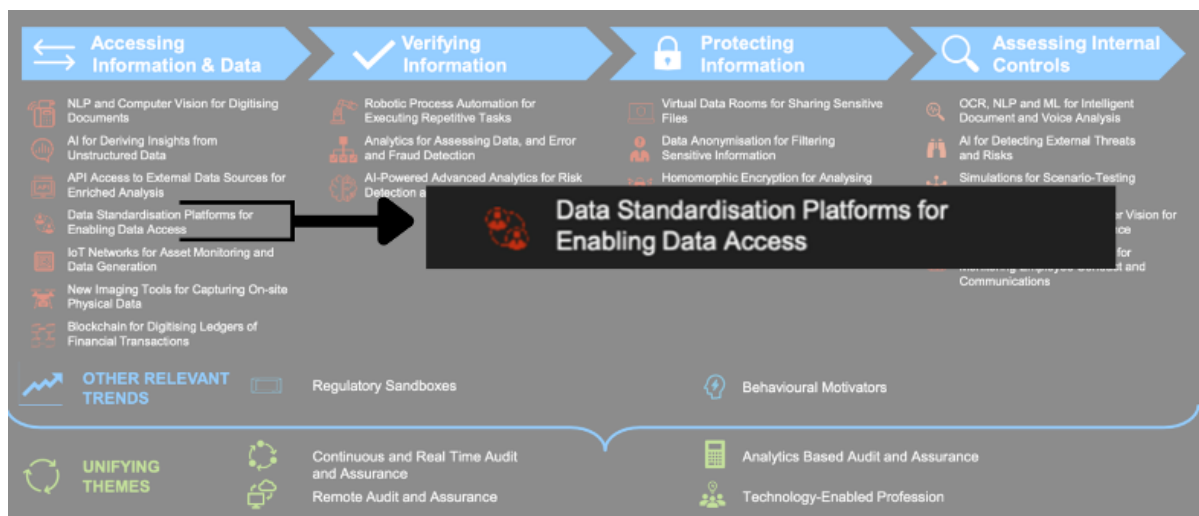


# IAASB DIGITAL TECHNOLOGY MARKET SCAN: DATA STANDARDIZATION

Oct 27, 2021 | English

Welcome to the first market scan prepared by the IAASB's Disruptive Technology team. Building on our previous work, which included the [Innovation Report](#) created with [Founders Intelligence](#) and discussed at the January 2021 IAASB Meeting, we will bring you a regular market scan focusing on various topics from the report around every two months. The market scans will consist of exciting trends in the area, including interesting developments on this topic, what this might mean for the IAASB, corporate and start-up innovation, and noteworthy investments.

In this market scan, we will explore **Data Standardization Platforms for Enabling Data Access**, which falls under the activity of **Accessing Information & Data**. We're starting with Data Standardization because establishing a common standard of how data is structured and accessed is a foundation block to the success and widespread adoption of other innovative technologies.



We will cover:

- What is Data Standardization and why it is important?
- What some of the latest exciting developments on this topic are, including the increasing maturity of Common Data Models and Knowledge Graphs.
- What this could mean for the IAASB.

## What is Data Standardization and why is it important?

**Data Standardization** is the process of converting data to a common format that allows users to better analyze and utilize the data, thereby enabling data collaboration, large-scale analytics and the use of more advanced tools to interrogate the data.

With exponential growth in the amount and variety of data that companies create and use, there is voluminous unstructured, or inconsistently structured, data in companies' repositories. This leads to data silos and data that is underutilized or unnecessarily hard to access.

One of the problems this has created, amongst others, is where auditors are spending more time on data management, particularly when trying to access, "map," and use the entity's data as well as when performing data analytics. When each entity uses different data models and systems or platforms to store, structure and extract data, the inefficiency is amplified. Some firms are developing internal tools to address these challenges, such as by building [ETL \(Extract, Transform, Load\) tools](#) to minimize this inefficiency e.g., [KPMG had 25 different ETL projects in 2019](#). Auditor's time spent on data management may be better utilized on other areas of the audit, and many firms and practitioners may have challenges with obtaining tools to access, manage and evaluate data relevant for auditing and assurance engagements.

The data management industry, firms, and regulators are exploring various approaches to help audit and assurance professionals and other professional service providers with these challenges. [The UK's Brydon review in 2019](#), for example, recommends initiatives to develop a standard method of data extraction covering both structured and unstructured data.

