Enterprise Architecture Analysis - A study of the IT landscape at AstraZeneca

Per Westerberg

Degree project in
ICS
Master thesis
Stockholm, Sweden 2015

XR-EE-ICS 2015:008
Preface This report is a Master Thesis that was started in January 2014. It has been done as a part-time project in collaboration with the Department of Industrial Information and Control System (ICS) at the Royal Institute of Technology and AstraZeneca. Thanks go to AstraZeneca and my colleagues there for all the support.

I would like to give a special thanks to my supervisor Robert Lagerström at Royal Institute of Technology for his time, support, patience and guidance.

And to Ruben, Leo, Stella and Carol for all love and support – Thank you.

Mariefred, June 2015

Per Westerberg
Abstract. A case study at a global pharmaceutical company has been conducted to analyse how the Hidden Structure method using the Enterprise Architecture Analyses (EAAT) tool, developed at KTH, can be used to visualize the IT architecture and to create a better understanding on which applications could increase a risk of the robustness of the architecture if changes was done to them. Also the measure of IT support metrics as incidents and changes per system has been analysed to understand if they can be used to understand the robustness of the incident. The tool was used to create a model of the enterprise architecture of the company and the analysis showed that the robustness of the IT architecture was good; the core applications that could create most damage were identified. The analysis of the IT support metrics showed that it was difficult to use the change records as indicators since the number of changes per systems were very few. The incident analysis showed that the systems with the highest number of incidents were classified as belonging core or peripheral category.

Keywords. Enterprise Architecture. EAAT. Case Study. IT Support Metrics. Design Structure Matrices
# Table of Contents

Table of Contents 4

1 Introduction 4
   1.1 AstraZeneca Operations 5
   1.2 Chapter Synopsis for the thesis 5

2 Scope, objective and delimitations 7
   2.1 Scope 7
   2.2 Objective 7
   2.3 Delimitations 7

3 Theory 8
   3.1 Enterprise Architecture 8
   3.2 Design Structure Matrices and Hidden Structure Method 9
   3.3 Architecture classification of Hidden Structure Method 10
   3.4 Enterprise Architecture Analysis Tool 11
   3.5 ANSI/ISA 95 Enterprise – Control Systems Integration 13

4 Method 14
   4.1 Method Overview 14
   4.2 Data Collection 14

5 Results 17
   5.1 EAAT model of manufacturing and quality domain architecture 17
   5.2 Enterprise Data Set results 18
   5.3 Operations Data Set results 19
   5.4 Quality Critical Data Set results 20
   5.5 IT Support metrics for the Quality Critical dataset 21

6 Analysis 22
   6.1 Analysis of Hidden Structure Method results 22
   6.2 Analysis of Incidents and Changes 24

7 Discussion 27

8 Conclusion 30

9 REFERENCES 31
1 Introduction

AstraZeneca is a global innovation-driven biopharmaceutical company specialising in the discovery, development, manufacturing and marketing of prescription medicines. The revenue for 2014 was $26 billion [1]. AstraZeneca is active in over 100 countries with a growing presence in emerging markets including China, Brazil, Mexico and Russia. The number of employees is over 57,500 people (46% in Europe, 31% in the Americas and 23% in Asia-Pacific).

A strong focus in the business model since 2012 has been to acquire small companies which have a drug in the later stages of development where AstraZeneca see that this product would be a good fit in the overall portfolio. This is also leading to more integration challenges such as different culture, different quality systems and more IT systems.

The IT area in AstraZeneca has undergone a number of major changes. Going back to 1990’s where all IT operation, development and maintenance was in-house and today almost all has been outsourced in different waves and since 6 month some parts is again being in-sourced. The operating model for IT has changed a number of times, the main reason for the changes has been to focus on core business for the company, i.e. pharmaceutical development and manufacturing, and not run a big IT division, to more recently much more focus on lowering the cost.

A lot of efforts and money has put 3 regional ERP systems in place, these are based on the same SAP platform, but they are configured differently. There are still many IT applications that are more than ten years old and efforts have been done to stream line and simplify the IT landscape. However there is still a large legacy of old systems. The reasons for the large amount of applications are

- There are regulatory requirements that the data must be stored for a long time, for example 25 years after the product has stopped being manufactured [2]. This makes it impossible to just switch off an application when some new application replaces it.
- Global vs. Site applications, many applications are at site level. So each site can have a unique application for the same business process.
- Silo thinking, not only between R&D, Operations and Marketing; but also within these business areas.

There is a need to have a good understanding of the IT landscape when planning projects, perform system upgrades and making changes or enhancements to existing system. What risks does the business face when introducing these kinds of changes?

The use of risk management is fundamental in the pharmaceutical industry; there is an expectation from regulators that the industry is using risk-based approach when making business and quality decision. In [3], the use of Pharmaceutical Quality Management system that includes a risk-based approach for decision making is clearly stated. Also in IT industry it has been recognised that the lack of proper Risk Management can cause failure. In [4], poor Risk Management is on the third place of classic mistakes on how to fail an IT project and the mitigation would be to appoint Risk Officer / Quality Assurance person on the project that plays the devil’s advocate.

The cost of not having a good understanding of the impact a business change can have can be high [5]. In Sweden some examples of IT project that failed or massively overrun are Försäkringskassan (National Insurance Agency) introduction of a new self service system [6] and the Swedish Police new record management system called Puts [7].

