

淺介聯徵中心Open Data

李彥錚 / 金融聯合徵信中心 研究部

我國為便利民眾共享及應用政府資料，促進政府資料加值應用，並結合民間創意以提升政府資料之價值，自2015年起深化推廣「開放資料」(Open Data)之政策，而本中心配合金管會之指示，自該年起即產出各類Open Data上架於本中心官網及政府資料開放平台，並於其後數年陸續新增各項資料之開放，截至2020年初，已累計有122項開放資料集。由於自本中心推出Open Data起迄今已滿五年，茲以本文說明本中心Open Data之推動歷程、內容、效益以及未來之展望。

背景概述

Open Data之浪潮來自於美國(2009)、英國(2010)，早期的推動力量來自民間，Open Data的倡議者認為，藉由資料的公開，可促進公民參與政治，進而使政府的政策更透明；另外，藉由釋放政府的大量資料，可以節省民間在收集資料時必須花費的時間與金錢成本，再加上民間的創意及資源，就可以發展出各式應用，為社會整體創造附加價值，而在英美兩國的政府都開始推動之後，此風潮即迅速在各國間蔓延。從開放資料的內容領域來看，有4類資料被認為與多數民眾的日常生活較攸關，因此最早被討論加以開放，分別為交通運輸、

氣象、醫療、地政，隨後再逐步擴增至其他領域。

我國政府自2012年起即開始陸續小規模推行開放資料，資料開放領域主要是地方市政、氣象、政府預算等，而自2015年起，因行政院為深化推廣此政策，宣示該年為「Open Data 深化應用元年」，至此即開始大量開放各部會之資料，本中心所隸屬之金融領域亦自當時開始進行資料開放。截至2020年初，政府所開放之資料集總計達4萬4千餘項，當中與金管會相關之資料集則約為1千5百餘項(含本中心所開放之資料集)。

為配合政府之資料開放政策，本中心於官網新增「資料開放專區」，並於其下分成2個區塊：

1. Open Data專區

以去識別化之彙整方式呈現，依資料內容之面向共可分為9類：個人授信統計資訊(17項)、房貸類統計資訊(9項)、個人信貸類統計資訊(4項)、車貸類統計資訊(4項)、信用卡統計資訊(13項)、循環信用及現金卡統計資訊(24項)、企業授信統計資訊(22項)、財報類統計資訊(18項)、其他公開資訊(11項)。截至2020年初，共計有122項開放資料集。

2. 運用大數據分析資訊

為讓開放資料運用更加普及，活化資料應用，本中心另建置「購置住宅貸款統計資訊」及「產業財務統計資訊」大數據應用平台，提供使用者政府資料開放平台房貸及產業財務相關資料集，並結合金融機構往來之授信資料、會計師簽證財務報告書等進行加值運用。本平台之特色在於針對使用介面上做了許多設計，如簡易索引、地圖化、制訂下拉清單供使用者自訂條件篩選等功能，讓使用者能以更方便、更快速地取得相關資訊。

圖1：本中心官網之資料開放專區



Open Data的一項重要基本精神，在於「機器可讀性」(machine-readable)，其意義在於它們的數據要能夠輕易被電腦程式擷取，讓資料可在不受技術限制的前提下，被加值使用甚至可依使用需求加以重組；相對而言，文字型的文件(如PDF檔)、或編排精美的報表(即使是以數字呈現的word檔)，通常不是機器可讀取的格式，因此針對Open Data的格式及品質，官方制定了許多相關規範，必須在符合此類規範的資料才能被當成是Open Data，而前述之「運用大數據分析資訊」，因以使用者視覺上的便利性為出發點，與機器可讀性的本質不同，所以並不歸類為Open Data的一部分，因此不在本文後續的說明範圍內。

Open Data之資料格式

本中心針對Open Data的格式制定及品質控管，皆依循金管會所制訂的「金融資料開放標準與作業要點」辦理。

本中心之Open Data，係以開放格式提供，使用者不需事先提出申請即可使用，且不收取費用，也不限制使用之目的、地區及期間。另為了確保資料的機器可讀性及品質，特依照下述3個構面進行資料的設定：

- 可取得性：在這個構面下，主要的重點在於確保資料連結的有效性，且使用者不需再另行登入或進行額外的操作。

- 易於被處理：這個構面強調的是資料必須是「結構化資料」，除了資料要被機器讀取外，檔案的開啟也不能被特定軟體所限制，對於各個不同的系統也要有高的相容性。因此資料格式本中心幾乎皆設定為CSV檔，並採用UTF-8的編碼格式。資料本身除欄位固定且只有單列標題外，尚必須無合併儲存格、無公式、無空行、無小計。
- 易於理解：這個構面強調的是必須有資料集之詮釋資料，例如「主要欄位說明」，讓資料內容能被系統化的管控，且可讓使用者能快速掌握資料集的所含內容。此外，資料集尚必須定時更新(更新頻率由發佈者自訂)，目前本中心針對Open Data的更新頻率，基本上為每季更新(但少數資料如財務報表，因一年才出具一次，所以更新頻率會較久)，資料更新的時間點為每季季末(3、6、9、12月)之下旬。

