

涵蓋真實世界數據之藥品與生物製 劑申請資料標準化指引

衛生福利部食品藥物管理署

中華民國 114 年 11 月

目錄 (Table of Contents)

一、前言	2
二、目的及範疇	3
三、應用數據標準於來自 RWD 之研究數據	4
(一) 真實世界數據標準化的挑戰	4
(二) 處理真實世界數據程序之紀錄	6
(三) 使真實世界數據符合目前支援之研究數據標準的考量	7
(四) 將真實世界數據映射(mapping)至研究數據檢送標準的考量 ..	8
(五) 數據轉換(Data Transformation)的考量	10
四、結語	11
五、參考文獻	12
六、附錄	13
將健康照護數據映射到 CDISC SDTM 的例子	13
詞彙表	19

一、前言

為了加速藥品研發，並將新發展的藥品更快速且有效率的帶給需要使用的病人，各國法規單位於近年間，紛紛發布各種在法規決策中使用真實世界證據(Real-World Evidence, RWE)的指引，台灣則自民國一百零九年開始，逐步公告關於真實世界數據/真實世界證據(Real-World Data/Real-World Evidence, RWD/RWE)的各項相關指引，當中包括利用 RWD 來支持法規決策的「真實世界數據-評估以電子健康紀錄和醫療保險資料支持藥品與生物製劑的法規決策指引」。

另一方面，為了因應電子化送件的時代，各國法規單位亦逐漸要求申請者須依照標準電子格式檢送各項藥品相關申請。除此之外，若能以電子化系統檢送研究數據，將可讓法規單位更有效接收、處理、審查和歸檔所檢送之文件，進而促使審查過程能更簡化並與時俱進。然而電子化傳輸數據須將電子格式數據標準化(data standardized)，以確保件送的資料在不同的電腦系統上都可以審查處理，並且使審查人員能夠透過多項研究數據的結合，進而探索更多研究議題，此外數據標準化亦有助於分析工具的使用，可以檢視各項藥品數據，並藉此凸顯須關注的領域。

本文提供了藥品進行相關申請時，以符合法規單位要求或是可以支援之數據標準的方式來檢送 RWD 作為研究數據的建議，此處所謂

的「支援(supported)」是指接收藥品申請單位已建立的流程與技術，使用特定標準以支持接收、處理、審查和歸檔文件；而「數據標準(data standards)」則是指已定義之規則、慣例、指引、特徵、方法、格式與術語，用以提供數據交換和數據使用時的結構與一致性，因此，本指引可作為 RWD 用於法規決策指引的一部分。

二、 目的及範疇

本文闡述了「含有自 RWD 來源之研究數據」進行藥品相關申請時，使用數據標準化的考量。本文將 RWD 定義為從各種來源收集且與病人健康狀況和/或醫療常規照護相關的數據，RWD 的例子包括電子健康紀錄(electronic health records, EHR)、醫療給付資料(medical claims data)、藥品和疾病登錄資料(registry data)、病人產生的數據(包括來自家庭使用狀況之數據，如數位健康技術)，及從其他可提供健康狀況來源(如行動式裝備)所收集的數據等。

美國食品藥物管理機關(US FDA)基於該國的法案規定，指定申請者在檢送特定適用類別之藥品申請案(包括新藥查驗登記、生物製劑查驗登記、臨床試驗申請等)時，需使用特定的標準化研究數據電子格式，相關資料可參考 US FDA 公告之指引(The guidance for industry Providing Regulatory Submissions in Electronic Format—Standardized

Study Data (Study Data Guidance))。而上述檢送標準化電子研究數據之要求，也適用於當這些適用的申請案使用 RWD 作為研究數據檢送的狀況，亦即 RWD 作為研究數據檢送時，必須使用法規單位可處理、審查、和存檔的電子格式送件。

