Common Reference Data – the foundation of e-Business

Matthew West, Shell Information Technology International, UK

1. Introduction

Business integration is one of the key business drivers today, fuelled by the ever-improving capabilities of IT. However, many attempts at business integration fail because they fail to take sufficient account for the need for business to use a common language in order to be able to integrate. The common language consists of Reference Data.

2. Why bother with Reference Data?

To understand the importance of Reference Data it is important to understand the role it plays, particularly in enabling the integration of data from different sources. This can be seen from looking at the Mickey Mouse diagram (look at the silhouette) below.

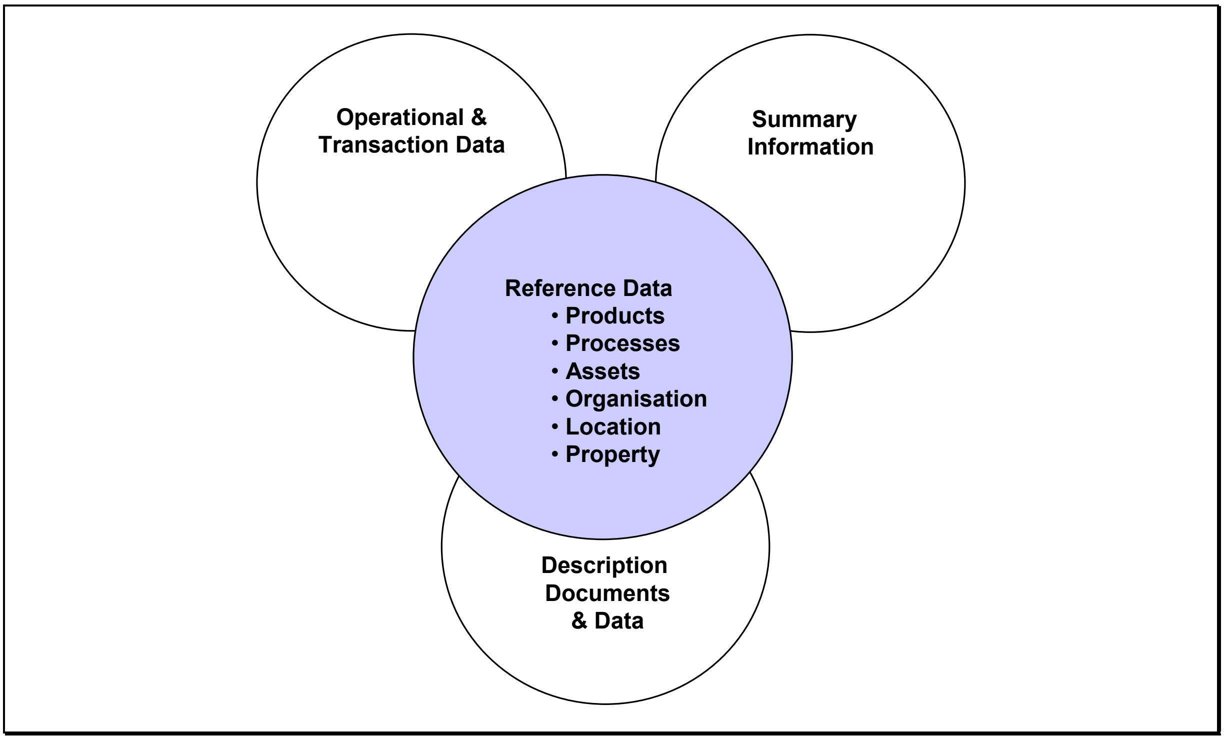

Figure 1: The Mickey Mouse diagram.

Operational and Transaction Data	The results of measurements, or the records of business transactions.
Summary Information	The aggregation of operational and transaction data, including the calculation of Key Performance Indicators, for performance management.
Description Documents & Data	Data and documents that describe or define such things as products, processes and assets.
Reference Data	The coding for such things as products, processes, assets, organisations, locations, and properties that enables these things to be unambiguously and consistently referred to across systems and organisations.

At the business level, information is shared as part of the interaction and integration of business activities. At the technical level, information is shared when it is created and used by different computer systems, people, or organisations. Figure 2 illustrates some possible flows of information between different parts of an organisation, and with the outside world. All of these mean that Reference Data needs to be shared and managed across the organisations concerned.

