

財團法人電信技術中心出國報告

出席英國 2023 年 歐洲資料治理暨主資料管理會議 出國報告書

單位名稱：財團法人電信技術中心

姓名職稱：劉宜菁 研究員

辜厚僑 副研究員

派赴國家：英國倫敦

出國期間：2023 年 5 月 8 日至 2023 年 5 月 18 日止

報告日期：2023 年 7 月 3 日

摘要

2023 年歐洲資料治理暨主資料管理會議（Master Data Management & Data Governance conference Europe 2023 London）由英國資料治理及商業智慧訓練公司 IRM UK 舉辦，會議自 2023 年 5 月 9 日開始至 5 月 12 日結束，為期 4 天，聚集英國銀行業者、國際應用生物科學中心 CABI、顧問公司（如 PwC 資誠、Bearingpoint、QuaData、MetadataWorks）等 42 位具資料治理經驗之領域專家共同參與，舉辦共 38 場次專題研討及 2 天之工作坊課程。本次會議中，聚焦於資料治理之現況發展以及未來趨勢推動，並邀請具豐富資料治理經驗之講者，分享資料治理之成功案例及經驗，透過本次研討會之參與，期蒐集國際資料治理發展最新動態，包含了解資料治理之方法與工具、成功案例之經驗蒐集、資料合規之運作，並建立與國際專家之鏈結。

目錄

| | |
|---------------------|----|
| 壹、 目的..... | 4 |
| 貳、 行程..... | 5 |
| 參、 會議過程及內容 | 7 |
| 肆、 心得及建議 | 55 |
| 伍、 附件：相關照片及資料 | 58 |

壹、目的

本中心為執行數位發展部「建構隱私強化技術與數據公益合規機制計畫」補助計畫案，了解國際在資料治理之發展趨勢，由劉宜蕎研究員及辜厚僑副研究員出席英國倫敦，參與由英國資料治理及商業智慧訓練公司 IRM UK¹舉辦之 2023 年歐洲資料治理暨主資料管理會議（Master Data Management & Data Governance conference Europe 2023 London）。

本次會議自 5 月 9 日開始至 5 月 12 日結束，為期 4 天，聚集英國銀行業者、國際應用生物科學中心 CABI、顧問公司（如 PwC 資誠、Bearingpoint、QuaData、MetadataWorks）等 42 位具資料治理經驗之領域專家，舉辦共 38 場次專題研討及 2 天之工作坊課程。

透過本次研討會之參與，期蒐集國際資料治理發展最新動態，包含了解資料治理之方法與工具、成功案例之經驗蒐集、資料合規之運作，並建立與國際專家之鏈結，期使後續執行本計畫之研究時，能針對資料政策法制及數據公益之應用發展，進行更深入之探討。

¹ IRM UK 為英國當地組織，專為企業裡的 IT 人員、決策者提供專業 IT 顧問，並提供課程或企業內訓課程，邀請適合之專家，協助企業分析技術或管理問題，並透過實體或網路研討會，讓企業組織代表與 IT 顧問得以直接交流，討論技術及管理所面臨之問題。

貳、行程

表 1：出國行程表

| 日期 | 行程 |
|----------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2023 年 5 月 8 日 09:35~16:30 | 差旅時間：臺灣桃園國際機場至倫敦希斯洛機場(搭乘中華航空 CI81) |
| 2023 年 5 月 9 至 12 日 9:00-18:00 | 參加 Master Data Management & Data Governance Conference Europe 四天議程 主要參與議題： <ul style="list-style-type: none"> • 歐盟資料治理經驗分享 • 資料治理之案例分享 • 資料治理之架構 • 資料治理及人工智慧之未來展望 |
| 2023 年 5 月 13 日至 15 日 | 陪同多元司長官參訪行程 |
| 2023 年 5 月 16 日 | 拜訪 TTC 曾合作之顧問公司—PLUM (私人行程) |
| 2023 年 5 月 17 日 22:00 - 5 月 18 日 18:30 | 差旅時間：倫敦希斯洛機場至臺灣桃園國際機場(搭乘中華航空 CI82) |

表 2：會議參與進程

| 2023 年 5 月 9 日 | |
|----------------|-----------------------------------------------------------------------------------------------------------|
| 時間 | 主題 |
| 09:10 - 10:05 | 開幕全體大會演講：準備好應用生成資料價值—DataGPT 了嗎？ (Plenary Keynote : Ready to apply Generative Data Value, DataGPT ?) |
| 10:05 - 10:20 | 資料治理頒獎 (Data Governance Awards) |
| 10:20 - 10:50 | 交流茶敘 |
| 10:50 - 11:40 | 釋放資料的力量 (Daring to Release the Power of Data) |
| 11:50 - 12:30 | 設計可執行的資料治理組織 (Designing an Actionable Data Governance Organization) |
| 12:30 - 14:00 | 午餐 |
| 14:05 - 14:45 | MDM (Master Data Management) 的未來 (The Future of MDM) |

| | |
|---------------|--------------------------------------------------------------------------------------------------------------------|
| 14:50 - 15:30 | 如何將資料治理的優勢與靈活資料治理結合？（How Can you Combine the Strength of Data Governance with the Flexibility and Speed of Agile？） |
| 15:30 - 16:00 | 交流茶敘 |
| 16:00 - 16:40 | 資料治理的變革（The Changing Face of Data Governance） |
| 16:50 - 17:30 | 閉幕演講：海盜運動（Be More Pirate） |

| 2023年5月10日 | |
|---------------|---------------------------------------------------------------------------------------------------------------------------------------|
| 時間 | 主題 |
| 0900 - 9:50 | 全體大會演講：資料可觀察性－人工智慧驅動資料治理之及時性與自動化（Plenary Keynote: Data Observability - A Catalyst for AI-Driven Real-time Data Governance Automation） |
| 10:20 - 11:00 | 利用 FAIR 提升資料治理效能（Leveraging FAIR for Better Data Governance） |
| 11:10 - 11:50 | 評估資料治理對於英國國家統計局的影響力（Measuring Data Impact with the Office for National Statistics） |
| 11:50 - 13:20 | 午餐 |
| 13:20 - 14:10 | 透過資料治理及參考資料促進環境政策執行與執法（Enabling Environmental Policy Execution And Enforcement Through Data Governance And Reference Data） |
| 14:10 - 15:00 | 讓資料發揮效果：如何不破壞您的資料治理計劃（Keynote: Making Data Work - How Not to Wreck Your Data Governance Programme） |
| 15:30 - 16:10 | 閉幕演講：資料及人工智慧將會為未來帶來哪些啟發？（Plenary Keynote: What the Future Will Bring Based on Data and Artificial Intelligence） |

| 2023年5月11日 | |
|---------------|----------------------------------------------------------------------------------------------------------|
| 時間 | 主題 |
| 09:00 - 12:15 | 工作坊一：資料網格及資料經緯－管理資料治理和資料品質（Data Mesh and Data Fabric - Managing Data Governance and Information Quality） |
| 13:30 - 16:45 | 工作坊二：元資料品質：忽略你的風險（Metadata Quality: Ignore at Your Own Risk） |
| 2023年5月12日 | |
| 時間 | 主題 |
| 0900 - 16:30 | 工作坊三：資料管理完全指南（Complete Guide to Data Stewardship） |

參、會議過程及內容

一、 第一天會議重點

(一) 開幕全體大會演講：準備好應用生成資料價值－DataGPT 了嗎？

(Plenary Keynote：Ready to apply Generative Data Value, DataGPT？)

為四天活動帶來第一場演講的是來自 Bearingpoint 顧問公司的 Jan Hendreyckx。Bearingpoint 是國際型商業顧問公司，主要業務包含商業策略與科技顧問、商業軟體產品及雲端增值服務、投資及企業併購分析服務。本場次的演講者 Jan 為 Bearingpoint 的比利時區域合夥人，專長包含資料治理及資料策略、資料架構及策略、資料品質、商業智慧及分析、資料管理等。曾擔任非營利組織國際資料治理協會 (Data Management Association International) 之理事。

在本場次的演講之中，Jan 主要圍繞在生成型預訓練變換模型 (Generative Pre-trained Transformer, GPT) 對資料應用、商業應用可能的影響，以及將來可能呈現的整體管制環境。

GPT (或稱 AI) 科技的出現，使得私人企業將資料 (Data) 轉換為營收的可能性浮現。過去，私人企業通常將資料當作是提供商業產品/服務之餘的副產品，以便利商店為例，商品的銷售為便利商店的主要收益來源，而各商品銷售的數量、購買客群等資料，是因為銷售而產生的副產品。這些副產品可以在經過資料探勘之後，做為調整未來銷售行為的依據，但整體而言，收益主要仍來自產品銷售，而資料僅用於輔助產品銷售。

在 GPT 技術逐漸成熟的情況下，資料可以轉換為資料產品，進而為公司爭取收益。Jan 認為，私人企業應該對 GPT 技術或甚至 AI 技術的使用有更多的了解，而這其中的關鍵在於，如何使得 GPT 模型更有效率。

Jan 於演講中提出三個使得 GPT 使用更有效率的方法，分別為：1. 以正確的資料集訓練模型、2. 問對的問題、以及 3. 確保獲得的答案不僅在統計上有相關性，並且對所追求的目的而言是精確的。

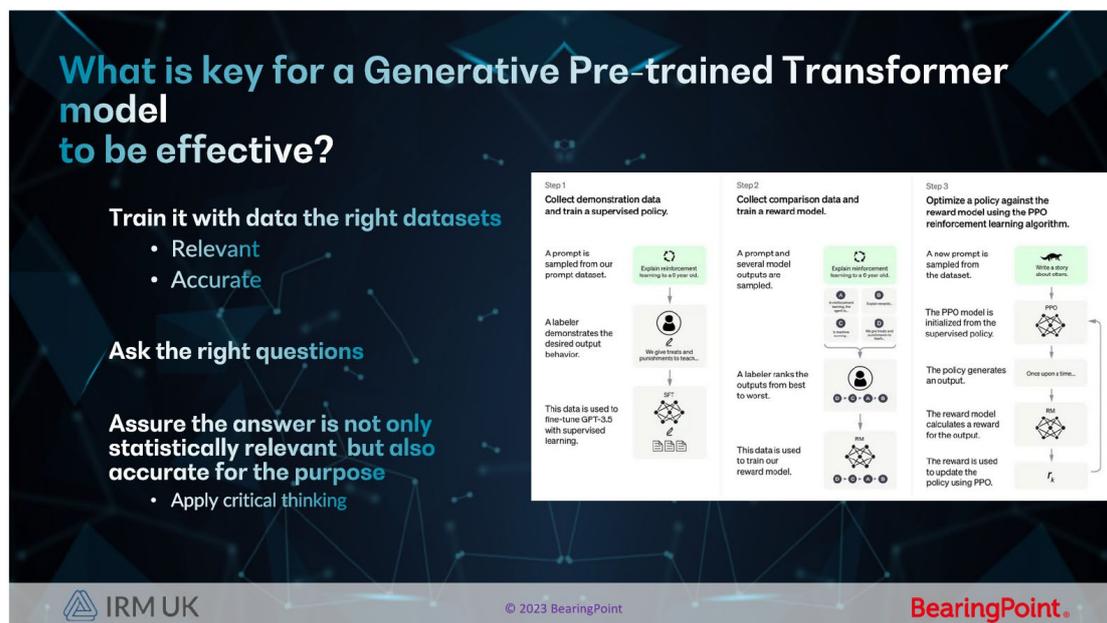


圖 1：三個使得 GPT 使用更有效率的方法
資料來源：BearingPoint

針對「以正確的資料集訓練模型」此一事項，又可細分為兩點，第一，用以訓練的資料集必須與所希望得到的結果有關聯性。舉例而言，若甲公司今天希望訓練出一個可以提供公司內部商業洞見的模型，那麼，甲公司應該優先以公司內部的資料製作訓練用資料集，而非其他公司的資料集。而除了甲公司內部資料以外，也應該將甲公司所處在之生態系的資料製作為訓練資料集。如此訓練出來的 GPT 模型，將與甲公司所處的生態系以及甲公司目前的營運情況緊密結合，進而能夠為甲公司提供最相關的商業洞見。第二，用以訓練的資料集必須要精確。為達到此目的，必須確保用以製作訓練用資料集的資料的品質。延續甲公司的例子，若甲公司內部的資料品質不佳，時常用誤植的情況發生，那麼以這些資料訓練出來的模型，亦會受到這些誤植的影響，進而降低所訓練出來 GPT 模型的可用性。

Jan 補充道，公司應注意自己的資料是否受到足夠的保護。若公司自己的商業資料任意提供給第三方 GPT 模型進行使用，可能造成公司商業資料成為其他公司「免費訓練」的材料。

The slide features a dark blue background with a network of white dots and lines. At the top left, the title 'Train it with relevant data sets' is written in white. Below it, a white text box contains the following content:

The quality of the outcome is greatly improved if the model is trained with relevant and qualitative data

- Feed it with your own data
- Feed it with specific data out of your ecosystem

Beware if you feed your data to commercial companies as "free training" as you will share your data point (hence value) with others.

Can you be sure that your data is properly protected?

On the right side, there is a green and white robot head icon labeled 'VEG3'. Next to it is a Bloomberg logo and the text 'BloombergGPT: A Large Language Model for Finance'. Below these is a diagram of the Microsoft 365 Copilot workflow. The diagram shows 'Microsoft 365 Apps' (Word, Excel, PowerPoint, Outlook) feeding into 'Microsoft 365 Copilot'. This process involves 'Grounding' (using 'Your data - emails, files, meetings, chats, calendar, and contacts') and 'Post-processing' to produce a 'LLM response' from a 'Large Language Model'.

At the bottom of the slide, there are logos for 'IRM UK', '© 2023 BearingPoint', and 'BearingPoint'.

圖 2：以相關之資料集訓練 GPT 模型
資料來源：BearingPoint

針對「問對的問題」此一事項，Jan 認為，公司必須使其員工理解 GPT 模型的特性，使員工在使用 GPT 模型時，能夠問對公司商業利益相關的問題，並且理解獲得有相關之輸出的方法。

針對「確保獲得的答案不僅在統計上有相關性，並且對所追求的目的而言是精確的」此一事項，Jan 認為，GPT 模型的使用者應該理解到，GPT 模型僅能夠提供與問題在統計上相關的答案，且無法指出答案與問題如何相關。因此，使用者應具備批判性思考的能力，要理解統計相關此一條件是否足夠做為自己的資料產品使用。並且針對組織內部，必須有對資料科學與 AI 的廣泛認知。

Apply critical thinking

- **What are the Trust requirements for you data product?**
 - Is statistical relevance (inferred or trained) appropriate for the use of your data product?
- **We need to extend the notion of data literacy to include both the question and the answer side of the equation.**
- **Plan for a broad data (science) and AI awareness session in your organisation in the long run and GPT solutions immediately.**



CRITICAL THINKING

Problem → Thinking → Reasoning → Analysing → Evaluating → Problem Solving → Decision Making → Solution

- 1 Avoid certain usage areas**
Avoid Using ChatGPT in a client project and avoid publishing documents and information created by ChatGPT in social media or otherwise.
- 2 Ask for consent**
Before uploading content or personal data to ChatGPT, always ask for consent.
- 3 Check the Output**
Always check the output for accuracy and completeness; never rely on the information provided by ChatGPT.
- 4 Be transparent!**
Always state that the content is AI generated and the content may be owned by third parties.

