Abstract - With the exponential increase of available, incoming data to a biopharmaceutical company, the need for consistent cross functional vocabularies is heightened. The use of effective vocabularies characterizing incoming data is one of the largest contributors to increasing data quality used for analytics and signal intelligence. This paper will describe challenges and methodologies used with business functions to develop sustainable vocabularies and vocabulary management processes. In addition, it will describe novel methods of developing cross functional vocabulary linkages for front-end processing and analytics as well as back-end signal intelligence.

Keywords – Terminology Management, Vocabulary Management, Cross Functional Signal Intelligence, Data Capture, Data Cleansing, Pharmacovigilance, Atomic Reference Layer

1. Introduction

The challenge of consistent medical coding terminology in healthcare environments has been well documented over the last two decades. With the increasing scrutiny over patient safety and exponential increase in available information, those in the pharmaceutical and medical industries are focusing on effective coding in different functions, systems, and processes, as well as alignment across those different schemes. In particular, for pharmaceutical companies, the recent trends around pharmacovigilance, that is “the science and activities relating to detection, assessment, understanding and prevention of adverse effects or any drug-related problem” (WHO, 2002), have increased focus on aggregate analysis of event data and overall risk management planning [1]. Additionally, the increase in diversity and amount of potential sources of data has exposed the lack of consistent standards in individual coding schemes as well as lack of satisfactory unified terminologies[2]. This provides distinct challenges in pharmacovigilance as potential additional adverse event data sources are increasing[1]. Therefore, terminologies that enable cross-operability between sources and functions improves the effectiveness of patient and drug safety as well as overall patient care.

Common terminologies allow reliable information to be shared and interpreted by both machine and human consumers to enable more evidence based practices and shared decision making[3].

While the vision of shared terminologies is grand and represents a definite gap emerging out of technological and process innovations, the development and agreement of those shared terminologies from potential stakeholders can be complex and difficult. Often, design sessions with stakeholders require many iterations and often exceeds time and resource expectations of those participants[4]. In addition to the initial development of the terminology, the on-going maintenance of terminologies over time requires significant resources from stakeholders to sustain consistent, usable terminologies[2].

In support of a cross-functional signal intelligence solution, we have worked with several functional areas representing disparate avenues of potential signal inputs, in order to reconcile, align, and clean existing functional coding terminologies. In addition, we have developed terminology solutions, known as the atomic reference layer, that enable cross-functional linkage between partially overlapping terminologies capturing synonymous, related, and causal relationships that support scalable machine analysis. The capture of this additional tacit knowledge in the form of terminology relationships has been essential in the providing extensibility of not only terminologies but business processes and resources.

This paper aims to provide practical experience in corporate terminology development and management. The first section will describe the aspects of terminology development, including the business context and challenges, stakeholders, and process. The second section will address maintenance issues, change management processes, and business ownership of terminologies. Finally, the third section will describe the challenges and solutions in terminology alignment and use in signal intelligence.
2. Terminology Development

2.1 Context of Business Terminologies

In any corporate environment, a company will engage with the outside world in a myriad of different methods and avenues. For a compliance driven pharmaceutical company, it is held accountable for ensuring proper storage of and action from many types of interactions from the Food and Drug Administration (FDA). The business terminologies that we worked with focused on the characterization of incoming interactions with the company. These business functions can include: sales calls, safety adverse event reporting, product failure reporting, etc. Specifically, these terminologies were descriptive, utilitarian, and custom in nature.

The purpose the terminologies is to effectively describe an interaction between an initiating party and the company. For instance, tier 1 call center agents, that is those with relatively limited subject matter expertise, are expected to assess the issue of the caller and assign appropriate terminology describing the issues. Occasionally, these issues can be escalated to more knowledgeable staff that have increased subject matter expertise and usually increased knowledge of the terminology schema.

The terminologies were hierarchical in structure for the purpose of organization as well as user navigation. However, contrary to traditional terminology structures where hierarchies are based upon broader-than or is-a relationships, these terminologies use a strict tree structure with different types of concepts at each level of the structure. For instance, the first level would be group by drug products, the second would be drug delivery method, the third level would be packaging type etc. This organization of concepts represents business understanding of the environment and is primarily used to facilitate understanding in coding and enable faster navigation through the structure.