For AstraZeneca this was made all too clear when in February 2012 the company had to replace an old bespoke ERP system for the Sweden production sites in Södertälje and Gärtuna. The old systems
were at the end of their life-cycle was to be replaced with a modern ERP system. At launch of the implementation, some unexpected difficulties occurred. The disruption of the supply chain had adverse financial impact, estimated to be as high as 280 MUSD [8].

Enterprise Architecture is a becoming a more established approach to manage business processes together with its IT systems and to model them on order to understand and improve how they perform [9].

The Enterprise Architecture Analysis Tool (EAAT) is a tool to develop models of enterprise and systems architecture [10]. The EAAT can be used to analyse the architecture in terms of hierarchical order and cycles and to create an architecture network representation using a method called Hidden Structure [11].

This report aim to show how enterprise architecture models, using the Hidden Structure method, can be used to mimic the IT landscape and display the degree of sensitivity the IT architecture is to changes.

For AstraZeneca these results can help to understand which applications need more testing during change implementation and which applications need more or potentially less monitoring from the Global Quality function.

1.1 AstraZeneca Operations

Focused on assured, cost-effective, efficient supply of medicines to patients, AstraZeneca’s Operations organization is the connection between the Research and Development lab and the pharmacy shelf. It has 10,300 employees working in AstraZeneca Operations organization at our Supply & Manufacturing sites in 20 countries.

AstraZeneca’s Operations business is organised into:

- **Three regions** – Supply Americas; Supply Europe, Middle East and Africa; Supply Asia Pacific,
  Analyzing supplier market information and leveraging this to support effective sourcing decisions
  Utilizing relevant AZ procurement systems
- **Three global functions** – Global External Sourcing; Global Quality; and Global Supply Chain & Strategy
- **Three corporate functions** – Procurement; Corporate IT and AZ Essentials

The level of autonomy of the sites has been one of the success factors in the past and something that is seen worth to preserve. This is also something that provides challenges when trying to run global change and simplification programs such as implementing common IT systems, but the trend is that the sites have started to understand the benefits of having common systems.

1.2 Chapter Synopsis for the thesis

1 Introduction
The introduction chapter gives the background to the thesis and shows the outline of the report.

2 Scope, Objective and delimitations
In this chapter the scope of the thesis and the objective of the work are specified. The delimitations of the work are also covered.
3 Theory
Theory chapter forms the basis of the academic depth of the thesis. The reader will gain understanding of Enterprise Architecture and the Hidden Matrices method.

4 Method
Chapter 4 is presenting how the work was organised to achieve the goals of the case study for the Master Thesis as described and the validity of the data.

5 Results
In this chapter the result from the Hidden Structure calculation is presented.

6 Analyses
Chapter 6 presents the analysis that is done using the results.

7 Discussions
Chapter 7 present a discussion on the findings and views of the author.

8 Conclusions
In the conclusion the reports and possible future work is presented.
2 Scope, objective and delimitations

The pharmaceutical industry is, as previously mentioned, heavily regulated by rules. The reason for these regulations is ultimately to protect the patient. The rules can be detailed, as in the when stating that all changes for a computerized system, that is used for pharmaceutical manufacturing and can have an impact on product, must be kept under change control and tested for intended use [2][12][13].

For AstraZeneca, with so many applications, there is a need to prioritise where the efforts for testing should be and where to assume more or less stringent control of changes to applications.

2.1 Scope

The systems landscape in AstraZeneca is supported in different ways. Systems that are managed locally on a site often have their servers are located in a local data centers. The centrally supported systems are supported by the central IT function and have their servers located in a few “world-class” data centers.

The analysis of the IT Architecture for this report is limited to the centrally supported systems, e.g. the systems are sitting on servers located in data centres managed by central IT functions. The scope of the project is limited in time; it should be equivalent to 20 weeks full time.

2.2 Objective

The project objective is to create a model of AstraZeneca enterprise application architecture using Hidden Structure Analysis. With the help of this model the coupling measures will be analysed and the architecture modularity will be visualized. The sensitivity of the IT Architecture can be used to understand what level of risk a change in any components could be to the robustness of the systems.

As a secondary objective the project should be to make a correlation analysis of the impact of modularity and coupling versus key business metrics on the systems.

And for AstraZeneca, can these results can help to understand which applications need more testing during change implementation and which applications need more or potentially less monitoring from the Global Quality function.

2.3 Delimitations

Systems that are managed locally are not in scope of this work. The concept of Risk Management is also excluded from this report.
3 Theory

With the introduction of integrated IT systems there is a need to have an overview picture of how these systems fit together to be able to manage the business needs and requirements in a good way [9].

In the area of enterprise architecture there is the work done by Zachman in that can be said to be the foundation of today’s research area. In [14] a concept of different architectural views is introduced to IT, by starting with looking at how the construction of a new building is set up with by different architectural representations such as

- Bubble charts – basic concepts of the building and architect/owner mutual understanding
- Architect's drawing – building as seen by the owner and architect/owner agreement on building
- Architect's plans – building as seen by the designer and basis for negotiation with contractor
- Contractor’s plans – building as seen by the builder and “how to build it” and direct the construction activates
- Shop plans – subcontractors design of a part/section and “how to build it”
- Building – physical building

What is shown is that there is a set of representation of the same object but depending on which role you have you need to look with your specific architectural view. So for information’s systems these would be the different view

- Scope / Objective – a ballpark perspective
- Model of business – owners perspective
- Model of information system – designer perspective
- Technology model – builders perspective
- Detailed description – builders perspective
- Machine language description – program code

So the introduction of using different views in these terms created the start for Enterprise Architecture.

