Defence Research and Development Canada
Recherche et développement pour la défense Canada

CoROSP Reference Architecture

Front End Analysis

Normand Pageau
Micheline Bélanger
DRDC Valcartier

Defence R&D Canada – Valcartier

Technical Memorandum
DRDC Valcartier TM 2013-307
January 2013

Canada

CoROSP Reference Architecture

Front End Analysis

Normand Pageau
Micheline Bélanger
DRDC Valcartier

Defence R&D Canada – Valcartier

Technical Memorandum
DRDC Valcartier TM 2013-307
January 2013

Abstract

This report summarizes the findings of three preliminary analysis activities conducted as part of the technical architecture orientation of the project Concepts for Recognized Operational Support Picture (CoROSP). It also identifies the key outcomes and puts forward some initial thoughts on how they could influence the system design. The aim of the CoROSP project is to demonstrate a new concept of information and decision-aid tools integration to assist the commander and the staff of Canadian Forces Operation Support team. Since it is envisioned that this new capacity will be integrated into the DND enterprise information system, this work will allow us to better understand the context, the conditions and how to leverage from the target environment.

ADM (IM) Information Sharing Centre of Excellence (ISCoE) is responsible to define and implement the DND Enterprise Reference Architecture as well as a list of standards and specifications to be observed in DND. ISCoE is also in charge to provide common services to DND and to identify, when appropriate, the tools that will sustain those core services. It is in that perspective that the first activity was to analyse the overall DND enterprise architecture while the second activity concentrates on Informatica PowerCenter, an important core tool for information integration. Finally, the third activity is to evaluate Endeca Latitude, a key component of an information system that provides analytical and business intelligence capabilities.

Résumé

Ce rapport résume les conclusions de trois activités d'analyse préliminaires menées dans le cadre de l'orientation de l'architecture technique du projet Concept pour l'Image Reconnue de Soutien Opérationnel (CIRSO). Il identifie également les principaux résultats et propose quelques réflexions initiales sur la manière dont ceux-ci pourraient influencer la conception du système. L'objectif du projet CIRSO est de démontrer un nouveau concept d'intégration d'outils d'information et aide à la décision pour aider le commandant et le personnel de l'équipe de Soutien Opérationnel des Forces Canadiennes. Considérant que cette nouvelle capacité est prévue pour être intégrée dans le système d'information d'entreprise du MDN, ces travaux permettront de mieux comprendre le contexte, les modalités et la manière de tirer parti de l'environnement cible.

SMA (GI) Centre d'Excellence de Partage d'Informations (CEPI) est chargé de définir et de mettre en œuvre l'Architecture d'Entreprise de Référence du MDN ainsi que la liste des normes et spécifications devant être observées au sein du MDN. CEPI est également en charge de fournir des services communs au MDN et d'identifier, au besoin, les outils qui supporteront ces services de base. C'est dans cette perspective que la première activité a consisté à analyser de façon globale l'architecture d'entreprise du MDN tandis que la seconde activité se concentre sur Informatica PowerCenter, un outil de base important pour l'intégration de l'information. Enfin, la troisième activité consiste à évaluer Endeca Latitude, un composant clé d'un système d'information qui fournit des capacités d'analyse et d'intelligence d'affaire.

This page intentionally left blank.

Executive summary

CoROSP Reference Architecture: Front End Analysis

Normand Pageau; Micheline Bélanger; DRDC Valcartier TM 2013-307; Defence R&D Canada – Valcartier; January 2013.

Introduction or background:

The purpose of this report is to summarize the findings of three preliminary analysis activities conducted as part of the technical architecture orientation of the project Concepts for Recognized Operational Support Picture (CoROSP). It also identifies the key outcomes and puts forward some initial thoughts on how these outcomes could influence the system design. The aim of the CoROSP project is to demonstrate a new concept of information and decision-aid tools integration to assist the commander and the staff of Canadian Forces Operation Support team. Since it is envisioned that this new capacity will be integrated into the DND enterprise information system, this work will allow us to better understand the context, the conditions and how to leverage from the target environment.

ADM (IM) Information Sharing Centre of Excellence (ISCoE) is responsible for defining and implementing the DND Enterprise Reference Architecture as well as a list of standards and specifications to be observed in DND. ISCoE is also in charge of providing common services to DND and identifying, when appropriate, the tools that will sustain those core services. It is from that perspective that the first activity was to analyse the overall DND enterprise architecture while the second activity concentrated on Informatica PowerCenter, an important core tool for information integration. Finally, the third activity was to evaluate Endeca Latitude, a key component of an information system that provides analytical and business intelligence capabilities.

Results:

Widely based on open standards and in line with emerging architecture concepts, mainly a Service Oriented approach, the ISCoE reference architecture is flexible enough to be adapted to most information systems of the DND enterprise, including CoROSP project system high-level requirements.

Informatica is a leader in this field and provides a wide range of product supporting data integration. PowerCenter tool suite demonstrated a very strong ability to provide Extract, Transform and Load (ETL) capability. The drawback is that users need some training to use the tool adequately, but it is worth it. The tool provides probably more capabilities than needed by the project but considering its availability and that it is part of the enterprise solution, it should be considered as a serious candidate for the CoROSP technology stack.

Endeca Latitude definitely demonstrates useful capabilities for manipulating data and metadata in a way that helps information discovery. It has highly capable graphic components to support visualization and a very flexible guided navigation. The combination of the Latitude Query Language (LQL) with the drill down and filtering capabilities gives the user very agile means to analyze the data. As part of the Endeca Latitude suite, the CloverETL tool offers good ETL

capabilities but is not as powerful as Informatica PowerCenter, which can handle more complex transformations and provides connectors to a wider range of data source types. Informatica PowerCenter is already part of the enterprise solution.

Future plans:

Being aligned with DND enterprise architecture should ease ROSP concept transition in the operational world. In that sense, future work should evaluate other key components of the DND technology stack based on ADM (IM) and ISCoE direction. One of the technologies to be reviewed is Autonomy IDOL that has been selected as the DND enterprise search engine and that can potentially play an important role in the ROSP capability.

Sommaire

CoROSP Reference Architecture: Front End Analysis

Normand Pageau; Micheline Bélanger; DRDC Valcartier TM 2013-307; R & D pour la défense Canada – Valcartier; janvier 2013.

Introduction ou contexte:

Le but du présent rapport est de résumer les résultats de trois activités d'analyse préliminaires menées dans le cadre de l'orientation de l'architecture technique du projet Concept pour l'Image Reconnue de Soutien Opérationnel (CIRSO). Il identifie également les principaux résultats et propose quelques réflexions initiales sur la manière dont ceux-ci pourraient influencer la conception du système. L'objectif du projet CIRSO est de démontrer un nouveau concept d'intégration d'outils d'information et aide à la décision pour aider le commandant et le personnel de l'équipe de Soutien Opérationnel des Forces Canadiennes. Considérant que cette nouvelle capacité est prévue pour être intégrée dans le système d'information d'entreprise du MDN, ces travaux permettront de mieux comprendre le contexte, les modalités et la manière de tirer parti de l'environnement cible.

SMA (GI) Centre d'Excellence de Partage d'Informations (CEPI) est chargé de définir et de mettre en œuvre l'Architecture d'Entreprise de Référence du MDN ainsi que la liste des normes et spécifications devant être observées au sein du MDN. CEPI est également en charge de fournir des services communs au MDN et d'identifier, au besoin, les outils qui supporteront ces services de base. C'est dans cette perspective que la première activité a consisté à analyser de façon globale l'architecture d'entreprise du MDN tandis que la seconde activité se concentre sur Informatica PowerCenter, un outil de base important pour l'intégration de l'information. Enfin, la troisième activité consiste à évaluer Endeca Latitude, un composant clé d'un système d'information qui fournit des capacités d'analyse et d'intelligence d'affaire.

Résultats:

Largement basée sur des standards ouverts et en ligne avec les concepts d'architecture émergents principalement d'une approche orientée service, l'architecture de référence CEPI est suffisamment flexible pour être adapté à la plupart des systèmes d'information d'entreprise du MDN, y compris les exigences système de haut niveau du projet CIRSO.

Informatica est un chef de file dans son domaine et offre une vaste gamme de produits supportant l'intégration de données. La suite d'outils PowerCenter a démontré une très grande aptitude à fournir une capacité d'ETL (Extract, Transform and Load). L'inconvénient est que l'outil nécessite de la formation pour être utilisé de manière adéquate, mais cet effort en vaut la peine. L'outil offre probablement plus de fonctionnalités que nécessaires au projet, mais compte tenu de sa disponibilité et qu'il fait partie de la solution d'entreprise, il doit être considéré comme un candidat sérieux de l'environnent technologique de CIRSO.

Endeca Latitude démontre assurément des capacités utiles pour manipuler les données et les métadonnées qui favorisent la découverte d'information. Il possède des composants graphiques hautement performants en appui à la visualisation et à une navigation guidée très flexible. La

combinaison du langage LQL avec les capacités de détailler et filtrer offre à l'utilisateur un moyen très souple d'analyser les données. Partie prenante de la suite Endeca Latitude, l'outil CloverETL offre de bonnes fonctionnalités ETL, mais n'est pas aussi puissant qu'Informatica PowerCenter qui peut gérer des transformations plus complexes et fournir des connecteurs à un plus large éventail de types de sources de données et qui fait déjà partie de la solution d'entreprise.

Perspectives:

Étant aligné avec l'architecture d'entreprise du MDN devrait faciliter la transition des concepts du IRSO dans le monde opérationnel. Dans cette optique, des travaux devraient être réalisés pour évaluer toutes autres composants clés de l'environnement technologique du MDN basée sur la direction du SMA (GI) et du CEPI. L'une de ces technologies est Autonomy IDOL qui a été choisi comme le moteur de recherche d'entreprise et qui peut potentiellement jouer un rôle important dans la capacité du IRSO.

Table of contents

List of figures

List of tables

Acknowledgements

The authors wish to thank Christabel Mak, Change and Configuration Manager at ISCoE, for her great support and collaboration in sharing DND enterprise tools and knowledge. The authors also wish to thank LCol Yan Francki and his team of DCSAS 5 (Developmental Lab - Capability and Structure Analysis Support) for sharing their insights and knowledge on Endeca Latitude.

This page intentionally left blank.

1 Introduction

1.1 Scope

This document covers the results of three different tasks in the course of the "Reference Architecture Analysis" activity to support the "Front End Analysis" phase of the Concepts for Recognized Operational Support Picture (CoROSP) project. The results and the range of technologies investigated have been limited by contractual facilities and prioritized based on available DND guidance.

Those three tasks are:

1. ISCoE Architecture Analysis
2. Informatica Power Center Assessment
3. Endeca Latitude Assessment

1.2 Objective

The main objective of this report is to summarize key findings from the three tasks and how they could influence CoROSP project reference architecture. Secondly, this report also provides some initial technology guidance and suggestions for the next phases of the project.

1.3 CoROSP

CoROSP is a Technology Demonstration Program (TDP) project that aims to demonstrate concepts supporting operational support domain awareness to assist commanders and staffs in decision-making. A net-enabled demonstration environment will be built leveraging on existing and new net-centric/service-oriented architecture concepts to provide information integration, knowledge management and decision support tools drawing information from all relevant sources of information. [1]

1.4 Document Overview

The first three sections describe tasks, execution approaches, and results. In conclusion, the overall outcome is put in context with project requirements, with a presentation of ways it could influence technologies environment evolution in support of the Recognized Operational Support Picture (ROSP).

This page intentionally left blank.

2 ISCoE Architecture Analysis

2.1 Context

The ADM (IM) Information Sharing Centre of Excellence (ISCoE) is responsible for defining and implementing a Reference Architecture, providing a foundation of generic services and functions, as well as a list of standards and specifications to be observed at DND. Considering the ROSP requirement to have access and manipulate information provided by different DND data sources and applications, there is a strong alignment between CoROSP TDP and ISCoE.

An initial discussion with an ISCoE representative on reference architecture blueprints confirmed mutual objectives and a potential collaboration. It became of interest to analyze more closely the benefits and limitations of this architecture with respect to the needs envisioned by CoROSP.

Some tools to support this architecture have already been identified and are currently used by the operational systems. ISCoE intent is to stay independent of those tools as much as possible instead basing the architecture requirements on open standards. However, the technical documentation that we used at the time of this analysis was very limited and in an early phase.

The results presented in this section are derived from the work realized in early 2011 and described in the contract reports [2] and [3].

2.2 Objective

The objective of this task is to analyze the current Enterprise Architecture (EA) proposed by the Information Sharing Centre of Excellence (ISCoE) and more precisely to evaluate its Reference Architecture capabilities and limitations in the context of the Recognized Operational Support Picture (ROSP).

