IBM Industry Models

Business user focused Vocabularies for IBM Industry Models

Contents

1 Introduction

There is a slow evolution occurring regarding the role of semantic content across the IT industry. The traditional role of such assets is either as relatively simple glossaries to provide supplemental business metadata to the more technical structural metadata definitions or as a set of taxonomies for use by certain architect and analytics during the early phases of a development lifecycle to establish a more precise understanding.

The rapid changing of the typical data management landscape in recent years with the arrival of various big data technologies such as Hadoop, streaming and virtualization means that many organizations are being forced to review their approach to the use of business vocabularies to ensure common understanding business meaning. The heterogeneous set of technologies now in place in many organizations means that there is an increased need for a usable and enterprise-wide capability to ensure common understanding of the various components in that landscape by the increasing range of different consumers of that landscape, ranging from business users, data scientists, citizen analysts, IT operations staff, and IT development staff.

In parallel, there are changes in the acceptance and practicality of new semantic technologies, which means it is likely that it is necessary to also look at the possible evolving role of ontologies in defining business focused vocabularies.

1.1 About this document

This document is intended to provide guidelines to organizations that are using IBM Industry Models to help in building vocabulary assets that are effective in accurately reflecting the meaning and full context of the key business terms that are needed to support the specific needs of the various business focused users.

This document was created with the assistance and feedback from a number of IBM® customers who are investigating ways of further evolving their overall semantic structures to provide more appropriate artifacts for use by the various business users.

1.2 The motivations for this document

IBM Industry Models have provided pre-defined model content for both business vocabulary style models (glossaries and taxonomies) and structured logical design models (ER , UML and process) for many years. Traditionally these models were intended for use as part of various design time activities and so were designed for use by specific personas (business analysts, enterprise architects, modelers). However, there is now a requirement for the business vocabulary content to also be made available to the various business users who are involved in the runtime production environment. These users range from business or departmental users with likely a limited scope of the overall vocabulary to data scientists and citizen analysts who require a broader scope.

This change poses questions about what are the considerations, approaches and best practices for using these business vocabulary models in this new context. This document is intended to assist organizations that are faced with such questions.

This document is intended to provide the necessary guidance, and considerations to organizations that wish to use IBM Industry Models as a basis for building various user-focused glossaries (local glossaries, departmental glossaries or common enterprise-wide glossaries for use with all business users). A key consideration is how to enable the creation and ongoing maintenance of such glossaries but also to retain the necessary linkage back to the broader IBM taxonomies, which provide more details and precision that is needed for the technical users and their activities.

1.3 The scope of the business vocabulary in this document

The concept of a business vocabulary can mean different things to different organizations. In many cases such a vocabulary could consists of a range of components such as business terms, categories, rules, policies, rules, etc. The focus of this document is the core set of Business Terms, the categorisation of those terms and the relationships both between terms and between terms other assets.

1.4 The underlying tooling used in this document

This document is intended to be independent of any specific underlying vocabulary tooling. The overall discussions and considerations in this document should be applicable to most typical vocabulary tools. In cases where it was necessary to show examples in a particular tool, IBM InfoSphere® Information Governance Catalog (IGC) is used. For IGC-specific considerations, these are provided in a separate document.[1] However in some cases in this document it may be necessary to reference specific sophisticated solutions that may be possible to implement only with certain dedicated metadata management software such as IGC.

1.5 Intended audience

This document is intended for anybody who is considering using IBM Industry Models to assist with the definition, and governance of a cross enterprise common vocabulary that reflects and underpins the various business activities.

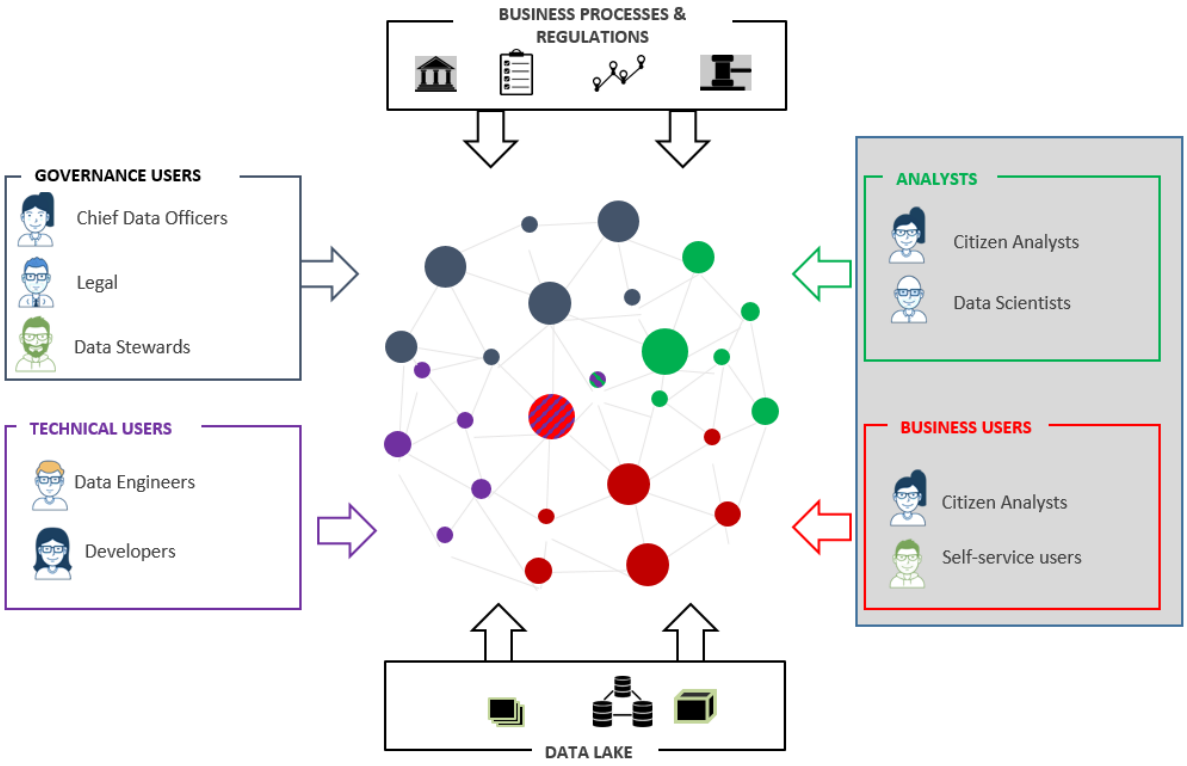
[1] Future document: “IBM Industry Models and InfoSphere Information Governance Catalog - Approaches and guidelines”.

2 The overall context in which vocabularies are used

Fundamentally the business vocabulary is the means by which the business language is created and maintained. This business vocabulary is intended to provide the common meaning for all users across the enterprise.

2.1 Addressing the needs of many users

That common business vocabulary needs to address the different needs of the various users across the organization.



As shown in the above diagram there would typically be a range of different users. These groups of users have potentially radically different needs in terms of the subset of the overall business vocabulary that they use as part of their day to day activities. The specific focus of this document is the analysts and business users in the shaded area on the right side of the above diagram.

There is a persistent requirement from organizations that a holistic view is taken in providing these different views, so that, for example, a simple glossary of terms being used by one set of business users is a subset of a broader more complex taxonomy or ontology that is used by more technical users. And when there is a change in the underlying ecosystem being described by the business vocabulary (either in terms of the underlying physical assets in the environment or new or changed business requirements or regulations), that this integrated business vocabulary can be efficiently adapted to ensure that the views of the different users are changed in a coordinated and consistent way.

2.2 The value of a business vocabulary

While the business vocabulary is often defined as a way to provide a business view to connect with more technical artifacts or as the starting point for defining the business scope to be used to guide the identification of a subset of logical model elements, the business vocabulary has significant value in its own right.

A properly defined business vocabulary that reflected the needs and terminology of the various users can be used to:

- Act as a common language across the different business users across the organization.
- Provide a reference point when aligning new or acquired businesses into the overall organization.
- Provide a basis the identification of gaps and overlaps between different projects or activities that the enterprise may be engaged in.

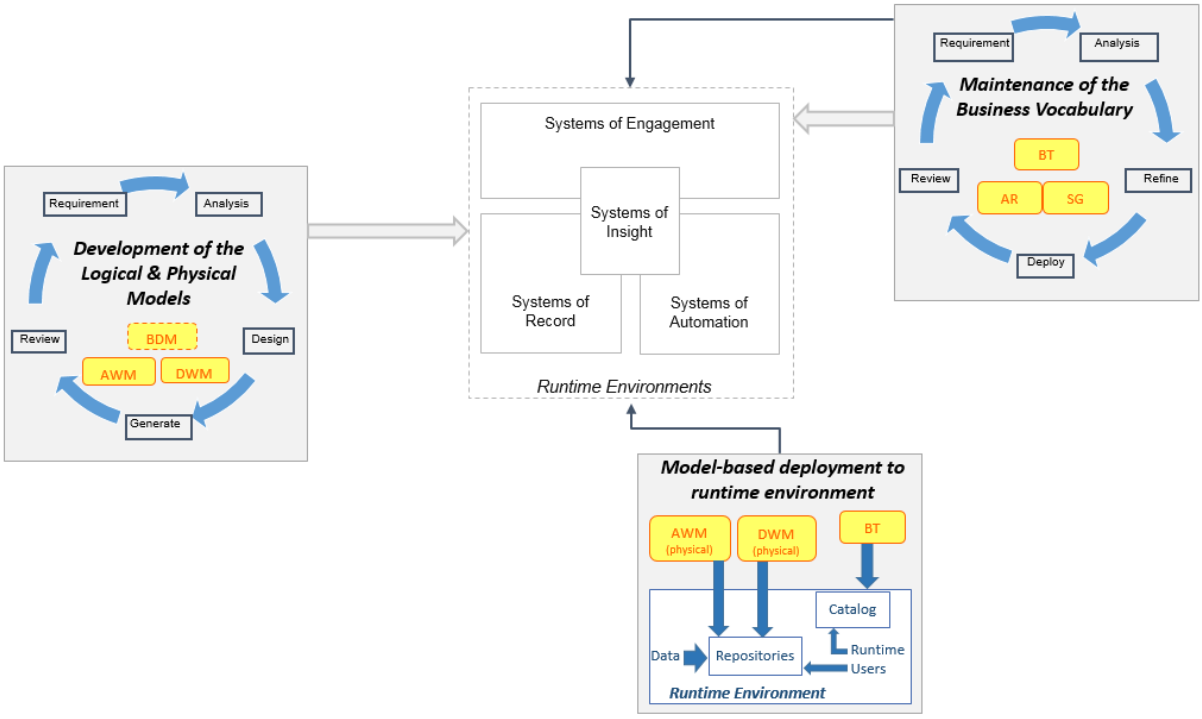
2.3 The enterprise data principles

The attitude of the enterprise towards the data that is being described is critical in shaping the business vocabulary. There are a number of key enterprise data principles or capabilities that drive the need for enterprise business vocabulary and its linkage to local (departmental) business vocabularies where they exist for many organizations. Such enterprise data principles include:

1. Data is an Asset: Data is an asset that has value to the enterprise and is managed accordingly.
2. Data is Defined: Data is defined consistently throughout the enterprise, and the definitions are understandable and available to all users.
3. Data is Shared: Users have access to the data necessary to perform their duties. Therefore, data is shared across enterprise functions and organisations.
4. Data is Accessible: Data is easily accessible for users to perform their functions.
5. Data is Governed: Each data element has a trustee accountable for data quality.
6. Data is Managed: All organizations in the enterprise participate in information management decisions needed to accomplish business objectives.
7. Data is Secured: Data is protected from unauthorized use and disclosure.
8. Data is in context: Data is related to their functions and scenarios.

2.4 The different lifecycles

From an IBM Industry Models perspective, the creation and maintenance of this business vocabulary is one distinct part of a broader landscape. This landscape is typically concerned with three different but related lifecycles.



The three lifecycles shown in the above diagram are:

Maintenance of the business vocabulary.

This is related to the creation and maintenance of the business language by using the business vocabularies from IBM Industry Models and other related artifacts. This lifecycle is responsible for defining the *business meaning* of the various artifacts being described, and as such this lifecycle needs to have a significant involvement of the relevant business organisations. If this business vocabulary becomes too technical or in other ways does not resonate with the business, then it will not succeed.

The development, governance processes and associated tooling for managing this lifecycle are typically oriented towards business or less technical personnel.

Development of the logical and physical models

This lifecycle is related to the creation of the associated data, process or services models which are intended to enable the *definition of the various structures* that are required to define the physical environment. While heavily influenced by the business, the personnel involved in the detail of this lifecycle (data modelers, data architects, business analysts) are of a more technical nature and associated processes and tooling underpinning are of a more technical nature. In fact, the normal IT artifact governance and management processes are employed here.

There should also be a feedback loop to the maintenance of the business vocabulary. In many cases it may be necessary to validate or correct some of the assumptions that were used in the definition of the business vocabulary.

Model-based deployment to runtime environment.

This lifecycle is related to the onboarding of the business vocabulary and the associated model-derived structures into the runtime environment and the feedback from that environment to the processes that manage the ongoing maintenance of the business vocabulary. For more details on the specific steps associated with this and the previous two lifecycles, refer to the separate document "*IBM Industry Models support for a data lake architecture".* [2]

[2] http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=IMW14877USEN

3 General considerations for defining the business vocabulary

3.1 What is a business user-focused glossary

Across the typical physical environment for which organizations need to govern with the use of a business vocabulary, there can be a range of different Vocabulary elements and grouping corresponding to the range of potentially different users (for example, business analysts, IT users, designers, modelers, data scientists, business users). These different users have different needs based on their technical knowledge, the scope of the environment they need to see, the frequency of usage and the types of access.

Addressing business users needs

This document is predominantly focused what is required for a glossary to support the needs of the Business Users. The characteristics of such business users would typically include:

- Primarily interested in the business-specific terms for a particular area of the landscape
- Is not of a technical nature, has no interest in the more technical aspects of the full glossary.
- Required to have the content presented in a simple easy to understand form.
- Might require a number of different ways of being able to visualize the results (for example simple lists of terms, or network graphs of related terms)
- Some Business Users might be interested in metadata KPIs (Reporting, Dashboard)
- The profile that is applied to a business user might change depending on the role that business user is adopting during a particular activity.

Characteristics of a glossary focused on business Users.

While it varies from organization to organization, the typical characteristics of a glossary that is focused on Business users include:

- A thin layer of business-focused categorizations that sits on top of a (more technical) taxonomy.
- A common set of categories across all departments.
- Primary focus is on the real business users - the people who consume/use the category.
- The Categories and contained terms must be business meaningful and understood by business users without the need for explanation.
- Likely that the structure/hierarchy of the categories depend on how the organization is structured

There is potentially a need for a separate set of considerations for the set of users who are authors or otherwise involved in the term development process. In this case, these users might need to see the broader taxonomy in order to make decisions on whether new or changed terms are suitable for promotion to be used by the broader set of business users.

Characteristics of functional views

An extension of the business user glossary is the concept of creating a set of Functional Views, which are set of business terms that are grouped according to specific functional topics. So, for example, an organization might group a set of terms that are pertinent to Credit Risk. The advantage of such views is that it promotes the use of common terms by business users across different departments. The business functional views should be flexible.

It is important to enable the growth of such views to address specific business areas or reporting functions. It might be possible to allow groups of users to define their own groupings - subject to the appropriate governance process.

The purpose of the business vocabulary is to provide a consistent set of terms that can be used by both business and technical teams to describe the information in their physical environment. The business vocabulary is the initial contact point for most users and so it is imperative that the language creates a positive first impression of the overall IT environment's usefulness and quality. Therefore, the effort to create and maintain the vocabulary it with relevant content is a key success factor.

3.2 Ownership of the business vocabulary

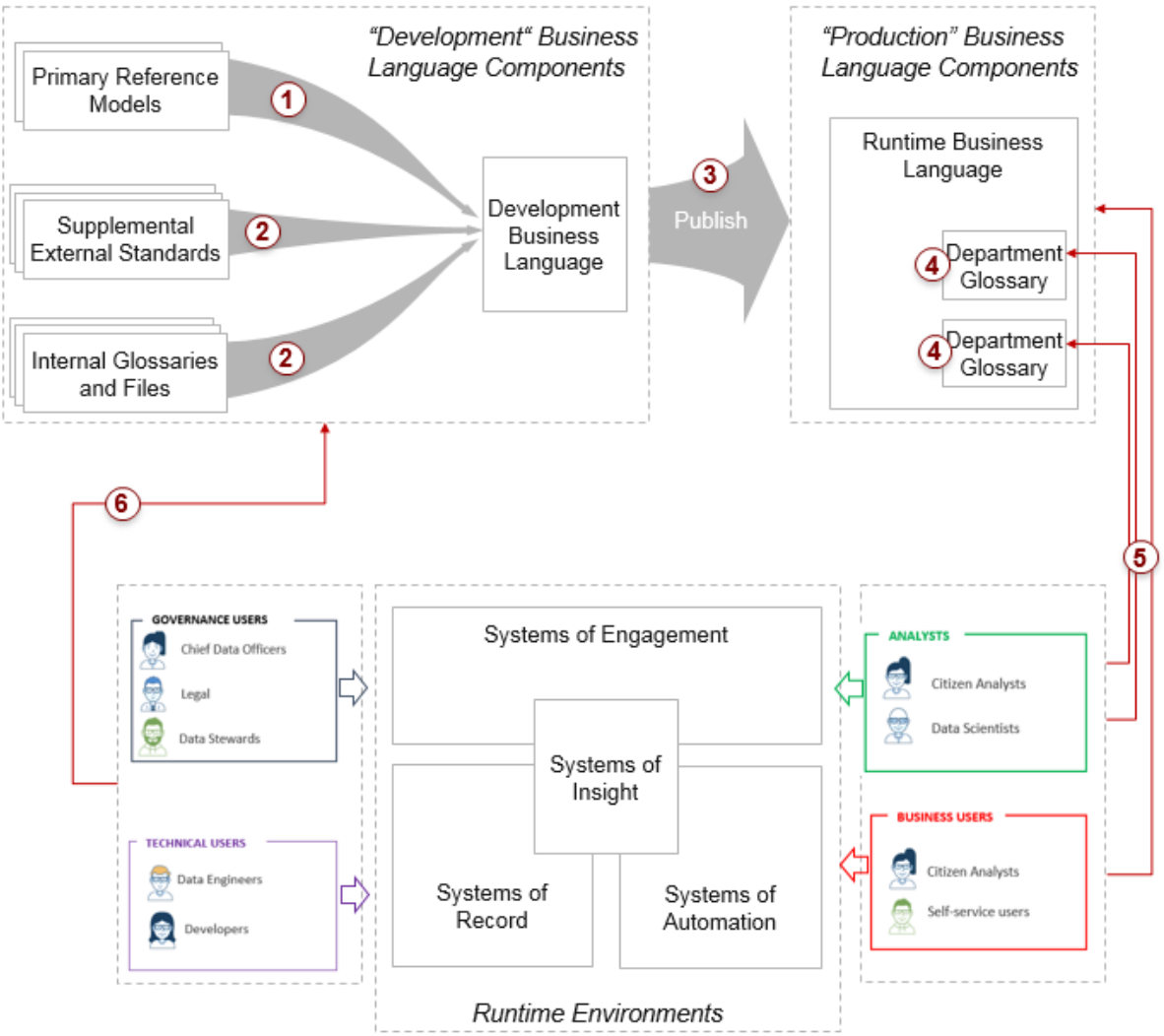
One key fundamental consideration is regarding the ownership of the central business vocabulary. Traditionally where such central business vocabularies exist they tend to be curated by the IT organization on behalf of the business. But, in reality, in many cases the different aspects of the business often had quite radically different senses of ownership of such an asset. At an executive level, there might have been a certain sponsorship or at least tolerance for the need for such a central vocabulary, but often it was difficult for the various user departments to see the benefit to them of such a vocabulary, especially when the application of the governance necessary to ensure/enforce commonality was seen to be slowing things down when it came to their individual projects, especially for projects which we were the initial pilots for the evolution of such a business vocabulary.