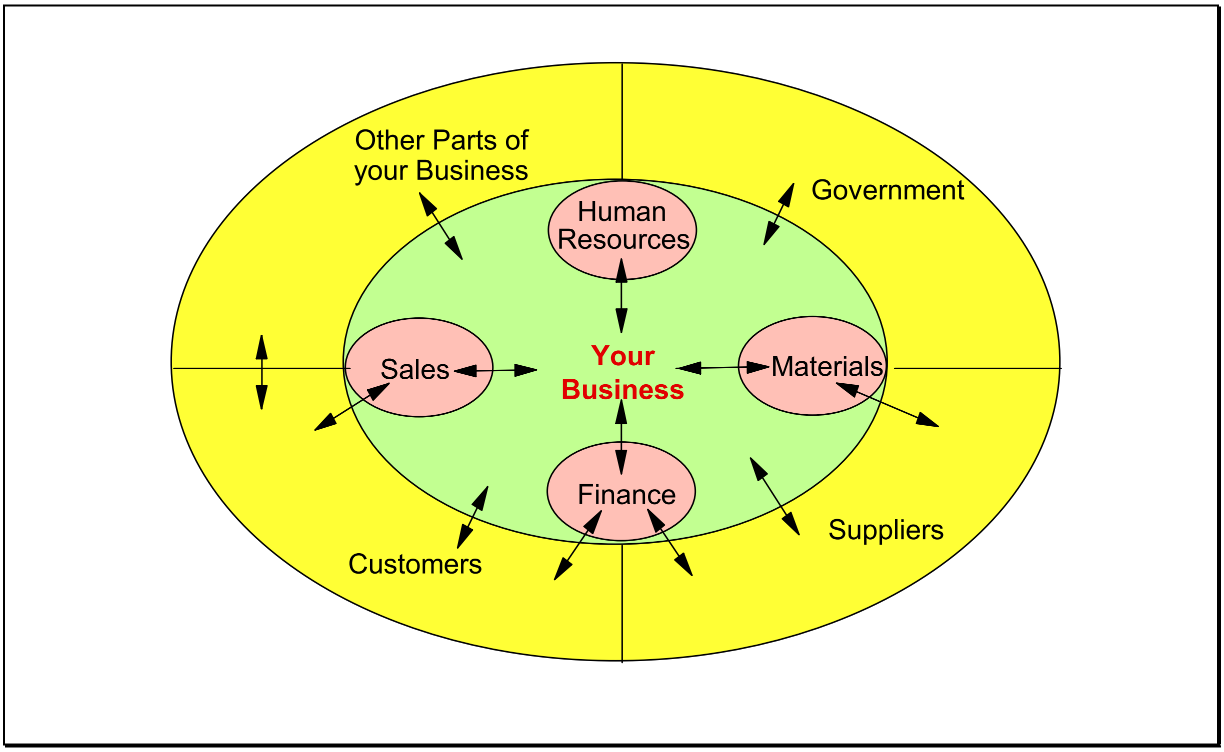

Figure 2: Sharing information.

Reconciling data becomes necessary when incompatible Reference Data are used, whilst business opportunities are lost from a failure to share data effectively.

The need to integrate business is the key driver for data integration and hence the development and management of Reference Data. Figure 3 below shows the three key dimensions to business integration.

