

Intelligent Foundations for Knowledge Based Systems

Mukundan Kandadai Agaram*

Delta Dental of Michigan, 4100 Okemos Road, Okemos, MI 48864, U.S.A.

ARTICLE INFO

Article history:

Received: 18 April, 2019

Accepted: 23 June, 2019

Online: 15 July, 2019

Keywords:

Knowledge Based Systems

Rule Based Information Systems

Intelligent Discovery

Data Visualization

Enterprise Decision Management

Master Data Management

Software Architecture

Enterprise Architecture

Domain Driven Design

Healthcare Adjudication

ABSTRACT

Knowledge Based Systems play a very important role, within Healthcare, with a primary goal of supporting a high quality service at an optimal cost. A widely accepted knowledge acquisition technique is through the use of Business Rules in a natural language like format. As clinical terminology is centric to HealthCare, a large percentage of these rules use industry standard codes to describe clinical concerns. Usually there is a large number of codes to denote the various clinical processes and procedures within Healthcare. These can translate to a combinatorially explosive number of Business Rules within the HealthCare IT system. It is common practice to leverage a Rule Engine to execute these rules and produce decisions within the system. A Rule Engine accepts a collection of rules called a ruleset. Within Healthcare Insurance, these rules and rulesets embody regulatory, policy and contractual concerns. To effectively manage this huge body of rules and rulesets, it is typical for Knowledge Based Systems to reuse rules across rulesets. Further, the knowledge required to author these rules and constructing comprehensive rulesets is specialized and requires deep expertise within the domain. Further, this also requires an expertise with authoring unambiguous business rules and operating the Rule Management Systems or Knowledge Based Systems. There is a need for the continuous Governance of these rule based knowledge assets for the organization. Typically knowledge experts in this field have a decade or more of experience. An emergent challenge or trend witnessed within the industry is that the experienced knowledge workers are retiring and their positions are being replaced by people not as experienced. This paper proposes techniques to build a layer of intelligent capabilities that can transform the methods for the automation, creation and maintenance of knowledge artifacts, that can aid and support the inexperienced knowledge workers with the effective management and administration of these assets within a Healthcare system.

1 Introduction

Healthcare domains are characterized by a rich knowledge accumulated through a combination of science, technology, experience, innovation within the industry and standardization through regulatory entities. Given the high cost of administration of health care, it is common practice to be reliant on a strong infrastructure of insurance to fund the high costs of administering healthcare. Further, there is a strong demand and pressure for absorbing scientific and technological advances within the field into operational practice within a short period of time. There is a high risk

of adverse practice effects, fueling the necessity for standardization of practice, within the industry. This paper is an extension of the paper titled *Intelligent discovery features for EDM and MDM Systems* originally presented at 2018 IEEE 22nd International Enterprise Distributed Object Computing Workshop (EDOCW) [1]. This paper further extends the concepts beyond the discovery features described in the original paper. It further formalizes these concepts into a set of foundational patterns, called *intelligent foundations*, with a view to providing automation towards hard to scale and hard to automate human labor intensive processes around business rule mining and knowledge manage-

*Mukundan Kandadai Agaram, Delta Dental of Michigan, 4100 Okemos Road, Okemos, MI 48864, U.S.A, Phone: +1 (517) 381-4736, Email: magaram@deltadentalmi.com

ment. This paper further lays down a formalized foundation of ingredients to construct recipes for building intelligent virtual knowledge working software agents to aid a new breed of digitally savvy, but functionally inexperienced workforce.

1.1 Healthcare in the United States

In the US, healthcare can be broadly categorized into the *Provider* i.e. the clinical entity viz., hospitals, doctor's offices, clinics etc. that provides the clinical services and the *Payor* i.e. the funding entity a.k.a insurance companies that pays for these services. The cost of healthcare is expensive and exorbitant to the general population. In addition it is highly susceptible to litigation, which further acutely drives up the cost. Healthcare insurance infrastructure is predominantly operated by the private sector, through a combination of "for profit" and "not for profit" entities. Stereotypically, healthcare insurance within the US is offered as a benefit, to employees of companies. The employers purchase health insurance for their employees through customized benefit plans, from healthcare payor companies. The health insurance costs are more affordable when bought in bulk for larger groups of people and is often referred to as the "group discount". This is due to the fact that the risk of "adverse selection" is significantly lowered. The cost of insurance is high for the average citizen seeking to buy it [2], [3], [4], [5]. Recent reforms in healthcare have attempted to introduce competitive pricing through the notion of healthcare exchanges at competitive rates. The general gist of this is to be able to provide insurance for healthcare to citizens, who otherwise cannot participate in employer health benefit plans either due to lack of employment or lack of availability of the benefit from the employer. In the US, there is a specialization of insurance products for *Medical* healthcare, *Dental* healthcare and *Vision* healthcare.

1.2 Background

The work presented in this paper and the original foundational paper is part of an evolving case study in the practical application of data science, analytics and visualization techniques that can automate the process of knowledge analysis and mining at a Dental Insurance company. Delta Dental of Michigan, Indiana and Ohio is the leading provider of dental benefits, serving the mid-western part of the US, with a subscriber base of nearly 15 million members throughout the states of Michigan, Indiana, Ohio and it's growing base of other affiliated states in other parts of the US. The company has witnessed a strong growth over the last decade of it's 60+ years of history. A key factor for this growth is it's highly automated insurance and benefits administration software platform, that achieves an industry enviable 95% auto adjudication rate (the percentage of claims that are entirely processed by the software platform with no human intervention). The cornerstone of the software platform is a sophisticated rule manage-

ment system called BRIDE, which enable the domain *knowledge workers* or Subject Matter Experts (SMEs), to author, test and manage the operational rules, using highly configurable *Domain Specific Languages* and a semantically rich *Business Vocabulary* [6]. The BRIDE platform is an innovative homegrown product which is the result of original collaborative work by the author and other fellow contributors, which is elaborated in a published conference paper [7]. The knowledge externalized into BRIDE as operational *Business Rules*, can be translated and deployed to any forward chaining production rule engine format [8].

1.3 Motivation

The process of deciding to pay a claim a.k.a *adjudication* in the healthcare claims arena, requires a deep understanding of healthcare clinical knowledge. Within the dental healthcare space, it is not uncommon to manually adjudicate complex and costly dental procedures, by dental experts who have had prior hands on experience with the clinical aspect of the science. Further a second dimension of expertise is to understand, the payor's business policy as it pertains to the interpretation of allowances and coverages. As a consequence, a typical knowledge worker has decades of experience within the industry and/or company. However, these experts are now reaching retirement age and will be transitioning their responsibilities to younger and lesser experience functional specialists. This provides strong motivation to build "*intelligent capabilities*" into the platform, that can automate laborious manual processes and to provide a decision support system for generating or recommending rules to the knowledge workers as they author these rules. This will be elaborated further in subsequent sections.

1.4 Organization

This paper is organized as follows. In 2, we will explore the characteristics of knowledge sources, their nature and consequences, within the context of healthcare claims processing and adjudication. We will cast a wider net to describe how the patterns presented by these problems can be applied to other functional verticals with similar characteristics. This will then provide a segue to 3, where the subject of *intelligent capabilities* will be explored with a viewpoint of automating manually intensive tasks and to provide seamless decision support to the act of *rule/knowledge mining*. In 4, we will explore the subject of building intelligent foundations and capabilities with the backdrop of the BRIDE [7] case study, that can support a growing and diverse eco-system of artificial intelligence. In 5 we will discuss inherent challenges in addressing the intelligent automation of knowledge/rule mining. This will naturally set the stage for discussion of future research/work, in 6. We will finally provide a meaningful conclusion for the work in 7.

2 Knowledge Characteristics

Knowledge/Rule Mining through traditional means can be a manually intensive process. Traditional Rule-based Information Systems, rely on the ability of human experts, to author high quality, non-ambiguous system enforceable rules. Further, knowledge rich areas such as clinical domains, yield a high number of such rules. These rules exhibit symptoms of combinatorial explosion, when certain influencing parameters change. These parameters are commonly influenced by external forces, such as customer needs, advances in science, market trends, regulatory mandates, adherence to standards etc. Second, the complexity of the rules can be high. In certain cases automation of the business logic, through procedural programming techniques can yield, very large (in the order of 10K lines of code), extremely complex programs [9]. The maintenance cost of such programs increase dramatically over time. In the face of such complexity, a system of externalized rules provides significant cost savings and improved agility to change. However, the system of rules that are mined out from large legacy procedural language programs is iterative in nature and is best achieved in a phased manner.

The original paper [1], outlines the complexities of the problem, especially around the mining, discovery and management of business domain knowledge as externalized business rules, into six different categories viz., *Volume and Growth Complexity, Organizational and Reuse Complexity, Rule Logic Complexity, Domain and Subject Matter Complexity, Healthcare Complexity, Healthcare Consolidation Complexity*. In this section we will contrast these with other intrinsic characteristics surrounding *Knowledge* and their relevance within surrounding eco-systems.

Healthcare systems operate around a hub of knowledge. This hub is surrounded by a triad of pillars, *People, Process* and *Technology*, which is depicted in 1. The *People* realm represents the Clinical knowledge workers, Business User Community, Claims/Insurance Practitioners etc., and their collective knowhow. The *Process* realm represents the organization's conscious act of documenting the knowhow with a view to standardizing, documenting and repeating them in a predictable fashion. The *Technology* realm reflects the collective collage of automation software, which in the case of the Healthcare Payor business, could be the Claims Processing System, Benefits Administration System, Accounting System etc. The company's knowledge assets live in the hub of these three pillars. We will now proceed to examine the characteristics of knowledge which intersects with each of these pillars. We will further examine the influences of introducing *intelligent foundations* to *Knowledge Management* and how it helps to preserve the most valuable assets of the organization i.e *knowledge*.

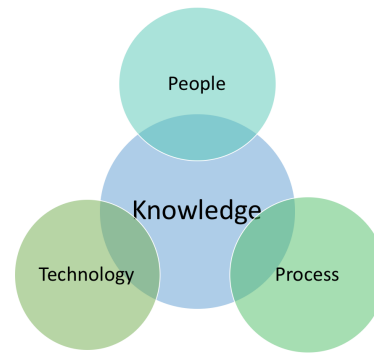


Figure 1: Knowledge Hub

2.1 Documented Knowledge

Documented Knowledge as the name suggests, represents all documented artifacts describing business processes, policies, contracts etc. Documented knowledge, exists in a variety of formats for different audiences. Healthcare is a highly regulated domain and is subject to a variety of audits. Non-compliance with regulation and audit findings present dire consequences, in the shape of costly penalties and potential litigation. Hence, it is crucial to maintain detailed documentation of processes, contracts and such. The basis of *Documented Knowledge* is depicted in 2.

