

Producing the Integrated Data Model

An overview of the information required for understanding the production of an integrated model of an organisations data in all its forms and how to approach the solution.

Adrian Miley

(Adrian.Miley@MileyWatts.Com)

Overview

1. **Scoping The Problem** – Establishing why the Integrated Data model is to be produced
2. **The Integrated Data Model** – Categorising the many data models that will already exist.
3. **Defining the Types of Models**
4. **Things That Need To Be Recorded**
5. **Model of a Domain Model** – Describes the meta-model of a data model.
6. **Completeness – Removing Optionality**
7. **Completeness – Things That Must Be True**
8. **Approaching the Problem** outlines a possible approach to making the problem manageable with staged delivery.

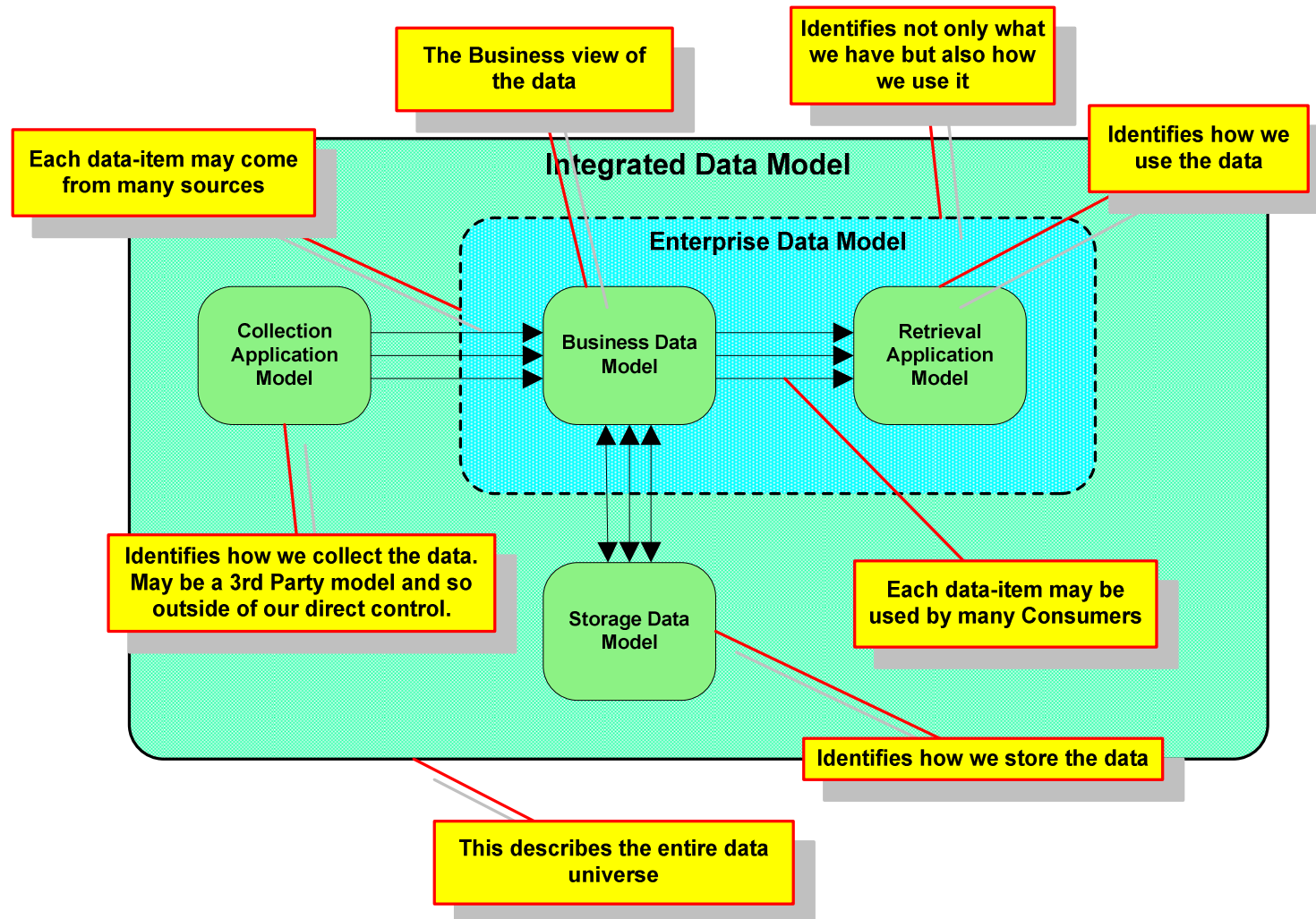
Scoping The Problem

Identifying the purpose of the Integrated Data Model is essential. Some scoping questions are:

- Why is the Data Model being produced?
 - Without a commercial purpose the company will eventually lose the willpower to continue. Success criteria = sponsorship.
- Is it the Business Content or the Company or both that is being modelled?
 - The Business Content Data is about what the company does;
 - The Company Information is about how it operates.
- Is it for external or internal consumption?
 - External consumption has legal implications for the content of the model.
- Is a Data Dictionary or Data Encyclopedia being created?
 - A Data Dictionary just documents terms
 - A Data Encyclopedia also provides indexing and context.
- Is the model describing behaviour or just characteristics?

Whilst these questions are unanswered the analysis should not be started.

The Integrated Data Model



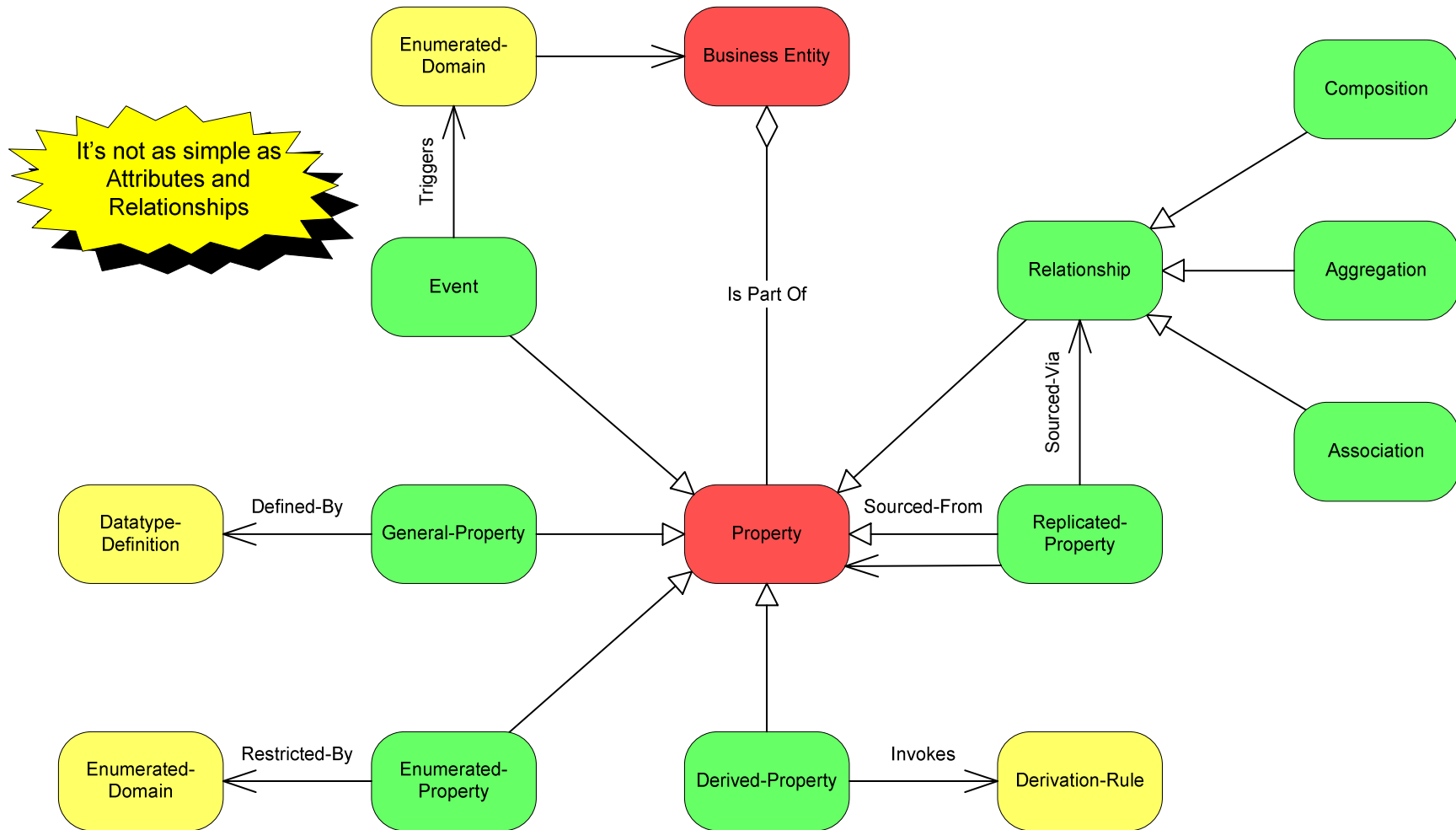
Defining The Types Of Models

- **Business Data Model** describes the data using business terms and semantics. A **Business Entity** is the artefact within a **Business Information Model**.
- **Storage Model** describes the private representation of the data as it exists in persistent storage e.g. database. This is the **Storage Form**.
- **Collection Model** describes the structure of the data as it is received. Each feed may maintain only a fragment of a Business Entity. This is the **Collection Form**
- **Retrieval Model** describes the data as it is used by a consuming application. All Retrieval Forms are read-only views of the Business Entity they are based on. This is the **Retrieval Form**.
- The **Enterprise Data Model** combines the Business Information Model and the Retrieval Data Models to provide a view of what the data is and where it is used.
- All the models together form the **Integrated Data Model** that describes the entire corporate universe coherently.

What Information Should Be Recorded?

- There is a lot of information that could be recorded against each data-item. A **Domain Model Meta-Model** is essential.
- The minimum set of mandatory attributes should be documented and enforced. These would generally include any information about:
 - Data-type definition
 - Data validation rules such as enumerated domains, value ranges etc.
 - Update restrictions
 - Help-Text and Usage Notes
- Who is responsible for each part of the model (the Dublin Core data-set).
- Any significant events / processes that operate on the data.

Model Of A Domain Model



Completeness – Removing Optionality

