

Data Architecture

Introduction

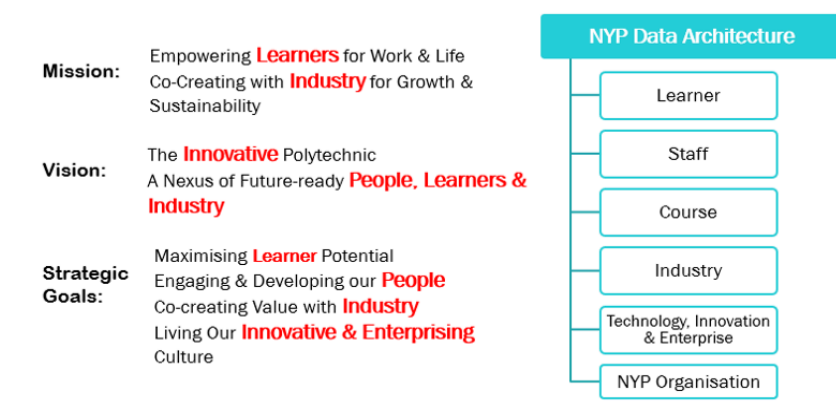
Data Architecture can be considered as a key enabler in the NYP Data Management Framework, making it possible for NYP to do more for its key stakeholders, with improved outcomes.

An essential consideration in developing data architecture of an organisation must be on how it can be used to effectively and efficiently represent the core business processes, where data is collected, created, processed, used, distributed, stored and disposed.

The importance of data architecture resembles that of building architecture which shows the layout of building boundaries, the water piping, electrical power supply, the lighting and the HVAC systems. It provides critical information on the what, where, who, when and how of data in an organisation.

A reasonable approach its alignment with NYP's Vision and Mission.

Alignment with NYP Vision and Mission



Enterprise Report Data

As part of its operations, NYP communicates its progress by tracking and reporting specific data at various channels and platforms. Such data need to be clearly and consistently defined to facilitate a common vocabulary in meetings and discussions.

Model Structure

There are many different definitions of data architecture with each description based on one's understanding and interpretation. The most succinct and all-encompassing version is given by The Open Group Architecture Framework (TOGAF) which says that data architecture is 'composed of models, policies, rules or standards that govern which data is collected, and how it is stored, arranged, integrated, and put to use in data systems and in organizations.'

With the amount of data generated online increasing by the day as a result of technological advances in internet connectivity, electronic storage and computing power, business leaders with wisdom and foresight will do well to ensure that their critical asset, data, is managed in a structured and organised manner, according to the definition of data architecture. In the simplest terms, the primary benefit of data architecture is this: it provides a systematic and well-defined way of managing organisational data assets.

A 4-level structure was adopted in our hierarchically arranged 'Categorical Model':

1. Subject Area - This is the core data of the organisation, NYP's life and blood.
2. Aspect - Perspectives or views of functional areas within each subject area whose names bear close resemblance to business processes. It also serves as containers to group lower level data into an organised structure.
3. Category - This is the actual data object in the data model.
4. Sub-category - This level represents the attributes of the data object – the raw values that determine the characteristics of the object

Data models at different levels can be used for different purposes. Generally, levels 1 and 2 are typically used for strategic and business discussions. It is possible that business strategy discussion could make it necessary to identify new data that does not exist, to facilitate new initiatives, or the improvement of existing delivery outcomes and services. For such cases, a broad overview of our operational processes and the associated data would be more useful.

On the other hand, application development and data infrastructure teams would require details about the data types, formats, size, volume etc to facilitate their technical implementation work. For these, data models at levels 3 and 4 would be necessary.