The advent of the role of a chief data officer provides many organizations with a C-level executive role who would typically be supportive of the need for such a common vocabulary, and therefore potentially provides a good non-IT champion for this asset.

However, even in such cases, there is still the need for the active executive sponsorship to enable the organization to invest the necessary costs and resources to create and maintain a vocabulary.

3.3 Define the components of the business vocabulary

In setting up the business vocabulary for the environment, it is important to consider the required components to manage both the provisioning of terms and associated artifacts and mappings regularly as well as the need to manage and control the access and usage of the contents of the business vocabulary.



The diagram above outlines what the possible components necessary for the setting up and maintenance of business vocabulary might be. The main components are:

- Development business vocabulary – This is the set of business vocabulary components under development by the organization and have not yet been published for use by the users of the runtime environment.
- External Input – This relates to external models, documents, taxonomies that the organization uses as input to the development of the business vocabulary. Examples of this might be IBM Industry Model content, regulatory taxonomies, or DBpedia.
- Runtime business vocabulary – This is the set of business vocabulary components that are certified by the relevant stewards/curators as suitable for use as part of the business vocabulary to describe the environment.
- Departmental glossaries – These are specific subsets of the business vocabulary that are identified as necessary to give groupings of business users the set of business terms that are specific to their business domain and by using their vernacular. A specific variation of departmental glossaries that will be covered later is the concept of Functional Views. Whereas departmental glossaries are a collection of terms that are aimed at the need of a particular department, functional views are where there is a need to enforce a common set of business terms for a particular function or subject across the enterprise.

The specific activities that are outlined in the above diagram are:

1 The use of a primary reference model to provide the overall framework on which the business vocabulary is based. This would typically be the central business terms that are provided with IBM Industry Models, for organizations using these models.
2 In addition to the primary reference model, the business vocabulary might also be supplemented or implemented by a range of other inputs such as other external standards and local internal glossaries. If it is expected that there will be periodic future updates from any source for a particular input, then it might be useful to retain an area where such inputs can be landed and examined.
3 Promotion to the runtime business vocabulary, normally once the normal governance lifecycle activities have been followed to ensure that the components of the business vocabulary adhere to any standards regarding terminology, structure, and so on.
4 There might be a need to create or update departmental glossaries where certain views of terms are created to support the needs of different groups of users. Or where possible, it might be possible to set up functional views to enforce a common view of certain functions for all users across the enterprise.
5 Access to and use of the business vocabulary by the business users
6 Feedback to the development business vocabularies regarding any changes/additions that are seen as necessary based on ongoing usage. In terms of the feedback/influence of the technical users, it is necessary to consider potentially different roles of data engineers and developers – while because both have access to both environments, however developers are more likely to have a more hands-on maintenance role. Examples of the consequences of this type of feedback include :
    - Requesting access to additional data elements for specific users or groups of users
    - Provisioning data based on the glossaries content to a separate data discovery sandbox

3.4 Determine the needs of the different users

A key consideration is that there are likely to be a range of different types of users all attempting to avail of this business vocabulary. These uses are likely to be addressing different requirements, they have different levels of business and technical knowledge and have different needs in terms of the breadth of the physical environment they are interested in. There is also the important consideration of security/ownership considerations of the data.

In general, the considerations pertaining to the different types of users and how the business vocabulary supports them are:

The scope of business vocabulary that is suitable for the user.

In some cases - such as the IT Operations personnel, the Data Curator and potentially Data Scientists - there might be a need for the user to have access to the full set of terms in the business vocabulary. In other cases, there might be a desire to restrict to certain users to only the subset of

the business vocabulary that pertains to their area. This would have the benefit of presenting to users just the terms that they understand and are familiar with and avoiding overloading any search results with terms that are alien to the users. However, the converse is that in some cases there might be some lost opportunities when these users are not aware of other terms in other domains that might be of use to them in their role.

Another scope-related consideration is to decide whether or to what extent to include links in the business vocabulary to the upstream Business Processes in the Systems of Record and of Engagement.

With the question of scope come the security and ownership considerations.

Experimental or Production areas of the IT environment.

In most cases, it is expected that environments such as data lakes will consist of data scientists and other analysts carrying out more experimental or investigative activities in addition to the business users more interested in the production data and associated queries. The data and query needs and behavior of these two sets of users are likely to be different and might influence the specific aspects of the business vocabulary being used by these users. More details on this consideration will be provided in *Chapter 5 – Governance Considerations*.

Whether there is a need for any local dialects.

In many cases in larger organizations, there might be cases where groups of users have their own vernacular of terms specific to them which are unknown or irrelevant outside of their group. Or perhaps there are synonyms for their terms that are called something different in other areas. It would be important to define the policy towards the support, or not, of such dialects. A lot would depend on the normal expectations and cultural norms in place in the organization. One approach might be to try to define a single set of terms for use by all users, alternatively to allow for the use of terms specific to a department, but in this case, it might be useful if such terms were linked as synonyms to the equivalent enterprise version of the term to ensure clarity of communication.

What is the level of structure that is required to be exposed?

There is likely to be varying degrees of hierarchical structure built into the business vocabulary from a simple flat glossary of business terms with little hierarchical structure to taxonomies with extensive and deep hierarchies of business terms. The latter might be the case in the Banking FM, Telco and Retail models where such deep taxonomies are provided.

The use of such hierarchies of business terms has the advantage of providing a degree of precision when it comes to understanding the semantic relationships between terms and the associated mappings those terms have to the physical assets. However, in general, such deep taxonomies can sometimes not resonate with the departmental business users, where they might prefer just to see a simple list of the relevant terms.

One option might be to retain the use of the taxonomies along with the deep hierarchies, but to ensure that these hierarchies are only visible to the users for whom such detail is welcome and relevant, with the mainstream business users just having visibility to the individual set of business terms that make sense in their specific context.

Business and technical terms.

There might be a need to have a collection of both business and technical terms in the vocabulary. In some cases, it might be necessary to ensure that detailed technical language is not exposed unnecessarily to business users. However, in other cases, for specific groups of users, the decision might be taken to provide additional more technical terms that to provide further context for the business term. An example of this would be the use of more generic or non-business terms to describe with more precision the location of business terms in an overall taxonomy.

What artifacts to include in the business vocabulary

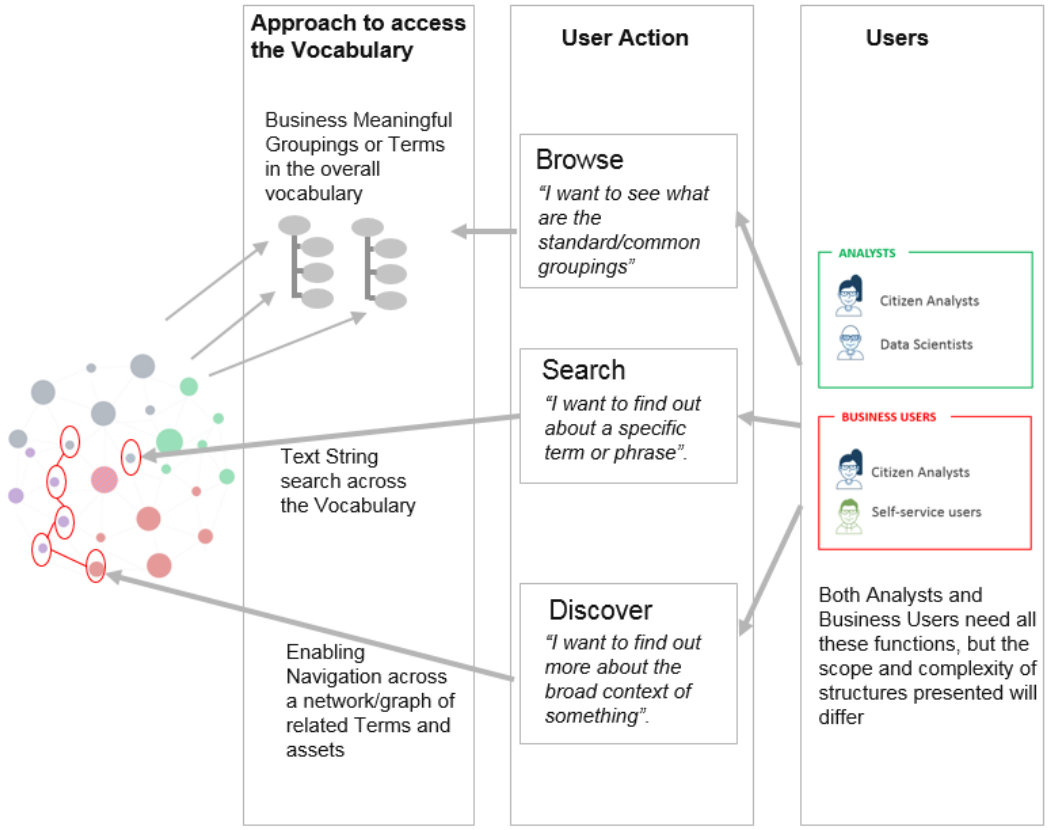
While the use of a set of business terms with appropriate definitions is likely to form the backbone of any business vocabulary, some consideration might also be given to what other artifacts to expose as part of the business vocabulary. Specifically, it is a question whether to include business metadata and technical metadata in the overall business vocabulary.