3.1 Enterprise Architecture

The art of Enterprise Architecture is to model not only the system but to also take in business and infrastructure layers to make a more complete model that can show how the business is impacted by the IT.

Since Enterprise Architecture is fairly new discipline, there are many definitions of Enterprise Architecture. As by [15] reads as follows:

**Enterprise Architecture**: a coherent whole of principles, methods, and modules that are used in the design and realisation of an enterprise organisational structure, business processes, information systems and infrastructure, where

**Enterprise**: any collection of organisations that has a common set of goals and/or a single bottom line.

Another definition that focuses on a model is [10] and that is used in this thesis:

**Enterprise Architecture** is an approach for managing the organisation’s information system portfolio and its relation and support to the business. At the base of the approach lies an architectural model, incorporating concepts such as software components, connectors, functions, information, business processes, organisational units and actors.
As can be seen in Table 1, there are different kinds of level for information systems architecture, for the level of Enterprise Architecture the impact is of more strategic outcome and you target all stakeholders, both business and IT.

Table 1 – The relationship between Enterprise Architecture and other architecture disciplines

<table>
<thead>
<tr>
<th>Level</th>
<th>Scope</th>
<th>Detail</th>
<th>Impact</th>
<th>Audience</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enterprise Architecture</td>
<td>Organisation</td>
<td>Low</td>
<td>Strategic Outcome</td>
<td>All stakeholders</td>
</tr>
<tr>
<td>Segment Architecture</td>
<td>Line of Business</td>
<td>Medium</td>
<td>Business Outcome</td>
<td>Business Owners</td>
</tr>
<tr>
<td>Solution Architecture</td>
<td>Function / Process</td>
<td>High</td>
<td>Operational Outcome</td>
<td>User and Developers</td>
</tr>
</tbody>
</table>

3.2 Design Structure Matrices and Hidden Structure Method

To effort to analyse a enterprise IT landscape can be huge, but by using a Hidden Structure method, developed by [11], the analysis can be done easier.

Hidden Structure method has been used as shown in [11] to visualise couplings between components in a complex architecture and analyse the hierarchical order and the cyclic groups.

The method starts by identifying the direct dependencies between elements in the network. These dependencies can be represented as links between nodes (elements) in a graph diagram as in Figure 1.

![Graph and its dependency matrix](image)

The next step is to consider not only direct dependencies but also indirect, this creates a Design Structure Matrix. This is done by doing a matrix multiplication and by saying that an element is affecting itself. This mean that there will be a diagonal of “1” that illustrates that an element is dependent on itself, see Figure 2a. The result is also binary, there is a dependency or there is not, where a “1” show a dependency and a “0” shows no dependency as seen in Figure 2b.
By analysing the matrix using the Hidden Structure method as described in [16], a number of measures can be found.

Visibility Fan-In, VFI, is the number of elements of direct and indirect dependency that depend on a specific element. This number is found by summing the entries in the column of the element.

Visibility Fan-Out, VFO, is the number of elements that a specific element is direct and indirect dependent on. This number is found by summing the entries in the row of the element.

Another metric is the Propagation Cost which can be defined as the density of the matrix. This value show how much of the architecture could be affected when a change is done on a random element.

\[
\text{Propagation Cost} = \frac{\sum_{i=1}^{N} VFI_i}{N^2} = \frac{\sum_{i=1}^{N} VFO_i}{N^2}
\]

3.3 Architecture classification of Hidden Structure Method

The elements in DSM can be classified in to four distinct types depending on as per follows

- Core - elements belonging to one large cyclic group
- Control - elements depending on other elements but not used themselves used by many
- Shared - elements that are used by other elements but do not depend on many other
- Periphery - elements that are not used by or depend on a large group of other elements

The Hidden Structure method follows the flow in Figure 3 to determine what kind of architecture classification comes from the analysis of the matrices.
Core-periphery architecture is defined when the largest cyclic group is larger than 5% or the overall architecture and if the largest cyclic group is 1.5x bigger than the second largest group. If there are several cyclic groups that are of similar size the architecture that called Multi-core, the strict definition is if the largest cyclic group is still more than 5% of the architecture but the largest is not 1.5x larger than the second largest. If the architecture contains only small cyclic groups than it is called Hierarchal, that is when the largest cyclic largest group is smaller than 5% of the architecture.

3.4 Enterprise Architecture Analysis Tool

The Enterprise Architecture Analysis Tool is software to develop models of enterprise and systems architecture and it is developed at KTH. The first version of the tool was available 2008. With the models many different kinds of analysis can be done. The EAAT is inspired by the graphical decision-theoretic methods such as Bayesian networks and influence diagrams. It has a graphical interface where a Multi-Attribute Prediction (MAP) class model is built following a three layer Meta model based on Archimate [15].