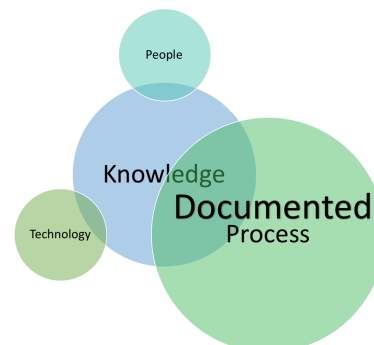


Figure 2: Documented Knowledge

Documented Knowledge, has the following consequences.

1. Footprint of *Documented Knowledge* can be huge and verbose. As it grows over time, the maintenance and upkeep of it becomes increasingly labor intensive and costly. As a result it is necessary to have quick and correlated discovery available to navigate the knowledge bases.
2. *Documented Knowledge* through traditional means is not *interactive* and *adaptable* to different audiences and contexts. Unlike *tribal* knowledge where knowledge workers can decide to provide a detailed explanation or high level overview depending on the appropriate audience and context, documented knowledge needs to be captured at varying levels of abstraction.
3. The knowledge can become outdated quickly, i.e. the knowledge bases if not updated periodically

can lose its accuracy, and hence its efficacy and value. To derive continual value from this source of knowledge, proper controls need to be put in place to effect the timely update of these knowledge resources.

4. Being textual and descriptive, semantics can get distorted. To mitigate this, the documented knowledge bases need to be reviewed and updated periodically by knowledge workers.

But if done right, this is an important source of knowledge and as such an asset to the enterprise. We will further explore how building a foundation of intelligent capabilities within the Knowledge Based Systems will capture the documented knowledge at the source of knowledge, through explicit means i.e. where the users can document aspects of pertinent contextual knowledge and implicit means where the underlying system learns, predicts and creates knowledge content.

2.2 Tribal Knowledge

Tribal knowledge is any undocumented information that is not commonly known by others within a company. It can also be thought of as the collective wisdom of the organization. It is the sum of all the knowledge and capabilities of all the people [10]. In a highly specialized functional domain such as healthcare, detailed nuances of claims adjudication procedures, interpretation of high level organizational policy or adaptation of external regulatory stances or standards can reside in this domain. Typical knowledge workers, with decades of experience in the field and more importantly experience administering the knowledge, within the domain within the same company are a very rich source of such knowledge. The origins of *Tribal Knowledge* in the *People* domain is depicted in 3.

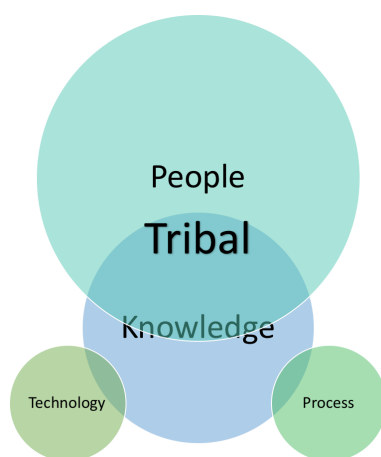


Figure 3: Tribal Knowledge

This form of knowledge has the following consequences.

1. There is a high risk of erosion, as the people holding them leave the organization through attrition or retirement. This presents a strong motivation

to build intelligent systemic capabilities that can tap into this rich source of tacit knowledge and be able to capture and leverage where possible, within the organization.

2. Traditional methods for capturing or documenting the knowledge i.e. transforming it into *Documented Knowledge*, in entirety is challenging at the least.
3. Tribal Knowledge can be clouded by subjective influences.
4. Some aspects of tribal knowledge pertains to the expertise or experience of the knowledge workers and their innate ability to leverage different sources of knowledge. This can be very contextual and situational by nature and unless well integrated into the knowledge management platform can become very hard to assimilate and correlate contextually. Hence, there is a very strong need to not only capture knowledge, but to artfully disseminate, correlate and index it for easy and fuzzy discovery.

The industry at large has been undergoing a significant generational workforce transition [11], which puts a heavy burden on the organization to effectively transition the tacit tribal knowledge held by the seasoned “baby boomer” generation, to a younger inexperienced generation. Special attention needs to be paid towards building a foundation of *Artificial Intelligence*, that can aid in the seamless capture from the older generation while effecting a smooth transition with interactive decision support capabilities back to the newer generation joining the workforce. On the other side of the coin, there is a strong need to attract and retain digitally-savvy, millennial and post-millennial generations of workers [13], while training them to be future knowledge workers.

2.3 Knowledge Systemization

We will now examine the concept of *systemization* of knowledge. This is the act of capturing the knowledge into a software system with a view of automating it. This can be achieved with one of two approaches.

1. In this approach, the knowledge is programmed and automated into software applications that capture the essence of the knowledge. We will call this the *Black Box* approach, for the obvious reason that the knowledge is now buried deep within the implementation of a software system, as depicted in 4.

This approach while being non-transparent to the non-programming business user community, at least does captures the knowledge within the bounds of a software. In theory this knowledge could be mined out (albeit painfully) and documented. However, there are some drawbacks.

- (a) The organization might not own the software system, which implements the business logic. In this scenario the organization has limited control on the mining effort at a later stage and has to entirely depend on the owning vendor of the software to mine out their business logic as metadata.
- (b) The software system is programmed in a legacy technology, with limited technical resources who understand the technology. This presents the same consequences challenges of erosion as posed by *Tribal Knowledge*, described earlier in 2.2

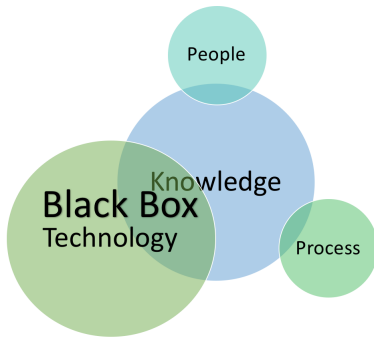


Figure 4: Blackbox Knowledge

2. This approach can be described as the act of formally capturing the knowledge in a *Knowledge Management System*. The most tangible forms of knowledge are *business rules* and *business policies* as these tend to be precise statements describing aspects of the business. Within the health-care payor space, these are typically captured as operational rules that can be automated. Such knowledge is *externalized* into a Rule Management System a.k.a Decision Management System as operational rules intended for consumption by a software system for automation, but more importantly externalized to be human readable and comprehensible by non-programming savvy users. We will call this the *White Box* approach, as this knowledge is *externalized* and is now transparently available to be understood and manipulated by the organization. The knowledge is externalized and managed under *Knowledge Management*, as shown in 5.

Externalization of the knowledge, presents the following benefits.

- (a) Externalization as described above, makes the organization more *agile*, as they now have the ability to change the operational business rules without making systemic changes which require a software release cycle. Further these changes can be directly controlled by the end business users, rather than the Information Technology community within the organization.

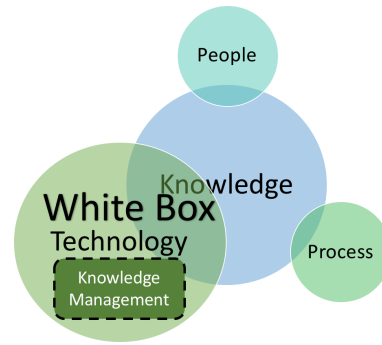


Figure 5: Whitebox Knowledge

3. Externalization of knowledge makes the organization's operation more transparent, but more importantly explainable as opposed to the *Black Box Approach*.

However, these are some consequences of knowledge externalization that need to be considered.

- (a) The number of externalized operational rules and policies could be large. Further, these rules are combined into large rulesets can present a combinatorially explosive [7], difficult to manage scenario over time. It takes highly experienced rule authors and knowledge workers well versed in the science and the organization's business operations, to effectively manage this, over time. This exposes the organization to the risk of knowledge erosion.
- (b) While the externalized knowledge is more transparent and explainable, to be able to comprehend far reaching cause and effect, the underlying system needs rich traceability capabilities [12], to provide detailed comprehensive explanations to the business users. This provides the deep motivation for *Intelligent Foundations* and providing an intelligent way to manage knowledge and sets up the business case for a *Virtual Knowledge Worker*, that is capable of answering deep *knowledge* questions, something that can aid and support the new younger generation of *digitally savvy* knowledge workers.

2.4 Intelligent Foundations

As discussed in 2.2, transition of the labor force [11], and catering to a diverse workforce with different knowledge assimilation requirements [13], presents a big challenge. This bolsters the case to build sophisticated and *intelligent capabilities* to address these needs. The intelligent capabilities will be elaborated further in 3. In this section, we will introduce the concept of *intelligent foundations*. Building such a foundation can then enable *intelligent capabilities*, that can rapidly train a new generation of young workforce, on the deep knowledge expertise. This is depicted in 6. The idea is as *Tribal Knowledge* erodes in an organization, the

new business users can learn rapidly, thereby restoring the tribal knowledge held by the people who leave the organization.

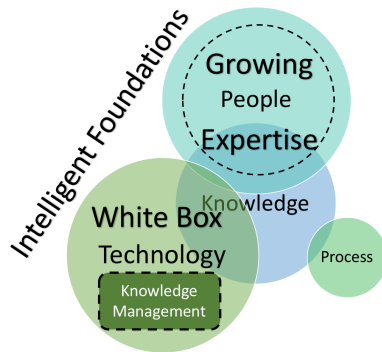


Figure 6: Intelligent Foundations

These *Intelligent Foundations*, can also be articulated as *capability* statements. We define a capability as a simple statement that describes an *ability* or feature of the system. Here are the ten foundational capabilities, at a high level, that enable building more specific intelligent capabilities around knowledge analysis, management and intelligent automation. These will be discussed in more depth in 4, against the backdrop of a case study, within BRIDE [7].