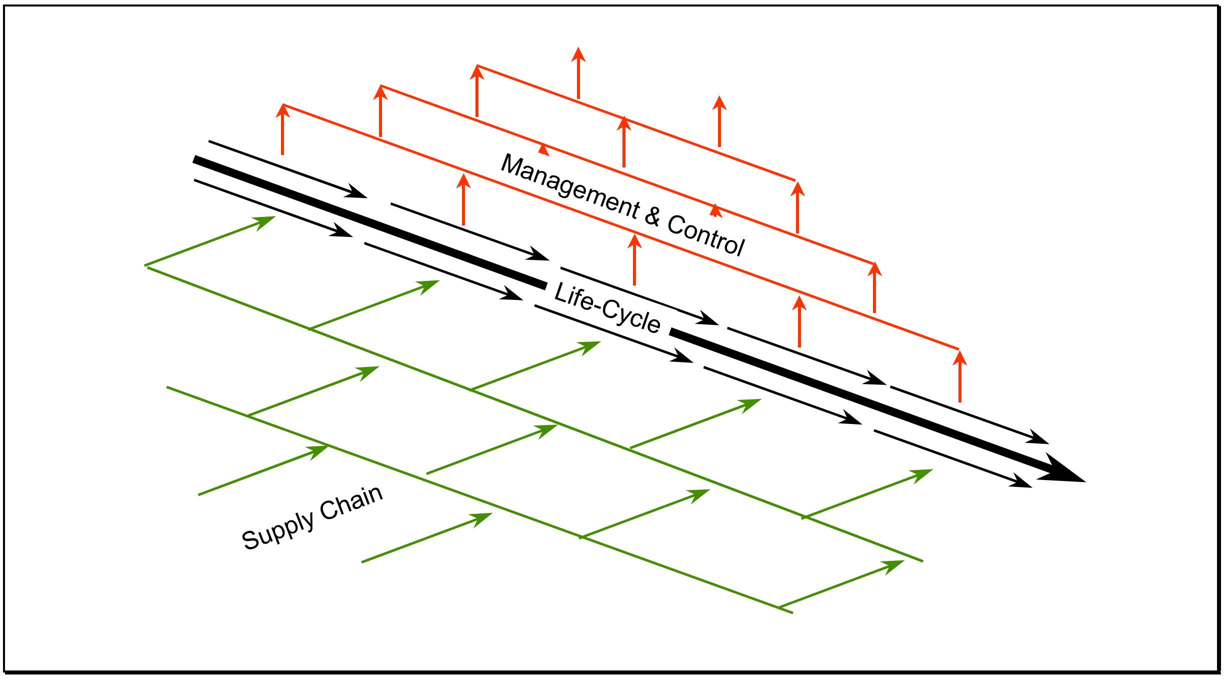

Figure 3: The dimensions to business integration.

Life-cycle:	All things (including activities) can be considered to have a life-cycle. The stages are: identify, specify, acquire, use/maintain, dispose. Information needs to be passed between these stages, which are often performed by different organisations.
Supply chain:	Activities have customers and suppliers. At least some of these are likely to be outside your direct control, either in different departments or external organisations. Business integration will require agreeing a common language so systems can talk to each other.
Management & Control:	Managing the performance of a business requires recording outcomes and Key Performance Indicators so that comparisons can be made, and best practice determined and implemented across the business.

In all cases business integration will require agreeing a common language so systems can talk to each other. This is defined through a data model and Reference Data. Sometimes this will require standardisation within an organisation, and other times it will require standardisation across different organisations or companies. The result of this data integration is that the exchange and integration of information can be automated. This reduces costs, and increases the speed of business, which leads to process intensification.

3. Standards for Reference Data

If Reference Data is a common language, the question is at what level is it defined. It is tempting to think that each company is responsible for its own reference data alone, but this does not address the need to communicate between companies for commercial purposes. Figure 4 shows who needs to define what kind of Reference Data for it to be effective in conducting business electronically.

Public Reference Data:	Concepts that are common across companies and are the basis of commerce. Examples include: Units of Measure, properties, generic products, currency and country codes.
Proprietary Reference Data:	Concepts that are legitimately defined by a company, but used by others as well as themselves. Examples include product model, and product serial numbers.
Private Reference Data:	Concepts that are defined and used only by one company. Examples are Key Performance Indicators.
Global Reference Data:	Concepts that are relevant on a global basis. Examples include Currency Codes and International Standard Products.
Regional Reference Data:	Concepts that are relevant within some region of the world. Examples include Value Added Tax, and the working days of the week, and Refinery Basic Data.
National Reference Data:	Concepts that are relevant within a country, such as reporting codes for public administration, local sales areas

