POLITECNICO DI MILANO

Facoltà di Ingegneria Industriale e dell'Informazione

Tesi di Laurea Magistrale in Ingegneria Gestionale

A MODEL TO DEPLOY ENTERPRISE SEARCH ENGINES FOR BIG DATA HANDLING IN THE PRODUCT DEVELOPMENT PROCESS

Relatore: Prof. Sergio TERZI

Gianluca FERRARI matr.814480

Anno Accademico 2014/2015

“The ultimate search engine would basically understand everything in the world, and it would always give you the right thing.
And we’re a long, long way from that”

Larry Page

Abstract - English

The world of manufacturing has so far not fully understood the value of enterprise search engines because software vendors, resellers and implementers have focused their attention on the technical features that make these tools useful to product architects and engineers. Enterprise search has been underrated in terms of technical capability and flexibility to adapt to any necessity a product developer may have when performing his/her job. In fact, the technology behind enterprise search engines is so vast that its benefits go well beyond the technical domain: it can support all activities and help anyone dealing with product development. This technology is detailed in the literature, and there is evidence that manufacturing enterprises do have problems in managing an increasing amount of big data in their product development process. However, no scholar has ever related enterprise search engines to the broad concept of product development.

The purpose of this research is to prove that an enterprise search engine is a suitable software system to handle the huge amount of data generated by products during their development process, as well as a valuable complement of the existing enterprise software infrastructure. This work also details the technical features of the search engines, focusing in particular on the ones suitable to support the specific activities of the product development process. In addition, a description of a state-of-the-art enterprise search engine (Exalead) is provided herewith with the scope of giving a better understanding of the main subject of the research.

The study subsequently puts forward a model for the correct deployment of enterprise search engines to manufacturing companies. The model is tested in three different scenarios in order to prove its robustness in terms of enrichment of the literature, of practical value for software vendors and orientation towards the future of the manufacturing world.

Abstract - Italiano

Il mondo del manifatturiero non ha sino ad oggi compreso appieno il valore dei motori di ricerca aziendali perché le aziende dell'Information Technology, i rivenditori di software e gli implementatori si sono fin qui concentrati unicamente sulle caratteristiche tecniche che rendono questi strumenti utili per i progettisti. Il settore dell'Enterprise Search è quindi rimasto sottostimato in termini di flessibilità tecnica e capacità di adattarsi a qualsivoglia esigenza nell'ambito dello sviluppo prodotto. In effetti, la tecnologia su cui si basano i motori di ricerca aziendali è così ampia che i benefici da essa derivanti vanno ben oltre il perimetro di un ufficio tecnico: essa è in grado di supportare qualsiasi attività e fornire sostegno a chiunque sia coinvolto nello sviluppo prodotto. La tecnologia è dettagliata nella letteratura, in essa sono rintracciabili testimonianze riguardo ai problemi che le aziende manifatturiere si trovano ad affrontare relativamente alla gestione di una crescente quantità di big data. Tuttavia, ad oggi nessuno studio ha mai collegato i motori di ricerca aziendali all'accezione più ampia di sviluppo prodotto.

Scopo di questo studio è quindi dimostrare che un motore di ricerca aziendale è un valido completamento dell'attuale infrastruttura software, atto a gestire la grande mole di dati generati dal processo di sviluppo dei prodotti. Inoltre, questo studio dettaglia le caratteristiche tecniche dei motori di ricerca stessi, evidenziandone altresì lo stato dell'arte tramite la descrizione di Exalead, esempio tangibile dell'oggetto principale dello studio.

Fatte queste premesse, lo studio propone un modello per la corretta proposizione del valore dei motori di ricerca aziendali alle imprese manifatturiere. Il modello è quindi testato in tre diversi scenari, così da provarne la solidità in termini di valore teorico per la letteratura, di valore pratico per i venditori di software e di orientamento verso il futuro del mondo manifatturiero.

Extended Abstract - English

Introduction to the study

In the world of manufacturing, product data is knowledge in the form of digital content, and a lot of it is generated during the development process of the product itself.

Manufacturing companies usually rely on complex data management infrastructures that take care of handling every kind of data transiting through an enterprise. Still, they are far from being at ease with data management. Over the last few years, software vendors have been developing one more system to get rid of the problems of the old infrastructure. This new tool is called "Enterprise Search Engine". Enterprise Search Engines (ESEs) exploit new and revolutionary searching capabilities to collect information from all around the company.

Manufacturing enterprises have started to take into consideration ESEs: they are curious about their capabilities. Therefore, it is up to enterprise software vendors to come up with the right market strategy in order to address the needs of manufacturing enterprises and deliver ESEs. In particular, vendors face the challenge of matching a new technology with an infrastructure that supports the product development process.

Reasons

ESEs are expected to become an important part of the software architecture in manufacturing enterprises. However, there are still three unanswered problems in three different scopes:

1) <u>Theory:</u> despite listing all the advantages that a powerful search engine brings and reporting the advancements of enterprise search technology, the literature shows no real trace of a link between ESEs and Product Development (PD).
2) <u>Practice:</u> scholars' studies as well as the enterprises' reports list several advantages following the implementation of an ESE, but no standard method is available on how to convince a customer of its value.
3) <u>Proposition:</u> when selling an ESE, enterprise software vendors tend to use the typical consultancy approach of "what problems can we solve?". ESEs do not just solve single problems, they foster a whole new way of conducting the PD process, and they may happen to solve problems customer enterprises are not even aware of.

Research Objectives

Objective 1: To critically review the literature about ESEs and the software for Big Data management in the manufacturing context.

Objective 2: To prove that ESEs are an effective tool for manufacturing companies to support the designing and engineering of their products.

Objective 3: To examine the technical characteristics of ESEs and shift their focus from single vertical market solutions to the technology as a whole.

Objective 4: To match the advantages and benefits of ESEs' technology with the requirements and needs of the enterprises.

Objective 5: To develop a standard model to propose the value of ESEs to manufacturing enterprises.

Objective 6: To validate the developed model.

Research scope

This research is expected to bring the following advancements:

- A detailed study of Big Data in the PD process of manufacturing enterprises.
- A detailed study of ESEs, along with their history, features, practical applications and future advancements.
- A better model for an effective value deployment of ESEs to manufacturing enterprises.

Overview of Big Data

Big Data is a collection of vast data sets having a great diversity of types so that it becomes difficult to process them by using state-of-the-art data processing approaches or traditional data processing platforms [2]. It is possible to measure how "big" data are according to the so-called 5Vs. We decline the 5Vs for the manufacturing context:

- *Volume*: typical data shared inside manufacturing enterprises are the ones related to the design and engineering of products. Depending on the complexity of the design, the files exchanged inside the enterprise can have a volume up to tens of giga-bytes.
- *Velocity:* the rate at which data are generated depends not only on the speed of the designers, but also on the frequency of data change (i.e.: product versioning) and exchange.

- *Variety:* since the design of a product requires the contribution of different people with different skills, it also requires the use of different software tools to manage it.
- *Veracity:* Big Data need to be stored in safe repositories and access has to be checked via the implementation of security protocols. The risk of granting full access to all users is that data can get jeopardized, thus compromising the final product.
- *Variability:* data can have a different meaning according to the person that is accessing them. Also, the larger the amount of information Big Data carry within, the larger the number of different interpretations they offer.

Big Data and Product Development

The new PD process differs from industry to industry and from firm to firm. However, it is possible to draw a general process that is suitable for every kind of firm: the Booz, Allen and Hamilton (BAH) Model, published in 1982 [15]. This is also the model that has been adopted for the purposes of this study because, despite being rather aged, it is still widely in use.

Image 1: Stages of new Product Development (Booz, Allen and Hamilton, 1982)

Enterprises implement the BAH Model following two main schemes. The first one consists of applying the model to the company structure as it is, relying on the enterprise's departments, thus letting each business unit take care of its own part of the process. The second approach consists of assembling a team of individuals from various functions for the duration of the development process and to allocate the task of making subsets of decisions among them [34].

In the end, the innovation of physical products is a process including technical design, manufacturing, requirements management, etc. In order to support these activities, manufacturing enterprises use a series of software tools: Computer Aided Design (CAD), Computer Aided Engineering (CAE), Computer-Aided Manufacturing (CAM) and Product Data Management (PDM).

The above-listed activities require the individuals of organizations to accomplish the learning and accumulation of new knowledge. An individual's knowledge has to be turned into information. This way, other members of the organization can use that information in order to apply it and create new values for the organization. For this reason and because of the extreme importance of data and information sharing, enterprises have adopted a particular software tool to manage the PD process: Product Lifecycle Management (PLM) software systems allow all of the other above-detailed systems to interact and exchange information. PLM itself is the discipline controlling the evolution of a product and providing the organization with the accurate product information at the right time in the right format during the entire product lifecycle [22].

The four above-listed software systems plus the PLM superstructure are the backbone of PD. However, as the volume of data grows, the challenge is now being able to access and analyze it effectively [7]. Besides, these systems are limited to the PLM systems databases, so they do not take into consideration all the other instruments companies use to support the development of products. For example, a mechanical engineer does not just spend his time designing clamps on his CAD, but he also exchanges emails with his co workers, sends drafts to his supervisor, checks for other clamp experts inside the company who may know more about the subject. This is also true for Excel files, Word documents, and every kind of file in which designers, engineers and developers put down their knowledge in the form of words, numbers and images.

There is a need for a tool serving as a search engine to retrieve, display and analyze information in the form of Big Data also outside the company's PLM software system: an ESE. The search engine must meet specific requirements in order to properly address the Big Data issue in manufacturing enterprises: *User Friendliness*, *Data Merging,* *Security* and *Presentation Capabilities* and *Integration*.

Enterprise Search Engines

The aim of enterprise search is to empower the members of an organization with all the information flowing throughout the organization itself. ESEs achieve this target following a standard process:

Collect: ESEs perform an extraction of complex structured data and associated metadata from every indexed data source. The indexing activity consists of pointing at the various data sources where the ESE goes looking for data to be retrieved.

Process: ESEs sort, summarize, aggregate and analyze the collected data.

Access: ESEs open the processed items and look into them in order to have direct access to the content. In doing this, the ESE checks the user's right to access the content.

Display: ESEs present the results of the search on the monitor depending on the kind of user, the role he/she has in the company and the level of detail he/she needs.

After years of evolution and refinement, the ESE technology is now in a position to match all the needs put forward by manufacturing enterprises:

User Friendliness: the first step towards user friendliness is designing the ESE to look as similar as possible to a web-based search engine. As a consequence, the Google-like search bar and the browser navigation page are both adjustments to make the entire system more similar to the world outside the enterprise. Furthermore, the use of these web browsers can eliminate the need to install and maintain specialized client software on each machine [6].

Regarding the searching activity itself, advanced ESEs propose semantic interpretation of natural language queries. Over the years, ESEs have been improved with semantics-sensitive technology. This translates into several advantages: query suggestions, auto-spelling correction, type-ahead query, fuzzy matching of results and faceted navigation.

Users can also perform analyses on the data they extract with the ESE. Big Data analytics is the process of using analysis algorithms running on powerful supporting platforms to uncover potentials concealed in Big Data, such as hidden patterns or unknown correlations. ESEs can also be empowered with contextual data, role- and rule-based views, user ratings and tags. Another concern for users of ESEs is how to identify relevant queries information out of the overwhelming amount of data displayed in return. In order to divide the documents into groups, the engines implement algorithms for clustering analysis.

As significant amount of knowledge generated during the design and manufacturing phases is associated with the 3D model, ESEs can retrieve similar 3D models based on their shape, thus retrieving also related knowledge that cannot be discovered by other means [50].

Data Merging: the index replaces a traditional relational database structure as the primary vehicle for information access and reporting, so ESEs automatically structure unstructured information, enrich semi-structured or unstructured data and correlate it with structured data.

Security: ESEs access the other enterprise software systems without making any change to the existing security protocols. During the process of generating results, the search engine matches the authenticated user's access rights with the retrieved documents' access rights defined in the Access Control Lists.

Presentation Capabilities: ESEs deliver a fully unified view of information, providing uniform access and reporting for heterogeneous data in a unified format with minimal duplication. The indexing modules use statistical, linguistic and semantic processing to transform heterogeneous, multi-source data into a unified base of actionable information intelligence. The ESE further performs qualitative and quantitative analytics on textual, numerical, temporal and geographical values. The final result of the search integrates intelligence from all relevant sources and supports both qualitative and quantitative analytics with a single platform. As a matter of fact, real time analytics play a key role in the presentation of results.

Integration: ESE are fully and totally integrated with all other PD software systems. However, ESEs do not track or modify the processes like PLM does. They rather act "on top" of existing systems without modifying their processes. Connectors allow ESEs to perform feed management, including tools to join, merge, and cascade feeds from all over the company: email alerts, push notifications and documents uploading from ERP, PLM and any other platform.

Model Proposition

The Model is a roadmap for the proper deployment of ESEs. Its aim is to track down the needs of product developers in manufacturing enterprises and put them in a framework matching all of the benefits brought by the engine.

<u>*Users of the Model:*</u> business developers, technical-sales and sales from ESEs vendors.

<u>*Framework:*</u>

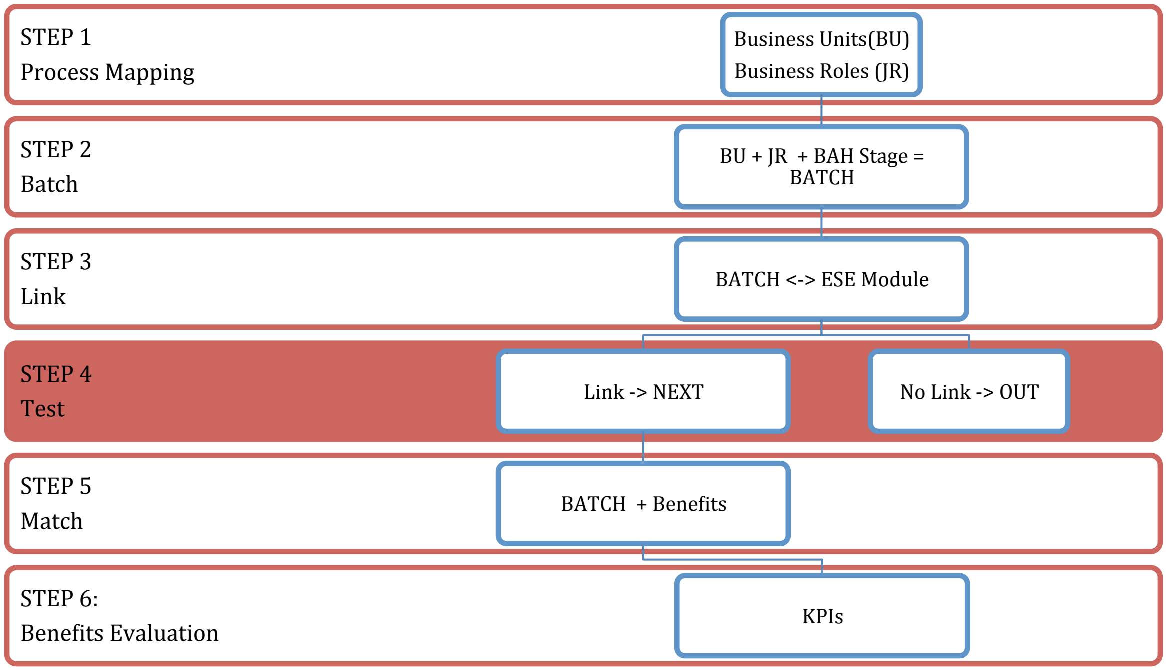


Image 2: Overview of the Model

STEP 1, Process Mapping: identify the Business Units (BUs) and Job Roles (JRs) involved in the company's PD process.

STEP 2, Batch: break down the PD process into the stages of the BAH Model and form a list of BATCHes structured as follows:

BATCH i	
BAH Model Stage	stage (1), ... , stage (7)
Business Unit	name (1), name (2), ... , name (7)
Job Role	name (1), name (2), ... , name (9)
Requirements	...

STEP 3, Link: link every BATCH with the proper vertical market software module in order to fulfill the requirements.

STEP 4, Test: the BATCHes that are consistent with BUs/JRs and modules proceed to STEP 5. The others exit the Model.

STEP 5, Match: matches each BATCH with the specific benefits the module brings to it.
STEP 6, Benefits Evaluation: proposes suitable KPIs to measure the benefits of every BATCH.

BATCH i	
BAH Model Stage	stage (1), ... , stage (7)
Business Unit	name (1), name (2), ... , name (7)
Job Role	name (1), name (2), ... , name (9)
Requirements	...
ESE Module	Name / Trigram / ID Code
Benefits	...
KPI	1) ... 2) ... n) ...

The BATCHes at the end of STEP 6 represent the offering that the user of the Model proposes to the customer.

Model Limitations

Prerequisites: the customer enterprise has a pre-installed software infrastructure.
Economic Feasibility: the Model does not take into consideration aspects related to the price of the ESE. This Model is to be used for the deployment only.
Limited KPIs: not all the BATCHes would have a matching KPI, because many benefits are not measurable through mathematic formulas and financial indicators.
Diversity among Enterprises: the lack of a clear organizational chart and the misleading perception of JRs could make the application of the Model more difficult.

Model Validation

In order to validate the Model, we tested it in three different scenarios:

1) A practical use-case, in which we select a manufacturing enterprise and compare the value proposition out of the Model with the one put forward by a best-in-class software developer and vendor company.
2) The analysis of the Model by a best-in-class ERP and enterprise search solutions reseller and implementer. We asked both a Sales and a Technical Sales person to express an opinion about the Model and its robustness.

3) The analysis of an expert independent consultant in the field of PLM.

Model Validation – USE-CASE

Vendor: Dassault Systèmes

Customer: COMAU (COnsorzio MAcchine Utensili)

Methodology: we performed a gap analysis between Dassault Systèmes' and our value proposition procedure in order to highlight the difference between the two in the result.

Findings: Dassault Systèmes' standard procedure is somehow shortsighted compared to the proposed Model, because it focuses only on the ESE's modules for product shaping and engineering. The Model takes into account the requirements of all the people involved in the PD process. The Model would allow Dassault Systèmes to put forward a far more comprehensive offering vis-à-vis the one presented with its standard procedure.

Model Validation – Sales Experts Interview

Interviewee: the sales manager and the technical sales manager from Altea S.p.a.

Methodology: separated, semi-structured interviews.

Findings: The experts' opinion is that the Model is generally robust, because:

- It has a strong theoretical background
- It is on the right path to become a best practice in the future, because it addresses the problem out of the research study in the proper way.

The Model does not always apply to a real context because:

- It is NOT scalable. It is suitable for big enterprises, but does not fit with the small and small/medium ones.

Model Validation – Independent Expert Interview

Interviewee: John Stark, president and head consultant for JSA Consulting.

Methodology: unstructured interview.

Findings: the independent expert believes the Model is robust because it addresses the two main issues in the right way. As a matter of fact, mapping the enterprise's roles is the best approach for an implementation model, whereas ESEs will become a key tool in the manufacturing enterprise of the future.

Conclusion & Future Research

Conclusion

Objective 1, 2, 3, 4, 5 have been fully achieved at the end of the study.

Objective 6: has been partially achieved at the end of the study, because the final target of the Model has been narrowed down to big enterprises.

Future Research

Big Data and Product Development: researchers should detail how product data, product data management and the support software systems change when manufacturing enterprises develop a Product-Service System.

Enterprise Search Engines: having access to data does not necessarily mean understanding the content of those data. When dealing with technical issues, in particular, we would prefer requesting an opinion from an expert rather than having access to some data about the issue. For this reason, one of the improvements that researchers want to apply to ESEs is to enable them to act as expertise search systems.

EXALEAD Market Solution Example: in this study we mention Dassault Systèmes' Exalead as an example of state-of-the-art ESE. However, there is one feature that would dramatically boost Exalead's performances and value: business intelligence. An add-on with business intelligence capabilities would allow Exalead to also compete at a corporate strategic level.

Model: The everyday life of enterprises is regulated by the repetitive operations that make them run efficiently. The Model could be improved by making it a procedure for ESEs value proposition to support the entire enterprise's environment.

Extended Abstract - Italiano

Introduzione

Nel mondo del manifatturiero, i dati legati al prodotto sono conoscenza pura nella forma di contenuti digitali, la maggior parte dei quali è generata durante lo sviluppo del prodotto stesso.

Le imprese manifatturiere sono solite dotarsi di complesse infrastrutture software, le quali non sono tuttavia ancora in grado di risolvere tutte le problematiche relative alla gestione dei dati che transitano attraverso l'azienda.

Negli ultimi anni le aziende dell'Information Technology hanno sviluppato un nuovo sistema per affrontare i problemi legati alla vecchia infrastruttura. Questo nuovo strumento prende il nome di "Enterprise Search Engine" (ESE), o "Motore di Ricerca Aziendale". Esso sfrutta nuove e rivoluzionarie capacità di ricerca per la raccolta di informazioni attraverso tutta l'azienda.

Le imprese manifatturiere hanno cominciato a intuirne le potenzialità e mostrano interesse per le loro possibili applicazioni. La sfida per i produttori e venditori di software è ora quella di pianificare una strategia di mercato per venire incontro alle necessità del manifatturiero e vendere gli ESE ai propri clienti. In modo particolare, i vendor affrontano la sfida di mettere in relazione una nuova tecnologia con una preesistente infrastruttura a supporto del un processo di sviluppo prodotto.

Ragioni dello studio

Si prevede che gli ESE diventino parte fondamentale dell'architettura software delle aziende manifatturiere. Tuttavia, vi sono ancora tre diversi aspetti insoluti in tre ambiti differenti:

1) Teoria: nonostante la letteratura elenchi le tecnologie rivoluzionarie e le infinite possibilità dei motori di ricerca, non vi è traccia in essa di un legame tra gli ESE e lo sviluppo prodotto.
2) Pratica: la letteratura e la reportistica dei vendor elencano i vantaggi derivanti dall'installazione di un ESE, ma non delineano un metodo standard per convincere i clienti dell'effettivo valore di queste soluzioni.
3) Proposizione: nel processo di vendita di software si tende a utilizzare l'approccio tipicamente consulenziale del "Quale problema possiamo risolvere?". Gli ESE non solo risolvono uno

specifico problema, ma incentivano un nuovo modo di condurre il processo di sviluppo prodotto, risolvendo a volte problemi di cui l'azienda cliente non era a conoscenza.

Obiettivi della Ricerca

Obiettivo 1: Effettuare uno studio critico della letteratura riguardante gli ESE e i software per la gestione dei Big Data nel contesto manifatturiero.

Obiettivo 2: Dimostrare che gli ESE sono uno strumento efficace per supportare la progettazione e l'ingegnerizzazione dei prodotti nelle aziende manifatturiere.

Obiettivo 3: Esaminare le caratteristiche tecniche degli ESE e spostare l'attenzione dal singolo verticale del software alla tecnologia comprensiva di base.

Obiettivo 4: Collegare vantaggi e benefici degli ESE con i requisiti e le necessità delle aziende.

Obiettivo 5: Sviluppare un modello per l'impiego degli ESE alle aziende manifatturiere.

Obiettivo 6: Validare il modello sviluppato.

Scopo della Ricerca

- Uno studio dettagliato dei Big Data nello sviluppo prodotto delle aziende manifatturiere.
- Uno studio dettagliato degli ESE inclusivo di storia, caratteristiche, applicazioni pratiche e migliorie future.
- Un modello più avanzato per la proposta di valore degli ESE alle aziende manifatturiere.

Panoramica sui Big Data

Il termine "Big Data" indica un ampio insieme di dati con caratteristiche tali da rendere arduo il loro processamento persino usando i migliori approcci e piattaforme [2]. È possibile misurare l'effettiva complessità di questi dati attraverso cinque variabili dette le 5V, che decliniamo nello specifico contesto manifatturiero:

- *Volume*: i Big Data condivisi nelle aziende manifatturiere sono per la maggior parte legati alla progettazione dei prodotti. Secondo la complessità del progetto, i file scambiati possono variare in volumi fino alla decina di giga-byte.
- *Velocity (velocità):* il ritmo cui sono generati i dati dipende non solo dalla rapidità dei progettisti, ma anche dalla frequenza con cui i dati sono modificati e scambiati.

- *Variety (eterogeneità):* la progettazione di un prodotto richiede la collaborazione di ruoli diversi con scopi diversi, e richiede quindi l'utilizzo di diversi strumenti software.
- *Veracity (veracità):* i Big Data devono essere salvati in repository sicuri, e l'accesso a essi deve essere garantito solo attraverso l'implementazione di protocolli di sicurezza. In caso contrario, il rischio è di compromettere i dati e, di conseguenza, il prodotto finale.
- *Variability (variabilità):* i dati assumono un diverso significato a seconda della figura professionale che vi accede.

Big Data e Sviluppo Prodotto

Il processo di sviluppo nuovo prodotto è diverso da settore a settore e da azienda ad azienda. Tuttavia, è possibile tracciare uno schema sostanziale valido per qualsiasi impresa: il modello di Booz, Allen and Hamilton del 1982 [15], altrimenti detto "BAH Model". Questo è anche il modello adottato nello studio perché, seppur datato, oggigiorno è ancora largamente impiegato.

Figura 1: Stadi dello sviluppo nuovo prodotto (Booz, Allen and Hamilton, 1982)

Le aziende implementano il modello BAH secondo due metodi differenti. Il primo consiste nell'applicare il modello all'organigramma classico dell'azienda, facendo cioè affidamento sui dipartimenti e lasciando che ogni divisione si occupi della propria parte del processo. Il secondo approccio invece prevede la formazione di team di individui provenienti da diverse divisioni durante la durata del processo di sviluppo prodotto. A ciascun membro del team verranno quindi assegnate attività in base a conoscenza, formazione e capacità individuali.

Fatte le dovute considerazioni, l'innovazione di prodotto è un processo che include attività come disegno tecnico, fabbricazione, gestione dei requisiti etc. Per supportare queste attività, le imprese manifatturiere usano una serie di strumenti software: Computer Aided Design (CAD), Computer Aided Engineering (CAE), Computer-Aided Manufacturing (CAM) e Product Data Management (PDM).

Tutte le attività sopra citate richiedono agli individui di un'organizzazione di imparare e assimilare nuova conoscenza. La conoscenza di ogni individuo deve poi essere poi trasformata in informazione. In questo modo, altri membri dell'organizzazione possono usare utilizzarla per creare nuovo valore per l'impresa. Per questo motivo e per l'estrema importanza che ricopre lo scambio di dati e informazioni, le aziende hanno adottato uno speciale strumento software per gestire l'intero processo di sviluppo prodotto: le piattaforme di Product Lifecycle Management (PLM) consentono ai sistemi sopra citati di interagire e scambiare informazioni. Il PLM in se è la scienza che controlla l'evoluzione di un prodotto e fornisce all'azienda le informazioni sul prodotto corrette, in tempo reale e nel formato opportuno durante tutto il ciclo di vita del prodotto stesso [22].

I quattro sistemi software elencati e il PLM costituiscono l'architrave dello sviluppo prodotto cui tuttavia, causa l'aumento esponenziale del volume di dati [7], diventa sempre più arduo accedere ed estrarre informazioni efficacemente. Inoltre, i sistemi attuali sono limitati al database del PLM, non tenendo quindi in considerazione tutti gli altri strumenti usati dalle aziende per svolgere attività collaterali allo sviluppo prodotto. Per esempio, un ingegnere non impiega il suo tempo esclusivamente progettando morsetti con il suo CAD; egli scambia email con i colleghi, invia bozze di progetto al suo supervisore e chiede l'opinione degli esperti di morsetti nell'azienda, che probabilmente hanno una maggiore conoscenza sull'argomento. Lo stesso ragionamento vale per i file Excel, i documenti Word e per tutti i tipi di file in cui disegnatori, ingegneri e sviluppatori ripongono la propria conoscenza in termini di testo, numeri e immagini.

Vi è la necessità di uno strumento per accumulare, ordinare, visualizzare e analizzare l'informazione sotto forma di Big Data anche al di fuori del sistema PLM: un ESE. I motori di ricerca devono tuttavia rispondere a determinate richieste dell'utente medio per indirizzare al meglio la gestione dei Big Data nel manifatturiero. Le richieste sono in termini di *User Friendliness*, *Merging dei Dati*, *Sicurezza*, *Presentazione Risultati* e *Integrazione*.

Enterprise Search Engines

Lo scopo ultimo di un motore di ricerca aziendale è di fornire ai membri dell'organizzazione tutte le informazioni conservate nell'impresa stessa. Gli ESE raggiungono questo obiettivo compiendo una serie di azioni standard:

Collect: estrazione di dati strutturati complessi da ogni sorgente dati indicizzata e associazione di questi ai rispettivi metadati. L'indicizzazione consiste nell'indicare all'ESE le sorgenti in cui cercare i dati da estrarre.

Process: filtraggio, riassunto, aggregazione e analisi dei dati raccolti.

Access: apertura e analisi degli oggetti processati per accedere direttamente al contenuto. Durante questa attività, gli ESE controllano i diritti d'accesso degli utenti.

Display: presentazione dei risultati della ricerca in formato diverso a seconda dell'utente, del suo ruolo in azienda e del livello di dettaglio desiderato.

Dopo anni di evoluzione e affinamento, la tecnologia degli ESE è ora in grado di indirizzare tutte le necessità espresse dalle aziende manifatturiere precedentemente elencate:

User Friendliness: il primo passo per ottenere un sistema facile da utilizzare è progettare l'ESE il più similmente possibile a un motore di ricerca per il Web; accorgimenti come la barra di ricerca "alla Google" e la navigazione attraverso un browser rendono lo strumento più simile a quelli usati nella vita quotidiana.

Gli ESE più avanzati supportano l'attività di ricerca con l'interpretazione semantica del linguaggio. Nel corso degli anni gli ESE sono stati infatti potenziati con tecnologie semantiche per renderli più adatti ad ogni genere di utente. Questo si traduce in diverse migliorie: suggerimenti per il completamento delle query, auto-correzioni grammaticali, ricerca fuzzy dei risultati e navigazione facettizzata.

Un altro tema caro agli utenti è l'identificazione delle informazioni veramente utili tra tutti i risultati riportati dopo una query. Per questo motivo gli ESE implementano procedure di clustering per la suddivisione dei documenti in categorie che ne facilitano l'organizzazione. Gli utenti possono inoltre analizzare i dati estratti grazie ad algoritmi di affinamento della ricerca atti a scoprire potenziali informazioni celate nei Big Data, come ad esempio pattern e correlazioni.

Gli ESE possono essere ulteriormente potenziati con l'aggiunta di dati contestuali, visualizzazioni role- o rule-based, meccanismi di rating e tag.
Infine, gli ESE consentono la ricerca di modelli 3D basandosi sulla loro geometria per supportare le fasi di progettazione e fabbricazione, in cui gran parte della conoscenza è proprio associata ai suddetti modelli 3D.

Merging dei Dati: l'indice sostituisce la tradizionale struttura relazionale del database come principale veicolo per accedere e riportare le informazioni. In tal modo, gli ESE strutturano automaticamente le informazioni non strutturate arricchendo i dati semi o non strutturati e mettendoli in relazione con quelli invece strutturati.

Sicurezza: gli ESE accedono agli altri sistemi software aziendali senza modificarne in alcun modo i protocolli di sicurezza preesistenti. Durante il processo di generazione dei risultati, il motore di ricerca confronta i diritti di accesso degli utenti con quelli relativi ai documenti raccolti definiti nella Lista di Controllo Accessi (ACL).

Presentazione Risultati: gli ESE riportano in una videata unica tutte le informazioni, riportando dati eterogenei in un formato uniforme e con un numero minimo di duplicati. I moduli di indicizzazione usano processi statistici, linguistici e semantici per trasformare dati eterogenei e multi-sorgente in una base unica e facilmente accessibile. Gli ESE conducono inoltre un'analisi qualitativa e quantitativa su valori testuali, numerici, temporali e geografici.

Integrazione: gli ESE sono completamente integrati con ogni altro sistema software per lo sviluppo prodotto, agendo tuttavia al di sopra (on top) dei sistemi esistenti, non modificandone quindi in alcun modo i processi. I connettori consentono agli ESE di gestire il flusso dei feed con strumenti per collegare, unire e sequenziare feed provenienti da tutta l'organizzazione: email, notifiche push e caricamento di documenti da ERP, PLM o altre piattaforme.

Proposta Modello

Il Modello è una guida strutturata per una corretta proposizione del valore e per l'impiego di un ESE. Lo scopo del Modello è identificare i requisiti degli sviluppatori del prodotto nelle imprese manifatturiere e inserirli in una cornice comprensiva di tutti i benefici derivanti dall'installazione di un motore di ricerca aziendale.

<u>*Utilizzatori del Modello:*</u> business developers, tecnici-commerciali e commerciali di aziende venditrici di ESE.

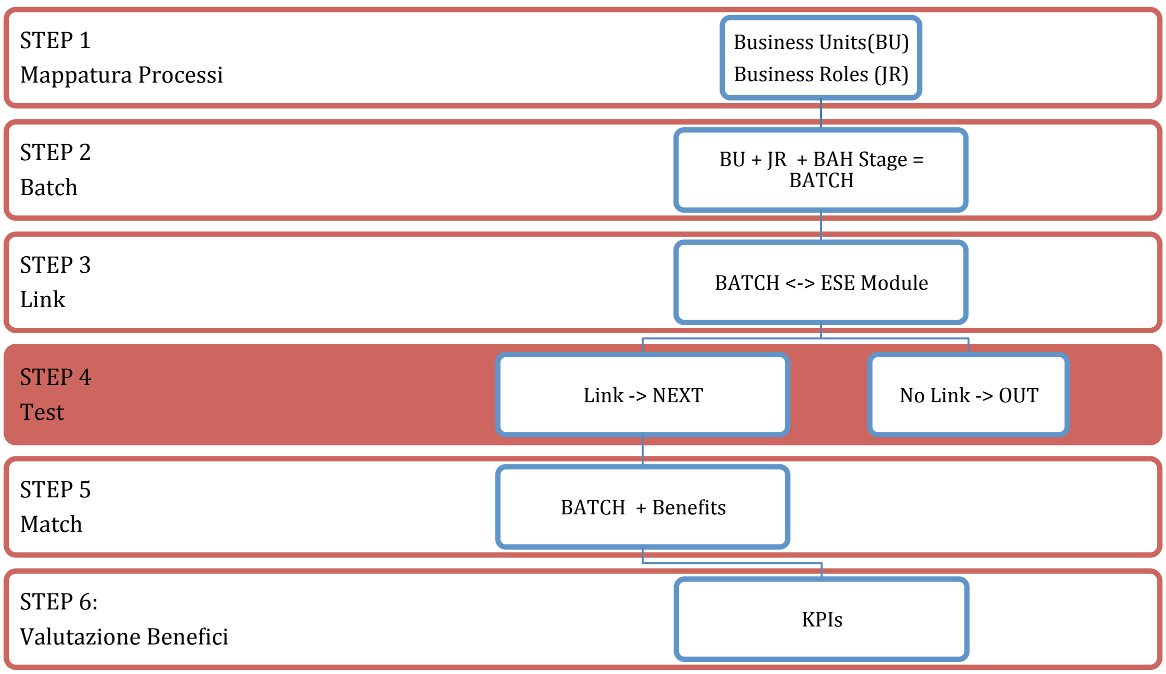


Figura 2: Struttura del Modello

STEP 1, Mappatura Processi: identificare le Business Unit (BUs) e i Ruoli (JRs) facenti parte del processo di sviluppo prodotto dell'azienda cliente.

STEP 2, Batch: dividere il processo di sviluppo prodotto negli stadi del modello BAH e creare una lista di BATCH strutturati come segue:

BATCH i	
Stadio BAH	stadio (1), ... , stadio (7)
Business Unit	nome (1), nome (2), ... , nome (7)
Job Role	nome (1), nome (2), ... , nome (9)
Requisiti	...

STEP 3, Link: collegare ciascun BATCH con il modulo software verticale appropriato per il soddisfacimento dei requisiti identificati.

STEP 4, Test: i BATCH comprensivi di BU/JR e modulo software verticale procedono nello STEP 5, gli altri escono dal processo.

STEP 5, Match: collegare ciascun BATCH agli specifici benefici portati dal modulo.

STEP 6, Valutazione Benefici: individuare KPI per la misura dei benefici di ciascun BATCH.

BATCH i	
Stadio BAH	stadio (1), ... , stadio (7)
Business Unit	nome (1), nome (2), ... , nome (7)
Job Role	nome (1), nome (2), ... , nome (9)
Requisiti	...
Modulo ESE	Nome / Trigram / Codice ID
Benefici	...
KPI	1) ... 2) ... n) ...

I BATCH che escono dallo STEP 6 rappresentano l'offerta che gli utilizzatori del Modello presentano al cliente finale.

Limiti del Modello

Prerequisiti: l'azienda cliente deve possedere un'infrastruttura software preinstallata.

Fattibilità Economica: il Modello non tiene conto del prezzo dei moduli dell'ESE. Il Modello è usato esclusivamente per la proposta tecnica del prodotto.

KPI Limitati: non tutti i BATCH avranno KPI associati, perché molti dei benefici non sono misurabili attraverso formule matematiche e indicatori finanziari.

Diversità tra le Aziende: la mancanza di un chiaro organigramma e la concezione approssimativa dei JRs può rendere difficoltosa l'applicazione del Modello.

Validazione Modello

Ai fini della validazione, il Modello è stato testato in tre diversi scenari:

1) Un use-case pratico, in cui la procedura proposta nel Modello è paragonata a quella di un'azienda top class tra i produttori e venditori di ESE.

2) L'analisi del Modello da parte di un'azienda top class tra i rivenditori e implementatori di soluzioni ERP e enterprise search. È stata chiesta a un a un commerciale e a un tecnico-commerciale l'opinione circa la solidità del Modello.
3) L'analisi di un esperto consulente indipendente nel campo del PLM.

Validazione Modello – USE-CASE

Vendor: Dassault Systèmes

Cliente: COMAU (COnsorzio MAcchine Utensili)

Metodologia: si effettua una analisi degli scostamenti tra i due diversi risultati ottenuti dalla procedura di Dassault Systèmes dal il Modello.

Risultati: la procedura standard di Dassault Systèmes è in una certa misura miope se paragonata al Modello proposto perché il vendor concentra la sua attenzione sulla proposta dei moduli concepiti esclusivamente per supportare attività di modellazione e ingegnerizzazione. Al contrario, il Modello tiene in considerazione le necessità di tutte le figure coinvolte nello sviluppo in senso lato del prodotto. Se adottato, il Modello consentirebbe a Dassault Systèmes di presentare un'offerta nettamente più comprensiva rispetto alla procedura ora adottata.

Validazione Modello – Intervista Rivenditore

Intervistato: il manager dei commerciali e il manager dei tecnico-commerciali di Altea S.p.a.

Metodologia: sono effettuate separatamente due interviste semi strutturate identiche.

Risultati: gli esperti intervistati valutano il Modello come valido in quanto:

- Ha un solido fondamento teorico.
- È correttamente orientato a diventare una pratica comune in futuro, poiché indirizza il problema emerso dalla letteratura nel modo corretto.

Validazione Modello – Intervista con un Consulente Indipendente

Intervistato: John Stark, presidente e consulente capo di JSA Consulting,

Metodologia: intervista non strutturata.

Risultati: l'esperto indipendente crede che il Modello sia valido in perché indirizza le due problemi chiave in modo corretto: la mappatura dei processi secondo i ruoli, che è il miglior approccio per un modello implementativo, e le necessità future delle imprese manifatturiere, che necessiteranno di uno strumento innovativo e flessibile come gli ESE.

Conclusione e Studi Futuri

Conclusione

Alla conclusione dello studio gli *Obiettivi 1, 2, 3, 4, 5* sono stati totalmente raggiunti.
L'*Obiettivo 6* è stato solo parzialmente raggiunto perché, contrariamente alle previsioni, il target finale del Modello si rivelato essere limitato alle grandi imprese.

Studi Futuri

Big Data e Sviluppo Prodotto: i ricercatori potrebbero dettagliare come la gestione dei dati del prodotto e i sistemi di supporto software cambiano quando le aziende manifatturiere sviluppano non solo un oggetto fisico ma un Sistema Prodotto-Servizio.

Motori di Ricerca Aziendali: avere accesso ai dati non significa automaticamente comprendere il significato di quegli stessi dati. In particolare, quando si affrontano difficoltà di natura tecnica si preferisce chiedere il parere di un esperto anziché avere accesso a dei banali dati circa il problema. Per questo motivo, uno dei temi ricorrenti nella ricerca di miglioramento per gli ESE è di dotarli di strumenti per la ricerca di esperti (expertise search).

Esempio di Motore di Ricerca - EXALEAD: in questo studio descriviamo Exalead di Dassault Systèmes come esempio di stato dell'arte dei motori di ricerca aziendali. Tuttavia, Exalead può essere ulteriormente migliorato con l'introduzione della tecnologia per attività di business intelligence, così da renderlo uno strumento decisionale a livello di strategia corporate.

Modello: il quotidiano delle aziende manifatturiere è regolato da attività di routine che, se svolte correttamente, rendono l'impresa efficiente. Il Modello può essere ampliato così da renderlo una procedura per proporre i vantaggi degli ESE a supporto di tutto il mondo delle operations.

Index of Thesis

List of Figures

List of Tables

List of Abbreviations

ACL: Access Control List
API: Application Programming Interface
BAH: Booz, Allen and Hamilton
BI: Business Intelligence
BOM: Bill Of Materials
BPM: Business Process Management
BU: Business Unit
CEO: Chief Executive Officer
CIO: Chief Information Officer
CMMS: Computerized Maintenance Management System
CMS: Content Management System
COMAU: COnsorzio MAcchine Utensili
CPM: Corporate Performance Management
CPU: Central Processing Unit
CRM: Customer Relationship Management
CSV: Comma-Separated Values
CTO: Chief Technology Officer
DBMS: Database Management System
DOS: Disk Operating System
DS: Dassault Systèmes
EBOM: Engineering Bill of Materials
ERP: Enterprise Resource Planning
ESE: Enterprise Search Engine
FSCM: Financial Supply Chain Management
GFS: Google File System
HDD: Hard Disk Drive
HTML: HyperText Markup Language
JR: Job Role
KPI: Key Performance Indicator
MBOM: Manufacturing Bill of Materials
MES: Manufacturing Execution System
OEM: Original Equipment Manufacturer
OLAP: Online Analytical Processing
PD: Product Development
PPS: Product – Service System
QMS: Quality Management System
R&D: Research and Development
RFID: Radio Frequency Identification
SB: Social Bookmarking
SBA: Search Based Application
SCM: Software Configuration Management
SME: Small-Medium Enterprise
WIP: Work In Progress
WMS: Warehouse Management System

XLS: Excel File
XML: Extensible Markup Language

1 Introduction to the Study

1.1 Introduction

The world of manufacturing has always been characterized by a remarkable amount of data.

Data cannot go by a single definition, but it is certainly accurate to say that data are knowledge in the form of digital content. Product data is everything we know about the product, and most of them are collected during the development process.

Francis Bacon once said that knowledge is power. If this statement is still valid nowadays (and it certainly is) then data is the pure essence of power in the Product Development phase of manufacturing enterprises. Product Development is the essence of manufacturing, for it can be defined as the transformation of a market opportunity and a set of assumptions on product technology into a product available for sale [7]. And selling goods is so far the ultimate goal of manufacturing companies all around the world.

However, possessing a great power (or, in this case, a great amount of data) does not necessarily means being able to use it to one's advantage.