2.3 Approach

The execution of this task is divided in four steps. The first step consists in conducting a limited survey of architecture analysis methods and proposing a tailored methodology adapted to the current task context. The second step is to assess all existing documentation on ISCoE architecture, in order to identify and find missing information elements for the analysis. The third step is to proceed with the ISCoE architecture analysis based on the CoROSP project requirements. In complement, the final step is to review and compare Joint Command Decision Support for the 21st Century (JCDS 21) project Core Test Bed (CTB) with ISCoE architecture to identify major differences and propose the most suitable approach to CoROSP needs.

2.4 Results

2.4.1 Analysis Method

After a short survey of architecture analysis methods, the following have been investigated in more detail to determine the most appropriate method for the current context:

- Software Architecture Analysis Method (SAAM) [4]
- Active Reviews of Intermediate Designs (ARID) [5]
- Architecture Trade-off Analysis Method (ATAM) [6]
- Architecture Level Modifiability Analysis (ALMA) [7]
- Active Design Review (ADR) [8]

This set of methods has been chosen based on their completeness and their popularity in the literature. SAAM and ALMA methods analyze the architecture impact using scenario variation. ATAM, like SAAM and ALMA, is based on a scenario technique to evaluate the behaviour of the architecture but uses multiple competing quality attributes such as modifiability, performance, availability and security. ADR has a different approach, using review questionnaires aimed at the validation of the design quality. This technique also validates the consistency of the components and standards compliance but requires detailed documentation and usually concentrates on specific parts of the system architecture. ARID is a hybrid of ATAM and ADR, allowing the behaviour of the architecture performance to be assessed through typical scenarios as well as validating the coherence of the design. A tailored version of the ARID methodology has been selected mainly because it covers more aspects of the architecture and it is still adequate for the preliminary state of ISCoE documentation. The original method is a nine step process divided into two phases, as illustrated in Figure 1.

Phase 1: Pre-meeting	Step 1: Identify reviewers
	Step 2: Prepare design presentation
	Step 3: Prepare seed scenarios
	Step 4: Prepare for the review meeting
Phase 2: Review meeting	Step 5: Present ARID method
	Step 6: Present design
	Step 7: Brainstorm and prioritize scenarios
	Step 8: Perform review
	Step 9: Present conclusions

Figure 1: Active Reviews of Intermediate Designs

The customized method is essentially based on Phase two of ARID. The execution of ARID requires a team of participants that involves stakeholders, lead architect, expected designers and a facilitator. Considering that two people were available for the review process and a limited set of scenarios were available, the method was consequently adapted. Since the selected reviewers were two members of the team, it was not necessary to prepare a presentation to explain the

design to the reviewer. It was decided to use the scenarios that were already defined and, accordingly, there were no need to have a review preparation meeting. In that context, Step 1 to Step 5 have been replaced by a single step to review the available scenarios. Figure 2 shows the steps of the resulting method.

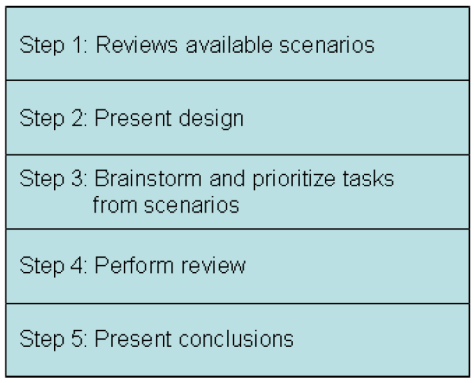

Figure 2: Tailored ARID method

The broad outlines of the process can be described as follows:

Step 1: Instead of creating new scenarios and trying to identify the most appropriate ones as stated in the original ARID method, this step uses existing defined ROSP scenarios and tries to extract the most pertinent ones to cover architecture requirements.

Step 2: Using a common set of material, the two reviewers go through the documentation and build a shared understanding without questioning the rational of the architecture decision.

Step 3: Since playing the scenario while simulating architecture behaviour can require a significant effort, it is necessary to prioritize and select more precise tasks in the scenarios that trigger specific architecture conditions. This selection can be based on task recurrence, complexity or stakeholder importance.

Step 4: This step is the core activity of the review. For each simulated task, this step involves highlighting the system activities, the required services, the data use and the load on the system. This allow issues and considerations to be documented concerning the architecture design.

Step 5: This final step is to document the conclusions deduced from observations analysis of the previous step.

2.4.2 Documentation

Considering the limited availability of ISCoE documentation, a new document has been produced that compiles and structures all the pertinent information collected through PowerPoint presentations, draft documentation and discussions with ISCoE team. This document [3] has been used as a reference for the next step of this analysis.

The two main views of this architecture are the System View (SV-1) [Figure 3] as defined by DND/CF Architecture Framework (DNDAF) and the Reference Architecture [Figure 4] based on a SOA architecture.

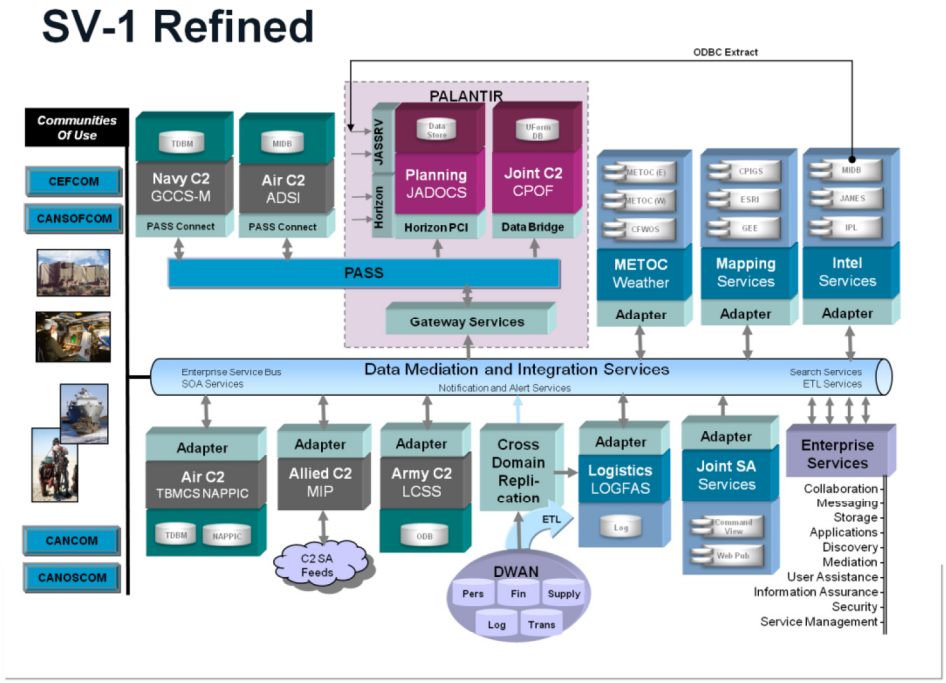

Figure 3: ISCoE SV-1

The core component of SV-1 is the Enterprise Service Bus (ESB). The definition of this concept may vary depending on the product vendors. It has also evolved over the years. The key role of the ESB is to enable enterprise integration by uncoupling and controlling a distributed service oriented environment. Based on open standard, the ISCoE concept is to use an application adapter to connect existing enterprise systems to the bus as the main C2 environment or common systems, like weather and mapping. The ESB ensures transparent communication between the systems but it also provides some common enterprise services such as authentication, quality of service, service level, service discovery and business rules management as well as some more specialized services such as monitoring, encryption and publish and subscribe management.

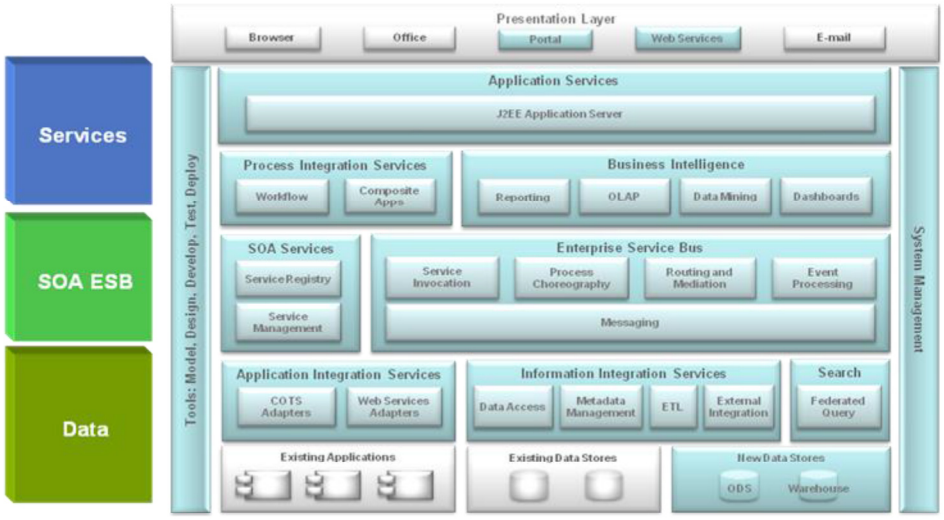

Figure 4: ISCoE Reference Architecture

The ISCoE Reference Architecture area of responsibility is composed of services categories represented in a layered bloc diagram. The detailed requirements specifications under those categories are not yet defined but refer to the generic concept that they represent. An adequate categorization of services will promote enterprise governance.

2.4.3 Analysis

The ARID method is based on scenarios. For this purpose, the two scenarios to be used were identified in [9]:

Deployed: A bridge on one of the supply routes has been blown up and needs reconstruction - equipment, supplies and personnel in theatre are insufficient.

Domestic: Request for 20 additional CF fire fighters is received - locate and deploy the available qualified resources.

The architecture has been analyzed by playing selected tasks of those scenarios and the proposed ISCoE architecture has demonstrated that it could support most of the system's needs. Even if this activity was limited in scope, considering the high level of the architecture description, it would be difficult to refine the resulting observations by expanding the rage of scenarios. However, the noted observations are described in Table 1.

Table 1: ISCoE Architecture Analysis Observations

Observations	Comments
In our example, Canadian Forces Supply System (CFSS) being replicated from the Unclassified, the available information could not be fully synchronized.	If such a situation is plausible, mechanisms should be established so that information accessed is as accurate as possible.
If connectivity is lost, information may become inaccessible.	Alternate Networks or communications roads should be considered.
Because the query has to go through a BUS this adds a layer of data manipulation that could be avoided.	To ensure efficiency in service levels, certain types of applications could connect directly without going through the Enterprise Service Bus (ESB)
BUS adds configuration and maintenance overhead.	Same comment as above.
Because of use of BUS and related application, this can diminish network responsiveness and slow down the processes. Such information could be obtained through direct system interfaces without use of Enterprise BUS.	Same comment as above.
Some systems are not fully integrated.	To ensure maximum efficiency, the applications will have to comply and integrate as much as possible with the target architecture.
Accessing a high volume of data (e.g. maps) can jeopardize the system's responsiveness.	Network potential issue that must be addressed.

2.4.4 JCDS 21 Test Bed Comparison

JCDS 21 Test Bed is an output of the of DRDC Valcartier (DRDC-V) JCDS 21 TD Project and consists in the development, integration, demonstration and experimentation vehicle for the project. The aim of the project was to demonstrate a Joint, Net-enabled Collaborative Environment to achieve Decision Superiority. The development of a JCDS 21 TD Test-bed supports systems integration from both a technological and human perspective, and provides the means to ensure interoperability among the participating systems when building the overall JCDS 21 TD system [11].

The JCDS 21 Test Bed is based on Service Oriented Architecture (SOA). As illustrated in Figure 5 [11], the Core Test Bed (CTB) is the central part of JCDS 21 architecture and uses an Enterprise Service Bus (ESB) as an integration platform. A Universal Description, Discovery and Integration (UDDI) service is used by the ESB to allow the services available on the system network to be discovered and integrated. ESB enables information exchange between services of four main groups: "Common" and "Information Access" services from the CTB and "Emulated Consolidated Secret Network Infrastructure (CSNI)" and "Decision Support" services to support business capabilities. Two other categories of services: "Stimuli" and "Metrics" allow supporting experimentation and evaluation of the Test Bed.