Historically, our company grew from business functions that worked independently from one another; this resulted in terminologies that were developed independently and based upon ad-hoc need. Terminologies were generally developed to accommodate immediate business need as well as implemented system constraints.

As business needs evolved over time, terminologies were adapted to support those changing needs. For instance, in one function, the need to capture information about particular medical conference or clinical trial resulted in the addition of concepts. This change management methodology results in inconsistent terminologies. Even separate branches of an individual vocabulary were managed and developed separately resulting in duplicate terms with differing names (e.g. Cancer – Breast, and Breast Cancer). Without overarching modeling rules and guidelines, changes to terminologies were not scrutinized and aligned resulting in many duplicate, overlapping, and overly specific or overly general concepts.

Though hierarchy levels were clearly defined, business requirements around the capture of information ultimately governed the use of terms. For instance, in many cases, the characterization of an incoming interaction would involve multiple products. In order to be captured by a single coding, the use of a “Non-Product-Specific” product would be required. Strictly speaking, “Non-Product-Specific” is not a company product and does not fit the semantic definition of product, but is a case of a concept artifact generated from business requirements.

Systems supporting collection of interactions would also impose restrictions and design implications on terminologies. Because there were strict technical requirements, such as the number of coding layers of a terminology, expanding terminologies to fit changing business need was not allowed. One instance of this produced a terminology that had great variability in the bottom, leaf layer of the terminology, resulting in a coding scheme that was dense and therefore difficult to use. The increased need to capture more granular concepts forced the terminology developers to improvise a solution: place increasingly granular terms in the bottom-most layer. Over time, this bottom layer became bloated with the sheer number of terms with varying granularity; coders found it difficult to find correct terms and understand appropriate usage of terms.

Custom business need and custom system requirements resulted in terminologies that were unique and difficult to reuse. Though some externally managed standard terminologies existed in this space, these standard terminologies were not utilized in development. Similar to the complaint of many who develop terminologies for use, it was determined that the requirements for the specific internal terminology did not match with what was available externally[5].

The consistent usage of each terminology was primarily to describe the nature of an interaction, specifically the nature of the inquiry, complaint, or report by the caller. This provided clear boundaries in terms of scope and provided key guidelines in developing and cleaning these terminologies.

2.2 Stakeholders for Business Terminologies

Typically, ownership of terminologies is shared across a business unit. Subsequently, there is less of a unified vision and influence on the terminology. The skill set of the responsible parties also focused on subject matter expertise with little background in terminology management, which was reflected in the state and management of the terminologies over time.
In some cases, there was a distinct disconnect between those responsible for developing and maintaining the terminology versus those who were expected to use it. These disconnects included: lack of understanding of potential usage, differences in business process, and language, culture and education barriers; understanding of all potential stakeholders of the terminology, from those that develop and maintain the terminology to those who would consume it, is essential for success in terminology development work[6].

One primary example is the distinct difference in United States and international business clients; not only are there language and culture differences, there exist fundamental differences in business process due to country specific privacy laws and regulatory agencies. One of the major differences was simply one of resources, United States based call agents were only responsible for a very narrow section of a terminology structure, however those internationally, because of resource constraints, were responsible for coding the entire terminology. Each differing experience will drive each party’s view in developing terminologies; those in the United States pushed toward more granular and complex terms while those in international desired simpler, more general terms.

Being a United States based company, it was important to overcome bias in terminology work by gaining representation and sponsorship from both areas. Without representation from potential consumers of the terminology, lack of adoption from clients becomes a real risk; without consistent, correct use of the terminology, the value of coding records with it diminishes greatly, even for those records coded correctly.

2.3 Terminology Development Process

One major challenge in developing terminology is related to identifying the appropriate business stakeholders and owners. Stakeholders are responsible for representing their business interest in the terminologies, with the ability to make decisions for their business units. This included occasions for dispute and compromise, however the stakeholder must be empowered by the business unit into this role. Owners are required to have ultimate responsibility for the purposes, and direction of the terminology as well as vision into other related terminologies.

