

下一代資料平台是 資料網格（下）

口述／孫培然 彙整／CIO編輯室

II 自助資料服務工具準則

在資料網格（Data Mesh）中提到自助資料服務工具的準則，首先就是資料服務供應，重於擷取不再講求用 ETL 來批次轉檔；其次是資料的發現和使用，重於提取和載入；再則是發布事件流，重於中心化資料管線的資料流；最後是資料集生態，重於中心化資料平台。

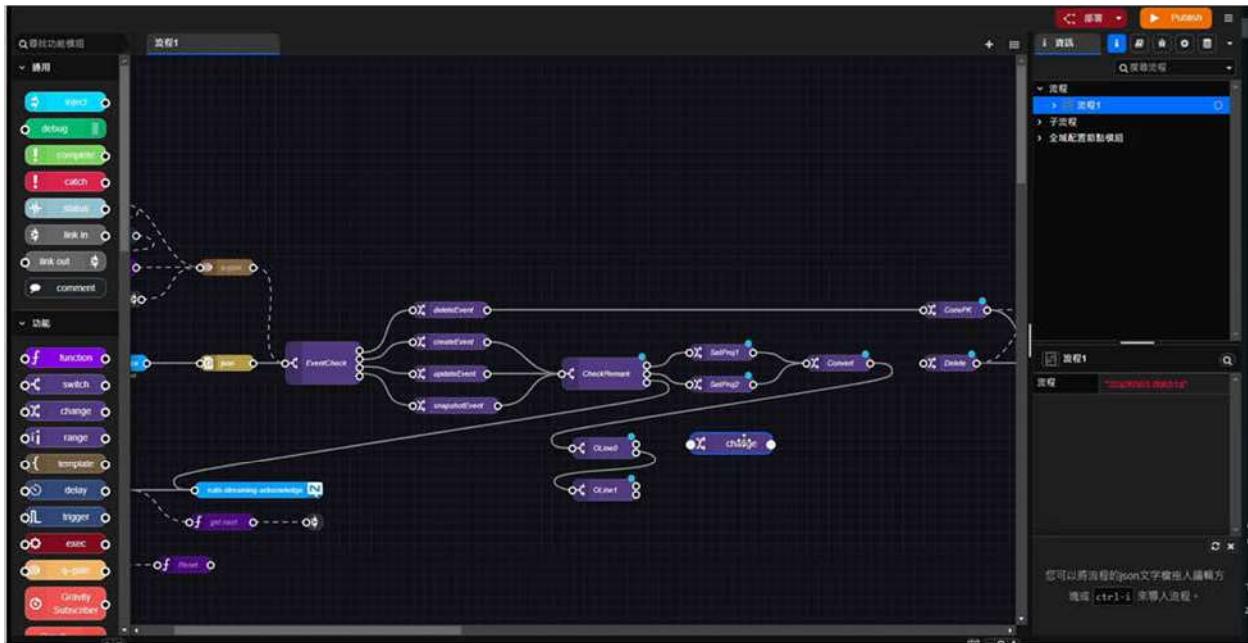
然而自助資料服務平台的建置，會因為不同結構組織而需求有所不同。主要有三種不同的使用者角色需求要考慮：

（1）潛在使用者，使用者必須能夠找到他們需要的資料產品，以自己的身份訂閱，並將資料存取到自己的領域中，他們也可以反過來從資料中創建自己的資料產品。

（2）資料產品生產者，生產者這些人希望使用自助資料服務平台來建構他們的資料產品。這包括自助計算、存儲和處理，以及建置、測試和部署的簡化方式。

（3）資料產品所有者，必須能夠長期管理他們的資料產品，包含通知現有使用者即將發生的變更、處理功能請求、發布有關重大變更的指導、管理警報和待命輪換以及管理資料產品生命週期，例如棄用和刪除。

接下來是要透過聯合計算治理，避免形成各單位的資訊孤島。首先是要資料治理去中心化，將領域資料自主權下放給使用者，遵循分散式系統架構，由獨立的資料產品組成架構量子，形成每個資料產品具有獨立的生命週期，通常會由該領域團隊，透過自助資料服務工具來建置和部署。



圖一、設定資料管線

再來就是要以資料即產品的思考，透過標準化保證互通性，早期的中心化是把實體集中，但資訊架構演進到至今則是將它抽象化成邏輯的集中，不再依靠實體來中心化，而是要依靠開放標準（Open Standard）來中心化。雖然架構量子分散在各地，但是資料產品的位置、API、元資料、所有權、文件、資料樣本都要有所規範及遵循，才不會久而久之沒有治理變成資訊孤島。