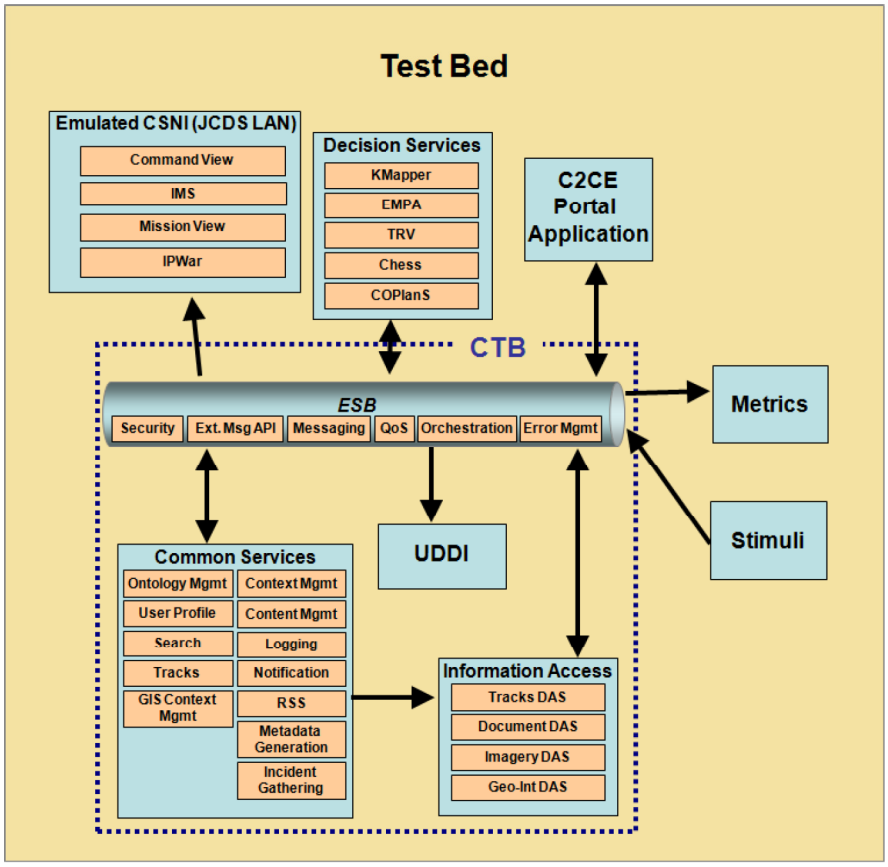

Figure 5: JCDS 21 Test Bed Architecture Overview

ISCoE Reference Architecture and JCDS21 Test Bed Architecture have been influenced by the Defence Information Services Broker (DISB) project [10] and their resulting approaches are similarly based on SOA architectures principles. Because the ISCoE architecture documentation does not provide the lower design details, the comparison has been performed using a higher level of information. Table 2 summarizes the main differences and similarities.

Table 2: ISCoE / JCDS21 CTB Architecture Comparison

	ISCOE	JCDS
Maturity	Only at conceptual level.	Has been proven functional and most aspects have been widely tested through multiple projects.
Documentation level	Almost non-existent	The Concept of Operation, the requirements and the core architecture are very well documented and most of the projects that were using the architecture components are also well documented..
Implementation	Not implemented yet from the target perspective. Some components are already used or deployed but nothing is integrated according to the target architecture.	Has been implemented on the RD network.
Approach	SOA	SOA
Open Standards	Probably yes but need to be validated.	Yes for the most part.
ESB	Yes	Yes
Flexibility / Reutilisability	Yes	Yes
ETL	Through the use of Informatica suite, ISCoE implements ETL services for extracting data from outside sources, transforms it to fit operational needs (which can also include quality verification) and finally loads the generated data into a target (database or data warehouse)	JCDS provides some services to extract data from multiple external sources and transfers it to the requesting services. but there are no transformation services or loading services to store into a target database.
Ontology	Such services are not envisioned in the actual architecture plans.	JCDS exploits an external ontology service tool that is used to generate the OWL (files) descriptions used by other services such as K-Mapper. For example, these ontologies are used to add qualitative attributions to the documents stored in the K-Mapper database.

2.4.5 Final Thought

The ISCoE reference architecture is flexible enough to be adapted to most information systems in the DND enterprise. Without having to replicate every piece of the architecture, it is possible to build a compliant architecture based on an open standard that respects the major guidelines. The main building blocks are in line with recognized emerging architectural design and are well supported by the industry. This can be a great benefit in the context of a Technology Demonstrator Project where some larger technologies or architecture components are less suitable or too demanding for this project's scale. For example, the Enterprise Service Bus is a very useful technology that contributes to a loosely coupled architecture and improves enterprise reliability. But a lesson learned from the JCDS21 project is that even considering those benefits, its role was minimal compared to the effort required to manage the technology. Business service orchestration is a key role of the ESB but considering the time and budget constraints of the project, less priority has been allocated to this higher level abstraction of service definition and on this technology facet expertise [12]. Service orchestration has ultimately not been implemented. The CoROSP demonstration system should assess the benefits of each of its architecture building blocks and prioritize their need within the project scope while maintaining a potential integration into ISCoE Reference Architecture.

Recognizing that CoROSP requirements and ISCoE architecture definition are at a high level and are continuously evolving, it is recommended that the two initiatives maintain close collaboration to refine the accuracy of this analysis.

This page intentionally left blank.

3 Informatica Power Center Assessment

3.1 Context

An important role of ISCoE architecture is to provide data warehouse management capability. A key function of this activity is to store enterprise strategic data for subsequent analysis. This work is commonly supported by an Extract, Transform and Load (ETL) process. ADM (IM) selects Informatica Power Center to perform this task. Considering the needs related to such capabilities in CoROSP architecture and that DND owns an enterprise license for the tool, it is important to analyze that tool and assess its strength and limitation. ISCoE envisions the use of ETL also to propagate and control required data at a higher security level network. This use of the tool is not covered in this study.

The results presented in this section are derived from the work realized in early 2012 and described in the contract report [13].

3.2 Objective

The objective of this work is to get a clear understanding of Informatica PowerCenter capabilities in order to determine its roles and how it can be integrated into the CoROSP technical solution.

3.3 Approach

The study related to Informatica has been divided in three parts. The first part includes the installation of the tool suite in order to evaluate the technical considerations and proceeds with a series of basic capabilities test cases. The second part focuses on the ability of Power Center to be integrated within a SOA architecture while the third part demonstrates and evaluates the feasibility through a proof-of-concept of this integration.

3.4 Results

3.4.1 Initial Tests

PowerCenter is a client-server application that comes in four editions: *Standard, Advanced, Real-Time and Cloud*. The edition license provided by ADM (IM) is the *Standard* one. The server deployment requires the installation of a database server to store its configuration parameters, Oracle 11g has been selected for this purpose. Some configurations are needed to be able to exploit the server, although some issues have been encountered, the documentation provides all the information needed to execute this duty. The client installation is a straightforward wizard mechanism that mainly installs four components:

- Repository Manager
- Designer
- Workflow Manager
- Workflow Monitor

Following the installation and configuration, some basic test cases allowed us to verify some stated features on Extraction and Transformation. Table 3 and Table 4 compile the corresponding results and observations.

Table 3: Extraction Features

Data Transfer	Activities	Importance	Status	Description
Table to Table	Connects to relational DB tables	High	Done	Two database tables can be easily integrated using GUI based mapping facility.
CSV/Txt to Table	Use of file source utility	High	Done	Text (txt) or delimited files (CSV) can be integrated using I/O reader for the ETL tool. Many transformation features are available.
XML to Table	Creating table from XML directly	High	Done	XML file can also be transformed to one or many tables and the data can be shifted to relational DB.
Source to many targets	One-to-many data integration	Medium	Done	Two types: a) Row level: data from one table can be shifted to 2 tables based on routing transformation utility b) Column level: columns from one or many tables can be pushed to one or table and vice versa.
Sources to one target	Many-to-one data integration	Medium	Done	

Issues Encountered	Details
PreSQL execution	Experienced difficulty pushing data to a non-empty table. It required pre-SQL execution to remove data prior to ETL process run and service turned unresponsive.
XML Integration Issue	XSD compliance to table is required. We are currently searching how xml (hierarchical) data can be pushed to many tables through one ETL process.

Table 4: Transformation Features

Name	Activities	Importance	Status	Description
Connected Lookup	Connects 2 relational DB tables	Medium	Done	Normal look up while integrating many tables.
Unconnected Lookup	Use of file source utility	Medium	Done	Used as function call with table or other data source.
Value/Expression transform	Fixed Value Addition	Medium	Done	Additional column addition and direct value or variable mapping is possible to this column value on the row
Router transformation	Conditional data movement	Medium	Done	Row Level splitting of one table to many tables.
Maplet Transformation	Mapping based transformation	Medium	Done	Usual mapping based transformation where source tables
Union Transformation	Many to one transformation	High	Done	Tabular data from more than data-source get composed into relevant single target based on primary and foreign keys
Rank Transformation	Top n records from sorted data	Medium	Done	Rank -based source to target data transformation; useful for gathering most important data from large tables.
Normalizer Transformation	For Cobol and mainframe data	Medium	Not Tested	Used for Cobol mainframe data transformation
Java Transformation	Use of Java Code	High	Done	Most powerful transformation with ability to use Java code during transformation
Saving ETL process	ETL process backup and save	High	Done	Two types of saving: back up inside the same server and transfer the whole work as xml file to another server

Issues Encountered	Details
Training Requirements	It requires sufficient training in design mapping using PowerCenter Designer tool.
Integration Performance	We experienced slow processing in Java-based transformation compared to other ones.
Java Functions	In our experience it requires designer to know Java functions separately in order to use them. Example: generateRow() to generate row in target table.

This work allows us to conclude that Informatica Power Center is a powerful tool and that the stated features are well supported. This product offers many capabilities but in return, it requires a greater effort to develop the necessary understanding to use it efficiently. In the context of large business, it is reasonable to claim that the workforce will be available to ensure the installation, configuration and maintenance of the tool.

3.4.2 Integration

Intended to be part of an enterprise architecture, PowerCenter provides SOA based capabilities to interact with other architecture building blocks. Data extraction and transformation techniques need to be designed through Power Center client applications but two main web services can be used to retrieve the workflow and data mapping information generated by the client and to control the workflows execution process. Respectively, those two services are PowerCenter Metadata Service and PowerCenter Data Integration Service. Business services can then be created and use those Power Center services to support the ETL needs, considering that the transformation engineering was made beforehand.

Figure 6 shows a simple implementation of this SOA integration capability of Informatica Power Center using an existing client application prototype, namely LogSeer, by developing two services: Mapping Service and Execution Flow Service.

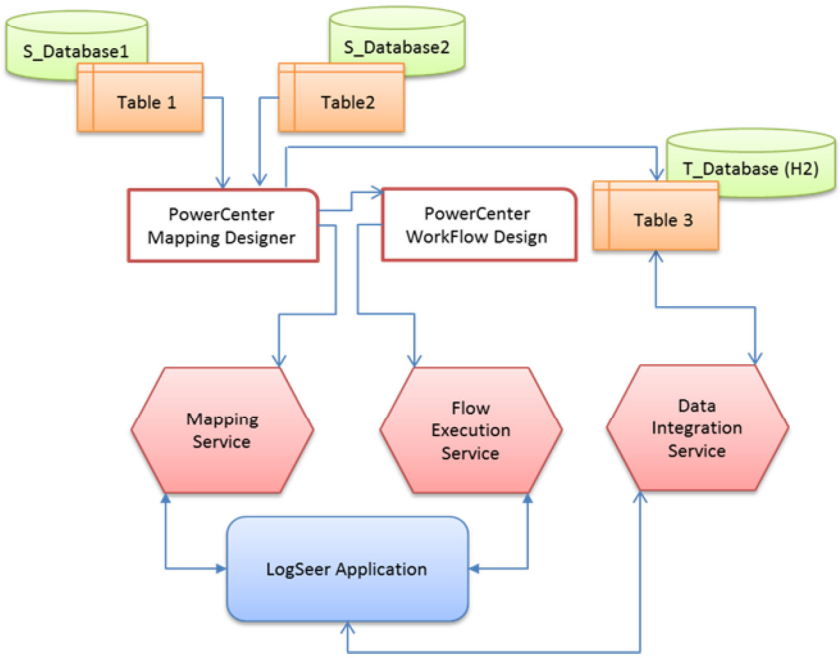

Figure 6: Service Oriented Architecture using PowerCenter application capabilities

This solution will be used in the proof-of-concept described in the next section.

3.4.3 Proof-of-Concept

A limited scenario has been developed to demonstrate and validate Informatica Power Center integration in the Enterprise Architecture. Usually, ETL services are involved in building the data warehouse to be dedicated to, among other things, business intelligence (BI). In this scenario, the intent is to use Power Center services to control the load of external sources of data into Latitude Endeca MDEX and to use Latitude Endeca BI tools to analyze the resulting information aggregation.

The approach is composed of the following steps:

1. Create Power Center Mapping
    a. Load data from different data source type
    b. Include Data Mapping to change the data structure
    c. Include some Java transformation on the data
    d. Merge data from multiple data sources in one data structure
    e. Generate XML from new data structure
    f. Call a web service with the new XML as input to push data in Endeca MDEX
2. Create a simple workflow that executes the defined mapping
3. Use the integration solution described in the previous section from LogSeer
    g. Get the workflow created
    h. Execute the workflow
4. Consult Endeca Latitude portal to explore and analyze the data

This scenario implementation has required the development of a utility web service that interprets the XML input and accordingly calls the Endeca Data Ingest web service. The resulting scenario diagram is depicted in Figure 7.