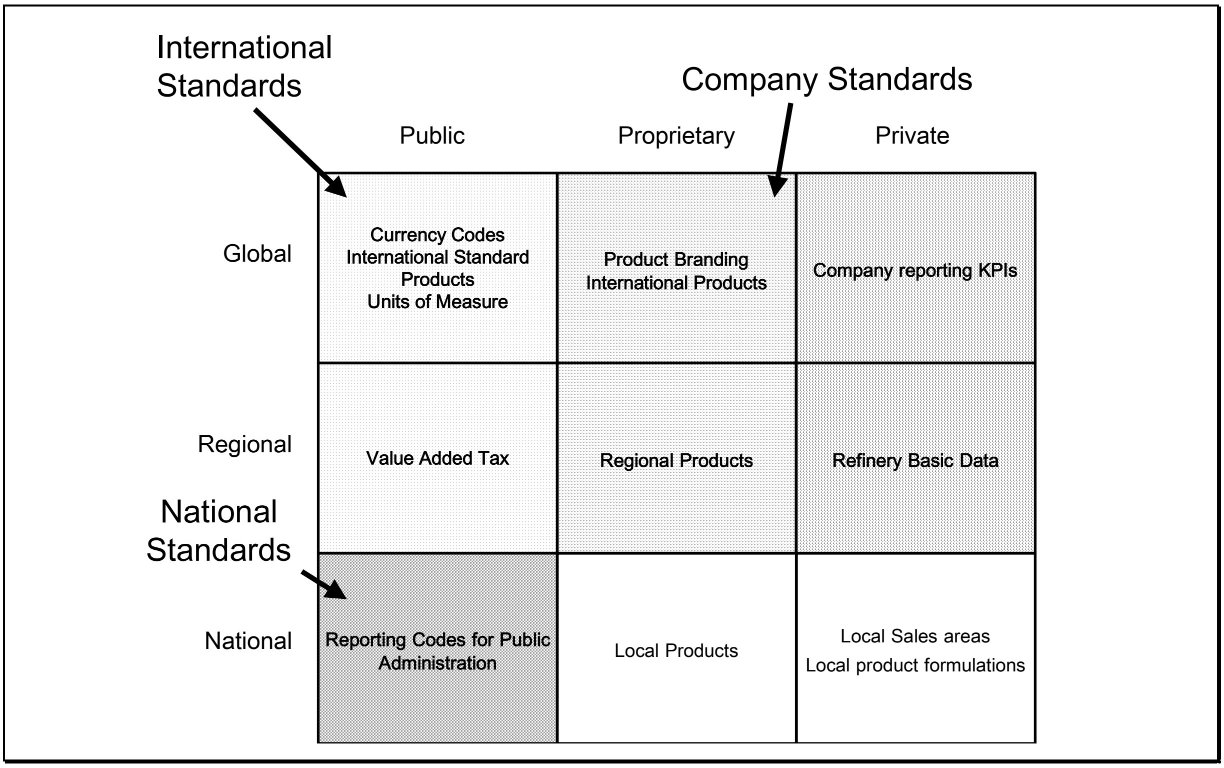

Figure 4: The levels for standards in Reference Data.

Figure 4 also shows the scope for standards at an international level, national level, and company level for Reference Data.

4. Using ISO 18876 and ISO 15926 to support Reference Data based Integration

Since we have shown that integration is a key business driver and that common Reference Data is essential to integration, both within and across companies, it is appropriate to address the issue through International Standards. Two standards that have been developed by the Process Industries specifically address this area:

- ISO 18876 – Integration of Industrial Data for Exchange Access and Sharing (IIDEAS) specifies a high-level integration architecture and methodology for integration. Part 1 [1] presents a general architecture for integration, whilst Part 2 [2] presents a methodology that can be adapted to a number of situations.

- ISO 15926 – Integration of lifecycle data for process plants including oil and gas production facilities is being developed to provide a series of parts that are a particular working out of ISO18876.

ISO18876 introduces a number of key concepts for integration models.

Integration model concepts

The concepts represented by an integration model can be classified as primitive, and derived concepts. Primitive concepts are the building blocks for the definition of other concepts, and can be further classified as foundation concepts, general concepts, or specific concepts.

This is represented in Figure 5 below.

