

4.0的基石—在利用數位系統以提高品質時，數據的品質將直接影響結果的品質

◎李麗女 編譯

1月時我預先註冊COVID-19疫苗接種的預約，從小就患有慢性疾病，使我面臨更高的併發症風險，所以當電子郵件最終到達時，我很高興我收到了一個連結以註冊我的時間段。（我很高興這不是網路的釣魚電子郵件，因此我以閃電般的速度點擊了該連結。）

但是有一個問題：實際上並沒有任何預約可用，但這並沒有阻止我在醒著的每個小時左右嘗試連結，不僅花費時間而且花費情感能量，而且對每一個死胡同越來越失望。

電子郵件告訴我現在是我的時間，所以我的預約必須在那個系統的某個地方！我的決心堅持但最終卻是徒勞無功的搜索。

幾天後，同一個連結將我帶到了一個新網頁。最後，我能夠輸入一個日期來請求預約。我現在的公司提供軟體和數據品質保證，所以當然我做的第一件事就是要求預約日期為2016年1月1日，看看會發生什麼。有什麼比提前做好準備更好的方法來戰勝大流行病呢？

不過，運氣不好：該日期沒有可用的預約。該系統讓我可以選擇在從那時到2026年12月31日之間的4,017天中的任何一天申請疫苗接種，因此肯定會在下週甚至下個月提供一些疫苗，但是什麼都沒有。

最終，我找到了另一種預約疫苗接種的方法：完全繞過這個系統。他們沒有嘗試更複雜的人工智慧(AI, Artificial Intelligence)/機器學習(ML, Machine Learning)系統來為人們配對正確品牌的疫苗和接種疫苗的最佳時間，因為他們將遠遠達不到該目標。

為什麼我的經歷如此令人沮喪？為什麼我要尋找不同的方法來完成這項工作？為什麼疫苗分發工作沒有做好透過AI或ML以推動具有價值的準備？因為我遇到的每個問題都有一個根本原因：數據品質的問題。

什麼情況下是足夠好？

數據品質在很多方面是具有挑戰性。問題可能會在很多地方被引入而惡化，並且隨著人們對數據的操作和交互使用（或讓它積聚灰塵）隨著時間的推移而復雜化。

例如，打錯字或手指滑落可能會影響入口點的數據品質。軟體的用戶界面可能被所編寫的程式辨識為不合法的數值（例如我在嘗試接種疫苗時遇到的日期選擇器），或者可能沒有足夠的驗證。入口點可能被所編寫的程式接受數據庫尚未準備接收的數值，被處理的數值之方式可以使得它們是為有用的或無用的。通過儀表板或商業智慧系統(BIS, Business Intelligence System)訪問數據的業務用戶可能會為其解決的問題選擇錯誤的欄位，或者誤解該欄位實際代表的內容。

此外，業務人員對數據品質的看法與資訊科技的IT人員並不相同；如果你站在業務面，你會問：“我可以使用這些數據嗎？我相信嗎？”如果你站在技術面，你會關心數據庫資料是否存在正確的約束或是有觸發器，或者是否滿足完整性、一致性和唯一性的要求。你很自然地適應了描述“好”之實際含義的不同觀點和語言。更複雜的是，並非所有人都對“好”這個數據有相同的要求。

想想客戶的記錄，營銷部門可能需要確保電子郵件地址是屬於目標市場的真實人物聯繫的資料；客戶主管可能會關心客戶記錄是否代表有興趣購買更多產品的人；產品經理可能只關心焦點小組成員的客戶記錄是否準確、完整且沒有重複性。我可以使用它嗎？我可以相信它嗎？這些數據適合使用嗎？關鍵利益相關者將對這些問題有自己獨特的答案。弄清楚“好”是什麼意思——並確保數據品質足以滿足每個人的需求——不必太過要求，也不需要花數百萬美元投資於數據長(CDO, Chief Data