1. *Discovery*: Ability to fuzzily discover diverse artifacts
2. *Tagging*: Ability to *tag* or decorate knowledge artifacts with manual and generated metadata.
3. *Visualization*: Ability to graphically visualize *affinity* between artifacts
4. *Classification*: Ability to classify artifacts based on characteristics
5. *Correlation*: Ability to build correlation between different artifacts
6. *Event Processing*: Ability to generate, process and react to system and user events, within the platform.
7. *Recommending*: Ability to recommend action to take based on correlation between artifacts
8. *Language Processing*: Ability to absorb, disseminate and score keywords from textual content/documentation at detailed inflection points
9. *Generation*: Ability to generate new content
10. *Learning*: Ability to learn from data patterns and from human feedback

2.5 Knowledge Governance

Healthcare as discussed earlier is a highly regulated, knowledge rich domain. There is a strong need to formalize, document and more importantly *govern* the

knowledge, which refers to choosing structures and mechanisms that can influence the processes of sharing and creating knowledge [14], surrounding the detailed aspects of administration, business rules, business policy and processing. Within healthcare the business rules around administration are driven by different motivations, varying from regulatory, industry standardization, organization policy to client contractual customizations. Understanding and cataloging the various business rules and associated knowledge artifacts, contributes to better organizational management of the same. This consequently enhances the documented process knowledge surrounding them. By building intelligent capabilities around integrating and automating aspects of *Knowledge Governance* into the Knowledge Management System, aids in better administration of the governance process. Further it centralizes and promotes the use of the externalized knowledge across all three pillars i.e. *People, Process and Technology*, in a seamless fashion. This is shown in 7.

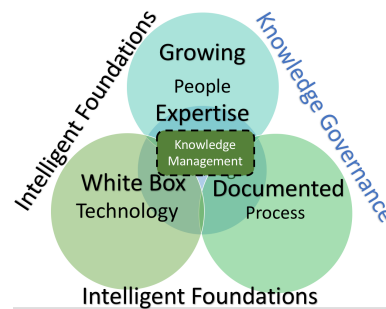


Figure 7: Intelligent Foundations with Governance

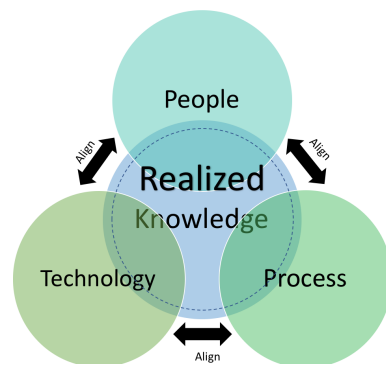


Figure 8: Realized Knowledge

A big goal is to better *align* knowledge sources across the three pillars of knowledge and unify them as a net sum of *Realized Knowledge*, as shown in 8. The introduction of *Intelligent Foundations*, coupled with *Knowledge Governance* strives to achieve that [15].

3 Intelligent Capabilities

In this section we will describe a subset of *Intelligent Capabilities* that are built on top of the set of ten foundational intelligent capabilities i.e. *intelligent foundations*

described in 2.4. These *Intelligent Capabilities* cover the vast gamut of functions needed to effectively manage a complex, vast, functionally diverse and growing footprint of externalized knowledge, as seen in healthcare. A logical viewpoint of this framework is shown in 9.

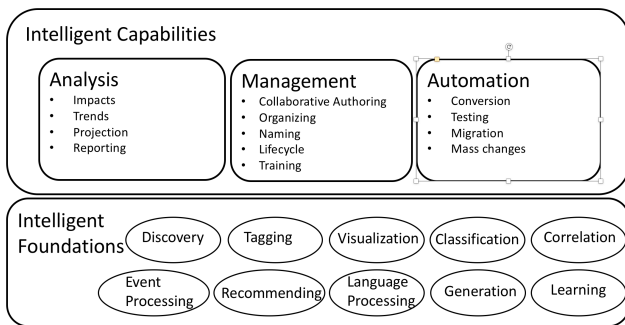


Figure 9: Intelligent Knowledge Management Framework

The capabilities are broken up into these broad categories viz., *Analysis*, *Management* and *Automation*. We will elaborate each of these broad categories of capabilities.

3.1 Analysis

Analysis covers the most important and far reaching capability across the organization. Practically every *business user* would need access to some or all aspects of this capability to perform their job effectively. As mentioned earlier, authoring operational *business rules* is a central part of knowledge externalization. With healthcare, some of the most complex rules are around *clinical terminology*.

To illustrate this further, we will contrast this with a Dental Insurance use case. With dental health insurance, clinical rules are specified using ADA (American Dental Association) mandated CDT (Current Dental Terminology) Procedure Codes. The knowledge workers are making changes to these rules as contractual benefits with customers change, or when the ADA releases a new version of the CDT, each year. Each Dental Procedure has a nomenclature that describes the procedure using key clinical terminology. When a certain aspect of dental clinical science changes, the ADA makes appropriate changes to the CDT to -

1. Introduce new Procedure Codes to reflect new clinical procedural advances
2. Retire obsolete Procedure Codes

In the United States, all healthcare business participants are mandated to be HIPAA (Health Insurance Portability and Accountability Act) compliant. In Dental Insurance, among other things, this means that Dental Providers (Dentists, Dental Offices) and Payors (Dental Insurance companies) are CDT compliant. This translates to the Dental Insurance adjusting thousands of dental clinical rules to reject or adjudicate claims based on the correct/appropriate dental procedure code being submitted.

These are some of the more fine-grained use cases within the *Analysis* capability.

1. *Impact Analysis*: Sometimes there are business rule changes across segments or the whole book of business, to specific areas of treatment, caused by a variety of reasons. These reasons, for example can be one of the following:
 - (a) Periodic ADA CDT version changes that are introduced on a yearly basis.
 - (b) Changes to a large client's contractual benefits on renewal, on a periodic basis.
 - (c) Introduction of a new line of business or product.
 - (d) Regulatory mandates.
 - (e) Corporate directional changes.

There is a strong need, for an ability to analyze and make micro-level operational rule changes based on macro-level shifts in function. To be able to do this the system needs the ability to correlate and analyze impacts across varying granulation of knowledge assets.

2. *Trend Analysis*: Trend Analysis can offer up useful insights to the organization and is intrinsic to the collective knowledge it holds. Here are some examples of trend analysis.
 - (a) *Reuse Trending*: As mentioned in 2, certain knowledge rich domains, can have combinatorially explosive rules under certain conditions. Therefore there is a need to provide appropriate tools for managing the analysis of trends to minimize the explosion of such rules through intelligent optimization and reuse.
 - (b) *Metric Trending*: The collection of accurate metrics leads to better knowledge governance and quality of rules/knowledge that yield better business decisions. For example, within the dental claims adjudication arena, we would like to analyze the most popular rules and be able to see seasonal or long term trends in their usage. Another useful analysis is to study the trending of the auto adjudication metric, which is the percentage of adjudicated claims that do not require any form of manual expert intervention, as this directly ties to the cost of administration.
 - (c) *Volatility Trending*: It is important to understand how frequently a knowledge artifact or rule changes. These can lead to deeper questions that would help analyze - are there seasonal influences to the change? Are these changes reflective of changes to the book of business? Can we use this as a future predictor of like phenomena?

- (d) *Variance Trending*: When looking at a collective body of knowledge artifacts for example, an entire set of rules, there is typically a variant component and an invariant component. This form of trending is again useful as a future predictor to study correlational causal effects with underlying books of business. For example the variance trending for Government Contracts can differ from Contracts, underwritten for small commercial clients, or clients within a certain vertical industry like car manufacturers or the demographic makeup of a certain group of underwritten people within a group.
- (e) *Complexity Trending*: Are knowledge artifacts (i.e. rules, rulesets, policy, key terms etc.) within a certain sub-domain of the business growing in complexity? This form of trend analysis helps identify and proactively address the management of such artifacts.
- (f) *Obsolescence Trending*: It is not uncommon to retire knowledge artifacts i.e. business rules or policies, over time and this form of trending can be an indicator in correlating the obsolescence of knowledge with evolving the book of business. In a certain sense this is similar to the volatility trending, discussed above, however obsolescence of knowledge can provide certain key insights and indicators around directional change for the organization, especially around obsolete business models or outdated processes, which are valuable.
3. *Projection Analysis*: Projection Analysis is the ability to analyze existing knowledge artifacts, to generate or “project” inductively, through inferred heuristics, gained from historical correlation of associated historical contextual changes to the corresponding changes of knowledge artifacts. This technique of analysis provides an approach for both the system and humans within the knowledge economy to project or generate knowledge artifacts based on the correlated data. This correlated data can further be used to generate features, to develop future state capabilities such as machine learning based on the generation and validation of projections based on correlation between characteristics of knowledge artifacts and high-level characteristics.
4. *Reporting Analysis Results*: With *Intelligent Capabilities*, built on AI foundations of fuzzy discovery, classification and correlation, business users can now be empowered to gather insights with deep drill down capabilities. It is important however to be able to package the results of ad-hoc analysis sessions into interactive reporting that can be exported and shared to a broader audience or community within the organization.

It is also important to be able to provide interactive *data visualization* features, packaged into this capability. By combining multiple *Intelligent Foundation* concepts, new users can be onboarded and trained on complex interrelated concepts in a more effective, timely and continual manner.

3.2 Management

Traditionally, management of knowledge assets within an organization requires deeply experienced multi-faceted knowledge workers, well versed in the domain terminology and the functional vertical. In addition, the knowledge workers tasked with authoring externalized business rules, need to possess traits that can translate complex functional subject matter into clearly articulated, unambiguous business rules and policies. Lastly, they need to be adept at navigating a complex knowledge management system, replete with a rich interconnected set of business vocabularies and complex features for other life cycle functions of knowledge management. The *Intelligent Capabilities* around management can be broken up into these following broad areas:

1. *Collaborative Knowledge Authoring*: This capability is at the heart of any Knowledge Management System, and to a large extent determines the success and efficacy of such a system. With traditional means i.e. with no *intelligent foundational* capabilities, a big part of the success and efficacy of the external knowledge hinges on the skills of the knowledge workers to deliver the quality. By building this capability on an intelligent foundation, new possibilities abound. For example here are a few plausible use cases, that leverage intelligent management capabilities. With a *collaborative* approach, the system collaborates seamlessly, with the users and facilitates collaboration amongst participating knowledge workers
 - (a) *The authoring tool provides a guided wizard like experience for complex rules*. For example, for new business rule authors who are not familiar with the intricacies of the detailed business vocabulary of a complex domain, the act of authoring a complex business rule that achieves a granular operational business decision can be a daunting task. In such scenarios, a high level, questionnaire driven wizard can step the business rule author through the act of authoring.
 - (b) *The authoring tool provides impromptu training sequences for guided authoring*. An extension of 1a, described above could show a step by step progression of an authored business rule to the new business rule author. This will help train that business rule author to learn the intricacies of the business domain vocabulary and it's associated key terms. Further, such sessions could be

recorded for future playback as interactive help topics.