本文將針對使用 RWD 作為藥品相關申請之臨床研究數據檢送時，進行數據標準化所需考量的事項，提供概念性介紹。此外，因應現行的大型跨國臨床試驗公司和世界主要國家法規單位，普遍採納臨床數據交換標準協會(Clinical Data Interchange Standards Consortium，CDISC)所發展的臨床數據標準格式，故本文中部分內容會涉及 CDISC 相關概念，若需要進一步了解 CDISC 相關細節，可參考 CDISC 組織網站(<https://www.cdisc.org/>)。

三、 應用數據標準於來自 RWD 之研究數據

(一)真實世界數據標準化的挑戰

將來自於 RWD 的研究數據標準化，以將其納入適用標準化研究數據電子格式檢送之藥品相關案件的檢送中，將涉及許多挑戰，這些挑戰包括但不限於以下各項：

1. RWD 來源的多樣性，以及其不一致的格式，例如 EHR、登錄資料使用的表格。

2. 在一項研究中可能使用超過一種類型的 RWD 來源，例如合併使用 EHR 和醫療給付資料。
3. 因相同或相似的數據元素(data elements)，可使用不同標準(standards)、不同術語(terminologies，指一種字彙主體，用以將特殊技術應用於檢送非臨床和臨床研究數據的標準化醫療詞彙)、和不同的交換格式(exchange formats，指一種可用以將檔案或數據庫從一種結構轉換為另一種結構的數據格式，例如 XML [Extensible Markup Language，可擴展標記語言]是一種經常使用的數據交換格式)表示，使得在區域範圍或全球範圍所擷取的來源數據產生差異。
4. 某些資訊只存在於非結構化的文件中，例如醫師自行記錄的病歷文本。
5. 有多種方式和演算法，可被用來建構想要匯聚(aggregate)數據的數據集(datasets)，例如從不同醫院抽取非結構化資料時可採用 Clinical BERT 或 NER 等不同演算法。
6. 健康照護資料的許多面向都可能影響數據的整體品質，包括執行業務的流程、數據庫(database)的結構、不一致的詞彙(vocabularies)和編碼系統(coding system)、及分享數據時用以保護病人隱私資料的去識別化方法(de-identification)

methodologies)等。

7. 申請者對任何研究中所使用的數據來源，有不同程度的存取權限(accessibility)。

用以應對上述這些挑戰的方法，將取決於所使用的特定數據和正在進行的研究，因此申請者應及早與法規單位相關部門討論此類議題，且最好在申請者執行研究之前即進行討論。

(二)處理真實世界數據程序之紀錄

在數據應用(data curation，指處理來源數據的過程，藉由將標準應用於交換、匯集、分享和檢索數據)和數據轉換(data transformation，指將數據從一種格式或結構轉換為另一種格式或結構的過程)期間，應制定適當的流程以確保對結果數據的信賴度(confidence)。這些過程的紀錄，可能包括但不限於從數據來源系統到最終研究分析數據集的數據添加(data additions)、清除(deletions)或更改(alterations)的電子紀錄(例如，稽核路徑(audit trails)、品質管控程序等)，申請者還應該在申請藥品相關檢送中，將為了符合目前數據標準要求而進行的更改過程加以記錄，並說明這些更改可能造成的潛在影響。

(三)使真實世界數據符合目前支援之研究數據標準的考量

當試著使 RWD 符合法規單位要求的檢送數據標準時，申請者應考量在產生符合申請藥品檢送要求格式的研究數據集時可能需要的程序，包括相關的數據轉換(data transformation)、數據變換(data conversion，將電子數據從一種格式變換為另一種格式)或數據映射(data mapping，指在兩個不同數據模型間，定義資料元素關係的過程)。

申請者有任何要在藥品申請中檢送來自 RWD 來源之研究數據的計畫，以及使用之數據映射及數據轉換方法，均應盡早與法規單位討論。此類討論應包括將數據從原始 RWD 格式轉換為中間格式(包括用以分析之格式)的方法，及轉換為最終準備檢送給法規單位的研究數據集的方法。申請者應在計畫書、數據管理計畫(data management plan)、和最終研究報告中描述這些方法。