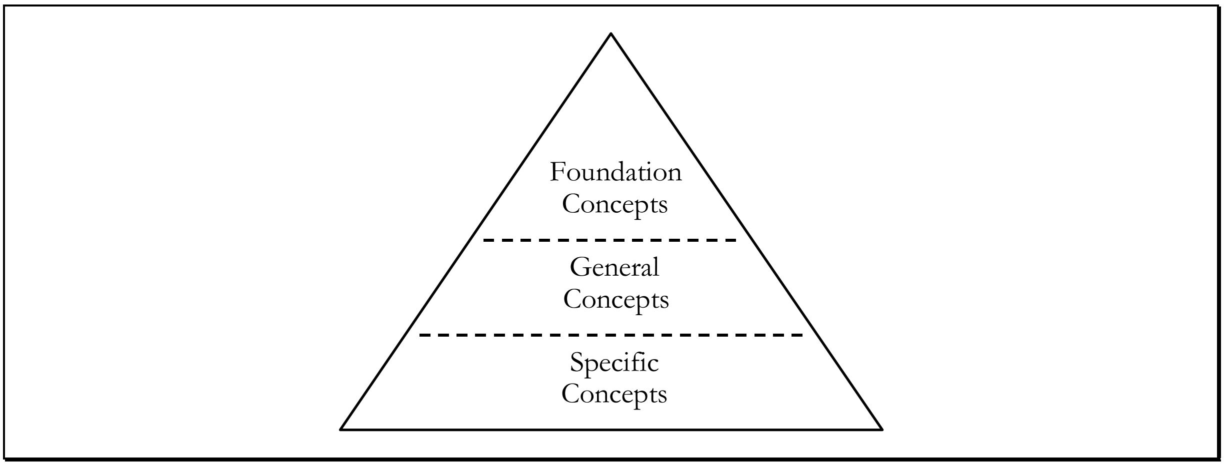

Figure 5: Primitive Concepts

Specific concepts are dependent on general concepts that are dependent on foundation concepts, since all the lower concepts rely on the existence of one or more higher-level concepts. For example, at the top level, an integration model should have foundation concepts like classification, connection and composition. General concepts might include those of physics. Discipline specific concepts would be those that are limited in their range of application.

A full integration model

A full integration model, as shown in Figure 6, is more than just primitive concepts; it includes derived concepts – useful and valid combinations of primitive concepts. Only derived concepts that are of interest need be recorded.

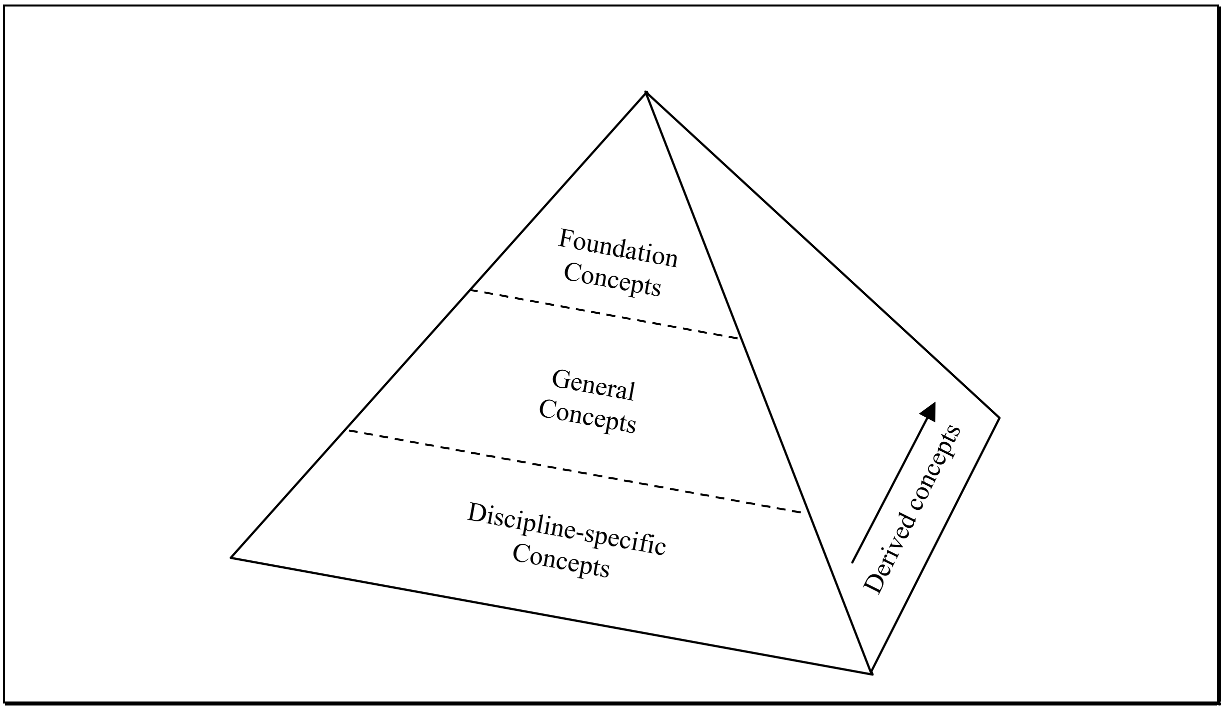

Figure 6: A full integration model

This architecture does not require that primitive concepts remain primitive forever. If a concept that is initially thought to be a primitive concept turns out not to be, then the concepts it is derived from can be identified/added, and the derivation added, so that it becomes a derived concept away from the front face of the pyramid. This allows flexibility to reflect an improved knowledge of the world, rather than reflecting knowledge of the world that is constrained by a modeller's knowledge at a point in time. Therefore, an integration model will need to be maintained and extended, and a mechanism for maintenance and extension will be necessary.