- (c) *The authoring tool provides help based on correlation with keywords used in key business vocabulary.* This feature is another interactive training feature for the authoring tool to train new users. Further, while this use case is centric to the day to day management of knowledge, it can be used as a key analysis feature, that can help the business rule authors to do integrated impact and correlation analysis as they author business rules, policies and other knowledge artifacts.
- (d) *The authoring tool disseminates complex authored rules to catalog and correlate keywords to contextual help topics.* As new knowledge assets i.e. Business Rules, Policies etc., are created inside the Knowledge Management System, new business terms and language could be introduced. A consequence of this is that only small segments of the business community are aware of the semantics and impact of the newly introduced terminology. This bolsters the case for cataloging and correlating this information at the point of entry, into the management system. This could make the system smarter through the continual layering of correlational meta-data.
- (e) *The authoring tool provides a means for a community of knowledge workers, to collaboratively author related knowledge artifacts, concurrently by providing push notifications of changes and preventing conflicting or inconsistent changes to occur.* This capability leverages the intelligent foundations in 6, (the seminal book [16] provides a sound introduction to the science of events). Using this, the platform is capable of sending push notifications of user edit events, to different users, analyzing sequences of events, to make inferential decisions about dynamically constraining certain situational edit operations for certain knowledge assets, by certain participating knowledge workers.

2. *Knowledge Organization:* To manage the combinatorially explosive growth of knowledge assets, there is a compelling need for a sophisticated organization method that promotes the optimal reuse of knowledge assets. To be able to effectively reuse common knowledge assets like business rules, the first barrier to overcome is the ability to discover (rather fuzzily) that such a business rule or a *similar* business rule does exist. In the absence of this, a lot of duplicate or near duplicate knowledge assets are created. Organization of knowledge is key to enabling knowledge recall, management, mining and last but not the least Governance. We will now consider the

organization of knowledge from the following vantage points.

- (a) *Structure and Granulation:* Here we need to consider how the knowledge is structured and abstracted. Typically the knowledge artifacts can be structured in a hierarchical fashion, from most granular to least granular (i.e. high-level to low-level). Further each level of granulation can have several viewpoints. For example in dental insurance, the highest level of knowledge can be emblematically represented in a client/customer contract, which in turn, can be represented as an ensemble of *business policy* statements, which are high level statements of business purpose, that are driven through a combination of regulatory, contractual and administrative concerns.

These business policies are implemented through sets of operational *business rule* statements, that enforce the intent of the policy, within the Claims Administration platform. The breakdown of artifacts from a high-level perspective to a low-level perspective has to be well thought through and it is desirable to follow some form of a structured dissemination, based on some pre-established categorization. For example, the clinical adjudication policies are administered through a set of well established *treatment categories* within BRIDE [7]. This helps in managing combinatorial explosion of rule artifacts through a controlled reuse of pre-existing configurations of sets of rules. In the absence of such a categorized structure, there would be significant redundancy and duplication in structure as each hierarchical structure would end up duplicating, sub-structural elements of its hierarchy. However, in spite of this form of structured organization duplication and redundancy can be introduced due to the lack of *discovery*, as discussed in 2.4.

Lastly, to achieve these hierarchical structures of granulated knowledge artifacts with reuse, the underlying domain structures share a many to many correlation amongst each other, which are elaborated in more length in [7].

- (b) *Ownership* This is an important aspect of *Knowledge Governance*, that can influence the level of reusability discussed in 2a. With a growing consciousness and need for privacy, customers and clients are increasingly unwilling to share common data. Within, the consolidating healthcare payor space, this presents some interesting consequences.

Within the dental insurance space, the collective knowledge i.e. policies and rules

gained to preempt Fraud, Waste and Abuse, can be a shared asset, as these knowledge assets are gained through collective experiences. But to effectively leverage the shared operational knowledge, the historical claims *data*, across customer boundaries are needed to effectively adjudicate them. But privacy considerations and constraints can explicitly govern and prevent us from leveraging cross-customer data and knowledge, across boundaries, thereby limiting the potential to act on opportunities, that can minimize claims utilization.

Another consequence of ownership is that there are strict rules, that can be driven by a contractual agreements that dictate, who *owns* the knowledge and who has the rights to view, use or change them, under specific conditions. This poses interesting challenges to sharing and pooling of knowledge assets

- (c) *Correlation and Connectivity* To effectively organize knowledge, we need the ability to *correlate* and *connect*, pieces of knowledge assets. This introduces the important notion of *traceability* [12], which provides the users the ability to perform key activities such as root cause analysis, answer audit questions and explain processing and/or adjudication reasoning to customers and other stakeholders. Correlation can be achieved through the following three means or mechanisms.
- i. *Structural* This is the strongest and most basic form of correlation and connectivity that can be built to correlate different knowledge assets, explicitly correlated through well defined data and object models relationships. Within the BRIDE eco-system [7], different knowledge assets such as Rule Stores, Rule Packages, Rule Packets, Rules etc. are correlated in this fashion, which lends to easy organization.
 - ii. *Manual or Explicit* This form of correlation is weaker, but equally important. This allows the business users to explicitly tag or store metadata about various knowledge assets. These can be used to connect and organize knowledge assets based on the explicitly defined metadata.
 - iii. *Derivative or Implicit* This form of correlation builds on the above two forms of correlation. Here, we leverage smart algorithms to derive, infer, classify, tag, index, store, visualize and report additional correlations.

3. *Knowledge Artifact Naming*: As mentioned earlier, the detailed or granular knowledge assets

i.e. operational business rules and policies can grow combinatorially. The business users usually would like to introduce some mnemonic naming conventions to better organize and catalog the knowledge base. Hence, introducing capabilities to automate naming of newly introduced knowledge assets that adhere to user introduced naming conventions make sense. Further smart automated algorithms that manage the naming and cataloging of the knowledge assets especially when knowledge assets split, morph or converge, make the continued governance of such assets, over time, error and conflict free. Here are some of the explicit advantages or automation opportunities that can be realized by automated enforced mnemonic naming conventions.

- (a) Infer stereotypical characteristics of knowledge assets based on the mnemonic name
- (b) Projection and generation of *new* knowledge content based on mnemonic name of close affinity.
- (c) Proactive discovery of semantically like knowledge assets based on mnemonic name similarities

4. *Knowledge Lifecycle Management*: From a systemic standpoint knowledge assets need to honor constraints placed by systems. In addition, knowledge externalization and development processes need to synthesize knowledge asset development and lifecycle with systemic software development, especially for emergent areas of knowledge externalization and management. In doing so, at a very high level, knowledge assets goes through the following main lifecycle phases:

- (a) *Creation*: This is where the knowledge asset is generated.
- (b) *Change*: This is when the semantics or content of the asset morphs.
- (c) *Culmination*: This is where the knowledge asset loses relevance and is retired or terminated.

Within an environment, certain knowledge assets, such as business rules and rulesets have their own miniature lifecycle, before they are utilized by the automated system. For example BRIDE [7], has the following lifecycle

- (a) *Author*: The business rule is authored or edited within the BRIDE Guided Editor.
- (b) *Translate*: The authored business rule is translated to a format that can be interpreted by an *Rule Engine* [8]. This action is implicit for the standard use cases within BRIDE i.e. the Business User does not need to explicitly translate the rule, but it is tied into the Authoring phase of the lifecycle.

- (c) *Deploy to Sandbox* The overarching *Rule Package*, that contains the *Rule Packet*, that contains the *Rule*, is deployed to a safe sandbox, where the business users can test the rules.
- (d) *Deploy to System* The overarching *Rule Package* is deployed to the System, where the operational business processes can consume the decisions, made by these rules.

It is a common practice for knowledge based systems in organizations to have different environments. Broadly the environments can be categorized into:

- (a) *Production*: This is where the knowledge is applied to day to day operational business processes. New business rules and policies for established areas i.e. business decisioning services can still be developed here by the business users without any software or systemic development impacts. The BRIDE system supports this type of configuration, wherein the Business Rule authors enjoy such autonomy, wherein they can fully develop, test and deploy new operational knowledge assets *Rule Packages* [7], to the running production Claim Administration Platform.
- (b) *Non-Production*: This is where the organization develops, tests and validates the effects of newly introduced knowledge on the overall system. For emergent areas of knowledge externalization, the process entails the development of a standardized set of *Business Vocabulary* [6], that are mapped to an underlying implementation model, that integrates well with the other systemic components. Non-production environments can vary in number and purpose depending upon the organization's software and automation capability development processes. For example, here is a list of plausible non-production environments:
 - i. *Production Dry-run and Training Environment*: Intended for the Business Users to run live simulations, where they can do large scale application of knowledge without impacting live production business processes.
 - ii. *User Acceptance Testing Environment*: Intended for Business Users to run full scale end to end software release testing regiments, applying emergent areas of knowledge externalization and applying existing large knowledge bases with large scale impact changes.
 - iii. *Integration Testing Environment* Intended for component level integration testing of new and existing knowledge bases with systemic components.

- iv. *Development Environment* Intended for the development of business vocabulary and unit testing it's effects in a limited sense.

Finally, as knowledge assets take birth in different environments i.e. *Production* or *Non-Production*, there is a strong need to be able to migrate these assets from one environment to another in a safe and secure manner, without loss of integrity and fidelity. The BRIDE system [7], supports these capabilities in full, which enable the business knowledge workers to autonomously migrate and administer these knowledge assets.

- 5. *Knowledge Training*: It is crucial to have succession planning amongst knowledge workers, to avoid large scale dissipation of knowledge. With this capability the Knowledge Management System, uses intelligent foundations 1, 2, 4, 5, 8 and 10 described in 2.4 to absorb knowledge topics about the use and context of knowledge assets, to fuzzily discover, classify, correlate, disseminate and analyze language, learn and recommend actions to the knowledge worker in a proactive fashion.