Let's imagine we set up an enterprise for the production of boxes. We would have to design, develop, manufacture parts and then assemble them in a final product, the box. Now, the box is one of the simplest tools ever used by mankind over the centuries. It was invented not only before computers, databases and data management systems, but even well before the birth of industrial manufacturing and production, when artisans used to assemble raw boards or fold gold foils into the shape of a box.

Getting back to our entrepreneurial idea, we would produce only one kind of box; let's say a standard 20 x 20 x 20 cm box to target basically any kind of market. Let's also imagine that, as we are new to the business and we want to keep it simple, we will only make steel boxes.

Let's now suppose that the box can be made of three parts:

- Cover (C)
- Body (B)
- Hinge (H)

Let's say we will use only three colors to make our boxes attractive to customers:

- white (w)
- black (b)
- grey (g)

As we can see from the following picture, in our data management system we now have 9 possible parts, each registered with one single code:

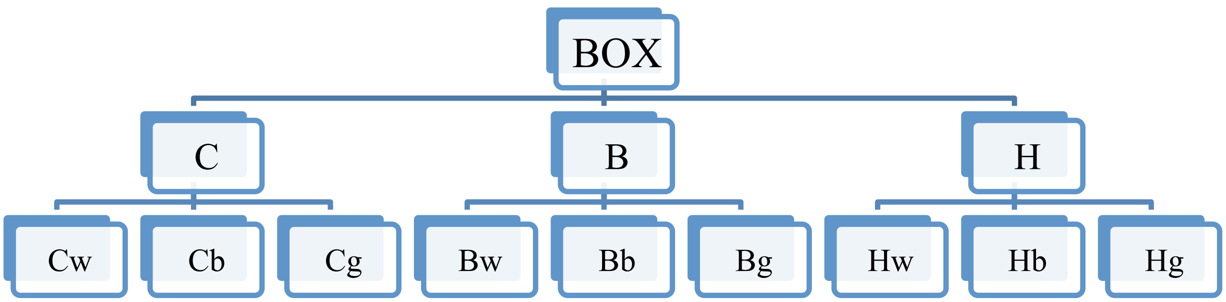


Figure 1.1: Product variants

We may be unable to manufacture all the parts by ourselves, whether for capacity or competence reasons. We may find a supplier that is specialized in making covers, thus saving us a lot of production costs. On the other hand, we may decide to purchase a certain percentage of parts called "Bb" but also to manufacture the rest by ourselves, because we want it to be made our way. So we decide to analyze the components and mark them with an "O" if they are to be outsourced or "I" if they will come from out production plants.

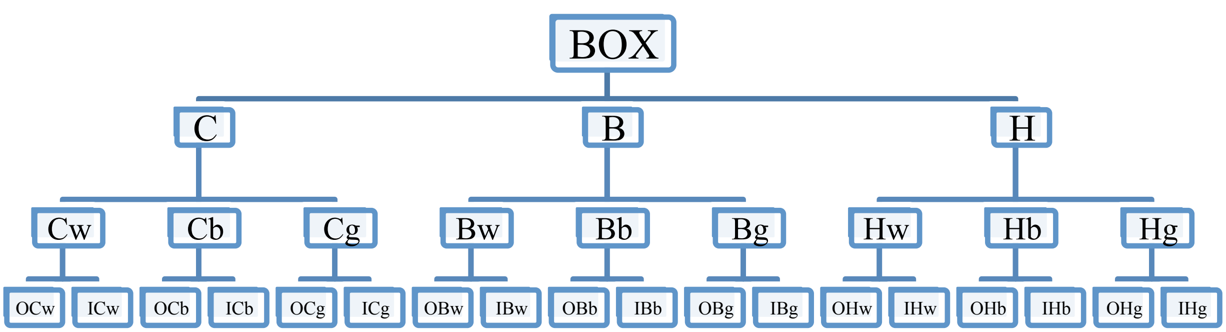


Figure 1.2: Outsource and insource

As we can see from Figure 1.2, there are now 18 product codes in out data management system.

At this point, we need to assemble the parts in order to build up the box, which is the final product. According to our simple calculations, we could put in our catalogue more than 50 possible variants of the same product "Box".
In this made up and simplified scenario, we have a large number of possible configurations. Still, we did not mention anything related to the data from the market or the feedback from the customers, the purchase of raw materials, the management of raw materials, parts, assemblies and WIP warehouses, the design and testing of products, the product teams and the possible concurrent designing activities, the norms and regulations applied to industrial products, the documentation that goes along the product (warranty certificate, brochures etc.) and much more. The issue can only get worse if we consider the complexity of products such as cars, ships and airplanes. These products are made of hundreds, thousands of parts and components and there are hundreds, thousands of people who contribute to their manufacturing process. All of the activities, decisions and tools that we just mentioned are part of the macro Product Development process, which represents the core of every manufacturing enterprise.

Getting back to the concept of the power of data, it is evident that nowadays the problem is no longer about getting data. The real matter is being able to handle data and transform them from an issue into the power each manufacturing enterprise need to be successful.
In the last three decades the advancement of technology has brought several instruments to help enterprises solve this problem. Nowadays, manufacturing companies can rely on complex data management infrastructures designed to take care of handling every kind of data transiting through the enterprise. Still, the world of manufacturing is far from being at ease with data management. These data infrastructures may be an assembly of different software systems purchased in different periods of the company's history, when the company was facing different challenges and had different budgets available. Then again, those software systems may be purchased from a single software vendor, from various software vendors or even custom made. Indeed, the software infrastructure helps the enterprise, but it does not solve all of the issues related to data.

In the last years, software vendors have been developing one more software system to get rid of the problems of the old infrastructure. This new item took the name of "Enterprise Search

Engine". Enterprise Search Engines exploit new and revolutionary searching capabilities to collect information from all around the company. Manufacturing enterprises have started to take into consideration Enterprise Search Engines because they think these software systems can solve the problems we mentioned before: the handling of uncountable data sources generating uncountable kinds of data, and the management of multiple software systems that seldom work well together.

At this point, it is up to enterprise software vendors to come up with the right market strategy in order to address the needs of manufacturing enterprises and deliver Enterprise Search Engines to their customers. In particular, they face the challenge of matching a new technology with an old infrastructure that supports an always changing and evolving Product Development process.

1.2 Research Background

1.2.1 Reasons for Research

Enterprise Search Engines are becoming an important part of the software architecture of manufacturing enterprises. However, there are still plenty of unanswered questions both in the literature and in the enterprises' know-how. In particular, the problem can be divided into three points:

1) Despite listing all the advantages that a powerful search engine brings and reporting the advancements of enterprise search technology, in the literature there is no link between Enterprise Search Engines and Product Development. All the previous studies consider them whether as stand-alone systems or as add-on patches to empower the ERP system, but they never mention a connection with the Product Development software infrastructure.
2) Scholars' studies as well as enterprises' reports list uncountable advantages following the implementation of an Enterprise Search Engine. Still, there is no trace of a standard method to convince customers of its value. Where there is a number of undeniable advantages and a list of common enterprises' needs, there must also be a systematic way to match the two.
3) When selling an Enterprise Search Engine, enterprise software vendors tend to use the typical consultancy approach of "what problems do you have?" and, as a consequence, "what problems can we solve?". This approach is not wrong by any means, but is certainly limited compared to what these tools can do. Enterprise Search Engines do not just solve single

problems, they rather foster a whole new way of conducting the PD process, and they may occur to solve problems the customer enterprises do not even realize to have.

1.2.2 Research Objectives

Objective 1: To critically review the literature about Enterprise Search Engines and the software for Big Data management in a manufacturing context.

Objective 2: To demonstrate that Enterprise Search Engines are an effective tool for manufacturing companies to support the designing and engineering of products.

Objective 3: To examine the characteristics of Enterprise Search Engines and shift the focus from single vertical market solutions to the technology as a whole.

Objective 4: To match the advantages and benefits of Enterprise Search Engines with the requirements and needs of the organizations.

Objective 5: To develop a standard model for the deployment of Enterprise Search Engines in manufacturing enterprises.

Objective 6: To validate the developed model.

1.2.3 Research Scope

The advancements this research is expected to bring are:

- A detailed study of Big Data in the Product Development process of manufacturing enterprises.
- A detailed study of Enterprise Search Engines, along with their history, features, practical applications and future advancements.
- A better model for an effective deployment of Enterprise Search Engines to manufacturing enterprises.

1.3 Research Methodology

Research is a systematic and methodical process of enquiry and investigation, which entails the use of various research methods. It produces and enriches the body of knowledge of a discipline or profession and only by rigorous employment of suitable methodology and relevant research methods can the research be carried out with confidence [64].

To achieve the overall research aim and the six research objectives, the research is conducted based on the following methodology.

1) The study begins with an exhaustive review of the literature about Big Data for manufacturing, software tools for manufacturing products development and Enterprise Search Engines. In addition, the study describes a state-of-the-art Enterprise Search Engine, thus providing a valuable example of the main subject of the research.

2) Starting from the review of the literature, the study points out the gap in past research and the possible benefits deriving from the creation of a new model. This model is dedicated to the deployment and implementation of Enterprise Search Engines supporting the Product Development in manufacturing enterprises.

3) The study details the logic, the background and the theoretical importance of the Model. Then, the study describes in detail the practical aspects of the Model, including its limitations and possible future improvements.

4) The developed Model is validated through three testing scenarios: the implementation of the Model in a manufacturing context, the interview with a software vendor company and the discussion with an expert in the field of PLM.

1.3.1 Premises of the Model

When setting up STEP 1 of the Model (6.3.1), the main issue has been applying the BAH model to a real industrial context. In fact, there is no trace in the literature of a manufacturing company listing its personnel and its BUs and stating how they are related to the BAH model. Also, as it often happens, real industrial life is rather different from the theory, especially when comparing that very theory with complex manufacturing environments. In order to solve this problem, we have analyzed twenty manufacturing enterprises: as far as they are disclosed, we mapped their business processes and organization charts. From the comparison of these companies, the study works out a list of Business Units (BUs) and job roles (JRs) involved in the PD process.

The names of the companies have been omitted for copyright reasons. Nonetheless, in order to provide information that is as clear as possible and to support the robustness of the study, it is assured that the enterprises have been chosen among the Top 100 World manufacturing companies by revenue listed by Fortune [1i]. They belong to the Automotive, Industrial Equipment, Consumer Goods, Aerospace & Defense and Construction Equipment industries. They have been chosen because of their solid position in the market over the years and their brand awareness worldwide. Also, they have been taken out of that particular list because we had make sure their size in terms of output was comparable to one another. In any case, this limitation does not prevent the Model itself from being scalable to medium, small-medium and small manufacturing companies. As a matter of fact, we also chose twenty small-medium enterprises (SMEs) to carry out the same kind of study. These SMEs belong to Automotive, Industrial Equipment and Construction Equipment industries. In particular, they are all tier 1 or tier 2 suppliers of original equipment manufacturers of the same industries, and they are all located in Northern Italy, between the cities of Milan and Turin. Despite the limited size of the sample and the consequent limitation brought to the completeness of the study, we are still confident about the robustness of the Model and the fact that most manufacturing enterprises worldwide develop their products following a procedure that is similar if not identical to the one detailed in this study. Consequently, we believe the proposed Model to be scalable to manufacturing enterprises of every kind and size.

1.3.2 Use-Case Validation

The first method used for testing the Model consists of drawing a comparison between our Model and the implementation model used by an ESE's vendor. It would be impossible and wrong for us to implement the Model in a real situation where a vendor has to sell an ESE to a new customer enterprise. It would be wrong because implementing a brand new model on a prospect customer proves neither the efficacy of the model nor the eventual added value of such a procedure. There would be too many variables at stake and too many of them would be out of our control. Then, it would be impossible because it would take too much time to see the results. A sales process may last over a year, because it requires detailed background knowledge of the customer, deep study of the present infrastructure and a high degree of bureaucracy. Also, no software vendor would implement a model that comes straight out of the theory. As a consequence, we opted for a simple comparison between two procedures. To do so, we have selected COMAU as a use-case. COMAU has already installed the ESE named "Exalead" from the software vendor DS, thus representing an ideal situation to test our Model. The methodology consists of performing a gap analysis between the value proposition process of a best in class vendor such as DS and the one we would conduct following our Model. The purpose of this method is to test the Model in practice, to see how it fits a real business situation.

COMAU AS-IS

The mapping of COMAU's PD process, the situation AS-IS and the identification of COMAU's needs have been figured out through a public event presentation held by COMAU's IT Manager. It was impossible to conduct a structured interview with COMAU's manager or to ask any detail information about the PD process. Therefore, we solely rely on the material provided by COMAU, the customer analysis by DS and the previous study by Gaudino (2005) [69]. Regarding the installation performed by DS, we interviewed Dassault Systèmes' Sales Representative for the product Exalead. In this case, we decided to perform an unstructured interview. As a matter of fact, the aim of the interview is not to have an opinion or an evaluation but to obtain a description of the actual situation. We therefore decided to let the interviewee report its experience with no bounds or constrains. In particular, our aim was to identify the real needs of COMAU's people who are involved in the PD process according to the analysis by DS. It was also important to avoid influencing the interviewee with personal evaluations circa our Model and the gap analysis between the two procedures. So, we performed the interview without disclosing the Model to DS' sales representative.

Gap Analysis

The unit we use to measure the gap between the DS' procedure and our Model is the number of modules (also known as vertical market software frame) that the approach can detect and propose to fulfill COMAU's requirements. The effective number of modules is calculated by counting the number of people it would be provided to.

1.3.3 Experts Validation: Altea

The second method chosen for testing the Model consists of interviewing a team of experts from a software reseller, consultant and implementer.

We decided to conduct a semi-structured interview for several reasons. First, in a semi-structured interview, interviewers begin with a small set of open-ended questions, but spend considerable time probing participant responses, encouraging them to provide detail and clarification; these data are generally analyzed qualitatively [63]. This is a fundamental point for our research, since we need the interviewees not to give out a digit or a schematic response, but rather to argument their answers, discuss in detail the Model and express their true opinion about it. In fact, we considered answering to a structured questionnaire to be a limitation to the real value of an expert's opinion. Indeed, questionnaire research can be seen as over-reliant on instruments and, thus, disconnected from everyday life, with measurement processes creating a spurious or artificial sense of accuracy [63]. In other words, questionnaires are more suitable for quantitative research that requires large numbers of interviewees and a small number of possible answers. In our case, we do not actually know what kind of answer the experts may give, therefore we wanted them to be as free as possible to express and detail their judgment, although this could mean exposing our Model to criticism. Actually, the purpose of the interview is also to expose the Model to other flaws that we have failed to detect by ourselves. Also, performing a semi-structured interview helped add value not only to the Model but also to the whole study. As a matter of fact, the people interviewed mentioned their own personal experience in the field and specific cases of enterprises where they performed an analysis similar to the one of the Model. To our view, this is the final proof that a semi-structured interview is indeed the best way to develop an analytical analysis of the Model, but also to enrich the whole study with interesting ideas.

The company Altea was chosen because of its long time experience and reputation in the field of enterprise search, and particularly in the field of enterprise software solutions for data and process management. We selected one sales and one technical-sales business figure because we want the Model to be tested by people who are directly responsible for the real application of the software. Indeed, Altea's experts are every day on the job, and they experience sales and implementation procedures to win real business challenges. Most of all, they have been among the first in Italy to propose the value of enterprise search technologies to their customers in the field of manufacturing. Also, Altea's sales and technical sale people commit themselves to the product because they are rewarded according to their sales and consulting performances.

Interview

The questions are organized to test not only the Model as a whole, but also its internal structure and even its single STEPs. In particular, according to the main theories about semi-structured interviews, we encouraged the interviewee to follow his own logic and even to digress if he thinks it may provide a better understanding of the topic. We then take full responsibility of separating the digression from the real answers to the questions.

The interview has been performed in Italian, due to the location of Altea and the mother tongue of both interviewees. The answers are translated in English and transcribed under each question.

The interviewees are asked the following questions:

1) *Do you think the management of a manufacturing enterprise would be ready to consider the value of an Enterprise Search Engine in all of its aspects, and not just as a detached vertical market solution?*

This question addresses the theory and assumptions behind the model. We want the experts' opinion about what their customers think of ESEs, but we have the necessity to check if our assumptions about ESEs' technology and vertical market solutions are right. The answer is thought to be a general disclaimer about the model's validity. In fact, according to the literature, ESEs have indeed more value as a whole than as an assembly of separated models. Does this assumption match the reality? What does the expert think about it? And then, does he find practical proof of it in his daily job?

2) *How useful do you think it is to have such a standard procedure to carry on the selling process of an Enterprise Search Engine?*

This question addresses the methodology behind the Model. We want the experts to express their opinion about the idea of building up a standard procedure to help sales and tech-sales in their job. Obviously, the literature already puts forward standard procedures, but it is also true that all those procedures have been developed in a different context. We need the experts to evaluate not only the concept behind the Model but also the structure itself, to check if such a procedure could be accepted by a software vendor and adopted as best practice.

3) Do you think the STEPs in this order focus on the right target to address the selling process?

This question addresses the relation between the Model's structure and its effectiveness. We wanted the experts' opinion about the priorities we gave in order to achieve the purpose of the Model. We decided that the procure must start from the people (JRs) and the people's organization (BUs); their activities in the process (BAH Stage) and their needs (modules) come afterwards. Is this a good way to address a customer enterprise? Is this the most effective way to work out a consulting and selling process? According one's personal experience, does the Model get what is important for the customer enterprise?

4) Do you think the KPIs are suitable to convince a potential customer enterprise?

As we mention in paragraph 6.4.3, the KPI are the key instrument to convince the management to adopt a solution. Anyway, these KPIs are not the only ones existing and the may not be the most suitable. An expert can give an opinion about the selected KPIs and maybe put forward some of the ones he uses. In general, the expert can describe his personal choice in terms of performance evaluation and state his reasons.

5) Would you apply this model when dealing with a prospect customer?

This question is a sort of conclusion of the interview. It is also the final assessment of the Model, because the expert has to decide whether to adopt the procedure or not. In fact, this question is used to let the expert describe not just if the Model is generally acceptable or generally wrong, but to point out how good or how bad it is, where it needs to be improved, its strengths and its weaknesses.

Asking open questions cannot be the only way to test the model. Although every answer can provide useful information, this is not enough for an analytical analysis. In order to give scientific and academic proof to the interview, we need to further analyze each answer according to three axes:

a = From 1 to 10, how good is the model in theory?

This axis gives a score of the theoretical aspect of the Model. In other words, it takes into consideration the opinion of the interviewee about how the Model is aligned with previous studies, the existing best practices of the industry and the scholars. We give a score on this axis basing on how close is the Model is to the ideal theoretical situation.

b = From1 to 10, how good is the model in practice?

This axis gives a score of the practical aspect of the Model. It takes into account the fact that the model has to be applied to a real industrial context that may strongly differ from the theory. When answering the question, the expert may report his past experience, which contributes to increase the score of the Model along this axis. Let apart whether the idea behind the Model is correct or not, can this actually work in real life?

c = From 1 to 10, how good will the model be in the future?

The last axis inspects the experts' opinion of the Model with an eye on the future. The experts have a deep knowledge of the history of ESEs and they experience every day how the world of ESEs goes. But then they also have opinions, views and predictions for the future. It is important to understand whether the Model is headed towards the future of ESEs or it is just a product of present knowledge and has no possible application in the future. According to the scores we perform the analysis as of paragraph 8.4 and we draw our conclusions.

It is important to stress that the interviewees are not asked to assign the scores by themselves. We want to keep the interview as personal and open as possible. We believe it is better to let the experts express their opinion about each topic and then construct the score by ourselves. The fact is, it is the score to be subject to the outcome of the interview, not vice versa. We need the score in order to perform the analytical study of certain aspects of the answers, but the value behind the answers goes beyond any score, and it has to be taken into account.

The purpose of this method is to have the Model tested by people who actually sell and implement ESEs. While in the use-case scenario we look for measuring the Model's performance, here we ask for an evaluation of likely users of the Model.

1.3.4 Experts Validation: Independent Expert

We decided to interview John Stark, president of John Stark Associates. This choice has been made because Stark is not actively involved in the sales process, so he has a more theoretical and academic view of the Model. In fact, both Dassault Systèmes (Chapter 7) and Altea (Chapter 8) are bound to contracts which force them to sell the ESE they own. Therefore, their opinion about the Model can be influenced by their eagerness to find more possible applications for ESE, and hence more licenses to sell. In addition to being fully independent in his judgment, Stark also works as a PLM consultant for severl world-class manufacturing enterprises and is therefore well aware of their needs in terms of both software infrastructure and problem solving.

This way, the validation of the Model is as complete as possible. We have the Model tested in a real manufacturing environment (COMAU, paragraph 1.3.2), by a vendor that is actively involved in the value proposition, sales and implementation processes (Altea, paragraph 1.3.3) and by an independent expert in the field of PLM.

1.4 Structure of the Work

This thesis is made of 8 Chapters, each dealing with a particular aspect of the research. The chapters can be grouped into three categories: Chapters 2, 3, 4 and 5 include the literature and material review, Chapters 6, 7 and 8 deal with the Model proposition and test, and Chapters 1 and 9 are respectively the introduction and the conclusion of the study. The content of each Chapter is briefly described as follows:

Chapter 1 is an introductive section in which to find a descriptive introduction to the study, the structure of the whole work and the research methodology, with an emphasis on the research background, the assumptions behind the Model and the method to test it.

Chapter 2 introduces the reader to the concept of Big Data and then declines the same concept to the specific field of manufacturing. The content of this Chapter consists of both academic literature review and practical studies/reports from the field.

Chapter 3 provides a description of the Product Development process for manufacturing enterprises as reported by the literature. Afterwards, the Chapter describes in detail the software systems that the companies use in order to support this process. In the end, there is a list of what the literature identifies as the most impellent needs of enterprises in terms of software solutions for the Product Development.

Chapter 4 contains an overview of enterprise search. The history, the technological background and the specific features for Product Development of Enterprise Search Engines are described in detail. Particular emphasis is given to the technical features to support product developers.

Chapter 5 provides an example of state-of-the-art Enterprise Search Engine available on the market. The Chapter details both the technical insight and the vertical market modules of this specific software solution.

Chapter 6 contains the Model proposition. Here the reader finds the purpose of the Model, the logic behind it and the literature gap it aims at filling. Then, the Chapter describes the full structure of the Model and lists its limitations.

Chapter 7 describes the first testing scenario for the Model. It contains the procedure, the gap analysis and the findings related to the COMAU use-case, where we compared the value proposition process of an Enterprise Search Engine vendor to the one of our Model.

Chapter 8 describes the second testing scenario for the Model. It contains the interviews with the experts in the field of enterprise search, a brief introduction of the company their work for, the evaluation of the answers and the findings.

Chapter 9 contains the conclusions from the tests in Chapter 7 and 8, and the conclusion of the whole study. It also contains some ideas for future research about enterprise search and for the improvement of the Model.

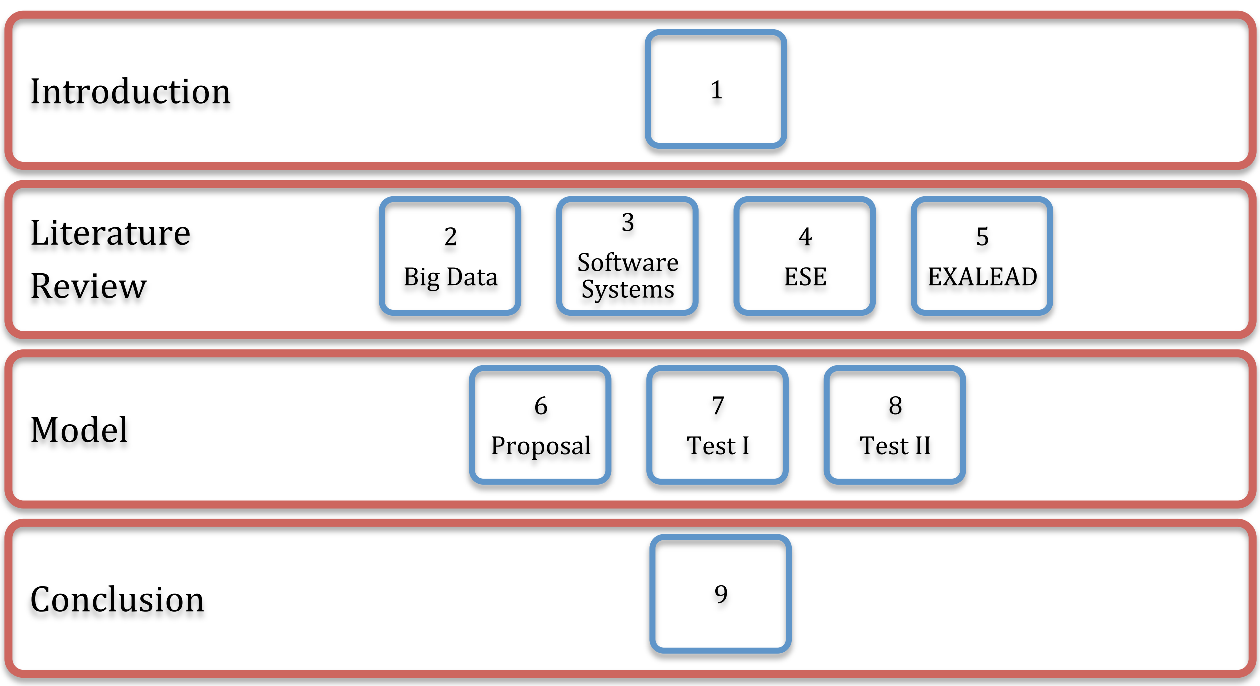


Figure 1.3: Structure of the Work

2 Overview of Big Data

2.1 What are Big Data?

2.1.1 Origin of Big Data

The term "Big Data" was first introduced to the computing world by Roger Magoulas from O'Reilly media in 2005 in order to define a great amount of data that traditional data management techniques cannot manage and process due to their complexity and size [18].

A study on the Evolution of Big Data as a Research and Scientific Topic shows that the term "Big Data" was present in research starting with the 1970s but was comprised in publications in 2008 [18]. In the distributed systems world, "Big Data" started to become a major issue in the late 1990s due to the impact of the worldwide Web and a resulting need to index and query its rapidly mushrooming content. The turn of the millennium brought further challenges as companies began to use information such as the topology of the Web and users' search histories in order to provide increasingly useful search results, as well as more effectively-targeted advertising to display alongside and fund those results [8]. Nowadays the Big Data concept is treated from different points of view covering its implications in many fields [18]. Actually, the term Big Data itself has entered abruptly into everyday language since the beginning of the years 2010s. It is even stated that Big Data continues to become increasingly important, even though the concept itself is at times confusing and under constant evolution [7]. The continuous development of the subject, gives way to countless definitions by several scholars and experts. The truth is, Big Data encompasses everything from click stream data from the web to genomic and proteomic data from biological research and medicines. Big Data is a heterogeneous mix of data both structured (traditional datasets in rows and columns like DBMS tables, CSV's and XLS's) and unstructured data like e-mail attachments, manuals, images, PDF documents, medical records such as x-rays, electrocardiogram and magnetic resonance images, forms, rich media like graphics, video and audio, contacts, forms and documents [8]. Such a broad category of items is the reason why Big Data are so hard to be defined clearly and synthetically.

2.1.2 5Vs classification model

While the potential benefits of Big Data are real and significant and some initial successes has already been achieved, many technical challenges still remain, and those must be addressed to fully realize this potential [66]. From a more practical point of view, Big Data is a collection of vast data sets with a great diversity of types so that it becomes difficult to process them by using state-of-the-art data processing approaches or traditional data processing platforms [2]. Indeed, the term "Big" derives from the fact that a given data set is composed by a huge amount of information. In particular, experts define the term "big" as the sum of five variables:

Volume: the amount of generated data representing the most classical measure unit for data. It is also common to talk about files dimensions and how much space a file takes in terms of kilo-, mega- or giga-bytes. But, when dealing with Big Data, data measurement is in tera-bytes (2^{40}) or even peta- bytes (2^{50}), and it is rapidly heading toward exabytes (2^{60}) [20].

Velocity: means both the rate at which data is generated [7] and the time in which Data must be acted upon [66] or, in other words, in which data occurs at very high [20] and which comes at a slow rate. Considering data velocity, it is to further complicate the matter, it is noticed that data can arrive and require processing at different speeds, as illustrated below:

- Batch: At time intervals
- Near-time: At small time intervals
- Real-time: Continuous input, process, output
- Streams: Data flows

While for some applications the arrival and processing of data can be performed in batch, other analytics applications require continuous and real-time analyses, sometimes requiring immediate action upon processing of incoming data streams. Increasingly often, data arriving via streams need to be analyzed and compared with historical information. Different data sources may use their own formats, a characteristic that makes it difficult to integrate data from multiple sources in an analytics solution. As highlighted in existing work, standard formats and interfaces are crucial so that solution providers can benefit from economies of scale derived from data integration capabilities that address the needs of a wide range of customers [19].

Variety: means the heterogeneity of data types, representation, and semantic interpretation [66]. In other words, the different formats in which data can appear [7]. Handling and analyzing these data poses several challenges as they can be of different types:

- Structured: Formal schema and data models
- Unstructured: No pre-defined data model
- Semi-structured: Lacks strict data model structure
- Mixed: Various types together

It is argued that a large part of data produced today is either unstructured or semi-structured [19].

While these three are important, this short list fails to include additional important requirements such as privacy and usability [66], that are included in the last two Vs.

Veracity: the quality of captured data. As a matter of fact, due to intermediary processing, diversity among data sources and in data evolution raises concerns about security, privacy, trust, and accountability, creating a need to verify secure data provenance [20]. Also, data must be accurate in order to have value if not meaning at all. Big Data also present noise, impurities especially when extracted from the Internet or from large databases that are not set for queries.

Variability: the inconsistency that data can show, due to the fact that data meaning is constantly changing. In particular, data can be extremely variable when their meaning is linked to language, semantics and language processing. Words have different meaning depending on the context and the time period. For example, searching for the word "tweet" on any popular Web search engine shows very different results if the search is performed in the late 1990s rather than in 2015.

2.1.3 Big Data Analysis

The above-listed definitions are the so-called 5V dimensions of Big Data. Depending on the subject, some dimensions may have greater importance compared to the others but still, when dealing with Big Data, all of them contribute to the challenge of managing data of such a kind. In fact, given the complexity of the topic, the literature has shown several important challenges to be faced by scholars and enterprises. Unfortunately, many people just focus on the analysis/modeling phase: although that phase is crucial, it is of little use without the other phases of the data analysis pipeline. Even in the analysis phase, which has received much attention, there are poorly understood complexities in the context of multi-tenanted clusters where several users' programs run concurrently. As shown in the picture below, many significant challenges extend beyond the analysis phase [66]:

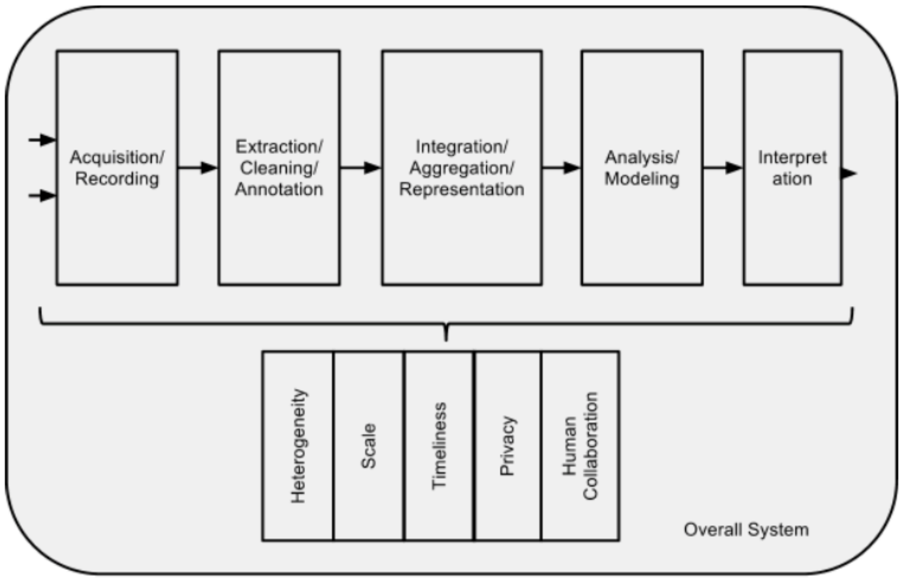


Figure 2.1: The Big Data Analysis Pipeline

The flow of boxes on the top shows the major steps in the analysis of Big Data, whereas the lower boxes point out the requirements that need to be faced when addressing the above steps. The analysis in explained in detail by Agrawal (2011) [66] and Assunção (2015) [19].

<u>*Data Acquisition and Recording*</u>

Big Data does not arise out of a vacuum: it is recorded from some data-generating source.

Much of this data is of no interest, and it can be filtered and compressed by orders of magnitude. One challenge is to define these filters in such a way that they do not discard useful information. In addition, the data collected by these sensors are most often spatially and temporally correlated. We need research in the science of data reduction that can intelligently process this raw data to a size its users can handle while not missing the needle in the haystack. Furthermore, we require "on-line" analysis techniques that can process such streaming data on the fly, since we cannot afford to store first and reduce afterward.

The second big challenge is to automatically generate the right metadata to describe what data is recorded and how it is recorded and measured. Metadata acquisition systems can minimize the human burden in recording metadata. Another important issue here is data provenance. Recording information about the data at its birth is not useful unless this information can be interpreted and carried along through the data analysis pipeline. We thus need research both into generating suitable metadata and into data systems that carry the provenance of data and its metadata through data analysis pipelines.

Information Extraction and Cleaning

Frequently, the information collected will not be in a format ready for analysis. For example, consider the collection of electronic health records in a hospital, comprising transcribed dictations from several physicians, structured data from sensors and measurements, and image data such as x-rays. We cannot leave the data in this form and still effectively analyze it. Rather we require an information extraction process pulling out the required information from the underlying sources and expressing it in a structured form suitable for analysis. Doing this correctly and completely is a continuing technical challenge. Note that Big Data also includes images and will in the future include videos; such extraction is often highly application dependent. In addition, due to the ubiquity of surveillance cameras and popularity of GPS-enabled mobile phones, cameras, and other portable devices, rich and high fidelity location and trajectory data can also be extracted. Existing work on data cleaning assumes well-recognized constraints on valid data or well-understood error models; for many emerging Big Data domains these do not exist.

Data Integration, Aggregation, and Representation

How to aggregate and correlate streaming data from multiple sources?

Given the heterogeneity of the flood of data, it insufficient to merely record it and throw it into a repository. Consider, for example, data from a range of scientific experiments. If we just have a bunch of data sets in a repository, it is unlikely anyone will ever be able to find, let alone reuse, any of this data. With adequate metadata, there is some hope, but even so, challenges will remain due to differences in experimental details and in data record structure.

Data analysis is considerably more challenging than simply locating, identifying, understanding, and citing data. For effective large-scale analysis all of this has to happen in a completely automated manner. This requires differences in data structure and semantics to be expressed in forms that are computer understandable, and then "robotically" resolvable. There is a strong body of work in data integration that can provide some of the answers. However, considerable additional work is required to achieve automated error-free difference resolution.

Even for simpler analyses that depend on only one data set, there still remains an important question of suitable database design. Usually, there will be many alternative ways to store the same information. Some designs will have advantages over others for certain purposes, and possibly drawbacks for other purposes. Database design is today an art, and is carefully executed in the enterprise context by highly paid professionals.

Data Modeling and Analysis

Methods for querying and mining Big Data are fundamentally different from traditional statistical analysis on small samples. Big Data is often noisy, dynamic, heterogeneous, inter-related and untrustworthy. Nevertheless, even noisy Big Data could be more valuable than tiny samples because general statistics obtained from frequent patterns and correlation analysis usually overpower individual fluctuations and often disclose more reliable hidden patterns and knowledge. Further, interconnected Big Data forms large heterogeneous information networks, with which information redundancy can be explored to compensate for missing data, to crosscheck conflicting cases, to validate trustworthy relationships, to disclose inherent clusters, and to uncover hidden relationships and models.

Mining requires integrated, cleaned, trustworthy, and efficiently accessible data, declarative query and mining interfaces, scalable mining algorithms, and Big Data computing environments. At the same time, data mining itself can also be used to help improve the quality and trustworthiness of the data, understand its semantics, and provide intelligent querying functions. On the flip side, knowledge developed from data can help in correcting errors and removing ambiguity.

Big Data is also enabling the next generation of interactive data analysis with real-time answers.

A problem with current Big Data analysis is the lack of coordination between database systems, which host the data and provide SQL querying, with analytics packages performing various forms of non-SQL processing, such as data mining and statistical analyses. Nowadays analysts are impeded by the schematic process of exporting data from the database, performing a non-SQL process and bringing the data back. This is an obstacle to carrying over the interactive elegance of the first generation of SQL-driven Online Analytical Processing (OLAP) systems into the data mining type of analysis that is in increasing demand.

Interpretation

Having the ability to analyze Big Data is of limited value if users cannot understand the analysis. Ultimately, a decision maker, provided with the result of analysis, has to interpret these results. This interpretation cannot happen in a vacuum. Usually, it involves examining all the assumptions made and retracing the analysis. Furthermore, as we saw above, there are many possible sources of error: computer systems can have bugs, models almost always have assumptions, and results can be based on erroneous data. For all of these reasons, no responsible user will cede authority to the computer system. This is a particular a challenge within Big Data due to its complexity. There are often crucial assumptions behind the recorded data. Analytical pipelines can often involve multiple steps, again with built in assumptions.

In short, it is rarely enough to provide just the results. Rather, one must provide supplementary information that explains how each result was derived, and based upon precisely what inputs. Such supplementary information is called the provenance of (result) data. By studying how best to capture, store, and query provenance, in conjunction with techniques to capture adequate metadata, we can create an infrastructure to provide users with the ability to both interpret the analytical results obtained and repeat the analysis with different assumptions, parameters, or data sets.

Furthermore, with a few clicks the user should be able to drill down into each piece of data that he sees and to understand its provenance, which is a key feature to understanding the data. This means that users need to be able to not just see the results, but also to understand why they are seeing them.

2.1.4 Challenges in Big Data Analysis

Having described the multiple phases in the Big Data analysis pipeline, the study now turns to some common challenges that underlie these phases. These are the ones listed in the five boxes at the bottom of Figure 2.1.

Heterogeneity and Incompleteness

When humans consume information, a great deal of heterogeneity is comfortably tolerated. In fact, the nuance and richness of natural language can provide valuable depth. However, machine analysis algorithms expect homogeneous data, and cannot understand the concept of nuance. So, especially when the data is unstructured, how to quickly extract meaningful content out of it? The only answer is: data must be carefully structured prior to data analysis. Greater structure is likely to be required by many (traditional) data analysis systems. In addition to this, computer systems work most efficiently if they can store multiple items that are all identical in size and structure. Efficient representation, access, and analysis of semi-structured data require further work.

Some incompleteness and some errors in data are likely to remain even after data cleaning and error corrections. Incompleteness and errors must be managed during data analysis.

Scale

Of course, the first thing coming out of one's mind with Big Data is their size. After all, the word "big" is there in the very name. How to efficiently recognize and store important information extracted from unstructured data? How to store large volumes of information in a way it can be timely retrieved? Are current file systems optimized for the volume and variety demanded by analytics applications?

Managing large and rapidly increasing volumes of data has been a challenging issue for many decades. In the past, this challenge was mitigated by processors getting faster, following Moore's law, to provide us with the resources needed to cope with increasing volumes of data. Yet, a fundamental shift is now underway: data volume is scaling faster than computer resources, and CPU speeds are static.

First, over the last five years the processor technology has made a dramatic shift: rather than processors doubling their clock cycle frequency every 18-24 months, now, due to power constraints, clock speeds have largely stalled and processors are being built with increasing numbers of cores. In the past, large data processing systems had to worry about parallelism across nodes in a cluster; now, they have to deal with parallelism within a single node. Unfortunately, parallel data processing techniques that were applied in the past for processing data across nodes don't directly apply to intra-node parallelism, because the architecture looks very different; for example, there are many more hardware resources such as processor caches and processor memory channels that are shared across cores in a single node. These unprecedented changes require us to rethink how we design, build and operate data processing components.

The second dramatic shift that is underway is the move towards cloud computing, which now aggregates multiple disparate workloads with varying performance goals into very large clusters. This level of sharing resources in expensive and large clusters requires new ways of determining how to run and execute data processing jobs so that the goals of each one can be cost-effectively met, and to deal with system failures, which occur more frequently as we operate on larger and larger clusters, that are required to deal with the rapid growth in data volumes. This places a premium on declarative approaches to expressing programs, even those doing complex machine learning tasks, since global optimization across multiple users' programs is necessary for good overall performance. Since users are unaware of other users' programs, reliance on user-driven program optimizations is likely to lead to poor cluster utilization. System-driven holistic optimization requires programs to be sufficiently transparent.

A third dramatic shift that is underway is the transformative change of the traditional I/O subsystem. How to handle an ever-increasing volume of data? For many decades, hard disk drivers (HDDs) have been used to store persistent data. HDDs had far slower random I/O performance than sequential I/O performance, and data processing engines formatted their data and designed their query processing methods to "work around" this limitation. But today HDDs are being increasingly replaced by solid-state drives, and other technologies such as Phase Change Memory are around the corner. These newer storage technologies do not have the same large spread in performance between the sequential and random I/O performance, which requires a rethinking of how we design storage subsystems for data processing systems. Implications of this changing storage subsystem potentially touch every aspect of data processing, including query processing algorithms, query scheduling, database design, concurrency control methods and recovery methods.

Timeliness

In a sort of Big Data trade-off, an increase in size means a reduction of speed. The larger the data set to be processed, the longer it will take to analyze them. The design of a system that effectively deals with size is also likely to result in a system that can process a given size of data set faster.

There are many situations in which the result of the analysis is required immediately. For example, if a fraudulent credit card transaction is suspected, it should ideally be flagged before the transaction is completed, thus potentially preventing the transaction from taking place at all.

Given a large data set, it is often necessary to find elements in it that meet a specified criterion. In the course of data analysis, this sort of search is likely to occur repeatedly. Scanning the entire data set to find suitable elements is obviously impractical. Rather, index structures are created in advance to allow the quick disclosure of qualifying elements. The problem is that each index structure is designed to support only some classes of criteria. With the new analyses performed using Big Data, there are new types of specified criteria, and a need to devise new index structures to support such criteria. Designing such structures becomes particularly challenging when the data volume is growing rapidly and the queries have tight response time limits.

Privacy

The privacy of data is another major concern, and one that increases in the context of Big Data. Managing privacy is in fact both a technical and a sociological problem, which must be jointly addressed from both perspectives to realize the promise of Big Data.

For example, banks need to protect their customers from losing information about their online accounts, governments cannot afford the leak of secret information and private companies do not want their competitors to steal their virtual projects from the company's repositories.

There are a number of additional challenging research problems. For example, we do not know yet how to share private data while limiting disclosure and ensuring sufficient data utility in the shared data. One of the most critical aspects is that today many online services require us to share private information, but beyond record-level access control we do not understand what sharing data means, how the shared data can be linked, and how to give users fine-grained control over this sharing.

Human Collaboration

In spite of the tremendous advances made in computational analysis, there are still many patterns that are easily detectable by humans but which computer algorithms still have a hard time finding. Ideally, analytics for Big Data will not be all computational, but they will rather be designed explicitly to have a human in the loop. The new sub-field of visual analytics is attempting to do this, at least with respect to the modeling and analysis phase in the pipeline. There is similar value to human input at all stages of the analysis pipeline.