法規單位明白可以使用一系列方法將所支援的研究數據標準(例如，CDISC's 的 Study Data Tabulation Model [SDTM]或 Analysis Data Model [ADaM])應用到 RWD 來源，如 EHR 或醫療給付資料，我們鼓勵申請者與法規單位討論。透過充分紀錄使用的一致性方法(conformance methods)和其合理性，來自 RWD 的研究數據可被轉換為 SDTM 和 ADaM 數據集，並以適用的型式檢送法規單位。

(四) 將真實世界數據映射(mapping)至研究數據檢送標準的 考量

法規單位意識到，幾乎所有的數據領域(data domain，指臨床研究中，從所有受試者收集到具有邏輯相關性且具有共同特定主題之觀察結果的集合，例如不良事件領域[adverse event domain]、併用藥品領域[concomitant medication domain])中，RWD 數據來源和法規單位所支援或要求的研究數據標準間，使用的術語和其精確意義存在有很大的差異，例子包括用於種族/民族(race/ethnicity)的含意與特定用詞、藥品的術語系統(terminology system)和對來自健康照護紀錄中生命徵象測量結果的解釋等。即使對於看似相似的變數(variables)，在 RWD 來源和法規單位所支援的數據標準間，這些變數的定義也可能存在差異，例如，性別(sex)作為變數，在 CDISC 的術語中可能會基於身體特徵概念做編碼，但在 EHRs 的某些情況下可能會使用性別認同做編碼(即身體特徵為男性，但自我性別認同為女性，因此在醫療紀錄中被記載為女性)。在這種情況下，申請者應記錄將性別變數或其他變數映射到 CDISC 術語中對研究結果的潛在影響。

申請者應選擇最符合數據特徵和研究性質的最適當的映射方法，且應在研究數據審查者指引(Study Data Reviewer's Guide，指申請者提供給法規單位，用以描述任何特殊考量、指示(directions)、或一致性

議題的文件，可協助於審查者使用所檢送的數據，且可以幫助審查者了解研究報告與數據間的關聯性)中解釋所選用方法之合理性。此外，申請者從目前支援的研究數據標準中選擇特定數據元素，來代表 RWD 來源的研究數據合理性的文件是至關重要的，申請者應在研究數據審查者指引中，提供數據映射之整體方法和預期影響的描述，以強調涉及的數據領域(data domain)。此外，申請者應在其中包含一個數據字典(data dictionary)，當中記錄來自 RWD 來源並映射到研究數據標準之每個數據元素的定義及所有相關資訊，例如與其他數據的關聯性、來源、用途和格式，研究數據審查者指引還可用於引導審查者進行詳細映射，這些資料可放置在 Define-XML 檔案(CDISC Case Report Tabulation Data Definition Specification，可傳輸描述任何表格數據集結構的詮釋資料[metadata]，當與 CDISC 基礎標準一起使用時，此檔案提供了使用 SDTM 和 Standard for Exchange of Nonclinical Data [SEND]標準的人類和動物模型數據集的詮釋資料，和使用 ADaM 的分析數據集)和數據領域檔案中。

(五)數據轉換(Data Transformation)的考量

申請者在將 RWD 轉換成為與法規單位支援的數據標準一致時可能會遇到挑戰，這些挑戰的例子包括(但不限於)：處理出現在健康紀錄中多個位置的語意概念(semantic concepts 或 terms)(例如，藥品資訊可能出現在醫師記載的病歷文本中、在處方單中、或配藥單中)、概念的不一致編碼或錯誤編碼(例如，藥品編碼有些使用全民健康保險藥品編碼，有些使用 ATC 編碼；腫瘤長在胃與食道的交界處，診斷碼可能下為 C16.0 [胃食道交界之胃惡性腫瘤]或 C15.5 [食道下 1/3 惡性腫瘤])、研究期間發生的數據收集改變或編碼改變(例如，從 International Classification of Disease-9 [ICD-9]改換為 ICD-10)、或缺失資訊(不論是因為該項資訊通常不會在健康照護中記錄[例如，自行在藥局購買的成藥]，或是因為錯誤的數據輸入所造成)。

