



# 國家級人體生物資料庫整合平台 醫療資訊整合說明暨討論會議 2020/10/30會議報告

楊奕馨

國家衛生研究院 癌症研究所

[Email: yhyang@nhri.edu.tw](mailto:yhyang@nhri.edu.tw)

<http://nicr.nhri.org.tw/pi/yangyh-cv/>

2

## 上次 2020/03/25 會議之後

感謝

北醫, 台大

中山醫, 中國醫, 中榮

義大, 高醫

# 國家級人體生物資料庫整合平台 共同資料欄位說明

# Biobank 遇上 醫療大數據

病人的醫療照護朝向避免 **undesired outcomes** 發生

4

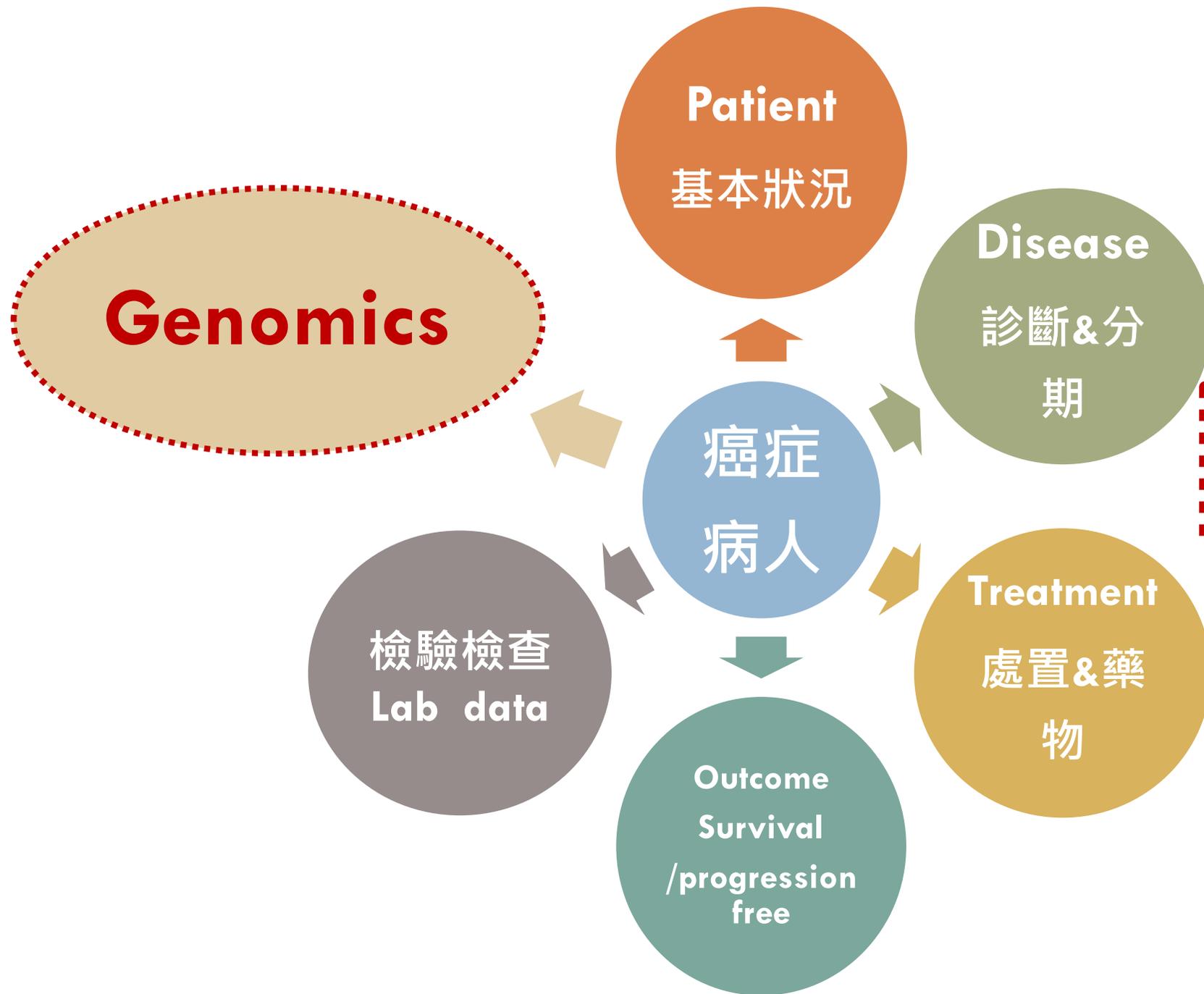


**Biobank**

genomic data 的  
重要來源

**醫療數據**

近幾年國際上許多先進國家的biobank, 納入醫療數據 (UK biobank, Biobank popgen, All of Us, Biobank Japan, ...)



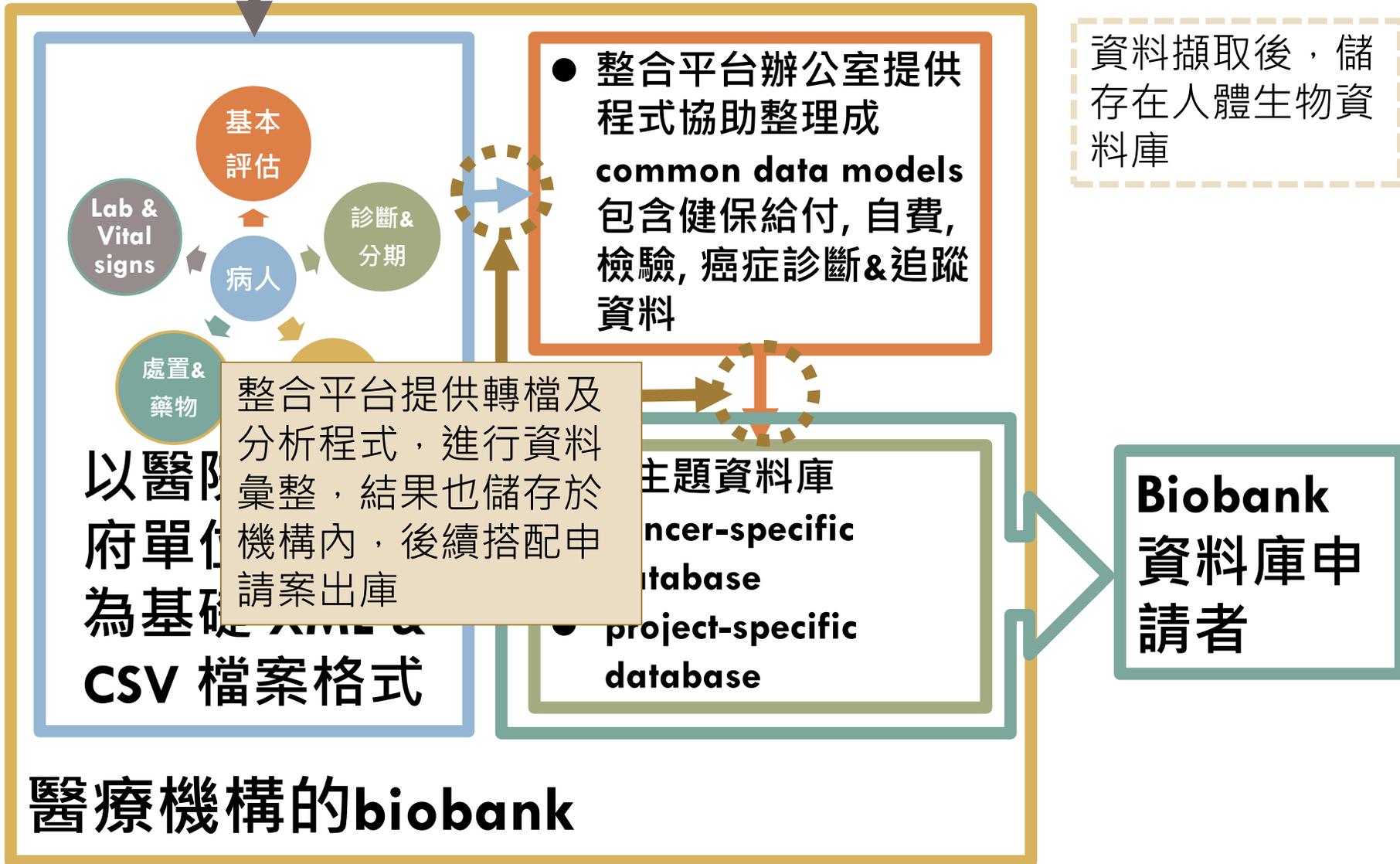
## 臨床醫療資料 (EMR):