Overview of the model integration process

The model integration process takes a number of application models and an integration model. It ensures that all the concepts of the application models are represented in the integration model, and develops a mapping specification between the integration model and each of the application models.

There are three possible cases for the integration process:

- the integration model and the application models exist before the integration process starts;
- the application models to be integrated exist before the integration process starts, but not the integration model;
- the integration model exists before the integration process starts, but the application model needs to be developed from some statement of requirements.

The first of these covers all the elements of the other two, and is described here in outline for one application model. The other two processes are described in more detail in ISO18876-2.

Integrating an application model with an integration model is illustrated in Figure 7 below. The goal of this integration process is to allow the same information that is represented in the application models to be represented in the integration model without loss of meaning, and to allow transformations between these representations. The result of the integration process is a mapping specification between the application model and a part of the integration model. In order to define this mapping spec ification it may be necessary to extend the integration model so that it precisely represents the concepts found in the application model.

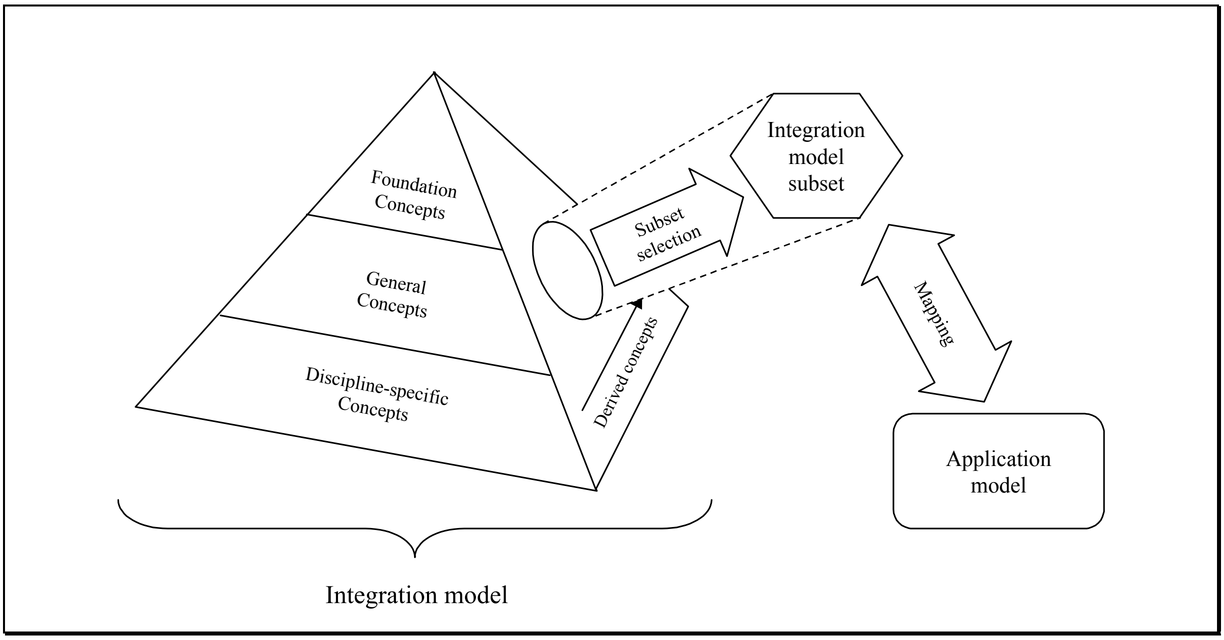

Figure 7: Integrating application models with an integration model

For a model, being an integration model derives from the role that it plays with respect to other models. The fundamental characteristic of an integration model is that it integrates two or more application models.

ISO 15926 – Integration of lifecycle data

ISO 15926 aims to implement the integration architecture of ISO 18876. The way that this is done is illustrated in Figure 8 below.