In the MAP Metamodel, as described in [10], the three layers separate the use of the elements in each layer, see Figure 4. In the Business Layer, which acts as the interface between the business and its customers, one or many Business Services can be consumed by external customers, services are then realised in the business by one or many Business Processes.

The Application Layer, which sits between in the Business Layer and the Infrastructure Layer, is where the software applications is performing it functions. An Application Service is realized by one or more Application Functions. An Application Component is a part of a software system, and performs the work of an Application Function; it is modular, deployable and replaceable. To model a collective behaviour in the Application Component there are Application Collaboration elements.

In the Infrastructure layer, the focus for the Metamodel has moved from business process and applications to the supporting infrastructure. Here there are Infrastructure Services that supports Application Functions in the Application Layer. Infrastructure Services are realised by one or many Infrastructure Functions. The Infrastructure Functions are performed in a Node, where a node could be a physical server.
The analysis can be done by using the attributes in the multi-attribute prediction class diagram, MAP. The MAP metamodel has different elements that can be used to model key quality attributes for IT systems. From [10] the key attributes in the model are defined as:

- **Application Modifiability**, determines how easy it is to make modifications to the IT systems.
- **Data Accuracy**, determines the likelihood of having complete, consistent, relevant and accurate data.
- **Application Usage**, measure Task-Technology Fit for a given application.
- **Service Availability** determines the probability of the model system to be available or not.
- **Interoperability**, determines the ability of two or more systems or components to exchange information and to use the exchange information.
3.5 ANSI/ISA 95 Enterprise – Control Systems Integration

ANSI / ISA 95 Enterprise - Control System Integration is a standard developed by manufacturing companies on how to integrate IT system on different levels in the company. By setting the systems at the hierarchical level as seen in Figure 5, the integration is then done between the levels. The standard is split in three parts:

- Part 1 - Models and Terminology [17]
- Part 2 - Object Model Attributes [18]
- Part 3 - Activity Models of Manufacturing Operations Management [19]

The model provides a good structure on how to view the IT landscape for a manufacturing company as seen in Figure 5.

![ISA 95 model with its 5 levels](image)

This model is used to understand what type of support a system needs. If an incident happens on a level 3 system, it needs to be resolved faster than if it happens on a level 4 system.

So the level that the systems belong to, can give information about risk to the product and patient if something would fail. In a similar way as the element type from the Hidden Structure method can give show if there is a risk that other systems would be affected by a failure to a specific system.
4 Method

4.1 Method Overview

The method to run the case study was done by following these steps

1. Set the goals for the case study
   The same goals as presented in section 2.2
2. Study the literature
3. Understand where and what kind of data can be collected
4. Perform data collection
5. Set up model in EAAT
6. Run the model with the data sets
7. Analyse the results
8. Report

To use the EAAT tool for DSM analysis, a work was done in [20] to enhance the functionality EAAT to encompass the calculation and the reporting of DSM. The metamodel for the DSM calculation is just one element, Application. There are no other elements that are of interest when we run this calculation. The input to the EAAT for DSM analysis is an Excel spread sheet where all applications are set up in matrices on rows and columns as in Table 3. If an application is dependent of another application that is then marked with Provide if the dependency is above the diagonal line, if the dependency is below the diagonal line it is marked with Depend. If there is no dependency it is marked with a 0.

Table 3 – An example of an input matrix for 6 applications to the DSM analysis

<table>
<thead>
<tr>
<th>Object Model Name</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a</td>
</tr>
<tr>
<td>Application</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>1</td>
</tr>
<tr>
<td>b</td>
<td>0</td>
</tr>
<tr>
<td>c</td>
<td>0</td>
</tr>
<tr>
<td>d</td>
<td>0</td>
</tr>
<tr>
<td>e</td>
<td>0</td>
</tr>
<tr>
<td>f</td>
<td>0</td>
</tr>
</tbody>
</table>

The output from the EEAT tool is a PDF-file where the analysis is presented in both graphical form and in table form. For each element in the matrices the DFI, DFO, VFI and VFO is presented and the propagation cost together with an analysis of what type the architecture is.

4.2 Data Collection

The main data of the 1112 systems that has been used for this report was collected from a Configuration Management Database (CMDB) system called Troux [21]. In this CMDB all systems / applications that are centrally supported are recorded with an application ID and application/system name. The centrally supported system consist of applications where the applications servers are located in data centers managed by a central IT function and if some problems occur with the
applications the users can call service desk and they would manage the incident. These applications are of multi-user type, however the user base can vary from very few users to 65 000 users for the system with the largest user base.

The primary details that are stored in the CMDB database are as follows:

- Unique Reference Number
- System/ Application Name
- GxP regulated Yes/No, is the system impacted by pharmaceutical manufacturing laws
- SOx regulated Yes/No, is the system impacted by Sarbanes Oxley laws
- Data Privacy Yes/No, is the system impacted by Data Privacy laws
- System Owner (and local representatives).
- Dependencies (inbound application / outbound application)

The data has been split up in three different groups:

1. Enterprise data set. Here are all applications from the CMDB.
3. Quality Critical data set. The applications that are supported from a quality perspective by the company’s global quality function.