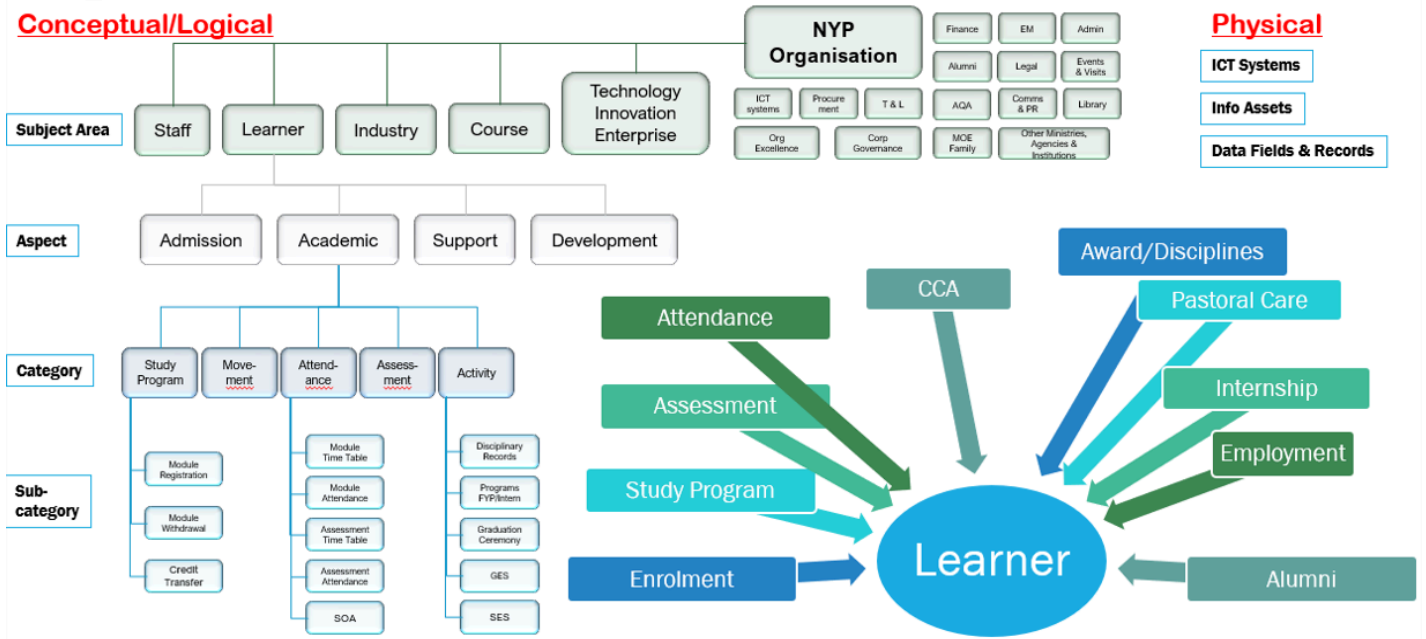
Conceptual/Logical Model - an overview

The figure below showed an overall view of the conceptual model, and the hierarchical expansion of the Subject Area of 'Learner'.

3

DATA ARCHITECTURE – REPRESENTING NYP'S CORE BUSINESS PROCESSES

Conceptual/Logical



Modelling Process

Six Subject Areas were first identified, based on their impact, resource constraints and knowledge of the processes. Other Subject Areas will be developed along the way.

The modelling process would typically involve the data ownership team, including the management, subject matter experts and data stewards. A typical process would include:

1. Data modelling team created a simple data model for the Subject Area, based on available information;
2. Co-development of the data model with the data stewards and subject matter experts. This is a discovery process where data stewards were encouraged to verbalise the process details as well as data entities involved. This would enhance the acceptance level of the eventual model;
3. A fine-tuning process with the data ownership team, followed by their endorsement. This process would enable the management team to look at the model from the logical and communication perspective. This would be useful from harmonisation of vocabulary, roles and responsibility perspective.

Work has been done on the following data models (of varying levels of details and completeness):

1. HR
2. Learner
3. Course
4. Estate Management
5. Procurement
6. Administration