

frilly / June 19, 2017 03:29PM

### [從DAMA出發，一個指標庫到底是如何煉成的？](#)

在數據管理領域，我們通常將數據分為：主數據、交易數據、參考數據、元數據和統計[[url=http://www.finereport.com/tw/](http://www.finereport.com/tw/)]數據分析[[url](#)]（指標），指標是[[url=http://www.finereport.com/tw/](http://www.finereport.com/tw/)]BI系統[[url](#)]裡面核心的概念，是一個企業數據運營關注的核心數據，一般以KPI和報表的形式體現。

從實踐來看，一個企業要進行數據治理，涉及了架構、安全等諸多層面，但最迫切的是提升數據質量，其中指標質量則是重中之重，一般業務上90%以上關於數據的疑問都從指標的質疑開始，只要你從事數據相關工作，就應該深有體會。

「這個指標好像跟業務發展實際不符，快去查查」，估計這是報表取數人員聽到的最多的一句話了。

下文就來談談如何從根本上去提升指標的數據質量，即實現指標的標準化，作為一個數據管理人員，不管你有多少能力，曾經解決了多少問題，當過多少回救火英雄，都應該從更為長遠的角度來思考這個問題。

指標標準化的核心價值在於實現「書同文，車同軌」，即通過針對指標的一系列管理過程，去提升指標準確性、一致性、敏捷性及開放性。

DAMA將數據治理放到核心地位，指標的標準化就是個典型的數據治理問題，治標是容易的，治本的代價則太高，但如果要實現進階，還是要站的高一點，多思考一下，想想是否有更好的方法，就從筆者多年前做過的指標標準化項目開始吧，分為[[b](#)]組織保障、報表梳理、指標整合、實現方式、功能架構、可視化引擎及管理流程[[b](#)]等七個方面。

#### [[b](#)]1、組織保障[[b](#)]

指標庫這類數據管理項目，或稱BI項目，一般業務部門參與的力度是不大的，這是大多BI項目實施效果不佳的一個深層次原因。

DAMA提到要實施數據治理活動，跨部門的數據治理委員會等是關鍵的組織，的確是這樣，指標跟全公司每個單位都相關，對於其進行規範化改造當然應該獲得大家的一致同意。

可惜的是，大多企業沒有這個理想條件，也不會有數據治理委員會，在數據還未成為真正的實質性資產前，比如納入財務部的資產目錄，很少有企業會設立這個數據組織，因為效益不明顯，因此，哪個企業都不大可能為指標出一個規範並且通令全公司貫徹執行，對於數據管理人員，指標庫這個事情也許意義不小，但對於全公司意義則小了，這是現狀。

在沒有公司層面的組織保障前，數據管理人員或BI部門大多得靠自己，通過自己來推動事情往前走，這是應有的態度，你不提，公司也沒有任何人會提，畢竟你是最大受益者，實施指標庫這個事情非常複雜，誰都沒有成功的把握，秉持小步快跑，試點探索的原則是不錯的。

筆者的這個指標庫項目獲得了分管領導的強力支持，這是項目能進行的現實組織保障，其實這類管理項目設立之初，很難讓業務部門和一線人員馬上認識到其價值並充分參與進來，這個溝通管理成本太高了，但無論如何，一個數據治理項目能否成功，公司的支持是第一要務，不僅僅是IT部門的事情，DAMA的很早就《DAMA數據管理知識體系指南》明確了數據治理的組織要點，以下是DAMA的數據治理組織架構圖，非常超前：

[[img](http://www.finereport.com/tw/wp-content/themes/BusinessNews/images/2017061516.png)]<http://www.finereport.com/tw/wp-content/themes/BusinessNews/images/2017061516.png>[[img](#)]

當然我覺得現實的組織演進也許如下圖更合適，但道理是一樣的，相關利益方需要對這個事情達成共識：

[[img](http://www.finereport.com/tw/wp-content/themes/BusinessNews/images/2017061517.png)]<http://www.finereport.com/tw/wp-content/themes/BusinessNews/images/2017061517.png>[[img](#)]

#### [[b](#)]2、報表梳理[[b](#)]

指標的主要表現形式是報表，因此第一要務就是報表梳理，公司的報表浩如煙海，因此這個項目設立之初就限制了範圍，主要針對一線市場部經理、終端管理、流量管理三類核心角色，共梳理了相關的39個彩信、48份郵件通報及數據集市上的733張報表。（筆者所在公司為某運營商）

[img]http://www.finereport.com/tw/wp-content/themes/BusinessNews/images/2017061518.jpg[/img]

[img]http://www.finereport.com/tw/wp-content/themes/BusinessNews/images/2017061519.jpg[/img]

