

# 衛生福利資料科學中心資料檔檢核標準程序

衛生福利資料科學中心新增資料檔時，依下列程序進行檔案檢核作業。

## 一、歷年(各月)資料筆數與欄位數

- **資料筆數**：同一檔案進行跨年度檢核，觀察各年度之資料筆數，若突然多或特別少，則原始提供的資料可能有誤，應洽資料原提供單位確認。
- **欄位數**：檢查檔案歷年(各月)欄位數，若歷年間有差異可能是新年度的欄位有部分增刪，應特別註記或洽資料原提供單位確認。

## 二、調整新年度資料檔各變項基本資料之一致性

- 各年度間同一變項應在名稱、型態及長度保持一致，但若配合政策改變而修改相關內容(如卡號)則不在此限。
- 資料庫與譯碼簿在變項名稱、資料型態及長度的部分應一致。
- 同一變項各資料原提供單位所提供之譯碼簿名稱有異，在資料庫與譯碼簿間差異較大者，應以資料庫為主修改譯碼簿相關內容。
  - **名稱**：包括中文欄名與英文欄名，檢視各年度是否有差異並一致化。
  - **型態**：檢視各年度是否有差異，須將 Char 與 Num 一致化。
  - **長度**：檢視各年度是否有差異，若自該年度後統一改變長度，則須加註。
  - **順序**：年度間同一檔案之各欄位順序應盡可能保持一致。

## 三、鍵值唯一化

- 檢核資料檔串檔鍵值應唯一化。

## 四、個別欄位分析

- **同檔同欄位譯碼分析**：
  - **類別欄位**：
    - ◆ **譯碼分布-合理性**：分析各欄位「組成(譯碼類別)」分布情形。另「總數」雖已於第一部份歷年(各月)資料筆數分析過，但與「組成(譯碼類別)」分布一起呈現有助於比較。
    - ◆ **資料品質-正確性**：留意各欄位之遺漏值比率，並註記於檢誤總表，遺漏值比率太高亦為資料品質不佳。另須留意無相對應之錯誤譯碼。
  - **日期欄位**：
    - ◆ **檢核年齡-合理性**：將生日換算成年齡，藉合理性以判定正確性。
    - ◆ **檢核日期-正確性**：透過轉換年齡時，可檢查日期錯誤情形。
    - ◆ **檢核年月分布**：計算檔案筆數年月分布，觀察歷年資料是否有不合理情形。
  - **數字欄位**：不一定是 Num 格式欄位，部分 Char 格式欄位亦為序位/連續資料，分析資料分布情形(如分析 Range、Mean、Mode、Max、75% Q3、Median、25% Q1、Min 等)。
- **同檔跨欄位交叉驗證**：資料庫品質問題僅由個別欄位分析無法真實呈現，需透過不同欄位

交叉檢核能進一步找出資料庫之問題，先針對個別檔案中可進行交叉檢核之欄位分析，如性別，比對由原始建檔與身份證號第一碼轉出之性別、就醫日期與生日相減的合理性等。

## 五、參照資料庫手冊增修

- 欄位增刪：年度間欄位的增減應於譯碼簿中同時更新並加註說明。該年度若有新增欄位，依其欄位特性，增至性質相近之欄位中；若減少欄位，則須註記於手冊中提醒使用者。
- 依上述進行系統性資料品質盤點後，據此針對標示資料品質待確認之重點項目，逐一查詢是否有新增譯碼一併編列、修正或提報提案單。
- 過程中如有關於資料檢誤條件可納入此SOP檔案中說明檢誤邏輯，使檢核模式更臻至完善。

## 六、製作資料檔檢誤總表

- 依上述檢核原則，製作個別檔案描述性報告，並彙整為資料檔檢誤總表。
- 盤點各欄位檢誤情形，標示資料品質較差、無誤、遺漏值情形等註記，有助於明瞭資料確認的進度與後續須關注資料品質較差者之重點項目。
- 資料檔檢誤總表如下：

衛生福利資料科學中心資料檔檢誤總表(1)