3.3 Automation

As the externalized knowledge bases, grow, evolve and diversify within the organization, there is a strong need to *automate*, aspects of analysis, management and lifecycle management. We will now explore the intelligent capabilities needed to support this.

There are several tasks in the lifecycle that the knowledge workers have to do to externalize and create knowledge assets and related collateral. The biggest task is the process of authoring business rules and policies. Business Rules in particular are more challenging, for the following reasons:

1. The business logic captured by a single business rule could be complex and if not authored properly will not enforce the policies accurately or worse, introduce errors in claims processing and adjudication. Another consequence is they could introduce unforeseen loopholes that can be exploited by Fraud, Waste and Abuse scenarios.
2. There could be a high number of business rules due to combinatorially explosive sets of conditions. The knowledge workers have to expend significant analysis effort to analyze if there are gaps or conflicts in the logic and implementation. As a consequence, there could be a lot of duplicate, equivalent or subsumed business rules that get created over time, by knowledge workers, as there were no sophisticated discovery mechanisms present to discover them.
3. The business rules need to be carefully assembled into sets of rules, to form executable rulesets.

Many of the rules in a ruleset can have overlapping conditions and prioritization of rules within a ruleset matter. Between business rules in a ruleset there could be partial or complete overlaps of logic called subsumption. The knowledge workers need to make sure that these are minimized and restricted, only allowing some legitimate partial overlaps. An error in configuration of these rulesets causes errors in processing. The BRIDE system [7], provides a sophisticated system of configurations such as *Rule Packages* and *Rule Packets*. The Rule Packages are deployed as executable rulesets, that can be executed by a Rule Engine [8].

Here are some of the more fine-grained use cases within the *Automation* capability.

1. *Automating Conversion*: Consolidation and Digital Transformation are driving organizations to consolidate and transform their business operations. Within the healthcare payor space, healthcare insurance companies are seeking to move their operations to either:
 - (a) Homegrown healthcare administration platforms of healthcare insurance companies.
 - (b) Commercially available healthcare administration platforms.

The process of mining, converting and automating the knowledge i.e. Business Rules and Policies, of an insurance company on to a new benefits administration platform, is a risky, time-consuming and expensive process. There is a high risk of failure for the business and if not implemented accurately, there are hidden and ongoing costs introduced due to sub-optimal claims utilization, which can run the company out of business. Traditional methods for mining the knowledge from legacy systems [9] are manual, subjective and error-prone. The conversion process is dependent on multiple documented knowledge sources and is heavily reliant on knowledge workers who have had prior experience with the systemized knowledge, within the legacy platform. The challenges posed by traditional manual methods are:

- (a) It is dependent on knowledge workers intimately familiar with knowledge sources in the legacy platform.
- (b) It is ad-hoc and situational to the conversion project or customer.
- (c) It is time-consuming, typically a couple of years or more.
- (d) It is labor intensive, i.e. needs focused effort from a dedicated team of skilled knowledge workers.
- (e) It is not scalable to handle conversions of large magnitude or multiple parallel conversion efforts.

Partially automated approaches to *mining* information from legacy platforms entail, writing customized conversion programs/scripts that perform extraction and transformation of the knowledge to a mapping schema or file. The knowledge workers can then, manually analyze the data in the mapping data source and manually author the knowledge assets, into the target platform. Unfortunately, this again poses the same challenges except for challenge 1a, which is automated through the conversion program.

This is where more comprehensive and intelligent automation capabilities help solve the problem. As a case-study, using intelligent automation capabilities, the conversion effort was dramatically reduced from two or more years to under six months. The case study leveraged building automation capabilities that relied on an ensemble of *intelligent foundations* viz., 1, 2, 3, 8, 5 and 9, from 2.4.

Figure 10, depicts the steps to automate the mining and conversion of knowledge.

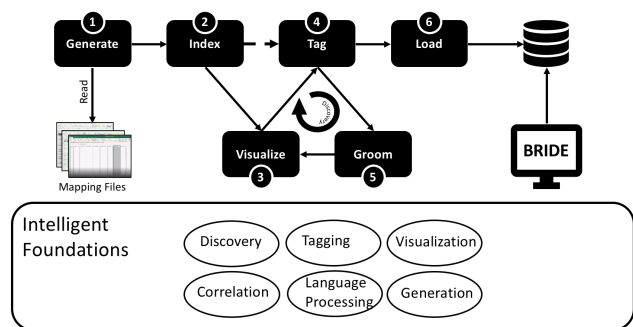


Figure 10: Automated Conversion Process

The conversion automation process is described through a sequence of six steps described below:

- (a) *Generate*: This is the first step, that analyzes the conversion data in the Mapping File and generates business rules, organizing them into rulesets, which for BRIDE [7] entail, organizing them into *Rule Packages* made up of reusable *Rule Packets*, containing the *generated* business rules. The generation algorithm follows the structural organizational rules mandated by BRIDE to stereotype the packets and rules with the appropriate *Treatment Category*. The generated artifacts are assigned a generated systemic naming convention in this step. The complex normalized data in the mapping files are put through a sequence of file transformations.
- (b) *Index*: The intermediate data that is generated by the above step is then loaded in an Elasticsearch database, which is a search engine based on the Lucene library, to enable

- them for easy and fuzzy discovery. This enables the intelligent foundation of *Discovery*, for the steps detailed below.
- (c) *Visualize*: Once loaded, custom data drill-down visualization programs aid the knowledge workers to analyze and proof the generated knowledge assets. In this step we use intelligent foundations of Visualization, Discovery and Correlation..
- (d) *Tag*: In this step, specialized algorithms analyze the generated business rules and elicit important parametric data to tag the knowledge asset. The two fundamental intelligent foundations used here are Tagging and Language Processing.
- (e) *Groom*: This is the phase of feedback for the algorithm to learn from knowledge workers, more specialized scenarios, to enable more accurate auto generation of knowledge assets. In addition, this is a process that is well integrated into the overall conversion capability to allow the knowledge workers to work collaboratively with the automated conversion process.
- (f) *Load*: This is the phase, where the generated rules are loaded into the BRIDE knowledge platform, into the *authored representation* [8]. Techniques such as rule parametrization (i.e. the variant information from rules are externalized as rule parameters) and mnemonic rule template selection are used to select the appropriate rule template and the parameters are substituted using macro substitution. Alternatively, more sophisticated intelligent builder approaches can be employed, wherein the authored representation for the rule can be generated based on structure of the generated rule, which can be gleaned by a command pattern generated during the *Generate* phase. Finally, the knowledge workers can resume normal knowledge management activity with the generated rules within BRIDE.

Figure 11, depicts the anatomy of an index for a Rule. This is created during the *Index* process, from the outputs generated from the *Generate* step. In addition to the generated rule, the additional metadata generated during the *Tag* phase sets the stage for the subsequent grooming and accurate and automated loading into BRIDE knowledge base.

Figure 12 depicts the anatomy of an index for a Rule Packet. The basic structural data for the Rule Packet is created during the the *Generate* phase. The intelligent conversion algorithm blends in manually authored rules by knowledge workers with the generated rules, to construct a fully functional subset of rules that constitute a

Rule Packet for a given Treatment Category. Further during the *Tag* phase, there is an abstraction phase of the algorithm that filters only relevant or important tags from rules that are emblematic of characterization for the Rule Packet. Lastly, the algorithm generates a descriptive name for the Rule Packet that facilitates easy discovery for knowledge workers for higher level activities such as contract implementations.

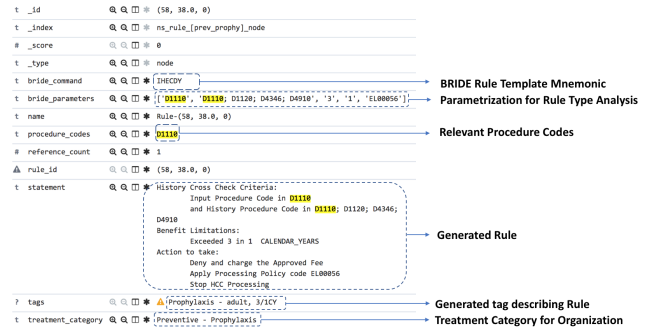


Figure 11: Anatomy of a Rule Index

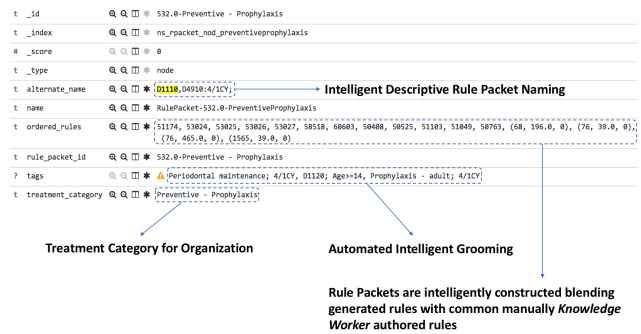


Figure 12: Anatomy of a Rule Packet Index

2. *Automating Testing* With rapid growth of knowledge assets within the platform, especially in the form of operational business rules, there is a compelling necessity for having an intelligent automation technique that can analyze rules within a ruleset and generate appropriate test scenarios. For example within the BRIDE [7] framework, there are over 17,000 test cases in just one area of adjudication rules. These are test cases that have been created by knowledge workers during their testing phase of knowledge management. A significant chunk of these test cases were generated from live claim processing data scenarios. However, with a strong and steady growth of operational business rules, more sophisticated techniques are needed to conquer the scale problem. Here are some use cases that explain this capability further:

- (a) An automated test scenario creation algorithm intelligently interprets semantic tags about the rule parameters collectively con-

tained within a ruleset to generate test data and scenarios.

- (b) As new rules are authored and included into a ruleset, an algorithm amends or augments test scenarios to test out new semantic consequences of the change.
- (c) As knowledge workers decide to edit rules, an algorithm that can analyze the impacts from a test scenario perspective interactively communicates consequences of the change to the knowledge worker. This is a capability that can grow in the degree of sophistication.