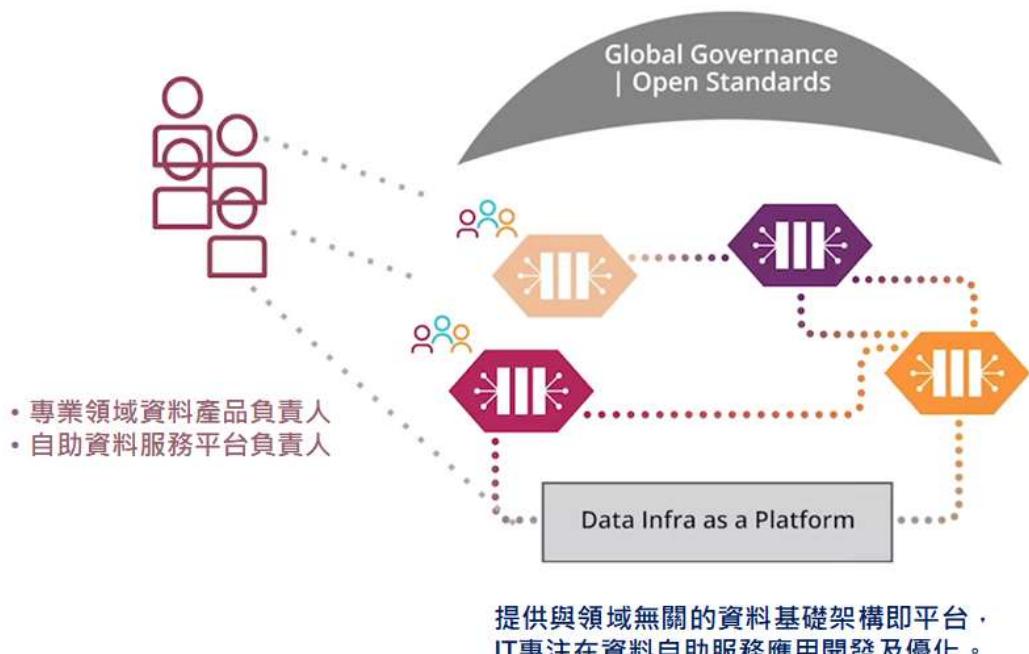
最後透過「自助資料服務基礎建設平台」的概念，進行自助資料服務，透過去中心化及領域自治，支持全域標準化的互通性及動態拓樸，希望做到 Low-

code/No-code 來設定資料轉置的方法，產生不同新資料的管線。以圖一為例，提供視覺化平台工具，以拖拉的方式來設定資料管線，就像前面提到的，把來源資料的西元年轉成民國年輸出，使用者也不用撰寫程式直接拖拉過來設定，就可以轉出使用者所要的資料，這也就是 Low-code/No-code 的概念。

其實，大家所熟悉的 BI 軟體工具（如：Power BI）就是屬於 Low-code/No-code 的概念，以使用者自助設計為導向。在以前還沒有使用 BI 軟體工具的時代，使用者需要一個報表時，就得需要花很多時間與資訊人員溝通報

表需求，以及解釋一大堆專有名詞及計算公式，而資訊人員在似懂非懂地，從 HIS 資料庫撈出資料產生一個報表。如果日後這個報表要再加幾個欄位或者更改計算公式，就得再麻煩資訊人員去更改程式，若資訊人員沒有空，就只能跟使用者說再要兩、三個禮拜，使用者就只能等好久。

若醫院有了 AI 中心以後，使用臨床資料的需求又會更多了，比如說使用者有時要 15 或 30 個欄位的資料內容，現在就不再個別客製欄位需求。而是以資料產品的概念，將資料包裹成一個產品的方式提供。假設該資料產品要有 50 個欄位的資料內



圖二、聯合計算治理的藍圖

容及相關元資料，則以交付產品的思維全部給使用者，讓使用者拿著資料產品去做相關 BI 、 AI 及創建新的資料。舉一個簡單的例子，醫院若有建構病人用藥歷程的資料產品，藥師就可利用 BI 軟體設計，自行抓取所需欄位，如科別、醫師別、診間、藥品碼等，也可以自行設計所需限制條件及調整，或許就可以紓緩資訊人員工作負荷。