IRM UK © 2023 BearingPoint BearingPoint

圖 3：培養批判性思考能力

資料來源：BearingPoint

Jan 認為，針對 AI 管制，歐盟目前採取與過去對資料隱私相似的路徑：

- 在不損害基本權利的前提下，確保社會因 AI 技術而受益；
- 於 2023 年 3 月 31 日，義大利資料保護主管機關宣布即刻禁止使用 ChatGPT，並且針對該工具所使用的個人資料進行調查。

面對即將到來的管制環境，Jan 認為公司應該確保 AI 遵守 FAT（Fair, Accountability, Transparency）三原則：

- **公平（Fair）：**
應尊重個人隱私、遵守智慧財產權相關法令（專利法、商標法、著作權法），並且確保模型中沒有偏差（Bias）與歧視（Discrimination）
- **可歸責（Accountability）：**
實施適當且有效的方法以確保所使用的方法可通過合規性檢驗。
辨認潛在的負面結果
- **透明（Transparency）：**

不僅是解釋決策，並且包含資料科學專案的全部階段。為保護營業秘密及隱私，不必然將所有細節揭露予終端使用者。特殊重要性（feature importance）分析可協助促進透明化。

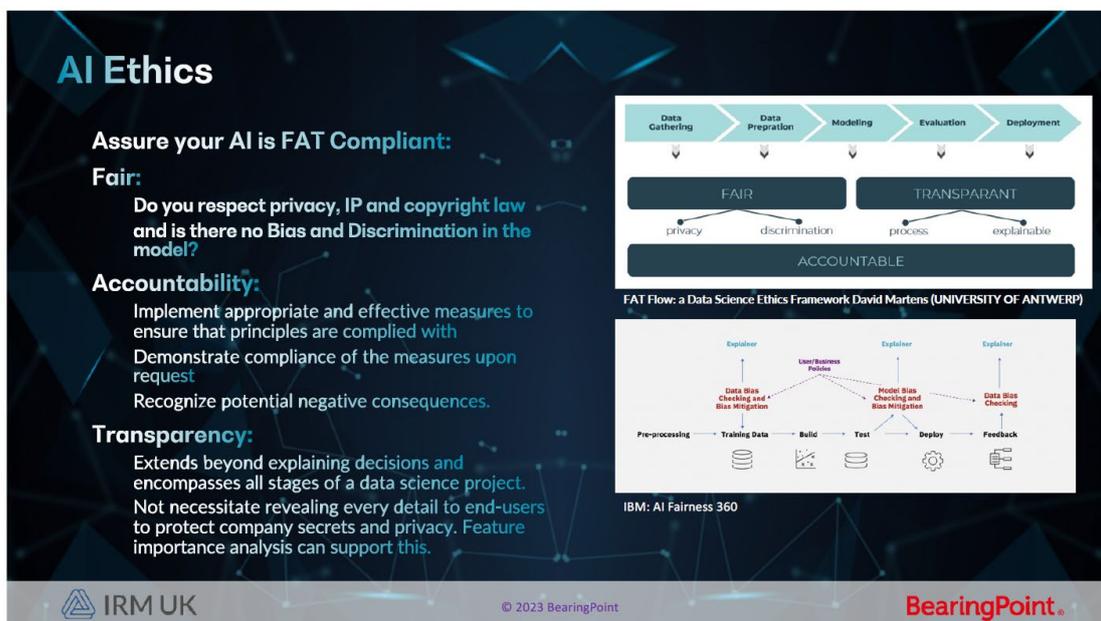


圖 4：AI 倫理三原則

資料來源：BearingPoint

(二) 釋放資料的力量 (Daring to Release the Power of Data)

本場次的演講者為來自 Volvo Penta 的 Valentina Niklasson 以及來自 Informed Decisions Consulting 的 Håkan Edvinsson。Volvo Penta 為 Volvo 集團旗下，生產遊艇及工業用引擎的製造商。Valentina 為 Volvo Penta 的商業資訊架構與資料治理主管，於本次演講中，偕同來自顧問公司 Informed Decisions Consulting 的 Håkan Edvinsson，一起為大家說明 Volvo Penta 如何進行數位治理，進而提供新的客戶服務。

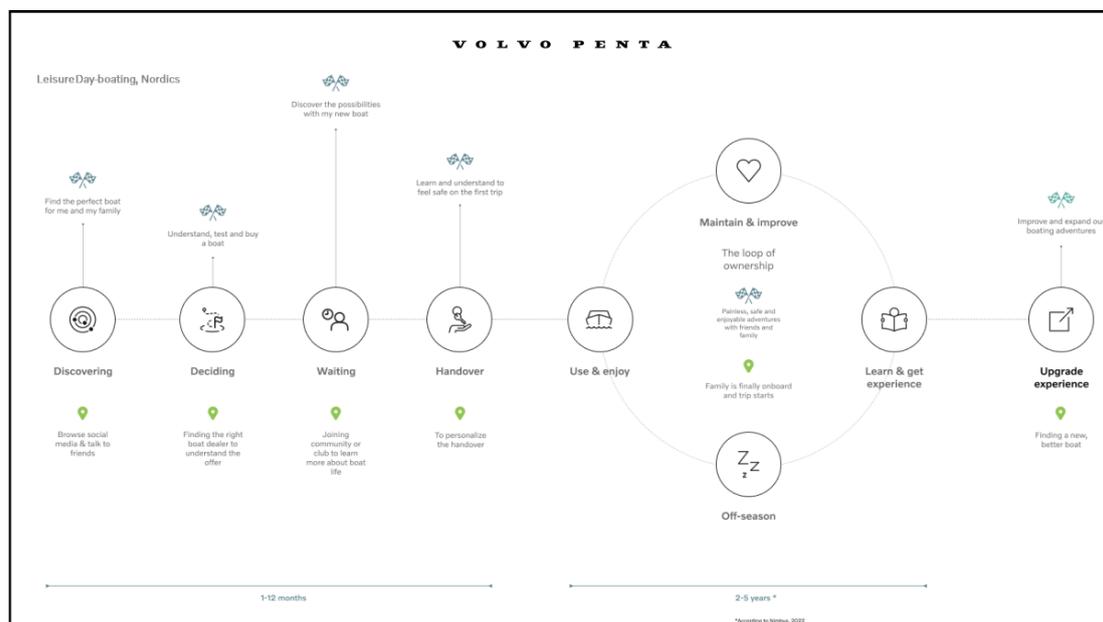


圖 5：Volvo Penta 所提供之整合型客戶服務

資料來源：Volvo Penta

Volvo Penta 過往專注於遊艇引擎以及相關零組件之製造，長時間下來，Volvo Penta 亦累積了大量的資料。於近年，Volvo Penta 認為這些資料除了提供做為組織內部維修或生產的參考外，更可以透過資料治理的方式，整合資料，並向客戶提供嶄新的加值服務。

因為上述的原因，Volvo Penta 與 Informed Decisions Consulting 合作，希望為這個百年企業帶來嶄新的風貌，進而為公司帶來長期穩定的收益。Volvo Penta 現在為遊艇船主帶來全方位的服務，從航程的規劃、停靠港補給、停靠港觀光、即時零組件監控、緊急情形通知、維修建議等皆包含在套裝服務之內。為達到這個目標，Volvo Penta 善用自身長期零組件製造所收集到的資料，在零組件設計時，便將可能使用到之資料、需即時監控的資料納入設計藍圖之中。因此，在打造船隻時，Volvo Penta 可以掌握整體船隻的情形，便針對必要資料加以監控，從而達到上述的整體套裝服務。Volvo Penta 為了達成加值服務之目標（如觀光資訊等），亦與相關產業結合，例如各大港口之旅遊業者，進而可以將 Volvo Penta 所掌握之資訊提供予客戶。

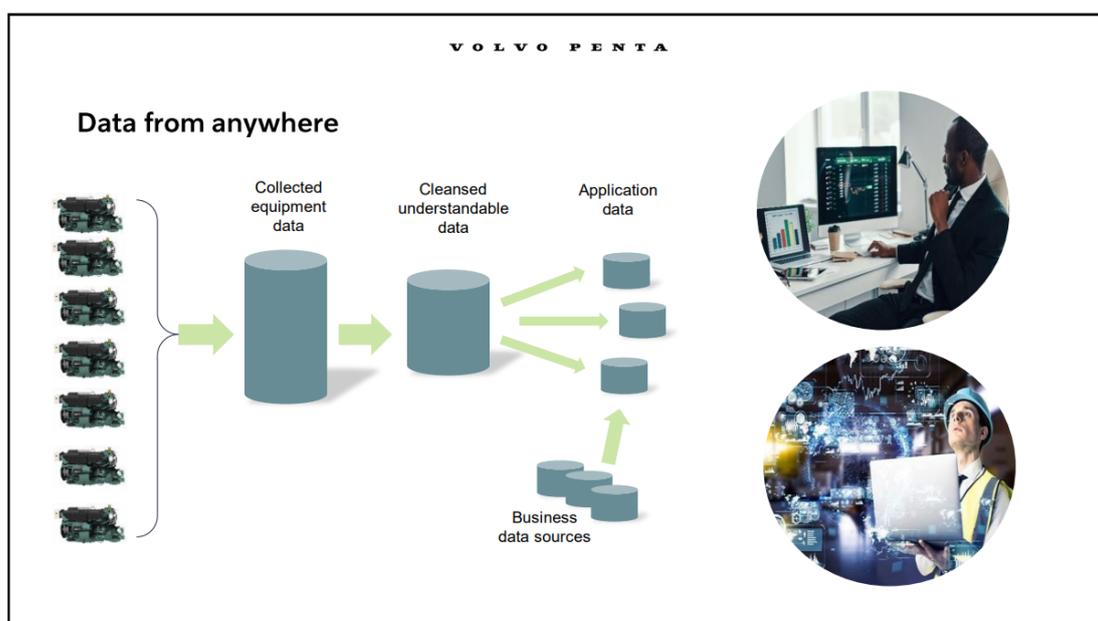


圖 6：為達成整體服務的資料架構

資料來源：Volvo Penta

最後，兩位講者強調企業在進行資料轉型時，要注意的原則：「商業資料設計就是商業設計（Business data design is business design）」。提醒希望進行資料轉型的企業，在設計商業資料之蒐集、應用機制時，要緊貼企業自身之商業目的，唯有面向好的商業設計，才能達成好的商業資料設計。

(三) 設計可執行的資料治理組織 (Designing an Actionable Data Governance Organization)

本場次由荷蘭發展銀行 (FMO – Dutch Development Bank) 的資深資料治理專員 Dana Julinschi 為大家帶來演講。荷蘭發展銀行設立於 1970 年，是一間官股銀行，其 51% 的股權為荷蘭政府所有。荷蘭發展銀行的主要任務，是為低收入及中收入地區中的私部門帶來額外的金融資助。

在本場次的演講中，Dana 為聽眾們介紹荷蘭發展銀行是如何進行資料治理，並且賦權各部門的大小職員，使整個組織「動起來」進行積極的行動性資料治理。

在過往推動資料治理的過程中，Dana 發現，通常在推動資料治理時，時常會發生一種現象——在初始時，組織內大部分的職員會為組織進行資料治理感到興奮且積極參與，但一段時間後，成員們多半感受到挫折且態度轉為消極。面對這樣的情形，Dana 認為，要進行持續性且有效的資料治理，必須要採取「行動性資料治理 (Actionable Data Governance)」。Dana 認為，要成就行動性資料治理，必須要結合五個要素：脈絡、願景、量能、人事、以及貫徹始終 (Delivery)。

執行資料治理的人員，要先為組織內的資料治理之脈絡做出明確的定義，首先必須釐清自己在組織內所處的角色以及要進行的事情。接著，針對資料治理提出價值提議 (value proposition)，亦即之所以要進行資料治理所能提供的誘因。第三，必須知道資料治理與組織之間的關係，並且釐清組織內外與資料治理相關的利害關係人。最後，必須了解資料治理與組織內其他目標的關係，亦即了解組織內哪些人可以成為資料治理的合作夥伴，而哪些人會是你的同儕。



圖 7：定義資料治理脈絡

資料來源：荷蘭發展銀行

在釐清資料治理的脈絡之後，資料治理推動者應進一步提出組織之資料治理願景。Dana 認為，一個成功的行動性資料治理願景，有三原則與二方向。資料治理推動者必須「找到問題」，對自己所要解決的問題有清楚的定義，這同時包含目前的問題以及未來可能的問題。第二，資料治理推動者必須「描繪成功」，一個只看到問題而不知何謂成功的資料治理，無法使組織內部積極參與，因此，資料治理推動者必須要對成功的圖像有所描繪，使組織內部能夠為了成功的圖像積極協助參與資料治理推動者所提出的資料治理目標。第三，資料治理推動者必須「解釋所採取的行動」，資料治理需要將大目標切分成許多細節的行動，讓組織內部的許多部門成員一同參與，而在細分的過程中，並不是所有的部門成員都能夠清楚看到細節的行動與整體希望達成的目標之間的關聯性是什麼，進而使得這些部門成員缺少積極參與資料治理的動力。因此，Dana 認為，資料治理推動者應該慢下腳步，向合作的各部門成員解釋之所以採取特定細節行動的原因，以及這些行動與成功圖像之間的關聯。在進行三原則時，資料治理推動者也要掌握二方向，在思考時，要「策略性思考」，不論是在思考問題時或是在描繪成功時，要不時以策略性、長期發展的角度從新思考。第二個方向，則是要「說人話」，因

為資料治理牽涉到的組織內部門甚廣，每個部門成員的背景領域並不相同，並不是所有人都對資料治理有所了解。資料治理推動人員在與其他部門成員溝通時，必須以實際與簡要的方式進行，避免長篇大論或高談闊論，使得部門成員無法清楚理解所要採取的行動。



圖 8：行動性資料治理三原則二方向

資料來源：荷蘭發展銀行

資料治理推動人員，必須要掌握組織內部推動資料治理的「量能（capabilities）」，資料治理需要組織內部的許多部門成員一起努力，並非一個人、一個單位可以一蹴可幾。因此資料治理推動者必須要根據不同的項目，委託不同的部門成員進行，而為了要有正確的工作委派，資料治理推動者亦需考量到組織整體在進行資料治理所能提供的量能，亦即多少程度的資料治理，是可以長時間在組織內部推動而不影響其日常任務的。Dana 建議，資料治理推動者可以就人事、技術、程序、資料等四個面向分別考量。並以資料產生、資料分配、資料使用三個階段考量。進而得出整體的資料治理營運模型。



圖 9：資料治理之量能與營運模型

資料來源：荷蘭發展銀行

在「人事」要素上，Dana 強調，資料治理推動者必須清楚自己的核心隊友有
 哪些人，並給予這些人適度的信任，並掌握延伸的隊友組成。除了隊友，資料治
 理推動者也需要了解組織內其他人與資料治理的關係，掌握各部門成員與資料之
 間的可歸責性，甚至必須掌握組織內部可能成為資料治理之推力的成員。

在「貫徹始終」要素上，Dana 認為，資料治理推動者必須專注於自己的目標，
 對資料治理整體情勢有所掌握，並且隨時保持韌性。

(四) MDM的未來 (The Future of MDM)

本場次的主講人為 Profisee 公司的資料策略長 Malcom T. Hawker。Profisee 是一間提供主資料管理服務的公司，其協助企業進行良好的主資料管理，提升資料品質，進而增加資料的可用性，使企業能夠創造更多的利益。同時，Profisee 是第一個也是唯一一個與微軟「Fabric」整合的主資料管理 (Master Data Management)，可令微軟「Azure」服務使用者的資料價值最大化。

講者 Malcom 有 29 年的產業經驗，曾擔任 Gartner 的資深執行分析師，為主資料管理與資料治理的專家。

在本場演講中，Malcom 演講的主題為主資料管理的未來。在近年來，許多新興的資料處理技術都宣稱可以取代過往的主資料管理。但時至 2023，Malcom 認為，隨著 AI 時代的到來，主資料管理的重要性反而愈發重要，從過去的「有很好 (nice to have) 轉變為「必須有 (must have)」，因為良好的主資料管理，將會是任何數位轉型倡議的基礎。