Officer) 和專門的工作人員身上。

在任何規模的公司，你可能都覺得自己無力成為關注數據品質變革的擁護者；但是，無論你的數據品質計劃是企業範圍內的還是基層的，你個人都可以採取措施來幫助你（以及你的CDO，如果有的話）以提高組織的數據品質：

- 像財務長(CFO, Chief Finance Officer)一樣思考——當數據品質提高時，您希望看到哪些業務成果？為什麼你需要更好的準確性、一致性、完整性或有效性呢？不要僅僅為了使數據更好而投資於提高數據品質。數據品質還應推動特定的、可衡量的業務品質成果。
- 應用柏拉圖原則——關注最重要的數據來源和數據流程的數據品質。對於許多組織而言，只有一小部分數據生態系統驅動了大部分業務價值。但首先，你必須找到它。
- 提高可見度——要找到它們，數據流程必須對使用它們的業務人員和工程師具有可見性，並且必須對概念和高級流程有共同的理解。映射傳入的原始數據、清理和處理數據的任務、處理過的數據出現的報告或其他位置以及利益相關者在檢查它們時提出的業務問題之間的連結。
- 縮短可解釋性的時間——共享數據字典是必不可少的，但它們往往難以理解。確保任何處理數據的人都能輕鬆了解數據的確切含義、數據是如何收集或創建的，以及誰對數據負責。
- 消除單向的車道——能夠返回並從錯誤中恢復很重要，因此請確保保留原始數據、乾淨數據和處理過的數據。特別是對於需要能夠在某個時間點獲取數據之簡要印象的審查，請確保你的流程不會刪除或覆蓋以後可能需要的數據。一些數據湖泊內置了這些功能。
- 不要存檔所有東西——儲存很便宜而且你現在可以收集的所有數據中肯定會有驚天動地的見解……當你有時間去挖掘它時。儘管保存一切可能很誘人，但不要這樣做。
- 建立精實管理——了解管理人員和數據交互方式的規則至關重要。誰可以在什麼情況下以及何時與你的數據進行交互作用？數據來自哪裡？為什麼要生產？它是如何被監測和被控制的？
- 認識到數據品質軟體不是立竿見影的勝

利——如果沒有其它的要點，不要掉入決定透過購買具有重大承諾的昂貴系統，來提高數據品質的陷阱。——尤其是對與數據交互作用的人有意義的數據管理流程——軟體將只是一種安慰劑。

始終與數據品質有關

以下是這些如何組合在一起的例子：我一直在與一個想要遷移和測試近200份報告以及生成它們的自動化任務的組織合作。透過從傳統BIS轉變為更靈活和可維護的BIS，該組織希望能夠更好地控制數據品質。我們與其墜入減緩數據品質惡化，不如讓數據流程更加可見，並從業務問題回溯到支持它們的系統部分。

採用這種方法可表明，組織數據生態系統中只有15份報告和一部分數據來源是必不可少的。在過去的幾年裡，我一直在幫助各種規模的組織為數位轉型計劃做好準備。這些組織渴望實現自動化，甚至更渴望使用數據科學和機器學習來催化持續改進。但是，如果在數據品質和數據管理方面沒有堅實的基礎，他們的倡議計劃之實現必然會低於預期或持平。數據品質是基石，當你利用數位系統來改善品質時，數據的品質將直接影響結果的品質。你無需投資數百萬美元即可從已改進的數據品質中受益，但你確實需要將其視為業務問題而不是技術問題。

作者：

Nicole Radziwill is senior vice president of quality and strategy at Ultronauts Inc. in New York. She holds a doctorate in technology management (quality systems) from Indiana State University in Terre Haute. She is an ASQ fellow and an ASQ-certified Six Sigma Black Belt and quality manager. Radziwill is the author of Connected, Intelligent, Automated: The Definitive Guide to Digital Transformation and Quality 4.0 (Quality Press, 2020) and the former editor of Software Quality Professional.

資料來源：

Quality Progress June 2021 Page 46-48
Reprinted with permission from Quality Progress ©2021 ASQ, www.asq.org
All rights reserved. No further distribution allowed without permission.