In today's complex world, it often takes multiple experts from different domains to really understand what is going on. Big Data analysis systems must support input from multiple human experts and shared exploration of results. These multiple experts may be separated in space and time when it is too expensive to assemble an entire team together in one room. The data system has to accept this distributed expertise's input, and support their collaboration.

The extra challenge within this topic is the inherent uncertainty of data collection devices, such as RFID tags, smartphones and so on. The fact that collected data are probably spatially and temporally correlated can be exploited to better assess their correctness.

Many more open questions surround the world of Big Data and for sure the number of questions will keep on rising. In order to address a specific field of competence to part of them, this study concentrates its efforts on Big Data in the world of Product Development.

2.2 Big Data in Manufacturing Enterprises

As this study focuses on Big Data in the field of manufacturing, the broad Big Data concept has to be narrowed down to this specific field. In particular, the 5Vs have certain characteristics to be taken into consideration:

2.2.1 5Vs of Data for Manufacturing

Volume: typical data shared inside manufacturing enterprises are the ones related to the design and engineering of products. The volume of data will depend on the characteristics of the product to be manufactured. In fact, depending on the complexity of the design, the files exchanged inside the enterprise can have volume up to giga-bytes.

The sheer size of the data, of course, is a major challenge, and is the one that is most easily recognized. However, there are others. Industry analysis companies like to point out that there are challenges not just in Volume, but also in Variety and Velocity, and that companies should not focus on just the first of these [66].

Velocity: the rate at which data are generated depends not only on the speed of designers, engineers, marketers and so on, but also on the frequency of data change and exchange inside the enterprise. Design and production are often collaborative activities that require the contribution of people from various areas of the enterprise. Besides, manufactured products are eventually subject to revisions and versioning. Therefore, data are generated but also mutated at high speed inside modern manufacturing enterprises.

Variety: as the design of a product requires the contribution of different people with different skills, it also requires the use of different tools. In Chapter 3, this study explains the different tools used inside manufacturing enterprises and the different ways to cope with the different file formats.

Veracity: Big Data in the manufacturing business are not made to be static. As explained before, the manufacturing process needs data to pass throughout the enterprise. Therefore, veracity of data is a key issue. Data need to be protected from external threats, as competitors may want to access the enterprise's know how or take advantage from a leak of information about, for example, the technical specifications of the latest high tech component. At the same time, data also need protection from internal threats. As a matter of fact, not all data users are equal and not all of them have the same level of access to information. Big Data need to be stored in safe repositories and access has to be checked via the implementation of security protocols. The risk of granting full access to all users is that data can get compromised at a certain point of the development or manufacturing process, thus compromising the final product.

Variability: data can have a different meaning according to the person that is accessing them. For example, the chief engineer receives in input the bill of materials (BOM) and uses it to set up the design process of the product, taking note of the parts, components and specification needed. On the other hand, the production manager uses the BOM to figure out what kind of machines should be used to manufacture the product, what tools have to be purchased and how to allocate

the manpower along the production shifts. In other words, the same document can be read in two different ways. In addition to this, we need to consider that the greater amount of information Big Data carry within, the higher number of different interpretations they offer.
One more aspect has to be considered. This study is fully aware of things such as cross-enterprise collaborative PD, supplier-to-vendor chain, enterprises collaboration etc. These are all sources of Big Data standing out of the enterprise's perimeter, thus implying that Data can be compromised in terms of veracity and variability. This study will show later on in Chapter 3 how the enterprise takes care of those data and what specific software system is implemented to ensure safety and trust.

The purpose of this study also creates the need for a proper definition of Big Data. Though quite broad, this research adopts the definition given by N. Silver, who refers to Big Data as "Digitized data created by people, machines, sensors, tools, and other mechanisms. This data acts like an audit trail or data footprint, telling a story about the events and interactions of humans, machines, markets, natural systems, and other entities" [1]. The key element in this definition is the interaction between humans and machines. The study already explained that Big Data is a heterogeneous mix of both structured and unstructured data like e-mail attachments, manuals, images, PDF documents, medical records such as x-rays, echocardiogram and magnetic resonance images, forms, rich media like graphics, video and audio, contacts, forms and documents [8]. We need to add to the list all of the data designers and engineers use to perform their job, all of the recording and tracking data and the instruments to control the entire process. The tools used to manipulate, access, analyze and present Big Data are also in continuous evolution. Design of data marts and tools for extraction, transformation, and load (ETL) are essential for converting and integrating enterprise-specific data. Database query, OLAP, and reporting tools based on intuitive, but simple, graphics are used to explore important data characteristics [17]. There is a wide category of software systems for data management that have become an essential part of manufacturing enterprises. These software systems are called Enterprise Resource Planning systems, or simply ERP.

2.2.2 Use of Big Data in Manufacturing

Manufacturing stores more data than any other sector: in 2010 the amount of new data stored was already close to 2 exabytes [65]. Nedelcu (2013) [65] developed an exhaustive study about data for manufacturing.

The manufacturing sector has been an early and intensive user of data to drive quality and efficiency, adopting information technology and automation to design, build, and distribute products since the dawn of the computer era. In the 1990s, manufacturing companies racked up impressive annual productivity gains because of both operational improvements that increased the efficiency of their manufacturing processes and improvements in the quality of products they manufactured. The manufacturing sector has been the backbone of many developed economies. Increasingly global and fragmented manufacturing value chains create new challenges that manufacturers must overcome to sustain productivity growth. In many cases, technological change and globalization have allowed countries to specialize in specific stages of the production process. In order to continue achieving high levels of productivity growth, manufacturers will need to leverage large datasets to drive efficiency across the extended enterprise and to design and market higher-quality products. Manufacturing companies will also have to build the capabilities needed to manage Big Data. Despite the fact that the sector has been dealing with large datasets for two decades, the rising volume of data from new sources along the supply chain and from end markets requires a new level of storage and computing power and deep analytical expertise if manufacturers are to harvest relevant information and insights.

But what kind of data does manufacturing have do deal with? How does this sector manage the data flow? What are the specific needs for a manufacturing enterprise in terms of Big Data?

First of all, after explaining the 5V model (paragraph 2.2.1), there is a need for further data classification.

Based on the dimension of data structure, we have:

- Structured data: all data from fixed fields such as spreadsheets and relational databases.
- Unstructured data: all data that do not reside in fixed fields, e.g. text from articles, email messages, untagged audio or video data and so on.
- Semi-structured data: the data that do not reside in fixed fields but use tags or other markers to capture elements of the data e.g. XML or HTML-tagged text.

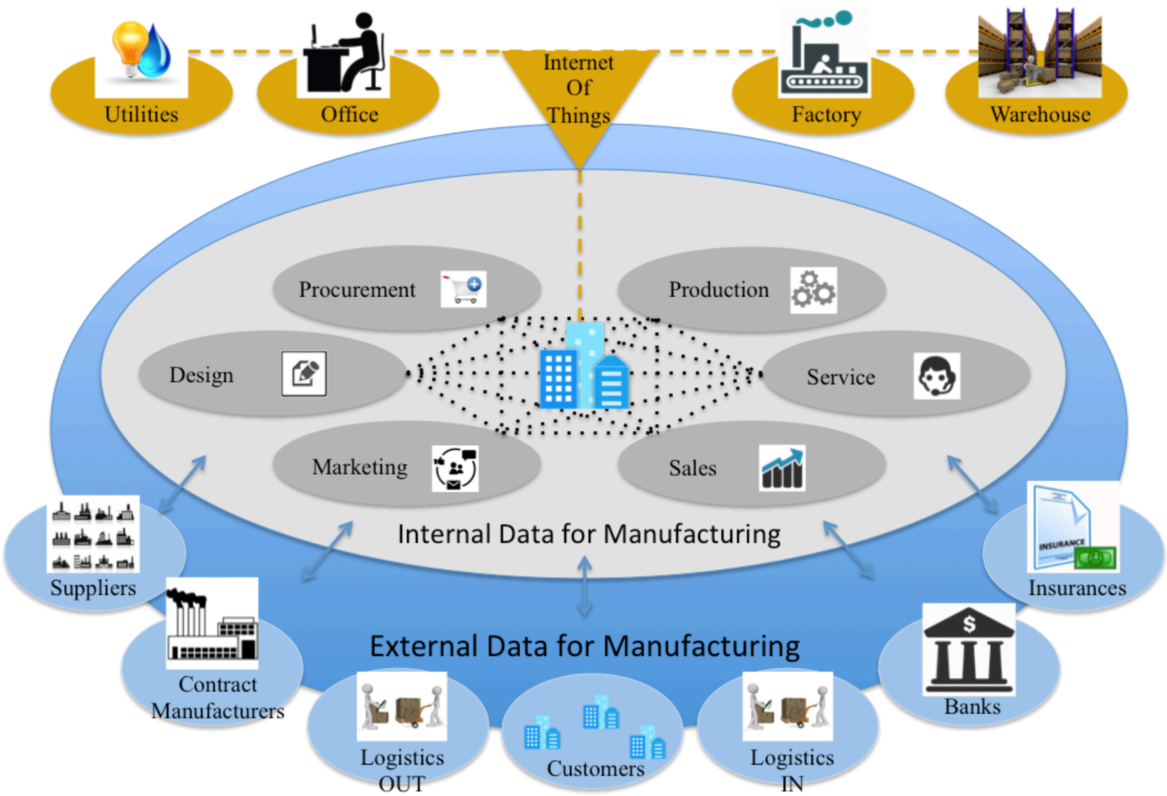


Figure 2.2: All data manufacturing enterprises deal with

Based on the location of data sources, we distinguish between external and internal data.

2.2.3 External data for Manufacturing

Every input of the company is a generator of external data. As we show in Figure 2.2, data come from sources that are directly connected to the manufacturing activities, such as the shipment of components and raw materials, but also from activities that support the enterprise as an organization, like banks and insurance companies. Here follows a list of the main sources of external data that are peculiar of manufacturing enterprises.

Raw Materials

The raw materials come from a supplier along with their documents and descriptions (extraction site, extraction date, chemical composition etc.). Checking for the veracity of input data is a key concern for manufacturing enterprises: from the moment the raw materials enter the value chain, the data they carry with them will also be accessed and processed.

Supply Chain Systems

The supply chain management systems manage all the movements of parts and assemblies coming from the outside (name of suppliers, lot number, part quantities, specifications, delivery status, etc.). Supply chain manufacturers, especially those producing fast-moving consumer goods, have significant additional opportunities to improve demand forecasting and supply chain planning. The volatility of demand has been a critical issue for manufacturers. Their retailing customers have pushed hard for increased flexibility and responsiveness from suppliers, given the diverging and ever-changing consumer preference. Other trends, such as the increasing use of promotions and tactical pricing, have only magnified the volatility issues suppliers will have to face. Manufacturers can improve their demand forecasting and supply planning by a better use of their own data. But far more value can be unlocked when companies are able to integrate data from other sources including data from retailers, such as promotion data (items, prices, sales...), launch data (specific items to be listed/delisted, ramp-up/ramp- down plans...) and inventory data (stock levels per warehouse, sales per store...). By taking into account data from across the value chain (potentially through collaborative supply chain management and planning), manufacturers can smooth spiky order patterns. The benefits of doing so will ripple through the value chain, helping manufacturers to use cash more effectively and to deliver a higher level of service.

Company – Customers Interaction

While obtaining customer input through market research has traditionally been a part of the product design process, many manufacturers have yet to systematically extract crucial insights from the increasing volume of customer data to refine existing designs and help develop specifications for new models and variants. Best-in-class manufacturers conduct joint analyses to determine how much customers are willing to pay for certain features and to understand which features are the most important in terms of market success. Manufacturing companies are using data from customer interactions not only to improve marketing and sales but also to drive product development decisions. Increasingly, it is economically feasible to embed sensors in products that can "phone home," generating data about actual product usage and performance. For example, thanks to systems that monitor the performance of products that have already been sold, Boeing knows that during a single cross-country flight, a Boeing 737 generates 240 terabytes of data. Manufacturers can now obtain real-time input on emerging defects and adjust the production process immediately. In addition, we need to consider that nowadays most of the companies' products are made according to an "engineer to order" logic. Understanding the behavior of repeat customers is critical to delivering in a timely and profitable manner. Most of its profitability analysis is to make sure that the company has good contracts in place. The company says its adoption of analytics has facilitated its shift to lean manufacturing, and has helped it determine which products and processes should be scrapped.

Marketing, Sales, After-Sales

There are many opportunities to leverage large datasets in the marketing, sales, and after-sales service activities. Opportunities range from the segmentation of customers to applying analytics in order to improve the effectiveness of sales forces. An increasingly important application for manufacturers is using sensor data from products once they are in use to improve service offerings. Networks can even connect physical products, enabling those products to report their own serial numbers, ship dates, number of times used etc. Thanks to these networks, a lot of companies have now access to valuable pools of data generated by their products and services.

Networking, Collaboration

The use of Big Data offers further opportunities to accelerate product development, help designers home in on the most important and valuable features based on concrete customer inputs as well as designs that minimize production costs, and harness consumer insights to reduce development costs through approaches including open innovation. In order to drive innovation and develop products that address emerging customer needs, manufacturers are relying increasingly on outside inputs through innovative channels. With the advent of Web 2.0, some manufacturers are inviting external stakeholders to submit ideas for innovations or even collaborate on product development via Web-based platforms. Also, all data gathered from sources outside a company such as third-party data providers or public social media sites are to be considered external data for innovation and customer needs' awareness.

2.2.4 Internal Data for Manufacturing

The core of manufacturing is still within the R&D function, where the real product is created, and in the Production Department, where the creation is brought to life (and to the market) through the operations performed by machines. These are part of transfer and assembly lines, flexible manufacturing cells, job shops and so on. As a matter of fact, this is where data have been bringing significant advantages for ages, and this is also the area where the experts are convinced that Big Data can play the biggest role. In other words, it is from product design, instrumented production machinery and process control that manufacturing enterprises gather the most valuable Big Data.

Usually, the production units of manufacturing enterprises manage their internal data with a structure similar to the one in Figure 2.3 below.

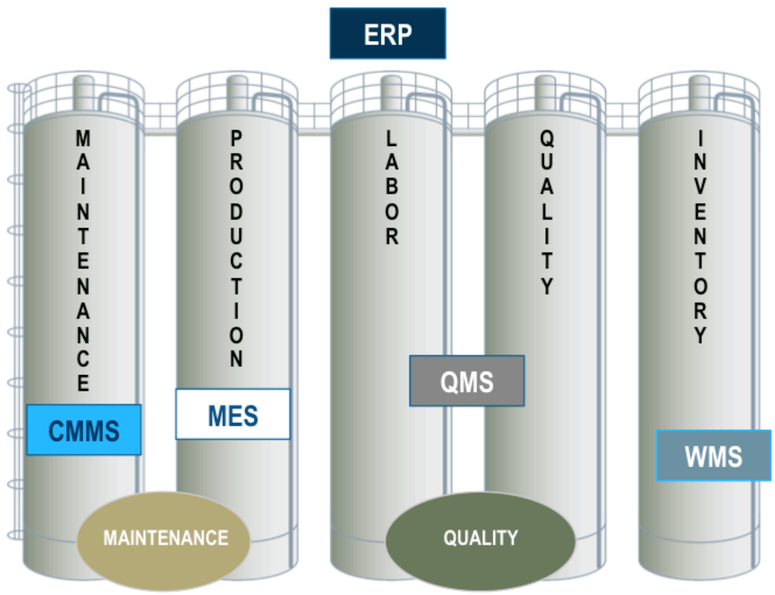


Figure 2.3: Software systems for managing internal Big Data

Both scholars and production managers believe the greatest opportunities of Big Data for their function are those to detect product defects, to boost quality and to improve supply planning. Better detection of defects in the manufacturing/production processes is next on the list.

When asked for their main requirements and top priorities, production managers point out:

- **Visibility:** into manufacturing and product supply operations, tightly linking plants to the enterprise.
- **Control:** over manufacturing operations through directed manufacturing, actionable dashboards, enforcement of best manufacturing practices and error proofing.
- **Synchronization:** across the customer's enterprise-wide production, quality, warehousing

and maintenance activities within their plants, warehouses and suppliers.

But then, there are several problems that come out when trying to achieve these goals.

First, it is tough to collect detailed product, genealogy and event data: products start being something concrete when they come out of the R&D department. If the R&D provides scattered and unstructured data or worse, if it does not keep track of the information, there is no method, procedure or system the production can use to monitor its own process and keep track of the product requirements.

Second, it is difficult to add multi-plant data from disparate systems into a single, global repository. As we can see from Figure 2.2, enterprises often take advantage of external suppliers and manufacturers; but, it is also common for them to split the production among several production plants that are actually located in different geographical areas. Many times it is also a matter of strategic acquisitions: an American corporation may purchase an existing production plant in Asia, and this plant may implement hardware and software solutions that are totally different from the ones the company uses in the other plants.

Third, a heterogeneous IT systems environment adds complexity. As we can see from the picture, the whole area dedicated to maintenance is managed with a Computerized Maintenance Management System (CMMS), whereas a Warehouse Management Systems (WMS) takes care of the logistics, a Quality Management System (QMS) checks whether the products and the working environment meet the standards and, finally, the Manufacturing Execution System (MES) keeps track of all the machine operations inside the factory. All of these software systems may be included in a larger ERP system, but they can also be stand-alone systems that require some kind of connector to work together. In addition, the environment described so far needs to interact with the R&D function, because it is from there that the main input for the production comes from.

To sum up, this Chapter makes it clear that manufacturing enterprises have a need for a complex software architecture that is fully dedicated to the PD activities.

3 Big Data and Product Development

Every product out of any industry has its own lifecycle. It can be divided into three periods: *beginning of life* (BOL), *middle of life* (MOL), and *end of life* (EOL). BOL is the period in which a product concept is generated, designed, and subsequently physically realized. MOL is the period when products are distributed, used, and maintained by customers or engineers. EOL is the period when products are recycled by manufacturers or disposed of by customers [24].

While a product's lifecycle is and will always be like the one just described, on the other hand products are constantly changing and evolving. With the evolution of products, product data are changing and evolutional in their structures and attributes [24]. Depending on the product stage of the lifecycle, product data describe conditions, maintenance or repairing needs, success on the market, the number of times the product has been revised etc. Indeed, products carry with them a significant amount of data that are the key to successful product management. The specific science that studies products lifecycle is called Product Lifecycle Management. Product Lifecycle Management (PLM) is the business activity of managing a company's products in the most effective way across their whole lifecycles, from the very first idea of a product until it is retired and disposed of [3]. It can also be seen as a strategic business approach that applies a large set of business solutions in support of the collaborative creation, management, dissemination, and use of product definition information across the extended enterprise from concept to end of life, thus integrating people, processes, business systems and information [22]. This definition gives the idea of how vast and complicated the subject is. The necessity of connecting people, processes and information implies the use of complex data and tools to handle them. In addition, a deeper study of the product lifecycle is required to better understand where all of these data are located.

Apart from the traditional BOL-MOL-EOL structure, according to PLM products lifecycle can be divided into phases more suitable to the manufacturer's point of view: *design, production, logistics, utility, maintenance, and recycle*. In the design phase, an idea in the designer's head is transferred into a detailed description. Subsequently, a product in its final shape is obtained in the production phase. Later, the product is stored in a warehouse and then transported to customers in the logistics phase. In the phase of utility, the customer uses the product whereas the manufacturer provides the remote service. If anything goes wrong, the product enters into the maintenance phase, and if it can no longer be used, it comes to the end of its life like recycle or disposal [24]. Therefore, this subdivision is still somehow linked to the classical BOL-MOL-EOL model. By contrast, it is more suitable to this study. In fact, the subject of this study is the first phase, the very beginning of the Lifecycle: the design phase, also know as Product Development (PD).

3.1 Product Development

New products are central to the growth and prosperity of a modern and innovative organization [36], and PD is defined as the transformation of a market opportunity into a product available for sale [34]. Therefore, a PD process is the sequence of steps in which an enterprise designs and commercializes a product, beginning with the perception of a market opportunity down to the production, sale and delivery of a product [23].

In fact, a new product that is introduced on the market evolves over a sequence of stages, beginning with an initial product concept or idea that is evaluated, developed, tested and launched on the market. The new PD process differs from industry to industry and from firm to firm. Actually, it should be adapted to each firm in order to meet specific company resources and needs [33]. Nonetheless, it is possible to draw a general process that is suitable for every kind of firm. One of the first developed models for the new PD process is the Booz, Allen and Hamilton (BAH) Model, published in 1982 [15]. This is also the model that has been adopted for the purposes of this study because, despite being rather old and raw, it is still largely used by companies nowadays. Indeed, it shows lots of characteristics that make it still valid.

First, it is the best-known model, because it underlies the new PD systems that have been put forward later [33]. In other words, this model encompasses all of the basic stages of models found in the literature and it is particularly important because all the models that have been developed afterwards have been inspired by this one. Second, the BAH model is based on extensive surveys, in depth interviews and case studies and, as such, appears to be a fairly good representation of prevailing practices in industry [33].

Figure 3.1: Stages of new Product Development (Booz, Allen and Hamilton, 1982)

Third, this model shows an outline made of stages. Each stage involves a series of activities, but none of these are bound to a specific enterprise structure in terms of organizational chart. This means that the adoption of this model does not force the company to adopt a certain structure, which is an element this study will come back to in the following paragraphs.

The stages of the BAH Model are analyzed in detail by Bhuiyan (2011). Here the study presents a brief summary mainly inspired by the exhaustive scholar's work, explaining the stages of the new PD process and their implications.

3.1.1 Stage 1 - New Product Strategy

The purpose of this stage is to provide guidance for the new product effort. It identifies the strategic business requirements the new product should comply with, and these are derived from the corporate objectives and strategy of the firm as a whole.

The importance of an appropriate corporate environment for new PD has long been recognized [37]. In fact, research shows that a growing body of evidence points to new product strategy linked to corporate objectives and strategies, top management support and appropriate organizational structures as factors in successful new PD [37]. What emerges from the previous reports is that the push for a new product always starts from the top management of the company. There is also a vital link between the organization's culture and product innovation: the development of a new product is a core activity, therefore it has to involve the whole organization, starting from the top.

3.1.2 Stage 2 - Idea Generation

This stage is the most creative and the one requiring not only imagination, but also a deep knowledge of the market and, more in general, of the environment in which an enterprise operates.

Firms that are effective at idea generation do not focus solely on the first source to generate ideas, i.e. ideas that are originated from inside the firm, but that concentrate on all potential idea sources [40]. Indeed, there is a multitude of sources as well as many different methods to generate ideas. The firm can derive new ideas from internal sources (i.e., employees, managers), from external sources (i.e., customers, competitors, distributors, and suppliers), and from implementing formal research and development. Brainstorming, morphological, analysis and gap analysis are most commonly employed methods for generating ideas [40]. Customers can also be an especially good place to start searching for new product ideas. A relatively high rate of success of product ideas originates from marketing personnel and customers [41]. For sure, shaping the product according to customer needs and desires is a great advantage in terms of competitiveness.

But what is the meaning of "product idea"? To generate a product idea means to list attributes that define the concept of the product. Indeed, a useful representation of a product is a vector of attributes (e.g., speed, price, reliability, capacity). In this study attributes refer to both customer needs (also referred to as customer attributes or customer requirements) and product specifications (also referred to as engineering characteristics or technical performance metrics) [34]. Attributes are an abstraction of a product. Concept development also involves the embodiment of these attributes into some kind of technological approach, which we call the core product concept. The decision as to which technological approach to pursue is often supported by two more focused activities: concept generation and concept selection [34]. These two activities allow product developers to first generate a great number of ideas and then cut down to the feasible ones, usually passing through a stage – gate process or exploiting divergent and convergent thinking to narrow down the number of solutions.

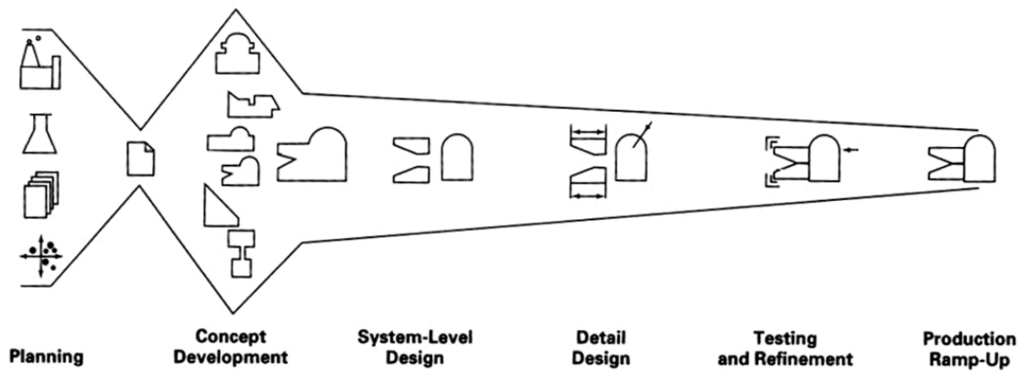


Figure 3.2: Divergent and convergent thinking (Ulrich, Karl and Eppinger, 2004)

Concept development decisions define not only the product specifications and the product's basic physical configuration, but also the extended product offerings such as life-cycle services and after-sale supplies [34], as well as the whole production system, as the study will show in the explanation of phase 7. As shown in Figure 3.2, when only feasible ideas are left in the hypothetical funnel, they are passed on to the next phases.

3.1.3 Stage 3 - Screening

Product strategy and planning involve decisions about the firm's target market, product mix, project prioritization, resource allocation, and technology selection. Scholars and researchers show that these factors have a significant influence on the probability of economic success. In structured development environments, product planning often results in mission statements for projects and in a product plan or roadmap, usually a diagram illustrating the timing of planned projects. This is what screening is about: scanning the market in order to make specific decisions, including the following. What is the firm's target market? What portfolio of product opportunities will be pursued? What is the timing of the PD projects? What assets will be shared across products? Which technologies will be employed in the planned products? Efforts are generally made to coordinate these decisions with the firm's corporate, marketing, and operations strategies. At the end of this stage, the concepts that do not meet the requirements listed before are eliminated, so that they will not absorb useful resources later on in the process. To be more specific, the end of stage 3 corresponds to a first product plan approval. The approval of the product plan is often based on how well it meets strategic goals, justification of the product opportunity, and how well the target market fits the company's image and vision (of whom it wants to serve). In a nutshell, product planning is the set of decisions ensuring that the firm pursues the right markets and products from a strategic viewpoint [34].

3.1.4 Stage 4 - Business Analysis

In the screening stage the analysis is done based on a new product showcase, resources and competition, thus making it a rather qualitative approach. On the other hand, in the business analysis stage, ideas are evaluated using quantitative performance criteria. After gathering enough new product ideas through various sources from the idea generation stage, the ideas to be pursued will be selected based on the business value they bring about. Making a good selection is critical to the future health and success of the business. The point is that product development costs rise substantially with each successive stage in the new PD process [15].

The ideas that have been classified as "Go" ideas must be screened further by using criteria set up by the top management [42] [43]. These ideas must be described on a standard form that can be accessed by a new product committee. The committee then assesses each idea against a set of criteria, which verify the attractiveness and visibility of the idea as well as its fit with the company's strategy, objectives and resources. The ultimate result of screening and evaluation is a ranking of new PD proposals, such as the resources can be allocated to the projects that seem most promising (Crawford, 1997; Wind, 1982).

After the screening, the business analysis is the detailed investigation stage that clearly defines the product and verifies the attractiveness of the project prior to heavy spending. Cooper's studies of new products showed that weakness in the upfront activities seriously compromises the project performance. Inadequate market analysis and a lack of market research, moving directly from an idea into a full-fledged development effort, as well as failure to spend time and money on the up-front steps, are familiar themes in product failures. For all of the above reasons it is important that at the end of this stage the enterprise comes to practical conclusions such as the likely selling price, a forecast of sales volumes, the target profitability and the break-even point.

3.1.5 Stage 5 - Design and Development

Once the results of the business case of the new product conform to company objectives, the new product team can move on to the development stage, which is made up of activities that range from prototype development to volume ramp-up and test marketing. The interaction between the program and project manager is no longer one of selling or buying the concept, but rather one of bringing the product to the market on time, within budget, and with the required specifications.

A product concept is generally brought to life through decisions about the physical form and appearance of the product. These decisions are part of an activity that is generally called industrial design [34]. At the step of product design, the data involved can be traced from the description of the needs to the specific product function description and finally to the detailed design specifications, like drawings of the product configurations, accurate programming codes for the automated manufacturing equipment, and all kinds of parameters [24]. Product design is an iterative, complex, decision-making engineering process. Generally, a design process consists of three phases: product design specification, conceptual design, and detailed design [25]. Moreover, it is necessary to make use of the maintenance and failure information to constantly modify and improve product design. Based on the history information of breakdowns and root causes, current products can be designed more efficiently and reliably [24].

In the development stage, business case plans are also translated into concrete deliverables. What is critical for success at this stage is to move through development, so as to launch the product as quickly as possible and to ensure that the prototype or final design does indeed meet customer requirements. This aspect in particular requires seeking customer input and feedback throughout the entire development stage.

We use the term "product design" in its narrow sense to refer to the detailed design phase, which constitutes of the specification of design parameters, of the determination of precedence relations in the assembly, and of the detail design of the components (including material and process selection). These decisions generally result in geometric models of assemblies and components, a bill of materials, and control documentation for production [34]. Given the number of variables involved, each design choice may be a tradeoff affecting many other design parameters [38].

It is important to gain competitive advantage and enjoy the product's revenues as soon as possible so that the enterprise can minimize the impact of a changing environment. Thus, as the product proceeds from one step of the development stage to the next, the personnel dedicated to the new product should reassess the market, position, product, and technology in order to increase chances of delivering a successful product [44] [45].

Closely related to the decision of which variants to offer is the decision about which components to share across products in a firm's portfolio. The ability to share components across products is partly determined by the product architecture, which is the scheme whereby a product's functionality is partitioned among components [34].

Such an intricate mix of variants and requirements that need to be taken into consideration has forced designers to adopt new techniques to shape products, the most important and revolutionary of these ones being the parametric design. The goal of parametric design is to decide values of design parameters while satisfying and/or optimizing some desired performance characteristics. Parametric design is generally performed after a basic product concept has been established, when the creation of a mathematical model of product performance is possible [34].

The output of stage 5 is a product project. Be it a virtual sketch or a physical prototype, after this phase the product has to be somehow tangible and ready to be tested.

3.1.6 Stage 6 - Testing

While detailed design decisions are being made and refined, the design is also prototyped to validate for fit, function, and fabrication. Ulrich and Eppinger [47] provide a comprehensive description of the prototyping process. Typically, the firm has a choice of developing prototypes sequentially or in parallel with different cost, benefit, and time implications [34]. Also, with the recent development of virtual prototyping technologies, a great part of the prototyping activities can be executed on the computer, thus making the process faster, cheaper and more efficient. After the prototype is ready, the core activity in this phase is to test it in all of its aspects. Testing means to verify the products capabilities, properties and performances. More than that, the purpose of this stage is to provide final and total validation of the entire project i.e. the commercial viability of the product, its production and its marketing [46]. To be precise, design and testing go hand in hand, with testing being conducted throughout the development stage. The information obtained during the tests is used to develop the product.

If a product contains designed components, decisions must be made on who will design these components and who will produce and test them [34]. As a consequence, suppliers may be actively involved in the testing phase, especially when products are made of distinct specific components: cars, ships, electronics and so on. Also, besides suppliers, other players external to the enterprise may be involved in this stage: governments. As a matter of fact, governments make regulations and standards on products, and enterprises have to take them into account.

This phase is extremely important in that it may dramatically decrease the chances of failure in launch, since it has the capacity of revealing flaws that could cause market failure [45].

3.1.7 Stage 7 - Commercialization

This last stage of the BAH model includes two activities: product launch and production ramp-up. A number of decisions must be made in association with these two activities. For instance, the firm must decide the degree to which test marketing should be done, and the sequence in which products are introduced into different markets [34]. Then, still regarding the marketing, an advertisement campaign must be organized which may also involve external experts. It is important for marketing experts to have access to all the information required in order to advertise what the product offers in terms of customers' needs fulfillment.

The activity of production also requires considerable effort on the side of the company. In fact, the phase of production is not just linked to the design but it is part of it. Indeed, designing a product also means designing its production phase, including machines to be used, process and dynamics. So, the production phase is nothing but putting into practice what has been planned during the design phase. If the flow of data between the stages of design and commercialization is not optimized, the whole production can be jeopardized. In practice, poor product-design decisions can also slow the rate of production ramp-up [34]. This is where the real complexity of production stands: the data during production are more than ever changing as the manufacturing processes are always dynamic and continuously produce real-time data. Numerous sensors are installed in the work environment to monitor the parameters of environment, equipment, and the products themselves. While some data might be stable, others are changing dramatically along with the flow of product manufacturing [24]. The perfect timing of product launch and product ramp-up guarantee not only the success of stage 7, but also of the entire new PD process.

As stated before, the BAH Model has been implemented by thousands of companies. Nonetheless, no model is perfect by itself. Often the problems do not lie with the scheme itself, but with how well or poorly it has been implemented [35]. In particular, enterprises implement the BAH Model following two main schemes. The first consists of applying the model to the company structure as it is, relying on the enterprise's functions, so letting each Business Unit taking care of its own part of the process. Manufacturing companies are created and structured in order to always develop new products, and their structure allows each member of the organization to fit into the process, bringing in its own set of skills and competences. The organization of a manufacturing firm into functions is particularly beneficial for managing an ongoing business with stable products, in which marketing is responsible for generating demand and operations is responsible for fulfilling that demand. The task of developing new products, however, presents an organizational challenge that introduces a discontinuity in ongoing operations. In other words, as we can deduce from the description of the BAH Model, when creating a new product there is a need for heterogeneous knowledge and flexibility, in order not only to develop the right product, but also to bring it to the last stage right on time. For this reason in particular (and for extraordinary business) functions may result too rigid. The alternative common approach is to assemble a team of individuals from various functions for the duration of the development process and to allocate the task of making subsets of decisions among all of them [34]. Firms are increasingly entrusting PD to cross-functional teams, which consist of members from several functional areas such as marketing, engineering, manufacturing and purchasing [39]. The structure of the team and its management depends on their tasks and on the kind of process they need to take care of. This study details three kinds of cross-functional teams that are largely exploited by manufacturing companies. The description of each team includes:

Team Name
- The kind of manager the team reports to
- What kind of resources the team controls
- The role of the team manager
- The responsibilities of the team manager

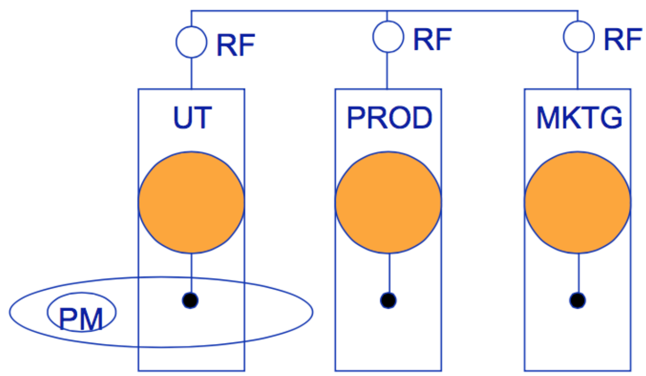
Team Collocation in the Organizational Chart

RF= Function Responsible
UT= Technical Department
PM= Product/Team Manager

PROD= Production Department
MKTG= Pre-Sales Marketing
MKT= After-Sales Marketing

<u>"Light" PD Team</u>

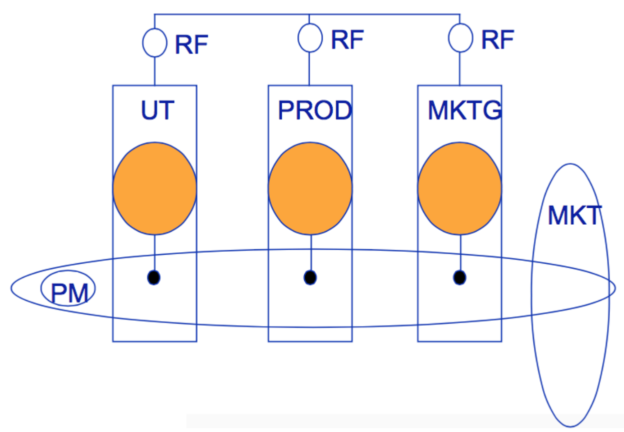
- Middle – Junior level Manager
- Key resources still under RF's control
- Facilitator role: enhances communication in the BU, but the team members still report to RF
- Responsible for the output pace



The "Light" PD Team is designed so that it does not have a strong impact on the functions. The team members keep on performing their task while the team helps the function it is assigned in terms of information flow and production pace. The "Light" Team's main advantage is to fasten the function and coordinate it from the inside. For this reason, this kind of team is not cross but rather intra-functional.

<u>"Heavy" PD Team</u>

- Senior level Manager
- Key resources and budget under his control
- The team members report to him
- Responsible for pace, cost and quality of the output

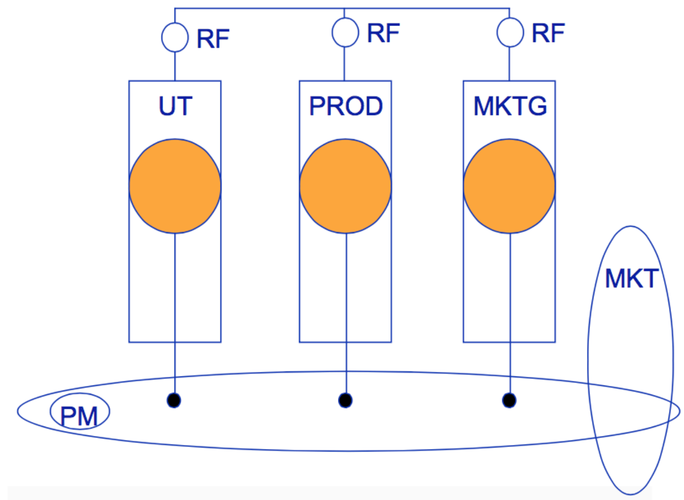


The "Heavy" PD Team is indeed cross-functional. It involves personnel from various functions according to their roles in the new PD process. For instance, in such a team, product positioning and market segmentation decisions are assigned to individuals coming from the marketing BU. Nonetheless, the "Heavy" team does not stand above the functions in the organizational chart.

The team members report both to the team manager and to their manager inside the function.

Task Force

- All team members are assigned to the project alone
- The team members are physically located near the PM in the firm's premises
- The team can ignore the enterprise's common procedures and bureaucracy



The Task Force is a kind of cross functional team that goes beyond the boundaries and procedural limitations of the functions. Its members are detached from their functions and they are assigned only to the teamwork. For this reason, the Task Force is used in exceptional situations, such as an urgent order or the development of a top priority product.

An additional aspect to be considered is the very definition of "new product". So far, this study has mentioned the new PD, which is a formal blueprint, roadmap, template or thought process for driving a new product project from the idea stage up to market launch and beyond [35]. But the truth is that most PD projects entail the redesign of existing items, rather than design of entirely new ones. Former literature has observed that in most large firms there is a huge investment in existing design procedures, often heavily bureaucratized [38]. In fact, enterprises systematically re use components and parts to develop new products. The reasons for such behavior can be the designing or stock cost of the components or just because the quality of a certain part is so high that it is worth re using. Research has shown that design and associated knowledge reuse is the key to reducing new product development time: only 20% of the parts initially thought to require new designs actually need them, 40% could be built from an existing design and 40% could be created by modifying an existing design. In addition, more recent studies estimates that more than 75% of the design activity comprises reuse of previous design knowledge to address a new design problem [50]. In this case the BAH Model remains valid, it is only a matter of defining what is really a new product. This is also the case of product versioning: when looking at the car market, it is common to observe that most models last for a decade. Usually once every two years the manufacturer re-designs the exterior parts or adds some more features, but the basis remains the same. The choice of product variants must balance heterogeneity in preferences among consumers and economies of standardization in design and production [34]. This is another example of how varied the concept of "new" PD can be.

To conclude, product innovation is a process that includes technical design, manufacturing, management and commercial activities involved in the marketing of new (or improved) products [36]. Learning and accumulation of (new) knowledge in an organization always start with the individual. An individual's personal knowledge has to be transformed into information that other members of the organization can use in their accumulation of knowledge in order to apply it to create (new) values for the organization. Learning and accumulation of new knowledge in an organization always require two transformation processes: one transformation process from data to information and another from information to (new) knowledge. This is so because only information, and not knowledge, can be shared and spread among the members of the organization [36]. For this reason and because of the extreme importance of data sharing among designers, marketing experts, common developers, team, functions and, more in general, in the whole company, enterprises have adopted software tools to manage PD data.

3.2 Traditional Software Architecture for Product Development

The data related to products are always in the biggest volumes and highest rates of change. When a product is being produced, to ensure the product quality is up to standard (particularly in precision manufacturing) all design demands such as size tolerance, geometric tolerance, and surface roughness should be strictly followed. Thus, configuration parameters like thickness and length, location parameters like coordinate, tolerance parameters like concentricity, and even the intensity of the material should be monitored at any time to decrease the risk of lower quality product. In conclusion, during the PD process the data from the stage of product design will be regarded as the production standards that the data from both the monitoring and the testing of products have processed to reach the standards [24].

3.2.1 CAD, CAE, CAM, PDM

The PD process phases 4 and 5 in particular are supported by several software instruments developed over the years. These tools can be either integrated modules or stand alone products, but all of them produce an output in the form of electronic files. Here follows a brief description of each type of file and the PLM software system component it comes from. It is a list of what this study calls the "traditional architecture" of PD, meaning that these four tools are nowadays well known and largely exploited by manufacturing enterprises:

CAD: Among the PLM tools for creating information, mechanical Computer Aided Design (CAD) systems are the most important ones [22] because, no matter the kind of product and its design complexity, product design needs to be facilitated by computer-aided systems to assist designers in the creation, analysis, and optimization of a design [26].

Regarding the birth and evolution of this system, it is argued that the first CAD system, SKETCHPAD, was developed at MIT by Ivan Sutherland in the early 1960s [26]. Over the years, CAD was utilizing the opportunities offered by general-purpose computing technologies, except some cases in the old past, when a few peripheral devices (like pen-plotters) were developed exclusively for the purpose of computer aided drawing or, later on, for interaction with systems [27]. Actually in the end CAD was doomed to become the perfect natural substitute of pen-plotters.

The first commercial applications of CAD systems were found in large enterprises, mainly in the automotive and aerospace industries. Back then, those were the only ones who could afford and justify the extremely high operation and maintenance costs of the early-day CAD systems. With the advancement of computer hardware and geometric modeling, CAD systems could be run on more affordable personal desktop computers and allowed for 3D solid modeling. With the advancement of the Internet and the client–server model, distributed CAD and the sharing of decentralized computing resources became possible. Later on, web-based CAD system based on the thin server–strong client architecture turned out to be hard to implement because of the heavy-weighted client mechanism; however, CAD systems based on the strong server–thin client architecture model are more effective and efficient in distributed and collaborative settings because of their light-weighted client mechanism [26].