Malcom 認為，如果一個組織允許業務部門獨立管理其程序或資料，則需要主資料管理以將跨部門的營運與洞見最佳化。Malcom 在本次的演講中，提出了三個主要論點：1. 業務需求迅速演變，而主資料管理仍然是數位企業的根本要素；2. 科技的進步允許主資料管理程式更加的達成這些不斷變化的商業需求；3. 新的主資料管理能力將加速支持主資料管理倡議的人與流程的質變。

隨著資料科學不斷進步，如 GPT 技術等更貼近人類的機器學習介面的出現，主資料管理的模式也會隨著技術與時代的演變持續進步。在過去，主資料管理主要是人工進行，而近年來，隨著 GPT 技術的出現，在主資料管理上，如微軟等大型企業也開始提供以機器學習技術輔助的主資料經緯（Master Data Fabric）系統，使企業/組織在進行主資料管理時，有人工智慧的協助，而提供更有效率的管理。

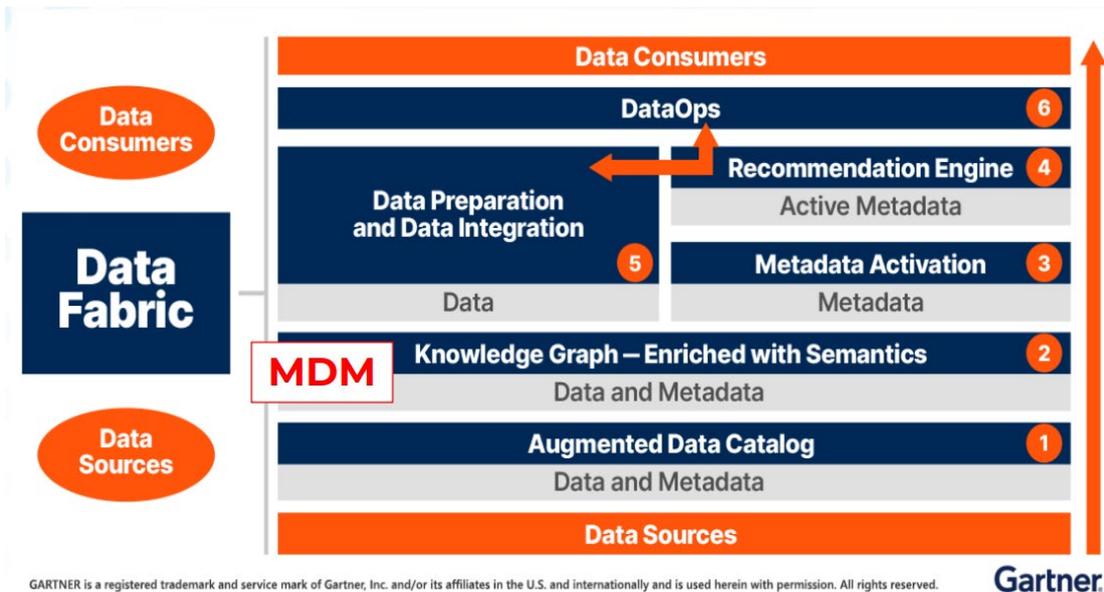


圖 10：主資料經緯系統之示意圖

資料來源：Gartner

因此，Malcom 針對主資料管理的未來發展趨勢，做出了下圖的預測：

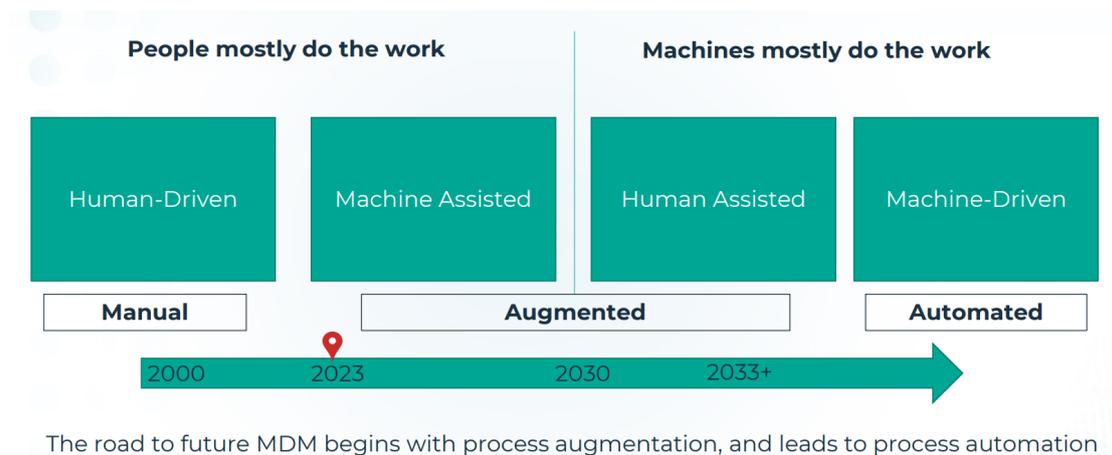


圖 11：主資料管理的未來發展預測

資料來源：Profisee

Malcom 認為，到了 2030 年，主資料管理將會有以下六點關鍵能力：

- 適應性主資料管理模型與治理（Adaptive MDM Data Modelling and Governance）
- 原位資料匹配（In Situ Data Matching）
- 與治理及目錄平臺深度整合（Deep Integration with Governance & Catalog Platforms）
- 外部資料共享與區塊鏈（External Data Sharing & Blockchain）
- 多次元階層（Multi-Dimensional Hierarchies）
- 人工智慧輔助，分散式管理關係（AI- Supported, Distributed Stewardship）

最後，Malcom 也提到，過去的資料治理多為規則遵循的模式。而因應更多元的資料應用型態以及人工智慧輔助工具，資料治理將轉曾規則遵循模式轉向為例外管理模式。亦即，在特殊情形出現時，才會觸發資料治理的審查。

(五) 如何將資料治理的優勢與靈活資料治理結合？（How Can you Combine the Strength of Data Governance with the Flexibility and Speed of Agile？）

本場次主講人為來自 QuaData 的 Sophie Angenot。QuaData 係位於比利時的顧問公司，主要業務為輔導企業客戶實施「靈活資料治理（Agile Data Governance）」進而改善組織之資料品質、資料管理、隱私及資料保護。Sophie 為創辦 QuaData 的管理合夥人，在資料品質與資料治理上有 15 年的經驗。在本場次的演講中，Sophie 為與會者介紹何謂靈活資料治理，以及其所帶來的好處。

何謂靈活資料治理？靈活資料治理的倡議者 Sophie 用以下的定義來說明所謂的靈活資料治理：「靈活資料治理促進會產生價值資料資產的管理程序，其採用負責任即可歸責的資料治理角色之網路，其中，該等資料治理角色係為了組織整體的利益而協作。靈活資料治理使用了已被證明為最佳實作的靈活方法達成可評量的結果（measurable results）。」

Sophie 認為，若一個組織成功導入靈活資料治理，將可以使得組織的資料資產達到以下幾點的成果：

- 滿足使用需求的品質
- 對尋找資料的人而言，資料是可找到的
- 對需要資料的人而言，資料是可取得的
- 對使用資料的人而言，資料是可理解的
- 適當地受到保護
- 以法規遵循的方式加以收集、管理及使用

Sophie 認為，要能成功達成靈活資料治理的關鍵，並不是創造複雜的法遵程序或使用特定的資料技術，而在於打造組織整體的資料治理文化。

傳統的資料治理強調創造資料治理的科層化結構，並且採金字塔型的由上而下的資料治理推動模型。Sophie 提到，傳統上的資料治理多會有三個層級，由上而下分別為：整體策略層級（資料治理議會）、個別戰術層級（資料治理執行委員會）、以及營運層級。而這樣的科層化結構，多半使得組織的資料治理不夠靈活，且無法得到大多數的成員積極支持。

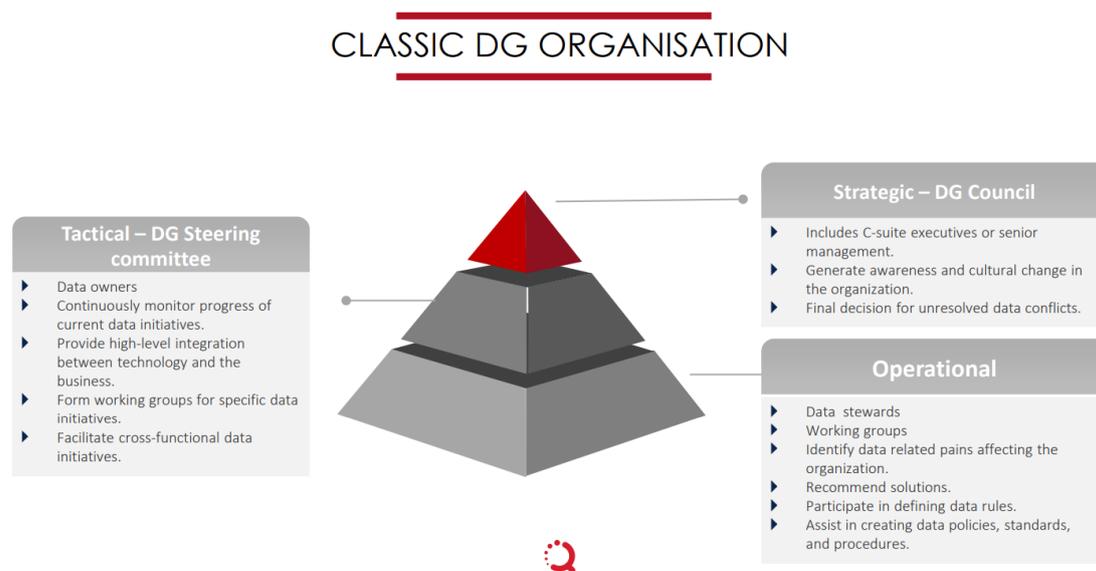


圖 12：傳統的資料治理組織

資料來源：QuaData

在本次會議中，Sophie 提到了如何使得資料治理組織更加靈活的四個大原則：

- 對組織內部成員（資料擁有者（data owners）及資料管家（data stewards））的授權應清楚明瞭，不要讓所有決定都須經由組織高層
- 以可評量的方式定義責任，使組織成員知道需要追求的方向
- 藉由創造透明性而促進信任
- 專注於產生價值的活動/為資料消費者（Data Consumers）排除障礙

CONSENT DECISION MAKING

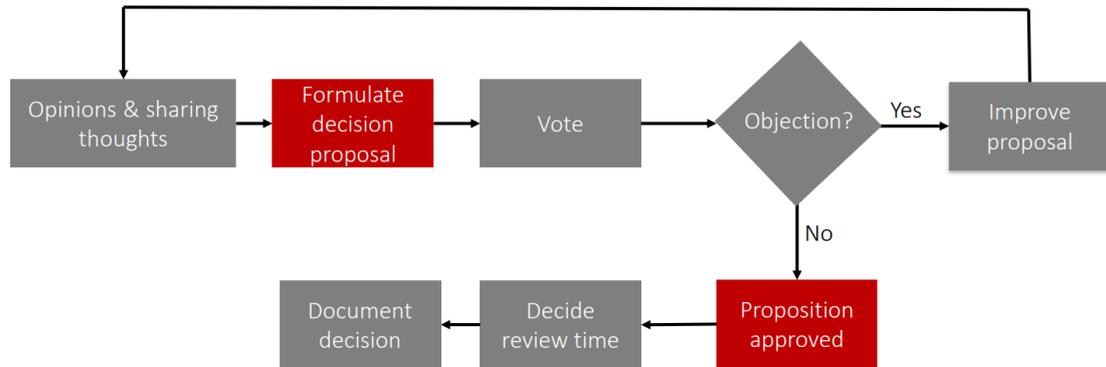


圖 13：靈活資料治理的組織內部同意決定機制示意圖

資料來源：QuaData

作為結論，Sophie 提及，資料治理並非一蹴可及的一次性工作，試著建構資料治理的組織應該接受資料治理之建立是一系列的改變，而資料治理的推動者應對其適當的管理，使其變動是適切的。

(六) 資料治理的變革 (The Changing Face of Data Governance)

本場次的演講者為來自伯明罕大學醫院 NHS 基金信託 (University Hospitals Birmingham NHS Foundation Trust) 的資料經理 Guy Harvey。Guy Harvey 雖然目前任職於伯明罕大學醫院 NHS 基金信託，但其主要是作為獨立顧問，協助有需要的機關進行資料治理。

在本次會議中，Guy Harvey 以其多年的獨立顧問經驗，分享對資料治理的反思。Guy Harvey 首先統整，其長年觀察到現行的資料治理情形，包含：普通的實施方式 (run-of-the-mill implementation)、通常的準備活動、以及帶動改變的驅力。

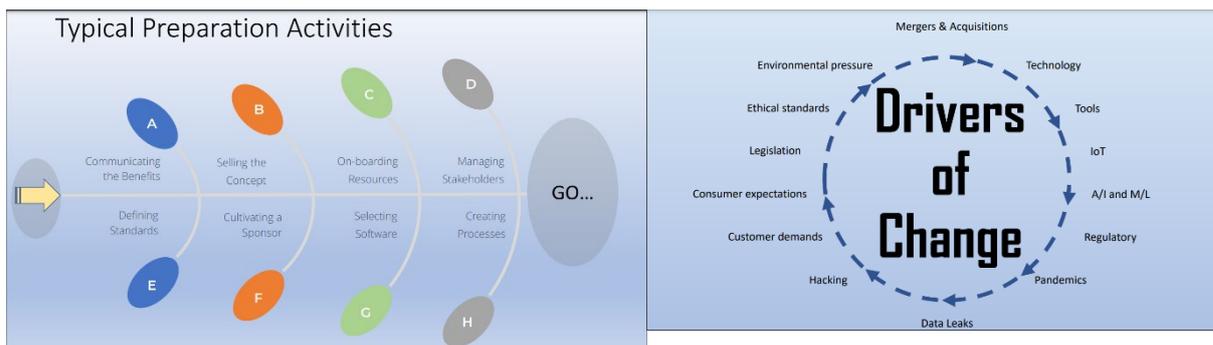


圖 14：現行的資料治理情形

資料來源：Guy Harvey

Guy 認為，資料治理的隨時代演進，可以歸納為「資料所有關係 (Data Ownership)」時代演進而產生型態的變化。根據定義，資料所有關係係指：對於

資訊的持有及責任。所有關係隱含權力以及控制。對資訊的控制不僅包含對資料進行以下行動的能力：存取、產生、修改、封包、從中受益、販賣或移除，也包含將上述行為之權限指派給他人的權力。

Guy 認為，若以資料治理成熟度以及組織成員的參與程度為軸線作圖，可以將資料所有關係分為四個類型：集權型（Totalitarian）、聯邦型（Federated）、在地化型（Localized）、協作型（Collaborative）。隨著資料治理成熟度以及參與程度提升，資料所有關係將從集權型逐漸演變為協作型。