For example, in some cases it might be useful for the business vocabulary to also include parts of the supporting conceptual, logical, and physical models in order to give further background or context to a particular term or set of terms. Some, or all, of the users might find it beneficial to have visibility to glossary elements that describe the broader landscape beyond their own particular area. The benefits of users having a wider visibility across the vocabulary include:

- Better enforcement of a common business language is possible when users are aware of other terms/synonyms in use across the organization.
- There is a greater likelihood of and encouragement of reuse when users are able to see what artifacts exist across the enterprise that they are available.
- The collaboration and joint activities across departments is assisted by the users have an understanding of the vocabulary in use in other areas of the business.
- The ability for the organization overall to be more adaptive and reactive is enhanced by business users in the various departments having an understanding of the terms/language in other areas.

3.5 Typical actions of the business users

In general, it is possible to broadly categorize the different types of actions that are carried out by the users into three different categories.



The diagram above outlines these three different types of action:

- Search – this is there the user knows the specific name (business or technical) of what they are looking for.
- Browse – this is where the user knows what they are looking for but may not know the precise name. They require to get a limited result set and might also want to look through the standard groupings or collections of terms. They are probably interested in seeing what is of interest to their colleagues in general.
- Discover – this is where the user wants to find new things, doesn't know exactly what they are looking for and mainly starting with entering context names (like Customer in general) or asking questions in natural language. They might also be looking to navigate or surf around for the terms and other artifacts that might be broadly connected to a concept.

3.6 IBM Industry Model artifacts involved

IBM Industry Model Business Vocabularies provide an accelerator for the creation of this language by providing a predefined language that describes that information that represented in the data models but is independent of the technical structure of the data models. IBM Industry Models have three components delivered in Information Governance Catalog (IGC) that can be used as the starting point of defining the business vocabulary.

Business Terms

The Business Terms component of IBM Industry Models typically provides the central enterprise-wide taxonomy of business terms intended for enforcing a common terminology across different IT and Business-focused users. Typically, these taxonomies can be large as they are intended to cover all possible aspects of an Industry (for example the Banking set of business terms in these models has more than 9000 defined terms). So, the typical method assumes that only the appropriate subset of these terms is used for any particular deployment.

Once this taxonomy of business Terms is customized to suit the organization's needs, it is then recommended good starting point for the creation of any business-user focused glossaries. In almost all cases, the approach is to:

- Identify the (often small) subset of terms from the IBM provided taxonomy that makes sense to a particular glossary for use with a specific group of business users
- Map any local terms that are used by the business which do not exactly match the enterprise terms and determine if/how to handle these terms – for example mapping them to the equivalent term in the taxonomy as a synonym.

Analytical Requirements

The Analytical Requirements components of IBM Industry Models comprise high-level reporting information and business measures along the axes of common dimensions. While the Analytical Requirements are primarily designed to allow business users to rapidly map reporting requirements to the data models, they can also be used as the part of a business vocabulary that is focused on supporting calculated values or KPIs.

In terms of building a business user focused glossary, this list of KPIs and associated dimensions could provide some valuable examples of the potential component terms of such a glossary, especially as such terms are defined to reflect business language and most of these terms have mappings back to the central taxonomy.

Supportive Glossary

The Supportive Glossary incorporate terminology that originates from external sources such as regulatory authorities and industry standard bodies. The terms in these structures are deliberately created to reflect the language that is used in the particular source application or regulation being described. So, there might be potential value in using such structures as a source of terms if there is a need to define a business users glossary that relates to the area they cover. It is more likely that such structures are most useful when defining a Functional View that equates to an area of the regulation that is already described in a particular supportive content structure.

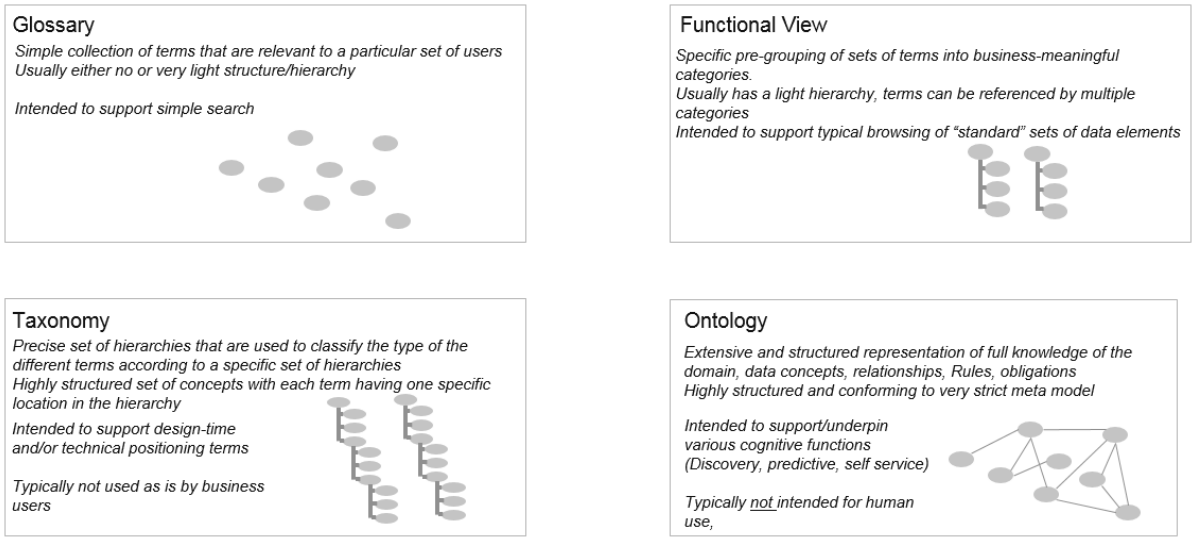
4 Building a business user focused vocabulary

This chapter describes the potential structural components of a business vocabulary and how they can be combined to create an overall enterprise vocabulary. This chapter also describes the different approaches to ensuring both that the specific needs of the business users are met and how the terms they are viewing are appropriately integrated with the broader vocabulary for broader use by other users.

4.1 Types of vocabulary components

The type of components of the business vocabulary varies across different organizations, but some common components can be as follows:

- Simple glossary of terms
- Taxonomy
- Ontology
- Functional View



The diagram above outlines the definitions, characteristics and usages of these different vocabulary components.

With most organizations today, simple glossaries and functional views are the predominant structures that are used for deploying vocabularies to the business users. Where taxonomies are used, it is typically only the small subsets of such taxonomies that are meaningful to business users, with the broader taxonomy being used by the technical users.

IBM Industry Models typically provide the following types of structures;

- Functional View – the Supportive Content and in some cases the Analytical Requirements in IBM Industry Models can provide the basis for functional views
- Taxonomy – The central set of Business Terms provided by IBM Industry Models would typically be used by organizations as the starting point for the definition of the central taxonomy in their overall Business vocabulary. The taxonomy is also the ultimate reference

point for business users who want to understand exactly the positioning of a particular term relative to other terms (for example, parent and child terms).

Finally, the role of ontologies in this landscape is something that is only just beginning. While pure ontologies, which take a formal and structured format, are typically not suitable for direct user by business users, there is a potentially significant role in such ontologies to underpin the applications used by business users to provide an improved self-service and guided experience.

4.2 Functional views v's departmental glossaries

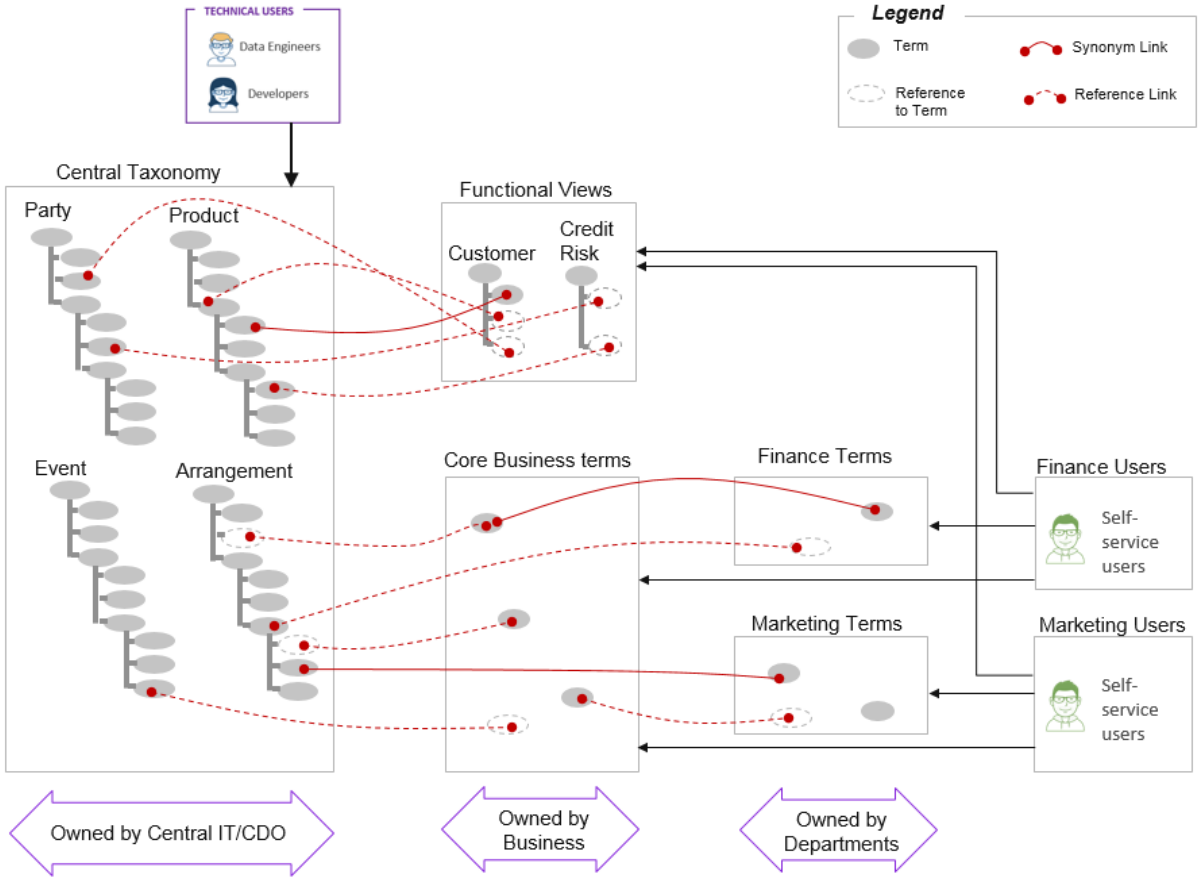
This document calls out two different but often overlapping concepts: Functional Views and Departmental glossaries.