金管會為確保開放資料之品質，每季皆會依國發會「政府資料品質提升機制運作指引」之檢測項目及頻率，自動化地檢測所有資料集，而本中心所開放的所有資料集，其品質皆符合國發會制訂最高等級之金標章水準¹，並協助金管會於2019年在國發會所舉辦的開放資料競賽中，於「資料開放金質獎」的項目下獲得第一名。

1 國發會於2020年起在「政府資料開放優質標章暨深化應用獎勵措施」內增訂「白金標章」(較金標章的標準更高)，惟未必每個機構的資料屬性皆可適用到白金標章的規範，故其定位暫訂為鼓勵性質的加分項目。

Open Data之內容

針對資料內容最初始的分類，係從消費金融與企業金融的區分做為出發點，與一般銀行實務相近。在消費金融下，我們可以觀察個人戶的授信以及信用卡資料，其中授信可依業務別(如房貸、車貸…等)再行分類，信用卡亦可再針對循環信用的分項下特別觀察；而在企業金融下，我們可以觀察企業戶的授信資料，此外財務報表資料為本中心資料庫的重要成分，亦特別將其內容獨立分項。

至於資料的呈現方式，除了每個資料時間點下的累積金額外(又稱為「存量」)，尚針對每個資料時間點下「新增」的金額加以觀察(又稱為「流量」)，以讓使用者瞭解資料的動態變化。

綜合以上的各種資料分類及呈現方式，即可組成多種資料集，然而由於本中心所開放

的資料集多達122項，實無法一一列舉，故特別選取幾種較常見的內容態樣加以說明，以使讀者能有大略的輪廓以明瞭本中心開放資料所內含的資訊，以及這些資料可被加以分析應用的方向。

態樣1：個人授信金額變動趨勢-存量

以表1為例，此類型的資料主要係揭露各時間點下，授信金額的加總、平均、以及分布狀況²。為了符合機器可讀性，數字格式不放千分位符號、亦不對小數點後之位數進行設定，呈現資料最原始的樣貌。此類資料可以觀察消金市場下的冷熱變化，為整體市場分析的參考資訊，亦可瞭解自身借款狀況所處的相對位置。

此種形式可以套用到不同的業務分類(如房貸、車貸…等)，藉此產生不同的資料集。

表1 個人總貸款狀況統計趨勢資料(資料編號1-4，部分顯示)

資料 年月	個人貸款 總人數	個人貸款 總金額 [仟元]	個人貸款 平均金額 [仟元]	個人貸款 金額第25 百分位數 [仟元]	個人貸款 金額第50 百分位數 [仟元]	個人貸款 金額第75 百分位數 [仟元]
201201	4742543	10431407540	2199.54	195	660	2249
201202	4728534	10426760702	2205.07	195	661	2250
201203	4716591	10449619069	2215.50	196	664	2255
201204	4714865	10492731783	2225.46	199	665	2261
201205	4712690	10554011059	2239.49	203	669	2271
201206	4703118	10612688184	2256.52	204	675	2285

2 因平均值有時會受極端值的影響，所以藉由四分位數資訊的揭露，讓使用者更能清楚整體資訊的分布狀況。

態樣2：個人授信金額變動趨勢-流量

以表2為例，此類型的資料主要係揭露各時間點下，新核准的授信金額加總、平均、以及分布狀況。相對存量的資料而言，流量的資料更能明顯反映各時間點下，所處環境及政策

之變化與影響。

此種形式同樣可以套用到不同的業務分類(如房貸、車貸…等)，藉此產生不同的資料集。

表2 新增個人總貸款狀況統計趨勢資料(資料編號1-20，部分顯示)

資料 年月	新增個人貸款總人數	新增個人貸款總金額[仟元]	新增個人貸款平均金額[仟元]	新增個人貸款金額第25百分位數[仟元]	新增個人貸款金額第50百分位數[仟元]	新增個人貸款金額第75百分位數[仟元]
201201	102705	255458543	2487.30	240	500	2000
201202	96298	251903581	2615.88	240	550	2000
201203	139583	338046843	2421.83	200	500	2000
201204	303014	350659947	1157.24	27	53	400
201205	282453	398773397	1411.82	35	60	560
201206	130421	369138069	2830.36	250	600	2345