Over the years, the tough competition among software vendors and the concentrated research efforts led not only to the enhancement of capabilities, but also to a diversification of the considered application domains (mechanical, electronic, architecture, etc.) and supported phases of product development. This diversification triggered an articulation of CAD functionalities. It gave floor to standard (cross-domain) functions such as computer-aided drawing, free-form curve/surface modeling, wireframe, solid and boundary modeling, assembly modeling, as well as to application domain-specific functions and domain-oriented implementations of standard functions such as computer aided process planning.

Nowadays, the CAD has become so important that in manufacturing companies the whole design process is carried out with Computer Aided Design software. In particular, one of the main advantages of a CAD software is that it provides engineers with the tools needed to perform their technical jobs efficiently and free them from the tedious and time consuming tasks that require little or no expertise [4].

With the inclusion of virtual engineering and rapid prototyping technologies, computer support of product design and engineering could expand from the mental domain of humans through the virtual domains created by computers down to the physical domain [27].

It seems that in the last period, namely from 2006 until today, the evolution of the CAD has reached the ceiling set by traditional computing technologies [28].

Whenever a project is started, the designer works on the virtual model, so that it can be modified, analyzed, optimized and, most of all, shared, because no kind of information and no data is easier to understand than an image. With the latter important feature the designer improves communication through documentation and creates a database for manufacturing. Official technical records and drawings can also be created through the CAD.

As a consequence, the output of a CAD system is an electronic file that conveys information such as materials, processes, dimensions and tolerances in the form of virtual reality images.

CAE: once the entity is designed, it requires the work of an Engineer to transform a sketch into a proper project ready to be manufactured. Computer Aided Engineering (CAE) software solutions allow engineers to make testing, calculations and to support them in performing engineering analysis tasks on virtual prototypes. The PD steps of simulation, validation and optimization of parts, components and products are executed through CAE systems. The following are typical operations made by CAE systems:

- Pre-processing: defining the model and environmental factors to be applied to it.
- Analysis solver.
- Post-processing of results.

CAE software covers the following engineering fields and areas:

- Stress analysis using Finite Element Analysis;
- Thermal and fluid flow analysis;
- Multi body dynamics and Kinematics;
- Analysis tools for process simulation for operations;
- Product or process Optimization.

CAE software systems are recognized by both literature and users as the most complex in the traditional architecture panorama. It is common thought that the success of computer software in engineering rarely depends on the performance of the hardware and the system-support software. Computer applications have to match engineering needs [29], but applications are designed by computer and software experts, whereas users may require specific features depending on their industry, involvement in the process and tasks.

The output of CAE software systems is an electronic file that coveys constraints, dimensions and tolerances.

CAM: Production Big data are driving additional efficiency in the production process with the application of simulation techniques to the already large volume of data generated by production. The increasing deployment of the "Internet of Things" is also allowing manufacturers to use real-time data from sensors to track parts, monitor machinery, and guide actual operations [65]. Computer-Aided Manufacturing (CAM) can be defined as the use of computer systems to plan, manage, and control the operations of a manufacturing plant through either direct or indirect computer interface with the plant's production resources [5]. In other words, the use of computer system in non-design activities but in manufacturing process is called CAM [30]. To put the definition into more practical terms, once all the PD Process is concluded, the product must be put into production, and CAM is the software system that has enabled the direct link between the three-dimensional (3D) CAD model and its production [30].

The evolution of virtual manufacturing has led to the creation of work-cell simulation tools that are capable of developing, simulating, and validating manufacturing processes. Moreover, off-line programming of multi-device robotic and automated processes (virtual commissioning) offer optimization functionalities, from the concept to the implementation phase [30].

With CAM systems it is possible to recreate a production plant in a virtual environment: the utilization of CAM enables the automation and computer support of all production activities on the shop floor, in order to manufacture parts designed with CAD and analyzed with computer-aided engineering (CAE). The equipment on the shop floor such as robots, controllers, machine tools, and machining centers, are controlled and operated using CAM systems [30]. For this reason, CAM software is often referred to as "Digital Manufacturing" or "Virtual Manufacturing" tool. As a consequence, when working in a digital environment, by taking inputs from product development and historical production data (e.g., order data, machine performance), manufacturers can apply advanced computational methods to create a digital model of the entire manufacturing process [65].

CAM software systems allow users to execute mainly the following operations:

- Set a system of coordinates to be used by machine tools;
- Select from a virtual model the portion of part to be machined;
- Define manufacturing operations to be repeated in the future;
- Set working parameters to determine finishing and tolerance levels;
- Set auxiliary working parameters, for example tool sorting, tool dimension and shape, working speed parameters and so on;
- Generate instructions for machine tools and save them using a post processor;
- Tool path graphic simulation;
- Send the data to Computer Numerical Control and machine tool;
- Interference control: tool-part or stirrup-part.

Synchronization among robots, vision systems, manufacturing cells, material handling systems, and other shop-floor tasks are challenging tasks that CAM addresses [30]. The proliferation of Internet of Things applications allows manufacturers to optimize operations by embedding real-time, highly granular data from networked sensors in the supply chain and production processes. These data allow ubiquitous process control and optimization to reduce waste and maximize yield or throughput. They even allow innovations in manufacturing that have so far been impossible, including nano-manufacturing [65]. As a consequence, in order to work properly, CAM software systems have to be supported by a bigger software infrastructure and connected to the rest of it. For example, the CAD databases must reflect the manufacturing requirements such as tolerances and features and the part drawings must be designed having CAM requirements in mind.

PDM: Product Data Management (PDM) is the discipline of controlling the evolution of a product and providing other procedures and tools with the accurate product information at the right time in the right format during the entire product lifecycle [22].

In any project it is important to document, so as to ensure that everybody is on track with the project's vision. In a product development project, data about the product need to be stored and kept available for the development team. The system responsible for managing product-related data is called PDM. The PDM software system is a software tool for the collection and organization of all data regarding the entire PD process.

3.2.2 PLM

The conclusion is that CAD, CAE, CAM and PDM systems allow designers and engineers to have a better working environment, but they do not take into consideration the key problem of data handling and information sharing (see paragraph 3.2.1). The main reason is that traditional computer-aided application tools were standalone systems and designed for single user without communicating and collaborating with others [26]. As a result, the problem is just shifted ahead, but it still persists: there is a need for information flow inside and throughout the company, and the information has the shape of data. In order to solve this problem, another software system has been created, standing above the four listed: PLM software systems allow all the other systems explained before to interact and exchange information. To tell the truth, PLM as a discipline emerged from tools such as CAD, CAE, CAM and PDM, but can be viewed as the integration of these tools with methods, people and the processes through all stages of a product's life [21]. But then, PLM as a software system and as a tool is meant to work as the hub for PD systems and tools, increasing reliability and facilitating the exchange of product data [22]. Combining the researches of Liu and Bergsjö, there are six basic functionalities in a PLM system:

1. Information warehouse/vault: The place where product data are securely stored
2. Document management: To manage, store, retrieve and use product information in the form of documents in an orderly manner. This includes document control capabilities, such as

check-in/check-out.

3. Product structure management: The arrangement of product definition information as created and changed during the product's lifecycle. Product definition information includes, requirement management, product configurations, associated versions and design variations. It facilitates the management of different BOMs: design BOM, production BOM, maintenance BOM etc.
4. Workflow and process management: The set-up of rules and procedures for work on the product definition and associated documents, managing engineering changes and revisions. It provides a mechanism to drive a business with information.
5. Parts classification management: Allows similar or standard parts, processes, and other design information to be grouped by common attributes and retrieved for reuse.
6. Program management: Provides work breakdown structures and allows coordination between processes, resource scheduling, and project tracking. [22,6]

PLM is a broader concept than PDM, which takes in the whole lifecycle as well as the tools used for authoring data. PDM remains the foundation of a PLM system, but the term PLM is used to consider the product lifecycle and collaboration aspects regarding product data [22].

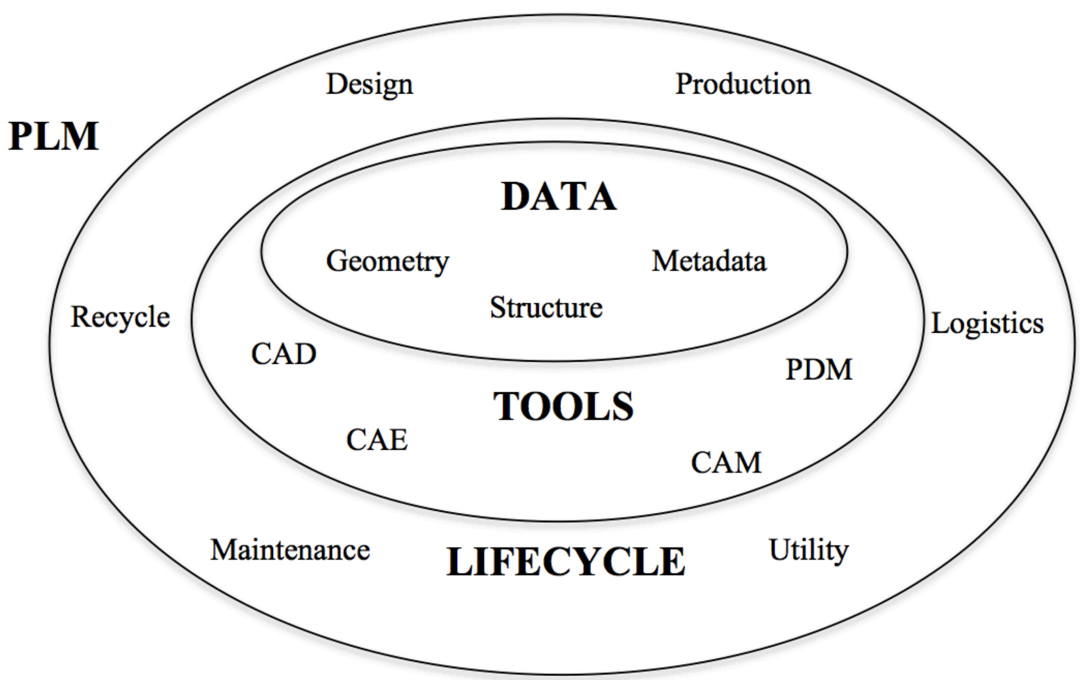


Figure 3.3: The concept of PLM

Besides, in order to be performing, PLM systems are to be connected to the company's Research and Development business unit. In particular, over the years companies have worked on the connection and interaction of PLM with Enterprise Resources Planning (ERP) and Customer Relations Management (CRM) software systems.

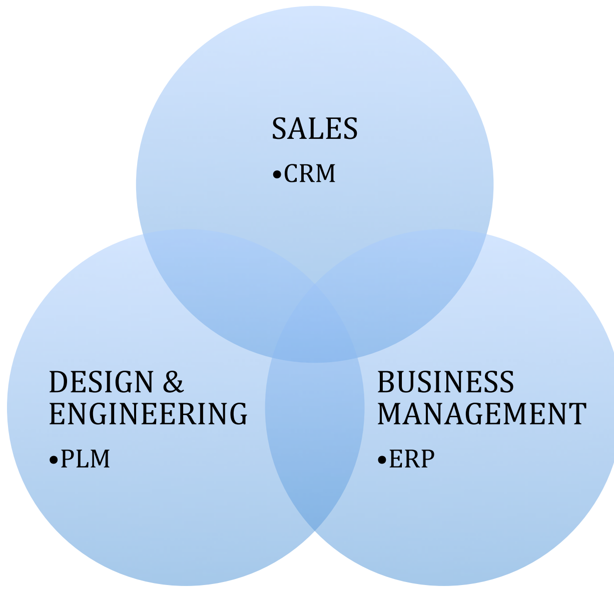


Figure 3.4: The main software systems in enterprises

This level of integration is of the utmost importance: the need for information management and structured collaboration is essential in the industry today, since there are many people involved in the PD [25]. Research has shown that there are four major factors to be considered in a successful integration: tools/technologies, processes, culture and people's behavior [25]. Achieving the perfect integration among the four is still an open challenge not only for the PD process but for the whole industrial world in general.

Before proceeding with a gap analysis of the traditional software architecture, it is important to add some detail to the explanation given so far. Both in the literature and the industrial world the difference between PDM and PLM it is often unclear. Sometimes they are thought to be the same software system; sometimes it is not clear where PDM ends and where the system scales up to the PLM level. In fact, the most popular commercial vendors mainly propose PLM as a high end PDM and, vice-versa, PDM as an old and low end version of a PLM.

	CAx		Digital Manufacturing	Collaborative Product Development	
	High End	Low End		High End	Low End
DASSAULT SYSTEMES	CATIA V5	Solid Works	DELMIA	ENOVIA	Smarteam
PTC	ProENGINEER		cMPM	Windchill	
SIEMENS PLM SOFTWARE SIEMENS	NX	Solid Edge	Tecnomatics	Teamcenter	Teamcenter Express
AUTODESK		Inventor			Buzzsaw

Figure 3.5: The four largest PLM software vendors per market share as of 2015

In order to avoid confusion and misconception, this study will refer to PDM as product data management software systems and to PLM as the overall software system for the interaction of the four elements of the traditional architecture.

In the next Chapter, the study proceeds with analyzing the gaps in the traditional architecture and to explain how both the software industry and researchers are trying to close this gap.

3.3 The Gap in the Traditional Architecture

3.3.1 Reasons for a New Tool

These four software systems plus the PLM superstructure are the backbone of PD or, as said before, they are the traditional software architecture for PD. They all represent the state of the art of PD technology, thus making PD much easier for designers and engineers. Nonetheless, there is still plenty of room for improvement in the area of Big Data management. Specifically, none of the above-listed systems have a peculiar feature for proper data retrieval. In fact, PDM systems are data management tools and do have some sort of search engine embedded. But their focus is mainly on data storage and data tracking. As the volume of data grows, the challenge comes in being able to store, access, and analyze it effectively [7]. Besides, these embedded systems are limited to the PDM systems databases, but they do not take into consideration all the other instruments the designers use to develop products. For example, a mechanical engineer does not just spend his time designing clamps on his computer; he also exchanges emails with his co workers, sends clamps drafts to his supervisor, checks for other clamp experts inside the company who may know more about the subject. This is also true for Excel files, Word documents, and every kind of file in which designers, engineers and developers put down their knowledge in the form of words, numbers and images. Talking about emails, the email exchange in particular is still the most popular mean of communication among colleagues, so it represents an enormous source of knowledge and information that is not fully accessible with tools from the standard architecture, because those tools do not take into consideration text analytics.

Text analytics has its academic roots in information retrieval and computational linguistics. In information retrieval, document representation and query processing are the foundations for developing the vector-space model, Boolean retrieval model, and probabilistic retrieval model, which in turn, became the basis for the modern digital libraries, search engines, and enterprise search systems [17]. The increased amount of available information, and the need to manage it from several and traditionally different engineering fields, have made it evident that it is no longer possible to design without solid knowledge of what is going on in related fields. With a different set of development tools, vocabulary and process traditions, the mechatronic field now has to confront the design tasks as an integrated design team. PLM systems and their PDM predecessors show promising signs of being able to support this. However, it is not (only) a matter of collecting and presenting information for designers. Suitable means must be found to categorize and identify the information that is needed [22]

Hence, the future challenge of software tools in the world of manufacturing is about how to retrieve Big Data and how to make them accessible and useful to the various operators throughout the company, because without data mining, we cannot deduce any meaning out of the high-dimensional data at our disposal [13].

Apart from the literature studies, in order to support this claim this work reports part of the latest Gartner's report on enterprise search, which states that [48]:

- By the end of 2017, 25% of workers will engage with search technology in business applications via natural expression at least five times a day.
- By the end of 2017, the best result for more than 50% of searches at leading global companies will not be a textual document.
- By the end of 2019, more than 10% of internal search results will not originate from explicit queries.

There is a need for a tool serving as a search engine to retrieve, display and analyze information in the form of Big Data out of the company's PLM software system. Both research and software vendors have thought of this new tool as a feature of an already existing product called Enterprise Search Engine (ESE).

Enterprise search allows users in an enterprise to retrieve the desired information through a simple search interface. It is widely viewed as an important productivity tool within an enterprise [32]. Instead, this study deals with ESE focusing on their contribution to the PD process, as it considers ESEs as the fifth software tool of the PD software architecture.

Nowadays ESEs have been improved and they have abandoned to old SQL-type way of searching. Nonetheless, they are still a very pervasive kind of tool inside enterprises, therefore they are well known also by designers and engineers. Now the challenge is to shift ESEs' focus on PD, making them more suitable to this new purpose.

3.3.2 Enterprises' Needs

Manufacturing Companies have a long series of needs, but in general terms, their ultimate goal remains that of making the greatest profit. Making profit means maximizing revenues and, at the same time, reducing costs not just in terms of money, but also in terms of time. In one word: competitiveness. Basically, when developing a new software system for manufacturing companies, it has to be translated into a tool to increase enterprises' competitiveness. This is also true for the PD process.

As explained before, the PD process represents the core of the creation of value for companies; therefore an ESE must be an additional tool to enhance productivity and competitiveness throughout the company, starting from its core.

Then, given the role played by Big Data and their central role in the PD process, it is clear that the early mentioned ESEs also need to respond to certain requirements in order to properly address the Big Data issue, namely:

User Friendliness: a key objective of every company is to provide its employees with tools that are easy to use. This means on one hand to have tools needing a minimum set up time and training in order to be used properly and, on the other, to have tools showing the result of their work in a simple way.

Regarding the first point, companies are afraid of facing the same problem they have when dealing with PDM software: users often have to spend a significant amount of time to become familiar with new software. Consequently, instead of concentrating on their daily tasks, they have to focus on how to use the new PDM product [6]. This may results in the user rejecting the tools. They may be perceived as an imposition, a complication or, even worse, an obstacle to the correct implementation of procedures the user used to be performed well without the tools.

Secondly, it is important to remember that ESEs not only search for information but, as stated before, that they are also visualization, presentation an analysis tools. It is known that presentation of the results and their interpretation by non-technical domain experts is crucial to extracting actionable knowledge, as most of the BI related jobs are handled by statisticians and not software experts [9].

The third need relates to the searching activity itself. Search provides an efficient way of finding relevant information from huge amounts of data only if the users know what to search for. In large corporations, queries can have large result sets. It is important to facilitate users in forming effective queries through browsing and navigating information in a relatively small, manageable space [11].

Data Merging: the ESE has to combine both structured and unstructured data in order to have a complete view of the information and knowledge.

Over 80% of enterprise data is considered to be unstructured [9], which means that the overwhelming majority of information in an enterprise is unstructured, i.e. it is not resident in relational databases that tabulate the data and transactions occurring throughout the enterprise [11]. Together with information in relational and proprietary databases, these documents constitute the enterprise's information ecosystem [11]. This is also true for data coming from the PD process; nowadays information is shared among designers using emails, videos, text documents, pictures, digital audio etc. These are all examples of data that are not described by a data model, and are therefore identified as unstructured data. In particular, images and video are structured for storage and display, but not for semantic content and search [9].

With such a great amount of information stored in the form of unstructured data, it is essential to provide high-performance parametric search allowing the user to navigate information through a flexible combination of structured and unstructured data [11].

The real need for enterprises is to add structure to unstructured data, so that it is possible to store them for analysis along with the already structured ones, thus creating an ideal merging of the two kinds into one Enterprise Database.

Security: privacy and security are also important challenges for Big Data. Because Big Data consists of a large amount of complex data, it is very difficult for a company to sort out this data on privacy levels and apply the according security [18]. In fact, every enterprise is made of people with different roles, responsibility levels and rights to access the company's knowledge. As a result, security access to data has become a top priority, since people are usually engaged in enterprise-specific behavior and are being compensated for their engagement [11]. Enterprises need the ESE to help personnel search for knowledge inside to company's intranet but also to protect themselves from unauthorized access.

In fact, enterprises need on one hand to encourage the designers to collaborate, share their ideas and knowledge as much as possible for a proper Collaborative Product Development. It is well known that some of the most valuable knowledge in an enterprise resides in the minds of its employees [11], but most of the times there it stays because knowledge is not well protected even inside the enterprise itself. At the same time, the ESE also requires a certain degree of flexibility in terms of security. As a matter of fact, people do change role and qualification inside the organization, so the ESE needs to recognize the new roles and to modify the permissions. A further step to enhance collaboration can be to open the ESE to external members, thus creating a temporary permission. To sum up, security is a key issue for enterprises both in terms of protection and knowledge sharing.

Presentation Capabilities: when dealing with *user friendliness* this study already mentioned that enterprises require the results of a query to be shown clearly also to the ones who are not software experts. This is only the first step of the much bigger challenge of presenting the results of the search. In fact, the key challenge here is to merge sets of results from all sources for a unified presentation. This must be done even though the sets typically have no documents in common and employ different scoring and ranking schemes [11]. The ESE also has to combine traditional PD data (CAD models, tables of content, BOM etc.) with new and unstructured data (see "*merging*") coming out of the search.

Another issue related to search result presentation regards the order in which the results are shown. When the average Internet surfer uses a search engine, he expects the engine to show him a list of results starting from the best or most fitting answer. Otherwise, when using an ESE users' input is much more specific, and they expect the right answer to their query. For example, in a boiler manufacturing enterprise the electrical engineer searches for the latest revision of the electrical box of model "201". He will expect the ESE to show him what he typed and then to present him some suggestions based on what it is similar or related to it.

Integration: when using big data analytics, the user is trying to discover new business facts that no one in the enterprise knew before, a better term would be "discovery analytics". To do that, the analyst needs large volumes of data with plenty of details [9]. The only way to obtain such a level of detail for such an amount of data is that the ESE has to be fully and totally integrated with all other PD software systems.

Integration with CAD, CAE, CAM, PDM and the whole intranet (emails, file systems, database etc.) allows the ESE to acquire and analyze data in real time. This is a key advantage in terms of competitiveness, because in a business context it is important for organizations to be able to analyze rapidly changing data in order to make good business decisions [7]. Integration is also a crucial need in terms of software upgrade. Enterprises want software systems that can upgrade quickly without compromising their proper functioning. Though a search service can incorporate new technologies in quick cycles, this is often not the case for enterprise deployments. Economic and time constraints sometimes hinder quick upgrade cycles. Often, enterprises use old software versions, although they are fully cognizant that they are not employing other available technologies. This is more evident when search software is embedded in third-party enterprise applications with extended release cycles. This sometimes leads to end users being disappointed with the quality of search provided within the enterprise [11].

This list requires further explanation. The needs are classified as "top priority" because they represent what enterprises look for in a new ESE available on the market. For sure, there are several other elements to be taken into consideration in the Big Data era: an ESE has to come at a low development cost, has to ensure high performance even at Big Data scale, to have low latency indexing whatever the volume of data, to scale linearly at a low hardware cost etc. In fact, these are all considered the basics of modern ESEs. Modern enterprises take for granted that nowadays ESEs are fast, scalable etc., because modern employees are used to Web-Based search engines and to state-of-the-art PD software systems. Therefore, according to literature and to all the considerations listed above, the following are the top priority needs for Enterprises when looking for the proper ESE:

- User Friendliness
- Data Merging
- Security
- Presentation Capabilities
- Integration

The study now proceeds presenting in order:

1. A description of ESEs, their functions and the technology behind them.
2. A modern ESE that best represents the ultimate software solution available on the market. This research will match the ESE with the enterprises needs listed above in order to compare the needs with the state-of-the-art and to point out possible gaps or improvements.
3. A model for implementing a state-of-the-art ESE in a modern manufacturing enterprise

4 Enterprise Search Engines

The aim of enterprise search is to empower the members of an organization with all the information contained throughout the enterprise itself. Over the years, the amount, form and quality of enterprise data has radically changed, and so has enterprise search. The history of enterprise search is bound to the evolution of computer technology and the technology average users, as well as to the World Wide Web.

4.1 Enterprise Search History

It is difficult to find detailed information about enterprise search, mainly because historians do not take into consideration the difference between web and enterprise search. In fact, up to the early 2000s, the two were actually the same thing. One of the most complete prospects comes from the Ballardvale Research [4i], which divides the history of enterprise search into four stages:

- 1985 - 1994: Expensive Proprietary Searching of Subject-Specific Databases
- 1995 - 1998: The Web Changes Everything
- 1999 - 2004: Academia Tackles Search -- Both Web and Enterprise
- 2005 - 2009: Embedded Personal Search

4.1.1 Moments of History

1985 – 1994: Expensive Proprietary Searching of Subject-Specific Databases

In the mid-1980s, computers started becoming popular with enterprises. That was the time when enterprise search also moved its first steps. In that period content suppliers offered clients the ability to search subject-specific databases. The search languages were proprietary and arcane, and it was the corporate librarian or another trained person who typically made the search, not a business end-user. In addition, these solutions were expensive because of the early stage of technology, so only the largest IT companies such as IBM® could afford them. As a result, ESEs were an elite tool that was just starting its development process and was still bound to the specific knowledge of computer science experts.

1995 – 1998: The Web Changes Everything

By the mid-1990s, the Web became increasingly popular within both enterprises and common users. Thanks to its user-friendliness, the "browser navigation" became a standard for all the new computer software that were put on the market. Mainly because of these two radical innovations, the ground was fertile for the birth of the first web search engines (i.e.: Google, Sept. 4th 1998) and for the development of keyword search rather than arcane, DOS-style commands. Every user became able to perform a search.

1999 – 2004: Academia Tackles Web and Enterprise Search

By the late-1990s, university computer science departments and think tanks began analyzing how to improve search and categorization. The universality of the Web and the need to index millions of documents suddenly made the problem intellectually interesting. This new research field boosted search companies such as the formerly cited Google, Yahoo and Altavista; they ultimately decided to concentrate on either Web search or enterprise search, but not both (Google and Microsoft are the exceptions) and they started developing algorithms for search optimization. The early 2000s were punctuated with intense debates on the underlying technology (semantics analysis, taxonomy, Bayesian inference etc.) as well as the mixing of Web search and enterprise search. The very outlook of enterprise search changed radically, with the progressive upgrade of arcane search charts and forms for the new search bars. These bars are still called "Google-like search bars" because worldwide users had become very familiar with them thanks to the popular Web search engine.

2005 – present days: Embedded Personal Search

The advancement of the Web 2.0 and the advent of an increasingly connected world forced search companies to further develop and expand their business. Search had become the key to access a volume of information that could no longer be managed without a combination of search and analytics tools. These tools came in the form of applications. Because of the evolution and popularity of the Web, enterprise search was now taking inspiration from Web search to create its own applications and trying to resemble the appearance of Web search engines: search bars, typing suggestions and, most of all, browser interfaces. In fact, the use of these browsers gives the user a real feeling of a web-based search engine rather than the traditional search tools.

A much richer view of information storage is fundamentally changing search capabilities and functionality. The indexing of content is becoming automatic, and the center of attention for search engine vendors is the user rather than the technology itself. Nowadays, the software systems considered in the present study include a series of functions that was not even imaginable during the first stage of their lives. For this reason, although they are called ESEs, they have become de facto Search Based Applications. The effort of improving ESEs and making them closer to users still goes on. In particular, one of the main focuses for current search companies is to embed human-like artificial intelligence (such as the well known Apple's Siri) in their ESEs.

4.1.2 Search Based Applications

In the above paragraph, the study mentions the so-called Search-Based Applications (SBAs). Software vendors as well as some scholars tend to consider the concept of ESE as somehow obsolete. The core functions of accessing and handling data and Big Data have empowered tools for business discovery, business intelligence, data mash-up, monitoring etc. In fact, all of these new features have transformed the traditional ESEs into the broader concept of Search Based Applications. Indeed, their evolution is still going on, as shown in Chapter 9 of this study.

4.2 The technology behind Enterprise Search

Although ESEs enroll a large number of data handling features, the basis for their functioning remains the ability of searching for data. The logic behind bare enterprise search systems is rather simple: when the user inputs a <u>query</u>, the ESE processes and <u>matches</u> it with the information <u>contained</u> in its index using <u>connectors</u>, then it shows the information to the user. The ESE shows the user this list of results according to the level of <u>precision</u> and <u>recall</u> set. Afterwards, if receiving an input from the user (for example, a "click"), the ESE links the information shown to the source of data. The whole operation is executed following the security criteria that are set when installing the ESE.

The process itself is of quick understanding. However, there are some fundamental technical terms that are worth explaining.

4.2.1 Technical Terms and Processes

Query Processing

Using an interface, the user issues a query to the system. The query consist of all the input given by the user with the aim of retrieving information from the system. As mentioned in paragraph 4.2, queries used to be specific strings of numbers and characters put in a standard format because of the limitations of early systems. Nowadays queries vary from the traditional text queries to pictures or geometries. As a matter of fact, in an enterprise there are various users and everyone can perform a different query in order to retrieve certain information. For example, military and security installations have been the first to use systems that can match an input picture with the similar ones in a database. Therefore, there are a large number of possible queries that the ESEs need first of all to process, then to match. In order to improve the ESEs query processing capabilities, the technology has made it possible for them to transform a query and broadcasting it to a group of disparate databases or external content sources with the appropriate syntax.

Matching

After processing the query, the ESE compares it to the stored index, then it returns the result of the query to the user. These results can be shown as a list of links to data or documents, as done by the Web search engine Google, or you can have the ESE present a preview of the documents or even the documents themselves. The way of presenting data depends not only on the installation settings and the ESE's capabilities, but also on the way the documents have been indexed.

Content awareness

In order to perform its duties, the ESE has to be linked to a specific content, but this content needs to be up to date. Therefore, ESEs need to be somehow aware of the content they show to the user. Content awareness can be executed following either a push or pull model. In the push model, the ESE integrates a source system so that it connects to it and pushes new content directly to its application programming interfaces (APIs). Such a model is used when it is important to have real time indexing. In the pull model, the software gathers the content from the sources using connectors.

Content processing and analysis

ESEs extract contents from different sources. For this reason, data may have many different formats, and document can be of different types. The problem has been already mentioned in Chapter 2 when discussing Big Data: XML, HTML, Microsoft Office Package document formats, plain text documents, pictures, videos, emails etc. In the most common case of text documents, the system needs to process the content in order to overcome the difficulties machines have in interpreting human language. First of all, the content is divided into tokens. A token is a pair consisting of a token name and an optional attribute value. The token name is an abstract symbol representing a kind of lexical unit, e.g., a particular keyword, or sequence of input characters denoting an identifier. The token names are the input symbols that the parser processes. Once the text is divided into tokens, it can be normalized to avoid case-sensitive search. Accents can also be normalized, so that recall is improved. The token is used as the basic matching unit for processing and analysis.

Connectors

A connector is a link between data inside a data structure. A data structure can be pictured as a set of nodes, whereas the connectors are the links between them. The connector is due to sound the source of data with pre-defined time intervals to look for content deletion, modification or update. However, connectors are also used to guarantee secure access to data. In fact, there are two main approaches to apply restricted access to data: early binding or late binding. In the late binding method, permissions are analyzed and assigned to documents at query stage. When the user inputs the query, the engine generates a data set. Before returning it to a user, this set is filtered based on the user's access rights. In the early binding method, permissions are analyzed and pre assigned to data at indexing stage. Connectors play a key role in the early binding method. As a matter of fact, in late binding the engine activates only the connectors that match the access rights of the user. So, the connectors do not act as a security filter. On the contrary, connectors are the main security implementers in early binding, because they are the ones checking the access rights: once there is an input from the user, the designated connector finds the requested data and checks the access rights. Then, if the access rights do not match, the connector inhibits the link and the data are not returned to the user. In this case, connectors have to be properly set when the installer starts the indexing process. Access control, usually in the form of an Access Control List (ACL), is often required to restrict access to documents based on individual user identities. There are many types of access control mechanisms for different content sources making this a complex task to address comprehensively in an enterprise search environment.

Indexing

The index is the list of content in a database. As mentioned before, the content is made of data and documents from a variety of repositories, therefore modern ESEs need adapters to index content from file systems, intranets, document management systems, e-mails, computers, databases and so on. An index is necessary because the ESE's database cannot contain all data; it has to refer to a summary. That is why the index is only a list of addresses. When implementing the ESE, the installer has to link the element of the index with its source. This activity is called indexing. In other words, indexing can be described as the collecting, parsing and storing of data to facilitate information retrieval. Indexing becomes more complicated when the system has to integrate structured and unstructured data in their collections; in this case, the ESE needs to implement specific algorithms for content search.

It is important to underline that ESEs are classified as vertical search engines, which means they are focused on a restricted area of content, in contrast to general engines such as the ones used for Web searches. Modern ESEs are embedded with adapters to index content from a variety of repositories, such as databases and content management systems

Precision and Recall

In information retrieval with binary classification, precision refers to the part of retrieved information and data the user considers relevant to his/her search. Recall is the part of relevant information and data that are finally retrieved. Both precision and recall are therefore based on an understanding and measure of relevance. As stated before, it is often necessary to normalize the content in various ways to improve recall or precision. Some of these ways are:

- Stemming: reducing derived, composed or inflected words to their root. Ex: cat, catty, kitten, cats stem from their root "cat".
- Lemmatization: grouping together different inflected forms of a word. Ex: the lemma of

"better" is "good".

- Synonym expansion: to match every word with all of its synonyms.
- Part-of-Speech tagging: also known as grammar tagging, it means to identify the words as nouns, adjectives, verbs etc.

Precision and recall are also two parameters used to evaluate search engines in general.

In addition, ESEs enlist a series of features that support the main functions and allow the entire system to be more efficient. These features have been developed over the years following the advancement of technology in the field.

4.2.2 Feature Enhancements for Enterprise Search

Parametric Refinement

Over the years, the amount of information available on the enterprises' hardware premises and in the Internet has been constantly increasing. Along with it, the users' requirements for queries typing have also increased: users need to perform advanced searches, to refine them along the process and often they happen to require fuzzy results. To be specific, users need to sort attribute values and the results coming from the query. Parametric refinement provides a solution to this problem by augmenting a full-text index with an auxiliary parametric index that allows fast search, navigation, and ranking of query results. The main issue with this powerful technology is that data preparation is very important, and organizations need to invest time in augmenting and normalizing the data. Classification and entity-extraction techniques will be used to augment information with attributes for improved search and navigation [11]. Parametric refinement is then classified as an improvement of user-friendliness and both precision and recall.

Semantics

Semantics is the branch of linguistics devoted to the investigation of linguistic meaning, the interpretation of expressions in a language system [14]. Semantics is the key to achieve impressive user-friendliness in the field of ESEs. For example, traditional databases are not designed to handle fuzzy constructs or unstructured text, so there is a need to enrich the system and make it more sensible to human language. The ability to dynamically construct virtual documents that can consist of relevant portions of many documents will be critical [11].

Extracting and finding information in large quantities of unstructured data requires assigning data into semantic classes, so that information can be filtered merged, and labeled [51].

In modern search engines, semantic dimensions are materialized as facets. In the field of semantics, facets are usually presented as a list of expandable categories. If, for example, you type ”wool” on a popular shopping website, you see the following ”Departments” on the left side of the screen: Arts & Crafts, Clothing & Accessories, Books, Home & Kitchen. Each of these ”Departments” is a facet, or semantic dimension, that allows you to divide the universe into two parts: things that belong to that dimension and things that do not [51]. Semantics also allows ESEs to understand and correct typing errors, suggest alternative queries (the Google-like “did you mean...?”) and perform an automatic sentence completion while the user is still typing. Semantics allow ESEs to perform entities extraction to locate and classify elements found in a text into predefined categories such as the names of persons, organizations, locations, expressions of times, quantities, monetary values, percentages, etc. This way, the system automatically provides a means to sort out the queries’ result sets. The degree of automation of the process is once again set by the user itself.

To sum up, semantics is what allows machines to get connected with human beings and create a mutual understanding between the two kinds.

Tags

In Computer Science a tag is a key word linked to a document. The tagging procedure allows users to label documents and data with keywords they can use as inputs for search queries. Tags are a further enrichment of metadata and are related to the progression of semantics in the field of enterprise search. Metadata in semi-structured documents bring tremendous value to content search and organization. Metadata can relate to a document (subject, author) or can apply to in-line contents (company, zip code, gene). Once documents are tagged, parametric search or OLAP-style analysis of multidimensional data sets is possible in order to reveal interesting details within a larger corpus [11]. Therefore, tags can be considered as a further tool to help users search for contents inside the intranet. The use of tags has been brought to life along with the concept of Social Bookmarking (SB). SB is the feature of Web 2.0 that allows users to create virtual bookmarks for their web pages. Bookmarks can also be shared with other users, thus empowering the social part of this concept. Anyway, the tagging mechanism can get in contrast with the semantic features. In fact, despite helping users categorize the items, tags are rather personal labels. As a consequence, they can be interpreted in different ways according to the role or the very personality of the user. Indeed, even the ESEs can interpret tags in different ways according to their linguistic and semantic settings. For example, a designer from R&D can tag some parts with the word “Apple”, but he/she may refer either to the popular IT company “Apple Inc.” (if he is designing some new computer components) or just to the fruit (in case he is dealing with a particular part of a harvesting machine).

Collaborative tagging systems help users searching for knowledge and information throughout both structured and semi-structured enterprise data, providing an alternative to the fixed categorization methodology.

Taxonomies

Taxonomies are related and connected to semantics, in the sense that semantics empowers ESEs to organize data into taxonomies. As a matter of fact, taxonomies are navigable categories and sub categories created to assist users in finding relevant information. The data categorization process is particularly important for users, since, as mentioned before, research has shown that presenting results in categories provides better usability [12].

Well-known examples from Web search engines include the directory structures of Yahoo! and the Open Directory Project [11], whereas in the world of ESEs the typical taxonomies are the geographic areas were the content has been created or uploaded in the system.

Taxonomy and syntactic generalization are applied to relevance improvement in search and text similarity assessment [53]. Classification rules for assigning documents into categories can be either manually constructed by knowledge workers or automatically learned from pre-classified training data. Once these classification rules are defined, documents can be automatically assigned to categories from multiple taxonomies.

In this area the state-of-the-art is still young and scientists foresee significant advancements in the field. However, it is reasonable to expect systems to generate taxonomies for refinement by domain experts, making the domain experts dramatically more productive by automatically mining the patterns and discovering associations in the data.

Faceted Search

Despite their importance, taxonomies are getting less popular because of the advance of faceted search. As the term "facet" has been quoted several times in this study, there is a need for a proper explanation. A faceted classification system classifies each information element along multiple explicit dimensions, called facets, enabling the classifications to be accessed and ordered in multiple ways rather than in a single, pre-determined, taxonomic order [16]. As a consequence, faceted search is a technique for accessing a collection of information represented using a faceted classification, allowing users to explore by filtering available information. Faceted search differs from taxonomies because of its dynamic nature: users can progressively narrow their search by applying filters on the way. It is nowadays considered a real alternative to taxonomies.

4.2.3 Ranking the Results of Enterprise Search

The core of ESEs remains the searching function, and a key to have performing search function is to list the results in the proper order. As a matter of fact, the ability to search for information throughout the whole enterprise is meaningless without a way to filter, organize, prioritize, and list the results. As usual, the user expects a list similar to the ones presented by web search engines. Anyway, the factors determining the relevance of search results within the context of an enterprise overlap with but are different from those applying to web search. In the past, ESEs were unable to take advantage of the rich link structure like the one on the Web's hypertext content. Nonetheless, since 2005 new generation ESEs have been more and more based on Web 2.0 technology, thus providing a hyperlink infrastructure inside the enterprise's content. As a consequence, ESEs can now employ the same ranking algorithms used for web search, such as PageRank.

Traditionally, the scoring and ranking of documents is based on the content in each document matching the query. However, the scoring and ranking depends on the kind of document search and, most of all, on the user who is performing the query.

Designers searching for pre-designed parts or components know exactly all of the characteristics of the item they are looking for. It is said that they perform a scoped research. Therefore, they expect the results to be either ranked according to their creation data (usually from the latest to the oldest) or even not ranked at all, i.e. a bare list of all the documents containing the word "Clamp" in their name. Then again, the system can keep track of the historical behavior of users and change the ranking by giving priority to the top searched documents. This adaptive ranking could be simplistic (boost a document's rank if a previous user elected to view/rate it after issuing the same search) or more sophisticated (boost rank if selected from the second results page for a similar query) [11].

On the other hand, a top level CEO would rather perform a query to find out what knowledge the company has about clamps. This is a totally different kind of search, also known as un-scoped search, which is more similar to the ones performed by a Web surfer rather than a vertical user. As a consequence, ESEs are empowered with algorithms such as PageRank. PageRank is an algorithm used by the Web search engine Google to rank websites when presenting the results of the query. PageRank was named after Larry Page, one of the founders of Google. He, along with other scholars [55] defines it as "a method for rating Web pages objectively and mechanically, effectively measuring the human interest and attention devoted to them". The algorithm is patented [56] and is an official trademark of Google, but it has also become a sort of standard in the field, being implemented and incorporated into more complex systems.

According to its inventor, the reason why PageRank is interesting is that there are many cases where simple citation counting does not correspond to our common snotion of importance [55]. Therefore, PageRank is based on a totally different belief: generally, highly linked pages are more "important" than pages with few links [55]. So, the algorithm exploits a hyperlink structure to assign authority to documents, and then it uses that same authority as a query-independent relevance factor.

The following is an intuitive description of PageRank: a page has high rank if the sum of the ranks of its backlinks is high. This covers both the case when a page has many backlinks and when it has a few highly ranked backlinks [55].

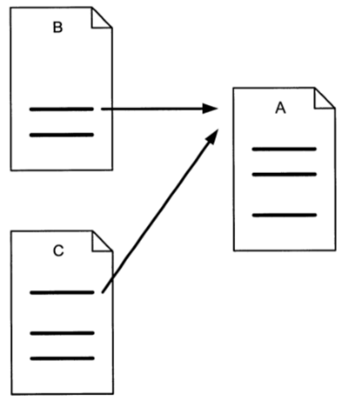


Figure 4.1: A and B are backlinks of C

Given the link structure of the Web, each Web page has some forward links (outedges) and backlinks (inedges) (see Figure 4.1). It is impossible to know all of the inedges of a page, but as soon as that page is downloaded, it possible to know all its inedges.

Therefore, given:

$u =$ web page

$F_u =$ set of pages u points to

$B_u =$ set of pages that point to u

$N_u = |F_u| =$ number of links from u

$c =$ factor for normalization (makes the total rank of all web pages constant)

Then, R is a slightly simplified version of PageRank:

$$R(u) = c \sum_{v \in B_u} \frac{R(v)}{N_v}$$

So, the results are listed from top to bottom according to their ranking score $R(u)$, which acts like a sort of "link counter".

The benefits of PageRank are the best for underspecified queries. For example, a query for "Stanford University" may return any number of web pages which mention Stanford (such as publication lists) on a conventional search engine, but using PageRank, the university's home page is listed as first [55]. However, this first and simplified version of PageRank does not take into consideration other aspects the user might find useful. For this reason, several studies have been performed in order to boost the performances of PageRank. In particular, Haveliwala (2002) points out a possible flaw of PageRank, which is that a single PageRank vector is computed to capture the relative "importance" of Web pages, independent of any particular search query [57].

For this reason, he proposes computing a set of PageRank vectors, biased using a set of representative topics, to capture more accurately the notion of importance with respect to a particular topic [57]. In this way, the query rankings are more accurate.

In contrast to the characteristics of PageRank, enterprises also use other query-independent factors, such as how recent or how popular a document is, along with query-dependent factors traditionally associated with information retrieval algorithms. Also, functionalities of enterprise search such as faceted search and clustering are both meant to let the user build his personal specific ranking, thus bypassing ranking systems. While faceted search has already been explained in detail in paragraph 4.2.2, the study now analyzes the latter functionality: clustering.