The functional views are a set of additional supplementary categorizations on top of the enterprise-wide business vocabulary which are intended to provide standard and consistent groupings of business terms for specific concepts. The functional views are the means by which an enterprise can enforce commonality for what is perceived to be standard concepts or functions. So, functional views are typically top down and are subject to the same governance as the other parts of the business vocabulary.

Departmental glossaries are the set of terms that are used by a particular grouping of business users. The departmental glossaries can use a combination of terms that are managed as part of the enterprise business vocabulary and terms that are purely local to that department. The main objective of such departmental glossaries is to provide the ability for organizations to achieve a compromise where there are differences in terms that are defined as standard by the enterprise and often-conflicting terms that may be traditionally in use by groups of users in different departments.

4.3 Combining the components to create an overall business vocabulary

In building a comprehensive enterprise business vocabulary that adequately addresses the needs of the different users and associated activities, it is likely that there is a need for a combination of these components to be integrated together.



The above diagram outlines what can be considered one potential implementation of a set of business focused glossaries. In this case, the decision has been taken to align the business glossaries according to the organizational/departmental groupings that already exist in the organization. So, a glossary for all the uses in the finance department, and so on.

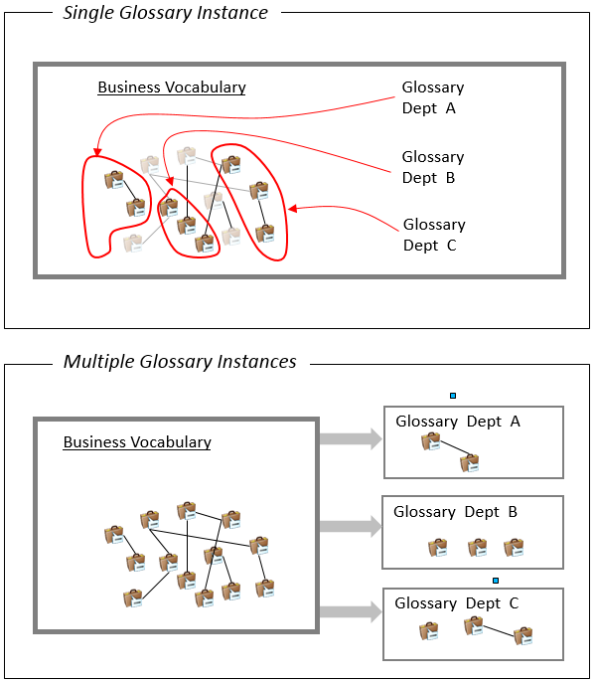
The requirement here is to ensure that the users in each department see only the subset of the overall set of terms which are meaningful to them, so that they are not flooded with terms from across the organization. To achieve this there are two layers of visibility determining the appropriate subset of terms which users can see:

- The separate departmental grouping (for example, finance specific terms) means it is possible to isolate terms that are only of interest just to that department. However, it would be important that these groupings are easily reconfigurable to reflect changes in the business and in the organizational structures.
- The supplemental cross-department cross business terms can be used to store all terms that are of interest to all users.

Finally, all more technical terms remain owned by IT and most likely not visible to business users. It should be possible, dependent on tooling setup, that even though these terms are grouped in these different categories, the inherent linkage between terms across grouping should remain. So, an IT user can see the full set of technical relationships to a term, even though a business user might just see that term in isolation as part of a simple flat glossary.

Level of integration of business vocabulary and departmental glossary

A main consideration in setting up departmental glossaries is the determination whether they are simply views across the broad enterprise business vocabulary or whether they are separate from but related to the enterprise business vocabulary.



The main advantage of a single overall glossary instance incorporating the enterprise business vocabulary and any associated views is that it is a simpler environment to set up with the least amount of initial and ongoing overhead. This would be the assumed default starting point for many organizations. In such a case, it is typically possible to enable different viewing profiles to ensure that users see only the relevant subset to them and the management of separate synonyms is possible.

The setting up of multiple glossaries is typically considered where there might be a need for different infrastructures (for example with different geographic/country locations) or where different governance processes/workflows are required for different departments. However, the setting up and ongoing maintenance of multiple glossary support in the current tooling environments will result in significant levels of complexity and overhead. This approach should be taken only after careful consideration of the potentially large levels of additional resources that are required and other implications in terms of potential duplication, governance issues, tooling limitation.

4.4 Approaches for defining a glossary suitable for business users

One other consideration in relation to the setting up of the business vocabulary is to determine the appropriate glossary of terms that is made available to the business users. In many cases, the full content of the business vocabulary containing many technical terms and potentially artificial constructs in taxonomies means that there is a need to identify a subset of the overall business vocabulary to expose to the business users. Indeed, it would be critical to the overall success that the subset of the business vocabulary that is used by the business users for their regular search and discovery activities is something that they understand and are comfortable with.

Some of the likely characteristics that such subsets of the business vocabulary should have are:

- Meaningful to the business users, both the term itself and any associated definitions and related artifacts
- Any grouping of terms of logical/natural to business users
- Subject oriented
- The glossary of terms for any set of business users to be of limited size
- Designed with consumption by business user in mind
- Should be able to find the item in reality (should not be simply abstract) - anchored on business reality
- Might be grouped based on industry standard groupings

Bottom up and top down

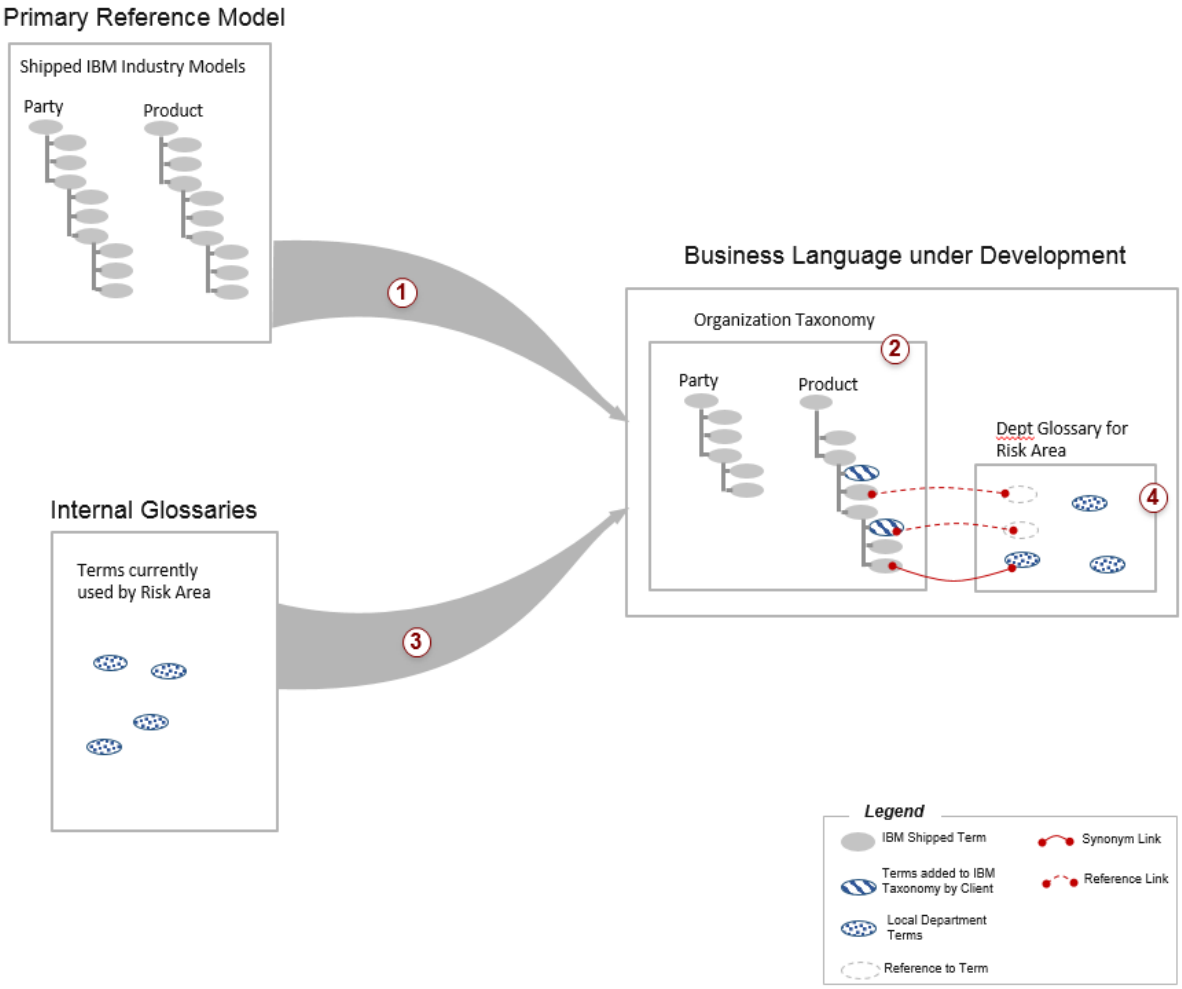
As part of defining the glossary for business users, there is likely to be a need to achieve a balance between influencing the glossary with the set of terminology as used by the business users (bottom up) with the definition of a standard set of terms and associated relationships to be propagated across the enterprise (Top down). If the glossary is overly influenced by the actual business terms that are used by each department, there might be a challenge with the establishment of a true cross enterprise business vocabulary. So, it is critical that there is executive support to enable the necessary tasks and activities to ensure that such a cross-enterprise vocabulary is the goal from the beginning.