II 聯合計算治理的藍圖

讓各領域自行負責生產資料產品，並要求他們有一定程度的自主權來選擇他們需要的工具，期望以最佳方式建置資料產品，

並將資料交付給使用者。倘若沒有任何資料治理政策及規範，組織將會面臨到許多資訊技術的蔓延導致形成技術債，以及必須支付高成本自助服務平台的風險。

所以聯合計算治理主要側重於在使用者的需求、資料產品所有者的自主權、業務合規性和安全性要求，以及全域性資料產品要求之間找到平衡點。聯合計算治理大致分為兩個主要任務，首先建立影響資料網格所有用戶的跨組織策略，包括資料產品的標準和資料處理要求；其次透過自助資料服務工具，提供一般共通工具來減少創建、發布、發現和使用資料產品的障礙。聯合計算

治理最適合收集廣泛的跨領域關注點的需求，例如監控、日誌記錄、存取控制、運算服務和存儲服務。這些要求必須規範提供給自助服務平台團隊，以建置必要的工具來支持資料網格。

接下來如圖二 所示，在透過聯合計算治理的藍圖，將架構量子也就是資料產品的資料統籌起來，生產者跟使用者就是領域專業的產品及自助資料服務平台的負責人，他們專門把這些架構量子去定義出相關資料管線、設定檔及元資料。另外，負責基礎架構平台的大數據工程師或資訊人員，就只需要專注地來維運資料網格的基礎建設平台即可。

在聯合計算治理下可以專業分工，每一個工作都是由不同專業的人所產生，中間可以透過資料的輸入及輸出互相溝通，這些資料的輸入跟輸出，當然就是通過 API 的方式來做溝通交互（Interoperability）。

在資料網格的聯合計算治理之下，整個醫療領域的資料所有權就會分散到不同領域的專科，比如說醫師針對高血壓跟糖尿病的病人所關心重點項目就會不一樣，每一科都有需要關注的領域跟資料。透過資料即產品的思考，每一科的領域都有自己各自獨立的單元，資料源經過資料管線處理後，再主動推送（Push）給使用者所需的資料需求。而不是像傳統方式被動地讓使用者的服務不斷地去輪詢拉取（Pull）資料。

也就是之前所提過「按讚、訂閱、分享並開啟小鈴鐺！」的概念。使用者只要訂閱了相關資料，利用分散式管線來做到即時性，就可以主動即時的提供給使用者去使用這些資料產品。這樣的架構量子就是每一個小小單體的微服務，再透過 DevOps 的持續整合、持續交付及持續部署到容器化（Container）上運行，形成資料即產品的聯合計算治理生

態系統。

II 資料網格對企業的幫助

在傳統集中式資料庫一直存在著一些問題及危機，一旦遇到高爆量時，資料平台就可能會崩潰。若將資料平台轉變成分散式架構以後，再導入領域導向的資料行為，經由自助資料服務的方式，提供資料供應管線，包含資料的清洗、組合，還有資料的所有權、隱私及安全等共通服務。

使用者只需要接受資料產品思維的設計跟管理，目的是強調在資料的所有權基礎上，我們要獲得規模化的資料交互能力，也就是說系統的應用不能集中在某個專業領域，而是要把資料處理抽象化、規模化到你只是純粹交付資料與領域無關。透過規模化的資料交付能力，就可以引入到不同領域的業務場景，並確保資料的品質、所有權跟安全性，保有技術棧的多樣化，吸引更多的資料領域人才願意投入。

反觀資料平台的演進，經歷了從資料倉儲（Data Warehouse）、到資料湖（Data Lake）、再到雲端資料平台（Cloud Data Platform）的過程，在這數十年以來，企業耗費巨資建設的資料平台卻難以獲得預期



專欄專家孫培然博士，私立醫療院所協會醫院資訊暨智慧醫療發展促進會會長，也任職於中國醫藥大學附設醫院資訊室副主任。

收益。因此，以微服務的分散式架構思維，正在取代單體的集中式架構，在資料領域正在引發一次典範式變革 - Data Mesh。

我預估資料網格大概三到五年後，就會受到相當高度的重視，套用一句古代兵法名言：「兵馬未動，糧草先行」的概念，將不同的科別所需要的資料產品，事先準備好放在不同科別的資料中台上，隨時等著使用者將資料提取/獲取到自己的領域中相關應用及研究分析。這也是在微服務架構中，所談到的事件驅動的觀念，而從被動化為主動地即時提供資料產品，以資料驅動（Data-driven）來帶動醫療服務品質的提升。