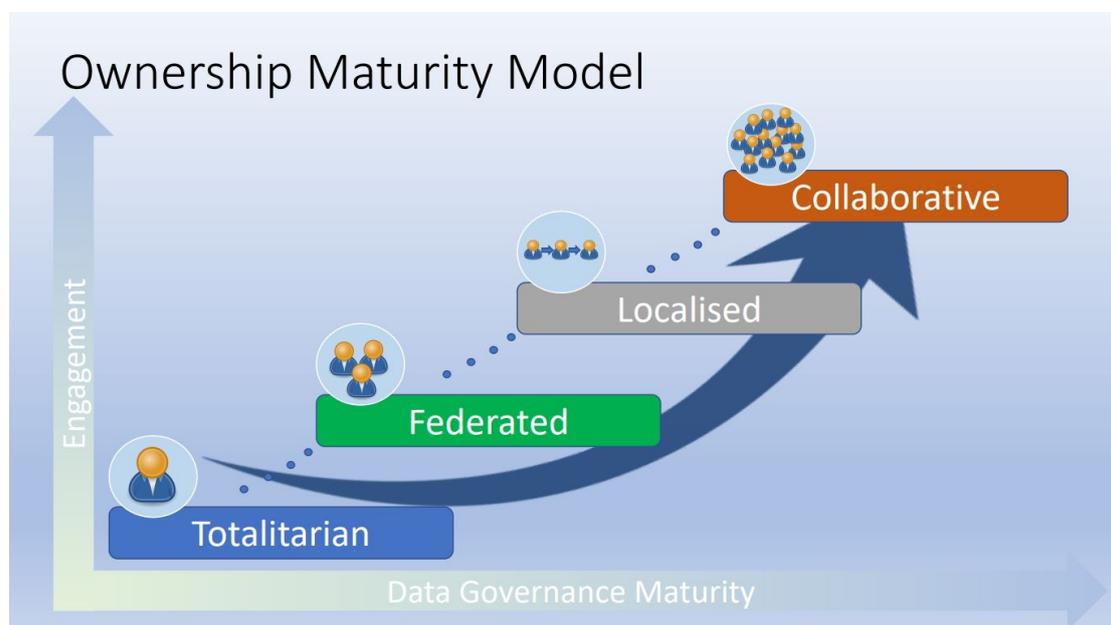


圖 15：資料所有關係之成熟度模型

資料來源：Guy Harvey

此四種資料所有關係，又可整理成以下表格：

表 1：四種資料所有關係

| | 基礎 | 成功要素 |
|-----------|---------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------|
| 集權型資料所有關係 | <ul style="list-style-type: none"> • 單一所有人負責全部資料領域 • 單一線的歸責關係 | <ul style="list-style-type: none"> • 強硬的治理結構 • 主要為簡單的資料領域 • 資料領域分處在少數位置中 • 資料平臺數少 |

| | | |
|-----------|----------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 聯邦型資料所有關係 | <ul style="list-style-type: none"> • 資料所有者一同作業 • 單一所有關係過於困難 | <ul style="list-style-type: none"> • 較不強硬的治理結構 • 少數幾個簡單的資料領域 • 資料領域分處數個位置 • 混合的資料平臺 |
| 在地化資料所有關係 | <ul style="list-style-type: none"> • 有傳輸中的資料以及靜止(Rest)資料 • 非線性的程序與所有關係 • 組織性結構 • 位置及地理 | <ul style="list-style-type: none"> • 普遍同意的規則與標準 • 可共享的知識庫 • 有意願一同作業 • 採接力形式 • 可回報「環節中最脆弱處」的合作方法 |
| 協作型資料所有關係 | <ul style="list-style-type: none"> • 共享的資料領域與子領域 • 共享的業務程序 | <ul style="list-style-type: none"> • 數個用於收集/儲存的位置 • 對資料倫理有共同信念 • 社群精神以及協作文化 • 整齊有序且可存取的元資料 (metadata) • 無階層式治理結構 |

資料來源：Guy Harvey

(七) 閉幕演講：海盜運動 (Be More Pirate)

本場次的演講為第一天研討會的閉幕演講，主辦單位請到「Be More Pirate」運動發起人之一的 Alex Barker 來為大家進行演說。「Be More Pirate」運動是一場由同名書籍所產生的運動風潮，該書的作者藉由講述未被大眾熟知的海盜的歷史，來說明非階層化的管理結構，如何在大航海時代大早一個小型的現代民主社會。組織當中的每一份子都有充分的發言權，並且彼此為集體利益著想而行動，並且積極友善包容每一個個體。講者及作者希望現代社會在看到書之後，也起身改變，成為一個更積極、友善、且富有團隊合作精神的人。

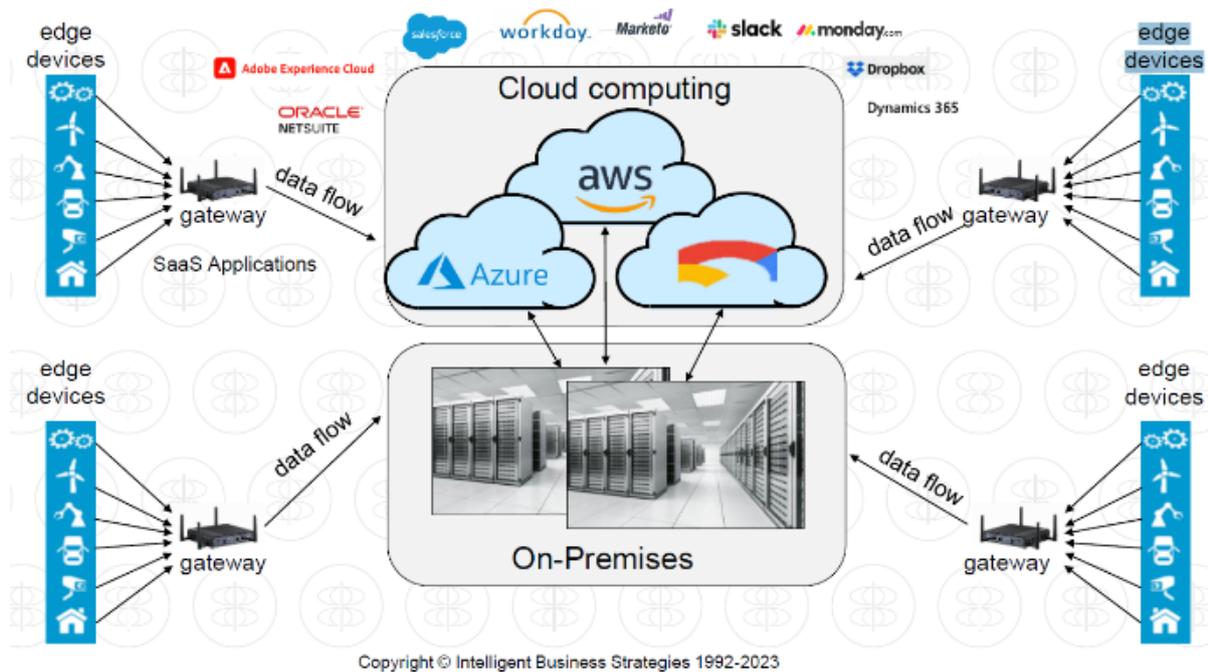
二、 第二天會議重點

(一) 全體大會演講：資料可觀察性－人工智慧驅動資料治理之及時性與自動化 (Plenary Keynote: Data Observability - A Catalyst for AI-Driven Real-time Data Governance Automation)

本場會議主講人為 Intelligent 商業策略公司總經理－Mike Ferguson。Mike 為獨立 IT 商業分析師及顧問，擁有 40 年的 IT 經驗。Mike 現今任職的 Intelligent 商業策略公司成立於 1992 年，公司總部位於英國，致力於 IT 諮詢與分析業務，如市場研究、資料治理、機器學習、產品規劃等。

在本次演講中，Mike 首先提到 1985 年至今，資料處理的轉變。在 1985 年代企業資料處理最終都會集中在單一台 IBM S/390 大型主機中，面臨的風險是資料處理過程中若出現難以發現的錯誤，會在累積到一定程度後，人們才會發現資料不正確，但此時資料庫已損毀，必須花費許多時間和人力才能找到原因，導致企業的資訊系統必須停擺，造成企業營業損失。

到了今日，隨著資料量日益增多及複雜化，資料之處理與存放也轉換到混合式運算環境 (Hybrid computing)，資料被分散在各處，除了本地系統外，還包含雲端運算 (Cloud computing)、邊緣裝置 (Edge device)、軟體即服務 (SaaS)。許多公司也使用不同的分析工具，來清理或分析不同系統中的資料，包含資料倉庫、資料湖、使用 Graph Analysis 圖形系統分析大規模的圖庫資料、使用資料挖掘工具 (Data mining tools) 進行資料科學任務 (Data science) 等。



Copyright © Intelligent Business Strategies 1992-2023

圖 16：資料在混合式運算環境

資料來源：Intelligent 商業策略公司

透過資料網格（Data Mesh）可使組織建立一個一致性的資料治理策略，確保資料產品保持一致性並達到資料共享，而企業則會透過開發資料管道（Data pipelines），來確保資料處理流程（資料提取、轉換、清洗、分析、儲存的過程）的準確可靠，達到有效的資料分析，但可能也面臨資料管道故障之風險，尤其資料發生異常時，因資料管道分散於不同系統及工具中，如何查詢紀錄、找到根本原因將是一大問題。

此時可透過資料可觀察性（Data Observability），進行資料管道和資料流程的監控及記錄，確保資料管道正常運作，提高資料的品質、完整性、即時性及可靠性。然而資料可觀察性也面臨一些問題，包含所用的資料沿襲（Data Lineage）工具設計不符合實際使用的期待，如目前管道執行過程中所創建的文件無法事先設定資料目錄（Data Catalogs）、資料延遲問題、資料管道通常涉及多個工具與技術，必須建立跨工具的協調機制（如資料格式標準），確保資料在不同管道中流動。



How Does Data Observability Work? - Monitor The Pipeline Orchestration, Each Pipeline Task And Data Flowing Through Each Task

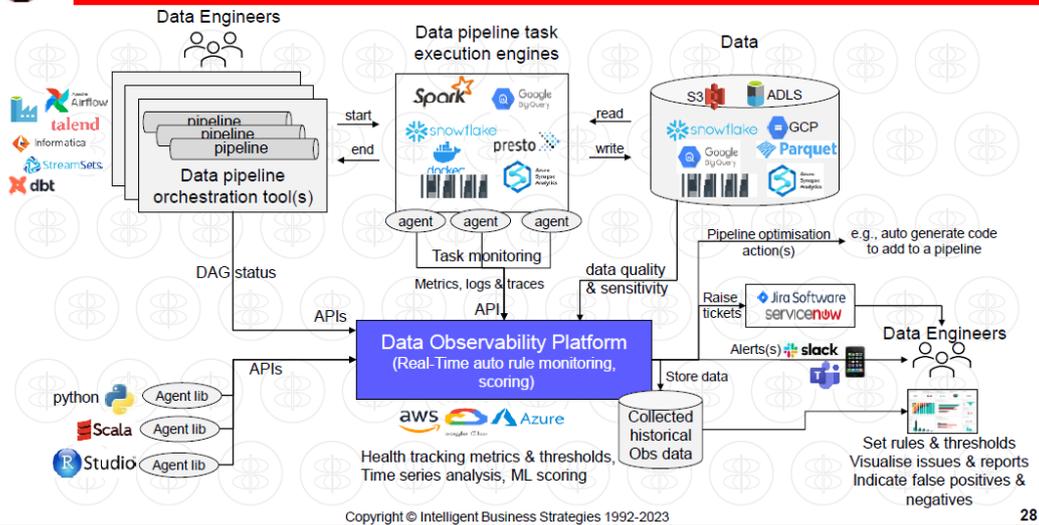


圖 17：資料可觀察性的運作機制

資料來源：Intelligent 商業策略公司

隨著企業擁有的資料日益複雜，資料分析服務已成為資料驅動業務的基石，目前許多大型公司都在嘗試透過機器學習、人工智慧、商業智慧（Business Intelligence）等技術，建立一個高品質且合規的資料產品，但遇到的困難是資料分散在不同地方，企業必須建立一個智慧商業策略的資料治理框架，框架之重點圍繞在資料生命週期的管理，從創建、提存（Data Ingestion）到資料刪除，其主要元素包含目標及策略、組織結構與責任、資料流程、資料品質一致性、制定政策（如符合不同地區的資料隱私政策、資料安全之保護措施、資料共享）等。

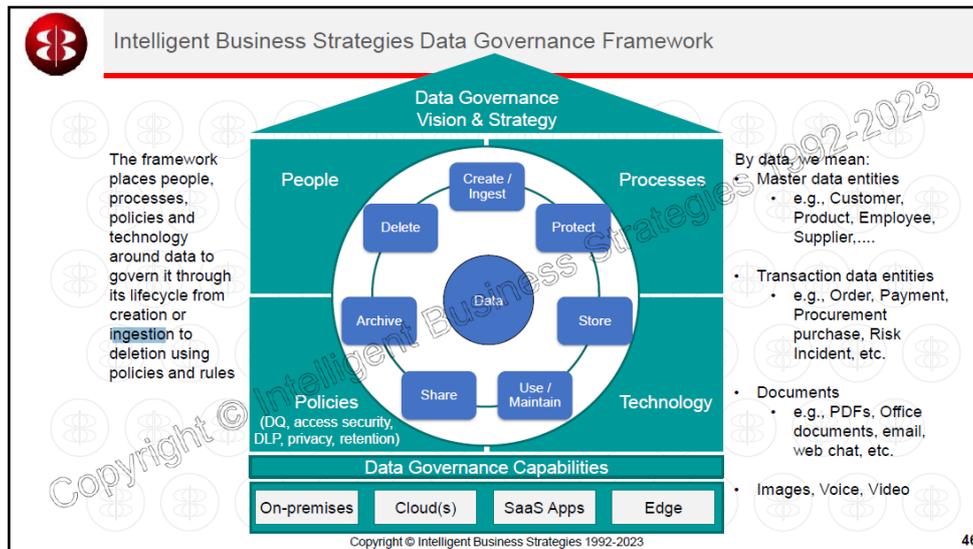


圖 18：智慧商業策略資料治理框架

資料來源：Intelligent 商業策略公司

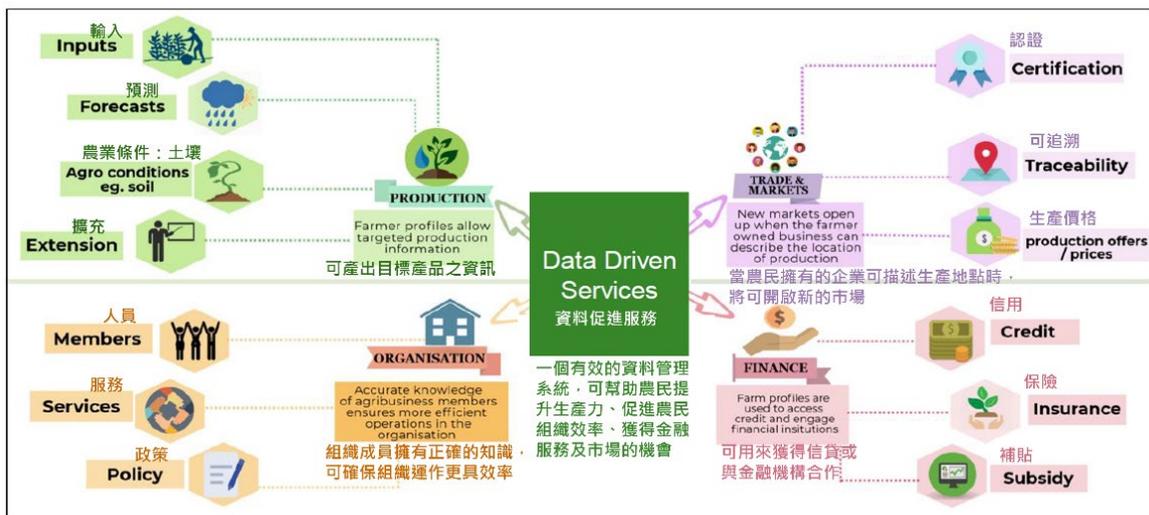
在資料治理的過程中，人們會透過資料可觀察工具監控資料，此外亦可透過資料目錄，為分散各處不同的資料進行分類、描述，以管理企業內部的資料資產、元資料（Metadata），目前已知的資料目錄工具如 Microsoft Purview 已可透過機器學習自動分類敏感性資料，另外 Azure 雲端資料湖平台（Data Lake Storage）亦可針對資料目錄中特定類別的資料創建資料管理策略（如訪問權限）。