態樣3：個人授信金額變動趨勢-交叉分析

以表3、表4為例，此類資料主要針對個人授信、以及各業務別之下，針對不同的時間點，加入「其他分類資訊」用以交叉分析(例

如：年齡、性別等)，來觀察授信金額的加總、及平均利率。此類資料集可以觀察多元的主題資訊，例如分析年輕族群的借款趨勢、女性借款量與利率的變化等。

表3 房貸借款人各年齡層下的授信金額及利率統計表(資料編號1-23，部分顯示)

資料 年月	借款人年齡區間	人數	授信餘額[仟元]	平均利率[%]
201201	20-30	106915	363806949	1.985133729
201201	30-40	609331	1951843676	2.008788370
201201	40-50	724545	2208874409	2.160330380
201201	50-60	475745	1465090944	2.283712019
201201	60以上	180377	495210262	2.400645968
201202	20-30	108222	367859681	1.987034671
201202	30-40	609755	1951739361	2.015655721
201202	40-50	722328	2199368250	2.170766723
201202	50-60	473277	1454519640	2.292990689
201202	60以上	178905	490025715	2.407802085

表4 信貸借款人不同性別下的授信金額及利率統計表(資料編號1-32，部分顯示)

資料 年月	性別	個人信貸總人數	個人信貸總金額[仟元]	個人信貸平均利率[%]
201201	01. 男	732637	313444085	6.17665
201201	02. 女	571492	171486714	6.04903
201202	01. 男	728059	311227045	6.17859
201202	02. 女	566821	169057511	6.06617
201203	01. 男	727012	311429621	6.18222
201203	02. 女	565096	169465623	6.07926
201204	01. 男	725839	312467076	6.17808
201204	02. 女	563173	169518433	6.08566
201205	01. 男	726136	315730851	6.17310
201205	02. 女	562407	171021647	6.08407
201206	01. 男	726211	318740693	6.17817
201206	02. 女	562071	173048071	6.09448

此種形式在不同的業務分類下，皆有不同
的資料集；且也會依「流量」、「存量」分別
建立不同的資料集。

態樣4：信用卡類資料

以表5為例，此類資料主要係針對信用卡
業務做統計，除觀察持卡張數外，也會另外針
對應繳金額、循環信用金額等資訊進行統計，
同樣也會加入「其他分類資訊」做交叉分析
(例如：性別、年齡等)，分別建立不同的資料
集。

態樣5：企業授信金額變動趨勢-交叉分析

以表6為例，此類資料主要係針對企業戶
者做統計，觀察在各時間點下，加入「其他分

表5 信貸借款人不同性別下的授信金額及利率
統計表(資料編號1-32，部分顯示)

資料 年月	信用卡平均每戶持卡張數
10308	3.85
10309	3.86
10310	3.85
10311	3.86
10312	3.85
10401	3.86
10402	3.86
10403	3.86
10404	3.86

表6 企業規模別貸款狀況統計趨勢資料(資料編號1-6，部分顯示)

資料 年月	規模別	企業規模貸款 總家數	企業規模貸款 總金額[仟元]	企業規模貸款 平均金額[仟元]
201205	0.上市	1183	3579297273	3025610.54
201205	1.大型	9113	5474891888	600778.22
201205	2.中型	20426	3355285825	164265.44
201205	3.小型	84705	557032049	6576.14
201206	0.上市	1179	3507988883	2975393.45
201206	1.大型	9147	5483037359	599435.59
201206	2.中型	20525	3214390747	156608.56
201206	3.小型	84916	561448494	6611.81

類資訊」用以交叉分析(例如：產業別、規模大小)，來看授信家數、授信金額加總、及平均金額等，因企業授信的異質性較個人授信為高，藉由區別不同類型始較能反映清晰的資訊意涵。企業資料與前述個人授信資料集的設定相

同，也會依「流量」、「存量」分別建立不同的資料集。

態樣6：財務報表類資料

以表7為例，此類資料主要係針對財務報表做統計，基本上係就重要會計科目(如資產、

表7 各產業淨利(損)總額及企業家數統計表(資料編號7-4，部分顯示)

資料 年月	產業別	淨利總和[仟元]	企業家數
201212	01民生必需品製造業	101829850	2427
201212	02金屬製造業	26063081	2308
201212	03非金屬製造業	46306559	2704
201212	04紙印刷業	5006384	443
201212	05器械業	91035050	3262
201212	06批發業	89560881	8842
201212	07零售業	44179332	1455
201212	08電子製造業	245227809	2210
201212	09服務業	133059768	4201
201212	10營建業	116246656	7929
201212	11投資業	72867681	3024
201212	12金融保險業	16247417	384