申請者對於在將數據轉換為法規單位支援之研究數據標準的過程中所遭遇到的數據挑戰，及其處理方法的合理性應該清楚記錄，以證明足以使用法規單位支援的研究數據標準檢送資料。用以將數據轉換為目前法規單位支援之研究數據標準所採用的術語映射與標準映射的詳細資料，可使用 Define-XML 和數據領域檔案處理。此外，對於更複雜的討論，也應該將敘述性的描述呈現在研究數據審查者指引中，不論是指引的主體或附錄，並為審查者提供適當指示以導引至記

載更多細節的 Define-XML 和數據集/領域檔案。

四、 結語

本指引提供了關於如何在進行藥品和生物製劑相關申請時，有效利用真實世界數據作為提交資料的概念性綜合考量，此外也說明了數據標準化的必要性，以及將 RWD 進行標準化過程中可能遇到的挑戰和可能的解決方法，另外也指出 RWD 在藥品法規審查中具有潛力 and 可能的挑戰，最後亦在附錄提供了數據映射和數據轉換的具體例子，以幫助理解過程中的操作。

藉由這些整體概念，申請者將可以讓作為藥品相關申請所檢送的 RWD 資料更準確、一致、並符合法規單位所支援的數據標準，進而增進以 RWD 作為藥品申請檢送資料的品質，如此將可加快藥品研發過程、增進檢送藥品申請資料的透明度、並改善檢送過程的效率，法規單位也可加速藥品的審查過程、適應新科技與數據科學發展的靈活性，最終達到使新發展的藥品可更快速且有效率的帶給需要使用的病人之目的。

然而因為將 RWD 進行數據標準化的過程可採用的方法眾多，為了確保藥品申請的檢送過程順利，且資料符合法規單位之規範，仍建議申請者在研發早期應該儘早與法規單位溝通。

五、 參考文獻

1. Data Standards for Drug and Biological Product Submissions Containing Real-World Data Guidance for Industry
2. The Guidance for Industry Providing Regulatory Submissions in Electronic Format—Standardized Study Data (Study Data Guidance)
3. Framework for FDA’s Real-World Evidence Program (December 2018) (<https://www.fda.gov/media/120060/download>)
4. Study Data Technical Conformance Guide – Technical Specifications Document
5. Website of Clinical Data Interchange Standards Consortium (<https://www.cdisc.org/>)
6. Description of Define-XML file in CDISC Website (<https://www.cdisc.org/standards/data-exchange/define-xml>)
7. 當代醫藥法規月刊第 95 期，美國非臨床試驗數據交換標準 (Standard for Exchange of Nonclinical Data, SEND)介紹

六、 附錄

將健康照護數據映射到 CDISC SDTM 的例子

RWD 和傳統臨床研究數據間編碼系統的差異，通常會在 Define-XML 檔案中說明，且此檔案會被納入在所有 SDTM 的檢送資料中。Define-XML 檔案提供一種將外部編碼系統轉換為適當的 SDTM 控制術語(controlled terminology，指有限的數值集，用以代表數據項目中唯一允許的數值)的機制。

此方法的一個例子是預期在不同 EHRs 之間，及 EHR 和 CDISC 之間的術語存在有異質性的種族/民族數據。臨床研究中建議使用至少 5 個種類來收集種族數據：

- (1) 美國印第安人或阿拉斯加原住民
- (2) 亞洲人
- (3) 黑人或非裔美國人
- (4) 夏威夷或其他太平洋島嶼原住民
- (5) 白人。

然而 RWD 來源可能不會遵循相同的編碼系統，因此申請者應將 RWD 術語映射到相關的 SDTM 術語。為了達到此目標，可使用 Define-XML 檔案中的「別名(Alias)」元素來記錄轉換為單一命名的資訊，以確保資料的可追溯性(traceability，指記錄整個數據生命週期中使用的