In addition to subject matter representation in the development process, it was critical to have cross-disciplinary roles: terminology development and management expertise, consuming information systems representation, and external standards awareness [7]. We had very strong subject matter expertise in our work with business functions, therefore providing terminology expertise was essential. It was also important to include representation from consuming information systems. One of the success factors in this process was having membership from these areas.

The development process was a hybrid model between Kremer’s model for terminology management[7] and traditional Information System development model. Most of the terminologies that we worked with already existed in some form, therefore there was a clear precedent of expected business process and usage to begin terminology development.

2.3.1 Domain Understanding

Once assembling the team, the first step was to understand the terminology domain. This included understanding the potential uses of the vocabulary; these uses ranged from characterization of incoming interactions, to reporting and analytics, to cross-functional signal intelligence. This process also included understanding current challenges and pain points from terminology users.

These initial sessions were relatively long due to the need to fully understand business process, workflow, and subject matter domain. Many of the conversations were not directly pertaining to the terminologies themselves, but rather the coding systems, business interactions and the nature of the work; our experience here is not dissimilar to those represented in the literature [4].

The design and function of terminologies satisfy business needs and reduce pain points in the business process, however it is one part of a holistic solution. It was important to distinguish which challenges and pain points were most appropriately addressed by the terminologies, and which could be addressed by training, system requirements, changes in business process etc.

For instance, a particular coding vocabulary was inconsistent in its branches causing confusion and unfamiliarity in coding. This coding vocabulary was implemented in a system that required linear, top-down system coding that caused users to guess and check what combinations of values would produce the lower level coding value that was desired, resulting in inefficient and inaccurate coding. The most apparent solution would be to clean the values of the vocabulary so that they are consistent from branch to branch, product to product. However, in addition to that, outcomes also included need for increased training as well as system requirements to enable users to search and code at any level of the hierarchy.

2.3.2 Develop Requirements

Based on initial working sessions with the business units resulted in the following requirements (focusing on the first three):

- terminology structure or model,
- terminology content,
- technical system coding,
• terminology governance process
• training

Many of the challenges and needs were addressed primarily in the first three sections. More detailed session were required for understanding governance and training processes.

Since many on the team did not have experience in principles of terminology management, it was important to communicate both the principle to abide to but also the value in doing so. For instance, Cimino outlines concept orientation as one of the principle characteristics of good terminology development; concept orientation is the notion that a term corresponds to one and only one meaning[5]. The values that were communicated to the team included cost savings in reclassification, increased fidelity in metrics, and increased accuracy in coding. Since discussions and iterations around terminologies can be onerous and long[4], it was important to be clear with the value of the end product and to have good sponsorship and buy-in from the business side.

One specific challenge in developing terminologies was that it revealed gaps and disjointed understanding across the business unit in terms of business process and workflow. This required those in the conversations to own the resolution of non-terminology issues and resolve those. In one area, we found that there was an aversion to documenting the exact definition of terms and when they should be used. From a terminology standpoint, clear definition, even those not defined by semantic relationships [5], are extremely valuable in ensuring correct usage of the coding. This is indicative of a larger issue in the business where subject matter expertise, specifically how to use coding terms, was “tribal knowledge” in that it was not well documented, passed along from one person to another and not scalable; knowledge transfer relied heavily on personal communication and one on one training of new staff. In addition, in international affiliates, where knowledge was not transferred well, adoption of coding mechanism and accuracy of coding was dramatically decreased.

2.3.3 Content Iterations

While previous steps in the development processes relied heavily on a business analyst or technical lead role, that is one who understands business process as well as taxonomic and technical considerations, working with the terminology values requires heavy involvement from taxonomists, those whose primary practice is the development of terminologies. The background of our primary taxonomist was pharmacy and terminology management; this was very appropriate given the nature of the terminology content and the work in reconciling and developing terms.

Strong business sponsorship is critical at this stage given the amount of meetings to discuss the specific content of the terminology. In our experiences, we were extremely fortunate to have strong business sponsorship and leadership in this area, which was a critical success factor.

Because of the amount of engagement required from the business as well as the dispersed ownership of the terminology, smaller working sessions with business unit representatives and the single taxonomy group was deemed most appropriate.