最後，Mike 提到對於資料治理的未來展望，其認為隨著人工智慧的發展，可驅動資料治理的相關技術或工具，達到資料即時及自動化之分析與監控，預計全球市場會從 2022 年的 32.8 億到 2026 年達到 74.2 億的增長。

(二) 利用 FAIR 提升資料治理效能(Leveraging FAIR for Better Data Governance)

本場會議由國際非營利性組織—國際農業及生物科學中心（Centre for Agriculture and Biosciences. International, CABI）分享其在成員國間推動之農業資料共享計畫。CABI 希望透過資料管理，改善農業價值鏈的各個環節，包含：

1. 提高農民生產產量：透過資料輸入分析預測天氣及土壤條件，達到精準農業之目標。
2. 發展健全組織：幫助成員夥伴和農業部門人員具備專業知識，讓組織運作更具效率，專注於可持續發展之農業目標。
3. 開啟新市場及信貸：透過認證、產品可追溯性及生產價格，讓農民有機會開啟新的市場，進而有機會向金融機構借貸或保險。



Source: CTA.int

圖 19：如何透過資料幫助農民？

資料來源：CABI

在農業資料共享計畫中，CABI 分享了農業資料共享計畫啟動至今，歷經了 3 階段：

【階段一（2018）：找出無法好好利用資料的原因】

CABI 花了 6 個月的時間訪談利害關係人，發現資料沒有妥善流動、實踐的原因，通常不是因為技術或科技問題，而與當地文化、政治有關，例如有些國家存在複雜的政治關係導致決策速度很慢，又或是人們恐懼資料共享所帶來之不利影響。對此，CABI 透過 FAIR 原則—資料可尋找性（Findable）、可訪問性

(Accessible)、可互操作性 (Interoperable) 及可重用性 (Reusable)，提升大家參與意願，如讓農民了解什麼資料可公開（如農藥配方和觀察結果可公開）、什麼資料不必公開（農民的財務狀況），並透過資料透明揭露之方式，讓每個人了解資料之可訪問性及可重用性，提升人們分享資料的意願，降低人們對資料共享之擔憂。

【階段二（2019-2021）：克服已識別的挑戰，使計畫的相關成員（如政府部門）開始實踐】

CABI 認為 FAIR 原則是個很有效的對話工具，但在執行上同樣也會面臨挑戰和阻礙，以語言為例，因不同國家及組織成員間使用之語言不同，為確保成員間之系統資料索引及項目一致，CABI 建議成員間應使用相同且簡單的語言表達，以確保大家往同一目標邁進。然而，CABI 仍然發現到有些國家的資料庫是離線的，或同一國家不同組織間不願對話，導致資料無法流通，或是資料受助方無權限取得資料等問題。此外，CABI 亦發現組織中不一定有專門的單位支持資料的共享，因此，CABI 開發資料共享工具包 (Data sharing toolkit)，該工具針對 7 種模組 (情境)，採用 FAIR 原則，進行課程、指南、案例研究、備忘錄、清單等供成員學習。



圖 20：CABI 開發之資料共享工具包
資料來源：CABI²

² CABI，資料共享工具包，<https://www.datasharingtoolkit.org/#how-to-use>

【階段三（2021-2025）：建立完善的資料生態系統、發展可複製的流程】

在階段三中，CABI 希望持續強化組織成員之知識及技能，進行 FAIR 原則之實踐；在以人為本之設計下，使資料受助方間有效交流對話，從實踐的過程中，持續修正資料共享工具包，建立完善的資料生態系統，使資料可持續被訪問、有效減少資料共享的不利影響，最終發展出可複製於其他組織之流程。目前 CABI 提出之流程包含：

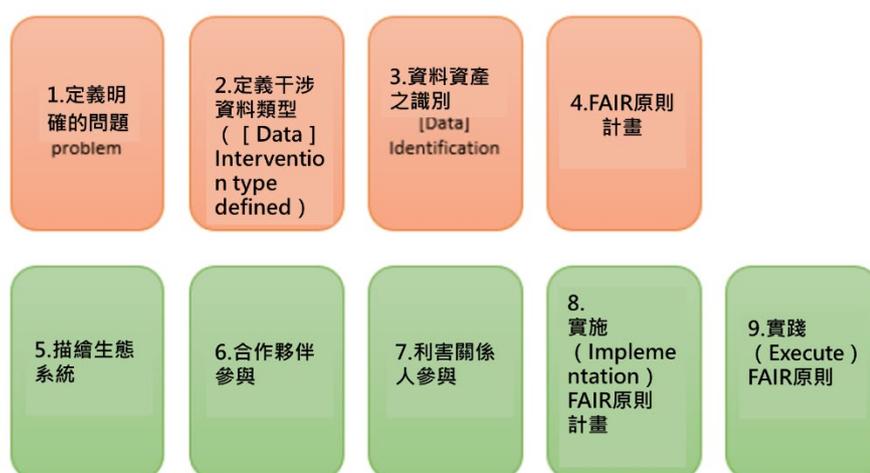


圖 21：CABI 開發之資料共享工具包

資料來源：CABI

(三) 評估資料治理對於英國國家統計局的影響力 (Measuring Data Impact with the Office for National Statistics)

本場會議主講人為 MataData Works 公司的執行長—Adam Milward。MataData Works 公司為 SaaS 的供應商，提供企業資料管理解決方案，幫助組織更有效率的管理資料。在本次演講中，主講者分享參與英國健康資料研究機構 (Health Data Research UK) 其中一個工作項目的資料整合經驗。

英國健康資料研究機構為國家級的健康資料科學研究機構，成立於 2018 年，組織成員涵蓋北愛爾蘭、蘇格蘭、威爾士、英格蘭之健康資料研究所、慈善機構、企業等。該機構成立之背景為英國雖然不乏諸多從事健康科學分析之研究人員、許多從事生命科學相關的公司 (如大型製藥公司)，亦擁有英國國家醫療服務系統 (National Health Service) 龐大的健康資料集，但在尋找或取得所需資料時，往往存在一大障礙，例如，若研究人員想研究罕見疾病，希望找到罕見疾病患者以獲取相關治療之研究，但由於罕見疾病患者相當少見，研究人員通常得花費許多人力及時間成本，包含聯繫各個醫院以找到正確的人並歷經請求使用資料之過程。此外研究人員也常發現獲得的資料與預期不符，例如奇怪的編碼或是資料有錯誤等問題。而為解決上述問題，讓人們可找到、獲得並理解所需資料，促進健康資料研究，英國健康資料研究機構開始進行資料系統之整頓，並找來了 Adam 所在的 MataData Works 公司，提供資料處理解決方案。

MataData Works 提供的解決方案，其重點圍繞資料的可發現性與資料的可理解性。在執行上可分為幾個階段，首先在前二週內，MataData Works 先建立了資料目錄，並訪談不同對象，以了解所需之資料內容。在第三週及第四週內對於資料架構達成一致性共識。在第五週到第八週後，將元資料之概念及相關工具導入系統中。在第 11 週至第 12 週中，除了系統穩定執行外，又增加的其他功能，例如在元資料的紀錄中增加更多實用的資訊，達到資料知識的共享。

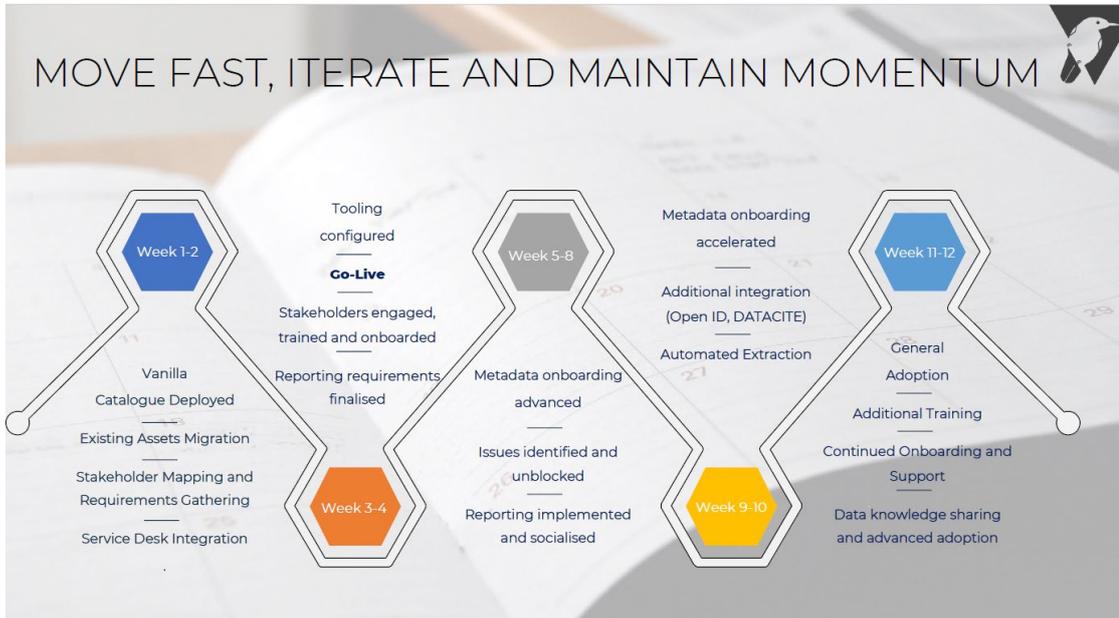


圖 22：MataData Works 提供英國健康資料研究機構之解決方案時程
資料來源：MataData Works

在這些過程中，最困難的地方為如何鼓勵大家參與並在系統上建立資料目錄，對此 MataData Works 採用貢獻排行榜的動態方式，當某一個組織發現排行榜中名次下滑，則會積極貢獻資料目錄，將整個過程遊戲化，最終達到充實資料目錄及資料集的目的。

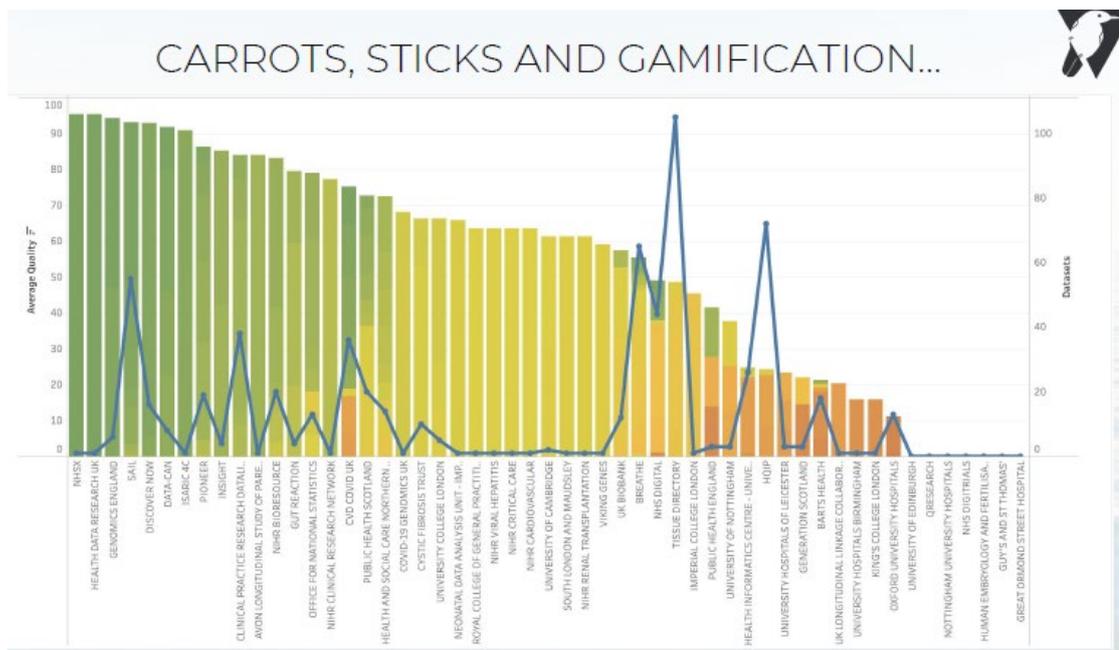


圖 23：MataData Works 透過排行榜鼓勵組織參與
資料來源：MataData Works

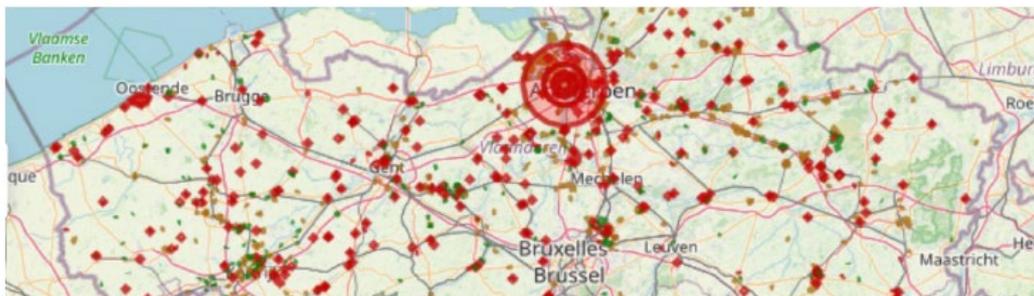
在創建了不同的資料集後，MataData Works 也開發了追蹤工具—數位物件識別符（Digital Object Identifier），以了解資料集使用情形，如哪些資料集被請求使用、哪些資料值得投資，並進一步評估資料品質和建立信任。在追蹤資料集的使用情形後，MataData Works 發現資料集已被用於多篇研究論文、不同項目以及國家對抗新冠疫情之因應措施中。

(四) 透過資料治理及參考資料促進環境政策執行與執法 (Enabling Environmental Policy Execution And Enforcement Through Data Governance And Reference Data)

本場會議主講人為比利時法蘭德斯 (Vlaanderen) 地方政府——環境及空間規劃部門的 Steven Geirnaert 資料治理主任 (Data Governance Officer) 及 Pascal Dussart 企業商業架構師 (Enterprise Business Architect) 。法蘭德斯環境與空間部門，負責法蘭德斯地區的環境、水質、廢棄物、氣候變化及能源等領域之監測及管理，並負責制定環境相關之政策及執法，目標是推動環境永續發展，以確保生態平衡。

本次演講中，主講人分享政府如何透過資料管理，有效處理化學品所導致之環境汙染問題。在法蘭德斯長期面臨全氟化合物 (PFAS) 化學品之汙染問題，PFAS 是一種廣泛被使用之化學物質，具有耐高溫之特性，常用於防水、防油、防汙、防沾黏之產品，如不沾鍋、食品包裝、防水防汙服裝等，但由於 PFAS 無法分解，且 PFAS 產品從生產、使用到丟棄階段，皆會釋放化學物質到環境中，造成人們巨大的健康問題，包含影響生殖能力、導致癌症等風險。在法蘭德斯地區，曾發現到人們的血液中驗出了 PFAS 物質，當地政府開始意識到化學物質對人體產生危害之嚴重性，進而思考如何有效解決化學品之汙染問題。