The data accuracy is overall good, but on the Dependencies there are some doubts. Some examples of missing dependencies have been discovered. These examples are mainly found in the non-Operations data-set, which is why the sub-set of Operations is seen as more reliable than Enterprise. The Quality Critical data set has been manually vetted and is expected to be more than 95% correct.

Interviews to verify the data accuracy has been conducted. The respondents had the roles of System Owners, Application Service Managers or Business Relationship Manager.

Excluded from the CDMS are individual PCs and any associated systems/applications that could reside on those local PCs. It also rules out locally managed systems, typically local laboratory systems, process control systems and HVAC systems. An estimate, based on the typical number of these local systems per site, would be that 10 000 – 15 000 systems would fall in this category. These systems are typically stand-alone; sometimes they would be configured in a local network.

The Incident and Change data for the Quality Critical sub-set was collected from 1st September 2014 to 30th January 2015, the reason for this time period was that from the 1 September the Service Desk change tool to Service Now, so before the introduction of this tool there was very hard to get incident data of the applications. Service Now is a Configuration Management Data Base from [22], there are different modules that support The company is planning to use Service Now to a large extent, even to register process control systems and local laboratory system. The aim is create one common system list for the whole company.

The Incident definition follows IT Infrastructure Library, ITIL [23]

Incident Management is the process for dealing with all incidents; this can include failures, questions or queries reported by the user (usually via a telephone call to the Service Desk), by technical staff, or automatically detected and reported by event monitoring tools.

The incident is logged in a Priority of 1-4, where 1 is the highest priority and the categories of the incident is logged in one of the following groups:

- Access/Permission
- Enquiry / Help
The changes has a number attributes as can be seen in Table 4.

Table 4 – The change data is categorised in the these statuses

<table>
<thead>
<tr>
<th>Priority</th>
<th>Change Type</th>
<th>Change State</th>
<th>Change Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 – Highest</td>
<td>Emergency</td>
<td>1. Create and Plan</td>
<td>Application Software</td>
</tr>
<tr>
<td>2</td>
<td>Normal</td>
<td>2. Review and Approve</td>
<td>Business Service</td>
</tr>
<tr>
<td>3</td>
<td>Standard</td>
<td>3. Build and Test</td>
<td>Documentation</td>
</tr>
<tr>
<td>4 - Lowest</td>
<td>Application</td>
<td>4. Deploy and Close</td>
<td>Hardware</td>
</tr>
<tr>
<td></td>
<td>Server Reboot</td>
<td></td>
<td>Network</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td></td>
<td>Other</td>
</tr>
<tr>
<td></td>
<td>Software</td>
<td></td>
<td>Server Reboot</td>
</tr>
<tr>
<td></td>
<td>System Software</td>
<td></td>
<td>Software</td>
</tr>
<tr>
<td></td>
<td>Telecom</td>
<td></td>
<td>System Software</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Telecom</td>
</tr>
</tbody>
</table>
5 Results

5.1 EAAT model of manufacturing and quality domain architecture

By using the EAAT tool, it was possible to build up a model of the IT landscape for the company. The input matrices, which had 1112 x 1112 elements, contained mainly of solitary applications. When using the DSM tool in EAAT there is no graphical schema, it's just a list of application instances and it gives the result in a PDF file. In Figur 6, there is however an example of how the global QA processes can be depicted using the tool.

![Diagram showing the EAAT model of manufacturing and quality domain architecture.](image)

Figure 6. Example of how the graphical interface can mimic the global QA processes.

The overall dataset from contained 1112 applications, which we call Enterprise data set. Since the Operations applications was of special interest the data was broken down in smaller data set. The Operations data set contained systems that were needed for manufacturing, excluding the systems that were owned by R&D and Commercial unless they were used in manufacturing. An even smaller sub-set was the named Quality Critical, these were systems that are critical to ensure the quality of the product and the supporting Quality Management System.
5.2 Enterprise Data Set results

The first data set was for the whole Enterprise and it consists of all applications for the company, which were 1112 applications. When run through the DSM tool in EAAT, as described in section 3.6, the result was that the largest cyclic group was 49 elements and the second largest cyclic group was 16.

When doing an element breakdown as described in section 3.4, the Core consisted of 49 elements, those are elements belonging to one large cyclic group. The number of Shared elements was 97; these are the elements that depending on other elements but not used themselves used by many other elements. There number of Peripheral elements in the Enterprise data set is 953 items, those are the elements that are not used by or depend on a large group of other elements. The number of Control elements is 13; those are elements depending on other elements but not used themselves used by many other. The propagation cost for the Enterprise data set was calculated as

\[
\text{Propagation Cost} = \frac{\sum_{i=1}^{49} V_{EI}}{1112^2} = \frac{\sum_{i=1}^{49} V_{EI}}{1112^2} = 10.4\%.
\]

The largest cluster in the Enterprise data set contained applications as for example

- Enterprise Resource Planning systems, there are 3 regional and 3 local ERP systems. Two scheduling and planning system where one is global and another local.
- Purchasing systems, there are four separate purchasing systems. Even though the purchasing function is used in the ERP systems but mainly for raw material, so these four applications are global, regional and local systems that are used for all other articles.
- Quality Control systems, these 3 global systems are for chromatography data repository, sample management and laboratory work instruction. And another sample management system for R&D and a local system still in use.
- Expense reporting systems, there are two systems for expense reporting.
- Medicaid claim system, used for the US market.
- Complaint Management System, the system for customer product complaints management reaches all market companies and all sites.
- Contracts Management system
- Middleware systems, there are two messaging systems in the cluster.
- Clinical trials systems, there are two planning systems for R&D
- Dangerous goods system
- Customer support systems, two customer support system are in the cluster.
Following the method for Hidden Structure the type of the Enterprise architecture is Hierarchal, the architecture only holds small cyclic groups.