第一階段

- 結構式資料
- 非結構式資料
  - 文字報告
  - 影像

# 電子病歷系統

第一階段是先將人體生物資料庫所收錄的病人之醫療資料先由醫院電子病歷資料庫及癌症登記資料庫中擷取出來



資料擷取後，儲存在人體生物資料庫

● 整合平台辦公室提供程式協助整理成 common data models 包含健保給付, 自費, 檢驗, 癌症診斷&追蹤資料

整合平台提供轉檔及分析程式, 進行資料彙整, 結果也儲存於機構內, 後續搭配申請案出庫

以醫院、政府單位、學術單位為基礎, 以XML & CSV 檔案格式

主題資料庫  
Cancer-specific database  
project-specific database

Biobank 資料庫申請者

醫療機構的biobank

# 電子病歷系統

門診醫療申報檔案 (TOTFA, XML檔)  
住院醫療申報檔案 (TOTFB, XML檔)  
檢驗檢查資料檔 (LAB, XML檔)  
自費藥品 & 自費檢驗資料  
癌症登記長表資料檔 (CRLF, TXT檔)  
癌症登記短表資料檔 (CRSF, TXT檔)  
死亡資料檔 (DEATH, TXT檔)  
癌症個案管理追蹤資料檔 (CASE, TXT檔)

**A**

醫療機構  
biobank

## 共同資料欄位 (common data elements):

門診醫療申報檔案總表 (TOTFAE)  
門診醫療申報檔案醫令 (健保給付: TOTFAO1, 自費: TOTFAO2)  
住院醫療申報檔案總表 (TOTFBE)  
住院醫療申報檔案醫令 (健保給付: TOTFBO1, 自費: TOTFBO2)  
檢驗檢查資料檔 (健保給付: LABH1, 自費: LABH2)  
癌症登記長表資料檔 (CRLF)  
癌症登記短表資料檔 (CRSF)  
死亡資料檔 (DEATH)  
癌症個案管理追蹤資料檔 (CASE)