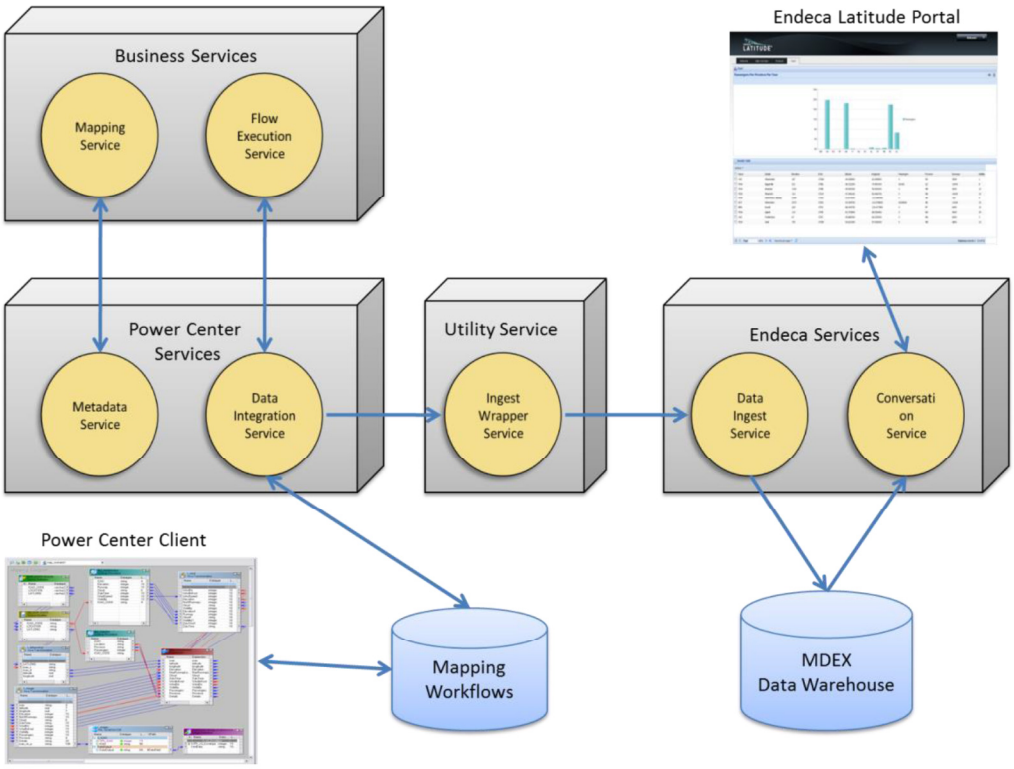

Figure 7: Proof-of-Concept Scenario

Even if the *Ingest Wrapper Service* is the primary design supporting this particular context scenario, it has been designed to be easily extended to fully support the *Endeca Data Ingest Service*.

3.4.4 Final Thoughts

Informatica is a leader in this field and provides a wide range of product supporting data integration. It is important to differentiate between features presented as part of Informatica solution and features that are product specific. The assessment study allowed us allows to better understand the capabilities related to Informatica PowerCenter Standard Edition tool for which DND as a corporate license. It is common to read references on Informatica whitepaper of data virtualization capability and real-time processing. These capabilities are not part of the version currently available by ADM (IM). There is a potential need for data virtualization in the CoROSP project and Informatica PowerCenter is not an appropriate tool for that requirement. Informatica provides a Data Services product for this particular purpose but there is no possibility for using this tool in a research and development environment at reasonable cost.

In this analysis, PowerCenter Standard tool suite demonstrated a very strong ability to provide ETL capability. The drawback is that the tool needs some training to be used adequately but it is worth it. The tool probably provides more capabilities than needed by the project but considering its availability and that it is part of the enterprise solution, it should be considered as a serious candidate for the CoROSP technology stack.

4 Endeca Latitude Assessment

4.1 Context

The Director Capability and Structure Analysis Support (DCSAS) 5 Developmental Lab is currently building a Center of Excellence (CoE) for Endeca Latitude expertise. This software is also identified as a potential tool for analytical processing in the context of logistic support. Considering that some of the capabilities delivered by this tool are likely to be part of the ROSP concept and that the military client already has a license for Endeca Latitude, it is appropriate to conduct further study of the tool and better target the capabilities it can provide to ROSP.

The results presented in this section are derived from the work realized in early 2012 and described in the contract report [14].

4.2 Objective

The objective of this work is to build a solid understanding of Endeca Latitude capabilities and to assess how it can be integrated into a ROSP solution.

4.3 Approach

This study is composed of four main phases. The first phase is to review the large bundle of documentation provided with the tool to develop a basic understanding of its capabilities, and to identify potential features of interest and interrogation that should be validated in subsequent phases. This leads to the installation phase to document the technical considerations, followed by a first set of test cases on the overall Latitude suite (phase two). Once the environment is confirmed in a fully working state, the third phase aims to build a proof-of-concept to evaluate key aspects of the tool. The final phase identifies, documents and analyzes those key results of the Endeca Latitude assessment that could influence the ROSP solution.

4.4 Results

4.4.1 Documentation

The documentation provided with Endeca Latitude describes all its aspects in considerable detail including installation, configuration, and use, as well as possible software development and integration. A high level view of the product is described in [15]:

“Endeca Latitude® is a complete platform for the agile BI of an enterprise. Latitude can be used to develop and manage analytic applications for any department or specialized use. The platform consists of the integrated elements shown in Figure 8 below:

- **Latitude Studio** – A highly-interactive, component-based environment for building and deploying enterprise-class analytic applications powered by the Endeca MDEX Engine.
- **The MDEX Engine** – a hybrid search-analytical database designed for agile BI.
- **Latitude Information Integration Suite** – A powerful platform with a high-performing ETL tool, system connectors, and content enrichment libraries for unifying diverse information, including structured data and unstructured content."

Figure 8: Endeca Latitude software overview

Figure 9 describes those three main components from a technical architecture perspective. The Endeca Latitude Studio is built over Liferay open source portal to which were added custom portlets based on open standards and Java Specification Requests (JSR), specially designed to manipulate and display in an efficient manner the data store in the MDEX engine. The portlets communicate with the MDEX using Discovery Service (DS) and Conversation Web Service (CS). The MDEX engine and data store mechanism is the core product of Endeca Latitude and few details are provided on its implementation in the available documentation. A small number of accesses allow controlling the MDEX, in particular to load data using the Data Ingest Web Service (DIWS) or the Bulk Load Port. The DIWS description is open and well described in the Endeca documentation to provide the capability to third party application to push data into the MDEX. On the other hand, the Bulk Load Port is Endeca internal implementation and can only be used in conjunction with Endeca components. This Bulk Load Port connector is included in a set of integrated components part of a commercial version CloverETL tool, to compose the Latitude Information Integration Suite.

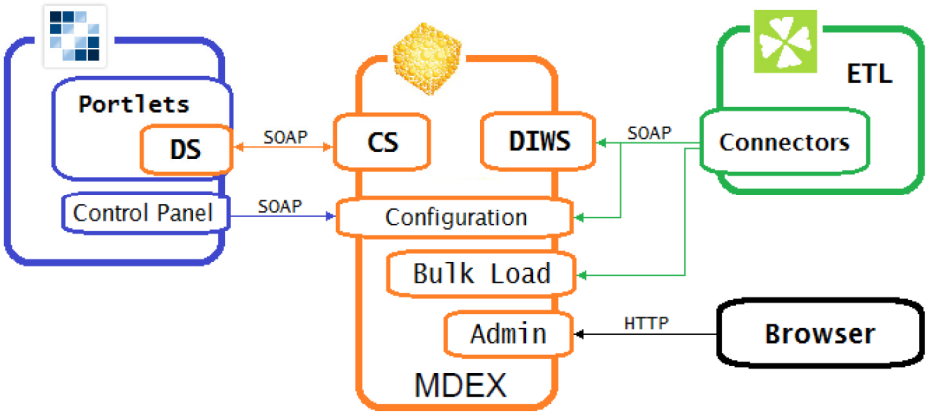

Figure 9: Complete view of Latitude components interactions

4.4.2 Installation

The software installation is pretty simple: assisted by a straightforward wizard, the overall installation take only few minutes and is suitable for most environments. It is also possible to proceed with a more complex setup to distribute components on different machines when required by the system load.

The initial setup comes with a useful QuickStart package that contains almost everything needed to experiment with all components of Endeca. The Bike Store sample covers all the steps from data extraction, metadata generation, data transformation, data load to the MDEX, data displays and analysis through the different portlets. Figure 10 presents an example of the resulting interactive chart and tabular reports.

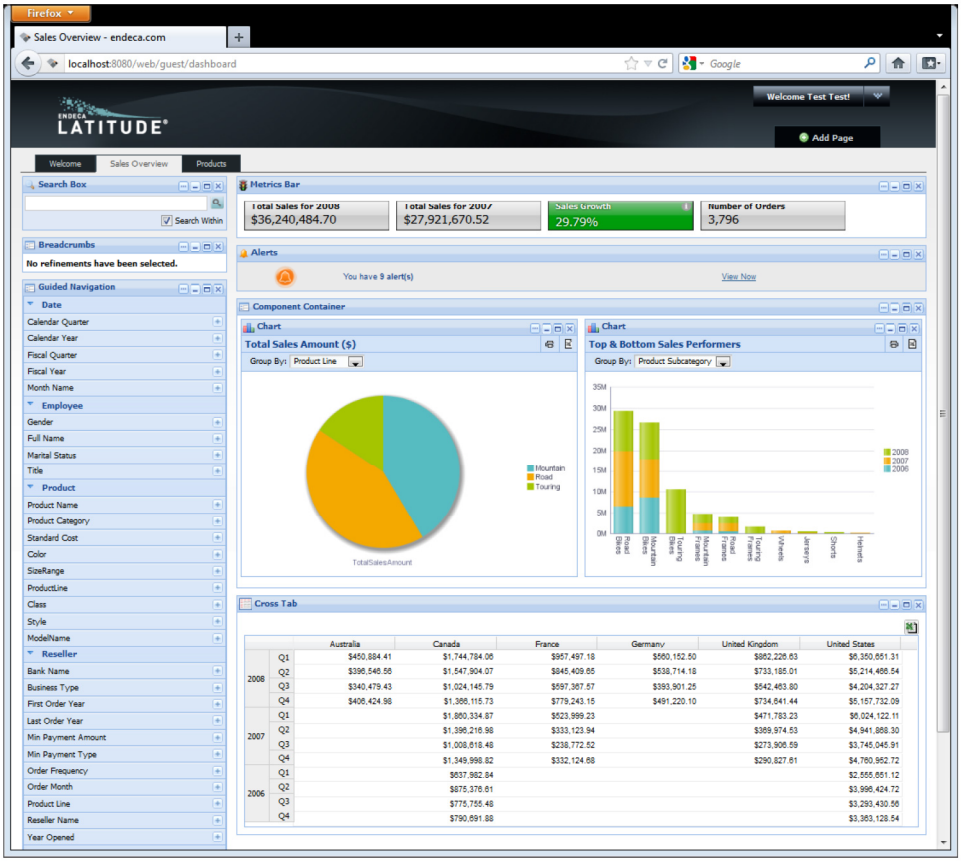

Figure 10: Latitude Studio - Bike Store Sample

Some initial considerations were derived by playing around with the QuickStart environment. Table 5 summarizes the main observations on the data model structure.

Table 5: Data Model Initial Observations

Element	Observation
Data schema	There is no notion of tables; all data are flattened under a record, like one big table. This also implies that one unique key needs to be identified, common for each record.
Data attributes	When the data origin from a relational table with one-to-many relation, the resulting record duplicates the attributes related to those multiple relations.
Data relation	Every piece of information on a record is an attribute; there is no more notion of relation.
Data transformation	The data format can be transformed in the new structure, for example the color code can be changed by the color name, the date format can be changed, two column first name and family name can be merged in one, etc.

These data model observations are common to data warehousing to provide an efficient way to manipulate the data in the context of BI. But this consideration needs to be well understood to benefit from the strength of Endeca analysis capabilities. It should be noted that these data model transformation functions are provided through the CloverETL tool embedded with Endeca Latitude suite but they can also be provided by any other ETL specialized tool.

4.4.3 Proof-of-Concept

The proof-of-concept used a simple existing R&D prototype to manage DRDC - Valcartier C2 systems to go through all the required processes to perform data analysis on the C2 systems information using Endeca Latitude Studio. The important steps of that process are described below.

Step 1: In preparation for record structure design and unique key definition, it is necessary to transform the original data model from the regular relational schema to a star schema representation. Figure 11 presents this transformation for the C2 system documentation model. In this case the "System View" is identified as the facet surrounded by information on the system (Metadata) identified as dimension of the model and the unique key is simply the "System ID".