However, if the glossary is overly influenced by a top down canonical structure, it might alienate some of the business users with the potential use of overly generic terms or terms they don't use or recognize.

The ideal is likely to be a balance involving the use of synonyms where the conflict of various local business terms is unavoidable. There is also a key role for the business vocabulary stakeholders to ensure that an appropriate management process is in place with provision to accommodating the requirements of the different groups of users.

Building a central taxonomy and departmental glossary from IBM Industry Models

As mentioned previously, it is possible to use IBM industry Models as a basis for the central taxonomy in an organization and as a reference point for various associated departmental glossaries.



The above diagram shows the potential steps that can be followed by using an IBM Industry Models vocabulary, specifically the central Business Terms. The starting point is the separate storage of the models as provided by IBM. These will remain stored separate from any derived taxonomy in order to:

- Retain a clean copy of IBM Industry Models in which to carry out any upgrades to subsequent versions of these models that are provided by IBM
- Ensure appropriate isolation of any terms that are created and reused by the client organization from the IBM-provided business terms to ensure any history and lineage for these different sets of terms can be maintained.

The possible steps that can be followed in the creation of a central taxonomy and associated departmental glossary are listed below. All of these steps are assumed to be subject to the appropriate level of governance that is put in place by the organization for managing the growth of the overall business vocabulary.

1. The appropriate subset of the terms needed by the enterprise are created in a separate development structure from the IBM provided models. Where applicable the terms and structures that used in the IBM models can be reused in the customer taxonomy.
2. The taxonomy is then updated to include any further customizations that are required where there is a need for additional terms or changes to the IBM provided models. This

customized taxonomy then becomes the central reference point across the enterprise for metadata mapping.

3. The set of potential business terms that need to be considered for a particular local or departmental glossary are defined/imported into the glossary development process. This means that these terms can then be reviewed for suitability and equivalence with the central taxonomy and potentially with other department glossaries.
4. As the terms are analyzed and approved for use, the development version of the proposed department glossary is created. This department glossary can contain: its own terms which might or might not have mappings to the taxonomy; references to terms in the taxonomy; and terms that are local but are synonyms of terms in the taxonomy.

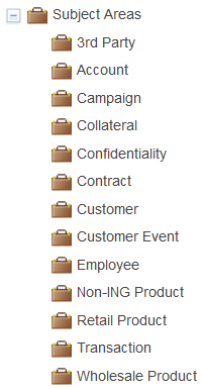
Once the departmental glossary terms have been successfully reviewed and approved, it is possible for the glossary to be published for use by the general set of business users for whom the glossary was intended. As per the workflow that is shown on the diagram in section 3.4, both the central taxonomy and the departmental glossary are subject to ongoing revision based on feedback from the various users.

4.5 How to group the terms

When determining how to create a suitable set of structures for the business vocabulary that addresses the varying needs of the different users, a key design decision is to determine the most appropriate means of grouping the terms for these different users. There are a number of options and considerations to keep in mind when defining such structure of term groupings. In general, it would be important to ensure that the category names and terms that are natural and meaningful to the business users.

Business groupings that already used by the business

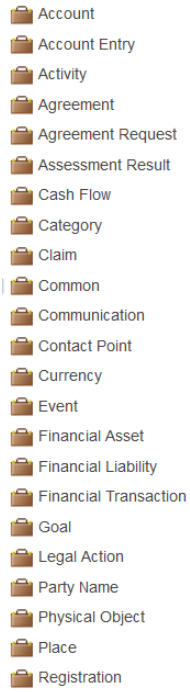
This would likely be the most obvious approach in many cases. This essentially is looking to reflect whatever natural groupings that are already in use by the business within the structure of the business glossary. It has the advantage of being something that is inherently meaningful to the business, but might result in reflecting any inconsistencies that are already inherent in these groupings.



The diagram shows a possible example of such categories that are driven by potential business-related groupings.

Data concept based

This approach might be considered where it is considered possible to expose the central data concept groupings that are used in the core business terms to the departmental business users. It has the advantage of resulting in a consistent and complete grouping scheme, however it would only be successful where these data concepts are already known to the business (or there is a willingness to educate the business users on these data concepts).



As seen in this diagram, the set of categories as provided in the core Business Terms area of the provided IBM models is an example of such a data concept grouping.

Function/Process based

A function or process based approach to grouping the terms means aligning the term groups with the specific business functions that are adopted by the organization. Very often such business processes are overlapping perhaps sharing terms, but would have the advantage of resonating with the business users as the groupings should be familiar with them

The hierarchy of the IBM Analytical Requirements is an example of such a Functional/Process category. However, as such groupings would by definition be closely aligned with each individual organization, it is likely that these groupings would need to come from within the organization.

Based on an external standard classification

In some cases, there might be external standard classification schemes that could be used. Similar to the data concept approach, it would be necessary that such external classification schemes are already known to the regular business users.

Examples of such external standards might be the Financial Industry Business Ontology (FIBO), APQC Process Classification Framework[3], or the BIAN (Banking Industry Architecture Framework) Service Landscape[4]. The diagram shows an example of the subset of the FIBO hierarchy.

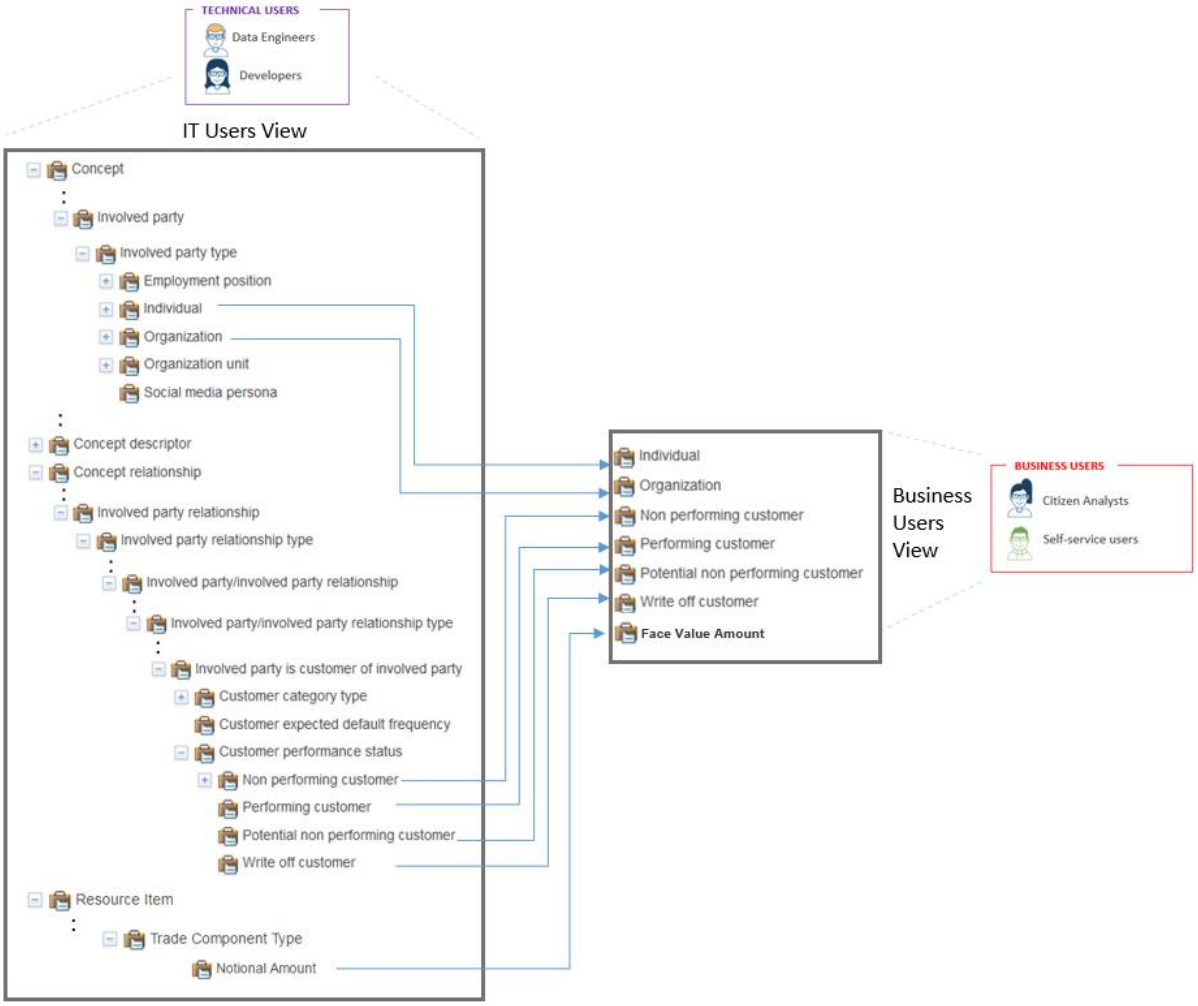
4.6 Integrating the business glossary with other business vocabulary components

Linkage to taxonomy - irrespective of whether actual terms or synonyms are used, it is important that all business user terms are linked to the central taxonomy. While it is not necessary for the normal business users to see this linkage, a greater the level of integration enables a higher level of communication between different business users and between business and IT.

Sample of departmental glossary

The diagram below shows how the set of elements available to the business users are a subset of the broader set of terms viewable by the technical users as part of an overall taxonomy. In this case the business users see the subset of the terms that are applicable/relevant to them.

[3] https://www.apqc.org/pcf
[4] https://bian.org/servicelandscape/



In addition, it is possible also set up Synonyms in cases where the terminology used by the particular set of business users is different from that in the central taxonomy. In this case the central term Notional Amount has a synonym called Face Value Amount for this particular set of business users as this term may be more reflective of the vernacular that is used by these users.