3. *Automating Migration*: As we discussed in 4, it is typical for knowledge workers to manage knowledge assets across multiple *production* and *non-production* environments. Within BRIDE, select knowledge workers can elevate their role to an administrator, to perform critical administrative functions of knowledge asset management. *Migration* of knowledge assets across environments is one such function, that needs to be done with care.

Here are a few challenges with administering this function, where the administrator is performing these tasks manually.

- (a) The number of physical environments can grow within an organization, over time. With consolidation the number of knowledge tenants in a multi-tenant environment, further impact the number of environments to manage.
- (b) The number of knowledge assets to migrate across environments grow. This puts an increased burden on the knowledge administrator to keep knowledge assets consistent across environments.
- (c) Not all environments, especially not all tenants may share the same assets i.e. business rules, policies, business vocabulary, test scenarios etc.. This introduces increased complexity for the migration path of these assets.
- (d) Maintenance in target environments can impede or impact the migration of knowledge artifacts. This problem is very similar to supply chain inventory and inventory movement, within Supply Chain Management.
- (e) Associated management functions such as deployment need to occur in target environments, post migration. This puts an extra burden on the administrator to ensure that these activities occur in a timely fashion.

Strong intelligent foundations can help with the automation of pieces of this important function. Here are a few of the scenarios, that can help

intelligently automate this function and reduce the dependency and burden on the knowledge worker.

- (a) Tag and inventory knowledge assets with migration specific metadata that is human and more importantly system consumable. This can enable intelligent algorithms to proactively audit, report and govern knowledge assets, migrating to target environments. Here intelligent foundation 2 is primarily leveraged.
- (b) Restrict the migration of knowledge artifacts to target environments that are undergoing maintenance or is being reserved for some focused activity. This prevents undesirable collisions and interference with activities in the target environment. Here the intelligent foundation 6 is leveraged.
- (c) Automatically include related artifacts impacted by the inclusion of a knowledge asset within a migration manifest. Knowledge Assets are intricately related, sometimes explicitly and structurally, but other times through correlation as explained in 2a. Hence movement of such knowledge assets should do the necessary impact analysis to ensure consistency and integrity of the knowledge assets. Here intelligent foundation 5 is primarily used. In addition, intelligent foundation 7 can be leveraged when the inclusion needs to be reported or recommended to the knowledge worker performing the administrative role.
- (d) The *Automated Migration Assistant* participates as a *virtual knowledge worker* in the target environment, as described in 1e. Here the intelligent foundation 6 is leveraged.

4. *Automating Mass Changes*: Sometimes it is necessary to make large scale or *mass* changes to knowledge assets. For instance, a large number of business rules need to be changed because a key business policy changed or the semantics of a key business term changed.

Through traditional methods, this would mean knowledge workers need to track down every single relevant knowledge asset that are impacted by the change and modify them to reflect the macro-change. A classic example of this are new yearly versions of the *Common Dental Terminology*, released by the *American Dental Association*.

Here by leveraging all the 10 intelligent foundations, intelligent capabilities that automate making mass changes can significantly reduce human labor, erroneous edits and significantly reduce cost.

4 BRIDE Case Study

4.1 Background

Delta Dental of Michigan, Indiana and Ohio is the leading provider of dental benefits within the midwestern part of United States. The company has been in business for over 60 years and has witnessed a dramatic growth within the last decade. The reason for this dramatic growth is a highly automated and sophisticated benefits administration software platform, powered by BRIDE [7], a powerful homegrown, elegant Enterprise Decision Management System. BRIDE houses the company's growing pool of complex, externalized operational business rules which allows the company to automate the vast majority of its claims adjudication process allowing it to scale its business footprint to unprecedented levels, while keeping operation costs at an optimal level. The current metric for measuring this automation level is called "Drop to pay" or "Auto Adjudication Rate", which is currently hovering at around 96%, which is an industry leading score.

To keep up with the fast growth, it was necessary for the BRIDE platform, at inception to introduce industry leading capabilities for surfacing full life cycle rule management functions, such as authoring, testing, deployment, migration, such that non-technical knowledge workers and subject matter experts could entirely manage these with little to no involvement from the IT Department. Over the last decade the BRIDE platform has kept up with innovation, by expanding the capabilities and staying a step ahead of the business needs.

The BRIDE ecosystem organizes its knowledge artifacts in a structured, highly reusable and componentized fashion. A logical view of the domain model for BRIDE at a high level is depicted in refbrideDomain-ModelFig.

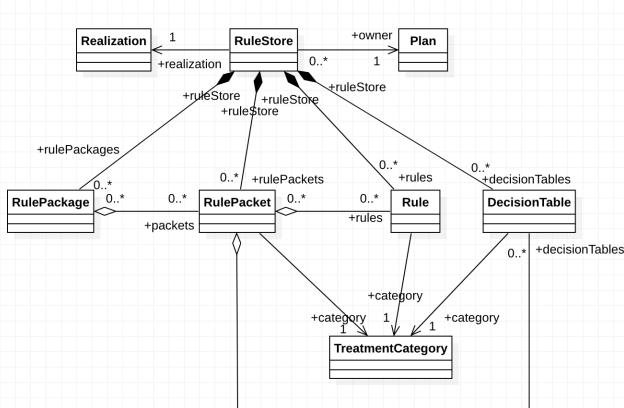


Figure 13: Highlevel BRIDE Domain Model

The core BRIDE constructs [7] are:

1. *Plan*: Represents a Dental Insurance Owner, that owns the rule artifacts.
2. *Rule Store*: A repository for storing rule artifacts, each rule store holds the reusable rule artifacts

that can be reused within it, in a structured and constrained manner.

3. *Realization*: Represents an area of business functionality that require specialized terminology to externalize and articulate operational business rules using a structured vocabulary. There are over 30 areas of business functionality for the dental insurance domain so far, ranging from several detailed aspects of claims administration to billing, rating and actuarial applications. This is a growing body of knowledge as opportunities for automation and externalization are identified.
4. *Rule Package*: Represents an *executable* set of rules, that can be executed in a forward chaining, fact inferential rules engine [8]
5. *Treatment Category*: A mechanism to categorize finely granular rule artifacts. Within the clinical domains, the treatment category construct allows the business to categorize specialized clinical rules to literal clinical treatment categories as recommended by clinical standard bodies such as American Dental Association. For non-clinical domains, the treatment category construct is repurposed to category sub functional constituent components of processing.
6. *Rule Packet*: Represents a subset of rules for a given treatment category. Each Rule Packet is stereotyped for a treatment category and should only contain Rules of that same treatment category.
7. *Rule*: Represents a reusable authored production rule (i.e. a rule containing conditions followed by actions) authored in a "Domain Specific Language" using an english-like vocabulary. It is also the most

The next wave of innovation within the BRIDE platform introduces the application of *Artificial Intelligence* techniques to effectively address a new, growing breed of current and anticipated challenges. The work done on developing some of these capabilities has gone towards shaping and developing the concepts discussed in this paper.

The approach to building this new wave of innovation, has been to implement a strong foundational architecture i.e. "*Intelligent Foundations*". We will discuss these foundational capabilities, their motivation, the architectural choices and application, subsequent sections.

4.2 Discovery

"Ability to fuzzily discover diverse artifacts" As the number of rule artifacts have grown over time, in spite of the ability to reuse rules a lot of unneeded redundancy and duplication have proliferated causing the knowledge base of operational rules and rulesets to

bloat. A major reason for this has been attributed to *discoverability*.

Within the original architecture, the BRIDE system provided SQL based searches against a relational database backdrop of the BRIDE schema. With this, the knowledge workers could search for specific rules with key terms and vocabularies and see their association with rule packets and rule packages. This was a bottom up search. They could also search for Rule Packages with specific Rule Packets in their configuration. With this they could search for full or partial rule package configurations, generate a report and then drill down. In spite of these rich search capabilities, discovery was not sufficient to prevent the proliferation of redundant rule assets overtime. Here are the main reasons why this alone is not sufficient:

1. *Layered Reuse Complexity*: BRIDE uses a layered reuse paradigm for constructing Rule Packages that eventually translate to executable rule-sets. While this approach in and off itself is a sound from a component reuse vantage point, the major complexity is the inability of the traditional search functions to get a comprehensive and 360° view of the Rule Package.
2. *Naming Complexity*: With the componentized approach promoted by BRIDE, the knowledge workers set about building reusable Rule Packets, which in essence were combinations of rules, within treatment categories. In the interest of providing meaningful names that described Rule Packets, the knowledge workers resorted to lengthy descriptive names that sought to summarize the semantics of emblematic rules contained within the Rule Packet. However, this approach is highly manual, lacks predictable standardization and is not sufficient to describe huge and complex Rule Packets.
3. *Variational Complexity*: The characteristics of specific clinical domains require blending of specific clinical terminology that by the very nature can lead to a combinatorial explosion of sorts. With the nuances and subtleties in books of business and the semantic authoring flavors introduced by different knowledge workers there were subtle variations of sequences of rules, individual rules that were introduced over time. Some of these were made deliberately to support the variational needs of the requirements. Yet, others were made unknowingly introducing inconsistencies in approach. These were challenging to discover.

ElasticSearch, is a search engine built on the Lucene library. Nowadays, it is a common trend with modern IT architectures to leverage this tool to enable and enhance fuzzy discoverability. Data can be loaded or *indexed* into ElasticSearch as JSON (JavaScript Object Notation) documents. Subsequently this data can be discovered through a textual search of terms or fields

contained in the JSON documents loaded and the tool will return all documents that match or nearly match textual patterns searched.

With complex natural language rules, this enables easy discoverability of these assets in a fuzzy fashion, that traditional relational queries do not allow. But this alone is not sufficient as the knowledge workers not only need to discover individual rules based on key business terms and vocabularies, but they also need to be able to discover the connections between rule artifacts, their levels of similarity and the characteristics of their similarities. With that there are five intelligent foundations that need to be applied in conjunction to fully realize the benefits of *discovery*. How these intelligent foundations collaborate is shown in 14. To understand these relationships:

1. *Language Processing* is used to disseminate the textual data in policies and rules, filtering key business terms and builds important metadata for *Correlation*.
2. *Classification* runs advanced algorithms, leveraging data created by *Correlation* to build more detailed correlation data.
3. *Tagging* is the default mechanism for storing metadata generated during *Correlation* and *Classification*
4. *Classification* and *Correlation* builds the foundation for *Discovery*.
5. *Visualization* helps makes sense out of *Discovery*, especially when there is a large amount of interdependent, connected data.