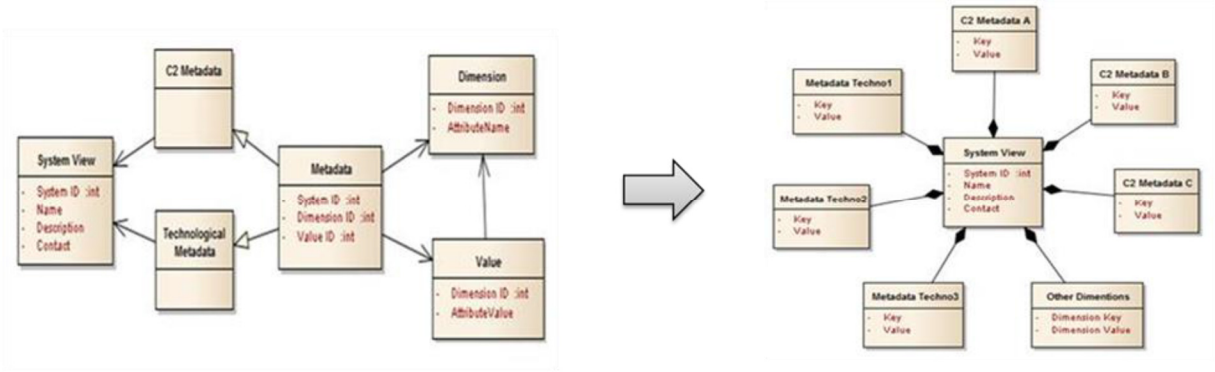

Figure 11: Data Model Transformation

Step 2: The next step is to proceed to the ETL process on the data. It should be noted that this case implies one simple data structure. Usually, in a complex environment such the ROSP, some initial data cleansing is required to efficiently perform this step. Consequently, it is assumed that the data architect has mastered an understanding of the different models involved. CloverETL provides the necessary components to execute data extraction of the MySQL database containing the information though a standard SQL query. Metadata should then be associated with collected fields in order to proceed with record structure creation. CloverETL provides a graphical environment to build the process flow and assist metadata creation. Figure 12 shows the user interface used to create metadata associated with one data source.

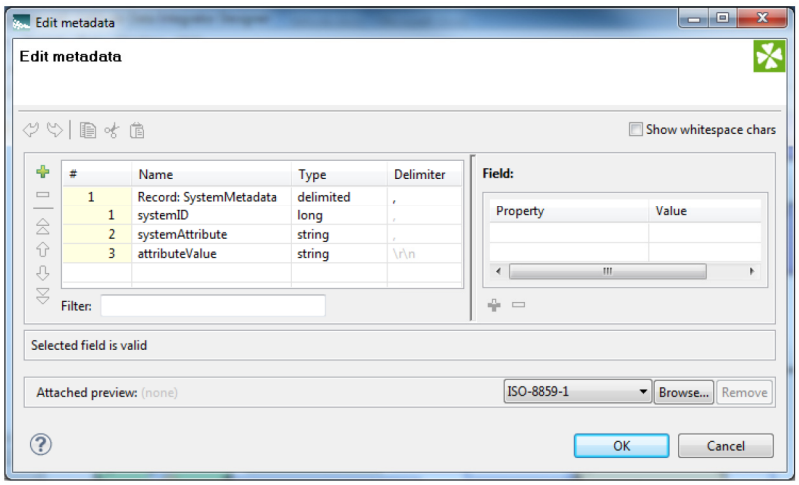

Figure 12: CloverETL Metadata Definition

Figure 13 depicts the extraction of the 14 C2 systems dimensions considered and the "Systems" facet. After being merged they are loaded into the MDEX through the "Add KVP" block.

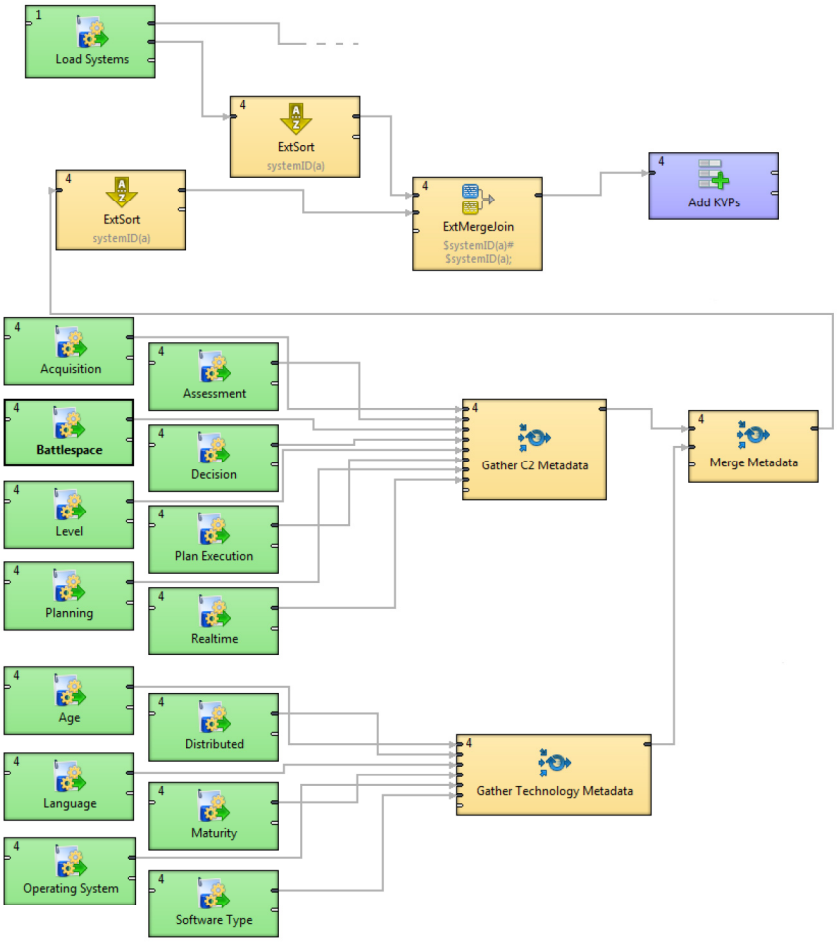

Figure 13: Record Meta Data Creation and Insertion in the MDEX

Step 3: The data ingested into the MDEX can now be accessed for analysis through Endeca Latitude Studio. Figure 14 is a screenshot of the resulting analysis interface, demonstrating filtering, navigation and report capabilities on the data.

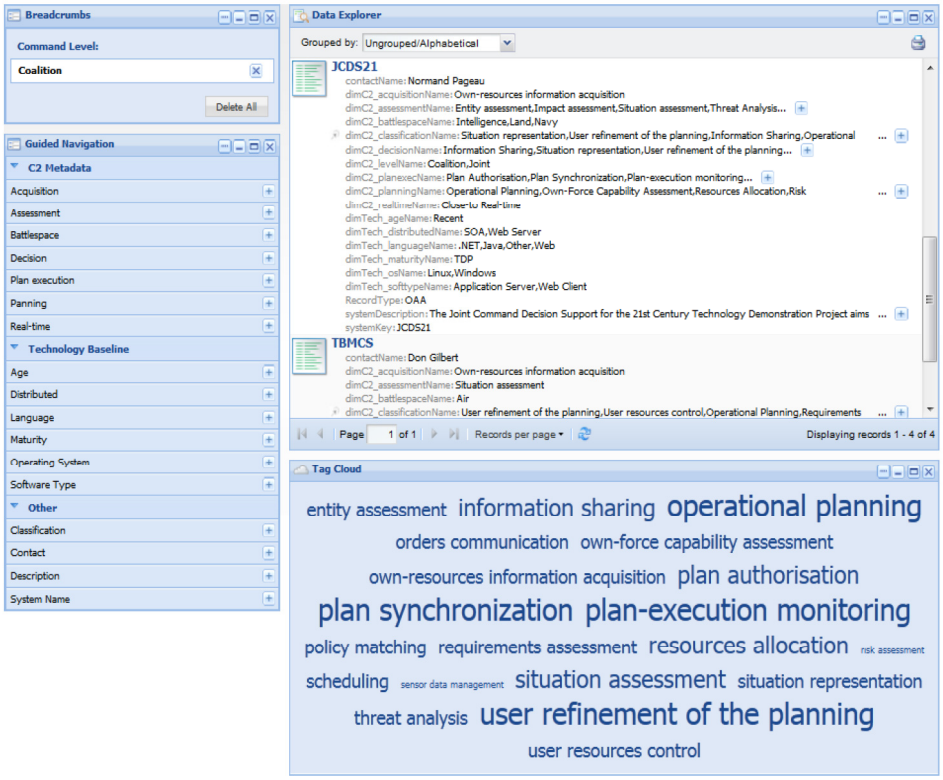

Figure 14: Endeca Latitude Studio for C2 Systems

4.4.4 Analysis Results

This section provides the most relevant results of the overall assessment of the tool.

Ease of use: The Endeca Latitude suite is well documented and provides a straightforward installation process that makes it easy to be up and running very quickly. The tools CloverETL and Latitude Studio provide graphical interfaces that in general provide great support for the user. The LQL query editor may be short on intuitive features, providing a small area for editing and creating a high risk of losing changes while navigating the portal. LQL editing with an external tool is recommended. The complexity behind this tool is related to the ETL process engineering – the analytical query mechanism generates the expected results. CloverETL interface is simple to use and the LQL query is powerful but they are targeted to a specialized user community. Decision makers should use the resulting product through the Latitude Studio portal, which is very intuitive.

Performance: The ETL is not a real-time process and depending on the complexity of the systems involved, it can take hours to execute once the process engineering and data cleansing is realized. Once the data is in the MDEX, the engine and the data structure is very efficient and give the end user a quick response time to complex queries. However, the data structure generates a large amount of redundant information and requires a high storage volume. In relation to the data

ingested by the engine, the MDEX is very greedy on memory consumption. In a complex environment such the ROSP, we recommend using a standalone machine to run the MDEX.

Flexibility: Having the key part of the environment scripted through ETL graphs and LQL query gives a lot of flexibility for customizing and reusing existing components without having to build a new version of the system. The CloverETL process abstracts the legacy system data sources and provides a flexible way to respond to changes in data models. The flexibility in the LQL queries mixed with the possibility of adding and configuring portlets on the fly, gives analysts and decision makers the possibility of adapting their working environment. The different services developed for interacting with the MDEX makes it possible to have custom applications developed to ingest and retrieve data and reduce the dependencies limitation with CloverETL and Endeca Latitude Studio. Finally, because the visualisation portal is based on an open source and open standard this facilitates the creation of new portlets specific to custom needs.

Capabilities: Latitude Studio offers a good set of configurable portlets that allow efficient visualization and navigation through the information. CloverETL offers good ETL capabilities but is not as powerful as other commercial tools like Informatica PowerCenter that can handle more complex transformations and provide connectors to a wider range of data source types.

4.4.5 Final Thought

Endeca Latitude definitely demonstrates useful capabilities for manipulating data and metadata in a way that helps information discovery. It has highly capable graphic components to support visualization and a very flexible guided navigation. The combination of the LQL language with the drill down and filtering capabilities gives the user a very agile means to analyze the data.

In the context of an enterprise architecture, it is important to provide flexible integration mechanisms, ideally based on established standard. Endeca Latitude has two main integration capabilities, first at the visualization layer where it uses portlet components and secondly it provides web services to interact with the MDEX engine. Both techniques are based on open standards. One weakness of this integration strategy is the performance of web services and XML when ingesting a large quantity of data into the MDEX. This is not an issue with CloverETL because Endeca provides a bulk load connector to perform this specific task efficiently.

Performance requirements are not yet entirely defined for the project. It should be taken into consideration that this kind of business intelligence tool is not well adapted to real-time data handling. Depending on the ROSP features, a certain time delay can be acceptable, for example in trend analysis based on historical data.

Endeca Latitude would need some improvement regarding the usability of the portlets configuration and a better assisted LQL query builder tool. It would also greatly benefit from a friendly MDEX administration tool assisted by performance metrics and MDEX information. Even if no formal load performance tests were conducted, it seems to require a significant use of resources. CloverETL is easy to use and provides many essential ETL capabilities but it would be useful to be able to connect to a more powerful ETL to include more complex workflows. Endeca provides web services for that issue but it will take considerable effort to reach the level of functionality already supported by CloverETL.

Endeca Latitude is an expensive product but a license is available to the project through the military client. On the other hand, even if this capability is identified in the global enterprise architecture, this software is not currently part of the technical solution of ADM (IM).

Recently, Oracle acquired the Endeca company and all their related tools and renamed it Oracle Endeca Information Discovery. Considering that Oracle has its own solution for some of the components of Endeca Latitude, like Oracle Data Integrator as ETL tool, it is natural to question the evolution of the Endeca suite and add a risk in investing in the current form of the technology solution.

5 Recommendations

5.1 Governance

Decision support systems rely on an accurate understanding of the situation. In a complex environment such as the military world, information on a situation is distributed through many systems with variable granularity, quality and availability. The orchestration of information sharing between these systems becomes a major issue. This orchestration is possible only by controlling the architecture from an overall perspective of the military organization. This control is ensured according to the best practice of an enterprise architecture implementation.