4.2.4 Data & Document Clustering

As the study already mentioned, one of the issues regarding queries in ESEs is how to identify relevant query information out of the overwhelming amount of data displayed in return, especially from a broad query. The study has also mentioned how taxonomies and, more recently, faceted search, help users group the retrieved information. However, in order to divide the documents into groups, ESEs need to incluside of algorithms for clustering analysis.

Clustering is a data mining method used to separate a heterogeneous population into a number of homogeneous subgroups without predefined classes. The purpose of clustering is to select elements that are as similar as possible within groups but as different as possible between groups [59]. When this logic is applied to the field of data mining, the experts talk about document clustering. Document clustering (also known as text clustering) automatically organizes a large document collection into distinct groups of similar documents and discerns general themes hidden within the corpus [61]. Following the general principles of clustering, the system divides the data into groups called "clusters", and the documents in the resultant clusters exhibit maximal similarity to those in the same cluster and, at the same time, share minimal similarity with documents from other clusters [61]. These groups are clustered based on entities' similarity according to specified variables and the meanings of clusters depend on the context of the analysis [59].

Various document clustering techniques have been proposed in the literature. However, the most common approach tries to partition the set of N objects in a dataset into C clusters so as to minimize the total within-cluster sum of squares about the C cluster centroids. This simple and intuitively appealing idea forms the basis for a large number of non-hierarchic methods that proceed directly from the object descriptions and that differ only slightly in the exact manner in which they are implemented [60].

Along with the clustering technique, there is another issue related to document clustering: most methods deal with monolingual documents (i.e., written in the same language). Nonetheless, as a result of increased globalization and advances in the Internet technology, an organization often maintains documents in different languages in its knowledge repositories, which necessitate multilingual document clustering to create organizational knowledge maps [61].

Text clustering groups the top several hundred search results into topics that are computed on the fly from the search-results descriptions, typically titles, excerpts (snippets), and metadata. This technique lets users navigate the content by topic rather than by the metadata that is used in faceting. Furthermore, clustering compensates for the problem of incompatible meta-data across multiple enterprise repositories, which hinders the usefulness of faceting [49].

The need to create software to perform and manage text clustering first came to life in the world of Web search engines. In order to address the challenge of web-scale data management and analysis, in the late 1990s Google created Google File System (GFS) and MapReduce programming model. In the mid-2000s, user-generated content, various sensors, and other ubiquitous data sources produced an overwhelming flow of mixed-structure data, which called for a paradigm shift in computing architecture and large-scale data processing mechanisms. As a matter of fact, the input data is usually large and the computations have to be distributed across hundreds or thousands of machines if they are to finish in a reasonable amount of time [62]. NoSQL databases, which are scheme-free, fast, highly scalable, and reliable, began to emerge to handle these data. Several studies also argued that the only way to cope with this paradigm was to develop a new generation of computing tools to manage, visualize and analyze the data deluge [49]. Indeed, Google was ready to face the challenge thanks to GFS and MapReduce, which can enable automatic data parallelization and the distribution of large-scale computation applications to large clusters of commodity servers. A system running GFS and MapReduce can scale up and out and is therefore able to process unlimited data [49].

Google File System

GFS is a proprietary file system created by Google cofounders Larry Page and Sergey Brin. Google generates an enormous amount of data and therefore needs unconventional distributed file systems in order to store and manage them. GFS is organized in a computer cluster composed of two kinds of nodes:

- Chunk: a server that stores a kind of file called "chunk". Chunks are 64MB files, similar to the files of standards file systems.
- Master: a server that stores all the metadata linked to the chunks, such as the maps of the position of the files, the recovered files or the copies of the chunks.

MapReduce

MapReduce is a programming model and an associated implementation for processing and generating large data sets [62]. The algorithm of the MapReduce library expresses the computation as two functions:

- A *Map* function, that takes an input key/value pair and produces a set of *intermediate* key/value pairs. The MapReduce library groups together all intermediate values associated with the same intermediate key and passes them to the *Reduce* function.
- A *Reduce* function, also written by the user, accepts as input an intermediate key and a set of values for that key. It merges together all these intermediate values associated with the same intermediate key to form a possibly smaller set of values. Typically, just zero or one output value is produced per *Reduce* invocation. The intermediate values are supplied to the user's reduce function via an iterator. This allows us to handle lists of values that are too large to fit in memory.

The core idea of MapReduce is that data are first divided into small chunks. Next, these chunks are processed in parallel and in a distributed manner to generate intermediate results. The final result is obtained by aggregating all intermediate results. This model schedules computation resources close to data location, which avoids the communication overhead of data transmission

[49]. As a consequence, MapReduce fits perfectly into the GFS, which provides the programming model with data and metadata pre-divided into chunks. The MapReduce model is simple and widely applied in bioinformatics, web mining, and machine learning [49].

Programs written in this functional style are automatically parallelized and executed on a large cluster of commodity machines. The run-time system takes care of the details of partitioning the input data, scheduling the program's execution across a set of machines, handling machine failures, and managing the required inter-machine communication. This allows programmers with no experience in parallel and distributed systems to easily utilize the resources of a large distributed system [62].

The framework of MapReduce is implemented in a master/worker configuration. It consists of a single master *JobTracker* and one worker *TaskTracker* per cluster node. The master is responsible for scheduling jobs for the workers, monitoring them and re-executing the failed tasks [49]. The slaves execute the tasks as directed by the master. A worker may be assigned a role of either *map worker* or *reduce worker* [6i]:

Step 1) *Input Reader:* The first step, and the key to massive parallelization in the next step, is to split the input into multiple pieces. Each piece is called a *split*, or *shard*. For *M* map workers, we want to have *M* shards, so that each worker will have something to work on. The number of workers is mostly a function of the amount of machines we have at our disposal.

The MapReduce library of the user program performs this split. The actual form of the split may be specific to the location and form of the data. MapReduce allows the use of custom readers to split a collection of inputs into shards, based on the specific format of the files.

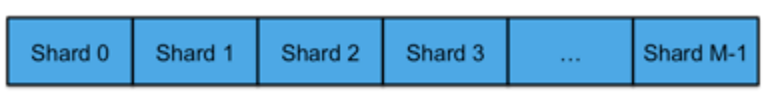


Figure 4.2: Split input into shards

Step 2) *Fork Processes:* The next step is to create the master and the workers. The master is responsible for dispatching jobs to workers, keeping track of the progress and returning results. The master picks idle workers and assigns them either a map task or a reduce task. A map task works on a single shard of the original data, a reduce task works on intermediate data generated by the map tasks. In all, there will be *M* map tasks and *R* reduce tasks. The number of reduce tasks is the number of partitions defined by the user. A worker is sent a message by the master identifying the program (map or reduce) it has to load and the data it has to read.

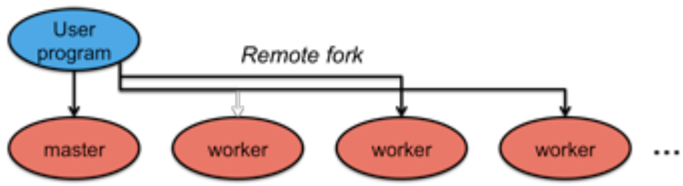


Figure 4.3: Remotely executed worker's processes

Step 3) *Map:* Each *map* task reads from the input shard that is assigned to it. It parses the data and generates *(key, value)* pairs for data of interest. In parsing the input, the *map* function is likely to get rid of a lot of data that is of no interest. By having many map workers do this in parallel, we can linearly scale the performance of the task of extracting data.

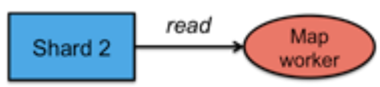


Figure 4.4: Map task

Step 4) *Partition:* The stream of *key/value* pairs that each worker generates is buffered in memory and periodically stored on the local disk of the map worker. This data is partitioned into *R* regions by a partitioning function.

The *partitioning function* is responsible for deciding which of the *R* reduce workers will work on a specific key. The default partitioning function is simply a hash of *key* modulo *R* but a user can replace this with a custom partition function if there is a need to have certain keys processed by a specific reduce worker.

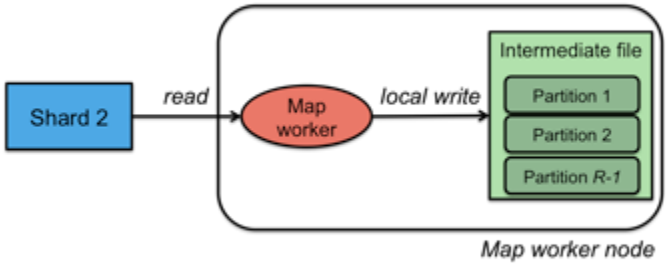


Figure 4.5: Create intermediate files

Step 5) *Sort / Shuffle:* When all the map workers have completed their work, the master notifies the *reduce workers* to start working. The first thing a reduce worker needs is to get the data it needs to present to the user's *reduce* function. The reduce worker contacts every map worker via remote procedure calls to get the *(key, value)* data that was targeted for its partition. This data is then sorted by keys. Sorting is needed since it will usually happen that there are many occurrences of the same key and many keys will map to the same reduce worker (same partition). After sorting, all occurrences of the same key are grouped together so that it is easy to grab all the data that is associated with a single key.

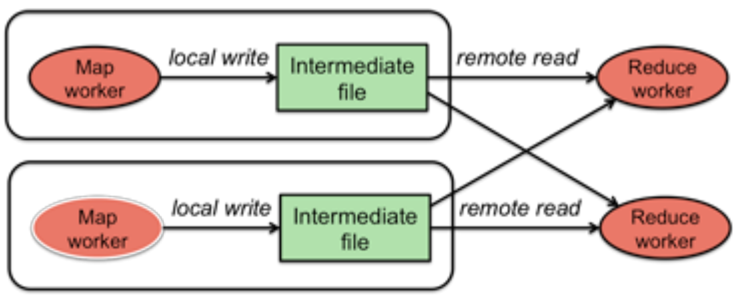


Figure 4.6: Sort and merge partitioned data

Step 6) *Reduce:* With data sorted by keys, the user's *Reduce* function can now be called. The reduce worker calls the *Reduce* function once for each unique key. The function is passed two parameters: the key and the list of intermediate values that are associated with the key.

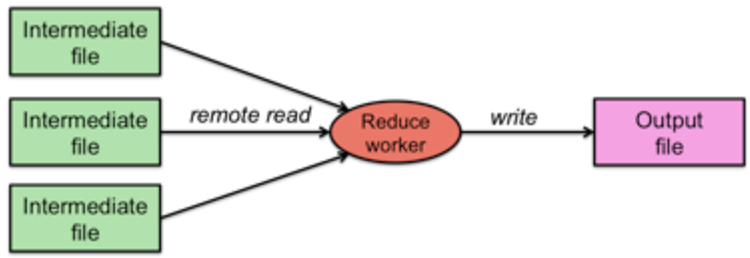


Figure 4.7: The Reduce function writes the output

Step 7) *Output:* When all the reduce workers have completed execution, the master passes control back to the user program. The output of MapReduce is stored in the *R* output files that the *R* reduce workers created.

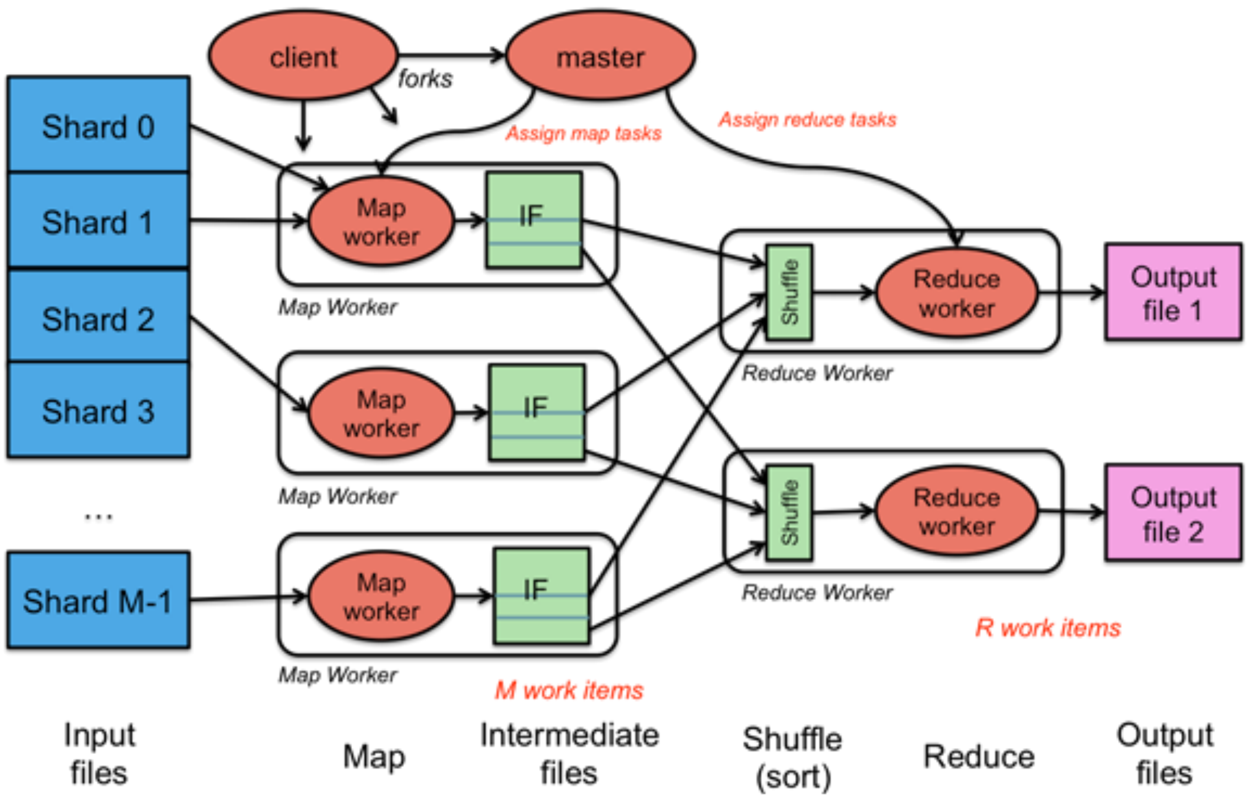


Figure 4.8: The framework of MapReduce

Figure 4.8 illustrates the entire MapReduce process. The client library initializes the shards and creates map workers, reduce workers, and a master. Map workers are assigned a shard to process. If there are more shards than map workers, a map worker will be assigned another shard when it is done. Map workers invoke the user's *Map* function to parse the data and write intermediate *key/value* results onto their local disks. This intermediate data is partitioned into *R* partitions according to a partitioning function. Each *R* reduce worker contacts all of the map workers and gets the set of *key/value* intermediate data that was targeted to its partition. It then calls the user's *Reduce* function once for each unique key and gives it a list of all values that were generated for that key. The *Reduce* function writes its final output to a file that the user's program can access once MapReduce has completed.

MapReduce was one of the first technologies developed for data and document clustering. Therefore, it has been improved and many times replaced by more advanced and less disk-oriented mechanisms, but it still remains a good reference for managing cluster computing all around the world.

4.2.5 Big Data Analytics

Big data analytics is the process of using analysis algorithms running on powerful supporting platforms to uncover potentials concealed in Big Data, such as hidden patterns or unknown correlations. Over the years, this kind of process has been incorporated in ESEs. The reason behind the merging is that ESEs are the ultimate data collectors inside the enterprises, so they are a natural source for useful data to be displayed for users' inspection. According to the processing time requirement, big data analytics can be categorized into two alternative paradigms [49]:

Streaming Processing

The starting point for the streaming processing paradigm is the assumption that the potential value of data depends on data freshness. Thus, the streaming processing paradigm analyzes data as soon as possible to derive its results. In this paradigm, data arrives in a stream. In its continuous arrival, because the stream is fast and carries enormous volume, only a small portion of the stream is stored in limited memory. One or few passes over the stream are made to find approximation results [49]. Commonly, according to the streaming processing paradigm, information has to be processed through pipeline structures so that the processing time is significantly reduced. Streaming processing theory and technology have been studied for decades. Representative open source systems include Storm, S4, and Kafka. The streaming processing paradigm is used for online applications, commonly at the second, or even millisecond, level. [49]

Batch-Processing

In the batch-processing paradigm, data are first stored and then analyzed. It is a paradigm that fosters the automatic execution of programs on a computer. The input data are collected into *batches*, then each *batch* is processed as a unit. This paradigm is the opposite of the one regulating interactive programs, which always require an input from the user. MapReduce has become the dominant batch-processing model.

At the end of data processing, the final purpose of Big Data analytics tools is merging the results collected from the sources and presenting them in a succinct and unified format with minimal duplication.

4.3 Image and Shape Searching

Following the needs mentioned in Chapter 3, ESEs need some specific adjustments in order to support designers and engineers creating a product. First and most important is the possibility of searching for 3D shapes, which are the basis of modern virtual prototyping.

Over the last four decades, scientists have tried to develop Artificial Intelligence (AI) techniques to enable computers to perform the same functions as the Human Visual Perception System, that is to determine similarities among 3D shapes from a large database of 3D shapes [50].

The 3D shape searching area has so far been dominated by research in vision and computer graphics, where researchers have extensively studied the 'shape matching' problem. However, CAD and engineering applications of 3D shape searching warrant considerations that go beyond the shape matching problem. For example, domain knowledge of manufacturing processes, cost and material selection play a very important role in the search process [50].

The simplest form of searching is by keyword in filenames, part numbers, or context attached to the CAD model. Product Lifecycle Management (PLM) systems allow part-name based searching of 3D models. However, this method is not robust, primarily for the following reasons:

- All models will not have a well-defined attached design/manufacturing context.
- Keywords such as project names or part names may be unknown to the user.
- The context may be too narrow or too broad to retrieve relevant models.

- The context changes with time, such as when designers or naming conventions change.

A significant amount of knowledge generated during the design and manufacturing phases is associated with the 3D model because of the growing interdependence of PLM, MRP and CAD systems. Most of this knowledge is geometry-related (manufacturing process details) or even geometry-dependent (analysis results). Therefore, a search system that is capable of retrieving similar 3D models based on their shape will also retrieve related knowledge that cannot be discovered by other means [50].

A significant amount of work has been done in the past two decades on text-based document retrieval. The Google Web search engine has become a de facto standard as a text-based search engine. More recently content-based retrieval systems have been developed for images, audio, and video. Content-based retrieval systems retrieve objects based on the integral similarity of objects. Retrieval by integral similarity totally differentiates content-based systems from typical search systems, which retrieve objects by some specific attributes of the objects [50].

According to Iyer et al [50], 3D shape searching techniques based on shape representations can be classified into the following categories: global feature-based, manufacturing feature recognition-based, graph-based, histogram-based, product information-based and 3D object recognition-based.

4.1.1 Global feature-based techniques

Global feature-based methods use global properties of the 3D model such as moments, invariants, Fourier descriptors, and geometry ratios. The key limitations of global feature-based methods are that they fail to capture the specific details of a shape and are not very robust, and fail to discriminate among locally dissimilar shapes.

While being computationally efficient, Global feature-based methods are unable to discriminate dissimilar shapes.

4.1.2 Manufacturing feature recognition-based techniques

The CAD community has worked on automated feature extraction since the '1980s, much earlier than the computer graphics community. An automated feature recognition program extracts instances of manufacturing features from an engineering design done with a CAD program or solid modeler. Most feature recognition techniques use a standard library of machining features.

Some techniques for feature recognition are rule-based, graph-based, and neural network-based. The shape representation of a 3D model is determined by various properties of extracted machining features. Most techniques are not robust in extracting features successfully and remain a significant challenge for research.

Manufacturing feature recognition-based methods do not decompose shapes uniquely. Additionally, they may require human intervention.

4.1.3 Graph-based techniques

The topology of 3D models is an important shape characteristic. Topology is typically represented in the form of a relational data structure such as graphs and trees. Subsequently, the similarity estimation problem is reduced to a graph or tree comparison problem. While tree comparison tends to be fast and easy compared to graph comparison, most engineering components cannot be represented as trees. On the other hand, graph comparison costs increase proportionally with graph size. Topological graphs can be matched based on exact or inexact matching techniques. While exact matching looks for noiseless matches between two graphs, inexact matching results in a measure of similarity between the graphs even in the presence of noise. However, the major advantage in representing 3D models as topological graphs is that it allows representation at multiple levels of details and facilitates matching of local geometry.

Some 3D shape matching approaches determine similarity based on the boundary representation (B-Rep), which is represented as a graph. Graph matching algorithms are employed to determine similarity between corresponding B-Rep graphs. However, a host of other approaches have emerged that convert the surface representation of a 3D model into a simpler topology preserving representation. These include Reeb graphs, shock graphs and skeletal graphs. All of these methods yield smaller graphs, thereby allowing faster comparisons, as compared to B-Rep graphs. However, these representations may often be an over- simplification of shape, leading to problems in pruning large 3D databases. Other approaches derive graph invariants such as number of nodes and edges, degrees of nodes, and eigenvalues (spectral graph theory) directly from the B-Rep graph for quick shape comparison.

4.1.4 Histogram-based techniques

Histogram-based techniques sample points on the surface of the 3D model and extract characteristics from the sampled points. These characteristics are organized in the form of histograms or distributions representing their frequency of occurrence. Similarity is determined by a comparison of histograms by means of a distance function. The accuracy and effectiveness of histogram-based techniques depend on the number of sampled points. A larger number of sampled points result in higher accuracy. However, the efficiency is inversely related to the number of sampled points.

4.1.5 Product information-based techniques

The techniques encompassing product information-based systems are specifically designed for the domain of engineering parts. Part designs are described based either on their manufacturing attributes or on their geometry. Although group technology-based representation schemes address both issues, they require the user to describe the shape of the part based on its drawing. On the other hand, 2D section image-based methods try to eliminate user input to a large extent by classifying the parts based on their 2D image attributes. Image processing techniques along with neural networks have been used to compare and cluster parts based on their silhouette images.

4.1.6 3D object recognition-based techniques

Three-dimensional object recognition techniques have been studied extensively by the computer vision community. Several methods have been developed for 3D object recognition. Some of them are based on aspect graphs, extended Gaussian images, superquadrics, spin images, and geometric hashing.

Topological graph-based techniques are intractable for large graphs because of the high complexity of graph/ subgraph matching problems. Skeletal graph-based methods are not applicable to all kinds of shapes. Histogram-based methods have a tradeoff between computational cost and number of sampled points. Sampling a lower number of points leads to very low accuracy. Product information-based techniques are not robust or require extensive human input. 3D object recognition techniques have been tested for limited shapes and have high storage/computational costs. In summary, each technique has its own advantages and disadvantages. The choice of the technique depends on the type of shape being searched and on the level of discriminating power required by the application.

5 EXALEAD Market Solution Example

After listing the basic terminology, the needs of Enterprises and a brief history of ESEs, this study presents Exalead as an example of state-of-the-art ESE serving the PD process.
Exalead was first of all a company founded in 2000 by François Bourdoncle and Patrice Bertin. The company was acquired by Dassault Systèmes (3DS) in 2010 for €135 million.

5.1 Exalead – Logic & Framework

Exalead appears on the computer's monitor as a widget, just like any other program installed on the machine. Once running, Exalead opens a web-browser page and the interface presents a central search bar just like the ones of the web search engines, though adding search suggestions at the bottom. Also, the page is made to be customized by the user.

5.1.1 Exalead - Framework

The following is a logic scheme of the architecture behind Exalead.

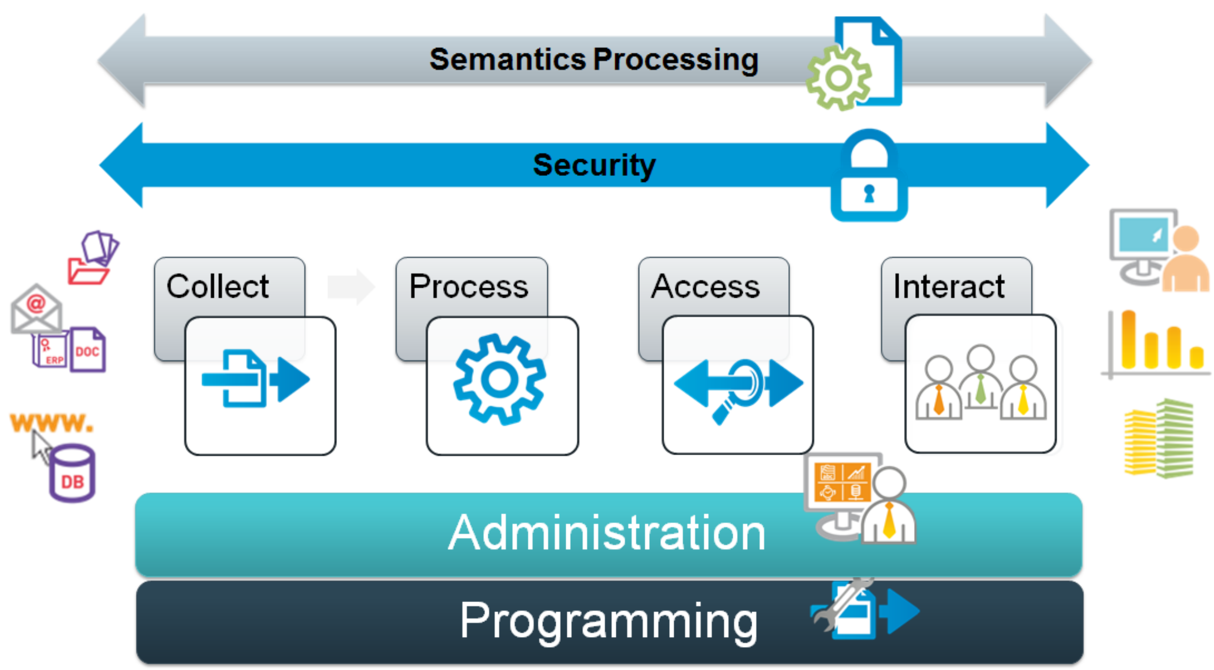
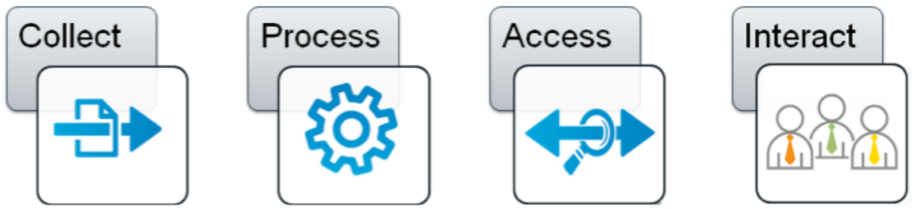


Figure 5.1: Exalead's framework

It is easy to divide it into three main parts: Core Workflow, Support Infrastructure, Master Infrastructure.

5.1.2 Core Workflow



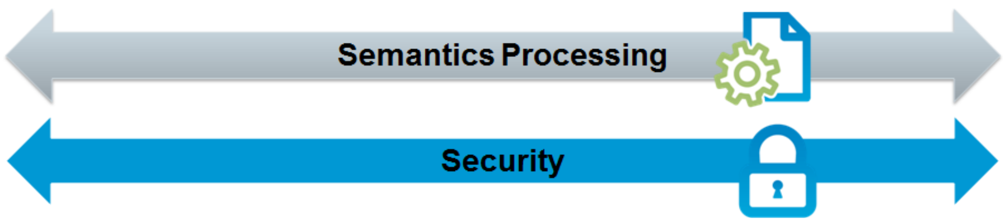
The Core Workflow is the group of the main activities performed by the ESE.

- **Collect**: the first activity is Data Collection. Using a group of pre-set connectors, Exalead performs an intelligent extraction of complex structured data and associated metadata (attributes, rules, relationships, etc.) from every indexed data source: the company database, the file system, the internet, the email archive or other software systems such as the ERP or the CRM. The fact that connectors are preset does not mean that they are fixed. It is actually possible to set new connectors and thus to customize the data sources to be taken into account.
- **Process**: the second step consists of processing the collected data. When executing this activity, Exalead manages to sort, summarize, aggregate and analyze the data collected during the "Collect" step. At this point, Exalead is still working at a shallow level, meaning it does not enter the single item yet. The entire processing activity is performed on data as files, as clusters of information, but Exalead does not look into the file content.
- **Access:** finally Exalead opens the processed items and looks into them in order to have direct access to the content. In this phase Exalead checks the user's right to access the content. If the access is verified, Exalead connects directly to the proper application in order to open the item and look into it. For example, when the user inputs to search for "clamp", Exalead searches for:
    1. All the documents whose title contains the word "clamp"
    2. All the files that that contain references to clamp: emails in which the supervisor discusses the latest clamp version, CAD files with clamp geometries included, word documents with clamp desciptions and so on.

Step n. 2 is performed only when Exalead recognizes the user's right to **access** the content of the item.

- **Display:** the last step is showing the user the result of Exalead's work. So, Exalead presents the results of the search on the monitor depending on the kind of user, the role he/she has in the company and the level of detail he/she needs. This is the first step in which the user is shown the activity, therefore it is also where customization plays a key role. Exalead has different options for data displaying: bare list of items, data charts, CAD model previews and so on.

5.1.3 Support Infrastructure



The logical scheme of Support Infrastructure highlights two features that support and empower the Core workflow. As the arrows show, these features are transersal and pervasive in the whole infrastructure.

Semantics Processing: Exalead is enriched with semantics-sensitive technology. It is a key feature in terms of user-friendliness and it includes several sub-features and tools studied and realized to make the tool more human-like. For this reason, over the years ESEs have been improved with semantics-sensitive technology, with the result of bringing them closer to users. Indeed, the purpose was to make the user like the tool and, at the same time, make the tool open to every kind of user, not only to data mining experts. As stated in Chapter 3, every user wants the ESE to understand as much as possible of the human language, so that the search would be facilitated. Exalead exploits this technology by offering the so-called Exalead Semantic Factory, a Semantic ETL (Extract, Transform and Load) of the "grey" data that constitutes the bulk of modern Big Data:

- Unstructured contents like documents, emails, call recordings and videos
- Semi-structured data such as HTML and XML, including social media contents

- Small but voluminous structured contents such as the machine data produced by smart meters, RFID readers, barcode scanners, weblogs and GPS tracking units

The semantic ETL processors within Exalead automatically structure unstructured information, and enrich semi-structured data. This is the list of all the specific activities performed by the Semantic ETL:

- Semantic data normalization (fuzzy data consolidation, cleansing, metadata alignment)
- Analog-to-digital conversion: voice-to-text, optical character recognition (OCR), etc.
- Qualitative analytics (sentiment analysis)
- Semantic query interpretation
- Fully modular and configurable pipeline
- Easy integration of business rules
- Multi-lingual support (124+ languages out-of-the-box; optional Extended Language pack for additional processors for 38 of these languages)

This translates into several advantages for the users.

First, there is no need for search strings. Once the user starts typing characters, Exalead suggests different possibilities, just as a web search engine (Query Suggestion). Second, semantics allow the software to be “smarter” when searching for certain data. In other words, Exalead is capable of interpreting what the user is looking for even if it is misspelt or incomplete. For example, when the user looks for all models of “clamp”, but he types “camp”, Exalead shows a few suggestions in order to adjust to the user’s mistake. Third, by merging structured and unstructured data the user has a unified view of all the information regarding his search. In other words, all items appear in a single page and they are all listed according to how well they semantically fit the search.

Security: Exalead guarantees a certain level of security inside the company. Not all users can have access to all data and not all data can be disclosed. Through the security system, each role is locked to a unique username and password. This study already mentioned the protection system when dealing with the *Access* activity of the *Core Workflow*. Actually, security is present in the entire software system and acts at different levels. The first matter regards secure access to files and is indeed connected to the *Access* activity. Then, during the process of generating results, a secure search engine needs to match the authenticated user's access rights with the retrieved documents' access rights defined in the Access Control Lists (ACLs).

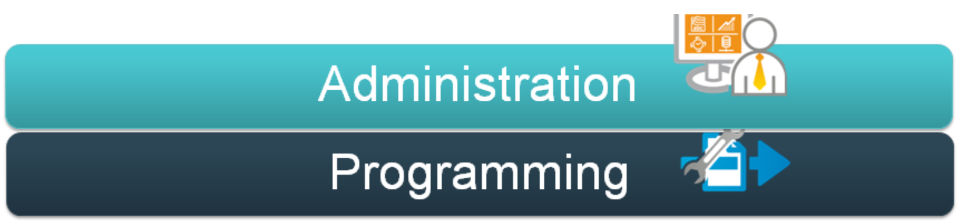
To determine a user's read rights for a particular resource when executing a search query, there are two basic options, each with a different impact on latency:

- Real-Time ACL Checking (sometimes referred to as "unmapped security" or "post-query filtering"): the search platform checks the ACL at query time for each individual search result returned by a query.
- ACL Indexing (also known as "mapped security" or "ACL mapping"): at indexing time, information regarding users and their rights is appended as metadata to the index entry for a document. Consequently, at query time, the search platform simply checks the metadata attached to each index entry rather than checking with the source application's ACL.

Because Real-Time ACL Checking is hard to manage and puts a heavy load on the server and incumbent systems, Exalead uses ACL Indexing as a standard approach. It is the fastest, most scalable option. It also offers a very high level of granularity.

However, there is latency in permission changes with ACL indexing, that may be unacceptable in certain contexts. Therefore, Exalead can optionally combine the two ACL checking methods to optimize performance while ensuring a responsive approach to changes in security access rights [3i].

5.1.4 Master Infrastructure



The Master Infrastructure stands above the Core Workflow and the Support Infrastructure. It is the tool for control, the interface with the software programmer and the control panel for all settings. It consists of two parts:

Programming: the activity of installation and indexing of Ealead. As stated before, like all other ESEs, Exalead also needs to be properly installed in a software environment (installation) and then set so that it can search for information in the intranet. The latter is the phase called "indexing". It consists of opening the control panel of Exalead and pointing at the various Data Sources in which the ESE goes looking for Data to retrieve. In order for Exalead to communicate with the Data Sources, it needs specific programs called "connectors". Packaged Exalead connectors are available for most enterprise sources, including file servers, eXtensible Markup Language (XML), Hadoop Distributed File Systems (HDFS), databases, email systems, directories, content management and collaboration systems, including all the PDM systems available on the market. In particular, a solid connector to the PDM software system allows Exalead to access all files from CAD, CAE, CAM software systems hence to properly support the PD process. Exalead extends this connectivity with advanced support for dozens of Big Data sources, including enterprise applications, data warehouses, Business Intelligence platforms, mainframes, NoSQL stores, real-time message queue data and social media networks. All of these packaged connectors are complemented by a robust Application Programming Interface (API) for connectivity to obsolete or custom (bespoke) repositories using standard protocols and languages (HTTP/Rest, Java, C#, etc.). The correct functioning of Exalead during its lifecycle mainly depends on correct *Programming*. In this phase the programmer implements the user's requests in terms of customization.

Administration: regards the everyday activity of controlling and maintaining Exalead, and is dedicated to users rather than software experts. A browser-based administration panel enables the user to add new resources and refine the index as needed. A statistics module further provides important feedback on users' needs and interests, reporting, for example, the most frequently consulted documents and top search requests. The *Administration* tool is used to monitor, control and customize Exalead in order to suit the changing of users' needs. The *Administration* function is performed by the Exalead Mashup Builder, a drag'n'drop interface for rapid prototyping and the deploying of search-based discovery applications. The user simply needs to drag and drop feeds from internal or external data sources into the Mashup Builder, then drag and drop multi-channel design widgets for search, presentation and/or reporting into a Web page template.

5.2 Exalead - Customers' Benefits

After the detailed description of the infrastructure and the logic behind Exalead, it is now possible to draw a matching between the state-of-the-art offering and the previously listed customers' needs.

5.2.1 User Friendliness

The first step towards user friendliness is designing the ESE to look as similar as possible to a web-based search engine. As a consequence, the Google-like search bar and the browser navigation page are both adjustments to make the entire system more similar to the world outside the enterprise. Actually Exalead is not a trendsetter in this subject. The first enterprise software systems to be web based are PDM systems, due to the exact same reason: user friendliness. The development of web-based front-end software has made PDM Systems more user-friendly and easier to interact with. As such, the web-based PDM systems require minimal training in comparison to other systems; hence the overall cost of system implementation is lower. Simple, inexpensive web browsers continue to expand the use of PDM to many people in the enterprise, who can barely find other equally convenient on- line access to such a broad information base. By enabling users to become active quickly, browsers provide a `jump start' in achieving early success with PDM in organizations that might otherwise lose enthusiasm [6]. Furthermore, the use of these web browsers can eliminate the need to install and maintain specialized `client' software on each machine [6]. User friendliness is also evident in the software installation and management. The outlook of a traditional widget helps the user getting accustomed to it, while customization makes the system suitable to the operator. First, the user interface includes a panel with optimization and debugging tools for the administration. Through the panel, the user can access the expansible library add to it some other out-of-the-box visual dashboarding widgets besides the ones included. Secondly, the more expert users can have two tools for more technical customization:

- Mashup Builder: an interface with a wide Layout Editor for creating complex page compositions, simplified menus, and live previews of widgets for application fine-tuning.
- Mashup Expression: an interface to modify language settings, pre-set syntax, semantics and facets.

There are also settings to help the user handle the integration with the rest of the PD software systems. As a matter of fact, the user can set a system of real-time alerts, with notifications triggered as soon as information matching stored queries are indexed. Flexible standard and custom options are available for publishing notifications: emails, enterprise communicator messages etc. To conclude with the administration aspects, Exalead employs diagnostic tools to help administrators detect and respond to configuration issues and ontology-driven data modeling for accelerated development (automatically applied to all underlying analysis, indexing and search configuration processes).

Regarding the searching activity itself, Exalead proposes semantic interpretation of natural language queries and query suggestions, auto-spelling correction, type-ahead query, fuzzy matching of results and faceted navigation. Specifically, Exalead employs massively scalable faceting (categorization and clustering) of textual, numerical, data and geographical values to allow users to explore a collection of information by applying multiple filters.

Another Exalead feature to empower non-specialist users to perform exploratory analytics on Big Data is the ability to operate in both "pull" and "push" modes.

In the "pull" mode, Exalead's semantic mining tools are used to identify the embedded relationships, patterns and meanings in data, with visualization tools, facets (dynamic clusters and categories) and natural language queries used to explore these connections in an ad-hoc manner.

In the "push" method, users can sequentially ask the data for answers to specific questions, or instruct it to perform certain operations (like sorting), to see what turns up.

It is an approach that allows ordinary users to tap into the potential of exploratory analytics: sampling it all in, seeing what shows up, and, depending on the situation, either acting on their discoveries or relaying the information to specialists for investigation or validation. Users can also integrate the preset analytics with contextual data (for example geo location), role- and rule-based views, user ratings and tags. The tagging mechanism is bound to the user's preferences, so there is no feature for automatic tagging. In this way, subject matter experts can be hired to tag or annotate documents manually. The main advantage of Exalead reflects exactly the needs of the user. As a matter of fact, the user has to be at ease in his working environment, and he wants to decide how to divide his material into groups and sub-groups of objects that are functional to his own stage of PD or his working habits, while automatic tagging is still too impersonal and restrictive. In the end, the tag mechanism's purpose is to help users. Manual tagging, however, does not scale to large volumes of information [11]; therefore, a certain level of automation is mandatory. Inside Exalead, tags are presented as virtual labels the user can attach to items searched, so that they further characterize the item. The user can then search for a tag instead of a name or a feature.

The final scope is providing a tool to search and explore extremely large data collections without using complex SQL queries, BSON objects or Map/Reduce functions.

5.2.2 Data Merging

Exalead provides specific solutions for this particular need. The already mentioned semantic processing pipeline meaningfully structures and enriches unstructured contents and correlates it with structured data. In detail: in the latter case, the Search-Based Application (SBA) index replaces a traditional relational database structure as the primary vehicle for information access and reporting, thus enriching structured data with valuable meaning from unstructured data. In the case of semi-structured data such as XML records and machine data produced by smart meters, RFID readers, barcode scanners, weblogs and GPS tracking units and unstructured content like documents, emails, call recordings and videos, Exalead uses the tools mentioned as semantic ETL processors. They automatically structure unstructured information and enrich semi-structured data, so that all the data out of the query result in a single piece of information. In the end, Exalead provides a merging of structured and unstructured data ready to be displayed as a whole.

5.2.3 Security

The study has already mentioned that Exalead has strict control over the roles and access rights of its users. In a sensible business process such as PD this is the first step to guarantee security over information inside the enterprise, but the second and even more important step for security is the possibility of enforcing the source-system rights. Every enterprise PDM has its security settings, and Exalead accesses it without making any change or modification. In other words, Exalead works above the PDM software systems, because it collects information without compromising it or breaking its protocols. In this way, being Exalead locked to each role, data collection becomes automated and non intrusive.

5.2.4 Presentation Capabilities

As far as the user interface is concerned, Exalead delivers a fully unified view of information, providing uniform access and reporting for heterogeneous data. The Exalead indexing module uses statistical, linguistic and semantic processing to transform heterogeneous, multi-source data flowing from the Connectivity and Semantic Factory modules into a unified base of actionable information intelligence. The Exalead index stores and maintains this data and processes incoming user's application queries against it. The module further performs qualitative and quantitative analytics on textual, numerical, temporal and geographical values, both during indexation and on-the-fly at query time. The final result of the search integrates intelligence from all relevant sources and supports both qualitative and quantitative analytics with a single platform. As a matter of fact, real time analytics play a key role in the results presentation, as well as the sub-division of results into categories. Searches within a category typically produce higher relevance results than unscoped searches [11], so the above-mentioned faceted search is also an important feature when showing the results to the final user. Research has shown that presenting results in categories provides better usability [12], so Exalead employs facets to divide the search into categories and then show it to the user that is not necessarily a business analyst. In particular, Exalead's presentation capabilities make it a tool that is also useful for the Sales and Marketing departments, thanks to dynamic pivot-style tables that join facets along multiple axes (e.g., Sales, Region, Date, Products, Vendors…) and Query-time definition of virtual fields (e.g., computing Sales on Unit Sold x Unit Price). To conclude, Exalead provides out-of-the-box content recommendation capabilities, the ability to automatically push specific content based on business rules and user context and behavior, thus making it a tool for connection between PD units and the rest of the enterprise.

5.2.5 Integration

Last but not least Exalead can connect with the rest of the PD software infrastructure and the rest of the enterprises intranet. In the previous paragraphs this study has presented Exalead's connectors and the way they let Exalead offer feed management, including tools to join, merge, and cascade feeds from all over the company (email alerts, push notifications, documents uploading etc.). The most important feature remains the integration with the enterprise's PDM system. Since Exalead is mainly a toll to help product developers, it is designed not only to be integrated with both commercial and custom PDM software systems, but also to be upgraded with a Java plugin framework to enrich all platform capabilities, so that it is virtually possible to seamlessly develop, test and deploy plugins. It even enables advanced users to go further in building robust Search-Based Applications from multi-source contents. These integration capabilities are designed to meet an increasing number of expert users. Again, users are more and more accustomed to software out of the enterprise. They are used to install apps on their smartphone or to modify their Google home page with their favorite widgets. The real challenge of integration is not to design an ESE that can be integrated with existing software but to make in such a way that it can be integrated with what will be designed in the future, be it commercial or custom made.

5.3 Offering

The framework put forward in the previous paragraphs is but the architecture that Exalead's programmers decided to adopt in order to create an ESE. The technology needs then to be reordered in an offering portfolio, so that the ESE can address the technology towards the specific requirements of the users.

In particular, ESEs are sold in modules. An ESE's module is a vertical market software frame, a particular tool created to shape the enterprise search technology in order to fit the specific needs of the target market.