Tackling PFAS-contaminated sites in Flanders since 2021



>6000 synthetic chemical compounds

Resistant against high temperatures, and water-, dirt-, and grease-repellent



<https://www.vlaanderen.be/pfas-vervuiling/>

https://en.wikipedia.org/wiki/Per-_and_polyfluoroalkyl_substances

圖 24：法蘭德斯 PFAS 汙染情形

資料來源：Vlaanderen Omgeving Departement

首先，法蘭德斯地方政府認為應廣泛處理歐盟 REACH 法規中所列出的高度關注物質（Substances of Very High Concern, SVHC），PFAS 僅是 SVHC 所列出對健康和環境有潛在危害風險的化學物質之一。在處理 SVHC 化學物質之方向為：減少及預防，像是逐步淘汰 PFAS 和具 PFAS 之產品、防止 PFAS 暴露於環境 3. 污染清理及廢棄物處理。他們提出的解決方案包含，制定政策、執法、許可證以及監測和研究，同時需要掌握跟多的資料，由於 SVHC 化學物質種類繁雜，需要透過資料的蒐集與共享，了解那些化學物質需要重點關注，以便執法單位有效進行檢查，另外也需要將許可證之內容資料化，使機器可讀，這樣才能使執法人員在執行的過程中，發現許可證授權之內容與實際狀況有何差異，尤其目前 SVHC 化學物質還正在研究中，有時候科學家可能已經發現某化學物質很有問題，但主管機關已經頒布許可證給業者，雙方欠缺溝通，而導致 SVHC 物質無法有效管理。

主講者提到他們制定了資料治理的框架，希望透過這些框架實現資料管理，另外他們也採用侵略性的資料治理方法（Invasive DG approach），使資料治理廣泛應用於組織中，確保參考資料品質和一致性，促進組織內部和跨組織的合作，確保資料的共享，有效解決化學品之污染問題。

Data Management Framework

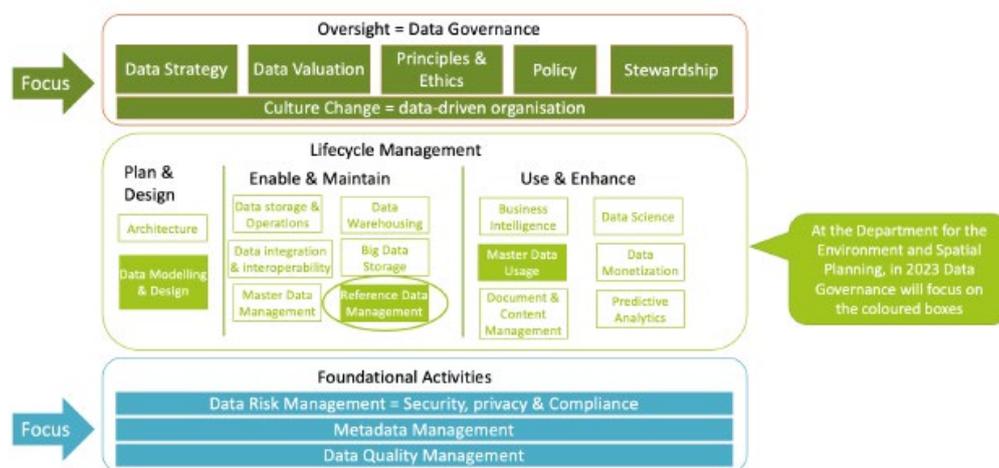


Figure adapted from DAMA Data Management Function Framework, DAMA – DMBOK Data Management Body of Knowledge 2nd Edition

圖 25：SVHC 的資料治理架構
資料來源：Vlaanderen Omgeving Departement

首先他們會蒐集不同的資料集，將不同種類的化學物質類型依照危險性和特性進行分類，製作不同的清單，並且讓不同利害關係人間能夠進行管理，並且透過利害關係人的訪談，了解大家的資料需求，從而建構化學物質的資料景觀(Data Landscape)，在收集化學物質的參考資料時，會從外部的資源（如美國國家生物技術資訊中心的免費化學物質資料庫 PubChem）結合內部的資源（如歐洲化學物質管理局的化學物質資料庫 ECHA DB）來蒐集資料，相關資料列表都是使用 PDF 文件，表示人們必須要研讀這些文件才能獲取所需的訊息，因此希望透過非結構化的資料呈現，製作新的化學品資料集。

另外，他們也導入了主資料管理方法，希望建立一個統一可信賴的主資料庫，讓來自不同領域背景的人員，如執行單位、研究人員可一同運用。在執行上，因資料庫來源眾多，像是 PubChem 中有超過 3 億種化學物質，ECHA 註冊了 2 萬 7 千種物質，還有每天會更新的物質註冊解析，資料會有重複問題需要解決，另外不同人對於化學物質的理解不同，因此需制定名詞定義、相同的名詞和符號。

最後提到，他們同時導入支持主資料系統的商業架構技術(Business Architect)，建立跨政策領域的資料清單治理機構，並且支援即將推出的環境十字路口銀行(Crossroads Bank for the Environment)，希望將許可證、執法、研究等方面的環境資料整理起來。並且透過價值流(Value Stream)分析 SVHC 主資料的運作，透過視覺化工具，了解整個主資料的運作步驟，並且進行優化，改進品質和效率。目前他們已經透過價值流，協助執法的進行流程：

1. 展開調查－優先調查可能具有 SVHC 的場所
2. 計畫調查－關注前 10 名的 SVHC 化學物質
3. 紀錄調查結果－記下 SVHC 和處理方式

4. 文件化—關注 SVHC 許可證與調查結果間的差異，找到改進的方法。

Identifying use cases supported by the MDM of SVHC (ZZS) through value stream stage analysis

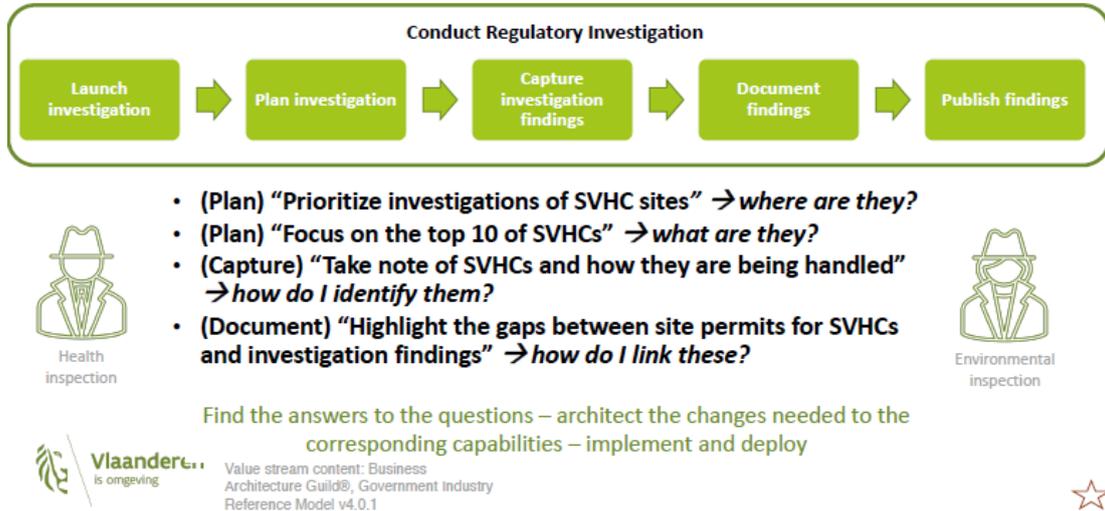


圖 26：如何透過資料流支援 SVHC 之主資料分析
 資料來源：Vlaanderen Omgeving Departement

(五) 讓資料發揮效果：如何不破壞您的資料治理計劃 (Keynote: Making Data Work - How Not to Wreck Your Data Governance Programme)

本課程的主講人為來自 Global Data Strategy 的 Nigel Turner。Global Data Strategy 是國際型資料管理顧問公司，專精於企業資料策略、資料治理、大數據分析與商業智慧、主資料管理 (Master Data Management)、資料品質、資料架構及資料模型等主題。Nigel 為 Global Data Strategy 的總資訊管理顧問，曾於英國電信 (British Telecommunications) 推動資訊品質改善計畫，於 Trillium Software 等企業擔任 IT 及資料顧問，並且為 DAMA UK 之理事。

在本次的演講中，Nigel 分享資料治理如何在企業中有效落實之經驗以及分享一些真實的案例說明企業資料治理計畫失敗的原因。依照調查顯示指出，90% 的企業首次在嘗試及執行 MDM 項目時都會失敗，導致失敗的原因通常是因為資料治理的框架不適合企業本身。另外只有 24% 的 MDM 計畫成功結束，有 76% 的 MDM 計畫會失敗。

成功的 MDM 必須結合資料、業務流程和管理，才能確保資料治理有效發揮作用，例如，不應由 IT 人員決定資料的是否正確，而應由業務人員，實際執行者建立業務規則，來進行資料治理。過去在英國的伯明翰機場中，便發生航空公司因未建立業務規則，委託外部的系統開發商開發系統，使用錯誤的重量假設，導致發生飛機載重計畫之錯誤，而導致飛機起飛事故。

而資料治理會發生的陷阱，首先是無法將資料治理計畫與組織目前嘗試要做的事情聯繫起來，通常會注定資料治理計畫的失敗。必須了解組織內部是靠什麼業務運作，確保資料治理的目的和實現組織目標保持一致性，例如提高收入、降低成本、引入客戶忠誠度計畫等。

其次，則是缺乏高階管理層的支持，導致資料治理無法被有效的推動，必須瞭解到組織的高階管理層對於財務、人力資源相當了解，但是很少人接受過資料治理的培訓，因此資料專業人員，應適度培訓高階管理層資料治理的精神與願景。

最後，則是讓資料治理計畫弄得太複雜及困難，必須長時間才能看到資料治理的效益，導致組織對於配合資料治理計畫意興闌珊，建議要多與組織內部不同

單位溝通訪談，製作吸引人的文宣，讓不同單位的人願意分享目前資料的優勢和劣勢，以及對於資料治理的期待。

在與組織內部成為溝通時，建議製作資料治理的路徑圖，不要拿出幾百個單獨的資料治理甘特圖，這樣會嚇跑大家。資料治理的路徑圖，應嘗試主題化，列出哪一些主題應該優先執行，以及說明中長期的資料治理策略。

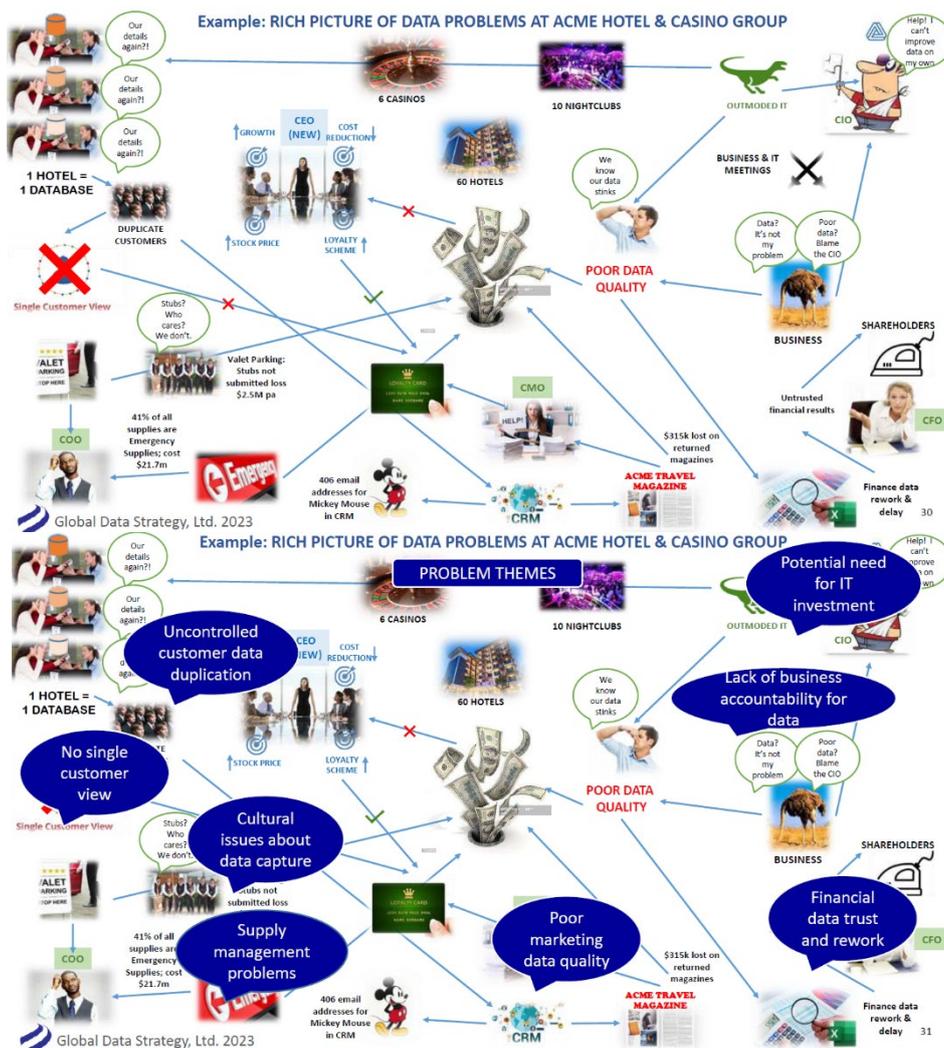


圖 27：資料治理路徑圖

資料來源：Global Data Strategy

另外 Nigel 也提醒資料專業能員，推動組織資料治理之建議，包含：

- 1.在與組織內部溝通資料治理時，要記得作為資料專業人員，工作目的是為非專業人士簡化資料，因此在討論時不要使用專業用詞，因為會使人們覺得資料治理與他們無關，無法效溝通。
- 2.要確保採用的資料治理方式與目前的趨勢相當，例如機器學習、人工智

慧、資料網格、資料分析、資料可觀察性等。

- 3.必須設定資料治理的優先事項和活動，先執行有機會快速獲得成功的項目，建立組織信心，為長期成功奠定基礎，

(六) 閉幕演講：資料及人工智慧將會為未來帶來哪些啟發？（Plenary Keynote: What the Future Will Bring Based on Data and Artificial Intelligence）

本場次的演講為第二天研討會的閉幕演講，主辦單位請到美國 McKnight 顧問公司的總裁 William McKnight 來為大家進行演說。分享目前常見的幾種 AI 技術應用，包含自動生成圖片程式、無人機、智慧醫療、基因改造技術、自駕車、ChatGPT 等。並分享了未來因 AI 技術將面臨的風險包含機器人取代人們的工作，導致的機會與不平等、真人模擬影像或機器生成文件導致錯誤訊息和假新聞發生、更多的網路安全問題、基因技術延長人類壽命但也面臨科學道德風險。

三、 第三天會議重點

(一) 工作坊一：資料網格及資料經緯－管理資料治理和資料品質 (Data Mesh and Data Fabric - Managing Data Governance and Information Quality)

本研討會的第三、四天為工作坊，而本場次的「Data Mesh and Data Fabric - Managing Data Governance and Information Quality」為工作坊的第一場課程。本課程的主講人為來自 Global Data Strategy 的 Nigel Turner。於本次的課程中，Turner 主要為大家介紹四個概念：資料網格 (Data Mesh)、資料經緯 (Data Fabric)、資料治理 (Data governance) 及資訊品質 (Information Quality)。其中，又以資料網格及資料經緯為本次課程的主要介紹對象，蓋此二概念為近年來資料工程業界出現的新興概念。而在介紹完此二新興概念之後，Turner 回到目前常見的資料治理以及資訊品質概念，綜合比較，希望讓參與課程的人員能夠掌握這四個概念之間的交互關係，進而可以針對未來是否於組織內部引入資料網格與資料經緯來解決組織內部的問題有所評估。