The CoROSP project would benefit from an enterprise architecture with access to the information needed to build operational support capability. The fact that such an enterprise architecture is not available amplifies the technical challenge related to the project. It is therefore proposed to build a proof-of-concept of the software implementation of the architecture based on the ISCoE guidance, but its complete implementation and governance is beyond the project mandate.

In an environment where several systems have to exchange information and functionalities, it is necessary to establish well defined communication mechanisms based on open standards and service contracts. The technologies or architecture choices for building this environment must enable the enterprise governance. A reliable open and accurate documentation of this choice, the interaction mechanisms and standards are a fundamental part of enterprise architecture frameworks good practices.

Governance is limited by dependence on third party technologies and knowledge whose evolution cannot be controlled by the organization. Strategic decisions must balance acquisitions versus internal production. Business functions should not depend on the implementation of a tool whose evolution, configuration and usage are not fully controlled by the organization. Otherwise, changes to the business model of the enterprise or a different evolution of the tools can compromise the overall enterprise system. In some cases, where mandatory systems are needed to operate specific platforms, this control is difficult to avoid but it should be taken into account in the design to abstract their interfaces. Also, wherever possible, the acquisition of different technologies that provide the same functionality should be limited. Clear and precise governance is required when several tools offering different capabilities or with an intersection of similar functions are part of the enterprise software solution. A typical case of this situation is the presence of two ERP solutions to manage different types of resources in the DND enterprise.

For this reason, the ROSP capability technology stack should be as much as possible in line with ADM (IM) selection. On the other hand, in a complex organization such DND, the selection is driven by highly complex requirements that can result in options that are too demanding to manage in a demonstration project such CoROSP. In such circumstances, it can be appropriate to rely on technologies and architecture based on the same standards. The associated reduction in the level of complexity in the management of the different technologies makes it worthwhile even if it means that these technologies may have fewer capabilities.

5.2 SOA Reference Architecture

ISCoE reference architecture is fundamentally SOA based. Service-oriented architectures offer a theoretical solution that idealizes the information systems supporting the enterprise business model. From this theoretical perspective, the proposed architecture effectiveness is unquestionable. But in practice, this theory is difficult to put in place. The main cause is that a common model, clear and approved by all communities of the enterprise is difficult to reach and the granularity of the decomposition of the business model leaves room for interpretation. Also, considering that the evolution of the enterprise can impact its business model and that the part of the enterprise modeled by the information system can be incomplete or changing, the architect needs to design the SOA with a current knowledge of the situation. Those factors can significantly affect service reusability benefit. Additionally, this benefit will be impacted by the nature of the CoROSP project; being focused on a specific field of the enterprise, the overall design of the architecture is out of scope and because ISCoE reference architecture is currently mainly theoretical, the core services that should be reused do not necessarily exist. A tailored version of the service architecture should be developed under a demonstration budget constraint project, but that will be difficult to reuse.

Still SOA is a good choice of architecture even in the context of the CoROSP project. It gives the flexibility to integrate iteratively new capabilities, to leverage on existing services and enable enterprise collaborative design in opposition to a stove pipe approach. Because SOA is part of the ISCoE foundation, it is an important aspect of the TDP to assess the ROSP viability in such an architecture. This ROSP demonstration could influence, at some extent, ISCoE generic services and potentially generate the requirements for new services.

Main SOA technology uses XML as a means of communication. Depending on the granularity of the services and the complexity of the system, this could generate performance issues, especially in a highly distributed environment. The flow of information should be carefully analyzed in the service design to guarantee the level of service required.

Figure 15 presents the starting point of a high level concept of SOA architecture influence with an ISCoE Reference Architecture diagram from the Open Group SOA Reference Architecture. The following sub-section discusses the key roles of those layers from a conceptual perspective.

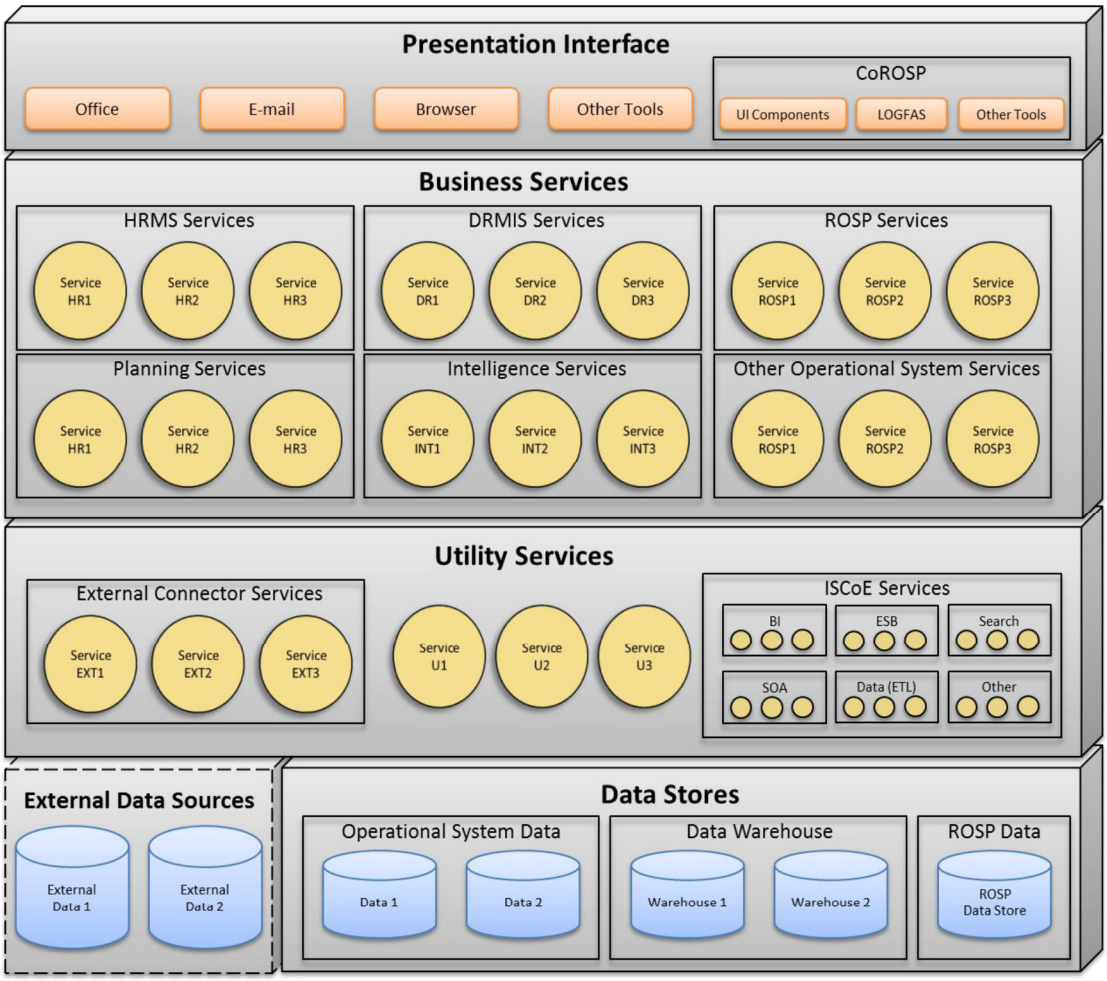

Figure 15: ROSP Reference Architecture Concept

5.2.1 Data Layer

Data is the basis of information systems; they can be generated or collected by applications, operators or sensors. Data can originate from two main sources, DND enterprise or external systems.

What characterizes external data is the fact that they are not under DND governance. This data can come from other government organizations or from public businesses providing equipment and services to operation support. For example, the availability of equipment spare parts or public infrastructure can be maintained by external systems. Data exchange with external systems enables collaboration between businesses.

The DND data store can be of various formats and distributed within the enterprise network. For the ROSP, the required data can be divided under three groups: the data specific to the ROSP capability, the data used and generated by other operational systems in the enterprise, and storage of strategic data to enterprise decision support. In this last category, data are normally duplicated

and loaded into a data warehouse on a different structure aimed at business intelligence capabilities. It is recommended to maintain such a structure for ROSP to provide efficient analytical capabilities.

The data layer is also responsible for storing metadata. Metadata provide additional information about the data like type, quality, accuracy, source, semantic, sensitivity and other relevant properties. Special attention should be placed on the metadata model design to support information management of the ROSP.

5.2.2 Utility Layer

The major responsibility of this layer is to abstract the representation and distribution of systems. It should provide the business services with a representation of the enterprise that is coherent with the ROSP ontology and it should be transparent as regards how and where the real physical data is stored, including data from external systems.

In a system integration environment, new information can be generated by combining different asynchronous data sources. This implies that we should pay particular attention to data source refresh rates and validate the fused information consistency. It is important to isolate the services dedicated to this duty from those in charge of business tasks. Creating new information by merging data can alter the sensitivity or classification of information and must be carefully considered.

Another role of this layer is to provide common enterprise services in support of the business layer. Those services should normally be part of ISCoE core responsibility since they must be centrally governed. The utility layer can be in fact subdivided into more than one layer but is simplified here for conceptual purposes. Figure 16 presents the SOA middleware as envisaged by the Open Group. This block diagram categorizes those enterprise services that should part of the utility layer. Notice that it does not include the “Business App Services” that should be represented in the business layer.

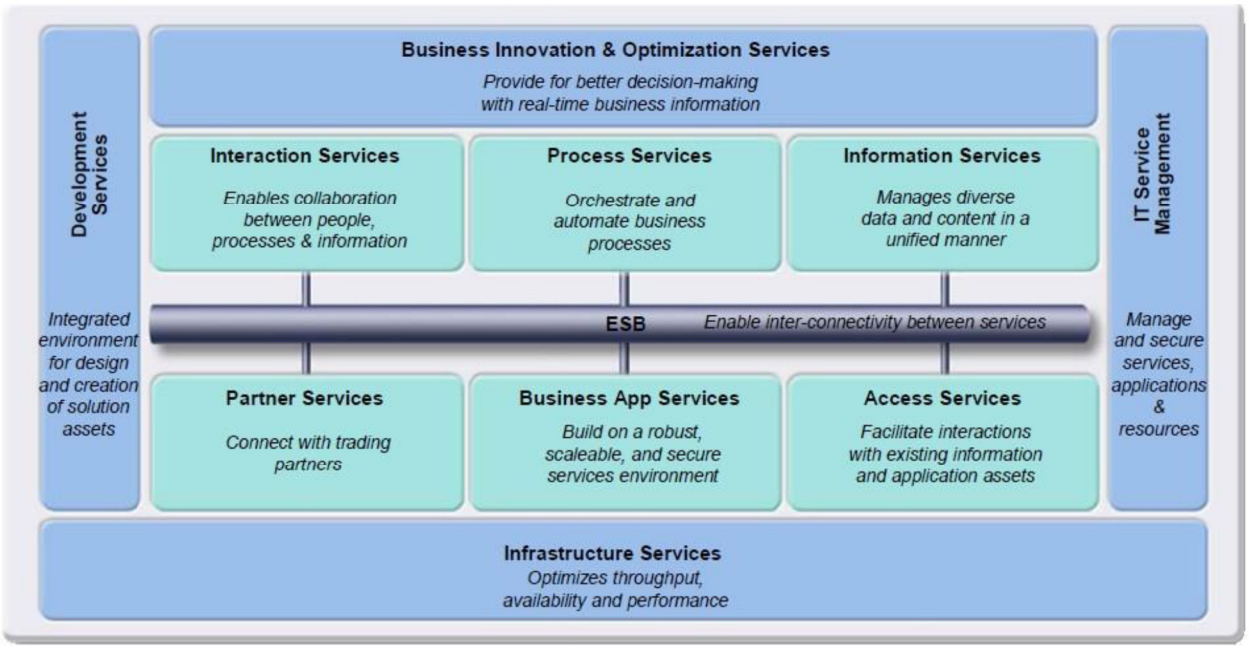

Figure 16: Open Group SAO Reference Architecture Middleware View

This layer should ensure the security and quality of services and provide reliable enterprise wide capabilities like search and collaboration services presented under "Enterprise Services" in Figure 3 of ISCoE architecture.

It is from that perspective that an ETL tool like Informatica PowerCenter, Endeca Latitude or another BI and Autonomy IDOL search engine would be made available by this layer and under the responsibility of AMD (IM) ISCoE.

5.2.3 Business Layer

In order to meet ROSP functional requirements, this layer primarily manages ROSP business services and process. A single service could determine the level of readiness of a unit but a more complex orchestration of services can support activity processes like new supply procurement. This layer will use the utility layer to access information and to share in the enterprise framework.

The ROSP capability will also need to re-use capabilities from other systems' business services. From example, it can use a service from the Human Resources system to know how many persons have a specific qualification.

This layer will enable capabilities exchanges between ROSP and other enterprise systems. The operational support picture is one view of the situation; it could be considered as a layer of the overall operating picture. Operational systems like Command View and CPOF should be able to access ROSP services to integrate key information in the common view.