An additional consideration here is to determine the appropriate level of the taxonomy to expose to business users. Where there is a need for the users to have a fuller level of knowledge of the context of any particular business term, then the greater the level of taxonomy that they should see.

4.7 Addressing the different user personas

The approach and capabilities for using the business vocabulary will vary according to the different personas of users.

	Glossary	Functional View	Taxonomy	Ontology
Citizen Analyst	Used	Used	Unlikely to be used	Use hidden via Apps
Self Service User	Used	Used	Not Used	Use hidden via Apps
Data Scientists	Used	Used	May be used	Use hidden via Apps

Data Stewards	Used	May be used for Reference	Likely to be used	Use hidden via Apps
Data Engineers	Used	Unlikely to be used	Used	May be used
Developers	Used	Unlikely to be used	Used	May be used

The above table outlines the different user personas and their interaction with and usage of the various business vocabulary components

4.8 Extending the business vocabulary

In addition to the artifacts provided as part of IBM Industry Models, there is also the possibility to extend or supplement the provided business vocabulary content with content from other sources, either artifacts that are internal to the organization or artifacts from other third party sources such as Regulators or Industry Standard models.

Increasingly, IBM Industry Models clients are looking to define an overall business vocabulary that consists of terms and elements from a range of sources.

Incorporation of Standard Industry Models

For many organizations, the full business vocabulary might be a combination of model content from a number of different external sources. Examples of such additional standards include FIBO from OMG, the BIAN Vocabulary, ACORD, FpML.

To date the typical approach (as shown in section 3.4 of this document) has been to selectively employ the relevant subsets of such standards to supplement the business vocabulary that might have been based primarily on content from IBM Industry Models. In this case, these standards are incorporated into the business vocabulary as part of the overall governance process which defines which subsets of such standards should be promoted for use.

Incorporation of Internal pre-existing glossaries

Another crucial source of supplemental content for the business vocabulary comes from the potentially many pre-existing glossaries and vocabularies already in use across the organization. These pre-existing artifacts are often developed in isolation by specific areas of the organization.

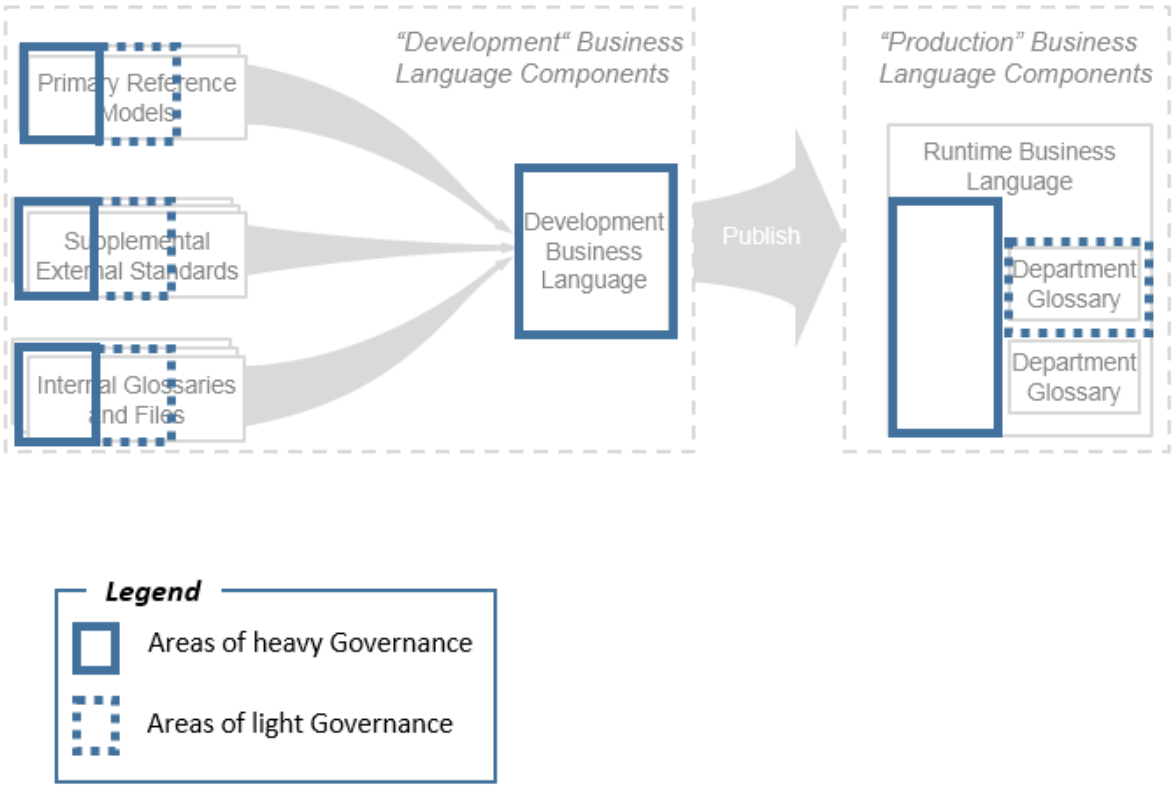
These glossaries often reflect the true vernacular in use by the different departments across the organization so it is important that the business language that is contained in these glossaries are correctly reflected as part of the overall business language. In many cases, the terminology that is contained in such glossaries can be meaningful to specific groups of users and is likely be used to populate the departmental glossaries with the appropriate synonym links to the central enterprise taxonomy. In such cases, it is still important that the use of such local or departmental synonyms is still subject to oversight by the normal governance process.

5 Governance considerations

In most cases, organizations that use IBM Industry Models to develop business vocabularies are doing so in the context of deployments for which there is a strong need for governance and management of the underlying artifacts (for example the deployment of data lakes).

Heavy and light governance

In the overall business vocabulary landscape, it might be useful to consider areas of heavy and light governance as shown in the diagram below.



For example, where there is a need to ensure a definitive enterprise-wide standard then there might be a need to enforce heavy governance covering a high level of authority that is needed to approve all artifacts and a tight focus on the standards in areas such as term definition, structure and relationships to other artifacts. However, in the case where there might be just departmental or some other smaller subset using components of the vocabulary then it might be sufficient to have a somewhat lower level of sign off, with less onerous requirements in terms of the standards applied to the term.

It might be helpful to consider the appropriate levels of governance that is required in terms of a number key areas:

- Governance team/workflow
- Local v's enterprise terms
- Dynamic v's Static governance

- Sensitivity of the Terms
- Experimental v's Production data

5.1 Governance team/workflow

It is clear that governance is not free and to achieve a level of good governance that is appropriate to the needs of an organization requires both upfront and ongoing investment in terms of executive support, commitment from the business users and participation from the central IT organization.

Executive/Business unit support

Given the focus is on defining vocabularies that are intended for use by the various business users, the support at the appropriate executive level from these units is critical to the success of the business vocabulary. The initial setup and ongoing growth and evolution of the business vocabulary requires representatives from the business to be allocated at least on a part-time basis and with the necessary authority to make decisions on behalf of their colleagues as to the contents of the vocabulary. For this to happen full support from their executives is critical.

In addition, where an organization has set up a chief data office role, then the involvement of that organization in the definition and development of the business vocabulary is critical, especially in terms of representing and championing this activity among the relevant business unit C-level execs.

Centralized governance council

Assuming that there is a need for the business vocabulary to be a mechanism for improved understanding and communication both between different business groups within the organization and between business and IT, then it is necessary that the creation and evolution of the business vocabulary is subject to a centralized governance council.

This governance council should be responsible for reviewing and approving all elements of the business vocabulary before their publication for use by the business users. This governance council would typically consist of representations from the main areas of the organization that are either creating or by using the business vocabulary.

The core objective of the governance council is to oversee the initial creation and the ongoing evolution of the business vocabulary. As it is likely that the environment in which the vocabulary is deployed is subject to constant change, the governance council is responsible for ensuring that the Vocabulary remains relevant and aligned with these changes. For an example of the typical roles that the members of such a governance council might play, see the IBM InfoSphere Information Governance Catalog Knowledge Center[5].

[5] IBM InfoSphere Information Governance Catalog - Planning, Designing, and Publishing the Catalog : http://www.ibm.com/support/knowledgecenter/SSZJPZ_11.5.0/com.ibm.swg.im.iis.bg.bestp.doc/topics/c_iadmgde_BuildingGlossary.html

5.2 Local v's enterprise terms

A critical consideration that influences the level of governance is whether the terms and associated structures are local to a particular department or are terms that are intended for use across the enterprise.

Where a term is seen to be one that is enterprise wide (for instance driven by a desire to ensure that all users in all departments have a common understanding of the business meaning of Revenue After Tax), then there is a need to ensure that all of the key business units are involved in agreeing the common definition and structure of that term. This would likely require a formal review and approval process with the agreement of all stakeholders being formally recorded. However, where a term is seen to be just of relevance and interest to a single department or group of users then it might be permissible for that term to be set up and used with little or no governance process. A key question here is how much an organization wants to invest in identifying and reviewing such terms to enhance potential future reuse across departments.

Local synonyms

One key aspect that the governance council needs to manage is the appropriate use of local synonyms. The current consensus is to allow the use of synonyms for specific groups as necessary, but it is critical that such synonyms should also be subject to the normal approval process by the central governance council. In this way, it is possible to enable to continuation of local synonyms that are meaningful to the business, but that there is a degree of standardization in terms of the quality and structure of the definitions, the links to supporting material and any links to the equivalent term in the central taxonomy. In this way, it is also possible to build out more enterprise standard terms, but also recognize that there are also enterprise synonyms - one term with specific approved/recognized synonyms.

The use of the same term as in central taxonomy means less management overhead and better enforcement of central single business vocabulary. Implications are that it might require some discussions/negotiations among different user groups to agree a single term and might also have implications for change to the taxonomy term.