在進入資料網格與資料經緯的介紹之前，Turner 首先提出他對此二概念的整體性比較觀察，Turner 認為，若以技術與管理此二軸向觀察資料網格與資料經緯，可以發現，資料網格此一概念中，約七成為管理，三成為技術；反之，在資料經緯此一概念中，約九成為技術，一成為管理。這也是在後續概念理解上需要注意的地方。

何謂資料網格？資料網格一詞於 2019 年由 Zhamak Dehgani 提出，旨在回應因資料倉儲 (Data Warehouse)、資料湖 (Data Lake) 等服務而出現的資料集中化現象。其以去中心化為基礎，將責任分散於最靠近該等資料之人員，俾以支持持續性改變及成長。

Zhamak 提出之資料網格，有四個核心原則：

- 領域所有關係 (Domain Ownership)
- 資料即產品 (Data as a Product)
- 自服務資料平臺 (Self-serve Data Platform)
- 聯邦式計算治理 (Federated Computational Governance)

其中，各核心原則的關係又能夠化成以下的關係圖：

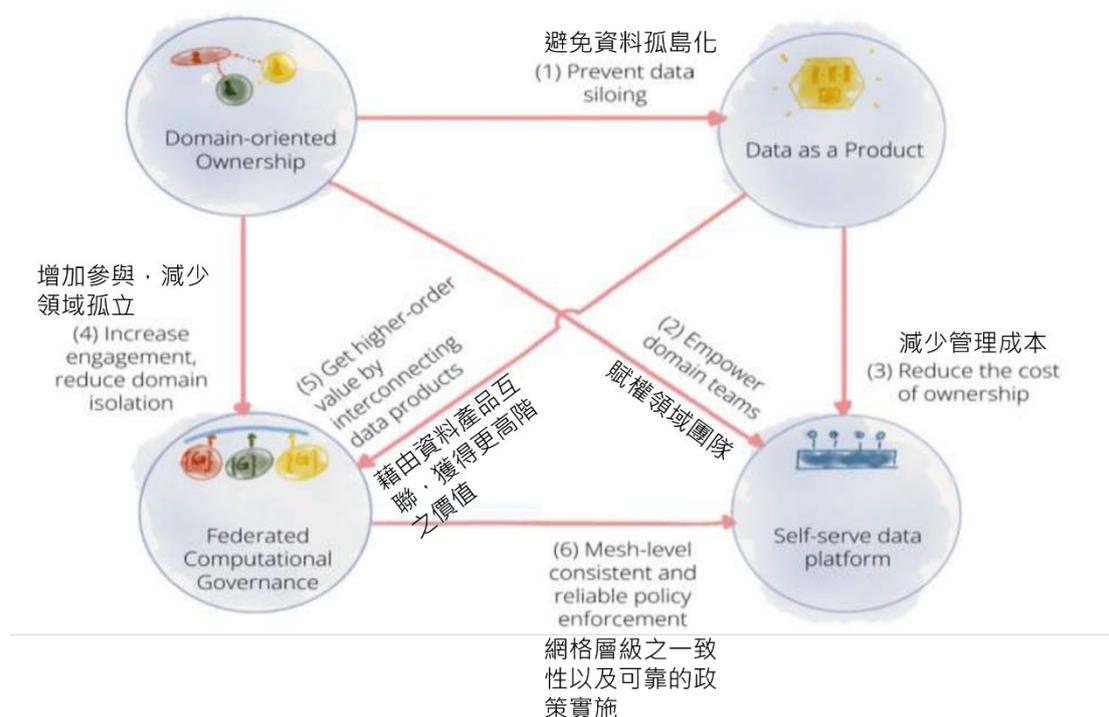


圖 28：資料網格的四大核心概念

資料來源：Global Data Strategy

資料網格主要係討論如何改變人們組織的方式，進而管理資料。藉由強調資料即是產品而非營運過程的副產品，資料網格肯認，在數位化組織中或希望成為數位化的組織中，資料至關重要。資料網格學說中，將資料的責任（例如治理、品質等）置於資料產生者的身上。

何謂資料經緯？與資料網格不同，資料經緯此一概念的出現並沒有明確的單一起源，許多重要的科技公司、商業顧問公司都對資料經緯有所定義。知名的資訊顧問公司 Gartner 對於資料經緯下了以下的定義：「資料經緯係一種新興的資料管理設計，其允許了橫跨各式各樣資料來源的輔助性資料整合與共享。」

為了整合各式各樣不同的資料來源，資料經緯希望發展一種「連接性虛擬纖維」，其使用各種不同的技術，俾以加強資料來源與資料消費者間的連結，以及資料庫與其他資料庫之間的連結。

根據所欲解決的問題，資料經緯概念中有著四個核心能力，分別為：元資料管理與資料目錄（Metadata Management & Data Catalog）、主資料與參考資料管理（Master & Reference Data Management）、資料整合與視覺化（Data Integration & Virtualisation）、資料傳遞（Data Delivery）。若以整個組織的資料架構觀之，資料經緯主要係作為資料來源與終端使用者服務之間的橋樑，其整體架構圖如下圖所示。

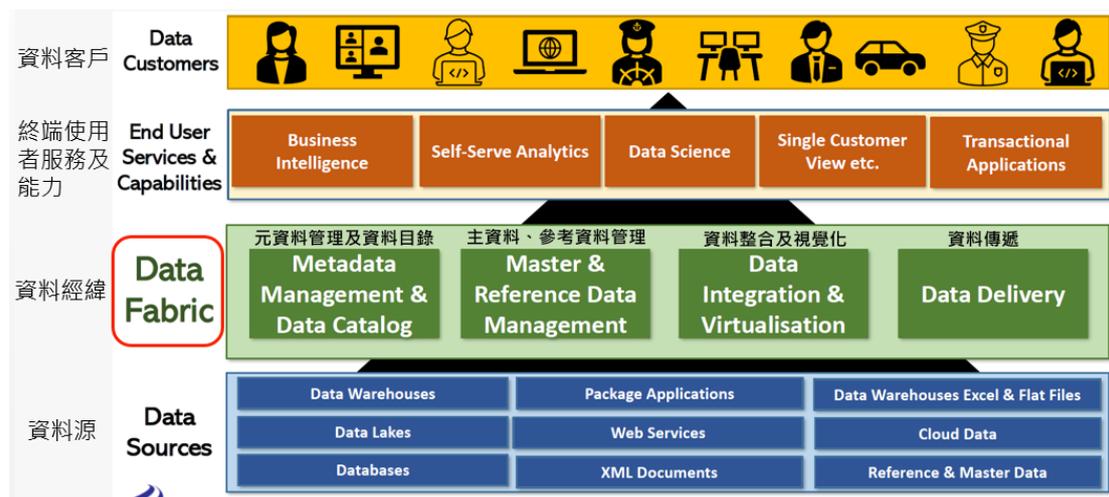


圖 29：資料經緯與整體資料架構

資料來源：Global Data Strategy

Turner 認為，資料網格與資料經緯為目前資訊產業界的「熱門趨勢」，未來若持續發展，對資料治理與資訊品質有所助益。Turner 將四個概念彼此間可以整合發會綜效的特性整理如下圖。

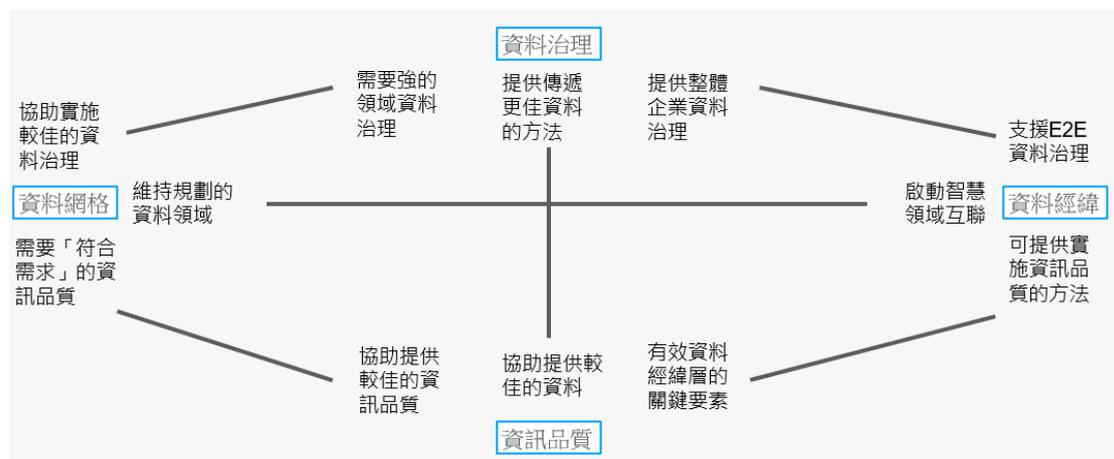


圖 30：四個概念的綜效

資料來源：Global Data Strategy

(二) 工作坊二：元資料品質：忽略你的風險 (Metadata Quality: Ignore at Your Own Risk)

本課程邀請了《Data Stewardship: An Actionable Guide to Data Management and Data Governance》一書的作者 David Plotkin 為與會者講解元資料品質。

元資料 (metadata) 之字元為 meta 加上 data，所謂 metadata 即是關於資料的資料，亦有人翻譯成後設資料、中介資料等。根據美國國家資訊標準組織的定義，元資料又可分為描述性、結構性與管理性³。由於現代社會，各組織內部分工逐漸專門且複雜，各部門所使用的資料也隨著各部門業務與分工的細碎化而有成為資料孤島的現象發生。因此，為了各部門之間的資料共享與溝通，品質良好的元資料——即關於資料的資料——的重要性益發重要。在本次課程中，David 即針對元資料之資料品質進行介紹。

David 認為，元資料品質 (Metadata Quality) 係確保資料品質的關鍵，其原因在於，元資料品質，將使得組織能夠在資料品質變差時，即時發現資料品質變差，並且知道資料品質為什麼變差。為了要針對元資料品質加以評估，組織應該要針對組織內部的元資料加以剖析，David 以下圖說明元資料品質評估流程以及與得出元資料品質規則 (Metadata Quality Rule) 表較之結果的流程圖。

³ National Information Standards Organization (2004), Understanding Metadata, <http://www.niso.org/publications/press/UnderstandingMetadata.pdf>

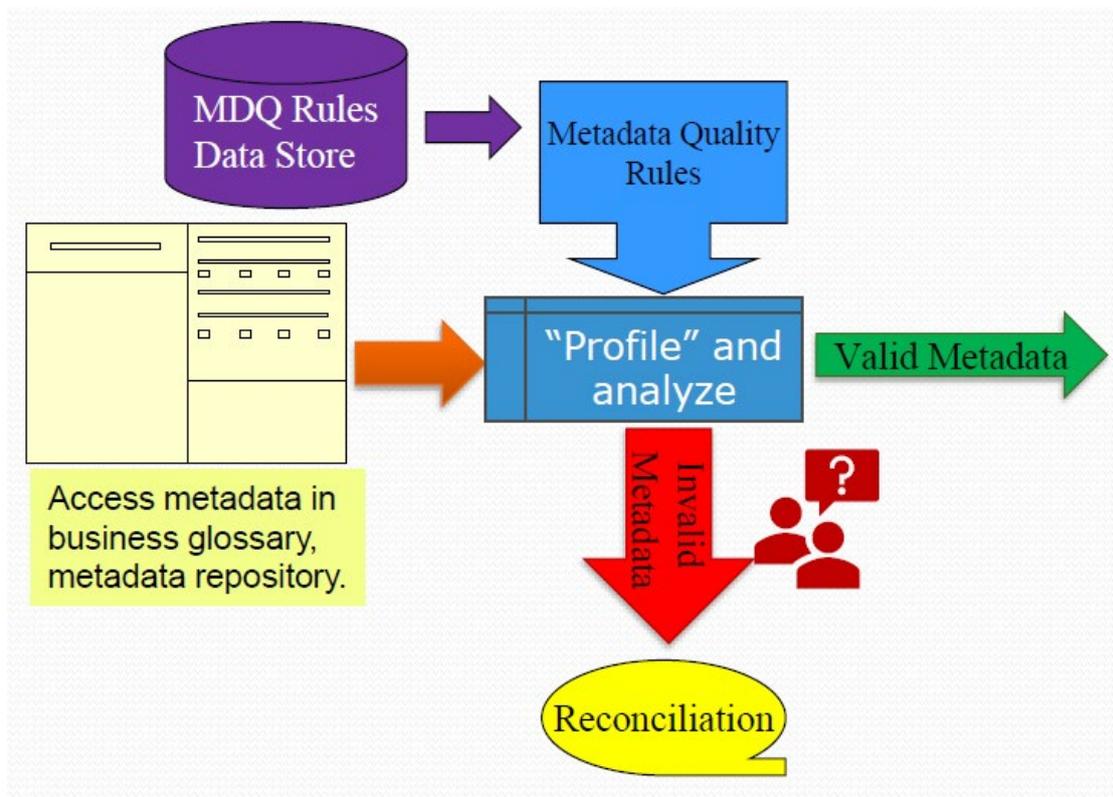


圖 31：元資料品質評估流程

資料來源：David Plotkin

元資料品質規則可以分為三種類型：

- 文字型：定義以及其他敘述性文字。組織成員必須遵守組織定義的標準，俾以產生明確且不模稜兩可的定義，限縮定義可被詮釋的空間。
- 結構型：連接了那些實體/物理位置。於結構型的元資料品質規則中，應敘明各事項與其他事項之間的關聯，並且針對各該關係是否被允許、必要、或禁止加以敘明。
- 屬性型：於屬性型的元資料品質規則中，應註明各屬性的數值之限制係如何決定（包含是否為選擇性的或強制性的）。

若於評估過程中，元資料與元資料品質規則比較的結果顯示元資料品質不符合元資料品質規則，則該等元資料係不合格的元資料。此時，應進入調和（Reconciliation）程序，該程序可見於下圖：

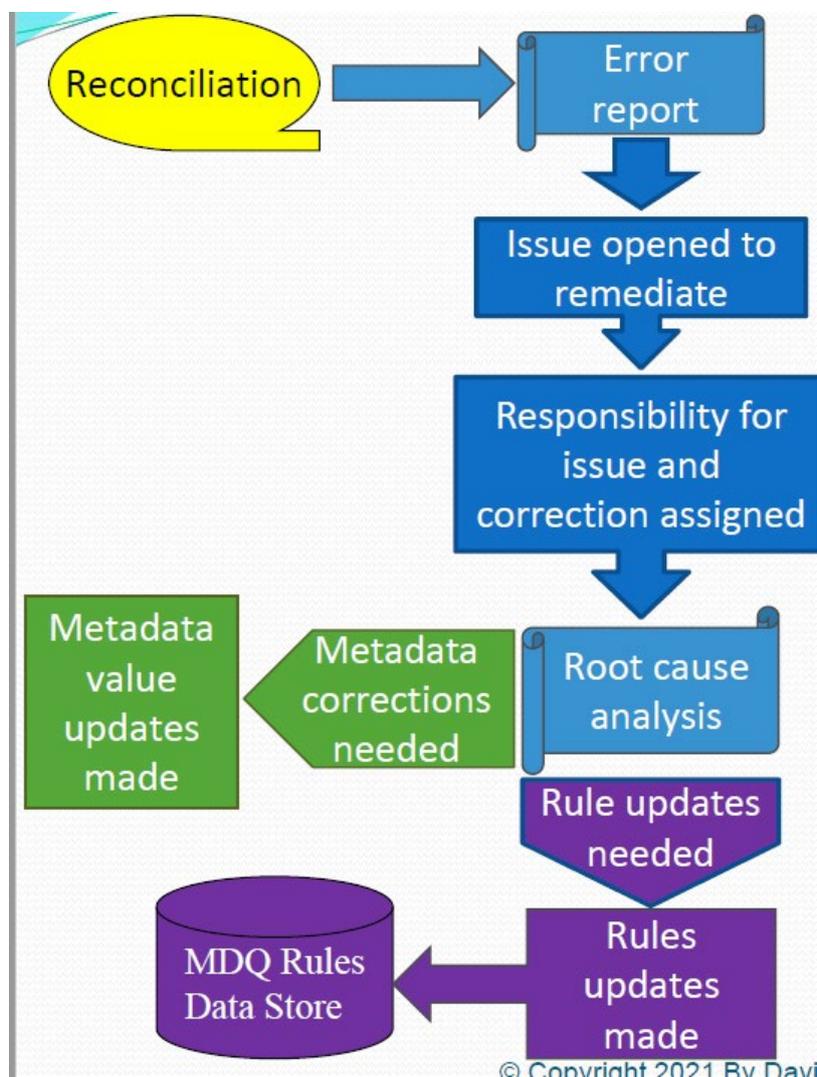


圖 32：元資料調和程序

資料來源：David Plotkin

除了組織內部的規則之外，David 也強調組織應注意各個主權範圍內的資料相關規範，例如歐洲個人資料保護法（General Data Protection Regulation, GDPR）、美國加州消費者隱私法（California Consumer Privacy Act, CCPA）以及加拿大個人資料保護與電子文件法（Personal Information Protection and Electronic Documents Act, PIPEDA）等。