Here follows the list of Exalead's modules [2i]:

5.3.1 EXALEAD CloudView

EXALEAD CloudView is an infrastructure-level search and information access platform used for both online and enterprise search. CloudView is the real core of the enterprise search technology embedded in Exalead. It provides the user with a tool to capture Big Data from all around the Enterprise, an instrument to aggregate them and a repository for their exploitation. It is the module for the capturing function. CloudView includes all the connectors with the rest of the software infrastructure and the Web.

It is also the scalable infrastructure on which four more Exalead's modules can be installed.

5.3.2 EXALEAD OnePart

EXALEAD Onepart is a solution for designers and engineers, allowing them to search for designs, documentation, specifications, standards, test results and related data for engineering, manufacturing and procurement activities. Also, once all the documents are correctly classified, this module is set to de-duplicate the existing parts. The final aim of this module is providing a tool to reuse similar or equivalent data instead of creating new ones from zero. As a matter of fact, it is programmed to classify and de-duplicate existing designs and documentation to save on stock and data storage costs. It is the module that includes search refinement tools such as faceted search, tags and parametric refinement.

5.3.3 EXALEAD Onecall

EXALEAD Onecall is a module for costumer relation and customer service. Its focus is on data collection from multiple sources, data merging and presentation. Onecall can connect to internal and external data type (CRM, ERP, Social Networks etc.) and aggregate information in a single desktop.

5.3.4 EXALEAD PLM Analytics

This module is programmed to provide the user with an interface to access and operate on top of a PLM system. PLM Analytics is a tool for process tracking and to process real time analysis. In fact, it collects all the information from the PLM in a single desktop and runs pre set KPIs so that the user has a real time view of the advancing process. However, this module is only a collector, an aggregator and interface, whereas the tracking and calculation activities are performed by the PLM. The module was invented to provide high-level data model knowledge to build analytics on top of a PLM authoring system.

5.3.5 Web Indexing

Web Indexing is the module to build the users' personal index of the Web. The crawler and semantic pipeline of CloudView, combined with the Web search engine and content APIs allows the user to build a targeted index of the Web.

6 Model

It is clear that the state-of-the-art technology provides a valid tool to fill the gap in the traditional enterprise software architecture for Product Development. This study has proved that an ESE (be it Exalead or any other commercial product) provides companies with an effective tool to support the designing and engineering of products. Indeed, ESEs are to be considered part of the software architecture dedicated to PD.

6.1 The Logic Behind the Model

6.1.1 Literature Gap

There are still a number of unanswered questions both in the literature and in the enterprises' know-how. In particular, the problem can be divided into three points:

1. Despite listing all the advantages that a powerful ESE brings and reporting the progress of enterprise search technology, no real trace can be found in the literature of a link between ESEs and PD. As stated in Chapter 3 and Chapter 4, studies link ESEs with the ERP and all of the software systems for the management of an enterprise. In Chapter 5, we also put forward evidence that the technology behind ESEs can be exploited to help designers in their daily job. However, we know from Chapter 3 that PD is a lot more than just sketching projects and virtual-testing components: there is a whole software infrastructure that needs to be accessed by different people from different departments, and ESEs could be the key to access and navigate that structure. Nonetheless, none of the previous studies applies ESEs to the whole PD process explained in the BAH Model, but they rather consider it as a stand-alone system.

2. The scholars' studies as well as enterprises' reports list uncountable advantages following the implementation of an ESE, but there is no mention of a standard method on how to convince a customer of its value. Obviously, every company carries out its own specific PD process and every company has a different supporting infrastructure. However, having a list of undeniable advantages and one of common enterprises' needs, there must also be a systematic way to match the two.

3. When selling an ESE, enterprise software vendors tend to use the typical consultancy approach of "what problems do you have?" and, as a consequence, "what problems can we solve?". This approach is not wrong by any mean, but is certainly limited compared to what an ESE can do. What the vendors' procedures do not take into account are the advantages that ESEs bring as a whole, as a new technology. ESEs do not just solve single problems, they foster a whole new way of conducting the PD process, and they may happen to solve problems customer enterprises are not even aware of. In other words, there must be a procedure to put forward the search technology first, and after that to shape the vertical market solutions on the customer's needs.

6.1.2 Model Proposition

What is still missing in the literature is a roadmap for the proper deployment and implementation of ESEs inside enterprises, a procedure to track down the general needs of product developers in manufacturing enterprises and insert them in a framework matching all of the ESE's benefits. This Model aims at being a tool for ESEs' vendor enterprises to identify the areas of a manufacturing enterprise where the ESE can bring value, and therefore explain to the decision makers of that company why they should install an ESE.

6.1.3 Purpose of the Model

The model is designed to be a tool for business development, technical-sales and sales people of information technology companies, software resellers and consulting enterprises that want their customers to understand the value of ESEs and then assist them in the implementation process.

Business Developer

His task is to raise the awareness and foster the culture, mentality and technological environment of customer enterprises so that, in the future, the other roles will be able to sell and implement an enterprise software systems for that customer.

Technical-Sales (Tech-Sales)

His role is to ensure the technical conditions that make the installation of the ESE feasible. He has to check if the customer enterprise has a proper hardware infrastructure and makes the existing software systems ready for a new addition. He is not just a computer technician, because his role also implies communicating with the customer enterprise's personnel in order to gather all the information needed. Furthermore, his activities do not involve only programming but also the logic design of software solutions.

Sales or Vendor

His role is to show the benefits of the software system he wants to sell and to carry on the sales process. He gets in contact with the various decision makers inside the customer enterprise and convinces them of the value of the product. He also discusses the terms of payment and eventually the delivery of additional modules or services.

The strong hypothesis of the model is that the above-listed people can take advantage of a structured and systematic approach to conduct a successful value proposition process. In fact, as the study will explain later in the Chapter, the proposed procedure is rather different from the one implemented so far.

6.2 Overview of the Model

In order to be a proper roadmap for the purchase and following implementation of ESEs, the model needs to be divided into pre-defined logic steps.

The first step of the Model consists of mapping the customer enterprise's PD process in order to point out how it is conducted. In particular, the user has to identify who carries out the PD process in terms of business unit (BU) or job role (JR) throughout the enterprise. In the second step, the Model identifies a list of BATCHes. Each BATCH is composed of one or more BUs and JRs that take part into the same Stage of the PD process. The output of this step is made of groups that enclose certain needs for a particular module belonging to the ESEs. This step is the core of the model, because judging from the aspect of the BATCHes an expert user of the Model can start evaluating the feasibility of the ESE's sale and implementation. It is the combination of BU, JR and PD Stage to generate the specific needs of each batch. In the third step, the Model links each BATCH with one or more modules of the ESE. After the third step, there is a test of the consistency of each BATCH, is called step four. In particular, the requirement that batches have to meet is being matched with one of the ESE's modules. If no module of the ESE can match that particular BATCH, it is given a "no go", and will be removed from the algorithm. On the other hand, each BATCH and its linked ESE's modules are put together and, from this moment on, they will be the only ones taken into consideration. Once there is list of defined and consolidated BATCHes, the fifth step is to associate the BATCHes with the benefits that their linked modules can bring them. The sixth and last step is to evaluate the benefits: for each BATCH, the Model proposes a method for calculating the benefits deriving from the installation of the module. As a consequence, at the end of step five, the user of the Model has a clear map stating which module of the ESE the specific customer enterprise may need and which benefits to submit to its judgment and approval.

Model users: business developers, techical-sales, vendors from IT enterprises, software resellers and consulting companies whose offering includes ESEs.

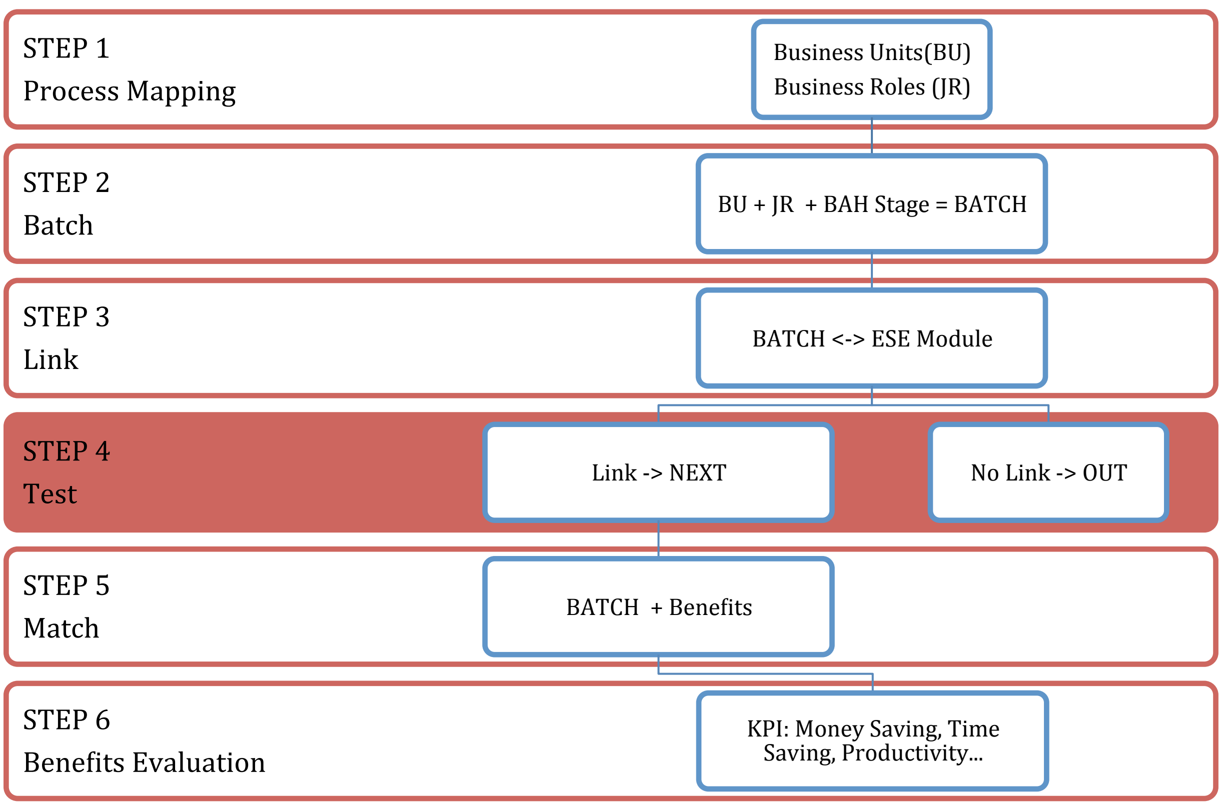


Figure 6.1: Overview of the Model

6.3 Model Framework

6.3.1 STEP 1, Process Mapping

The first aspect to be considered when setting up the model is the central role played by the PD process.

As explained in Chapter 3, given the value created by the PD process and its pervasive nature, it is of the utmost importance to get the whole enterprise involved in the process. Therefore, the first STEP is to identify all the BUs involved in the customer's PD process. Here follows the list of the traditional seven main Business Units (BUs) taking part in the PD process of manufacturing enterprises:

Corporate Office

The Corporate Office is where the most important decisions are taken. A push for new product ideas can come straight from here, because the Corporate Office is responsible for several activities that can ignite the process. The Corporate Office monitors the competitors, so it can tell whether they are up to launch new products or whether it is time to anticipate their moves and start developing a new concept. It analyzes the trends and mega trend of the moment, so it decides whether the environment is ready for a new launch and the technology is mature enough. The Corporate Office is also supposed to have a full view of the enterprise, thus knowing if the internal resources are ready for the development of a new product. In particular, this office monitors the financial status of the enterprise, which is the ultimate constraint when deciding to invest in new products.

Marketing

Marketing takes part in the PD process in two separated phases. The first one is called pre-sales marketing and takes place at the beginning of the process. Pre-sales marketing activities include the monitoring of the competitors, identification of a potential target market, speculation on what the penetrated market will be in the end and so on. In other words, pre-sales marketing involves all the activities devoted to support the creation of the product and its future, but before it is put into production.

The second phase is called after-sales marketing, and it takes care of the promotion of the product once it is available for the customers: advertisement, events, public demonstrations, mailers etc.

Research and Development

R&D is the core BU in the PD process because it is due to turn ideas into real objects. R&D can whether start from the green field and sketch out of ideas or just revise an old product to make a new version out of it. Also, the R&D is responsible for studying the structure of the product and exploding the engineering bill of materials (EBOM). According to the EBOM, the R&D BU decides if there is a need for engineering new parts, if the engineering is feasible or if it is necessary to purchase them on the market. In addition, the R&D BU also has access to the parts and components repository. In fact, the R&D BU needs access to the available documents regarding all the parts and components as well as to the warehouse, so that it can check if the components are available. Also, this BU is responsible for the testing the product, both virtually and practically, along with all the documentation that goes along with a new product. The criticality of the R&D department is that it needs to be always aligned with all the BUs of the PD process because it needs to create something that is related to the company's strategy (*Corporate Office*), whose production is feasible (*Production*) and that can be sold (*Sales*).

Sales

Sales operations come when the product is a definite object or idea ready to be launched on the market. The Sales BU keeps in touch with the customers and manages the sales process, which is often a long one, particularly when the manufacturer works on commissions and has to carry out a bidding process.

Production

Usually it is hard to identify the whole entities involved in the production of the products in just one BU. However, when analyzing manufacturing companies, the tendency is to group all plants, machinery and production lines under the same responsible and call it Production BU. It is the BU where all studies, projects and prototypes are exploited to realize the final product that will be sold. This BU has access to the manufacturing bill of materials (MBOM).

Purchase Department

This is a BU that does not have an active role in but contributes to the PD process. As a matter of fact, the Purchasing Department is responsible for purchasing raw materials, components, parts, assemblies, machines, software and everything enterprises need to carry out the PD process. In addition, the Purchasing Department decides whether to purchase the ESE or not. It is up to the Purchasing Department to evaluate the ESE and determine its usefullness to the enterprise in terms of benefits/price ratio. As a consequence, it is of the utmost importance to show the Purchasing Department the features of the ESE and convince it of the value.

Technical Department

Like the Purchasing Department, the Technical Department is also a BU supporting the PD process without having an active role in generating new ideas. It is more about putting into practice the ideas generated by the R&D. However, especially in Small and SMEs, it is hard to tell the difference between the Technical Department and the R&D, as they often overlap. In the logic of an ESE vendor, the Technical Department performs two functions:

- Practical Function: it is responsible for physically designing, sketching, drawing the ideas coming from the R&D or whoever is the creative mind within the enterprise. It is also responsible for keeping track of the PD and production process. In other words, it is the core practical side of the PD process.
- Control function: it is responsible for all the hardware and software used in the enterprise's premises. As a consequence, this BU is also responsible for evaluating the ESE's potential, but this time under a technical point of view. The Technical Department determines whether the ESE is a useful supporting tool for the PD process and what BU or BR may need it.

As we have seen in Chapter 3, there is no standard way for enterprises to carry out the PD process. Indeed, despite following the logic of the BAH model, enterprises seldom have also a standard procedure to develop new products. Depending on the nature, the history, the contingencies and the cost of the product, some enterprises would rather skip the rigid sequence of BUs and assembling teams for PD. More than that, each enterprise can also interpret JRs in its own personal way, assigning people to more than one BU or putting them through a matrix-looking organizational chart. Therefore, the model has to take into account this very important aspect: sometimes the stages of PD are carried out by single individuals or teams having a specific job role, but that act independently from a standard BU. For example, as stated above, the Technical Department and the R&D may overlap and be included in the same company function, but the roles of the personnel may vary according to the task they perform. In this case, the user of the model should concentrate on mapping the JRs rather than on the BUs. The following is thus the list of the nine job roles (JRs) that can be possibly involved in the PD process. We take for granted that each manufacturing enterprise can establish a whole new JR which is specific for its own context:

C-Level Roles: usually part of the Corporate Office, they represent the top management of an enterprise. Due to the nature of their role, they are seldom involved in PD teams. Nonetheless, especially in small-medium enterprises, they are responsible for the main idea, the beginning of the whole process. C-Level Roles are the CEO, CIO, CTO etc.

Pre-Sales Marketing Expert, Marketing Planner, Market Analyst: normally part of the Marketing BU, they take care of the first phase of marketing (Pre-Sales Marketing). The enterprise may even hire independent professionals or even an external company specialized in marketing. However, these JRs are often part of PD teams because of their knowledge of market needs. As we will see later on in the model explanation, these BRs perform similar activities and have the same requirements in terms of software tools and systems.

After-Sales Marketing Expert, Advertisement Expert: usually part of the Marketing BU, they take care of the second phase of marketing (After-Sales Marketing). As for the first phase of marketing, they can also be hired as professionals from independent companies. They are often part of PD teams, but have totally different skills and tasks from the Pre-Sales marketers. As we will see later on in the model explanation, After-Sales Marketing Experts and Advertisement Experts perform similar activities and have the same needs in terms of software tools and systems.

Designers: normally part of the R&D BU or Technical Department, they are the most likely roles to be involved in a PD team. Depending on their background, skills and attitudes, they may be extremely specialized in a certain kind of item or even feature. For example, the Aerospace and Defense industry includes designers who are specialized in aerodynamic surfaces, and their only task is to design these items. Therefore, depending on the product, a different designer may be involved. Despite Designers and Architects are two distinct jobs, inside manufacturing enterprises they assume the same JR.

Product Engineers: normally part of the R&D BU, they are also highly involved in PD teams. Their main task is testing the object designed by their BU and performing the relative simulation analysis.

Vendors: normally part of the Sales BU, they initiate and finalize the sales process. They are responsible for keeping a good relationship with customers, for identifying the best business opportunities and closing the deals. In case of manufacturing companies that have to deal with bids, they are also responsible for taking care of the bidding process.

Process Engineers: normally part of the Production BU, Technical Department or R&D, they take care of the production process. They are the link between the designers or the R&D and the personnel working in the production plant. Also, they often take care of designing the process to manufacture the physical products planned by Designers and Product Engineers. In long-term PD projects, they often perform the role of project managers because they have a full vision of all the activities of the PD process.

Production Managers: Normally part of the Production BU or Technical Department, they are responsible for the production feasibility and timetable, as they manage the production time, facilities and manpower. They work in close contact with the Process Engineers.

Specialized Technicians: normally part of the Production BU or Technical Department, they are directly responsible for the machines used in the production plant. Nowadays factory workers have become more than just machine operators. As machines are now embedded with software, technicians are now special professional roles requiring specific training and a vast knowledge in the field of operation. As a result, they have access to a larger amount of information and are often hardly replaceable.

<u>*Others:*</u> it is possible to find standard BUs being replicated in all manufacturing enterprises. They may have different names, but the ones we listed are definitely present in every company. On the other hand, JRs are far more connected to the background, business strategy and business model of every single Enterprise. The ones listed in this paragraph are the most common, but the Model takes into consideration all possible JRs that can occur during the screening of a company.

6.3.2 STEP 2, Batch

In Chapter 3, this study has considered PD as a macro business process and analyzed it from a conceptual and theoretical point of view given out by scholars over the years. In this STEP of the Model we need to break down the PD process into the stages identified by Booz, Allen & Hamilton [15] in the so called BAH Model, re proposed in Figure (6.2).

Figure 6.2: Stages of new Product Development (Booz, Allen and Hamilton, 1982)

The aim of this STEP is to provide the user with a list of BATCHes. Every BATCH is structured as follows:

BATCH i	
BAH Model Stage	stage (1), ... , stage (7)
Business Unit	name (1), name (2), ... , name (7)
Job Role	name (1), name (2), ... , name (9)

Requirements	...

The BATCHes are intended to give the user an instrument that relates the people of the customer enterprise to the stage of the PD process in which they are involved, whether they act through their BU or independently. As a consequence, it is possible to complete the BATCH with a full list of all the requirements of the BUs and JRs. Indeed, the real purpose of the BATCH is to relate the people with their role in the PD process, and then do what they need in order to perform the tasks and activities they have been assigned. The following is therefore the list of BATCHes the Model puts forward.

BATCH A	
BAH Model Stage	New Product Strategy (1), Business Analysis (4)
Business Unit	Corporate Office
Job Role	C-Level
Requirements	A tool to keep pace with market trends and mega trends, to access information throughout both the internet and the enterprise's intranet. It is necessary to be constantly up to date both on the external world and the enterprise itself. Depending on how much the C-level person is involved in the PD process, he may need a different degree of access to detailed information. This BATCH's main requirement is for a business tool keeping track of the business strategies related to the PD. It is also necessary to check whether the PD process is on schedule, and possibly to spot where the issue is.

BATCH B	
BAH Model Stage	Idea Generation (2), Screening & Evaluation (3)
Business Unit	Marketing
Job Role	Pre-Sales Marketing Expert, Marketing Planner, Market Analyst
Requirements	This phase of the marketing has to manage a large quantity of data coming from competitors, suppliers and customers. Also, these data

	come in different formats (emails, text documents etc.) and, most important, from different software systems. Furthermore, all these data have mostly to be accessed simultaneously because they need to be sorted, crossed and processed.

BATCH C	
BAH Model Stage	Commercialization (7)
Business Unit	Marketing
Job Role	After-Sales Marketing Expert, Advertisement Expert
Requirements	A tool that instantly connects the product’s characteristics and specifications with the software systems used to advertise it. Also, a tool to analyze the real time results of the advertisement activity: real time analysis of the performances, warm leads tracking and telemarketing are also performed each by a different tool but, again, the information must be accessed simultaneously.

BATCH D	
BAH Model Stage	Idea Generation (2), Design & Development (5), Testing (6)
Business Unit	R&D, Production
Job Role	Designer, Product Engineer, Specialized Technician
Requirements	A tool supporting the pure activity of designing the product, a software system helping designers save time and find the information they need at once. Besides, this tool needs to access the databases of both R&D and Production BUs, because, as we have seen in Chapter 2, the two are strictly connected. In particular, the access to information is vital. In fact, it is impossible to fasten the design or testing process by itself, because that really depends on the skills of the designer. What a software system can do is eliminating the time wasted searching for information such as CAD and CEA files or re-designing projects of parts that already exist inside the enterprise’s repositories.

	At the same time, the workers in the factory have exactly the same needs, but instead of information about sketches and CAD files they would rather search for CAM files, stored molds, pallets etc.

BATCH E	
BAH Model Stage	Idea Generation (2), Design & Development (5), Testing (6)
Business Unit	R&D, Production
Job Role	Process Engineer, Production Manager
Requirements	A tool supporting the design of the process, a software system helping engineers save time and find the information they need at once. Besides, this tool needs to access the databases of both R&D and Production BUs, because, as mentioned Chapter 3, the two are strictly connected. In particular, the access to information is vital. In fact, it is impossible to fasten the design of a process by itself, because that really depends on the skills of the designer. What a software system can do is eliminating the time wasted searching for information such as previous layout projects, templates, technological features etc. Production Managers have similar needs: they require rapid access to all the information to start up the production process and manage their resources.

BATCH F	
BAH Model Stage	Commercialization (7)
Business Unit	Sales
Job Role	Vendor
Requirements	A tool to keep track of customers, contact and manage the sales and bidding process. The main required connectors are with CRM and Customer Relation Analytics software. Besides, vendors are used to work hard on emails, Excel Files and Power Point. These are all detached tools, and they need a data aggregator.

BATCH G	
BAH Model Stage	None – Support the PD process
Business Unit	Technical Department
Job Role	Designer, Specialized Technician
Requirements	For its role, the Technical Department needs a tool to access the information about warehousing of products, parts, components, WIPs, tools etc. As we have seen in Chapter 3, all of the data is managed by an ERP system that communicates with the PLM system. The Technical Department needs to access both of them at the same time and to have all the information from the two to be visualized at once. In addition, the Technical Department is also due to approve all of the technical instruments used inside the enterprise. Therefore, it needs to check the requirements from other BUs in addition to its own ones.

BATCH H	
BAH Model Stage	None – Support the PD process
Business Unit	Purchasing Department
Job Role	No pre-defined role
Requirements	The Purchasing Department has no part in the PD process. In its supporting activities, however, it needs to have a view of what the company needs to purchase in order to develop new products. Again, this means combining the information coming from the ERP with all the instruments used for purchasing and billing inside the enterprise. In addition, the Purchasing Department is gives the final approval for the purchase of every instrument and tool used inside the enterprise. As a consequence, it needs to have a full view of the software solutions.

6.3.3 STEP 3, Link

As we have seen in Chapter 5, IT vendors offer various ESEs, but they all come in the same structure: there is a basic search technology and a software framework that allows the ESE to access the information, which means setting the connectors and generating an index. Then, this basic infrastructure needs to be specialized to perform certain activities, so the customer will look for specific modules. Thus, in STEP 3 the user of the model has to link every BATCH with the proper module so as to fulfill the requirements expressed by the BATCH. The procedure is as follows:

BATCH i	
ESE Module	Name / Trigram / ID Code:

For $\forall i \in A, B, C, D, F, G$

This STEP is valuable from a sales point of view. ESE modules may also be further divided into sub-modules, so it is extremely important that the user identifies exactly the software features he links to the BATCH.

Obviously, each ESE vendor has a different offering in terms of ESE capabilities, so they may not be able to cover all of the BATCHes with specific modules, although Chapter 4 has proved that state-of-the-art enterprise search technology can do it.

6.3.4 STEP 4, Test

At this point of the algorithm, the user has two kinds of BATCHes: the ones linked to a module of the ESE, and the ones that found no link in STEP 3. The latter group represents the BUs and JRs for whom the user does not have a software solution to fulfill their requirements. In some cases, it is also possible that the user cannot identify any BU or JR of such a kind in that customer enterprise. Therefore, these incomplete BATCHes need to be eliminated from the algorithm.

On the other hand, BATCHes that are consistent with BUs or/and JRs and that are linked to a module stay in the process and proceed to STEP 5.

6.3.5 STEP 5, Match

The user has now to list all the benefits that the installation from that module brings to the BATCH. It is not just a mere demonstration of the fulfillment of the requirements but rather a detailed explanation of all the implications that specific module will bring to the BATCH. In fact, from this STEP on, the user also needs to involve the technical sales experts. Selling an ESE and, in general, an enterprise software system is not just a matter of signing a contract. Once the product is sold, it has to be installed and customized according to the customer's specific requirements in terms of software settings and hardware capabilities. STEP five matches the BATCH with specific benefits so that:

- The customer can realize how his requirements are fulfilled.
- The customer can have a view of the impact on the whole organization.
- The user knows what to present to the customer and what to put forward when proposing a proof of concept.

Since it is not pre-determined whether the entire BU will have access to the module or it will be granted only to some JRs, the model will refer to the generic ESE's user with the term "subject". The model provides the following structure:

BATCH A	
BAH Model Stage	New Product Strategy (1), Business Analysis (4)
Business Unit	Corporate Office
Job Role	C-Level
Requirements	...
ESE Module	Name / Trigram / ID Code
Benefits	Full search capabilities, with both web and enterprise focus. Time saving thanks to the merging capabilities of the ESE: the subject can keep up to date on the latest trend and mega trends by receiving news feeds from blogs and industry websites, social communities, analysts and his personal team. It is possible to monitor the competitors by connecting the ESE to the competitor's websites and decide what kind of information is worth mentioning. For example, the subject can spot in a single view feeds about the keyword "smartphones" from Linkedin, The Financial Times, Apple.com and the enterprise's internal social community. The key benefit is the amount of information and the scalability of the amount retrieved. Also, the subject has a view on and keeps track of the full PD process from the same application.

BATCH B	
BAH Model Stage	Idea Generation (2), Screening & Evaluation (3)
Business Unit	Marketing
Job Role	Pre-Sales Marketing Expert, Marketing Planner, Market Analyst
Requirements	...
ESE Module	Name / Trigram / ID Code
Benefits	The installed module acts mainly as a pure search tool and connector. The subject can use the search functions to look for information in all of the repositories. As a connector, the module allows the subject to connect the specific software environments he uses to perform his market searches with his emails.

BATCH C	
BAH Model Stage	Commercialization (7)
Business Unit	Marketing
Job Role	After-Sales Marketing Expert, Advertisement Expert
Requirements	...
ESE Module	Name / Trigram / ID Code
Benefits:	The module acts mainly as a connector and information merger. The subject is able to connect with the product information repositories so that he can extract and process the information he needs to prepare an advertisement campaign. He can also have direct access to pictures, sketches, product features and characteristics so that he can include the real value of the product inside the ad campaign. Having a view of both the product's information and the customer's requirements allows the subject to target the market with maximum precision.

BATCH D	
BAH Model Stage	Idea Generation (2), Design & Development (5), Testing (6)
Business Unit	R&D, Production

Job Role	Designer, Product Engineer, Specialized Technician
Requirements	...
ESE Module	Name / Trigram / ID Code
Benefits	The subject can search for CAD files containing designed parts, components, assemblies and re-use what he needs instead of re-designing everything out of the green field. The module is a real interface to better understand and retrieve information from the traditional PD software infrastructure mentioned in Chapter 3. Also engineers can access to files containing case histories. Stress analysis is often a tough activity to carry on. Having access to previous studies may help saving time when performing the analysis on similar parts. The same logic can be applied to production managers and workers, who can then have immediate access to information without searching separately in the ERP, PLM and local machine software. In addition, the module can also be a tool to boost creativity and collaboration. The ESE can connect designers to the Web or the intranet and let them have access to communities, thus helping them find new ideas and solutions. The benefits translate into time (and money) saving as well as quality improvement.

BATCH E	
BAH Model Stage	Idea Generation (2), Design & Development (5), Testing (6)
Business Unit	R&D, Production
Job Role	Process Engineer, Production Manager
Requirements	...
ESE Module	Name / Trigram / ID Code
Benefits	The subject has a module to track the PD process, analyze its course and take care of the upcoming issues. Also engineers can access to files containing case histories. The subject of the PD process is often just the versioning of a previous product, so it may be useful to search back the

	production process applied in the past and adjust them for the new product. The same logic can be applied to production managers and workers, who can then have immediate access to information without searching separately in the ERP, PLM and local machine software. They can take advantage of case history and see how problems were solved in the past, and can have a real time analysis of the production on a simple screen page.

BATCH F	
BAH Model Stage	Commercialization (7)
Business Unit	Sales
Job Role	Vendor
Requirements	...
ESE Module	Name / Trigram / ID Code
Benefits	Also in this case, the main advantage derives from the searching and displaying capabilities of the ESE. It can group all the information about the customers taken from the CRM and put them into a custom format that is suitable for the vendor. Furthermore, the module provides functions for Business Analytics.

BATCH G	
BAH Model Stage	None – Support the PD process
Business Unit	Technical Department
Job Role	Designer, Specialized Technician
Requirements	...
ESE Module	Name / Trigram / ID Code
Benefits	Linking the modules with the ERP, the subject has a user-friendly tool to communicate with the warehouse and keep track of the parts, WIPs, components, raw materials etc. The main benefit is once again the re-use of all of those elements so as to save money out of warehouse space,

	parts turnover, over all storage efficiency. Besides, in this BATCH there are also JRs such as Designers that look for the same benefits of BATCH D.

BATCH H	
BAH Model Stage	None – Support the PD process
Business Unit	Purchasing Department
Job Role	No pre-defined role
Requirements	...
ESE Module	Name / Trigram / ID Code
Benefits	As explained before, BATCH G is an exception in the model. The subjects here do not need a specific module of the ESE, but they would rather have a full view of the advantages the solutions brings to all the other BATCHes.

6.3.6 STEP 6, Benefits Evaluation

STEP 6 is also the last STEP in the model. The input is a list of BATCHes correlated with the customer's benefits. In this STEP, the model points out the KPIs suitable for measuring the benefits of every BATCH. As matter of fact, all benefits end up being a matter of saving time, money or both. The truth is, manufacturing enterprises need to quantify those benefits and, most of all, to compare the situation AS-IS with the new perspective of having installed an ESE.

Despite their undeniable utility, KPIs are not always available, at least not for every BATCH. Some of the benefits brought by an ESE are not easily measurable. This study details this aspect in paragraph 6.3.7.

BATCHes are completed with the KPI label:

BATCH A	
BAH Model Stage	New Product Strategy (1), Business Analysis (4)
Business Unit	Corporate Office
Job Role	C-Level

Requirements	...
ESE Module	Name / Trigram / ID Code
Benefits	...
KPI	• $\frac{\#\ Links\ followed}{\#\ Links\ presented}$: the ESE is useful if the subject can actually find something useful out of the newsfeed. So, the KPI shows a percentage of how many links have been clicked compared to the total number of links presented by the module. • $\frac{\#\ Issue\ spotted\ before\ the\ alert}{\#\ Issue\ reported}$: the software tool is useful only if the subject has real time updates on the PD process. Once such a high level in the enterprise gets to know there is a problem in the process, it is already too late. This KPI measures the productivity of the ESE in terms of time gap between problem spotting on site and problem reporting to the management office.

BATCH C	
BAH Model Stage	Commercialization (7)
Business Unit	Marketing
Job Role	After-Sales Marketing Expert, Advertisement Expert
Requirements	...
ESE Module	Name / Trigram / ID Code
Benefits:	...
KPI	• $(query\ time\ AS - IS) - (query\ time\ with\ ESE)$: this KPI is nothing more than a measure of the ESE performances. It measures how fast the ESE can perform a query compared to the time required to perform the same query with standard instruments (call the colleague, sort the Excel files etc.)

BATCH D	
BAH Model Stage	Idea Generation (2), Design & Development (5), Testing (6)

Business Unit	R&D, Production
Job Role	Designer, Product Engineer, Specialized Technician
Requirements	...
ESE Module	Name / Trigram / ID Code
Benefits	...
KPI	• $(\#parts) * (\%\ duplicates) * (part\ cost)$: To measure the savings, this KIP is the net amount of money saved thanks to the benefits of the ESE. Out of this KPI the customer company can also work out a ROI calculation, which is even more explicative than the bare digit given out by this KPI: • $ROI = \frac{(\#parts)*(\%\ duplicates)*(part\ cost)}{Installation\ Cost}$

BATCH E	
BAH Model Stage	Idea Generation (2), Design & Development (5), Testing (6)
Business Unit	R&D, Production
Job Role	Process Engineer, Production Manager
Requirements	...
ESE Module	Name / Trigram / ID Code
Benefits	...
KPI	• $\frac{\#\ Issue\ spotted\ before\ the\ alert}{\#\ Issue\ reported}$: following the same logic of BATCH A, this KPI measures the productivity of the ESE in terms of time gap between problem spotting on site and problem reporting to the manager in charge.

BATCH G	
BAH Model Stage	None – Support the PD process
Business Unit	Technical Department

Job Role	Designer, Specialized Technician
Requirements	...
ESE Module	Name / Trigram / ID Code
Benefits	...
KPI	• $(\#parts) * (\%\ duplicates) * (part\ cost)$: the KPI is exactly the same used for BATCH D. As a matter of fact, the subject here is the Technical Department. So, this is the best indicator to show the performances of the ESE, because it points out the amount of money the Technical Department can save through the ESE. As we did for BATCH D, we also put the ROI as a KPI.

BATCH H	
BAH Model Stage	None – Support the PD process
Business Unit	Purchasing Department
Job Role	No pre-defined role
Requirements	...
ESE Module	Name / Trigram / ID Code
Benefits	...
KPI	All of the listed KPIs have to be submitted to the subjects of BATCH H. The KPIs will determine the opinion of the Purchasing Office about the ESE.

6.4 Model Limitations

The model shows several limitations and weak points. Nonetheless, this paragraph aims at explaining the weaknesses and points out the reasons why the model can still be considered solid.

6.4.1 Prerequisites

The model requires one fundamental prerequisite: the customer manufacturing enterprise needs to have a pre installed traditional software structure as per paragraph 3.2. If the customer company decides to store its whole enterprise knowledge on paper, there is no sense in installing and ESE. There has to be a hardware and software infrastructure to store and manage data. However, despite being a strong limitation to the model, it is also true that nowadays almost all enterprises are imbued with both advanced hardware and software. Also, if the customer enterprise has a PD traditional software infrastructure like the one detailed in Chapter 3, it is easier to show the utility of an ESE, since it can be regarded as a natural complement of that very infrastructure.

6.4.2 Economic Feasibility

First and most evident is the fact that, despite being a roadmap for a sales process, the Model does not take into consideration the aspects related to the costs of the ESE. The reason for this deficiency is that this model is focused on the technical aspect of the ESE's sale. The roadmap is valid no matter the number of modules that are sold to a customer enterprise or their actual cost. It is up to the user of the model whether to perform the economical feasibility analysis before or after implementing the algorithm. In the first case the cost is a prerequisite to the implementation of the model. In the second case the cost is used as a sort of conclusive STEP 7 of the Model: once the customer enterprise is ready to be proposed the ESE, the user of the model checks its feasibility from an economic point of view.

On more aspect to be considered is that it can be rather difficult to determine the real cost of an ESE. None of the of the largest software vendors listed in paragraph 3.2.2 provides a public list for its software solutions, and their ESEs are no exception. Therefore, the price of an ESE is a matter of contracts and deals between companies, a matter that is too complex to be included in a technical roadmap.

6.4.3 Limited KPIs

Secondly, not all the BATCHes have a matching KPI, thus making the analysis of the benefits weaker. For example, BATCH B and F are left without any KPI. The reason behind this deficiency is that many benefits cannot be measured through mathematic formulas and indicators. PLM vendors often have to face the same problem: enterprise software systems radically change the entire companies' way of doing business, so it is hard to quantify exactly the ROI of every single module or every single feature. PLM vendors usually bring a list of

successful use-cases in order to convince the customer. Besides, one of the prerequisites for a correct installation of ESEs is the customer enterprise to having a solid pre-installed traditional software infrastructure (paragraph 6.4.1). It is therefore hard to determine the benefits brought just by the ESE, because they can be due not only to this ESE by itself but also by the matching of the ESE with the existing infrastructure. This aspect requires a particularly mature mentality in a company willing to purchase the ESE, particularly in the role of the decision maker.
As it happens for PLM software systems, the user of this model has to show the benefits of BATCH B and BATCH F through success stories and use-cases.
Another issue related to the KPIs is that they may be interpreted in a different way by big or small enterprises. As a matter of fact, costs and savings have a different impact on big or small enterprises, not just because of their value but also because of their nature. For example, Altea's experts (see Chapter 8) point out that a big enterprise may be ready to fire personnel because some positions in the organization would become unnecessary due to the introduction of an ESE. Just as it happened with computers years ago, ESEs enable searching activities that were once performed by human experts. Nonetheless, small enterprises may have a different perception of the value of people. An entrepreneur may decide not to implement a software solution just because it would certainly turn into job losses. To sum up, the results of the Model do not simply depend on the structure of the model itself or on its applicability, but also on the very enterprise it is applied to. For sure, the Model cannot take into consideration every single scenario, and it is therefore set on the theory that enterprises' ultimate goal is to maximize their profit according to existing government regulations.

6.4.4 Diversity among Enterprises

In paragraph 1.3.1 we mentioned the scalability of the Model from big to small-medium and small enterprises. Indeed, the limit of the Model is not really the size of the customer company itself, but it is rather the different mentality between big and small enterprises. In particular, big and particularly multi-national enterprises tend to be organized in articulated organizational charts that make it easier to identify standard BUs and JRs, whereas in small enterprises we may find people who do not even have a fixed role inside the organization. The lack of a clear organizational chart and the misleading perception of JRs inside small enterprises make the application of the Model difficult. One more reason for distinguishing between big and small enterprises is the presence of a role dedicated to the IT infrastructure. Usually, big enterprises have at least one person who is fully dedicated to the software tools employed by the company. In the case, the user of the Model has an ideal interlocutor, who is skilled on software and eager to acquire new knowledge of the subject. Also, he/she would be the ultimate expert of the enterprise's software-regulated processes.

However, according to our predictions it is not impossible but just more difficult to apply the Model, which means it simply may not express its full potential.

7 Model Validation - USE CASE

The model is given value in two different ways. The first one consists of bringing forward an ESE technical sales and installation use-case.

7.1 Summary

The Model has yet to be adopted by an ESE vendor, making it impossible to measure its performance in a real business context. Therefore, in order to prove its value, we have decided to analyze a case study where an ESE was implemented following a professional vendor's own roadmap. We then performed a gap analysis to highlight the differences between the standard procedure and the proposed model. The results show that the model brings more value to the enterprise in terms of need fulfillment. In particular, the standard vendor's sales approach focuses only on the ESE's modules for product design and engineering. On the other hand, the model takes into account more than just the R&D function, thus ignoring the requirements of all the people involved in the PD process. In other words, we can state that the standard procedure is somehow shortsighted compared to the proposed model.

7.2 COMAU – AS-IS

7.2.1 The Company

COMAU (COnsorzio MAcchine Utensili) is an Italian multinational company based in Turin, Italy and is part of the FCA (Fiat Chrysler Automobiles) Group. It is a global supplier of industrial automation systems and services mainly for the automotive manufacturing sector. With 40 years of experience in advance manufacturing, 25 locations in 15 countries and 15 production sites, COMAU is a proven worldwide leader in sustainable automation and service solutions with a truly global presence. Starting from a small group of manufacturers in Turin, over the years COMAU has become a global company owning more than 100 patents thanks to the acquisition and integration of other companies. At present all major automotive manufacturers use COMAU systems and machinery in their facilities all around Europe, America and Asia. The aim of COMAU is to bring technology to a new level in order to meet the changing needs of their costumers. COMAU's mission is to realize industrial solutions engineered by design to achieve results, and to support their costumers from the idea to the project, and beyond. Anticipating costumers' needs, assisting customers at every stage of the business, overcoming obstacles and producing winning solutions are the COMAU's key competitive advantages. Moreover, awareness and commitment to environmental sustainability are the reasons for ongoing investment in research and development on energy efficient automation solutions, and methods to reduce energy consumption. COMAU has six Business Units: Body Welding & Assembly, Powertrain Machining & Assembly, Robotics & Maintenance Services, Aerospace Production Systems, Adaptive Solutions and eComau. The high quality and innovative solution achieved by COMAU have impacted a wide rage of industries and applications that run on process automation technologies, including automotive, aerospace, petrochemical, military, shipbuilding and energy efficiency consultancy. COMAU is a 360° company that oversees its product through the entire process, from product design to product system and maintenance services, and offers industrial asset management, project management and eComau services in order to succeed in the highly competitive market [64].

7.2.2 Business Model and Output

As stated in paragraph 6.2.1, COMAU is made of six BUs [7i].

