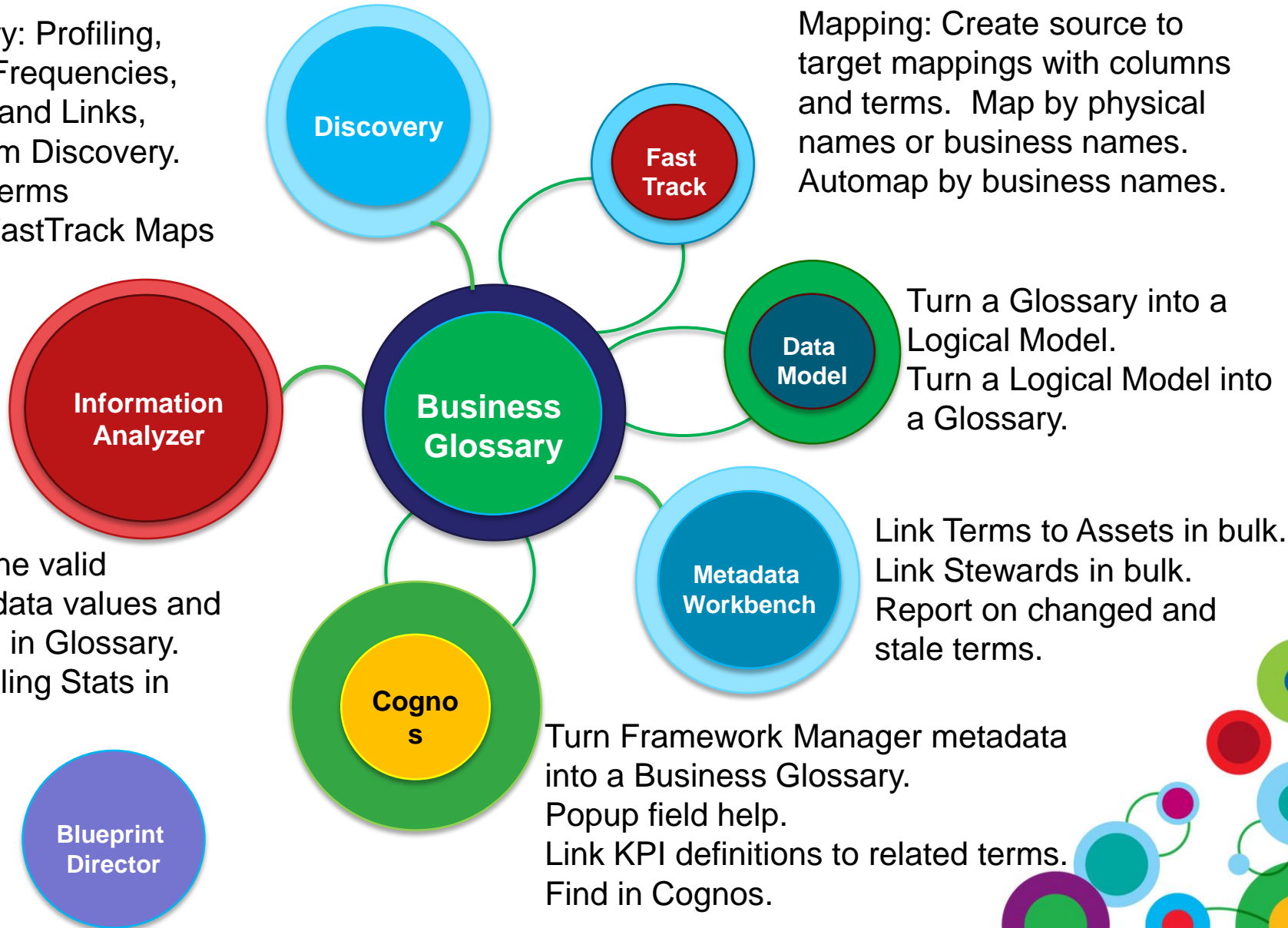
Modified On — 2010-10-08 13:29:20



Unified Metadata Approach

Discovery: Profiling,
Values, Frequencies,
Overlap and Links,
Transform Discovery.
Assign Terms
Create FastTrack Maps

Mapping: Create source to
target mappings with columns
and terms. Map by physical
names or business names.
Automap by business names.



Assigned Assets

- External Assets – Given Name is linked to external HTTP links such as documents in Sharepoint or Intranet Pages.
- Metadata Assets – Given Name has been explicitly linked to a FIRST_NAME column in the Warehouse.