The alternative approach is that the scope of the central governance council is only focused on the terms that are truly used across different departments and that such local terms that are used by just one department are not managed by the council. A critical implication of this alternative approach is that the proliferation of many such ungoverned local departmental glossaries increases the complexity of the overall environment with knock on issues in terms of training of staff to fully understand all of this local groupings. Plus, this might potentially undermine the build out of a common business language.

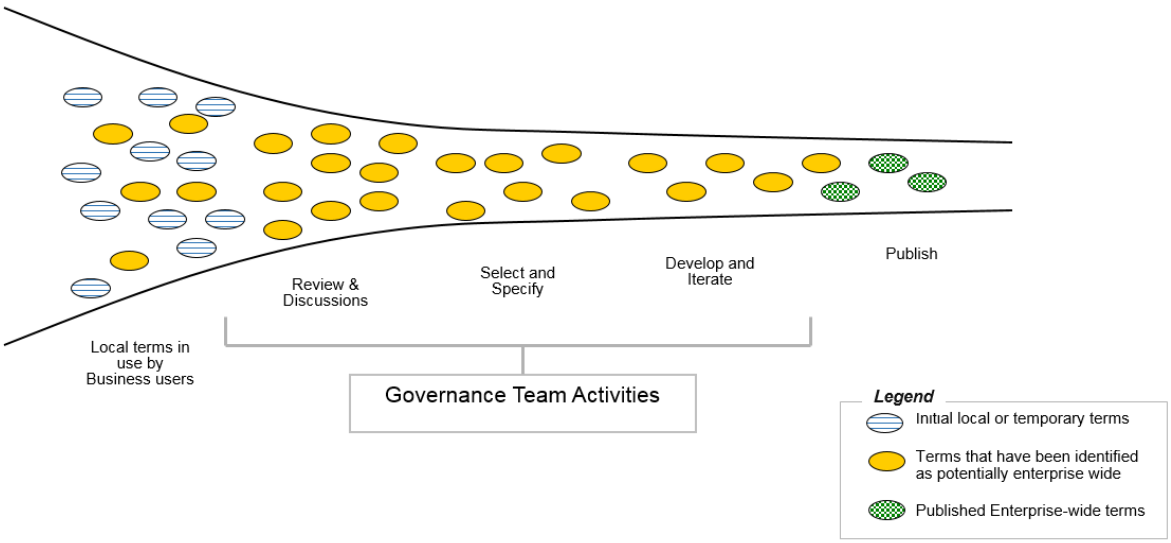
5.3 Dynamic v's static governance

In many cases, there is an advantage to having a dynamic approach to governance where the various users are encouraged to suggest/brainstorm the various terms that they would like to see in a glossary for their area. In such cases where the glossary terms are crowdsourced from the users, there is benefit in not apply a rigorous approach to governance as such an approach might impede the creation of a glossary that truly reflects the reality for these users. This needs to be balanced

with a need for strong governance that would typically enforce a higher level of standardization and consistency across the terms but potentially at a cost of the terms continuing to reflect the vernacular of the business users.

In the case of a purely local or departmental glossary, an organization might decide to continue with such a dynamic approach and to leave the business users to self-regulate their own set of terms. With the role of the governance team that is limited to recording these local terms and potentially assigning synonyms to enterprise terms where appropriate.

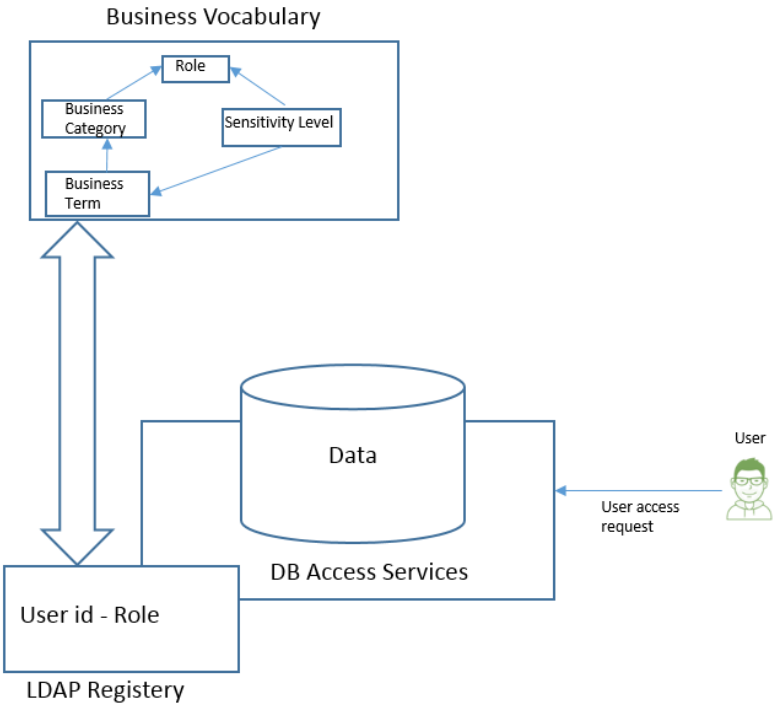
However, in the case of terms that are intended for broader or enterprise-wide usage, then how can the benefits of a dynamic approach be brought to bear? In such cases, it might be possible to consider varying the degree of static/dynamic governance according to the lifecycle of the evolution/development of that term.



The above diagram shows how local terms, which might initially be subject to a dynamic governance are gradually evolved to become subject to more static governance as they evolve towards enterprise terms. These terms are initially created (and used) as part of a dynamic and organic process among a certain set of business users. However, as a certain stage in the term lifetime it might be decided to move them to be subject to a more static governance process.

5.4 Sensitivity of the terms

A key element of the overall governance process is the sensitivity of data and combinations of data and the associated determination and enforcement of the appropriate levels of access to the different users.

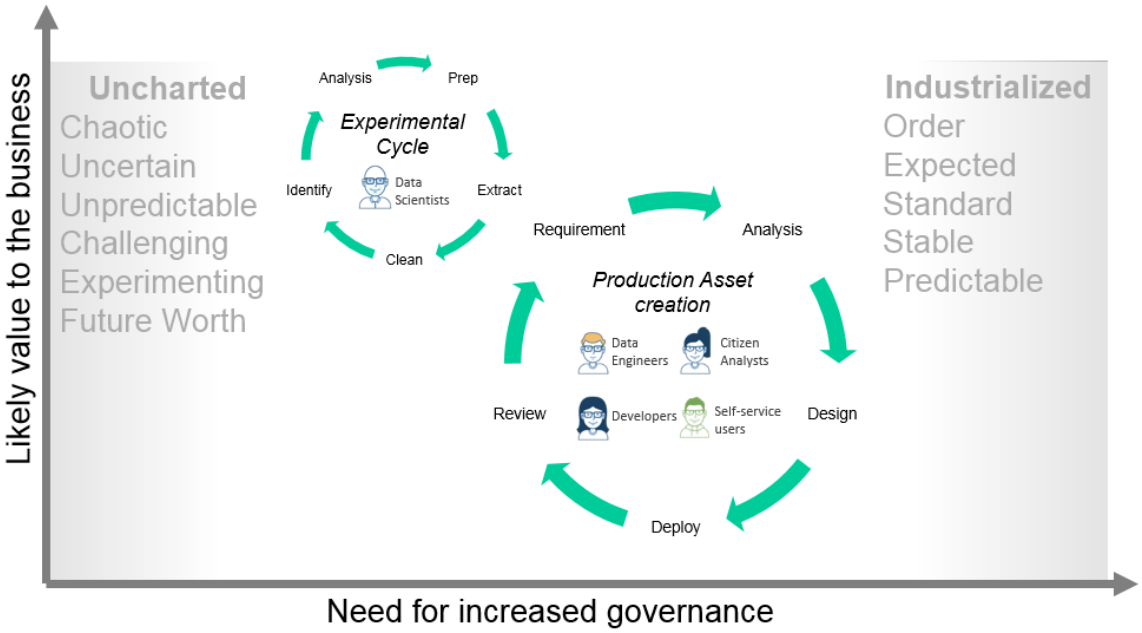


The above diagram shows an example of how a well governed business vocabulary can take an active role in the enforcement of the adequate level of security to sensitive data. In this case, the various user Roles, the sensitivity levels (for example, Open, Internal Use, Confidential, and so on) are modeled in the business vocabulary along with the business terms and categories.

Such a set-up would mean that when a user queries the data structures being described by the business vocabulary that the underlying security process, validates the role that user id has at the time of the query and what are the combinations of data that is allowed.

5.5 Experimental v's production terms

In many cases, a primary motivation for organizations to create a business-user focused vocabulary is in the context of the creation of a data lake and the need to bring the appropriate level of governance across such a wide and heterogeneous set of physical repositories. Such data lake environments would consist of the set of repositories that are intended to support the investigation and discovery activities of the data scientists. This phase can be referred to as an experimental phase, where the outputs of these experiments would influence and be used by the self-service business users in their more production oriented activities.



The above diagram shows the potential relationship between these different lifecycles when it comes to governance, including governance of the respective glossary components.

The experimental lifecycle might be quite a dynamic and unpredictable environment, where the results of such experiments are more often discarded than taken on for use by the production lifecycle. In this environment, it might be decided that the level of governance on the glossary terms that would be used to describe it would be of a much lighter nature to that used for the more conventional production activities.

However, the range of the level of governance of such terms in the experimental phase can vary across enterprises according to a number of factors:

- The overall cultural attitude towards the role of data scientists, specifically in terms of the level of control that is wanted.
- The attitude towards the potential reuse of the various data scientist sandboxes and other artifacts. The greater the level of reuse, the greater the argument for higher levels of governance over the definition of glossary terms
- The level of experience of the data science personnel. Where there is a lower level of experience, then such personnel are likely to benefit from the availability of a glossary of terms that describe the various data science sandboxes available to them.

In terms of the production lifecycle, it is assumed that there would be a need for a higher level of governance for the associated metadata assets such as the business glossary.

IBM Corporation
IBM Analytics
Route 100
Somers, NY 10589

Produced in the United States of America
December 2016

IMW14911-IEEN-00