5.3 Operations Data Set results

To investigate the data further, due to the unexpected result in having a Hierarchal architecture type, another dataset was created as a sub-set of the Enterprise data-set. The Operations data set consists of 806 applications that have their ownership in the Operations business function. The result from the DSM analysis was that the largest cyclic group was 18 elements and the second largest cyclic group was 6.

When doing an element breakdown the Core consisted of 18 elements, those are elements belonging to one large cyclic group. The number of Shared elements was 19; these are the elements that depending on other elements but not used themselves used by many other elements. There number of Peripheral elements in the Enterprise data set is 719 items, those are the elements that are not used by or depend on a large group of other elements. The number of Control elements is 50; those are elements depending on other elements but not used themselves used by many other. The propagation cost for the Enterprise data set was calculated as

\[
\text{Propagation Cost} = \frac{\sum_{i=1}^{806} VF_i}{806^2} = \frac{\sum_{i=1}^{806} VF_O}{806^2} = 0.6\% 
\]
5.4 Quality Critical Data Set results

The data set for Enterprise consists of all applications for the company, which were 136 applications. When run through the DSM tool in EAAT the result was that the largest cyclic group was 7 elements and the second largest cyclic group was 2.

When doing an element breakdown as described in section 3.4, the Core consisted of 7 elements, those are elements belonging to one large cyclic group. The number of Shared elements was 34; these are the elements that depending on other elements but not used themselves used by many other elements. There number of Peripheral elements in the Enterprise data set is 89 items, those are the elements that are not used by or depend on a large group of other elements. The number of Control elements is 6; those are elements depending on other elements but not used themselves used by many other. The propagation cost for the Enterprise data set was calculated as

\[
\text{Propagation Cost} = \frac{\sum_{i=1}^{136} V_{FI_i}}{136^2} = \frac{\sum_{i=1}^{136} V_{FO_i}}{136^2} = 3.8\%
\]

The largest cluster in the Quality critical data set contains as mentioned before 7 applications. These applications are

- Enterprise Resource Planning systems, there are 3 regional and 2 local ERP systems. And one global planning system.
- Complaint management system
In Figure 9 the rearranged matrices can be seen where the seen

![Figure 9. Rearranged view for Quality data set.](image)

Following the method for DSM the type of the Quality Critical sub set is Core-periphery.

5.5 IT Support metrics for the Quality Critical dataset

The number of changes on the systems are few, 33 changes in the group of the Quality Critical systems, see Table 5. The data was taken from 1 September 2014 – 30 January 2015. The definition of a change here is changes related to the system, not data changes.

Table 5 – The number of changes on the Quality Critical data set.

<table>
<thead>
<tr>
<th>System</th>
<th>Number of changes Sept 14 – Jan 15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vendor Quality system</td>
<td>14</td>
</tr>
<tr>
<td>Global Document Management system</td>
<td>7</td>
</tr>
<tr>
<td>Planning system</td>
<td>3</td>
</tr>
<tr>
<td>Global Sample Management system</td>
<td>2</td>
</tr>
<tr>
<td>ERP EMEA, ASIA, US</td>
<td>2 changes per system</td>
</tr>
<tr>
<td>The rest of 125 systems with no changes</td>
<td>0</td>
</tr>
</tbody>
</table>

When a user calls in to Service Desk it is logged as an incident. The total number of incident reported was 10721 for all systems during the period; the breakdown of incidents per category can be seen in Table 6.

Table 6 – The number of incidents per category of the quality critical system

<table>
<thead>
<tr>
<th>Access / Permissions</th>
<th>Enquiry/Help</th>
<th>Outage/Failure</th>
<th>Performance</th>
<th>Error Message</th>
<th>Backup Failure</th>
<th>Intermittent</th>
</tr>
</thead>
<tbody>
<tr>
<td>3306</td>
<td>3254</td>
<td>1947</td>
<td>1231</td>
<td>922</td>
<td>56</td>
<td>5</td>
</tr>
</tbody>
</table>
6 Analysis

6.1 Analysis of Hidden Structure Method results

The results from the study were not as expected when it comes to the Type for architecture as in Table 7. Through earlier studies [25][11], the Type has been Core-periphery for those examined companies, but this study suggests a Hierarchical architecture type for AstraZeneca.

Table 7 – The results from the DSM run for the three data sets.