Content development work is typically very granular requiring constant iteration and clarification of understanding. The learning curve of the domain and specific term usage in business context was generally high resulting in slower progress in the beginning. However, there is a clear turning point in the development of the content where there is shared trust and understanding between the business units and taxonomists. In our experience, the turning point was characterized by increased understanding of reasoning behind business process and context by terminologists as well increased comfort from business units in terminologist demonstrated by less detailed review and scrutiny.

2.3.4 Business Approval of Terminologies

Because of the constant engagement with a wide array of business clients throughout the terminology development process, the approval process for the terminologies was relatively iterative. The one of the keys to success in any terminology development project is the approval and acceptance of the terminology with those who will use it. This involves not only the use of the terminology, but the correct and appropriate use of the terminology.

Representatives from business units were required to make decisions and communicate terminology decisions back to their represented group. We also made an intentional effort in constant communication of progress and decisions to ensure approval and acceptance.

2.4 Cleaning Terminologies

Another type of terminology work included the cleaning of existing terminologies based on business use. As with any development work, actual business use may change slightly over time creating artifacts in the terminology.

One of the tenets of Cimino’s terminology management describes terminologies to resist using Rejected “Not Elsewhere Classified” terms, otherwise known as “Other”[5]. Since these types of terms are inherently defined by what is not represented in other existing terms, there is no sense of what is contained within the “Other” categorization and over time as other concepts are added, the definition of “Other” changes.

However, what we found in our development of terminologies was that “Other” terminology was required because terminologies were used for coding; this seemed
to be the case in other coding terminologies, e.g. ICD-9 [6]. Since terminologies are developed upon predetermined needs and understanding, and there cannot be the expectation that developers would conceive of every possible situation, there needs to be an option to capture situations that do not fit predetermined coding. This reduces misclassification and increases accuracy. However, there is a balance in this situation, providing an “Other” option enables coders to become lazy or default to “Other” when convenient. Therefore, this also speaks to the need of strong training and definition around surrounding terms in order to reduce the use of “Other” unless absolutely necessary.

The solution then around reconciling “Other” terms was to develop strict business processes in regular review and assessment of records categorized as “Other”. In one case of reconciling “Other” records that had been neglected for three years, we found that 30% of the records were categorized as “Other”. To the credit of the coders, our analysis discovered that each “Other” terminology contained two to three major terms that either needed to be merged with existing terms or added to the terminology. This regular review requires detailed scrutiny of records to understand the reasoning behind the categorization; it requires discipline and commitment to maintain clean data.

3. Terminology Maintenance

3.1 Adoption Process

Successful adoption of terminology solutions is closely tied to business approval and sponsorship. We found that delivering terminologies successful is not simply the loading of terminology into coding systems but rather also the development of business process, effective terminology change management practices, training in coding terminology, and training in maintenance of that terminology.

In terms of facilitating training, given the changes in the terminology, it was essential to carefully document the definition and usage of each term. Even only in text or narrative form, this asset provided the basis for training users of the new terminology. In some cases, terminologies were only changed by 2-5% requiring a simple notification of changes, while other terminologies were changed by 50-70% requiring much more robust training. Although clear definitions provide greater value, there was still opposition from the business in exposing that on the front end of coding mechanism, fearing the requirement to maintain definitions would be too great.

One of the lessons learned from the adoption process was the necessity to implement into “production” or usage with urgency. Once terminologies have been developed and locked, the terminology immediately begins to become stale. The longer the terminology sits without usage and without proper maintenance by the business, the greater need there is to reconcile the terminology back to current state. We encountered this issue in one business area where adoption of the terminology management tool lagged behind the completion of the terminology development work. Because of the length of time, the business was not actively using and managing the terminology, subsequently more time and resource it would be required to reconcile and approve the new terminology.

In the previous example, one aspect that exacerbated the situation was the rigor required due to compliance regulation. Certain business functions in a pharmaceutical company fall under strict compliance regulation requiring increased time and resources to approve and implement changes. It is important to account for these through the adoption process in order to limit the amount of rework required.

3.2 Terminology Maintenance Considerations

In the same method of representing value to the business around the intensive development process, presenting the value of maintaining the investment into the terminology as an asset was required.

Cimino states the requirement for terminologies to evolve gracefully, by clearly documenting changes and reasoning for changes[5]. Because of the necessity for terminologies to grow and alter over time, good reasons for change need to be documented and understood while bad reasons for change need to be avoided.