5.2.4 User Interface Layer

This layer is presented as a group of tools and graphical components available to the ROSP user community to interact with the system. It is mostly composed of graphical interface components that display in a flexible and user friendly manner the information that the user needs to process tasks and also the interaction mechanism that trigger parts of the business process.

This layer needs to be independent from the business layer in order to allow different components customized for user or role needs to interact with the same business capabilities. This layer should also permit access to business capabilities with device specific requirements. For example, desktop, tablet, smartphone or wall displays with specific characteristics such as device performance, screen size and resolution, screen rotation, touch screen capabilities, etc.

5.3 Experimentation Environment (Testbed)

In a R&D environment, it is important to give the architecture the flexibility to integrate services in support of the experimentation and evaluation of the system. This mainly implies the capability to collect metrics from system execution, to be able to simulate some part of the environment that can't be replicated and to determine some assumptions about the limitations in scope of a R&D project. This is normally referred as the project "Testbed". This environment can also play a key role in the development process as a "Sandbox" for the development team to validate and test the system.

5.3.1 Metrics

The services design should integrate mechanisms that will allow metrics extraction from system usage to validate architecture concept and implementation strategies. Those metrics should also give the user interaction information to evaluate system navigation and interfaces. Metrics plays a key role in validating and to optimizing the system. It can give important indicators to the scientific and development team on how the system can help user tasks and where improvement is needed. Metrics can also be extrapolated to estimate performance in operational conditions. Time to perform task is normally the most important factor considering the time pressure associated with military activities.

By looking at users' recurring tasks, it is possible to log the data and services involved and evaluate the impact of a low bandwidth on highly distributed services or important data volume. This could also indicate where the system would benefit from automation. From that perspective, the design of the CoROSP experimentation environment should consider metrics management that includes, but is not limited to, system performance, situation awareness and decision support.

5.3.2 Simulation

For the purpose of project tests and demonstrations, real environment conditions, systems or data feed can be replaced by a simulation module. The two primary objectives are first to be able to reuse existing simulated behaviours and second to overcome real systems and information sources not maintained by the project.

It should be transparent to the system if the data comes from a real operational system, if it has been generated from a ground true scenario simulator, if it comes from data recorded during past execution or if it is simply pre-packaged datasets. The data utility services should support different configurations.

5.3.3 Assumptions

In order to specify the technical scope of the CoROSP TDP, it is important to specify and document some assumptions. This section proposes some potential assumptions that need to be extended, refined and validated before the initial design of the solution.

5.3.3.1 Network

- Real network latency condition will not be simulated; however bandwidth usage metrics will be available to predict numbers.
- System distribution over cross-domains is assumed to be managed by other project.
- Network security level is considered the same for all available services

5.3.3.2 Security

- Network security is not managed by this project
- Computer security is not managed by this project
- The sensitivity of combined unclassified data sources will remain unclassified.

5.3.3.3 Robustness

- Even if the architecture is well designed to support redundancy, this aspect will not be validated through this project.
- Data available through different systems is considered accurate.
- Network is considered available all the time

5.3.3.4 Performance

- Global performance will be evaluated in a strategic and operational tempo
- Hardware performance will not represent current operational environment
- Project effort will not put emphasis on performance optimization

5.3.3.5 Technical

- This project will encourage but will not be limited to the use of open standard and technology guidance from existing operational systems except for hard requirements.
- Hardware will not be limited to current operational use.

- Even if the architecture supports multiple languages, only English is supported in this project.
- Demonstration conditions will be the same as test environment conditions.
- There is no technology limitation for the client and server development environment

5.3.3.6 Functional

- Some core services will be provided as part of the project to support demonstrated functions but are considered to be provided in the current operational environment. Some of those envisioned required Core Services are:
  - Error Management
  - Service Discovery
  - Quality of Services
  - Service Orchestration (workflow and business process management)
  - Authentication mechanism
  - Notification
  - Search
  - Collaboration tools
  - Content Management
- Demonstrated ROSP functionalities will be only covered partially based on client priority
- Some existing operational system capabilities may be simulated for ease of demonstration

5.4 ROSP Transition

Being aligned with DND enterprise architecture should ease ROSP concept transition into the operational world. Considering the in-progress status of ISCoE architecture implementation, it is thought likely that ROSP capabilities will transition incrementally.

To build an understanding of the operational support picture, there are two main requirements: to access and to analyze the related information. The first requirement is to access information by the aggregation of structured and unstructured data mainly localized on operational systems. Then, this information can be analyzed either by an analyst or via a collection of advanced decision aid features.

Data aggregation is well supported by the ETL process that extracts data from the operational systems, transforms the data into a new efficient format and structure and finally loads it into a data warehouse. Our study of Informatica PowerCenter leads us to conclude that this ETL tool is

very powerful and rich in features. Taking into account that it is also part of the DND enterprise solution, we recommend using it to support this ROSP requirement.

A situation is analysed through specific requests to the data warehouse where the retrieved information can be visualized by an analyst or processed by innovative operational support functions. Endeca Latitude Studio provides tools to support advanced capabilities to request and visualize the information, known as information discovery. Moreover, Endeca Latitude Suite (now Oracle Endeca Information Discovery) also offers a comprehensive set of features covering ETL and data warehousing. The ETL capability delivered with Endeca is provided by the CloverETL tool that overlaps with Informatica PowerCenter features. Endeca provides an open Web Services API which allows for direct data integration from other tools, such as Oracle Data Integrator and Informatica PowerCenter, but the development of such a connector implies in-house support to any evolution of Endeca. Considering the new acquisition of Endeca by Oracle, it is reasonable to assume that the product will be subject to adaptation in upcoming releases. Thus, it is recommended, as a short term solution, to keep CloverETL and wait for an enterprise solution. Even if it implies a redundant ETL process, this concept, illustrated in Figure 17, enables a quick deployment of an initial ROSP capability.

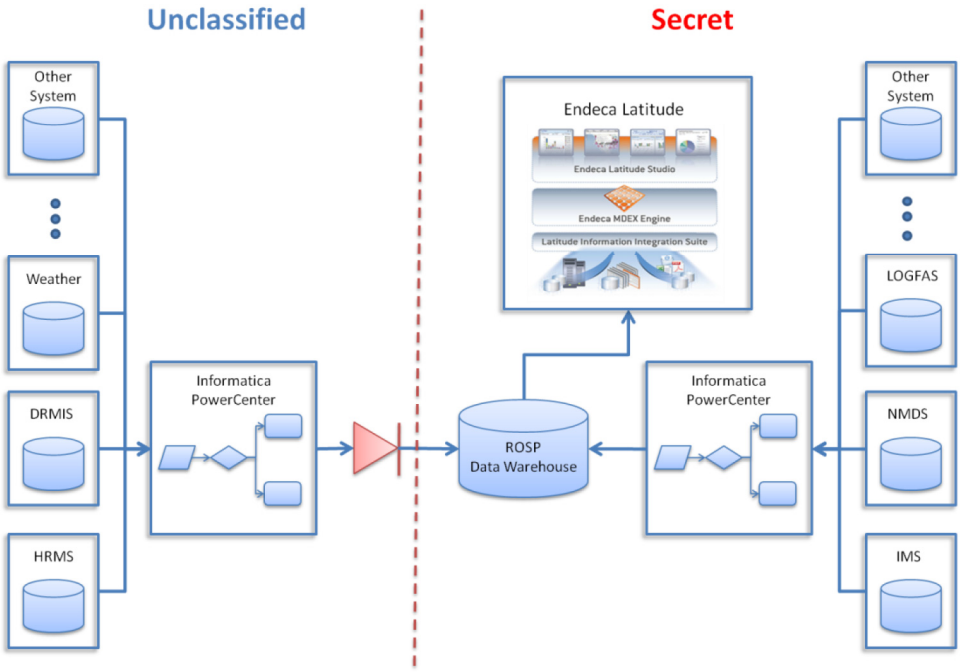

Figure 17: Operational System View - Endeca Latitude

The main aim of this solution is to acquire access to operational system data sources and to model the new ROSP data warehouse. The difficulty of this task can increase if operational systems database documentation is missing. It is recommended to keep the ROSP data warehouse on the classified secret network. The main reason is to be able to integrate data from the secret network

but also because even unclassified data integration may increase the sensitivity of the resulting information. Information from operational systems of lower-level security networks, like DRMIS and HRMS, can be loaded to higher-level networks using Informatica PowerCenter ETL through a data diode as proposed by ISCoE architecture.

While ISCoE enterprise infrastructure implementation evolves, this solution will allows the ROSP capability to integrate advanced decision aid functions. In the event that DND selects an enterprise tool to cover the features currently offered by Endeca, it will be appropriate to remove the ETL process duplication. CoROSP to be developed can follow an incremental transition based on priorities as shown on Figure 18.

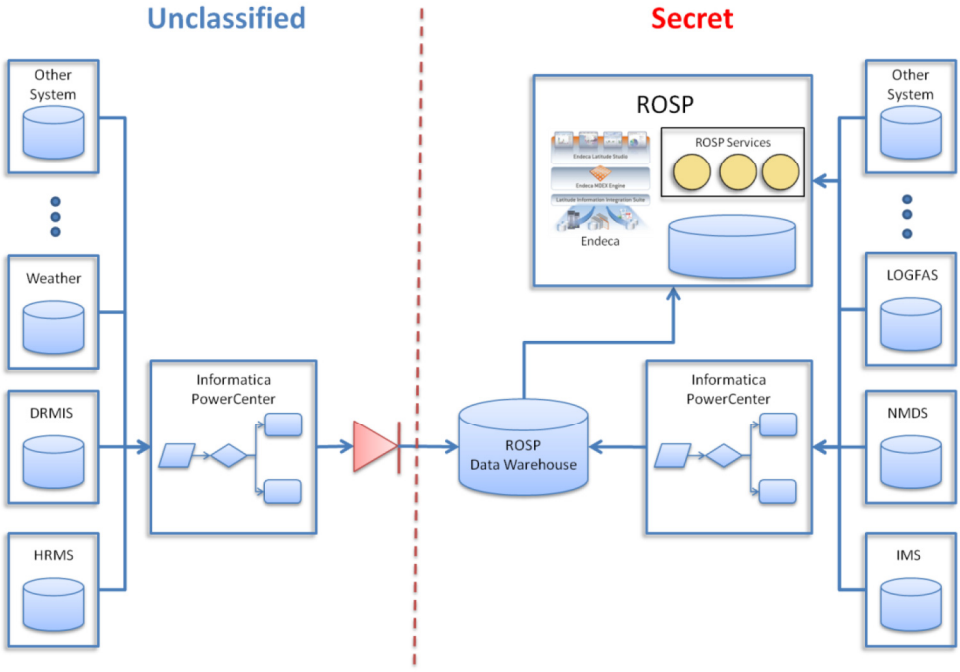

Figure 18: Operational System View - ROSP

Of course, additional advanced features cannot depend exclusively on the data warehouse information. Thus, connectivity will also be necessary with operational systems to access existing services and support functions requiring more up-to-date or near real-time data. Unfortunately, this real-time access will be limited to data and services on the secret network.

In order to reach ISCoE target architecture, the ROSP solution should be designed as a system that offers a new range of services to the enterprise that covers a well-defined domain and that promotes services and components reusability as well as information sharing. Ultimately, the ROSP integration, as illustrated by Figure 19, should take advantage of the Enterprise Service

Bus to access enterprise operational system data and services but also to share its own data, services and graphical components to the entire enterprise.