資料檔代碼	資料檔名稱	資料期間	使用手冊最新更新日期	描述性報告紀錄	資料提供單位	備註
Health-01	全民健保處方及治療明細檔_門急診	1998-2015	2017/6/20	99年	健保署	
Health-02	全民健保處方及治療明細檔_住院	1998-2015	2017/7/19	99年	健保署	
Health-03	全民健保處方及治療明細檔_藥局	2000-2015	2017/6/6	99年	健保署	
Health-04	全民健保處方及治療醫令明細檔_門急診	2000-2015	2017/6/6	99年	健保署	
Health-05	全民健保處方及治療醫令明細檔_住院	2000-2015	2017/6/6	99年	健保署	
Health-06	全民健保處方及治療醫令明細檔_藥局	2000-2015	2017/6/6	99年	健保署	
Health-07	全民健保承保檔	2000-2015	2017/6/6	99年	健保署	
Health-08	全民健保重大傷病檔	2001-2015	2017/6/6	101年	健保署	註1
Health-09	出生通報檔	2001-2015	2017/5/11	101年	國民健康署	
Health-10	死因統計檔	1971-2016	2017/6/28	101年	統計處	
Health-11	醫療院所評鑑等級	1998-2015	2012/7/23	101年	統計處	
Health-12	醫院醫療服務量檔	1994-2016	2017/8/30	101年	統計處	
Health-13	醫療機構現況檔	1994-2016	2017/8/29	101年	統計處	
Health-14	癌症登記檔LF	2007-2014	2017/6/6	101年	國民健康署	
Health-15	癌症登記檔SF	2007-2014	2017/6/6	101年	國民健康署	
Health-16	癌症登記檔TCDB	2002-2006	2017/6/6		國民健康署	
Health-25	醫事機構基本資料檔	2002-2016	2017/5/1	100、105年	醫事司	
Health-26	原住民出生通報檔	2006-2013	2016/3/15	99年	原民會	註1
Health-27	原住民死亡檔	2006-2013	2010/12/31	101年	原民會	註1
Health-28	專科醫師證書主檔	2010-2016	2017/3/17	101、105年	醫事司	
Health-29	醫事人員基本資料檔	2010-2016	2017/2/22	101、105年	醫事司	
Health-30	多重死因檔	2008-2016	2017/6/28		統計處	
Health-31	延遲申報死因檔	1994-2016	2017/6/28		統計處	
Health-32	交通事故檔(BAS、PSN)	2003-2015	2017/6/6	101年	內政部警政署	註2
Health-44	三高調查檔	2002-2007	2013/1/4		國民健康署	註3
Health-45	癌症登記年報檔		2017/6/16		國民健康署	
Health-48	台灣出生世代調查	1979-2014	2017/5/10		國民健康署	
Health-49	家庭與生育率研究調查	2005	2013/1/29		國民健康署	註3
Health-50	青少年健康行為調查	1965-2008	2013/1/31		國民健康署	
Health-51	罕見疾病通報資料庫	2006-2010	2013/1/25		國民健康署	
Health-52	人工生殖資料庫	2012	2014/12/19	103年	國民健康署	
Health-53	延遲申報多重死因檔	1998-2010	2017/6/28		統計處	
Health-54	癌症篩檢_子宮頸抹片篩檢檔	2010-2016	2017/6/6		國民健康署	
Health-55	癌症篩檢_大腸癌篩檢檔	2004~2013	2017/6/6	103年	國民健康署	
Health-56	癌症篩檢_乳癌篩檢檔	2010~2013	2017/6/6	103年	國民健康署	
Health-57	癌症篩檢_口腔癌篩檢檔	2004~2013	2017/6/6		國民健康署	
Health-58	健康行為危險因子監測調查	2010~2013	2015/2/16		國民健康署	
Health-60	婦幼主題式資料庫	2007~2012	2017/4/27		統計處	
Health-61	法定傳染病個案通報系統資料庫-防疫資料庫	2004-2014	2016/9/14	105年	疾病管制署	註1
Health-62	法定傳染病個案通報系統資料庫-機敏資料庫	1991-2015	2017/6/7	105年	疾病管制署	註1

衛生福利資料科學中心資料檔檢誤總表(2)