Historically, business units maintained little documentation on changes made to terminologies, therefore it was essential to communicate clearly the need to protect the investment of development. In addition, providing mechanisms and clear processes to promote and encourage this behavior eased the impact of this requirement.

3.3 Business Ownership of Terminology Maintenance

Since terminologies were focused on coding and very business centric, it was important to encourage business ownership of the terminologies. Though this was not an issue of contention through the development process, because of our recognized role in terminology expertise, we found the need to continually communicate that stance.

Although the main function of the terminologies was the coding of interactions, there was also a secondary usage of the terminologies around cross-functional signal intelligence[1]. Therefore, while ownership continued to be a business responsibility, additional stakeholders were now required to interface with the terminologies.
3.4 Technical Solutions for Shared Maintenance

One of the challenges in implementing a terminology solution is enabling the management of that terminology. Specifically, providing a solution to better manage change in many of the projects, clients were moving from management of terminologies in Excel spreadsheets. With terminologies containing thousands of terms, change management in Excel provided great opportunity for errors to be introduced by reviewers due to deleted rows, filtered results, etc. Though there existed general process change management, such as roles and approval process, it was difficult to implement in practice using email and Excel spreadsheet versions.

An additional value-add in delivering the terminology solution included a much more robust change management model. This included specifically defined roles for terminologies, branches of terminologies, and even down to individual terms. The duty of each role in changes in the terminology was dictated by the governance contract, whether users were required to approve, able to veto, or simply notified of changes.

The terminology management tool that was implemented was SchemaLogic because of the robust terminology modeling and change management capabilities. The terminology modeling capabilities enabled relatively complex terminology relationships and business process to be captured in the system. Subsequently, paired with the robust governance tools, these elements can be placed under change control.

Other options were considered that offered more user friendly interfaces and search capabilities more appropriate for terminology development, however the criteria for select centered around the modeling and change management capabilities. Ultimately, the terminology management tool must match the business needs around terminology work.

4. Terminology Alignment Solutions

4.1 Terminology Alignment Challenges

Although terminologies we worked with across different functions were fairly independent with little overlap in their expected use, there were some shared entities in each terminology. One of the drivers for the terminology work centered around the centralization of the systems used by these functions, invariably the question around alignment of entities arose.

As described above, terminologies were not strict hierarchical terminologies with broader-than relationships, but rather hierarchical terminologies utilizing a business context relationship system, including: product name, packaging type, indication for usage. Because each terminology was developed separately and for a specific business use, it contains many artifacts in these entities. In addition, there is clear business process engrained within those artifacts that are legitimate.

According to terminology best practices, it would be advisable to reconcile all of these values; in experience, we did receive much pressure also from Information System colleagues to “simply” align the values. However, it would be short sighted to recommend that choice without fully considering the implications of that process. Historically, business functions worked independently from one another resulting in disparate systems, resources, content, and business processes. Therefore, there existed a host of business processes and systems engraved within even terminologies that had been developed for these business functions.

Within a given scope, timeline and budget for terminology projects, it was sometimes unfeasible to strive for total alignment. In addition, we lacked the influence required to enact such broad sweeping changes, especially with functions that were highly regulated and governed by compliance protocols.

4.2 Atomic Reference Layer

The need to have separate terminologies with some overlapping terms and concepts required that we develop a separate terminology that would interface and provide connection between with all appropriate functional terminologies. The concept is similar to the Unified Medical Language System (UMLS) in its attempt to unify medical terminologies and facilitate information sharing across systems[8]. This allows for functional terminologies to remain functional, addressing specific business needs, while also providing a means for alignment and analysis.

One of the challenges with the UMLS system was the massive scope and breadth of domain coverage which resulted in an extremely cumbersome terminology to understand and manage. This was realized when creating the atomic reference layer for the first time; taxonomists were given little direction in terms of what relationships and linkages to build into the interface terminology. This resulted in a rather larger terminology that became unwieldy to manage. Though the initial concept was beneficial, providing relationship between terminologies that were inherently separate, the usefulness was greatly diminished.