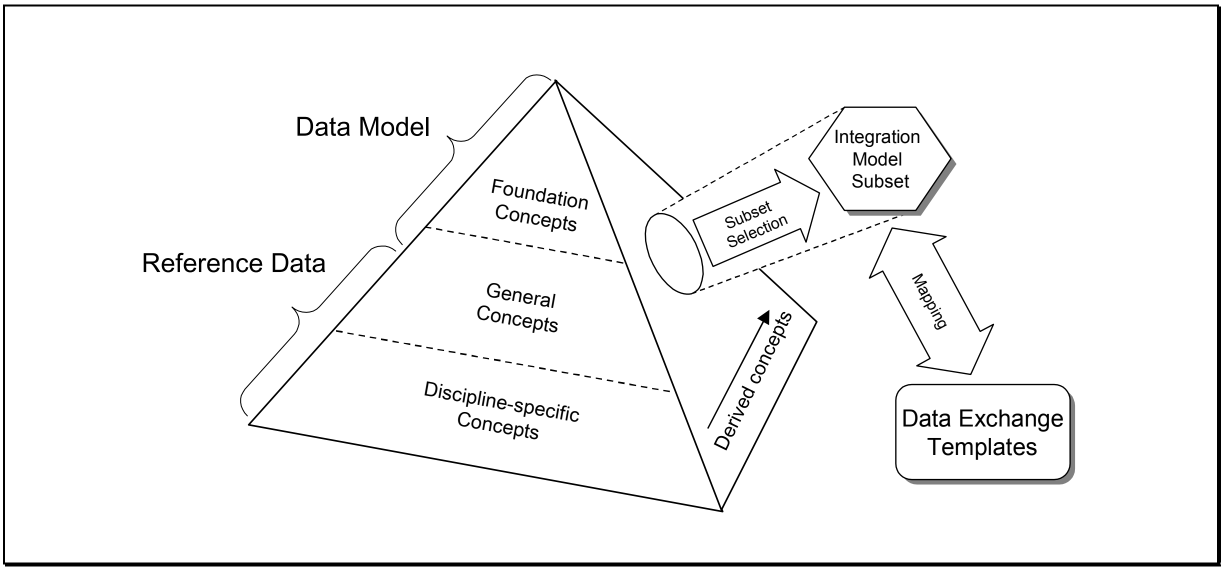

Figure 8: Mapping of ISO 15926 to the ISO18876 Integration Architecture.

The Integration Model is divided into two parts:

1. A data model that covers Foundation Concepts and some General Concepts. This is being standardised as ISO 15926-2 [3]. This consists of 201 concepts.

2. A Reference Data Library (RDL) that covers the remaining General Concepts, and Discipline Specific Concepts. This is being standardised as Parts 4, 5, and 6 of ISO 15926 where Part 4 is initial content for the Reference Data Library, and Parts 5 and 6 govern the administration of the Reference Data Library and the way additions are made. The RDL is planned to be implemented as an ISO Register, available on the internet.

As well as the Integration Model ISO 15926 provides for the definition of Data Exchange Templates that are the definition of data that is required to be exchanged for some specific purpose, such as an equipment data sheet, defined in terms of, and mapped to, the Integration Model, so that the data from a template can be integrated with data from other templates. This capability is being developed as Part 7 of ISO 15926.

5. Conclusions

Common Reference Data is critical to business integration. For much key Reference Data, it is desirable that it is defined through International Standards because its use is across companies to support commerce. For example for e-business to flourish we need electronic catalogues, but more than that we need these catalogues to be compatible so that product specification data from the catalogue can be brought into a design and merged with other design data, to discover, for example the total weight of an offshore platform. This cannot be done unless all the catalogues use the same coding for values, measures, and units of measure – for example. These are Reference data items, and if public reference data for these is not made available, product designers will be left with re-keying catalogue data, or building interfaces to each catalogue they use. This is potentially a major barrier to e-commerce.

ISO 18876 and ISO 15926 are two standards that are designed to support these requirements.

References

[1] ISO TS 18876-1:2003, *Industrial automation systems and integration — Integration of industrial data for exchange, access and sharing — Part 1: Architecture overview and description.*

[2] ISO TS 18876-2:2003, *Industrial automation systems and integration — Integration of industrial data for exchange, access and sharing — Part 2: Integration and mapping methodology.*

[3] ISO FDIS 15926-2, *Industrial automation systems and integration — Integration of life-cycle data for process plants including oil and gas production facilities — Part 2: Data model.*