資料檔代碼	資料檔名稱	資料期間	使用手冊最新更新日期	描述性報告紀錄	資料提供單位	備註
Health-63	症候群重症系統資料庫_防疫資料庫	1991-2015	2016/9/13	105年	疾病管制署	註1
Health-64	症候群重症系統資料庫_機敏資料庫	2002-2008	2017/6/7	105年	疾病管制署	註1
Health-65	即時疫情監視及預警系統資料庫	2002-2008	2016/9/13	105年	疾病管制署	註1
Health-66	症狀通報系統資料庫-類流感及腹瀉群聚	2007-2015	2017/6/7	105年	疾病管制署	註1
Health-67	症狀通報系統資料庫-水痘群聚	2005-2015	2017/6/7	105年	疾病管制署	註1
Health-68	症狀通報系統資料庫- H5N1流感調查病例	2014-2015	2017/6/7	105年	疾病管制署	註1
Health-69	定點醫師通報資料庫	2005-2014	2016/9/30	105年	疾病管制署	註1
Health-70	學校傳染病監視通報資訊系統資料庫	1999-2009	2017/6/7	105年	疾病管制署	註1
Health-71	人口密集機構傳染病個案通報檔-防疫資料庫	2004-2015	2016/9/10	105年	疾病管制署	註1
Health-72	人口密集機構傳染病個案通報檔-機敏資料庫	2004-2015	2017/6/7	105年	疾病管制署	註1
Health-73	結核病資料庫	2004-2015	2017/6/7	105年	疾病管制署	註1
Health-74	愛滋病資料庫	2002-2014	2017/6/7	105年	疾病管制署	註1
Health-75	外籍勞工健康檢查資料	1984-2015	2016/9/21	105年	疾病管制署	註1
Health-76	病毒合約實驗室檢驗結果明細檔	1999-2015	2017/6/7	105年	疾病管制署	註1
Health-77	真菌菌株資料庫	1999-2015	2016/9/30	105年	疾病管制署	註1
Health-78	淋病雙球菌菌株資料庫	1997-2006	2016/9/30	105年	疾病管制署	註1
Health-79	全國性預防接種資料庫	2006-2012	2017/6/7	105年	疾病管制署	註1
Health-80	台灣院內感染監視系統資料庫	2004~ 2014	2017/9/13	105年	疾病管制署	註1
Health-81	全民健保特約醫事機構資料檔	2007-2014	2017/6/7		健保署	
Society-07	原住民族戶籍資料庫	2000~2016	2010/12/31	99年	原民會	註1
Society-09	個人屬性檔(抽樣檔)	2006-2013	2017/6/7		統計處	
Society-10	國民健康訪問調查檔	2000-2010	2017/9/14	98年度資料	國民健康署	
Society-12	中老年調查檔	2001-2013	2013/1/11		國民健康署	註3
Society-13	吸菸調查檔	1989-2007	2013/1/11		國民健康署	
Welfare-1	少年身心狀況調查	2004-2009	2014/6		統計處	
Welfare-2	居家服務補助使用者狀況調查	2007	2015/1		統計處	
Welfare-3	居家服務補助使用者滿意度調查	2011	2015/1/5		統計處	
Welfare-4	低收入戶生活狀況調查	2013	2015/1/5		統計處	
Welfare-5	老人狀況調查	2009-2013	2014/10/31	94及98年度資料	統計處	
Welfare-6	身心障礙者生活需求調查	2000-2011	2017/2/2	89及100年度資料	統計處	
Welfare-7	單親家庭狀況調查	2010	2014/12/18		統計處	
Welfare-8	婦女生活狀況調查	1998-2011	2014/12/19		統計處	
Welfare-9	老人福利機構需求概況調查	1999	2015/3/6		統計處	
Welfare-10	身心障礙資料庫	2014-2016	2017/6/15	104年	社會及家庭署	註1
Welfare-11	低收入戶及中低收入戶資料庫	2012-2015	2017/6/7		社會救助及社工司	註1
Welfare-12	家暴通報明細檔	2011-2016	2017/7/18	104年	保護服務司	註1
Welfare-14	兒童及少年保護通報明細檔	2011-2016	2017/7/18	104年	保護服務司	註1
Welfare-15	性侵害通報明細檔	2011-2016	2017/7/19	104年	保護服務司	註1
Welfare-16	育兒津貼核定名冊	2013	2017/2/8	106年	社會及家庭署	註1
Welfare-17	弱勢兒童及少年生活扶助核定名冊	2013	2017/2/6	106年	社會及家庭署	註1
Welfare-18	弱勢家庭兒童及少年緊急生活扶助核定名冊	2013	2017/2/10	106年	社會及家庭署	註1
Health-59	遺傳疾病資料庫					

註1：特定檔案之申請，需另備檔案提供單位授權使用之公文。

註2：需提供成果報告予內政部警政署。

註3：Health-44提供年度：2002、2007。Health-49提供年度：1965、1967、1970、1973、1980、1986、1992、1998、2004、2008。