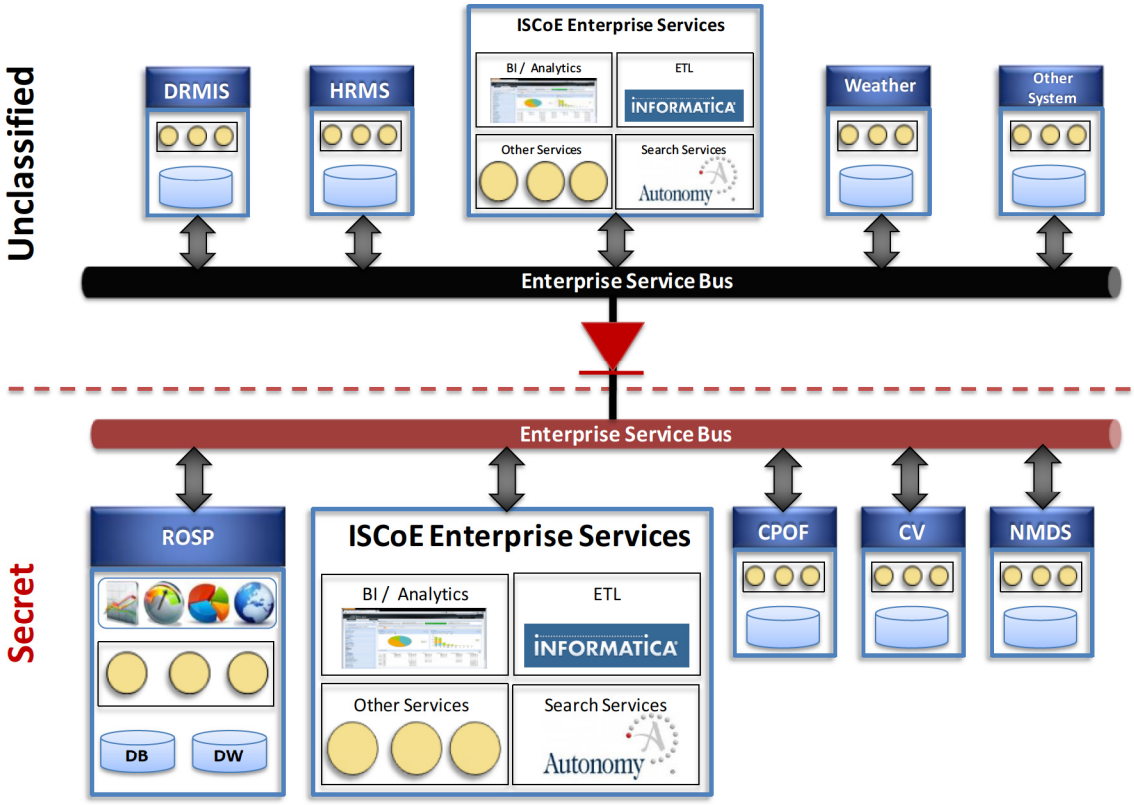

Figure 19: ROSP Operational System View

While each system plays a unique role as orchestrated by the enterprise architecture, ISCoE should provide the core and common services to contribute to system uniformity and coherence. More specifically, the ROSP system should provide unique graphical components, advanced decision-aid services and local data stores but use common enterprise services as they are made available by ISCoE like Informatica PowerCenter and BI/Analytics tools chosen by DND. Also, further investigation should be conducted to evaluate the use and the integration of search service into the ROSP capability provided by the Autonomy IDOL product. On the other hand, the developed ROSP services and graphical components should be sufficiently flexible to be integrated into different types of ROSP client applications but also to be integrated or used by other existing systems. It is envisioned that the ROSP information layer would let other systems like CPOF and Command View integrate it into a more global context. Finally, if it is considered appropriate within the DND direction, the developed ROSP capability can be integrated and reside into an existing operational system.

This page intentionally left blank.

6 Conclusion

To demonstrate concepts supporting operational support domain awareness to assist commanders and staffs in decision-making, CoROSP TDP has to define and develop a net-enabled demonstration environment leveraging on existing and new net-centric/service-oriented architecture concepts to provide information integration, knowledge management and decision support tools drawing information from all relevant sources of information. In this report we described key findings that should be considered while designing CoROSP project reference architecture. We also provided some initial technology guidance and suggestions related to the use of ISCoE architecture, Informatica and Endeca for the next phases of the project.

This page intentionally left blank.

References

[1] DRDC Valcartier, CoROSP TDP Synopsis Sheet, October 2010.

[2] FUJITSU Consulting (Canada), Sharing Center of Excellence Architecture Analysis, DRDC Valcartier CR 2012-044, May 2011.

[3] FUJITSU Consulting (Canada), ISCoE's Documentation, DRDC Valcartier, May 2011.

[4] R. KAZMAN, L. BASS, G. ABOWD, M. WEBB, SAAM: A Method for Analyzing the Properties of Software Architectures, Proceedings of 16th International Conference on Software Engineering, May 1994.

[5] P. CLEMENTS, Active Reviews for Intermediate Designs, Carnegie Mellon University Technical Note CMU/SEI-2000-TN-009, August 2000.

[6] R. KAZMAN, M. KLEIN, M. BARBACCI, T. LONGSTAFF, H. LIPSON, J. CARRIERE, The architecture Tradeoff analysis method, ICECCS '98. Proceedings of fourth IEEE International Conference on Engineering of Complex Computer Systems, August 1998.

[7] P. BENGTSSON, N. LASSING, J, BOSCH, and H. VAN VLIET, Architecture-level modifiability analysis (ALMA), The Journal of Systems and Software 69, 2004.

[8] D. PARNAS and D. WEISS, Active Design Reviews: Principles and Practice, Proceedings of eighth International Conference on Software Engineering, 1985, pp132-136.

[9] Private Communication, Human Factor 4 (HF-4) & Human Factor 5 (HF-5) Potential Operational Support Injects to the Experimentation Scenarios, DRDC Valcartier

[10] FUJITSU Consulting (Canada), JCDS 21 Core Test Bed System Requirements Specification, DRDC Valcartier, April 2008.

[11] FUJITSU Consulting (Canada), JCDS 21 CTB Software Architecture Description (SAD), DRDC Valcartier, January 2009.

[12] FUJITSU Consulting (Canada), JCDS 21 TD Release 3 Lessons Learned (DRAFT), DRDC Valcartier, January 2009.

[13] TrustSeer Corporation, Final Report On Installation, Features & Performance Analysis of Informatica Power Center, DRDC Valcartier, March 2012.

[14] LTI Software & Engineering, Endeca Latitude Software Integration Analysis, DRDC Valcartier, April 2012.

[15] Endeca Technologies, Endeca Latitude Technical Overview, Endeca Whitepaper for Business Intelligence, 2011

This page intentionally left blank.

List of symbols/abbreviations/acronyms/initialisms

ADM	Architecture Development Method
ADM (IM)	Assistant Deputy Minister (Information Management)
ADR	Active Design Review
ALMA	Architecture-level modifiability analysis
ARID	Active Reviews for Intermediate Designs
ATAM	Architecture Trade-off Analysis Method
AV	All Viewpoint
B2B	Business-to-Business
BI	Business Intelligence
BPM	Business Process Management
C2	Command and Control
C4ISR	Command, Control, Communications, Computers, Intelligence, Surveillance, and Reconnaissance
CapV	Capability View
CF	Canadian Forces
CFSS	Canadian Forces Supply System
CoE	Center of Excellence
CoROSP	Concepts for Recognized Operational Support Picture
COTS	Commercial Off-The-Shelf
CPOF	Command Post of the Future
CRM	Customer Relationship Management
CS	Conversation Web Service
CSNI	Consolidated Secret Network Infrastructure
CSV	Comma-Separated Values
CTB	Core Test-bed
CV	Command View
CV	Capability Viewpoint
CV	Common View
DB	Database

DCSAS	Director Capability and Structure Analysis Support
DEA	Director of Enterprise Architecture
DISB	Defence Information Services Broker
DIWS	Data Ingest Web Service
DND	Department of National Defence
DNDAF	DND/CF Architecture Framework
DOD	Department of Defense
DODAF	DOD Architecture Framework
DRMIS	Defence Resource Management Information System
DRDC	Defence Research & Development Canada
DRDC-V	DRDC Valcartier
DS	Discovery Service
DW	Data Warehouse
EA	Enterprise Architecture
EAP	Enterprise Architecture Programme
EABOK	Enterprise Architecture Body of Knowledge
ERP	Enterprise Resource Planning
ESB	Enterprise Service Bus
ETL	Extract, Transform and Load
HF-4	Human Factor 4
HF-5	Human Factor 5
HRMS	Human Resources Management System
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
IMS	Incident Management System
ISCoE	Information Sharing Center of Excellence
ISO	International Organization for Standardization
IT	Information Technology
JCA	Joint Capability Areas
JCDS	Joint Command Decision Support
JSR	Java Specification Requests

KVP	Key-Value Pair
LOGFAS	Logistics Functional Area Services
LQL	Endeca Latitude Query Language
MDA	Multi-Dimensional Analytical
NATO	North Atlantic Treaty Organization
NMDS	National Movement and Distribution System
OLAP	Online Analytical Processing
OLTP	Online Transaction Processing
OWL	Web Ontology Language
QoS	Quality of Service
ROI	Return on Investment
ROSP	Recognizes Operational Support Picture
R&D	Research & Development
SAAM	Software Architecture Analysis Method
SIB	Standards Information Base
SOA	Service-oriented Architecture
SOA RA	SOA Reference Architecture
SQL	Structured Query Language
SV	System View
TAFIM	Technical Architecture Framework for Information Management
TD	Technology Demonstrator
TDP	Technology Demonstrator Program
TOGAF	The Open Group Architectural Framework
UDDI	Universal Description, Discovery and Integration
UK	United Kingdom
XML	eXtensible Markup Language
XSD	XML Schema Definition

This page intentionally left blank.

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (The name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Centre sponsoring a contractor's report, or tasking agency, are entered in section 8.)

Defence R&D Canada – Valcartier
2459 Pie-XI Blvd North
Quebec (Quebec)
G3J 1X5 Canada

2. SECURITY CLASSIFICATION (Overall security classification of the document including special warning terms if applicable.)

UNCLASSIFIED
(NON-CONTROLLED GOODS)
DMC A
REVIEW: GCEC April 2011

3. TITLE (The complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S, C or U) in parentheses after the title.)

CoROSP Reference Architecture: Front End Analysis

4. AUTHORS (last name, followed by initials – ranks, titles, etc. not to be used)

Pageau, N.; Bélanger, M.

5. DATE OF PUBLICATION (Month and year of publication of document.)	6a. NO. OF PAGES (Total containing information, including Annexes, Appendices, etc.)	6b. NO. OF REFS (Total cited in document.)
January 2013	68	15

7. DESCRIPTIVE NOTES (The category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

Technical Memorandum

8. SPONSORING ACTIVITY (The name of the department project office or laboratory sponsoring the research and development – include address.)

Defence R&D Canada – Valcartier
2459 Pie-XI Blvd North
Quebec (Quebec)
G3J 1X5 Canada

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.)	9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.)

10a. ORIGINATOR'S DOCUMENT NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document.)	10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.)
DRDC Valcartier TM 2013-307	

11. DOCUMENT AVAILABILITY (Any limitations on further dissemination of the document, other than those imposed by security classification.)

Unlimited

12. DOCUMENT ANNOUNCEMENT (Any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in (11) is possible, a wider announcement audience may be selected.))

Unlimited

13. ABSTRACT (A brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual.)

This report summarizes the findings of three preliminary analysis activities conducted as part of the technical architecture orientation of the project Concepts for Recognized Operational Support Picture (CoROSP). It also identifies the key outcomes and puts forward some initial thoughts on how they could influence the system design. The aim of the CoROSP project is to demonstrate a new concept of information and decision-aid tools integration to assist the commander and the staff of Canadian Forces Operation Support team. Since it is envisioned that this new capacity will be integrated into the DND enterprise information system, this work will allow us to better understand the context, the conditions and how to leverage from the target environment.

ADM (IM) Information Sharing Centre of Excellence (ISCoE) is responsible to define and implement the DND Enterprise Reference Architecture as well as a list of standards and specifications to be observed in DND. ISCoE is also in charge to provide common services to DND and to identify, when appropriate, the tools that will sustain those core services. It is in that perspective that the first activity was to analyse the overall DND enterprise architecture while the second activity concentrates on Informatica PowerCenter, an important core tool for information integration. Finally, the third activity is to evaluate Endeca Latitude, a key component of an information system that provides analytical and business intelligence capabilities.

Ce rapport résume les conclusions de trois activités d'analyse préliminaires menées dans le cadre de l'orientation de l'architecture technique du projet Concept pour l'Image Reconnue de Soutien Opérationnel (CIRSO). Il identifie également les principaux résultats et propose quelques réflexions initiales sur la manière dont ceux-ci pourraient influencer la conception du système. L'objectif du projet CIRSO est de démontrer un nouveau concept d'intégration d'outils d'information et aide à la décision pour aider le commandant et le personnel de l'équipe de Soutien Opérationnel des Forces Canadiennes. Considérant que cette nouvelle capacité est prévue pour être intégrée dans le système d'information d'entreprise du MDN, ces travaux permettront de mieux comprendre le contexte, les modalités et la manière de tirer parti de l'environnement cible.

SMA (GI) Centre d'Excellence de Partage d'Informations (CEPI) est chargé de définir et de mettre en œuvre l'Architecture d'Entreprise de Référence du MDN ainsi que la liste des normes et spécifications devant être observées au sein du MDN. CEPI est également en charge de fournir des services communs au MDN et d'identifier, au besoin, les outils qui supporteront ces services de base. C'est dans cette perspective que la première activité a consisté à analyser de façon globale l'architecture d'entreprise du MDN tandis que la seconde activité se concentre sur Informatica PowerCenter, un outil de base important pour l'intégration de l'information. Enfin, la troisième activité consiste à évaluer Endeca Latitude, un composant clé d'un système d'information qui fournit des capacités d'analyse et d'intelligence d'affaire.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

CoROSP

Defence R&D Canada

Canada's Leader in Defence
and National Security
Science and Technology

R & D pour la défense Canada

Chef de file au Canada en matière
de science et de technologie pour
la défense et la sécurité nationale

www.drdc-rddc.gc.ca