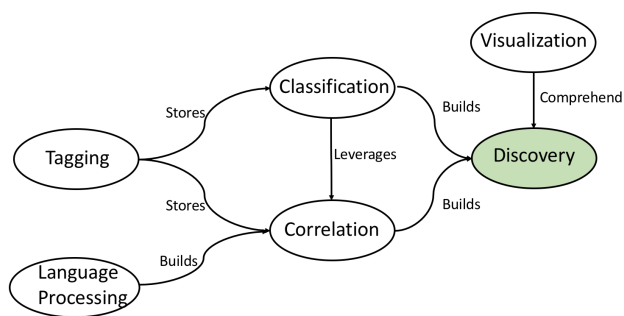


Figure 14: Collaborative Foundations for Discovery

4.3 Tagging

“Ability to tag or decorate knowledge artifacts with manual and generated metadata.”

The basic implementational notion of a tag is to assign a key and description tuple to an artifact of interest. However sometimes, there can be more than one tag to decorate the artifact with. Therefore, it makes sense to introduce these key concepts:

1. *Tag Aggregation (Tagregation)*: A Tag Aggregation is a composite structure of sub aggregations and Tags.

2. *Tag*: A Tag is simple key and description.
3. *Taggable Artifact (ITagArtifact)* A Taggable Artifact is any entity that needs tagging metadata.

The logical domain models for the basic tag constructs and taggable artifact constructs are shown in 15 and 16. The actual implementation of tags with the artifacts is more relevant when indexed into ElasticSearch for discoverability. A JSON representation of a Rule with a sample of extracted tags, used for discovery, is shown in 17. In this example, the tags for dental procedure codes used in the rule are, extracted to the tag “procedureCodes”. Similarly, other useful correlational metadata could be extracted.

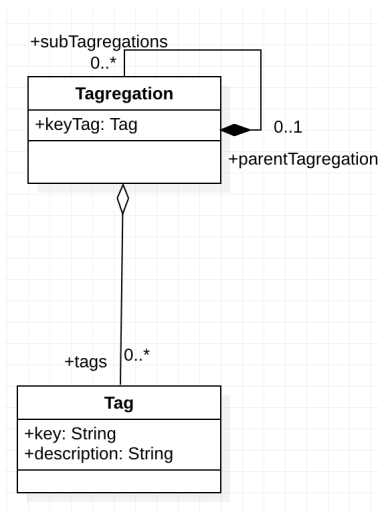


Figure 15: Tagging Domain

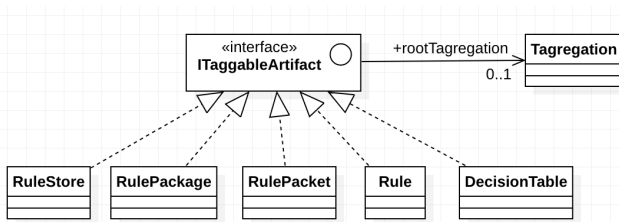


Figure 16: Taggable Artifact Domain

```

{
  "artifact_id":204,
  "artifact_type":"Rule",
  "organized_tags":{
    "procedureCodes":[
      "D6205-D6252",
      "D6545-D6634",
      "D6710-D6794",
      "D2410-D2810",
      "D6205-D6794",
      "D6999"
    ]
  },
  "flattened_tags": [
    "D6205-D6252",
    "D6545-D6634",
    "D6710-D6794",
    "D2410-D2810",
    "D6205-D6794",
    "D6999"
  ],
  "name":"HCA62815",
  "statement":"History Cross Check Criteria:\n\t Input Procedure Code in D6205-D6252; D6545-D6634; D6710-
  "packet_ids": [
}
    
```

Figure 17: JSON of a Tagged Artifact

4.4 Visualization

“Ability to graphically visualize affinity between artifacts”

To enhance the discoverability of the knowledge assets data visualization is an effective method of analysis. Here a few different visualization techniques have been leveraged, with the correlation and classification data collected through tagging. Knowledge assets as a graph visualization with color coded edges to depict the different types of classifications, is depicted in 18. This is an effective way to discover common trends and patterns. The original workshop paper [1], outlined six different affinity classification measures, which are shown in this visual with different colors. This kind of visual is very useful for identifying common clusters of affinitive rules within an area and the nature of their affinity. There has been significant research and work on the efficacy of applying topological data analysis techniques, to use cases within healthcare [17]

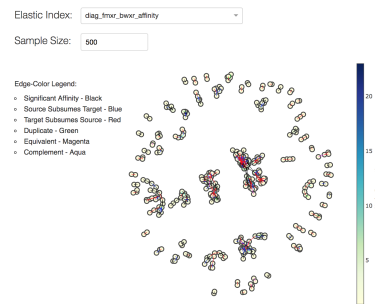


Figure 18: Color Coded Clustered Affinity Visualization

An edge configuration for the same visual is shown in 19. This shows the nodes on the circumference and different colored edges criss-cross connecting the different nodes which represent the rule artifacts. The advantage of this visual is that useful, metadata can be shown on the edges, on hover that aid immensely during knowledge analysis.

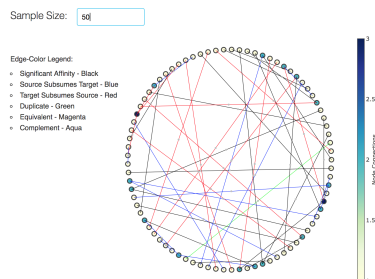


Figure 19: Color Coded Edge Affinity Visualization

A common themed rule asset, is picked out from a crowd of assets, in 20. This form of visualization is useful in identifying an emblematic asset for a certain area.

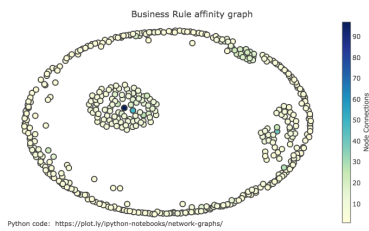


Figure 20: Clustering to identify baseline rule

A term cloud visualization is shown in 21. This is a very useful visualization for performing impact analysis based on key term correlation. In conjunction with correlation of key terms across rule and knowledge assets, it is possible to do end to end impact analysis.

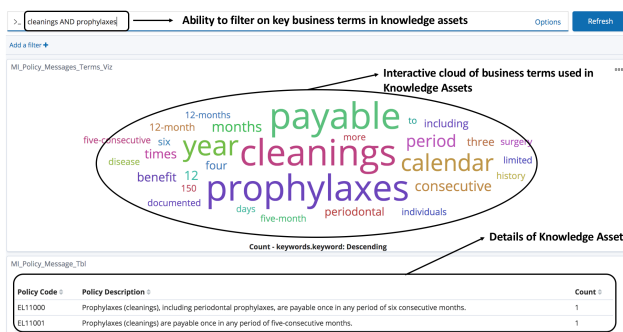


Figure 21: Interactive Term Cloud Visualization

A few observations when building data visualizations are:

1. It is useful to provide a way to filter based on category. This helps with narrowing sample sets and separation of concerns.
2. It is useful to provide a way to configure the size of a sample set. With certain types of analysis the knowledge workers are seeking understand common trends, to make informed decisions about management and governance.
3. It is useful to provide a mechanism for searching or filtering for combinations of terms. This aids in narrowing down the analysis. The visualization in 21, provides this mechanism.
4. It is useful to provide a detailed drill down capability on visualizations where possible. This allows the knowledge worker to identify the concrete knowledge assets with a single visual.

4.5 Classification

“Ability to classify artifacts based on characteristics”

Several sophisticated algorithms were developed to study the affinities between business rules and to classify them. Techniques to classify rule and knowledge assets ranged from naive simple bag-of-words to more sophisticated variations that combined bag-of-words with correlation to relevant master data assets.

In the original workshop paper six different classifications were developed, comparing business rules, which were:

1. *Duplicate*: When two rules were exact duplicates with different names. These were top candidates for elimination and consolidation.
2. *Equivalent*: When two rules were semantically the same but had conditions that were in a different order. The classification algorithm here was shallow and only checked for conditions or actions in a different order. Deeper semantic analysis of rules will be able to identify equivalent rules beyond these narrow confines, creating the opportunity for better cleanup of these redundant rules.
3. *Complement*: These were specific to this area of rules where clinical procedures were complemented across the input and history procedures, to provide a complete adjudication coverage. Often times, claims are submitted out of order and these kind of rules analyze logical completeness. The scope for these can be expanded to cover other clinical criteria such as teeth, tooth surfaces etc. in the future. Further this pattern can be applied to other functional verticals to analyze logical completeness.
4. *Subsumption*: Here there were captured as two classifications i.e. Source subsumes Target and Target subsumes Source. Here again deeper semantic analysis can analyze more subtle aspects of subsumption, like analyzing inclusive procedure patterns etc. In the absence of these deeper analysis capabilities, come subsumptions could miss this classification.
5. *Significant Affinity*: Here the algorithm used a threshold of score similarity to identify rules that were closely affinitive with each other. However this was pretty effective and did capture rules that were significantly affinitive with each other.

4.6 Correlation

“Ability to build correlation between different artifacts”

This has been a centric intelligent foundation, for connecting different concepts together, to perform associative analysis and impact analysis. The implementation of correlation relied on algorithms, that analyzed, indexed and decorated knowledge assets with metadata in the form of tags and/or first class correlative attributes, in specific cases. Within the BRIDE case study we implemented correlation for three forms of correlation and connectivity as explained in 2c. Here are a few architectural aspects to consider, while integrating this intelligent foundation within the Knowledge based system.

1. Correlative algorithms need to run close to point of knowledge. There is an inherent risk of running correlation algorithms in a batch fashion

as morphing data conditions could either miss certain temporal correlation or miss transient changes that can be critical information to capture.

2. The initial correlations should be captured at well designed intermediate sink points and should be verbose enough to not miss seemingly inconsequential details, that might add value at a later point. Intermediate logging could be an option for such an architectural implementation. Generating and storing changes as events is another option.
3. The correlation algorithms should be designed in an idempotent fashion where possible, especially with the dimension of temporality. There are use cases that might require the regeneration and storage of correlation, especially during software upgrades and patching.
4. Even though, for the most part correlation of knowledge assets do not include PHI (Personal Health Information) or PII (Personal Identification Information), there are some corner cases where indirect correlation of operational data that have such concerns might crop up. One example of this is the reverse engineering capability of BRIDE for test cases from live data. In such cases de-identification and regular purge of such correlation need to occur to address security concerns.