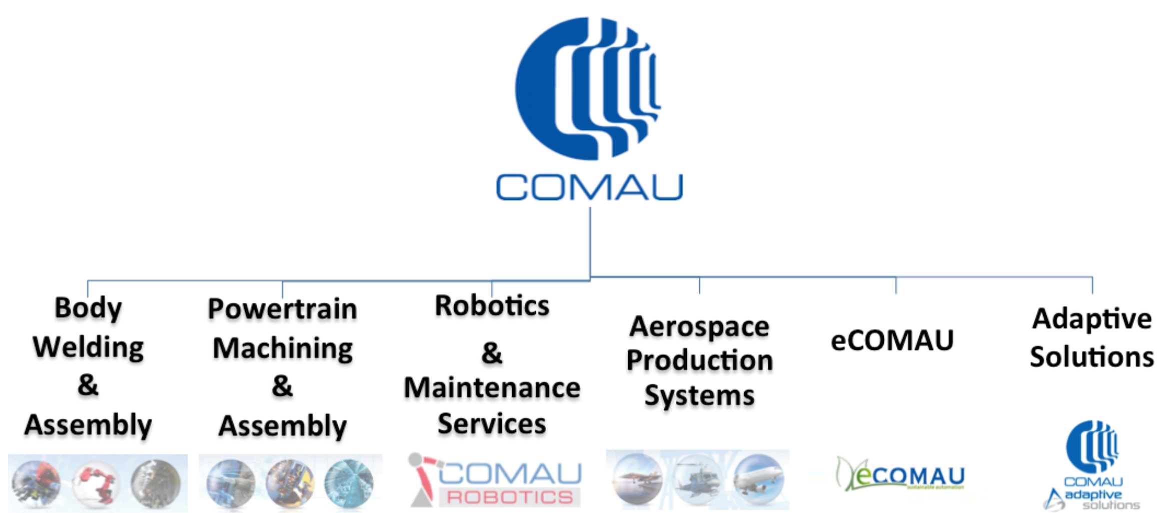


Figure 7.1: COMAU's offering and competences

Two of the above Units do not develop physical products but are service providers:

- eCOMAU: energy efficiency consulting.
- Adaptive Solutions: consulting for designing and building customized manufacturing solutions.

This study focuses on the other four BUs, i.e. the ones generating COMAU's core output: production systems. Actually, COMAU's customers are Original Equipment Manufacturers (OEMs), whom COMAU sells both physical and intangible products. The physical products are robots and machines that the OEM then uses to manufacture other products. The intangibles are the processes and layouts to transform robots into a manufacturing or assembly line or cell. Again, as stated in the company's website, "Our business is to engineer and manufacture production systems". The following are the BUs dedicated to this kind of output:

- Body Welding & Assembly: medium and high-volume body welding and assembly systems.
- Powertrain Machining & Assembly: development of custom flexible solutions for the production and/or assembly of powertrain units. The range of products goes from manual assembly to fully automated lines.
- Robotics & Maintenance Services: multi-axes articulated industrial robots, robotized solutions and related maintenance services.
- Aerospace Production Systems: engineering and manufacturing of production systems for the aerospace industry. The range of products goes from simple system integration (consulting) to full turnkey projects (full factory plant).

In addition, COMAU provides product-related after sales services that are also provided by the above BUs.

7.2.3 Product Development Process

In order to achieve the target production described in paragraph 7.2.2, COMAU implements a product development process that is common to Business to Business enterprises. As a matter of fact, there is no proper Phase 1 of the BAH Model, because the push for a new product comes directly from the customer OEM. It is not COMAU itself to generate a new idea, but rather the OEM that expresses its own requirements, thus pushing COMAU to find a solution. COMAU's PD process is shown in Figure 7.2 and explained in the following paragraphs:

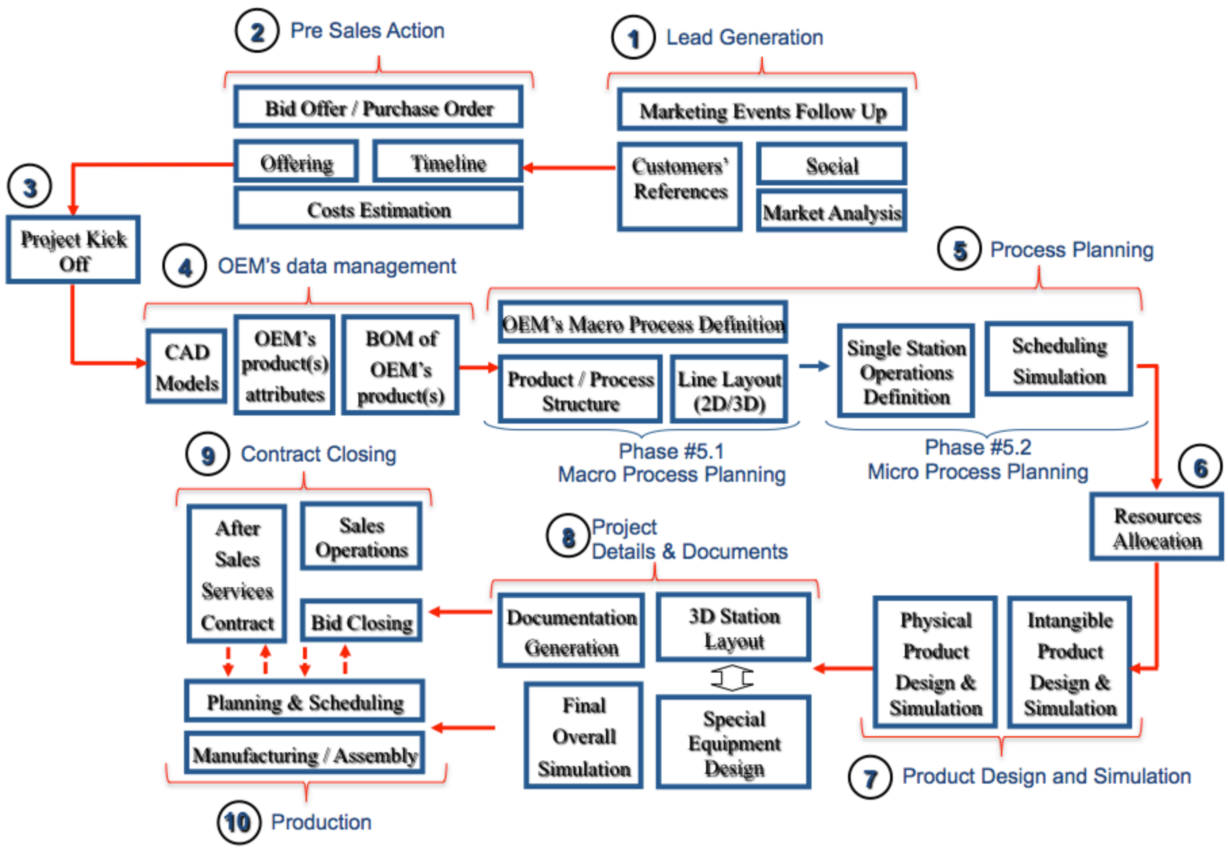


Figure 7.2: COMAU's Product Development process

Phase 1: Lead Generation

As stated before, the push for a new product starts from the customer's requirements. In this phase the marketing department makes analyses on new prospect customers and collects the requests from the existing OEM that would like to purchase a solution from COMAU (Customer's references). COMAU also organizes marketing events in order to keep in touch with the OEMs and attract new customers. From the follow up of these events the marketing department often generates new leads and opportunities. Also, over the last few years COMAU has been boosting its marketing activities in the field of social networks, thus significantly improving its brand awareness and attracting new prospect customers.

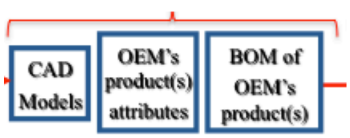
Phase 2: Pre-Sales Action

In this phase, the sales JRs are in charge of transforming the leads into real business opportunities. Once the customer receives COMAU's complete products offer, it starts discussing the most appropriate solution, the price and the timeline for the project deployment. If the OEM asks for a custom solution, COMAU fills in a special bid offer; if a standard solution is purchased straight out of the catalogue (i.e. a single robot) the customer would just send a purchase order. It is in this phase that the customer is shown its expense and delivery time estimate. In fact, both of these elements are included in the contract to be signed in Phase 3.

Phase 3: Project Kick Off

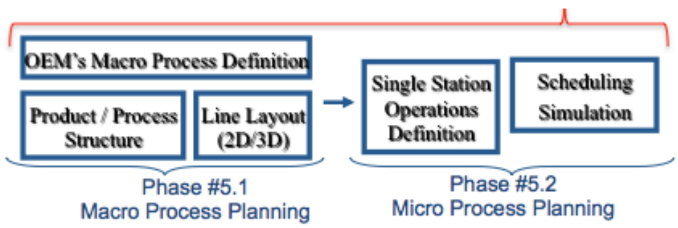
With the signature of the contract the OEM gives the green light for the designing of the product. This phase marks the start of the real R&D and technical activities.

Phase 4: OEM's Data Management



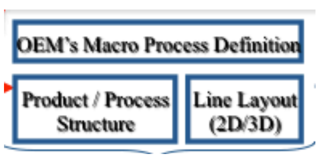
This is the first phase that involves the R&D BU. The OEM sends its requirements for COMAU to develop a solution around them. These requirements are in the form of data. They include the CAD assembly tree and geometry of the final product the OEM wants to put into production. This way COMAU can shape the production line around it. These data have to be analyzed and converted from the native to the local CAD system. The data include also a description of the product's characteristics, specific attributes and peculiarities, manufacturing challenges, physical limitations etc. Attached to those data, the OEM sends the engineering BOM of the product. For example, if the customer OEM is a car manufacturer and its final product is the bodywork of cars, the OEM would send COMAU a CAD model of the car it want to assemble along with the specifications of the bodywork: the material it want to use to manufacture the bodywork along with the desired durability, luster, elastic stiffness, strength, thickness, heat treatments etc. Then, the OEM gives the information on the manufacturing process: required takt time, level of line automation, safety requirements, maximum size of the production plant etc. In this phase, COMAU's engineers process all data to detect all of the products features: they list the standard components and input them into the Components Library and they list the custom components and input them into the Super Components Library. These two libraries are managed by the Librarians through the PLM system and Excel files. The two libraries together form the MBOM that is passed to the next phase.

Phase 5: Process Planning



In this phase engineers and designers analyze the requirements and start outlining the solution following a top-down approach. Here, designers and engineers act like architects, shaping the outlines of the plant and the stations, but they do not physically engineer them from a mechanical point of view. The phase includes two sub-phases so that the engineers can concentrate on a different level of detail.

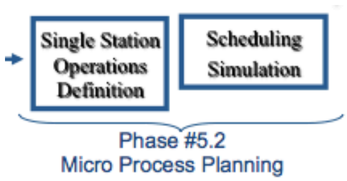
Phase #5.1-Macro Process Planning



In this phase engineers study and map the OEM's macro production cycle to identify the most critical issues to be considered. They import the CAD layouts and drill down the best feasible production process given the MBOM and the customer's process requirements. To proceed further with this phase, it is compulsory to have solid preliminary information on the customer's production process. As a matter of fact, the engineers identify the stream of macro operations and the flow of materials and WIP accordingly, so that they can divide the whole process into single stations and outline their sequence. For every station, they define the assigned parts and the cycle time assessment. Then, the engineers define the size of each station to balance the entire production line according to the customer's pre-defined production rhythm. So, according to the bottleneck station, the entire line is shaped to fulfill the customer's production standard.

Both the PLM and the individual file systems are used to manage the information described above. For example, the Component Libraries are often enclosed in an Excel file the engineers exchange within a project team, and the same happens for the MBOMs. The plant and Macro cycle are shaped with the CAD system.

Phase #5.2-Micro Process Planning



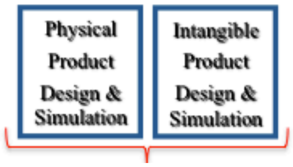
In this phase engineers and designers focus in detail on the single stations. They define the list of micro operations to be executed inside the station, their sequence and cycle time. Also in this sub-phase, engineers do not design the single mechanical components of the station but they rather focus on the layout and feasibility of the process. Once the layout is complete, the engineers simulate the schedule they fixed for the station using the CAM system. They also take into account the saturation of the station in order to have a balance not only on a macro, but also on a micro process level.

Once sub-phases #5.1 and #5.2 are completed, the PLM system updates all the documents. First, the input for the next phase will be an EBOM, which is inclusive of all the libraries and structure layouts to physically design the cells and assemble them in a plant. Second, the EBOM will include not only the parts and components to be designed but also the ones to be purchased. In fact, several standard components are purchased from suppliers, other components are assemblies of both standard and custom components and the fully custom components have to be designed from the green field. Third, there is a list of attachments such as text descriptions and notes, handmade sketches, pictures and so on that need to be sent to the designers along with the CAD files. At last, in this phase the engineers or the engineers' managers prepare the documents that will go along the process up to the completion of the product and its sale.

Phase 6: Resources Allocation

In this phase the personnel from the Purchase and Technical department fill in the Purchase BOM. It includes the groups of components, parts and pre-assembled parts to be purchased, the required quantities and quality specifications. But the purchase list could also include full robots to be bought from suppliers. As a matter of fact, we already pointed out in paragraph 7.2.2 that the real value of COMAU's final output is in the knowledge behind the plant's layout. Therefore, sometimes COMAU just purchases the whole hardware from the suppliers. In conclusion, the purchase list features the raw materials, tools and equipment specified in the EBOM and necessary for the realization of the final product. It is filled starting according to the Libraries from Phases 4 and 5 or creating new ones for the special equipment. According to this phase, the managers define the workload for the designers and engineers in Phase 7.

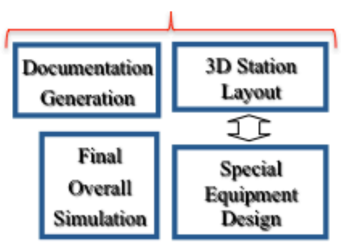
Phase 7: Product Design & Simulation



In this phase designers and engineers concentrate solely on the CAD design of the product. They use the traditional modeling and digital mock-up CAD tools to shape the physical product, which is the hardware (robots, pallets, transporters etc.), but they also exploit the system engineering features included in the same CAD environment to write down and draw the production process and the description and sequence of the operations.

In both cases, the engineers test the result of their work. At first, the subject of the simulation is the physical product. In this case, the engineers use CAE systems for the virtual testing of the physical qualities of the product. Then, the subject of the simulation becomes the intangible product, i.e. the production process. In this case engineers rely on CAM and discrete event simulation software tools. All of the tests follow both international and internal standards, therefore they are registered in the official product documentation.

Phase 8: Project Details & Documents



In this phase the R&D completes the last details of the project. Designers and engineers complete the CAD designs of the special equipment and include it in the layout of the plant. Once this activity is completed, they run a final overall simulation of the whole system, which is the final test to check whether it is ready for delivery to the final customer. The other main activity in this phase is the collection and aggregation of all documents. As a matter of fact, the final customer needs a long list of documents:

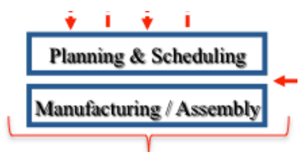
- The certifications of quality according to national, international and company-owned standards.
- The simulations that guarantee that the product fulfills the requirements and the warranty for potential substitutions.
- The technical documents, including CAD projects, CAD layouts and BOMs.
- The documents to finalize the purchase order, the costs certifications and the after sales services.

All of the above documents are created both from the PLM and ERP systems, because the former tracks down the activities from the PD phases along with the applied modifications, whereas the latter is the instrument used to supply and sell.

Phase 9: Contract Closing

In this phase the vendor leverages the closing of the sales process. The main tool involved is COMAU's CRM. During this phase the technical and commercial business people approve all of the documents listed before and store the project in COMAU's repositories. In addition, this is the phase to sign a contract for after-sale services, which may include technical (maintenance, upgrades, machinery additions, etc.) or consulting (best practices, energy saving methods, training, etc.) activities. It is important that this phase is aligned with the following Phase 10, as to meet the deadline agreed with the production.

Phase 10: Production



In this phase the physical product that was designed and simulated in Phase 7 is finally put into production. First, the production managers schedule the production throughout the production plants worldwide. Then, the technicians take care of whether manufacturing the new parts or assembling the ones that were due for purchasing in Phase 6. The factory personnel needs direct access to the whole product documentation and to the CAM data. It is important that this phase to is aligned with Phase 9, in order to meet the deadline set along with the final customer.

7.2.4 COMAU's Pains

In spite of its global success, COMAU feels it is still suffering unresolved problems in terms of PD software infrastructure. According to both DS's analysis and the IT manager's opinion, COMAU's pains can be listed as follows:

Data Location

Despite having one ERP and one PLM system for the entire group worldwide, a lot of data related to PD projects is still out of these two systems. According to COMAU's IT manager, 70 to 80% of the files used every day are inside the PLM environment, but the rest is on the users' file systems, on the customer's hard disks and on a Content Management System (Microsoft SharePoint).

The explanation for this situation is that the CMS was already present before the installation of the PLM. As usual, people inside the enterprise tend to stick to the tools they have been using since the very beginning of their experience. As already stated in previous the Chapters, most users do not like to change their tools, because they perceive innovation as a duty and the training to learn new instruments as a waste of time. This is particularly true in the use of file systems. Users still feel their files are more protected and easier to be managed if they have them stored in their own computer's repository. Of course, COMAU has been changed this mentality, but the problem of data location will persist until all of files have been included in the PLM frame.

Duplicate Files

Because of the many file repositories, COMAU's PD environment is characterized by several duplicate libraries, hence by a large number of duplicate files.

Sometimes the data are the same, but the related metadata are different. For example, the designers may find two CAD files containing exactly the same part but having different names and labels, and registered with a different project manager. Sometimes the metadata are the same, but the files are different. For example, two CAD files have exactly the same name, author, creation date etc., but they have been generated with different CAD programs. Therefore, they are actually two different files. In the end, some files ended up being duplicated just because of trivial users' mistakes. The most famous example happens when filling out the form for the bidding process. Some authors would call the subject, for example, "John's Company", some others "John Company" others again "John's Company Ltd" etc. This generates a huge amount of duplicates both users and administrators lose track of while they weight on the enterprises' systems and impact the users' working performances. This trouble is particularly amplified in the PD process because the technical product libraries are simultaneously in the PLM, in the designers' Sharepoint folders, in PDF files and in many other repositories that users cannot even remember. Once again, this problem can be easily solved through an enterprise PLM structure, but the problem will persist until COMAU has managed to implement it for all the people involved in the PD process.

Design Software Environment

COMAU has an internal PD software systems standard infrastructure. It has one CAD, CAE, CAM and PLM for all users worldwide. Nonetheless, each of COMAU's customers may use a different kind of CAD and/or CAM according to its own policies. The problem is, when COMAU's customers ask for the virtual projects of the product they are purchasing, they want to have it in their own format. Also, some CAD tools may be better than others at performing certain activities. As a consequence, COMAU ends up having at least one seat of each of the most popular CAD systems in the world: designers and engineers use Autocad from Autodesk to shape the 2D layouts of the plants, NX from Siemens PLM to convert their CAD projects in the NX format and send them to their main customer "FIAT", and they use both SolidWorks and CATIA from DS to design in 3D.

Because of this problematic situation, COMAU find itself being obliged to manage key files in different formats. As a matter of fact, CAD files contain the core product of the design activities, and COMAU has a problem in sharing this knowledge throughout the enterprise.

7.2.5 Explicit Needs

According to its IT manager, COMAU needs to reorganize and uniform its software tools supporting the PD process. This general statement can be translated into specific technical needs:

A tool to search and compare data

COMAU needs a tool to compare data of all types and formats. This tool has to be flexible and adaptable, as it must be integrated with all the types of CAD, CAE and CAM files. Besides, this tool also has to be to search for any kind of document throughout the entire software structure of COMAU. In fact, COMAU needs a tool to check all files and detect the duplicates, so that the IT operators can sort them out and finally store them in the PLM. Therefore, this tool has not only to search for documents but also to open them to unveil their content. In fact, searching for the documents' name would not solve any problem in this scenario.

A safe and user-friendly tool

COMAU needs a tool that is almost "out of the box". As the enterprise already has a software infrastructure, it does not want the new tool to go in contrast with the present software and further complicate the environment. So, the tool must first of all be easy to be employed by IT specialists. However, COMAU requires some user-friendliness standards for the users as well. Particularly, since the tool will be used to search for different files according to the user, it will have to be customized for special queries and to be equipped with custom filters to refine the very queries.

In addition, the tool must guarantee access to the whole company's repository from all around the world, whether PLM or not, and also to users having no knowledge of the PLM.

In case the tool guarantees the above top priority needs, COMAU's IT manager foresees a future need for new tools inside the enterprise.

Business Intelligence

Should the new tool prove to be useful and efficient, it would be implemented throughout the enterprise. Then, the new tool would need to be improved with some Business Intelligence features. In fact, the users would like not just to search for documents and compare them, but also to make some analysis and track them down. Besides, once all of the data is virtually managed through the PLM, the need will also arise for an instrument to implement some PLM analytics. The PLM system is already capable of providing analytics and indicators, but COMAU would like to aggregate them with the rest of the data coming from the ERP and to show them in good-looking and user-friendly dashboards.

7.3 Enterprise Search Engine's implementation

The situation described so far shows the perfect environment for the installation of an ESE. As stated in Chapter 3 and Chapter 4, ESEs provide the technology to solve all of COMAU's needs. Therefore, in early June 2015 COMAU started a screening of the ESEs available on the market. The choice came down to DS's Exalead. The person in charge of evaluating the ESE taking the final decision regarding a potential installation is called "decision maker", which at COMAU is the Worldwide IT Manager. Hereafter we analyze two different approaches to the installation.

First, we describe the procedure adopted by DS and then we implement the model on the same premises.

7.3.1 DS's approach

STEP 1, Industry Identification

DS approached COMAU's need for an ESE with the standard procedure established by the Corporation. DS' market strategy consists of targeting the market by Industry. In particular, COMAU is classified as a manufacturing enterprise in the Industry called "Transportation and Mobility", because COMAU's customers are the original car manufacturers.

STEP 2, Vertical Solution Selection

Based on COMAU's Industry, DS has identified the vertical market solutions called "Exalead Onepart" and "PLM Analytics" (paragraph 5.3.4) as the best-fitting software tools for COMAU's needs. As a matter of fact, DS addresses the vertical solutions according to the Industry the customer belongs to. The logic behind this go-to-market strategy is that every Industry is specialized in standard specific processes and therefore all the enterprises in that Industry need a vertical solution addressing those processes.

STEP 3, Vertical Solution Fitting

DS identifies the JRs that need Exalead Onepart and PLM Analytics. Again, this matching is based on studies of the Industry and on the best practices that DS has recorded in years of activity. According to DS, Exalead Onepart and PLM Analytics are to be addressed as explained in the Table below:

BU	Job Role	Module
R&D	*Product Engineer;* *Process Engineer;* *Designer*	*Exalead Onepart*
Technical Department	*Product Engineer;* *Designer*	*Exalead Onepart;* *PLM Analytics*
Purchase Department	/	*Exalead Onepart*

Table 7.1: DS' Vertical Solutions Fitting

STEP 4, Customer Matching

DS and COMAU's decision makers meet to discuss the possible implementation of the ESE. In this case, COMAU's decision maker is the IT Manager. He identifies in particular the JR "Librarian" within the Technical Department as the main user of Exalead Onepart. Librarians are specific JRs that COMAU allocates to the Technical Department. Their job consists of being the human interface between the PLM system and the other roles involved in the design. Librarians' main task is inputting all of the product and production data into the software systems. They are the managers of the ERP, PLM and of all the rest of the software.

As a consequence, Table 7.2 shows the final output of DS' model to implement an ESE in COMAU. We highlight in red COMAU's contribution to the final output of the procedure.

BU	Job Role	Module
R&D	*Product Engineer;* *Process Engineer;* *Designer*	*Exalead Onepart*
Technical Department	*Librarian; Designer*	*Exalead Onepart*
Technical Department	*Product Engineer*	*Exalead Onepart;* *PLM Analytics*
Purchase Department	*/*	*Exalead Onepart*

Table 7.2: DS' Customer Matching

7.3.2 Model's approach

In this paragraph we act as sales and tech-sales personnel from DS and we use the Model explained in Chapter 6 to convince COMAU that the ESE Exalead would boost the company's performances when running the PD process.

STEP 1, Process Mapping

Following the mapping as of paragraph 6.2.3, we detected the following BUs that have a role in COMAU's PD process.

Corporate Office: the Corporate Office plays no important role in this PD process. As mentioned when introducing the company, COMAU is part of the bigger FCA group, therefore COMAU's Corporate Office does not have the autonomy and empowerment it could have in a standalone enterprise. Its role is seeking for new customers at a high level, involving the C-Level key figures of the prospect customers. Therefore, COMAU's C-Level roles are not directly involved in the PD process but they rather foster the beginning of the PD process.

Marketing: COMAU's marketing is deeply involved in the PD process. In particular *Market Analysts* play an important role in Phase 1 (Lead Generation), because they work on the main source of new customers. *Advertisement Experts* and *Marketing Planners* take care of the remaining activities, which can be classified as after-sales marketing. Specifically, COMAU keeps track of the follow up of marketing events and social media communication. All of the JRs from the marketing need access to the ERP and CRM systems.

Research and Development: the R&D is the core of COMAU's PD process. From Phase 4 up to Phase 8 COMAU relies on *Designers* and *Engineers* to give life to the product. They receive information and transform it in an industrial solution thanks to their expertise, but helped by a solid software infrastructure. In Phase 4, 5, 7, 8 *Designers*, *Product Engineers* and *Process Engineers* work in close contact and have a continuous exchange of information. In Phase 6 *Process Engineers* are the key JRs because it is the phase where they perform their core activities. This is also the phase in which they require the most complete access to process information.

Sales: in COMAU the Sales BUs are involved in the activities of Phases 2 and 9. *Vendors* take care of the bidding process both in the Pre (Phase 2) and After-Sales (Phase 9) part. Therefore, they need access to the CRM but also to the PLM, as they need to have some visibility on the process advancements.

Production: in COMAU's PD process the Production BU is not always considered the most important activity. It may seem a nonsense that a world class manufacturing company does not put production among its top priorities. However, this behavior is common in modern Western manufacturing enterprises. As a matter of fact, the real value of COMAU's products is the embedded study and the knowledge of the intangible part of the product. Quite often COMAU sells only the intangible part and delegates the production of the hardware directly to the OEMs. In any case, COMAU still needs highly skilled *Specialize Technicians* to produce or assembly complex robots and standards parts, and experienced *Production Managers*, because the production is not only peculiar because of the output, but also challenging because of the deadlines set by the *Vendors*. *Technicians* in particular need access to the information of contained in the PLM, but they do not possess the skills to operate such a software system.

Purchase Department: this BUs is involved in Phase 6 of the PD process. It works in close contact with the Technical Department to assure the quality of the purchased parts, components and tools. Also, they are responsible for those items to be delivered on time. In order to address these challenges they need to have deep insight of both the suppliers and the final product's characteristics.

Technical Department: this BU has multiple responsibilities. On one hand it has to ensure the quality of parts, components, assemblies and tools that it decides to purchase together with the Purchase Department. On the other, the Technical Department is responsible for the whole documentation relative to both the physical and intangible product. Therefore, this BU also includes Designers who take care of this delicate task and Product Engineers, because they are the ones who master the knowledge of the intangible product. In addition, the Technical Department includes the specific JR of the Librarians. The one of the "Librarian" is a specific JR that COMAU allocates to the Technical Department. His job consists of being the human interface between the PLM system and the rest of the roles involved in the design. Librarians take care of inputting all of the product and production data in the software systems. They are the managers of the ERP, PLM and of all the rest of the software. The name "Librarian" comes from the fact that the main subjects of their job are the technical product libraries.

To sum up, STEP 1 identifies the following BUs and JRs involved in COMAU's PD process:

Business Unit	Job Role
Marketing	*Market Analyst; Advertisement Expert; Marketing Planner*
Research and Development	*Designer; Product Engineer; Process Engineer*
Sales	*Vendor*
Production	*Production Manager; Specialized Technician*
Purchase Department	*/*
Technical Department	*Designer; Product Engineer; Librarian*

Table 7.3: STEP 1 COMAU wrap up

STEP 2, Batch

BATCH B	
BAH Model Stage	Idea Generation (2), Screening & Evaluation (3)
Business Unit	Marketing
Job Role	Marketing Planner, Market Analyst

In this case, the "Idea Generation" Stage has a different connotation. In fact, it is not the Marketing itself to generate the idea, but indeed it collects the customer ideas and input them into the PD process. Therefore, from the point of view of the other BUs and BRs involved, the Marketing is the new idea generator.

BATCH C	
BAH Model Stage	Commercialization (7)
Business Unit	Marketing
Job Role	Advertisement Expert

BATCH D	
BAH Model Stage	Idea Generation (2), Design & Development (5), Testing (6)
Business Unit	R&D
Job Role	Designer, Product Engineer, Process Engineer

BATCH E	
BAH Model Stage	Testing (6), Commercialization (7)
Business Unit	Production
Job Role	Production Manager, Specialized Technician

Also in this case, we must clarify the meaning of "Idea Generation". We already pointed out that in COMAU it is the marketing that collects the customers' requests and therefore generates the idea for a new product. But then it is the R&D department that transforms the idea from "Here we need a new product" to "This is the new product". In other words, the marketing reports an idea, whereas the R&D gives structure to that idea.

The BAH Stage of "Testing" is present in both BATCH D and BATCH E because COMAU performs different kinds of testing: the tests performed in the R&D BU aim at checking if the product works in terms of stress bearing, system functioning engineering and technical feasibility. In other words, it is a test of the project. On the other hand, the testing activities done in the Production BU aim at checking if the manufacturing and assembly activities have been correctly performed and at certifying the product according to the rigid Industry and company quality standards.

BATCH F	
BAH Model Stage	Commercialization (7)
Business Unit	Sales
Job Role	Vendor

BATCH G	
BAH Model Stage	None - Support the PD process
Business Unit	Technical Department
Job Role	Designer, Product Engineer, Librarian

BATCH H	

BAH Model Stage	None - Support the PD process
Business Unit	Purchase Department
Job Role	No pre-defined role

STEP 2 of the model identifies seven BATCHes out of eight. BATCH A is missing because, again, the C-Level JRs do not take part in COMAU's PD process, and as a consequence they do not have a stage of the BAH Model to be linked to.
To sum up, STEP 2 identifies the following BATCHes involved in COMAU's PD process:

BATCH	BU	JR	BAH
B	*Marketing*	Marketing Planner Market Analyst	Idea Generation (2) Screening & Evaluation (3)
C	*Marketing*	Advertisement Expert	Commercialization (7)
D	*R&D*	Designer Product Engineer Process Engineer	Idea Generation (2) Design & Development (5) Testing (6)
E	*Production*	Production Manager Specialized Technician	Testing (6) Commercialization (7)
F	*Sales*	Vendor	Commercialization (7)
G	*Technical Department*	Designer Product Engineer Librarian	Support the PD process
H	*Purchase Department*	/	Support the PD process

Table 7.4: STEP 2 COMAU wrap up

<u>*STEP 3, Link*</u>

In the following table, we list the modules of Exalead that DS should bring inside COMAU's organization. The module "CloudView" is included in the final offering but is not listed. The reason is that, as explained in Chapter 5, Exalead is structured so that the vertical market software modules always need a base to be installed on, and this base is the very CloudView module. Therefore, we take its installation for granted.

BATCH	Exalead's Module
B	Exalead Onecall; Web Indexing
C	Exalead Onecall
D	Exalead Onepart; PLM Analytics
E	Exalead Onepart; PLM Analytics
F	Exalead Onecall
G	Exalead Onepart
H	Exalead Onecall; Exalead Onepart

Table 7.5: COMAU STEP 3 wrap up

STEP 4, Test

In STEP 1 and 2, the model identified seven out of eight possible BATCHes. Then, in STEP 3 we found at least one module of Exalead for each one of those seven BATCHes. Therefore, the test of STEP 4 filters none of the BATCHes, and all of them proceed in the Model.

STEP 5, Match

In this step for each BATCH we match the benefits Exalead's modules are bringing about. We do not list specific benefits for CloudView because, as stated before, its benefits are embedded with the modules it supports.

BATCH B	
BAH Model Stage	Idea Generation (2), Screening & Evaluation (3)
Business Unit	Marketing
Job Role	Pre-Sales Marketing Expert, Marketing Planner, Market Analyst
Requirements	...
ESE Module	Exalead Onecall; Web Indexing
Benefits	These two instruments allow the pre-sales marketing to make the most out of the Internet by creating a personal index made of useful websites, social networks, analysts' blogs, etc. In addition, Onecall allows marketers to aggregate the information out of both the Internet and the

	specific marketing instruments such as the CRM, the customers' libraries etc.

BATCH C	
BAH Model Stage	Commercialization (7)
Business Unit	Marketing
Job Role	After-Sales Marketing Expert, Advertisement Expert
Requirements	...
ESE Module	Exalead Onecall
Benefits:	Onecall provides data aggregation capabilities that are ideal to manage advertisement campaigns, especially if one has to deal with multiple advertisement channels and multiple sources of feedback.

BATCH D	
BAH Model Stage	Idea Generation (2), Design & Development (5), Testing (6)
Business Unit	R&D, Production
Job Role	Designer, Product Engineer, Specialized Technician
Requirements	...
ESE Module	Exalead Onepart; PLM Analytics
Benefits	Onepart was created to help designers, engineers and whoever in the enterprise may need to search for parts, components and assemblies that have already been designed before. The primary purpose of this tool is indeed to let the personnel know what kind of projects they already have in store so that they can whether re-use it or start designing a new project out of the previous one, thus saving time and costs. In addition to this, Onepart can be particularly helpful for chief designers and project managers to search into COMAU's case history and find out previous projects similar to the one they are commissioned to develop. The key driver to save time is to avoid starting from the green field. PLM Analytics tell designers and engineers about the advancement of the

	project, especially if they work in different team according to a concurrent engineering schedule. The combination of these two modules allow Process Engineers to have access to all of the past projects and specific layouts stored both in the PLM system and in the other repositories. Also, PLM Analytics gives them a view on the PD process, its advancement and its performance.

BATCH E	
BAH Model Stage	Idea Generation (2), Design & Development (5), Testing (6)
Business Unit	R&D, Production
Job Role	Process Engineer, Production Manager
Requirements	...
ESE Module	Exalead Onepart; PLM Analytics
Benefits	The Production Managers can also have access to the past production scheduling practices and use them instead of relying on their memory or re-designing a production plan from zero. Also, the Technicians who need to assemble the products can use Onepart to extract technical documents from the PLM system. They are not computer technicians and they are not familiar with software, therefore an intuitive and user-friendly interface can help them find the assembly instructions quickly and easily.

BATCH F	
BAH Model Stage	Commercialization (7)
Business Unit	Sales
Job Role	Vendor
Requirements	...
ESE Module	Exalead Onecall
Benefits	In COMAU's PD process vendors do not just deal with the financial and commercial part of the process. They also need access to the technical

	specifications of the products and to the PD timeline. In other words, they need a software tool to connect the CRM module of the ERP and the PLM systems. Therefore, Onecall allows vendors to have the information from the two systems in only one dashboard.

BATCH G	
BAH Model Stage	None – Support the PD process
Business Unit	Technical Department
Job Role	Designer, Specialized Technician
Requirements	...
ESE Module	Exalead Onepart
Benefits	Onepart helps the Technical Department by giving access to all the parts, assemblies, WIP and final products that are listed in the enterprise' software systems. Then, the Technical Department would use Onepart as a tool for document and project search and for the whole document management activity. The Librarians in particular need Onepart to search for meaningless duplicates or misaligned files to be corrected and then sent to designers, engineers or the production.

BATCH H	
BAH Model Stage	None – Support the PD process
Business Unit	Purchasing Department
Job Role	No pre-defined role
Requirements	...
ESE Module	Exalead Onecall; Exalead Onepart
Benefits	The Purchase Department needs both modules because Onepart gives a view on the insight of the enterprise, while Onecall is ideal to manage the information coming from the suppliers. The combination of these two modules makes the Purchase Department aware of what needs to be bought and what suppliers it should be purchase from.

STEP 6, Benefits Evaluation

BATCH B	
BAH Model Stage	Idea Generation (2), Screening & Evaluation (3)
Business Unit	Marketing
Job Role	Pre-Sales Marketing Expert, Marketing Planner, Market Analyst
Requirements	...
ESE Module	Exalead Onecall; Web Indexing
Benefits	...
KPI	• $\frac{\#\,Links\ followed}{\#\,Links\ presented}$: • $\frac{\#\,Issue\ spotted\ before\ the\ alert}{\#\,Issue\ reported}$:

BATCH C	
BAH Model Stage	Commercialization (7)
Business Unit	Marketing
Job Role	After-Sales Marketing Expert, Advertisement Expert
Requirements	...
ESE Module	Exalead Onecall
Benefits:	...
KPI	• $(query\ time\ AS - IS) - (query\ time\ with\ ESE)$:

BATCH D	
BAH Model Stage	Idea Generation (2), Design & Development (5), Testing (6)
Business Unit	R&D, Production
Job Role	Designer, Product Engineer, Specialized Technician
Requirements	...
ESE Module	Exalead Onepart; PLM Analytics
Benefits	...
KPI	• $(\#parts) * (\%\ duplicates) * (part\ cost)$

	• $ROI = \frac{(\#parts)*(\%\ duplicates)*(part\ cost)}{Installation\ Cost}$

BATCH E	
BAH Model Stage	Idea Generation (2), Design & Development (5), Testing (6)
Business Unit	R&D, Production
Job Role	Process Engineer, Production Manager
Requirements	...
ESE Module	Exalead Onepart; PLM Analytics
Benefits	...
KPI	• $\frac{\#\ Issue\ spotted\ before\ the\ alert}{\#\ Issue\ reported}$

BATCH G	
BAH Model Stage	None – Support the PD process
Business Unit	Technical Department
Job Role	Designer, Specialized Technician
Requirements	...
ESE Module	Exalead Onepart
Benefits	...
KPI	• $(\#parts) * (\%\ duplicates) * (part\ cost)$

BATCH H	
BAH Model Stage	None – Support the PD process
Business Unit	Purchasing Department
Job Role	No pre-defined role
Requirements	...
ESE Module	Exalead Onecall; Exalead Onepart
Benefits	...
KPI	All of the KPIs listed have to be submitted to the subjects of BATCH H.

	The KPIs will determine the opinion of the Purchasing Office about the ESE.

7.4 Results

7.4.1 Gap Analysis

At the end of DS' sales procedure, the company details the following offering for COMAU:

- Exalead Onepart * (#Designers + #Product Engineers + #Process Engineers + #Librarians + #people in the Purchase Department)
- PLM Analytics * #Product Engineers

At the end of the Model's implementation, DS would detail the following offering for COMAU:

- Exalead Onepart * (#Designers + #Product Engineers + #Process Engineers + #Librarians + #Production Managers + #Specialized Technicians + #people in the Purchase Department)
- PLM Analytics * (#Product Engineers + #Production Managers)
- Web Indexing * (#Marketing Planners + #Market Analysts)
- Exalead Onecall * (#Marketing Planners + #Market Analysts + #Advertisement Experts + #Vendors + #people in the Purchase Department)

The gap between the tow offerings is the following:

- Exalead Onepart * (#Production Managers + #Specialized Technicians)
- PLM Analytics * #Production Managers
- Web Indexing * (#Marketing Planners + #Market Analysts)
- Exalead Onecall * (#Marketing Planners + #Market Analysts + #Advertisement Experts + #Vendors + #people in the Purchase Department)

7.4.2 Findings

The main difference between the two approaches resides in the way of thinking the ESE. DS adopts a bottom up approach: starting from the way COMAU is classified, DS matches it with the vertical market solution that is specifically pre-designed to address that industry. On the other hand, the only necessary pre-requisite of the Model is that COMAU is a manufacturing enterprise by all means. Then, the Model does not match COMAU with a specific vertical software solution but it would rather link the company to the technology of ESEs. In fact, the Model does not take into consideration the brand of the ESE until STEP 3 (Link), when the details of the ESE are needed in order to propose an offering.

According to the gap analysis of paragraph 7.4.1, the Model allows DS to put forward a far more comprehensive offering than the one presented with DS' standard procedure. The key difference is that a systematic approach based on less restrictive premises is a better way for a software vendor to approach an ESE proposal. The Model's schematic mapping of the PD process would help DS identify the needs of every single role, thus proposing a larger offering. Out of the Model's scope is the purchasing power of COMAU, and we do not address the issue in this study.

What the Model aims at is suggesting the best proposal to implement an ESE in COMAU. This use-case confirms that the knowledge and the processes are embedded with the JRs and BUs, and the Model addresses them correctly.

8 Model Validation - Experts' Opinion

8.1 Summary

The second way that was chosen to test the model was to interview a pool of experts in the field of ESEs. As stated in Chapter 6, this model is meant to be a tool for sales and pre sales people to evaluate a potential customer of an ESE, but it is also the result of research and literature review. As a consequence, we chose two different kinds of interviewees:

- The software reseller and implementer Altea. In particular, we interviewed two business profiles: the head of vendors and the head of tech-sales. After an introduction to the context of the study, to the model and its limitations, they were both asked the same questions in the same order. We chose three dimensions, three points of view to evaluate each answer, and we assigned a score from 1 to 10 to evaluate each dimension. Then, we calculated the arithmetic mean for every question and we put the scores in a schematic framework for analysis and to draw our conclusions. The conclusion shows the opinion of the interviewees on the theoretical and practical robustness of the Model. We believe the opinion is particularly valuable because a likely user of the Model gives it out. Also, we evaluate the answers in terms of future evolution of the field of ESE and future application of the Model.
- An independent expert in the field of ESEs. In particular, we asked to a long time consultant in the field of PLM to give us an opinion on the Model and the whole study. We state his opinion to be "independent" because he does not sell software, but he is just focused on the study of implementation and application methods. We sent him an extended abstract of the study and we a remote-device interview in which we presented the Model. We then asked him to evaluate the Model in terms of theoretical and practical value and how he would consider it in a future perspective.

8.2 Enterprise Software Vendor and Implementer

8.2.1 ALTEA – The Company

Atea S.p.a. is an Italian group of complementary organizations based in Baveno, Italy. With over 20 years of experience in system integration and organization, 6 locations in Italy, 400 employees and more than 950 customers worldwide, Altea is a steadily growing organization and a proven leader in the field of technology-related services. Altea operates in close contact with industrial and service enterprises, basing its vision on innovation and sustainability, with a 360° approach on three guidelines: people, system and organization. Human resources are Altea's fundation whhereby the company expresses its professionalism in fulfilling customers' requirements. Using the latest Enterprise Social Solutions, Altea fosters collaboration and knowledge sharing, improving the information flow inside the enterprises. Over the years, Altea has developed strong partnerships with the main international vendors of software systems, such as SAP, Microsoft and Infor, showing specific focus on technological, organizational and change management aspects. Altea's broad offering includes knowledge and competences in applicative solutions for enterprise's management (ERP, CRM, CPM, SCM, BPM and business analytics), in collaboration technologies (Mobile, Cloud, Social, Big Data). In addition, Altea's offering is supported by specific vertical market software for every market and industry. Altea's own organization model is based on investments and strategic alliances with highly innovative and specialized enterprises. Altea's group of organizations includes, among others, ALTERNA, leader in Microsoft's Solutions and Technologies, PLM SYSTEMS, focused on the technical and managerial feasibility of new product development projects, BRIEFING, competence center for SAP CRM (Customer Relationship Management) and SAP FSCM (Financial Supply Chain Management), HELIANTHOS, reference partner for Organizational and Managerial Consulting, and CYRIUS, competence center for the Automotive Industry and Supply Chain Management Solutions. These are all the actors that enrich and broaden Altea's offering [6i].

8.2.2 ALTEA – Interview

The interview is made of an introduction to the model and six open questions. The following is the full text of the interview.