Completeness never happens – there is always more to do.

- All data-items are mandatory at some time during the life of a Business Entity. So by the end of the entities natural life all attributes must be populated. This is the **Entity Life History**.
- It does not mean that all attributes are mandatory at the point of creation. This is a common misconception.
- If a data-item is not populated by the end of the entity's life then one of the following must be true:
 - There is at least one more inbound process to analyze
 - The Business Requirements have been incorrectly stated and the data-item is for information only.
 - There is an additional level of sub-type analysis required in the data model. One sub-type will have the data-item as mandatory and the other will not possess it at all.

Until optionality is removed the analysis is incomplete.

Completeness – Things That Must Be True

All Data-Items in the Integrated Data Model must have:

- At least one source in a Collection Model. This establishes how we acquire the information. Even Derived Information is still sourced from somewhere.
- At least one consumer in a Retrieval Model. This establishes why we require the information.
- At least one place in a Storage Model where the data-item, or its underlying constituents, is stored. This establishes how we manage the data.
- If more than one point of storage then one and only one storage location must be identified as the **Database of Record** for all instances of that data-item. This establishes trust.

Until all these things are true the analysis is incomplete.

Approaching The Problem

- Establish the data modelling notation to be used and what each symbol means. Even the standard notations are ambiguous.
- Decide in advance on the success criteria and stop as soon as it is achieved.
- Produce the Integrated Data model in the following order:
 - Identify the main entry points into the content (the **Primary Business Entities**) and start with these.
 - Databases of Record – identifying these is essential to establishing completeness of Business Entities within scope.
 - New Retrieval Applications – these are the future of the company or business.
 - Legacy Applications – these tend to be the big revenue generators so will be fixed and long-lasting.
 - Data Collection Applications
- Add detail as it is required. Analysis Paralysis sets in if we try to achieve perfection at the first attempt rather than perfect the content over time.