應用和轉換過程)。

表格一說明如何將種族從非標準化數據轉換為法規單位支援的數據標準，其中別名(Alias)欄位顯示 EHR 系統中所顯示的原始編碼，Code 欄位則顯示在法規單位支援的 CDISC 控制術語中的相關映射用詞。

表格 一、使用 Define-XML 以指示將非標準化數據(種族數據)轉換為標準化數據(即 SDTM 和 ADaM)所涉及之決策的方法

此為一項顯示將編碼系統(本例為種族數據)交叉映射到 Define-XML 檔案中 CDISC 編碼的方法之例子。本表格並非建議如何將編碼系統映射到 CDISC 術語，此外，本表格僅顯示了多種用來呈現映射選擇方法中的一種可能。

種族 [RACE, C74457]	
允許值(編碼) Permitted Value (Code)*	顯示值(別名) Display Value (Alias)**
美國印第安人或阿拉斯加原住民[C41259]	美國印第安人或阿拉斯加原住民，美國原住民、阿拉斯加原住民；情境：EHR 映射 (EHR mapping)
亞洲人[C41260]	亞洲人、華人；情境: EHR 映射
黑人或非裔美國人[C16352]	非人或非裔美國人，黑人；情境：EHR 映射
夏威夷或其他太平洋島嶼原住民[C41219]	夏威夷或其他太平洋島嶼原住民，薩摩亞人；情境: EHR 映射
其他[*]	其他；情境：EHR 映射
白人[C41261]	白人，墨西哥人；情境: EHR 映射
*Permitted value (Code): 在研究數據表中提供的詞彙，且符合 CDISC 的控制術語	
**Display value (Alias): 在原始數據集中所使用的詞彙(即 EHR 值)。 Code/Alias: CDISC 中相應的元素	

當 RWD 系統使用的控制術語和法規單位要求之檢送數據標準不同時，可能會使術語映射(mapping terminology)成為一大挑戰，此外法規單位也意識到，在非介入性研究(non-interventional study，此處是指病人根據計畫書不會被分配到研究組別，而是在常規臨床照護中接受到感興趣的藥品)中，申請者可能會使用從多個 RWD 系統所匯聚的數據，由於不同 RWD 來源(例如不同醫院的 EHRs)可能對相同的概念使用不同的編碼系統(例如，藥品的編碼系統可能使用全民健康保險藥品編碼，或 ATC 藥品編碼)，或可能使用相同編碼系統，但對相同的項目使用不同的預設編碼(default codes)(例如，醫院 A 在性別的預設編碼上，女性預設為 F [female]，但醫院 B 則預設為 W [woman])，因此會使術語的使用更趨複雜。

在檢送藥品申請時使用 RWD 有許多可採用的方法，潛在方法的例子有：(1)將編碼轉譯為其映射的結構化定義，接著映射到適當的 CDISC 控制術語，這提供了最詳細的資訊但需要耗費大量人力；或者另外的方法(2)將所有原始編碼映射到最不精細的類比編碼(the least granular analogue codes)，接著將其映射到 CDISC 的控制術語，如此可耗費較少人力，但會導致更具體的分類細節將不會被呈現在檢送的標準化數據集中。中介數據轉換的最佳方法應由申請者決定，並依此記錄並論述其方法的合理性，然而，若是缺乏對評估藥品的安全性與

有效性所必須的細節，則後一種所述的方法就可能不適當。無論使用哪種方法，將別名(Alias)應用到 CDISC 的標準控制術語，是將命名規範標準化為 CDISC 所研發之格式的一種機制。