Introduction

The interviewee is introduced first to the study as a whole, and then to the Model and its explanation by reporting paragraphs 6.1 and 6.2. After this brief introduction, he is asked the following questions:

1) *Do you think the management of a manufacturing enterprise would be ready to consider the value of an Enterprise Search Engine in all of its aspects, and not just as a detached vertical market solution?*
2) *How useful do you think it is to have such a standard procedure to carry on the selling process of an Enterprise Search Engine?*
3) *Do you think the STEPs in this order focus on the right target to address the selling process?*
4) *Do you think the KPIs are suitable to convince a potential customer enterprise?*
5) *Would you apply this model when dealing with a prospect customer?*

The interviewee is asked each of the listed questions. Each answer is then evaluated according to three axes:
a = From 1 to 10, how good is the model in theory?
b = From1 to 10, how good is the model in practice?
c = From 1 to 10, how good will the model be in the future?
This subdivision is then reported in the final summary chart. See paragraph 1.3.3 for further explanation on the methodology behind the interview and its evaluation.

8.2.3 Interview – Sales

<u>Question 1:</u> Do you think the management of a manufacturing enterprise would be ready to consider the value of an Enterprise Search Engine in all of its aspects, and not just as a detached vertical market solution?
I would love to say "yes", because I do think that ESEs are not to be considered detached vertical solutions but powerful multiple information retrievers. As we are talking from a sales point of view, I would say that the average Italian manufacturing enterprise still looks for the solution that is specific for their business rather than a technology that enhances the enterprises' performances as a whole. For example, when I approach tiles manufacturers, they stress the importance of their specific business area, they ask for customer references from that business area and they want specific solutions for that business area. If you do not show them the value for their specific tasks, they would say that your software is worthless. Another example is the industry of pumps. There are several kinds of pumps: centrifuge pumps, air pumps and so on. After all, they are all pumps and the PD process to create them is almost the same. But still,

manufacturers claim they make air pumps and not water pumps, so they need something that helps them making that specific kind of pump and not just pumps in general. To my view, it is wrong to reason in this way. If you talk to a Technical Department, they should know that the concept is the same behind any pump. Yet, this is what happens in most cases on the Italian market. I am sure abroad the situation is quite different. On the other hand, if you focus your research on the selling process (i.e., Phase 7 of the BAH model), then managers and directors can appreciate this solution, because this technology has yet to be applied to the world of sales operations, no matter the industry. It is also important to notice that it is not just me thinking that the Designers should become less isolated in the PD process, but it is a trend that has already started to take place. For example, one of my customers manufactures welding machines. Until a few years ago, its R&D and Technical Department used to waste a lot of money in products they would never put on the market. This firm is a rich one, so in the beginning the owner did not care about this issue, but then he and the management started to force a change. They wanted the Technical Department to be aligned first of all with the Production Department, because I cannot waste resources in what is impossible to be realized, and then with the Vendors, because they know the market and can decide whether a product is good for commercialization or not.

a = From 1 to 10, how good is the model in theory? 8

b = From 1 to 10, how good is the model in practice? 5

c = From 1 to 10, how good will the model be in the future? 7

Score: 6.7/10

<u>*Question 2:*</u> *How useful do you think it is to have such a standard procedure to carry on the selling process of an Enterprise Search Engine?*

I think the good of the standard is that it also makes it general. Nowadays, enterprises cannot be classified into single industries. That kind of rigidity belongs to the past. Therefore, the mapping of the processes has to be drawn back to the single organization. It is right to start from the processes and map them down starting from a general scheme that is not too specific for a certain industry or a certain structure. Companies look for alternative ways of doing business in order to gain competitiveness and to be flexible in a fast changing market. It is wrong to start from a vertical solution designed on a pre-defined process. For example, one of my customers manufactures handles for windows and doors, but during the last two years he has been

progressively entering the photovoltaic business. In the past, mapping a software solution on this customer would have been trouble, because you needed the vertical systems for the industry "Industrial Equipment" to be integrated with the vertical for "Energy and Power", which were two worlds apart. It was an absolute mess. So, if I have a standard procedure that starts from the general, I no longer have this problem. On the other hand, according to my experience, in every kind of enterprise, especially in small enterprises, everything related to design and planning is totally detached from sales operations and production. I would say they are not just two worlds apart, but they are really antagonists. When you stated that the two should be tied and linked, you are absolutely right, because this is how the PD process should be structured. But the truth is, in small enterprises the R&D is the core of the innovation, then the sales test its applicability, and this is the end of the PD process. So, the PD process is extremely different between small and big enterprises. This element must be taken into consideration when building a procedure.
a = From 1 to 10, how good is the model in theory? 8
b = From 1 to 10, how good is the model in practice? 5
c = From 1 to 10, how good will the model be in the future? 8
Score: 7/10

Question 3: *Do you think the STEPs in this order focus on the right target to address the selling process?*
Related to what we said before, I think it is right to start from the business processes in a general way. To be sincere, I think that the third STEP is already a rather old-fashioned idea. Nowadays, the real challenge is making the technology so intelligent that, in the future, Enterprise Search Engines will be able to certify by themselves what kind of industry the enterprise belongs to, and then set themselves up to the processes of that industry. To be specific, the ESE could use the queries performed by the users in order to address the macro business area of the enterprise. Therefore, it will be the ESE to make itself vertical according to the needs of the users. Just to give you a simple example, if I install an ESE and the users search ten times the word "pills", I would like the ESE to get that the enterprise belongs to the industry "Medicals".
In the past this was considered science fiction because we did not have the technology to make it happen. We had the DOS, the 286 series microprocessors, but nowadays we do have the technology to make this happen, we can search through terabytes of data and have a real time

response, especially when we talk about something like Exalead, which is very "deductive". I would say that I want to start from something general, but then I want the technology to be tailored on the enterprise, to be suitable for its own processes. Again, the idea behind these two STEPs is good, but they will be soon part of the technology, not of the procedure to apply it.

a = From 1 to 10, how good is the model in theory? 7

b = From1 to 10, how good is the model in practice? 7

c = From 1 to 10, how good will the model be in the future? 5

Score: 6.3/10

<u>Question 4:</u> Do you think the KPIs are suitable to convince a potential customer enterprise?

It really depends on the KPI you set. For example, if the KPI gives me straight away the amount of money I save when designing a product, of course the management will take it into account. Nowadays manufacturing enterprises play their competition game on the profit, on the margin they have between market price and costs. The same thing for the PD process costs: how much does the infrastructure to design a product cost? If you can answer this question with a KPI, the management is going to be convinced. But then again, the ones to be effectively convinced would be the top managers. If you go down the organizational chart, everybody starts caring about their own matters rather then the enterprise as a collective. For example, you may have one or two people dedicated to the calculus of the product cost. Then, they realize that an ESE combined with a PLM system takes less then ten seconds to calculate that cost. As a consequence, the ESE becomes a thereat to them, the threat of losing their job. Again, this happens especially in small enterprises, but it is an aspect not to be under estimated.

a = From 1 to 10, how good is the model in theory? 7

b = From 1 to 10, how good is the model in practice? 5

c = From 1 to 10, how good will the model be in the future? 9

Average score: 7/10

<u>Question 5:</u> Would you apply this model when dealing with a prospect customer?

The model addresses the problem in the right way, it is technically correct and it follows the technological trend of the latest years. The technology of enterprise search follows the same historical evolution of the ERP systems: they were first thought as a comprehensive technology to be declined on the specific needs of the enterprises. With the passing of time, ERP systems have been progressively split into modules until those modules became no more part of a system, but single detached products. As a consequence, the customers started to purchase the single micro models from different vendors, thus making the customer enterprises' structure like a puzzle of software systems. At the same time, the system integrators became specialized in the implementation of single modules for single Industries. This model was not sustainable in a world were it gets harder every day to classify enterprises in single rigid industries. The splitting process reached a point that it was impossible to integrate the modules from the same vendor because they had been developed for a different industry and, in the mind of the R&D, they were not to be mixed. Since the late 2000s, the general trend has changed. The main ERP vendors decided to get back to the original idea, which is to sell a technology with two or three core competences, and then develop single module solutions according to the process of the customers, and not to the pre-defined industries. I like the idea of a model that follows this concept: a technology that supports a whole process, and then the declination of the technology on the single needs. This is the right idea to follow and the proof is that the world is moving in that direction. Still, I sense the model is too raw to be applied as it is. It is still too scholastic and too general to be applied on my average customers: I think the problem is not in the model itself but rather in the fact that each small enterprise is different from the other, so I would rather use it for big or multinational enterprises.

a = From 1 to 10, how good is the model in theory? 9

b = From 1 to 10, how good is the model in practice? 4

c = From 1 to 10, how good will the model be in the future? 9

Average score: 7.3/10

8.2.4 Interview – Technical Sales

<u>*Question 1:*</u> *Do you think the management of a manufacturing enterprise would be ready to consider the value of an Enterprise Search Engine in all of its aspects, and not just as a detached vertical market solution?*

The answer is yes, and the reason for me to answer this way is that usually the top management or whoever is in charge of making the final purchase of the ESE does not understand the advantages of the single vertical software solutions. For example, a company's CEO rarely takes into account the everyday problems of a junior designer in the Technical Department. The CEO has a different background, a different perspective and likely he has never used a CAD in his life. On the other hand, if you tell the CEO that you have a vertical solution that can help all the designers so that the company will save money, now that is a driver he understands. The point is: the management wants to see the advantages of a software systems from above, covering the whole enterprise and translated in a logic of strategic competitiveness. In the future, this concept will be even more important because the top management will not even expect the vendor to sell a software system, but he will want to have a software solution, no matter the verticals behind it.

a = From 1 to 10, how good is the model in theory? 9

b = From 1 to 10, how good is the model in practice? 9

c = From 1 to 10, how good will the model be in the future? 9

Score: 9/10

<u>*Question 2:*</u> *Do you think the STEPs in this order focus on the right target to address the selling process?*

As I said before, the procedure is indeed useful, but not for every kind of enterprise. Big enterprises would prefer such an approach because of the way they are structured and because it is easier to address them following such a procedure. On the other hand, small or SMEs are not used to this approach, they would not understand it. They need to be addressed directly with a vertical software solution to solve a specific problem, no matter the job roles or the activities performed by the personnel.

a = From 1 to 10, how good is the model in theory? 6

b = From 1 to 10, how good is the model in practice? 4

c = From 1 to 10, how good will the model be in the future? 5

Score: 5/10

<u>*Question 3:*</u> *Do you think putting the STEPs in this order allows the Model to rightly address the selling process?*

It depends on the kind of enterprise and on the visibility the company wants to give to the software vendor. I would say that it totally fits the theory. If a company would disclose its entire structure to me I would set up 5 or 6 scenarios of typical "issue" of a certain job role, and then I would explain the solution through an ESE. In real life, it is hard to find such a situation. Some customers agree with disclosing the information, so we do know everything to go with the procedure you put forward. Some others, especially the prospects, are more reticent.

a = From 1 to 10, how good is the model in theory? 9

b = From 1 to 10, how good is the model in practice? 5

c = From 1 to 10, how good will the model be in the future? 7

Score: 7/10

<u>*Question 4:*</u> *Do you think the KPIs are suitable to convince a potential customer enterprise?*

It depends on the need you are addressing. For example, if I were trying to sell Geolus or Exalead Onepart, it would be easy to convince the customer that it is a good deal. You can just show him how much he saves on the reuse of parts. But if we talk about an ESE in general, it may be hard to show the results with a KPI. It is more about convincing the customer of the limitless possibilities of a new technology.

a = From 1 to 10, how good is the model in theory? 5

b = From 1 to 10, how good is the model in practice? 4

c = From 1 to 10, how good will the model be in the future? 5

Average score: 4.7/10

<u>*Question 5:*</u> *Would you apply such a model when dealing with a prospect customer?*

I would say the procedure is valid and gets to the point. Once more, it really depends on the customer. I would absolutely use such a procedure if I were dealing with a big manufacturing enterprise that I already have in my customer list, I would definitely apply a similar procedure. However, I do not think your Model is suitable for a SME. I would say that, from the point of view of a university student this Model perfectly addresses the problem. At the moment the real world is still a step behind this.

a = From 1 to 10, how good is the model in theory? 9

b = From 1 to 10, how good is the model in practice? 5

c = From 1 to 10, how good will the model be in the future? 7

Average score: 7/10

8.3 Independent Expert Interview

The interviewee is John Stark, president of the John Stark Associates.

Stark is a leading developer and implementer of the business theory of Product Lifecycle Management. He is the author of two books about PLM ("Product Lifecycle Management: 21st Century Paradigm for Product Realization" and "Global Product: Strategy, Product Lifecycle Management and the Billion Customer Question"), has written many articles on the subject, and is a frequent keynote speaker at PLM conferences. In his long consulting career, Stark has worked with many companies including ABB, BMW, Braun, Coca-Cola, Eaton, Ford Motor Company, HP, IBM, Kodak, Legrand, Nestle, PSA, Renault, Saab, Schindler, Sikorsky, Valeo and Xerox [8i].

Stark has been involved in an unstructured interview in which we asked him to evaluate the Model according to its theoretical and practical value. In addition to this, we asked his opinion about the Model's addressing the future of the manufacturing and enterprise software fields.

8.4 Results

8.4.1 Scores

The chart below shows the results from the interview with Altea in terms of score for each question, score for each axis and average scores for all the categories.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Question	**1**			**2**			**3**			**4**			**5**			Average			
2		**a**	**b**	**c**	**a**	**b**	**c**	**a**	**b**	**c**	**a**	**b**	**c**	**a**	**b**	**c**	**a**	**b**	**c**	
3	**Sales**	8	5	7	8	5	8	7	7	5	7	5	9	9	4	9	7.8	5.2	7.6	6.9
4		6.7			7			6.3			7			7.3						
5	**Tech Sales**	9	9	9	6	4	5	9	5	7	5	4	5	9	5	7	7.6	5.4	6.6	6.5
6		9			5			7			4.7			7						
7		**a**	**b**	**c**	**a**	**b**	**c**	**a**	**b**	**c**	**a**	**b**	**c**	**a**	**b**	**c**	**a**	**b**	**c**	
8	Average	8.5	7	8	7	4.5	6.5	8	6	6	6	4.5	7	9	4.5	8	7.7	5.3	7.1	**6.7**
9		7.8			6			6.7			5.8			7.2						

Table 8.1: Scores from the interview of Altea

8.4.2 Altea – Interview Findings

Referring to Table 8.1 (Row 1, 2, 3.... ; Column a, b, c...), we can deduce the following findings about the Model.

The Sales and Technical Sales experts did not have the chance to talk about the interview before it took place or to take any agreement about how to answer, yet their opinions are matching.

According to the Sales expert, the Model has a strong theoretical background (3, 4 ; Q) and is likely to become a best practice in the future (3, 4 ; S). On the other hand, at the moment the Model lacks of realism, as it would be hard to implement it in real life (3, 4 ; R). To sum up the Sales expert's opinion, the Model is generally robust (3, 4 ; T), because it is good in theory and it moves in likely to become a standard in the future, but for the moment it is still too far apart from the small enterprises' mentality.

According to the Technical Sales expert the Model is also strong in theory (5, 6 ; Q) and is set on the right path for the future (5, 6 ; S), although the Technical Sales person is less optimistic than the Sales expert (6.6 score vs 7.6). However, also according to the Technical Sales expert, the Model is still not suitable for the real world (5, 6 ; R).

Overall, the applicability of the Model in the real world obtains has scored 5.3 (8, 9 ; R). According to the interviewees, this is mainly due to the fact that the Model is more suitable to implement ESEs in big enterprises rather than in small-medium and small ones.

Question 4 is about the KPI selection and application, and it is the question that received the lowest scores of all: 5.8 (9 ; K, L, M). This is due to the fact that ESEs for the support of the PD process are associated with PLM software systems. These types of software systems have always been difficult to evaluate using performance or financial KPIs, because many of the advantages they bring to organizations are hard to measure. However, the score is still above 5 out of 10, because to the experts' view, the Model fits well in big organizations.

Question 1 is about how open minded the management of manufacturing enterprises is and in what measure the Model is useful to convince them of purchasing an ESE. This question received the highest score of 7.8 (9 ; B, C, D). In fact, the Sales expert assigned a medium score of 6.7 (4 ; B, C, D), but the Technical Sales person gave a high score of 9 (6 ; B, C, D). Surprisingly, the Technical Sales person is even more convinced than the Sales expert that ESEs have to be adopted as a whole and not just in a single vertical solution. Due to his technical background in computer science, the Technical Sales expert states he can certify the technology is ready for an enterprise level implementation, at least in the context of big manufacturing enterprises.

A further support to the value of the Model comes from the fact that only two answers out of ten received a score equal or lower than 5, and from the Technical Sales expert: (6 ; E, F, G) and (6, K, L, M).

In conclusion, the experts' opinion is that the Model is generally robust, because it gets an overall score of 6.7 (9 ; T). In other words, the experts do validate the Model.

8.4.3 Independent Expert Interview Findings

According to Stark's opinion, ESEs have being discussed in this field only for a couple of years. He has been working in the context of PLM for more than two decades, but still he feels the need to read other experts' documents and reports in order to get prepared for the interview. Therefore, he points out that the Model deals with a very pressing topic that it is destined to increase its importance over the next few years, thus making the Model very future-oriented.

Stark thinks the Model addresses the problem in the right way. In his own consulting approach, people and processes come first, so the first thing to do in a deployment process is to map the enterprises' JRs and BUs in order to have a clear idea of the criticalities. The main issue is having the customer enterprise disclose all of its PD process structure. It is his first requirement for the customers prior to his intervention.

Stark also states in his everyday job the reasons for proposing the Model are indeed problematic: big enterprise software vendors focus solely on their customers' Industry, thus losing the big picture. Manufacturing enterprises do not only deal with production or mechanical design issues, but they are also deeply involved in the operations field just like any other profit-oriented enterprise. As a consequence, software vendors must propose ESEs as a whole and not just as single vertical market solution modules.

For the completeness of the study, we must mention that Stark's focus is only on big enterprises. Therefore, he evaluates the Model according solely to the point of view of a consultant for big enterprises, where usually PLM and the other enterprise software systems are already a solid part of the company's life.

9 Conclusion & Future Research

9.1 Conclusion

According to the literature review, the proposed Model and its testing, we conclude the study by stating the following objectives as of paragraph 1.2, namely:

Objective 1: To critically review the literature about Enterprise Search Engines and the software for Big Data management in a manufacturing context.
Objective 2: To demonstrate that Enterprise Search Engines are an effective tool for manufacturing companies to support the designing and engineering of products.
Objective 3: To examine the characteristics of Enterprise Search Engines and shift the focus from single vertical market solutions to the technology as a whole.
Objective 4: To match the advantages and benefits of Enterprise Search Engines with the requirements and needs of the organizations.
Objective 5: To develop a standard model for the deployment of Enterprise Search Engines to manufacturing enterprises.

have been fully achieved. On the other hand:

Objective 6: To validate the developed model.

has been only partially achieved. In the following paragraphs we detail our fulfillment of the objectives.

9.1.1 Theoretical Contribution

This study describes in detail the reasons for a link between ESEs and PD. In Chapter 2 we give proof of the importance of Big Data in the manufacturing world and describe the complexity of Big Data handling as a consequence of a steadily increasing technological environment. In Chapter 3 we describe how the traditional software architecture deals with Big Data handling and how this infrastructure still needs to be advanced in order to fulfill the requirements of modern manufacturing enterprises. We thus provide a connection between the PD software infrastructure and ESEs (*Objective 1*). In Chapter 4 we give a detailed description of ESEs along with their history and practical applications. In particular, we focus on their features to support all the stages of the PD process in manufacturing enterprises, from the pure designing and engineering of products to collateral activities. In Chapter 5 we emphasize the concepts of Chapter 4 by describing a state-of-the-art ESE (*Objective 2* and *Objective3*). In these two Chapters we also match one by one all the manufacturing enterprises' needs listed in Chapter 3 with all the benefits of ESEs that we detected in our study (*Objective 4*).

We can state that this study puts forward some significant advancements for the literature and the knowledge in the field of ESEs.

9.1.2 Practical Contribution

In Chapter 6 we put forward a proposal for a systematic method to match the list of common enterprises' needs with the characteristics of ESEs. This Model works as an effective procedure to convince manufacturing enterprises of the benefits of ESEs (*Objective 5*). In addition, the Model is structured in order to vehicle the value of ESEs to all the people involved in the PD process. In other words, the Model does not focus on the single vertical market solutions but on the technology as a whole (*Objective 3*).

The practical value of the Model is tested in two distinct scenarios. In Chapter 7 we describe a use-case in which the Model is compared to the procedure of a best-in-class software vendor. In Chapter 8 we asked a Sales and a Technical Sales person of another software vendor and implementer to express their opinion about the Model. In both cases the Model demonstrated to have practical value and to correctly address the current issues of manufacturing enterprises. In particular, the Model proves to be more effective than the status quo, thus being a useful and innovative tool for software vendors and implementers (*Objective 6*).

9.1.3 Future Applicability

Both experts from Altea and the independent expert agree on the Model being not just something new in the field, but also a procedure that correctly addresses the future needs of manufacturing enterprises. Both the Sales and Technical Sales people state that manufacturing enterprises will focus on comprehensive tools that help them connect all the software systems they have implemented so far. Besides, enterprises need to realize the importance of ESEs as a whole and not just as solvers of a single specific problem in a given industry. The independent consultant in particular foresees ESEs will follow the same evolution path PLM systems have already experienced: the theory came a long time ago, but nowadays enterprises still have problems understanding their value and importance. However, ESEs are likely to become the best tool to make the ones of the old infrastructure connect and work well together. In particular, their great advantage is that they are already set to be integrated with any other software system that will possibly be invented in the future.

9.1.4 Lacks & Flaws

From the Model validation process, the main emerging flaw is that the Model fits well big enterprises but is hardly scalable to small and small-medium ones. The experts think this happens for several reasons. First, compared to big companies, small enterprises have a totally different way of approaching the PD process. Very often small enterprises have nothing like a fixed process or even a standard infrastructure made of departments, units and functions. Especially in Italy, small enterprises are built around a close group of individuals, idea generators that come out with a key innovation leading the entire business. The Model is based on a fixed scheme that is typical of the industrial world, it is therefore impossible to fit the Model where no scheme or structure can be found. Second, even well structured small enterprises seldom have the cultural background to accept new and innovative software tools. Just like CAD in the 1980s and PLM nowadays, ESEs require training prior to their installation. At the moment IT, CAD and PLM managers of small enterprises see no utility in ESEs. On the other hand, as proven by Altea, they are interested in the single vertical market solutions of each ESE.

Third, small enterprises are not accustomed to complex PLM or ERP environments. As a matter of fact, small companies still rarely implement comprehensive software infrastructures, preferring to rely on single CAD or PLM seats without connecting them through a process manager such as the PLM. At the same time, they do train people to use CRM, CMMS and other ERP modules, but they do not build the enterprise around the entire ERP infrastructure. For this reason, small enterprises feel the benefits of ESEs would be undermined by the implementation time and costs. They are afraid that ESEs would be a complex tool that would change their way of working just to let them search for data for which they already know the location.

To conclude, the Model is not suitable for small and small-medium enterprises.

9.2 Future Research

9.2.1 Big Data and Product Development

In this study, we concentrated our efforts on the PD process of manufacturing enterprises. Therefore, we took for granted the paradigm of the traditional enterprise that acts just as product manufacturer. The truth is, particularly in the Western world, nowadays companies seldom develop mere objects, but they would rather combine the simple hardware with some other aspects in order to differentiate the final output from the one of their competitors. We already mentioned the case of COMAU (Chapter 7), whose output is made of both industrial robots and the customer's production process as well as plant layout. In fact, COMAU often outsources its hardware part of the product and delivers only the layouts and processes for the final production plant. This is the typical case of Product-Service System (PSS) Design, when the product is matched with a related service.

Hardware & Software

Manufacturing enterprises from the field of IT have long since developed products that combine hardware and software. Nowadays, a lot of enterprises in the field of Industrial Equipment and High Tech also decide to co-develop both hardware and software in their machines. For example, SMEs that produce mechanical lathes now tend to design also the software for the computer numerical control. As a consequence, the PD process includes new BUs and, most important, JRs that contribute to the final output of the enterprise.

The Model this study puts forward is tailored on generic manufacturing enterprises, but it would be interesting to improve it so it would be able to address the needs of software and service designers. This way, the Model would also foster a new study of the PD process. As a matter of fact, the BAH Model is a standard and generic study, while reshaping it to fit the designing of services would mean a considerable advancement for the literature.

Cross-Industry Enterprises

More and more enterprises are developing new ideas of doing business. Some enterprises just design products, leaving the production to other entities, some other would only deal with the after sales services, and so on. The point is, nowadays there are several enterprises that do business with the world of manufacturing without having the standard structure of manufacturing enterprises. Also, there is a tendency to associate the term "manufacturing" with the field of mechanics (Aerospace, Automotive, Industrial Equipment etc.), yet medical, food and beverage companies also manufacture certain products. Therefore, the same Model we put forward in Chapter 6 can be applied to a variety of enterprises, but it needs to be adjusted for the purpose. The entire Model is based on the match between JRs and BUs involved in the PD process and the standard structure of the same process provided by the BAH Model. In order for the Model to work for realities other than manufacturing, it would need to be matched with a different model for PD, as well as other JRs and BUs need to be identified and mapped.

9.2.2 Enterprise Search Engines

The first hint for future research comes from the field of ESEs. As this study mentioned in Chapter 4, ESEs have evolved during their short life and they are still subject to innovation. In particular, the study mentions the fact that searching for data means searching for information, thus aiming at acquiring new knowledge. Nonetheless, what the end user will want in the future is not just a matching document, but something that represents an answer [11]. Besides, having access to data does not necessarily mean understanding its content. Especially when dealing with technical issues, we would prefer requesting an expert 's opinion rather than have access to some data about the issue. When looking for expert knowledge, we usually turn to people we can ely on for quick information or recommendations. However, personal networks are sometimes not sufficiently diversified to reach the right contacts [71]. For this reason, one of the improvements that researchers want to apply to ESEs is to enable them to act as experts search systems.

Expertise is the embodiment of knowledge and skills within individuals. An individual may have different levels of expertise about different topics, and the expertise distinguishes experts from less experienced people and novices. Expertise search describes the process of seeking to find the people who might have the desired knowledge and skills [72]. The idea of expertise search to be integrated with ESEs is based on the fact that enterprise corpora contain evidence of what employees work on and therefore can be used to automatically find experts on a given topic [70]. Indeed, enterprises use to have a sort of catalogue containing all the information about its employees. This catalogue is often part of a software system for personnel's data storage, protection and management. Some systems use the profiles that indicate their expertise; such profiles may be obtained from different information sources such as curriculum vitae, publications, blogs, web sites and research project details. Combining such different data together may require gluing of vocabulary [72]. An expert recommender is a system which identifies people with particular skills and experience, can be a valuable management tool to promote collaboration and increase productivity by supporting knowledge sharing and transfer within and across organizations [70].

The idea itself has already been developed both in theory and in practice. Initial approaches to expert finding employed a database housing the skills and knowledge of each individual in the organization. The database would be manually constructed and consulted. Such approaches required considerable effort to set up and maintain. Consequently, various automated approaches have been devised to mine the information repositories of organizations from which to build a profile of an employee's expertise [31]. Recent work on automatic expert finders has formulated the problem of determining who has knowledge in a particular area as a retrieval task to rank people given a query topic [70]. Still, according to Altea's (Chapter 8) experts, actual ESEs are not advanced enough to support a proper search for enterprise expertise. In fact, ESEs can be addressed to look for experts within the enterprise, but this does not necessarily mean they can help users in finding the answers they need. First of all, ESEs look into some kind of fixed database: resumes, research projects and so on are documents that are seldom updated, and they always describe someone's expertise "up to some point". What is really interesting is to provide a tool that tells the user who is working on a certain topic at that very moment when the user is performing the query. In other words, other than telling who is the expert according to his background, ESEs should be able to tell who is working on that topic in that period of time and who already searched for certain information. This could save a lot of time when assembling a team for a new project. Also, this can encourage spontaneous teamwork inside the enterprise. For example, if a designer is told to design a brand new clamp, he could use the local ESE to check if someone else inputted the query "clamp" in the last month. In this way, he may be able to find out who is the most update person in the field of clamp designing, who is interested in clamps and why, who could provide him further information and who could have special suggestions about the design.

Again, the key feature of ESEs remains the capability of acting real time. If the ESE can only check for resumes that are a decade old or it takes days to process the query and find the experts, the tool would have absolutely no value in a business world that demands for quickness and flexibility.

The Model this study puts forward is specifically designed for addressing ESEs as support tools for the PD process. However, despite being the core of manufacturing enterprises, PD is not the only process for that kind of companies. The everyday life of enterprises is regulated by the repetitive operations that make them running efficiently. If the world of PD is associated to PLM software systems, the world of management and operations is matched with the ERP. As

Why only PD process? From the model test we found out that many other BUs and functions may benefit. But, a whole new study, a whole new KPIs calculation. Boost the limitations: economic aspect, organization aspect, more people to be envolved, more diversity -> harder to find a standard model.

9.2.3 EXALEAD

As any other ESE, Exalead can also be improved with the addition of other expertise search tools. However, there is also another feature that would dramatically boost Exalead's performances and value: Business Intelligence (BI).

In the industry reports and competitors analysis, Exalead is reported as a tool for BI. In fact, Exalead is capable of collecting data from various systems and repositories as well as displaying them in graphs, charts and dashboards. In particular, Exalead can collect all the information about the advancement and cost of the PD process from all of the software systems dedicated to it. Then, Exalead can display them in issue-tracking dashboards. Anyway, these features do not automatically make it a tool for BI, because BI is more than this.

Business Intelligence (BI) can be defined as the process of turning data into information and then into knowledge. Knowledge is typically obtained about customer needs, customer decision making processes, the competition, conditions in the industry, and general economic, technological, and cultural trends. BI was born within the industrial world in the early 90's, to satisfy the managers' request for efficiently and effectively analyzing the enterprise data in order to better understand the situation of their business and improving the decision process [73]. Therefore, BI software is mainly used by the management of the company, and it expects such a tool to perform calculations, analysis and predictions out of the extracted data. As a matter of fact, a critical component for the success of the modern enterprise is its ability to take advantage of all available information. This challenge becomes more difficult with the constantly increasing volume of information, both internal and external to an enterprise. It is further exacerbated because many enterprises are becoming increasingly “knowledge-centric”, and therefore a larger number of employees need access to a great variety of information to be effective [74]. Therefore, the new requirement of managers is to ensure that all processes are effective by continuously measuring their performance through Key Performance Indicators (KPIs) and score cards [73]. Adding KPIs, score cards and, most of all, the calculus capabilities to generate them would make Exalead a far more competitive Enterprise Search Engine.

Several authorities in the field of technology and management trends support the theory that BI and BI analytics will play a key role in the future. The IBM Tech Trends Report (2011) identified business analytics as one of the four major technology trends in the 2010s. In a survey of the state of business analytics by Bloomberg Businessweek (2011), 97 percent of companies with revenues exceeding $100 million were found to use some form of business analytics. A report by the McKinsey Global Institute predicted that by 2018, the United States alone will face a shortage of 140,000 to 190,000 people with deep analytical skills, as well as a shortfall of 1.5 million data-savvy managers with the know-how to analyze big data to make effective decisions [17].

References

[1] Silver, N. (2012). *The signal and the noise: Why so many predictions fail-but some don't*. Penguin.

[2] Chen, C. P., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences*, *275*, 314-347.

[3] Stark, J. (2015). *Product lifecycle management* (pp. 1-29). Springer International Publishing.

[4] SARCAR, M., Rao, K. M., & Narayan, K. L. (2008). *Computer aided design and manufacturing*. PHI Learning Pvt. Ltd..

[5] Groover, M., & Zimmers, E. W. J. R. (1983). *CAD/CAM: computer-aided design and manufacturing*. Pearson Education.

[6] Liu, D. T., & Xu, X. W. (2001). A review of web-based product data management systems. *Computers in industry*, *44*(3), 251-262.

[7] Frydenberg, M. (2015). Introducing Big Data Concepts in an Introductory Technology Course. *Information Systems Education Journal*, *13*(5), 12.

[8] Duggal, P. S., & Paul, S. (2013). Big Data Analysis: Challenges and Solutions. In *International Conference on Cloud, Big Data and Trust* (pp. 13-15).

[9] Information System & Management, ISM Book, 1st Edition 2010, EMC2, Wiley Publishing

[10] Savas, O., Sagduyu, Y., Deng, J., & Li, J. (2014). Tactical big data analytics: challenges, use cases, and solutions. *ACM SIGMETRICS Performance Evaluation Review*, *41*(4), 86-89.

[11] Mukherjee, R., & Mao, J. (2004). Enterprise search: Tough stuff. *Queue*, *2*(2), 36.

[12] Dumais, S., Cutrell, E., & Chen, H. (2001, March). Optimizing search by showing results in context. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 277-284). ACM.

[13] Lomotey, R. K., & Deters, R. (2013, June). Terms extraction from unstructured data silos. In *System of Systems Engineering (SoSE), 2013 8th International Conference on* (pp. 19-24). IEEE.

[14] Chierchia, G., & McConnell-Ginet, S. (2000). *Meaning and grammar: An introduction to semantics*. MIT press.

[15] Booz, & Allen & Hamilton. (1982). *New products management for the 1980s*. Booz, Allen & Hamilton.

[16] Tunkelang, D. (2009). Faceted search. *Synthesis lectures on information concepts, retrieval, and services*, *1*(1), 1-80.

[17] Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS quarterly*, *36*(4), 1165-1188.

[18] Ularu, E. G., Puican, F. C., Apostu, A., & Velicanu, M. (2012). Perspectives on Big Data and Big Data Analytics. *Database Systems Journal*, *3*(4), 3-14.

[19] Assunção, M. D., Calheiros, R. N., Bianchi, S., Netto, M. A., & Buyya, R. (2015). Big Data computing and clouds: Trends and future directions. *Journal of Parallel and Distributed Computing*, *79*, 3-15.

[20] Gudivada, V. N., Baeza-Yates, R., & Raghavan, V. V. (2015). Big Data: Promises and Problems. *Computer*, (3), 20-23.

[21] Teresko, J. (2004). Auto Report: Information Technology The PLM Revolution. *INDUSTRY WEEK-CLEVELAND OHIO-*, *253*(2), 32-38.

[22] Bergsjö, D. (2009). *Product Lifecycle Management–Architectural and Organisational Perspectives*. Chalmers University of Technology.

[23] Ulrich, K. T., & Eppinger, S. D. Product design and development. 2004.

[24] Li, Y., Wan, L., & Xiong, T. (2011). Product data model for PLM system. *The International Journal of Advanced Manufacturing Technology*, *55*(9-12), 1149-1158.

[25] Hsu, W., & Woon, I. M. (1998). Current research in the conceptual design of mechanical products. *Computer-Aided Design*, *30*(5), 377-389.

[26] Wu, D., Rosen, D. W., Wang, L., & Schaefer, D. (2015). Cloud-based design and manufacturing: A new paradigm in digital manufacturing and design innovation. *Computer-Aided Design*, *59*, 1-14.

[27] Horváth, I., & Vroom, R. W. (2015). Ubiquitous computer aided design: A broken promise or a Sleeping Beauty?. *Computer-Aided Design*, *59*, 161-175.

[28] Pohl, J., Chapman, A., & Pohl, K. (2000, August). Computer-aided design systems for the 21st century: Some design guidelines. In *5th International Conference on Design and Decision-Support Systems for Architecture and Urban Planning (August 2000)*.

[29] Raphael, B., & Smith, I. F. (2003). *Fundamentals of computer-aided engineering*. John Wiley & Sons.

[30] Makris, S., Mourtzis, D., & Chryssolouris, G. (2014). Computer-Aided Manufacturing. In *CIRP Encyclopedia of Production Engineering* (pp. 254-266). Springer Berlin Heidelberg.

[31] Balog, K., Azzopardi, L., & De Rijke, M. (2006, August). Formal models for expert finding

in enterprise corpora. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 43-50). ACM.

[32] Li, Y., Liu, Z., & Zhu, H. (2014). Enterprise search in the big data era: recent developments and open challenges. *Proceedings of the VLDB Endowment*, *7*(13), 1717-1718.

[33] Bhuiyan, N. (2011). A framework for successful new product development. *Journal of Industrial Engineering and Management*, *4*(4), 746-770.

[34] Krishnan, V., & Ulrich, K. T. (2001). Product development decisions: A review of the literature. *Management science*, *47*(1), 1-21.

[35] Cooper, R. G. (1994). Third-generation new product processes. *Journal of Product Innovation Management*, *11*(1), 3-14.

[36] Liepė, Ž., & Sakalas, A. (2015). The three-loop learning model appliance in new product development. *Engineering Economics*, *58*(3).

[37] Dwyer, L., & Mellor, R. (1991). Organizational environment, new product process activities, and project outcomes. *Journal of Product Innovation Management*, *8*(1), 39-48.

[38] Eppinger, S. D., Whitney, D. E., Smith, R. P., & Gebala, D. A. (1994). A model-based method for organizing tasks in product development. *Research in Engineering Design*, *6*(1), 1-13.

[39] Sethi, R., Smith, D. C., & Park, C. W. (2001). Cross-functional product development teams, creativity, and the innovativeness of new consumer products. *Journal of Marketing Research*, *38*(1), 73-85.

[40] Crawford, A. (1997). Ideas and objects: the arts and crafts movement in Britain. *Design Issues*, 15-26.

[41] Souder, W. E. (1987). *Managing new product innovations*. Lexington, MA: Lexington books.

[42] Cooper, R. G., & De Brentani, U. (1984). Criteria for screening new industrial products. *Industrial Marketing Management*, *13*(3), 149-156.

[43] Brentani, U. (1989). Success and failure in new industrial services. *Journal of Product Innovation Management*, *6*(4), 239-258.

[44] Cooper, R. G. (2001). *Winning at new products: Accelerating the process from idea to launch*. Basic Books.

[45] Urban, G. L., Hauser, J. R., & Urban, G. L. (1993). *Design and marketing of new products* (Vol. 2). Englewood Cliffs, NJ: Prentice Hall.

[46] Cooper, R. G., & Kleinschmidt, E. J. (1993). Stage gate systems for new product success.

Marketing Management, *1*(4), 20-29.

[47] Ulrich, K. T., & Eppinger, S. D. (2000). Product design and manufacturing.

[48] ANDREWS, W.; KOEHLER-KRUENER, H. Magic Quadrant For Enterprise Search. 19th August 2015 ID: G00269182, @2015 Gartner, Inc. and/or its Affiliates.

[49] Hu, H., Wen, Y., Chua, T. S., & Li, X. (2014). Toward scalable systems for big data analytics: A technology tutorial. *Access, IEEE*, *2*, 652-687.

[50] Iyer, N., Jayanti, S., Lou, K., Kalyanaraman, Y., & Ramani, K. (2005). Three-dimensional shape searching: state-of-the-art review and future trends. *Computer-Aided Design*, *37*(5), 509-530.

[51] Grefenstette, G. (2015). Personal Semantics. In *Language Production, Cognition, and the Lexicon* (pp. 203-219). Springer International Publishing.

[52] Aho, A. V., Sethi, R., & Ullman, J. D. (1986). *Compilers, Principles, Techniques*. Addison wesley.

[53] Galitsky, B. A. (2013). Transfer learning of syntactic structures for building taxonomies for search engines. *Engineering Applications of Artificial Intelligence*, *26*(10), 2504-2515.

[54] Arvidsson, F., & Flycht-Eriksson, A. (2008). Ontologies i. *PDF). http://www. ida. liu. se/janma/SemWeb/Slides/ontologies1. pdf. Retrieved*, *26*.

[55] Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). The PageRank citation ranking: bringing order to the Web.

[56] Page, L. (2001). *U.S. Patent No. 6,285,999*. Washington, DC: U.S. Patent and Trademark Office.

[57] Haveliwala, T. H. (2002, May). Topic-sensitive pagerank. In *Proceedings of the 11th international conference on World Wide Web* (pp. 517-526). ACM.

[58] Willett, P. (1988). Recent trends in hierarchic document clustering: a critical review. *Information Processing & Management*, *24*(5), 577-597.

[59] Hsu, F. C., Trappey, A. J., Trappey, C. V., Hou, J. L., & Liu, S. J. (2006). Technology and knowledge document cluster analysis for enterprise R&D strategic planning. *International Journal of Technology Management*, *36*(4), 336-353.

[60] Willett, P. (1988). Recent trends in hierarchic document clustering: a critical review. *Information Processing & Management*, *24*(5), 577-597.

[61] Wei, C. P., Yang, C. C., & Lin, C. M. (2008). A Latent Semantic Indexing-based approach to multilingual document clustering. *Decision Support Systems*, *45*(3), 606-620.

[62] Dean, J., & Ghemawat, S. (2008). MapReduce: simplified data processing on large clusters. *Communications of the ACM*, *51*(1), 107-113.

[63] Harris, L. R., & BROWN, G. T. L. (2010). Mixing interview and questionnaire methods: Practical problems in aligning data.

[64] COCCO, M. (2013). Development and implementation of a life cycle optimization model.

[65] Nedelcu, B. (2013). About Big Data and its Challenges and Benefits in Manufacturing. *Database Systems Journal*, *4*(3), 10-19.

[66] Agrawal, D., Bernstein, P., Bertino, E., Davidson, S., Dayal, U., Franklin, M., ... & Widom, J. (2011). Challenges and Opportunities with Big Data 2011-1.

[68] Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L., & Siegel, D. (2014). Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications. *Mechanical Systems and Signal Processing*, *42*(1), 314-334.

[69] C. Gaudino (2005). Processo produttivo orientato al digital factory: impatto economico sul bilancio aziendale.

[70] Petkova, D., & Croft, W. B. (2008). Hierarchical language models for expert finding in enterprise corpora. *International Journal on Artificial Intelligence Tools*, *17*(01), 5-18.

[71] Ehrlich, K., & Shami, N. S. (2008, April). Searching for expertise. In *Proceedings of the SIGCHI Conference on Human factors in Computing Systems* (pp. 1093-1096). ACM.

[72] Punnarut, R., & Sriharee, G. (2010, January). A researcher expertise search system using ontology-based data mining. In *Proceedings of the Seventh Asia-Pacific Conference on Conceptual Modelling-Volume 110* (pp. 71-78). Australian Computer Society, Inc.

[73] Golfarelli, M., Rizzi, S., & Cella, I. (2004, November). Beyond data warehousing: what's next in business intelligence?. In *Proceedings of the 7th ACM international workshop on Data warehousing and OLAP* (pp. 1-6). ACM.

[74] Cody, W. F., Kreulen, J. T., Krishna, V., & Spangler, W. S. (2002). The integration of business intelligence and knowledge management. *IBM systems journal*, *41*(4), 697-713.

Web References

[1i] www.fortune.com

[2i] http://www.3ds.com/products-services/exalead

[3i] http://blog.exalead.com/2008/02/20/exalead-oneenterprise-a-secure-search-solution/

[4i] http://www.ballardvale.com/free/SearchHistory.htm

[5i] http://arnoldit.com/wordpress/2014/05/06/trifles-in-enterprise-search-history/

[6i] http://www.alteanet.it/mission_e_valori

[7i] http://www.comau.com/eng/offering_competence

[8i] http://www.johnstark.com

Ackowledgements

I am extremely grateful to all that contributed to this work with their time, care, passion, experience, professionalism and, most of all, their beyond measure patience.

I wish to thank:

- Professor Sergio Terzi, my academic supervisor
- Samuele, Adriano and Stefano from Dassault Systèmes
- Massimo and Michele from ALTEA
- Daniele from COMAU
- Michael from the University of Iowa
- John from John Stark Associates
- Daniele and Matteo from Politecnico di Milano

To my Mother and my Grandmother.

致我所有的中国朋友，因为他们永远地改变了我，因为他们让我的生命充满精彩和爱。