A linked Word Document













Linked DB Columns



Change History



Assigned Assets

-  [Data Privacy Policy](#) **TYPE:** MS Word Document **CONTEXT:** Compliance **LOCATION:** Sharepoint
This describes the privacy requirements for client data.
-  [FIRST_NAME](#) ( [bankconnect](#) »  [BANKDEMO](#) »  [BANK](#) »  [MASTER](#))
-  [Name](#) ( [Input](#))
-  [Name](#) ( [StandardizedData](#))

Notes

History

Select a property to add to the table... Collapse Rows

Date	Edited By	Comment	Guidelines
2010-10-08 11:06:54	isadmin		Not applicable
2010-10-07 15:04:59	isadmin		Not applicable
2010-10-07 15:04:59	isadmin	Created	

Custom Search and Browse

- Using the Glossary API to write our own Author search and browse

ROO
Spring

CATEGORY

Create new Category

List all Categories

TERM

Create new Term

List all Terms

Find by Term

Name:

FIND **CLEAR**

List all Terms

Name	Short Description	Context			
5 Year age groups	Five year age groups.	Department of Health > Collection > Victorian Admitted Episode Dataset (VAED) > Data Elements			
Aboriginal or Torres Strait Islander Origin but No	Aboriginal or Torres Strait Islander origin, but the Country of Birth is not a code specific to Australia (1100 – 1102, 1199).	Department of Health > Collection > Victorian Emergency Minimum Dataset (VEMD) > Edits			
ACAS Status	The type of involvement of the Aged Care Assessment Service (ACAS) in patient separation.	Department of Health > Collection > Victorian Admitted Episode Dataset (VAED) > Data Elements			
Accommodation Type (a)	(a) The accommodation type or types occupied by the patient during their admission, including changes to this item during the episode.	Department of Health > Collection > Victorian Admitted Episode Dataset (VAED) > Data Elements			
Accommodation Type on Separation (b)	(b) The accommodation type occupied by the patient on their last (counted) patient day.	Department of Health > Collection > Victorian Admitted Episode Dataset (VAED) > Data Elements			
Account Class (a)	(a) The agency/individual chargeable for this episode, and associated sub categories, for this episode of care, including changes to this item during the episode.	Department of Health > Collection > Victorian Admitted Episode Dataset (VAED) > Data Elements			
Account Class on Separation (b)	(b) The agency/individual chargeable for this episode, and associated sub categories, on the last (counted) patient day.	Department of Health > Collection > Victorian Admitted Episode Dataset (VAED) > Data Elements			

Simpler Author Menu

35

Custom Browse and Data Entry Forms

- Using the Glossary API to write our own authoring forms

ROO
Spring

CATEGORY

Create new Category

List all Categories

TERM

Create new Term

List all Terms

▼ Create new Category

Name:

Parent Category:

Short Description:

Long Description:

Abbreviation:

Administrative status:

BR Start Date: Enter BR Start Date (required)

Version:

Home | Language:

Sponsored by SpringSource

Different Column Order

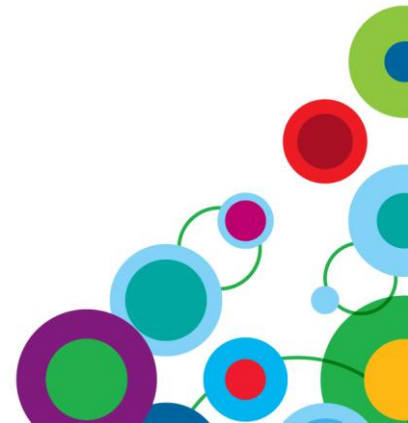
Better Validation

Better Date entry

September						
S	M	T	W	T	F	S
29	30	31	1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	1	2
3	4	5	6	7	8	9
2009			2010	2011		

BUSINESS GLOSSARY CLOSING REMARKS

- Final comments from the prosecution and the defence.



Steven Adler supporting Data Governance

Director of IBM Data Governance Solutions at IBM.

I meet a lot of people who ask me about the Data Governance products or roadmaps organizations should buy. The best products you can buy are the ones that tell you what you don't already know. To govern effectively, you need to know what's going on in the context to when it is happening, what it means, and how it relates to other things.



Steven Adler opposing Data Governance

Former drummer for Guns N' Roses.

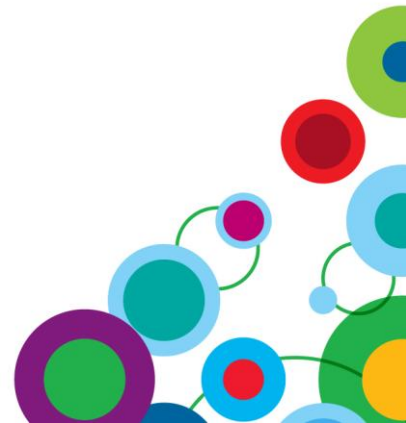
“We never did anything the sane, sensible way. We never went by the rules and never conformed to an accepted path to success. The way we came up with our songs, insisted on total artistic freedom, the way we practiced and played — no one did it like we did.”



Vincent McBurney - Judicial Verdict

Never been to a Guns N' Roses concert.

Strike a balance. Data Governance via Business Glossary is not about coming down hard on people but about providing information in context that is accurate, complete, insightful and timely.



Getting Started with Data Governance

Six things everyone can do today:

1. Define your desired outcomes from Data Governance
2. Be clear about the problems you are solving
3. Define a realistic organisational structure for your environment
4. Focus on a DG pilot program that can deliver outcomes with business benefits
5. Take advantage of best practices and models from organisations like the Data Governance Council and MIKE 2.0
6. Be real with organisational challenges, funding requirements, scope and duration of deliverables





WIN *an Apple® iPad*

Please remember to complete your session evaluation online at the Communication Station or point your Smart Phone browser to:

www.spss.com/goldcoast



For a chance to win an Apple® iPad