各部法律中，規範有所異同，一般而言，資料主體有以下的權利：

- 有權知悉公司持有哪些關於資料主體之資訊
- 有權要求公司在滿足某些條件的情況下刪除特定資訊
- 有權退出該資料庫之買賣
- 有權同意或拒絕特定資料使用

組織必須隨時注意國際上資料相關法令之規定，並根據該等規定修正現行的元資料規則。

四、 第四天會議重點

工作坊三：資料管理完全指南（Complete Guide to Data Stewardship）

本次課程主講人為 David Plotkin，目前擔任美國 Union Bank 資料管理經理，過去曾擔任美國金融保險公司 AIG 全球資料治理主管，並在不同公司擔任資料顧問。另出版《Data Stewardship: An Actionable Guide to Data Management and Data Governance》書籍。

本次會議中，David 提到了資料管理（Data stewardship）在組織扮演的角色並且分享資料管理之方式，包含資料品質（Data Quality）、資料倉庫（Data Warehouse）、主檔資料管理(MDM)，以及分享實際管理資料時組織可能面臨之挑戰及因應方式。若組織不執行資料管理，將導致不同部門使用自己的資料定義，導致資料品質不一，不同部門間的資料要進行 MDM、資料倉庫時，將會發現困難重重。

所謂資料管理是指由特定人員執行與資料相關的決策過程，其重點包含：

- 資料管理者，需管理並理解其負責之資料及資料存放地點，確保部門內部及跨部門所用資料定義及用法保持一致，簡化資料決策的過程及解決問題、了解資料品質的工具。另須留意資料管理者應非新聘的人員，而是組織內部既有熟悉資料之人員來擔任。
- 建立良好的資料品質：包含了解目前之資料品質、品質需求及如何改善資料品質以獲得更多價值，使公司的資料被充分理解並用於預期目的。
- 為資料之決策管理制定框架，透過書面、批准及重複可執行之管理流程來制定關鍵的業務用語、紀錄分析確定優先處理的資料問題及資料品質、優先支持資料治理的項目等。
- 正式的組織結構，組成資料治理組織，使資料管理融入企業的流程中，不同的業務可指定資料管理員，對於資料的責任追究必須明確清楚，並且制定額外的獎勵和認可機制。另提醒，好的資料治理應有組織由上到下的支持與參與，組織成員應了解資料的價值。

另須留意組織內會依業務不同而有不同的資料管理者，資料管理者間必須齊

心協力，例如業務資料管家必須考慮到企業利害關係人的利益；對於營運部門的資料管理員，必須接受營運部門的資料反饋，並提供回饋意見；而技術部門的資料管家則需要接收並回答業務資料管理員的技術需求。



圖 33：資料管理者間之合作

資料來源：David Plotkin

組織在執行資料治理時，應了解資料治理可選擇使用之方式及重點：

- 主資料管理：指管理資料的解決方案，使企業可進行分析。在執行上首先應識別主資料之定義，例如銷售額又可細分不同地區銷售額、含稅銷售額以及特定產品之銷售額，應有所區別。了解主資料存放的地點、使用主資料之目的、有哪些資料分析工具可使用，以及創建資料原則，例如 ISO 標準、GDPR。而此可透過資料治理之術語表達成。
- 資料沿襲（Data Lineage）：透過系統紀錄、幫助提供不同資料的細節，可確保資料完整性，執行資料沿襲的好處是，若未來有外部人士提出資料之質疑時，可提出資料沿襲之系統紀錄。
- 資料倉庫：制定資料轉換規則，使資料在不同系統間得以轉換。但應留意，資料倉庫汰換時與新資料倉庫間之轉換與區隔，避免資料重複問題。
- 提高資料品質：定義有資料價值的資料內容，例如企業紀錄客戶資料時，若特別紀錄喪偶與否，此是否企業有價值。另外透過統計「資料現況描

述 (Data Profiling) 結果」, 可更好的了解及評估資料集, 進行資料分析或資料清理。

- 資料安全：界定敏感性資料，對於資料之安全等級進行分類。另可執行隱私及合規性步驟，首先應成立隱私合規小組，制定組織政策，將資料進行分類，決定應受監管之資料分類，並將資料分類文檔紀錄到組織之資料詞彙表中，並由 IT 人員設置應經批准始訪問之資料，並開始實施並維護整個資料隱私及合規工作流程。
- 資料治理路線圖 (Data Governance Roadmap)：資料治理會隨著時間的推移, 常見的路線圖包含總體路線圖(通常約 18 個月)、工具實施路線圖、新資料路線圖 (onboarding data roading)。

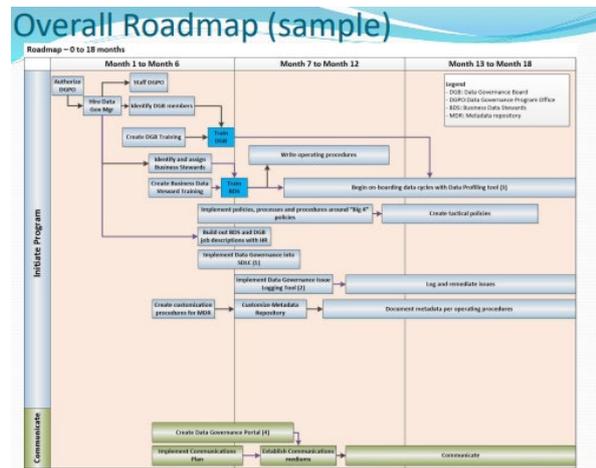


圖 34：資料治理路線圖－總體路線圖
資料來源：David Plotkin

肆、心得及建議

- 一、隨著企業的資料量日益增多及複雜化，資料之處理與存放不再侷限於企業本地端，而是分散在各處，例如雲端、邊緣裝置、軟體即服務（SaaS）等。由於資料分散各處且複雜，因此如何管理資料生命週期，確保資料之品質、完整性、即時性、安全性、隱私保障即相當重要，因此企業應制定符合自身之資料治理框架，框架重點包含目標及策略、組織結構與責任、資料流程、資料品質一致性、資料政策等。
- 二、資料治理常見之概念或工具，包含主資料管理、資料沿襲、資料管道、資料可觀察性、資料目錄、資料網格、資料經緯、資料品質等。目前資料治理的技術工具相當多元，企業應注意不同工具間的整合，並要留意採用之資料治理方式與目前趨勢是否相當，例如引入機器學習、人工智慧在資料治理上。
- 三、目前國際上已有許多資料治理之成功案例，並有許多資料治理之顧問公司協助達成組織之目標，例如：
 - （一）英國為整合國內之健康資料，特別成立英國健康資料研究機構，使產官學界皆能參與合作，促進健康科學領域之發展，並釋出八百多筆資料集，供研究人員、企業、組織能夠以付費或免費之方式申請取得。未來我國若欲發展健康資料之共享平台，建議可參考英國之作法成立獨立之機構，統籌健康資料之整合與取得方式。
 - （二）比利時法蘭德斯的环境及空間規劃部門透過建置 SVHC 化學物質的資料管理系統，依照化學物質的危害程度和特性區分不同的資料集，協助研究人員到執法單位間，可有效利用資料進行溝通，優先調查可能具有 SVHC 的場所，以利執法人員更有效率的進行環淨污染源的稽查。對於未來我國處理環境污染問題時，乃值得參考之作法。

(三) 國際非營利性組織 CABI 透過農業資料共享計畫，以增加農民收入及農作物之產量，而推出資料共享工具包，以 FAIR 原則—資料可尋找性 (Findable)、可訪問性 (Accessible)、可互操作性 (Interoperable) 及可重用性 (Reusable) 設計 7 種情境，提供組織成員學習，透過強化組織成員之知識與技能，提供資料擁有者與資料所有者間，得以交流對話。CABI 所提出之資料共享工具包作法，或可成為我國未來推動數據公益生態系之方向，以 FAIR 原則設計不同情境，以知識教育方式，使人們願意參與資料共享。

(四) 荷蘭發展銀行是如何進行資料治理，並且賦權各部門的大小職員，使整個組織「動起來」進行積極的行動性資料治理。

四、成功的資料治理，應確保資料治理目的與業務目的一致。資料治理推動者必須有良好的規劃能力以及溝通能力，掌握組織內部其他部門成員的工作目標以及想法，得到組織內部其他部門成員的協助與支持。同時應留意使用易懂的語言或圖像進行溝通，在規劃資料治理路徑圖時，宜先執行短期內能看到效益的項目。

五、本次研討會第三天及第四天是以工作坊形式進行，以第三天工作坊之教學方式為例，講者 Turner 在介紹完每一個概念後，會安排小組討論及分享活動。如此將小組討論穿插於課程中間，使得來自各個不同產業的資訊人員，能第一時間分享各自組織面臨的問題以及解決方法，更能促進整體課程的學習效果。建議未來若有安排工作坊之課程，亦可採取相同的形式，以利業務推廣。

六、對數位發展部未來資料治理之心得與建議：目前數位發展部正積極推動數據公益機制，希望鼓勵民眾及數據持有者開放無關個資之數據，供公益目的的使用。在推動數據公益時，應留意數據持有者本身是否已導入資料治理機制（包含資料治理工具、健全的組織、資料治理政策等），有基本的資料治理基礎後，才有進一步發展數據公益之可能。目前較適合加入數據公益

機制之數據，包含：醫療、交通、金融消費、環保、健康、社會福利等，在推動數據公益機制時，建議區分短中長期規劃，短期內可先輔導較具成熟資料治理之數據持有者（如上市上櫃公司、公部門），作為數據公益之示範案例。中長期則可輔導欠缺資料治理之數據持有者，先解決數據持有者組織內部存在資料治理不完備之痛點，再輔導其推動數據公益機制。此外，由於資料治理同時涉及我國數位轉型發展，是以，各部會亦應共同商討如何協助產業（尤其中小企業）利用資料治理方式，導入人工智慧、大數據分析，以利加速數位轉型。

七、對 TTC 未來發展資料治理與應用之心得與建議：在本次的研討會中，有講者分享到組織在規劃資料治理時，應與組織欲達成之目的相連，並優先針對可在短期內看到成果之項目先進行資料治理。因此若未來 TTC 欲發展資料治理，建議可由各組盤點欲達成之目標，再決定是否有導入資料治理之必要，另外亦可思考是否能透過資料治理，整合資料，提供嶄新的加值服務。以檢測暨網通技術組為例，其業務包含資通訊產品及綠能元件檢測並受理 NCC 委託電信終端設備、電信管制射頻器材及有線廣播電視終端設備審驗服務，具有諸多設備之檢測資料，若可進一步透過資料治理分析相關資料，或可從中探詢可能之加值應用。而以研究企劃組為例，其業務以承接主管機關計畫為主，定期蒐集各國通訊傳播監理政策、數位治理議題，若可將蒐集之各國資料透過資料治理之工具加以分析，並加入 AI、GPT 技術，或可協助未來研究人員之研究量能，並有機會成為新的服務產品。

伍、附件：相關照片及資料



圖 35：參與人員於開幕典禮合照留影（資料來源：現場拍攝）



圖 36：研討會第一天會議開場，主講人 Jan Hendreyckx 分享生成資料對資料應用之影響（資料來源：現場拍攝）



圖 37：研討會第二天會議開場，主講人 Mike Ferguson 分享資料治理與人工智慧之關係（資料來源：現場拍攝）



圖 38：研討會第二天閉幕演講，主講人 William McKnight 分享資料及人工智慧將會為未來帶來哪些啟發（資料來源：現場拍攝）



圖 39：研討會第三天工作坊，主講人 Nigel Turner 分享資料網格與資料經緯之差異（資料來源：現場拍攝）

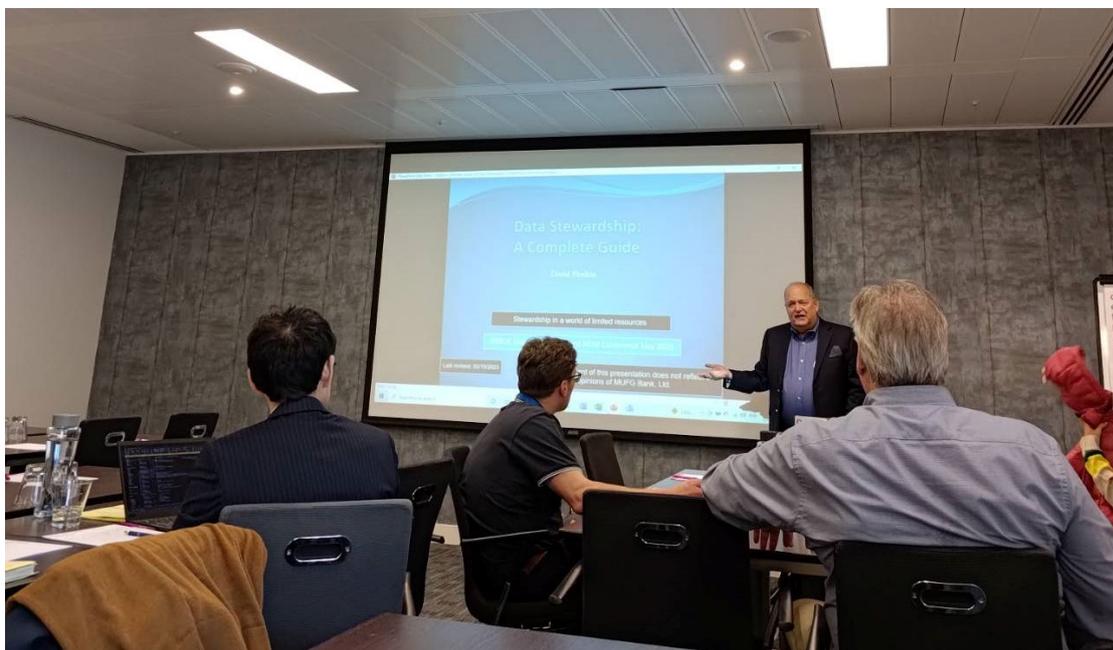


圖 40：研討會第四天工作坊，主講人 David Plotkin 分享資料管理完全指南（資料來源：現場拍攝）