The second iteration of the atomic reference layer became much more focused and narrowly scoped. We carefully understood the primary purpose of the atomic reference layer, in that it was a mechanism to enable cross-functional signal intelligence for facilitating patient and drug safety. This provided much clearer direction in which to create relationships between terms.
In addition to capturing semantically similar terms, robust relationship modeling allowed for rich relationships to be created between terms. This allowed capture of tacit knowledge that previously resided only in the experiences and personal wisdom of functional staff members. For example, in order to understand analyze safety signals from different functional areas, staff members from each separate area gather once a month to review trending data and analysis, relying heavily on staff experience and expertise. This is methodology does not provide a scalable or sustainable means to understand signals from separate functional areas.

The atomic reference layer provides a means to capture relationships that are also logical, causal, and experiential, enabling that knowledge to be dissimilated and analyzed more effectively.

5. Discussion

In many examples in the literature, terminologies were designed independent of a specific information solution [9]. However, in practice, developing terminologies in a corporate environment is never separate from system requirements and constraints. Therefore, there is a need to balance business requirements and semantic correctness, even if there are some undesirable implications, e.g. increased maintenance, some lack of semantic uniqueness, reduced alignment of shared entities.

In addition, some of the solutions proposed and implemented in publically available or academically developed terminologies result in relationships that are semantically complex. While these complex models provide the most optimal solution, it is infeasible to expect business colleagues with no terminology experience to accept and maintain. Because in our implementations business users are required to interact so heavily with the entire breadth of the terminology, many more complex concepts in terminology development, such as complex relationship modeling and terminology abstraction, were not practical in terms of usability. In addition, with the expectation that the business would manage terminologies, more abstract and intricate terminologies models are not feasible. Note even taxonomist struggle with the management of complex terminologies such as SNOMED and ICD-9[10].

The other aspect that perhaps exaggerates the issue is the need for accountability, traceability and completeness in regulated processes for functions that are more downstream in the drug development process. Therefore some automated methods that provide solutions that “eliminate a large percentage of potential errors” but can never be sure that automated terminology maintenance is completely correct [2], would be difficult to validate in a heavily compliance driven function.

When developing terminologies, knowledge of entire business process and use of the terminology is important; downstream implications and uses of the terms can affect the choices that a coder will make in classification. For example, in one application, coding of an interaction directly drives the return of appropriate documents that the call agent will use to respond. The behavior that we noticed was that call agents knew the response document that was required, and would code accordingly, regardless if that code accurately represented the interaction. This was a case where the usage of the codes directly affected the accuracy and interpretation the terms. Similarly, researchers found that since insurance payor reimbursement was driven by the International Classifications of Deaths (ICD) coding on patient records, coders would classify patient records as to maximize reimbursement while payors would reclassify to minimize payment [6].

While some of these experiences would not affect the development of the terminology, it may affect the training, documentation and business process around the usage of the terminology that is so closely tied to a completed solution. In another experience, the terminology development team was found to be relatively sheltered in throughout the development process, focusing only on those that would use the terminology for coding, and not on those that would consume those coded records. We found that the larger the influence and affect of the terminologies, the greater need for understanding and scrutiny of the entire life cycle of a terminology’s effect. Therefore, it is important to be cognizant of these issues in development to enable the most effective terminology.

6. Conclusion

As with any development or implementation work in a corporate environment, one must critically understand the climate of the company as well as what it values. Terminology development is a tool in facilitating strong cross functional communication, good data quality, and effective data capture. Therefore, there is a balance between: the best practices and tenets of terminology development and management, and the constraints from the business environment and stakeholders.

Well developed and maintained terminologies can be an extremely valuable asset to a business. However, it is critical to understand where the terminology fits within the business process and where it can add value. Though many terminology best practices will be useful, the nature of the business environment will dictate the effectiveness of those practices, with the ultimate goal of adding business value.

Specifically for pharmacovigilance and cross-functional signal intelligence, clean and accurate capture of interactions using terminologies is essential. The
characterization of the interaction using terminologies is one of the most easily accessible and profitable sources of information in signal intelligence analysis. Therefore, the success of pharmacovigilance and cross-functional signal intelligence is highly dependent, in part, to a significant, upfront investment in terminology development and ongoing commitment to terminology maintenance.

References