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Enterprise</th>
<th>Operations</th>
<th>Quality Critical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data set size [no of elements]</td>
<td>1112</td>
<td>806</td>
<td>136</td>
</tr>
<tr>
<td>Core [no of elements]</td>
<td>49</td>
<td>18</td>
<td>7</td>
</tr>
<tr>
<td>Shared [no of elements]</td>
<td>97</td>
<td>19</td>
<td>34</td>
</tr>
<tr>
<td>Peripheral [no of elements]</td>
<td>953</td>
<td>719</td>
<td>89</td>
</tr>
<tr>
<td>Control [no of elements]</td>
<td>13</td>
<td>50</td>
<td>6</td>
</tr>
<tr>
<td>Largest cluster [no of elements]</td>
<td>49</td>
<td>18</td>
<td>7</td>
</tr>
<tr>
<td>Second Largest cluster [no of elements]</td>
<td>16</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Largest cluster &gt;5% of system</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Largest cluster &gt;1.5x of next cluster</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Propagation cost</td>
<td>10.4%</td>
<td>0.6%</td>
<td>3.8%</td>
</tr>
<tr>
<td>Type</td>
<td>Hierarchical</td>
<td>Hierarchical</td>
<td>Core-periphery</td>
</tr>
</tbody>
</table>

From the Figure 10, the breakdown per element type for the Enterprise data-set can be seen. The core is 4.41 %, which is not that far from coming over the threshold of 5% of becoming a Core-periphery. Given that the data quality can be poor on the dependency attributes, since not all systems owners have updated the records, the true Type can possibly be Core-periphery, although a weak one.

The high degree of Peripheral applications can be explained by all applications that have no dependency to other systems, VFO = 1 and VFI = 1.
When analysing the Operations data set, the degree of Peripheral is even higher than for the Enterprise data set. This is not surprising, since Operations has the highest degree of stand-alone laboratory computer systems. When removing the R&D and Commercial systems from the data set also the core is reduced since Purchasing, Expense Reporting and Clinical Trial systems are out.

An analysis was also done to understand what element type the Quality Critical had, see Figure 12. The data accuracy is excellent on this data set since it has been manually vetted. One of the criteria to be in the Quality Critical data set is that the applications are important for the supply of products, a shortage of products on a market can be critical for the patients, this is why supply chain systems is in the data set. Some of the local ERP systems are not managed from a quality perspective by the global team; hence they are excluded from the list.
Figure 12. Breakdown of element type for the Quality Critical data set.

The core in the Quality Critical data set is 5.15%, as seen in Figure 12. This then pushes this data-set to a Core-peripheral type.

6.2 Analysis of Incidents and Changes

The incident data is coming from the calls to the global Service Desk. The big proportion of Access/Permission and Enquiry/Help calls is made more visible by a pie chart as in Figure 13.

The Access/Permission was taken out from following analysis since this is not a parameter that would be seen as connected to the robustness of the system, since these are mostly questions about forgotten password.

However the Enquiry/Help category can be an indicator of the systems robustness. A well functioning, smoothly running system with trained users should have fewer Enquiry / Help calls.
When the category of Access/Permissions was removed from the data set, the number of incident for the top ten applications can be seen in Table 8. The majority of the applications in the list have a lot of users; the ERP’s and Document Management System are expected to be in there for those reasons.

Table 8 – The top ten applications in the Quality Critical data set with most incidents over 5 months.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Application</th>
<th>Type</th>
<th>No of incidents</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ERP ASIA</td>
<td>Core</td>
<td>2058</td>
</tr>
<tr>
<td>2</td>
<td>Global Document Management system</td>
<td></td>
<td>1546</td>
</tr>
<tr>
<td>3</td>
<td>ERP EMEA</td>
<td>Core</td>
<td>1395</td>
</tr>
<tr>
<td>4</td>
<td>ERP UK</td>
<td>Core</td>
<td>353</td>
</tr>
<tr>
<td>5</td>
<td>Vendor Quality System</td>
<td>Peripheral</td>
<td>139</td>
</tr>
<tr>
<td>6</td>
<td>ERP Germany</td>
<td>Peripheral</td>
<td>137</td>
</tr>
<tr>
<td>7</td>
<td>Global Planning and Scheduling</td>
<td>Core</td>
<td>90</td>
</tr>
<tr>
<td>8</td>
<td>ERP EMEA - business warehouse</td>
<td>Peripheral</td>
<td>76</td>
</tr>
<tr>
<td>9</td>
<td>ERP – Supply Chain</td>
<td>Peripheral</td>
<td>67</td>
</tr>
<tr>
<td>10</td>
<td>Global Labelling system</td>
<td>Peripheral</td>
<td>52</td>
</tr>
</tbody>
</table>

From the data in Table 8, we can see that these applications are Core or Peripheral type. Since many of them are ERP, it is expected that they would be a part of Core. The Global Document Management system had a phased go-live, starting in June 2014 and was finished in November 2014. The analysis of element type was done in August 2014 and the Document Management system was not included. If the analysis would be done again the type could be Core for the Document Management systems since has interfaces to the ERP systems for US, EMEA, UK as well as the Global Change Management system.
Interestingly, the ERP ASIA is in top. This application is older, smaller and has fewer users than ERP EMEA. The finding would suggest that it was something wrong with the application or that the users are not fully trained yet. It is recommended that this is investigated further.

The numbers of changes in the systems from the Quality Critical data set are shown in Table 9. Overall the changes are few; there are some possible explanations to this. One being that the system on the data set is robust; another is that the company is trying to get into a quarterly release pattern for the systems. The last being that the sample period was from 1st September to 30 January, the budget for some systems can have been spent the first half year.