4.7 Event Processing

“Ability to generate, process and react to system and user events, within the platform”

This intelligent foundation forms the basis of a very important infrastructural capability. As stated earlier, it can be leveraged for Collaborative Authoring. But it can also be leveraged for Correlation and Connectivity. Within the BRIDE case study, this is currently on the roadmap but has not yet been implemented. However, a lot of work is projected on the roadmap to innovate this intelligent foundation, especially for capturing fine grained operational execution metrics to aid correlation, for machine learning and machine training purposes.

4.8 Recommending

“Ability to recommend action to take based on correlation between artifacts”

Within the BRIDE case study this was implemented as an explorative proof of concept, wherein the system could recommend close rules based on the current authored content. While this could be foreseen as a useful feature in and off itself, this is barely scratching the surface. The BRIDE case study, hopes to leverage this intelligent foundation for deeper more sophisticated scenarios, which can be the subject of another paper.

4.9 Language Processing

“Ability to absorb, disseminate and score keywords from textual content/documentation at detailed inflection points”

Within the BRIDE Case Study, this intelligent foundation has been used to disseminate policy descriptions, and procedure code descriptions, to correlate business rules and other artifacts such as Rule Packages, Rule Packets, Rules, Benefit Plans etc, with key term correlation. The Language Processing capabilities with Python and NLTK give a good starting point to filter out key business terminology, which can then be used to correlate. Again there is not more to report here from a case study perspective as we are in very early stages of fully realizing and leveraging this intelligent foundation.

4.10 Generation

“Ability to generate new content” Within the BRIDE Case Study, this intelligent foundation was used to automate conversion as discussed in 1. However, this is just one use case or application of this intelligent foundation. However, there are some other scenarios where Generation can help.

1. Perform mass changes to rules. As mentioned earlier the ADA, publishes a new version of the CDT every year. With the ability to generate new or change existing content, based on high level heuristics, it will be possible to automate the mass changes to Rules.
2. Generate new rule templates based on mnemonic naming patterns that are generated based on high level heuristics. Within the BRIDE Case Study, this was exercised in a limited purview, as new rules had to be generated based on the conversion data.
3. Automatically generate test cases based on rule content. Based on semantic analysis of knowledge and rule artifacts, intelligent generators can automate the act of building test cases on the fly and saving them for regression purposes.

4.11 Learning

“Ability to learn from data patterns and from human feedback”

This has not yet been implemented within BRIDE, but it is there in the roadmap. This is an area of future work and builds on the other intelligent foundations.

5 Automation challenges for Knowledge-intensive systems

In this section we will consider some of the challenges around automation, for knowledge intensive systems. In 3.3, automating the task of authoring was identified as a the biggest opportunity for automation. Within

the BRIDE case study, automating conversion was one of the biggest opportunities. Automating the process of knowledge acquisition, reduced the time effort from 2 years to a few months. However, there are some inherent challenges with automation of the rule mining and authoring process. They are:

1. *Standardization* : For an automated rule conversion process to work repeatedly, the data for that area of functionality needs to be standardized into a well thought out canonical format.
2. *Specificity*: It is not uncommon for the quality of knowledge mined out from legacy blackbox systems to contain a lot of noise and inconsistencies that are specific to the legacy system. Further the nature of the inconsistencies can vary across different target systems and conversion initiatives. The automation processes while being streamlined should allow for custom transformation rules that can be externalized and customized for each conversion effort.
3. *Semantics*: The interpretation of standard benefit clauses implemented in rules in the legacy source system could have different semantics and might need specialized handling or transformation into the target system's knowledge base.
4. *Inconsistencies*: Sometimes there are misinterpretations of documented high level policies and contracts in the source legacy system, that are being converted. Technically one point of view amongst knowledge workers is to rectify these during the conversion process. Yet, this can have dire behavioral regressions on the target system, post-conversion. These kind of situations, need to be identified during the automated conversions, traced back to the source and have a good explanation facility. Whether this method *can* be formalized consistently to bolster the efficacy of such automation efforts, remains a tantalizing proposition.
5. *Universal applicability*: For the BRIDE case study, the automated mining and generation process is fairly well established for a specific area of adjudication rules, that had the most number of rules and artifacts, that could not have been authored manually. But, whether this automated process can be streamlined and applied universally to other areas of rule mining is the subject of future work. This could be an area of research in the future.

6 Future Work

The paradigm of *intelligent foundations* to knowledge based systems holds great promise and potential. Especially within healthcare these techniques are making some significant differences. However there are areas that need further research and elaboration. Current

state of the art AI techniques involve an ensemble of methods. Intelligent Foundations, strives to formalize and document a set of reusable patterns, language and processes to effectively apply them on a large and universally diverse setting. This paper was born out of a case study to address seemingly unsolvable and hard to automate and humanly unscalable problems. Here are some high level areas for future work.

1. *Universal application of techniques*: One area of future work is to continue the rigorous formalization of the techniques described here for universal application to other diverse verticals.
2. *Identifying more intelligent foundations*: While this paper attempted to do a reasonable job in identifying key intelligent foundational capabilities and building a framework for intelligent knowledge base systems, there is a significant opportunity for more research and elaboration, to further refine and enhance this concept and framework. Future work should continue to investigate other intelligent foundations that can introduce other intelligent capabilities that make the reality of a *virtual knowledge worker* real.
3. *Exploration of intelligent capabilities*: While this paper, outlined a few of the capabilities that can be built atop these intelligent foundations, there are more to be discovered and documented. This will motivate AI researchers to apply techniques to effectively solve the one real problem i.e. make healthcare cost effective to the general populace.
4. *Automation of conversion and generation of knowledge content*: Lastly, this is a major challenge as current methods are fairly ad-hoc, human and expert intensive. While prevalent data science methods and machine learning techniques seek to solve "*well defined*" problems, the challenge of replacing or augmenting the loss of a human subject matter expert by *emergent* system capabilities is an alien concept. The automation of conversion, rule mining and generation holds great promise and should open up rich areas of research.

7 Conclusion

As stated earlier, this paper came out of a case study, but introduces concepts that hold great potential. The work motivating this paper is a project to modernize and improve the *state of the art* of the BRIDE platform. The case study for automation and generation of rules and knowledge artifacts into the BRIDE platform has been instrumental in practically implementing and proving out a majority of the *intelligent foundations* discussed in this paper. The primary goal has been to, introduce elements of Artificial Intelligence at a grass root level that can provide an ambitious shift in capabilities. *Intelligent Foundations for Knowledge*

Based Systems, provide a perfect foil for solving the giant problem of generational shift, which seems to be driving Digital Transformation.

Conflict of Interest The author declares no conflict of interest.

Acknowledgment The author would like to acknowledge the executive leadership at Delta Dental for sponsoring and fueling the innovation, and for their continual commitment and support for the BRIDE platform, which remains a source of inspiration and innovation to the organization.

References

- [1] M. K. Agaram, "Intelligent Discovery Features for EDM and MDM Systems," 2018 IEEE 22nd International Enterprise Distributed Object Computing Workshop (EDOCW), Stockholm, 2018, pp. 135-144. doi: 10.1109/EDOCW.2018.00028
- [2] A. Finkelstein, et. al. "Moral Hazard in Health Insurance", Columbia University Press, 2015.
- [3] D.M.Cutler, R.J.Zeckhauser "Adverse Selection in HealthCare", 1997
- [4] D.M.Cutler, S.J.Reber "Paying For Health Insurance: The Trade-off Between Competition And Adverse Selection"
- [5] T. Hall, "'Monte Carlo Estimation of the Impact of Adverse Selection on Multiple Option Dental Plans", 2010.
- [6] M. K. Agaram and C. Liu, "Vocabulary Model Requirements for Production Rule Systems," 2012 IEEE 16th International Enterprise Distributed Object Computing Conference Workshops, Beijing, 2012, pp. 132-139. doi: 10.1109/EDOCW.2012.26
- [7] M. K. Agaram and B. Laird, "A Componentized Architecture for Externalized Business Rules," 2010 14th IEEE International Enterprise Distributed Object Computing Conference, Vitoria, 2010, pp. 175-183. doi: 10.1109/EDOC.2010.26
- [8] M. K. Agaram and C. Liu, "An Engine-Independent Framework for Business Rules Development," 2011 IEEE 15th International Enterprise Distributed Object Computing Conference, Helsinki, 2011, pp. 75-84. doi: 10.1109/EDOC.2011.20
- [9] M. K. Agaram "The Phased Approach to Mining Business Rules" *Business Rules Journal* Vol. 8, No. 5, (May 2007)" URL: <http://www.brcommunity.com/a2007/b344.html>
- [10] L.F.Bertain, G.Sibbald "The Tribal Knowledge Paradigm", 2012
- [11] Adam Starks PhD (2013) *The Forthcoming Generational Workforce Transition and Rethinking Organizational Knowledge Transfer*, *Journal of Intergenerational Relationships*, 11:3, 223-237, DOI: 10.1080/15350770.2013.810494
- [12] M.K.Agaram "Traceability Aspects for Business Rules Management" *Business Rules Journal* Vol. 9, No. 1, (Jan. 2008)" URL: <http://www.brcommunity.com/a2008/b384.html>
- [13] Cahill, Terrence F., EdD, FACHE; Sedrak, Mona, PhD, PA *Leading a Multigenerational Workforce: Strategies for Attracting and Retaining Millennials*, *Frontiers of Health Services Management*: October 2012 - Volume 29 - Issue 1 - p 315
- [14] Foss et. al. "Knowledge Governance Processes and Perspectives", Oxford University Press, 2009.
- [15] Attila Bruni, Silvia Gherardi, Laura Lucia Parolin *Knowing in a System of Fragmented Knowledge, Mind, Culture, and Activity*, 14:1-2, 83-102, 2007
- [16] D.C.Luckham "The Power of Events: An Introduction to Complex Event Processing in Distributed Enterprise Systems", Addison-Wesley Longman Publishing Co., 2001
- [17] F.X.Campion, G. Carlsson "Machine Intelligence for Healthcare", 2017