另一個概念和術語無法準確且直接映射的例子是 SDTM 介入領域(intervention domain)中擷取處方藥品的項目，包含處方藥品數據的領域可被映射到 SDTM 介入領域的模板，此外，除非申請者使用保留處方和配藥資訊的 EHR 數據，或選擇連結 EHR 資訊與醫療給付資料，否則實際處方的不確定性(例如，是否替換了學名藥品)將持續存在。在此狀況下，申請者可應用 Define-XML 中的別名(Alias)元素來標準化編碼系統(如表格二所示)，其最終數據集(見表格三)將具有處方劑量資訊，以作為較不具體的處方編碼系統之一致性的說明。

標準模型將附有 Define-XML 編碼列表，如下所示：

表格 二、使用 Define-XML 以指示將非標準化數據(處方藥品)轉換為標準化數據(即 SDTM 和 ADaM)所涉及之決策的方法

此為一項顯示將編碼系統(本例為處方數據)交叉映射到 Define-XML 檔案中 CDISC 編碼的方法之例子。本表格並非建議如何將編碼系統映射到 CDISC 術語，此外，本表格僅顯示了多種用來呈現映射選擇方法中的一種可能。

允許值(編碼)*	顯示值(別名)**
FLUoxetine	FLUoxetine 40 毫克口服膠囊，RxCUI=383919，0093-7198-56，FLUoxetine 40 毫克口服膠囊允許的學名藥；情境：EHR 映射 (EHR mapping)

PROzac	PROzac 40 毫克口服膠囊；RxCUI = 313989，0777-3107-30；情境：EHR 映射
* Permitted value (Code): 在研究數據表中提供的詞彙 ** Display value (Alias): 在原始數據集中所使用的詞彙(即 EHR 值)。 Code/Alias: CDISC 中相應的元素	

在領域檔案中處方數據的標準模型將呈現如下(表格三):

表格 三、映射後藥品的處方數據相應領域的例子

此為一項將數據中的值(本例為處方數據)呈現為相關數據領域檔案中之 CDISC 模式的例子。本表格並非建議如何將數據映射到 CDISC 標準，此外，本表格僅顯示了多種用來呈現映射選擇方法中的一種可能。表格的欄位標題代表了 CDISC 數據元素，更多相關資訊可參考 <https://www.cdisc.org/standards/foundational/sdtm>。

原始數值 (來自原始 數據集)	--TRT ¹	--MODIFY ²	--Alias	--DOSE ³	--DOSU ⁴	--ROUTE ⁵	--DOSFRM ⁶
FLUoxetine 40 毫克口 服膠囊	FLUoxetine	由申請者填 寫以協助依 標準術語進 行編碼	WHO 藥 典中的學 名藥品名 稱(可能來 自原始數 據系統或 由醫學 編碼指定)	40	毫克	口服	膠囊
FLUoxetine 40 毫克口 服膠囊允 許的學名 藥	FLUoxetine	同上	同上	40	毫克	口服	膠囊
RxCUI ⁷ = 383919	FLUoxetine	同上	同上				

0093-7198-56 ⁸	FLUoxetine	同上	同上				
PROzac 40 毫克口服膠囊	PROzac	同上	同上	40	毫克	口服	膠囊
RxCUI ⁷ =313989	PROzac	同上	同上				
0777-3107-30 ⁸	PROzac	同上	同上				

欄位標題說明

1. --TRT：為介入領域當中的主題變數(topic variable)，此處即指治療項目，前方的"--"可套入 CDISC 的數據集層級的詮釋資料(dataset-level metadata)縮寫，例如 EXTRT 指暴露的治療、CMTRT 指併用藥品的治療等，且此項目只會有藥品名，不包含劑量劑型等其他資訊。
2. --MODIFY：若介入領域中的逐字主題變數被修改過用以幫助編碼，則修改後的文字將儲存在此欄位。
3. --DOSE：用以描述主題變數的劑量資訊，前方的"--"可套入詮釋資料縮寫，例如 EXDOSE 指暴露劑量、CMDOSE 指併用藥品的劑量等。
4. --DOSU：為--DOSE 的變數限定符號(variable qualifiers，用來在一般觀察類別中進一步描述特定變數，且只有在特定的變數上下文中才有意義)，此處代表 dose unit。前方的"--"可套入詮釋資料縮寫，例如 EXDOSU 指暴露劑量單位、CMDOSE 指併用藥品的劑量單位等。
5. --ROUTE：此處代表給藥途徑。前方的"--"可套入詮釋資料縮寫，例如 EXROUTE 指暴露的給藥途徑、CMROUTE 指併用藥品的給藥途徑等。
6. --DOSFRM：此處代表 dose form。前方的"--"可套入詮釋資料縮寫，例如 EXDOSFRM 指暴露劑型、CMDOSFRM 指併用藥品的劑型等。