Table 9 – The top eight applications in the Quality Critical data set with most change over 5 months.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Application</th>
<th>Type</th>
<th>No of changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Vendor Quality System</td>
<td>Peripheral</td>
<td>14</td>
</tr>
<tr>
<td>2</td>
<td>Global Document Management system</td>
<td>Core</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>Global Planning and Scheduling</td>
<td>Core</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>Global Laboratory ebook</td>
<td>Core</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>ERP EMEA</td>
<td>Core</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>ERP ASIAPAC</td>
<td>Core</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>ERP US</td>
<td>Core</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>Complaints Management</td>
<td>Peripheral</td>
<td>1</td>
</tr>
</tbody>
</table>

Interestingly, there are 125 systems out of 136 in the data set that have no changes at all. The Vendor Quality system, that has the highest number of changes, went live the 1st August, with quite a few issues during go-live. Hence a number of emergency changes had to be implemented. Likewise for the Global Document Management system, as mentioned in the incidents section above, this system went live.
7 Discussion

The IT architecture in the company has been under massive change over the last 15 years. The focus has been to decommission old IT system and replace them with standard systems and outsourcing of IT. Many of the old Astra systems were bespoke and developed during the 1970-80’s and with the merger with Zeneca in 1999 many duplications of systems came in. Zeneca had a number of local SAP systems that came in to the overall architecture.

So the primary focus after immediately after the merger for IT was to make the financial teams, using the software Hyperion to collect underlying data, set-up it for performance management, whilst connecting the infrastructure between the companies.

For the manufacturing sites in Sweden, the bespoke Oracle-based ERP systems, were coming out of support around 2005. The systems had their own messaging bus that sent messages between the different systems, see Figure 14. Each of these systems had their own system registered name and had allocated system support team in the business.

Figure 14. The bespoke systems for the Sweden manufacturing sites prior to 2012
In February 2012 the systems in Figure 14 where replaced with a common European SAP system, this system was built on an existing SAP instance so no new system registry was needed. As shown in Figure 15, the reduction in number of systems is big and the complexity reduced. Similar projects, perhaps not as big as the Sweden example, have been done at many sites. By building global or regional system solutions the application rationalisation has been huge. And there are not so many systems that have dependencies with each other.

![Diagram of systems](image)

**Figure 15. Only 4 systems to replace the bespoke systems for AZ Sweden manufacturing**

Of course there is a lot of complexity hidden just by having only one system registry for huge systems as European ERP, but still it looks good when you put it up as in Figure 15.

There are a high number of stand-alone systems in the architecture. The reasons for this is manifold

- It is hard to decommission a regulated system, since data sometimes need to be saved. This can be for up to the life time of the product + 10 years. As an extreme, take the well-known drug Xylocaïne, developed and manufactured by Astra since 1949. The regulation has not been in for more than 20 years, but still there is a lot of data to store.
- The dedicated laboratory systems that have very few users but still needed on each laboratory.
- Many replacement system projects have not included the decommissioning of the old system, so they are just standing there blowing air.

The analysis has showed that AstraZenecas IT architecture is not too sensitive to changes, with a propagation cost of 10% its suggest that only 10 % of the architecture would be effected by a change to a randomly selected application. For the Enterprise data set this is lower than other studies, where a Telecom company had 25%[16] and a Biopharma company had 23%[24].
The present study of the key IT support metrics for the Quality Critical data set shows that the element type for the top 10 incidents application and the top 8 change applications belong mainly to Core, but also to Peripheral group. However caution must be applied, as these findings might be explained by other causes, like the number of users for the incidents. Never-the-less, it shows that the applications belonging to Core are to be managed carefully.

So when analysing the Quality Critical data set, have this study showed any systems that should be added or taken off the list? It is clear that the Enterprise largest cluster list has highlighted a number of applications that need to be investigated further, mainly to see if they should be added to the list. The criteria for being on the list are a bit floating, since organisational and other parameters play in.

One application though, that has already been added to the Quality Critical list is called Dangerous Goods. This application was not in the Quality Critical list because it was not marked as GxP Regulated = Yes. This system is used to monitor the transportation of dangerous samples, dispatch of miscellaneous materials (including chemical, biological, radioactive and novel pharmaceutical compounds) according to the rules, recommendations and regulations for national and international transportation, supply and use of these materials. As such, it would make sense to manage this application according to the internal procedures with higher standards for testing and monitoring.
8 Conclusion

This study set out to determine the usability of the Hidden Structure Method using EAAT to understand which applications need more monitoring and testing from a quality perspective. The results have showed that this approach can quickly give a view of the critical applications in a complex and large Enterprise Architecture. The identified Core elements from the analysis are important to monitor from a change and performance perspective.

The study of the key IT support metrics for the Quality Critical data set shows that the element type for the top 10 incidents application and the top 8 change applications belong mainly Core, but also Peripheral group. Interestingly, the ERP ASIA is in top for Incident calls. This application is older, smaller and has fewer users then ERP EMEA. The finding would suggest that it was something wrong with the application or/and that the users are not fully trained yet. It is recommended that this is investigated further. However caution must be applied, as these findings might be explained by other causes.

Further studies for AstraZeneca on the Core applications data set is recommended, these applications are shown important to the classification of the architecture and they are in the top categories for Incidents and Changes. Can just the fact of being a Core application justify them to be in the Quality Critical list?
REFERENCES