負債、營收、淨利(損))計算總和，並加入「其他分類資訊」做交叉分析(例如：產業別、規模別、發行狀態等)，分別建立不同的資料集。本中心為中小企業財務報表的重要資料庫，藉由此類資料的公開，可以讓使用者知道中小企業財務狀況的變化趨勢。

Open Data之效益

本中心之Open Data自2015年2月開始上架以來，截至2020年初已累積下載約44.5萬次，各類資料的下載次數如表8，以消費金融類型的資料有較多的下載數，原因應係其與民眾日常生活較相關，且該類資料的資料量大兼面向多元，使用者可以從中發掘更多元素加以組合觀察。

表8 Open Data累計下載次數統計
(截至2020/03)

資料類別 (項目數)	下載次數
個人授信統計資訊(17項)	7.7萬次
房貸類統計資訊(9項)	6.3萬次
個人信貸類統計資訊(4項)	1.7萬次
車貸類統計資訊(4項)	1.1萬次
信用卡統計資訊(13項)	5.3萬次
循環信用及現金卡統計資訊(24項)	5.5萬次
企業授信統計資訊(22項)	6.8萬次
財報類統計資訊(18項)	4.9萬次
其他公開資訊(11項)	5.2萬次

在大數據的時代下，對於資料的需求及應用面向都相當多樣化，開放資料係在做好資料去識別化、且妥善保護個資的前提下，藉由集結眾人之力，各取所需地應用資料，藉此更能充分發揮資料的能量。本中心之Open Data，可做為瞭解金融整體環境現況與變動趨勢之入門資料，用以串接社會各界的資料需求，將資料可發揮的效益推及到民間。

對於金融業者而言，本中心Open Data可協助其進行市場調查，觀察消費者需求的變動(如貸款量的增減)、以及供給者的回應(如利率的變化)，藉此找尋潛在客戶、調整既有商品或開發新商品。而對於新創的FinTech金融科技業者，尚有額外的應用方式，舉例而言，部分FinTech業者係從事讓民眾能有效率地獲得各類金融資訊之業務，則此類資料即可在其網站或APP中進行加值應用；至於部分業者則係以做為借貸雙方的中介平台為主，則此類資料即可協助其評估在該平台上應該置放哪類的商品為主(即評估市場需求)、以及利率制定的參考(即評估市場供給現況)。

對於非金融業者而言，本中心Open Data可協助其觀察時事。例如研究學者在撰寫論文或研究報告時，可藉由瞭解信用環境的狀況，藉以評估如何挖掘主題、選擇主題，亦可在其研究結果的分析中，找尋各類解釋的可能性；又例如大眾媒體在報導時事，亦可藉此資料觀

察社會趨勢脈動，舉例而言，過去曾經報導過的主題，至少包括「老年人口的刷卡行為」³、「青年人口的信貸變化」⁴、「男女理財行為差異」⁵、「因應降息環境的信貸利率趨勢推估」⁶...等，皆曾在其報導內引用本中心Open Data之資料，在各媒體通路上加以討論，供社會大眾參考。

對於民眾而言，本中心Open Data除可滿足其自身研究興趣外，尚可協助其更瞭解自身信用狀況。儘管目前民眾已可透過向本中心申請個人信用報告，以瞭解自己的負債狀況及繳款狀況，但藉由Open Data則可以知道自己在這個市場裡的相對位置(例如：自己借的金額跟利率，跟別人比起來是多還是少)，繼而對於自己的信用狀況有更立體的瞭解。

未來之展望

於2019年6月時，行政院推行「智慧政府行動方案」，其推動策略為極大化政府開放資料加值應用，致力提升開放資料品質，訂定資料交換標準，期能達到開放資料透明，極大化加值應用之目標。在此前提下，本中心將配合金管會之指示，盡力協助政策之順利推行。

在開放的資料集數量上，本中心將配合新增資料集，將當前國人較關心的主題(如普惠金融等)，研擬有無適合的資料內容加以開放；在資料品質上，亦將盡力協助使金管會之開放資料品質合格率，能持續在政府機關的評比名列前茅；另外未來規劃將本中心資料集採用符合OAS標準的Open API⁷型式，讓使用者能更有效率地運用資料，做更多的加值應用。

3 2018/10/14，「經濟生活惡化？「下流老人」刷卡借貸增」，東森新聞雲。

4 2019/12/15，「台灣青年信貸金額 7年來平均增11萬」聯合新聞網。

5 2016/6/26，「金融大數據 男是信貸大戶 女更常當卡奴」，聯合新聞網。

6 2020/3/29，「國銀衝信貸 祭出超殺利率」，經濟日報。

7 API是Application Programming Interface（應用程式介面）的縮寫。藉由Open API，使用者可以透過開發額外的應用程式，以更有效率的方式存取資料。