### [b]3、指標整合[/b]

各類報表及相關指標表達各不相同，梳理前應該給出一個描述指標的標準框架，包括指標大類、子類、維度、周期、歸屬、命名規範等等，曾經由於框架漏了一些要素導致返工現象，這個頂層設計一定要做好，以下是示例：

命名規範：業務限定詞+業務名稱+量值限定詞+量值描述（量、收、用）

舉例1：兩網有效用戶到達數

舉例2：自建有線寬頻出賬用戶數

下圖列出了大致的梳理步驟，主要以省公司報表和彩信KPI為基礎確定基準指標，各地市指標剔除個性指標後，合并到省公司的基準指標中，形成本次的最終指標範圍。

[img]http://www.finereport.com/tw/wp-content/themes/BusinessNews/images/2017061520.png[/img]

全省指標共計6841個（未剔重），經過歸併整合，得到基礎共性指標2306個，如下圖所示：

[img]http://www.finereport.com/tw/wp-content/themes/BusinessNews/images/2017061521.jpg[/img]

此項工作耗時巨大，以下是成果的示意：

[img]http://www.finereport.com/tw/wp-content/themes/BusinessNews/images/2017061522.jpg[/img]

[i][u]數據分析，報表實例，專業的人都在這裡！加入[url=https://www.facebook.com/twfinereport/]FineReport臉書粉絲團[/url]！[/u][/i]

### [b]4、實現方式[/b]

根據指標性質不同可以分為3類，即基礎指標1046個、計算指標652個和通用行銷類指標303個。

[img]http://www.finereport.com/tw/wp-content/themes/BusinessNews/images/2017061523.png[/img]

### [b]5、功能架構[/b]

為了支撐指標快速，標準化實現，通過增強數據管理平台來實現指標的快速開發、部署和管理，主要包括指標信息維護、指標開發、運維管理、指標質量管理等功能。

比如指標庫每月需要新增超過9.5億行的數據，存儲周期按12+1，即123億行，以傳統關係型資料庫的查詢能力無法支撐，這裡就採用Hbase架構支撐海量指標的快速查詢。

[img]http://www.finereport.com/tw/wp-content/themes/BusinessNews/images/2017061524.png[/img]

### [b]6、可視化引擎[/b]

為了支撐指標組裝報表與配置報表的快速開發，使用數據可視化引擎產品，主要包括指標組裝、報表開發、報表展現功能，現在的這類產品很多了，但定製化給予一個創新性項目更大的自由度。

[img]http://www.finereport.com/tw/wp-content/themes/BusinessNews/images/2017061525.png[/img]

指標組裝報表製作工具是區別傳統基於SQL配置報表的靈活度更高的報表配置方式，主要提供基於指標選擇組裝生成報表。

[img]http://www.finereport.com/tw/wp-content/themes/BusinessNews/images/2017061526.png[/img]

[b]7、管理流程[/b]

指標的建設只是走完了數據治理的第一步，為了確保指標庫長期可用，必須要有一套針對的指標管理機制和流程，否則建設的結束就是混亂的開始，理想的做法當然是發布一套公司級別的指標管理規範，但這個時候時機往往並不成熟，比如系統可用性到底如何，因此，我們當時就確立了一個簡單原則，一條開發鐵律：不重複開發，能用指標實現的不允許單獨開發報表，當然這非常考驗數據管理的藝術，極大依賴於團隊的業務和數據能力，但有主見的數據管理團隊一定要懂得如何與業務人員進行博弈，記得你才是全公司數據的管理者，而不僅僅是個開發者。

[img]http://www.finereport.com/tw/wp-content/themes/BusinessNews/images/2017061527.png[/img]

筆者在關於指標庫的實現簡要談完了，但我對於大多企業搞指標庫卻是持悲觀態度的，傳統BI部門面對浩海的數據需求時，往往是沒有管理原則的，因為公司對你的數據管理授權是不明確的，我們不得不以犧牲長遠來滿足當前，其實BI每接收一個不規範（比如胡亂的指標命名和定義）的報表需求就要承擔由此帶來的管理成本，而不僅僅是開發成本，這為後續數據管理的混亂埋下了禍根。

但存在的又是合理的，因為搞個指標庫在開始的時候，無論是管理及運維成本都不低，關鍵是短期來看效益還不明顯，這也許是成功案例不多的一個原因。

因此，當我們在抱怨業務指標口徑一塌糊塗的時候，要記得是企業沒有數據管理的原則導致了這個現象，也是你的不作為導致了這個現象，這跟公司的文化、機制及流程是息息相關的，頂層設計沒解決，也許只能將就了，或者，你就要付出百倍的努力去改變或優化這個設計吧，這需要巨大的決心和毅力。

DAMA談數據治理首當其衝談組織設置，顯然是非常睿智的，奇怪的是在知乎上關於DAMA數據治理的討論幾乎沒有，這倒是值得思考的問題。

文 | 傅一平

原文自：微信公眾號 與數據同行

---