Data standardization complements this approach by helping to address the root cause of difficulties by converting the data to be extracted into a common format. In particular, digital multi-party platforms are gaining traction as a solution to provide standardized data structure and mechanisms to access data silos with non-uniform formats, thereby facilitating the sharing of data to unlock new values both internally and externally. One exciting development is the development of Common Data Models (CDM). A CDM is a shared data language, allowing standardized metadata and its meaning to be shared across applications easily. Unfortunately, at present, there is no one model adopted globally across specific industries or jurisdictions.

## Recent Noteworthy Developments in Data Standardization

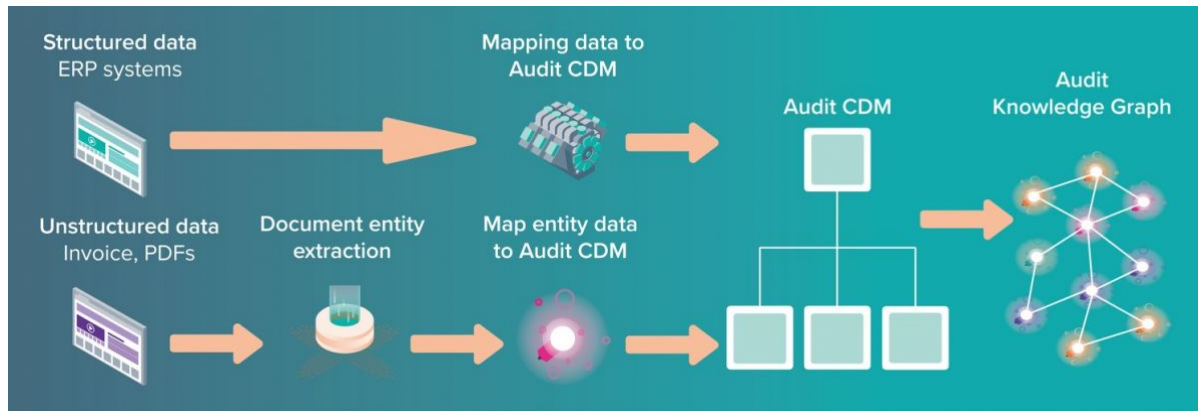
*This section is designed to provide examples of recent developments that may signal future disruption in this area. It is not a complete list of all activities in the field of data standardization.*

### 1. Disruptive start-ups are gaining traction

#### 1. [Engine B](#) receives new financial and board-level investment

- Institute of Chartered Accountants in England and Wales (ICAEW) upped its investment in [Engine B](#) to 10% and has taken a board seat.

- Engine B is partnered with key organizations, including Microsoft and ICAEW, to create an [Audit Common Data Model](#).
- Part of this project involves creating an Intelligent Data Access Platform designed to be installed in a client environment and ingest corporate data, both structured and unstructured, to map it on an [Audit CDM](#). This platform attempts to replace the need for complex ETL tools in favor of open data standards that facilitate the sharing of clean standardized data.



- Working with 13 audit firms, Engine B aims to roll out its assets in late 2021, first in the UK, then the US, as it collaborates with the AICPA. It aims to become a widely adopted infrastructure like Open Banking globally.

### Other Data Standard Initiatives

Data standards exist in various forms already. Another example is [ISO 21378, Audit Data Collection](#), issued by the International Organization for Standardization (ISO). This standard leveraged the American Institute of CPA American Institute of Certified Public Accountants Audit Data Standards.

### Key Venture Capital & Investment Terms

Venture capital (VC) is a form of financing where capital is invested into a company, usually a start-up or small business, in exchange for equity in the company. VC funding stages can be useful signals for the maturity of the start-up and its products or services – the later the stage, the more developed and established in market the startup typically is. VC funding can also be a useful barometer for the interest in a particular technology and how influential it may be in the market as well as an indicator of potential wider adoption.

For simplicity it can be useful to group start-up funding stages into the below broad categories:

- **Pre-seed and Seed:** start-ups in these stages are very early-stage start-ups, often prior to launching a product in market or with a few initial customers. They are typically seeking investments from various types of investors (e.g., individuals as well as firms) to establish themselves, build out the product, hire the core team and acquire more of their initial customer base.
- **Series A and Series B:** start-ups in these stages are in growth mode, usually with a product/service that is market-ready and launched, with some revenue being generated. They are usually looking for funding to fuel the continued growth of the start-up.
- **Series C and beyond:** start-ups in these stages are more mature, typically with products/services in market that have strong demand and likely have solid revenues and profits. Series C can be the last stage of VC financing (e.g., before an IPO) however many companies opt to raise more VC rounds such as Series D, E, etc. Funding at this stage is likely to be used to scale up operations and continue growth through entering new markets, R&D or making acquisitions.

For a deeper dive see [Venture Capital Jargon Buster](#) by Founders Intelligence and MJ Hudson.