綜合以上資料，若在 EHR 中擷取到併用藥品的處方籤資料為 FLUoxetine 40 毫克口服膠囊，轉換為可檢送的 SDTM 模式則須表示為 CMTRT=FLUoxetine，CMDOSE=40，CMDOSU=mg，CMROUTE=oral，CMDOSFRM=capsule，即如表格的一個橫列所呈現的資訊

原始數值使用的編碼例子說明

7. RxCUI：為美國國家藥品圖書館所發布的 RxNorm 命名系統中，用來代表分配給特定藥品的獨一、明確的標示符號，且被用在所有與該藥品相

關的事物中。

8. 0093-7198-56、0777-3107-30：二者均為美國國家藥品編碼(National Drug Code, NDC)系統的特定藥品辨識碼，由 10 或 11 位數、3 段數字組成，第一段為標籤資訊(labeler)，如製造廠、包裝廠等；第二段為產品碼，可辨識特定的強度、劑型等；第三段則為包裝碼，可辨識包裝大小與種類。

儘管此處僅呈現了少數例子，申請者應使用 Define-XML 檔案及相關領域數據檔案的元素，來說明如何將所有數據領域的健康照護術語轉換為一致的 CDISC 標準術語。如上方所示的例子，所有轉換的技術性細節均需放置在數據檔案中。

詞彙表

詞彙	說明
術語 (terminologies)	指一種字彙主體，用以將特殊技術應用於檢送非臨床和臨床研究數據的標準化醫療詞彙
交換格式 (exchange formats)	指一種可用以將檔案或數據庫從一種結構轉換為另一種結構的數據格式，例如 XML (Extensible Markup Language，可擴展標記語言)是一種經常使用的數據交換格式
數據應用(data curation)	指處理來源數據的過程，藉由將標準應用於交換、匯集、分享和檢索數據
數據轉換(data	指將數據從一種格式或結構轉換為另一種格式

transformation)	或結構的過程，包含資料提取(extract)、資料清理(data cleansing/scrubbing)、整合、去識別化過程
數據映射(data mapping)	指在兩個不同數據模型間，定義資料元素關係的過程
數據領域(data domain)	指臨床研究中，從所有受試者收集到具有邏輯相關性且具有共同特定主題之觀察結果的集合，例如不良事件領域(adverse event domain)、併用藥品領域(concomitant medication domain)
研究數據審查者指引(Study Data Reviewer's Guide)	指申請者提供給法規單位，用以描述任何特殊考量、指示(directions)、或一致性議題的文件，可協助於審查者使用所檢送的數據，且可以幫助審查者了解研究報告與數據間的關聯性
Define-XML 檔案	CDISC Case Report Tabulation Data Definition Specification，可傳輸描述任何表格數據集結構的詮釋資料(metadata)，當與 CDISC 基礎標準一起使用時，此檔案提供了使用 SDTM 和 Standard for Exchange of Nonclinical Data (SEND)標準的人類和動物模型數據集的詮釋資料，和使用 ADaM 的分析數據集

控制術語 (controlled terminology)	指有限的數值集，用以代表數據項目中唯一允許的數值
可追溯性 (traceability)	指記錄整個數據生命週期中使用的廢用和轉換過程
在非介入性研究 (non- interventional study)	此處是指病人根據計畫書不會被分配到研究組別，而是在常規臨床照護中接受到感興趣的藥品