1. InfoSum's 'non-movement of data' technology enables companies to connect their data (both internally and externally) to unlock new customer value. This works by having companies standardize their data (i.e., map their data) according to InfoSum's Global Schema rules and upload it to InfoSum's platform. **InfoSum raised a further \$65m in their Series B to scale its privacy-focused data collaboration platform:**

What does this mean? It means investors see an opportunity for data standardization when companies want to [collaborate and share data without moving data outside their companies](#). It also signals the growing maturity of the data collaboration space as InfoSum is a leading start-up and is raising large sums of money to upscale its operations.

## 2. Relevant industry players are taking more interest in data standardization

### I. A leading US accounting firm custom-built a common data model

- Besides a growing industry consortium for Engine B, a US accounting firm has partnered with [Orion Innovation](#) to build a CDM that enables uniformity in the data from hundreds of different ERP systems and technology platforms
- The CDM made the data more understandable and useful to all its business activities, particularly auditing, data analytics, and advisory. Apart from unlocking advanced data analytics on the full population of data, it also unlocked automation options, which may provide greater consistency of audit quality.
- See the case study [here](#) for more detail.

## II. The **EDM Council**, a global association created to elevate the practice of Data Management, is leading the development of an open-source semantic data standard

- The EDM Council has published the [Data Management Capability Assessment Model \(DCAM\)](#). DCAM defines the scope of capabilities required for an entity to establish, enable and sustain a mature Data Management discipline. It addresses the strategies, organizational structures, technology and operational best practices needed to drive Data Management across the organization, and ensures the data can support digital transformation, advanced analytics such as artificial intelligence and machine learning, and data ethics.
- The EDM Council partnered with CPA Canada to provide an overview of how DCAM can be leveraged for audit and business controls. A [recording is available](#).
- Since 2020, the EDM Council has been leading the Financial Industry Business Ontology ([FIBO](#)) initiative, which provides descriptions of the structure and contractual obligations of financial instruments and financial processes, to give meaning to the data.
- The fundamental aim of the standard is to harmonize data across disparate repositories to validate data quality and improve risk analysis by making links between datasets that are understandable to both humans and software, i.e., resolve data silos.

## 3. Knowledge graphs for audit use cases is showing promising progress

### I. Engine B's audit knowledge graph hopes to improve the quality of audits

- [Knowledge graphs](#) developed by Engine B provide contextual relevance of data by looking at the relationships between all data elements (both structured and unstructured), which allow auditors to make context-driven decisions. In particular, they are looking at anomaly detection and fraud detection as their initial use cases.
- These knowledge graphs can sit on top of their Audit CDM to perform visual and contextual data analysis on all relevant transactions. Here is an [explainer video](#) from the developers.

## II. The EDM Council is creating an Open Knowledge Graph Lab (OKGL) on top of their FIBO initiative

- Since late 2020, EDM Council has been developing [OKGL](#) as the infrastructure of knowledge graphs for application across different sectors. Particularly for the financial services sector, the EDM Council is exploring use cases in fraud, risk and anti-money laundering. The EDM Council is also currently preparing the rollout of a cloud sandbox to serve as a testbed to develop prototypes.

### What might this mean for the IAASB?

The IAASB has an interest in improving the data available to assurance practitioners as this may enable the performance of more advanced analytics and otherwise improve areas of the audit that use data (e.g., evaluating models and related controls). Data standardization also enables collaboration between the entity and others, including auditors or assurance practitioners. Data standards are of particular interest in the sustainability space as a tool for entities to satisfy different reporting standards.

Data standardization is not within the IAASB's remit because it is fundamentally a matter for how the entity manages its data. Local law or regulation requiring the maintenance of books and records may be most relevant. However, because of the benefits to audit and assurance quality, the IAASB should stay close to the topic and take opportunities to raise it with other stakeholders, such as regulators, preparers, and assurance practitioners.

While the developments on data standardization are promising, there is still a considerable way to go before it is widely adopted by entities and therefore able to fully benefit audit and assurance. In the absence of a widely adopted CDM or another method to standardize data, the gap in data management capabilities between differently equipped firms and practitioners, including between jurisdictions, may grow. Furthermore, widespread adoption of innovative automated audit tools and techniques will be inhibited when data is not structured in a standard format.

As prominent CDMs are more widely adopted and supported by entities and made available to firms, there may be a need for standards on assurance services on whether data is compliant with the relevant data standard.

### What do you think about this bulletin?

Please take the time to fill out our [quick survey](#) to let us know your thoughts about this bulletin, how it can be improved and what you would like to hear about going forward.

### What next?

Our next Market Scan bulletin will be distributed by January 2022.