**B**

## 主題資料庫:

- disease-specific database **C**
- project-specific database

8

# 執行步驟

9

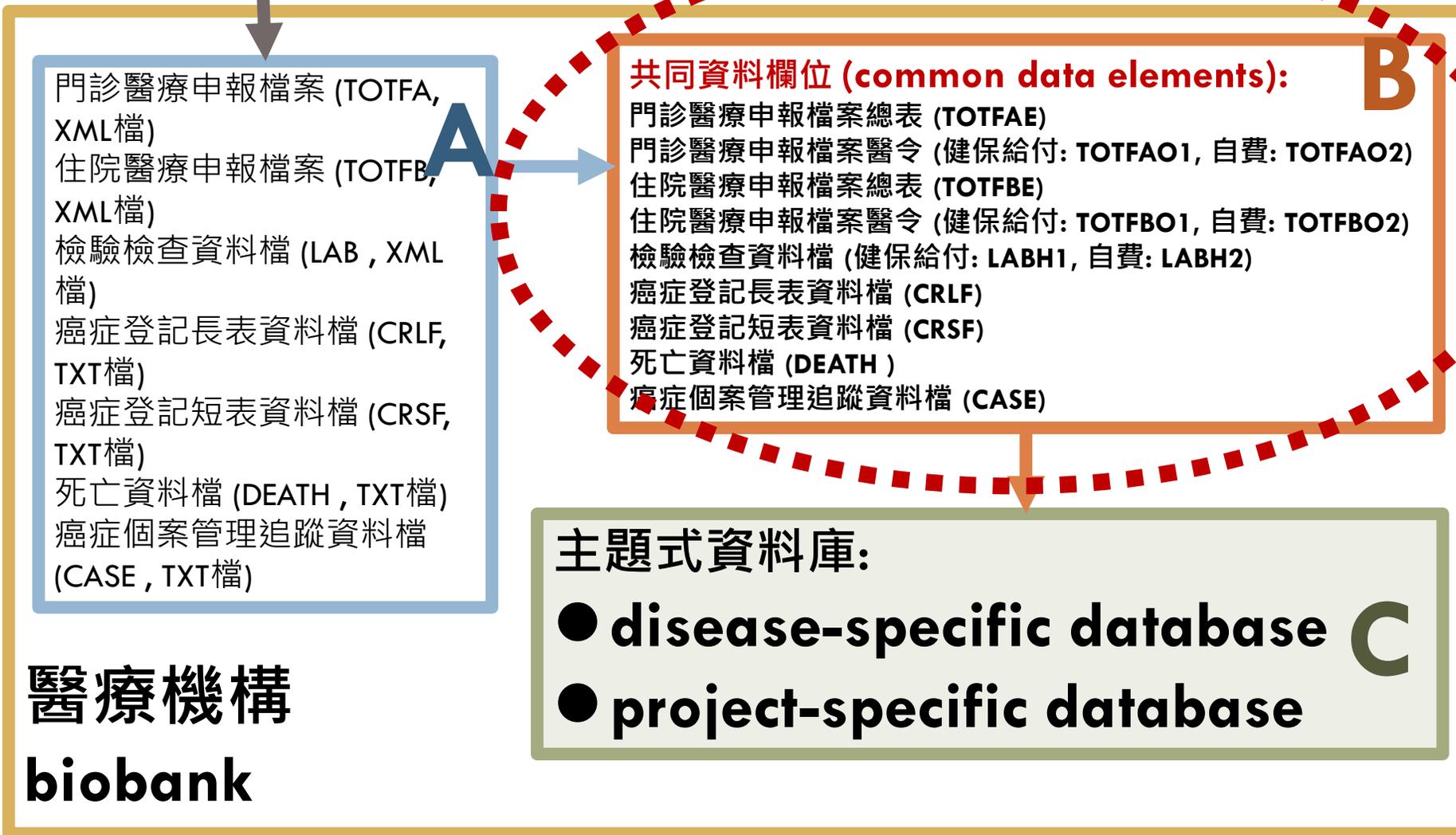
## 步驟一：決定匯入臨床資料的方案

方案一：資訊室提供 A框(上傳健保署&國健署資料檔形式)給自己醫院 biobank, 國衛院再協助轉成B框 & C框

方案二：資訊室依據B框的欄位定義, 將資料整理成CSV資料檔形式給自己醫院 biobank, 國衛院再協助轉成 C框

# 電子病歷系統

方案一：資訊室提供 A框的資料檔形式給自己醫院 biobank, 國衛院再協助轉成B框 & C框



# 電子病歷系統

方案二：資訊室依據B框的欄位定義，將資料整理成CSV資料檔形式給自己醫院 biobank, 國衛院再協助轉成 C框

## 共同資料欄位 (common data elements):

門診醫療申報檔案總表 (TOTFAE)  
門診醫療申報檔案醫令 (健保給付: TOTFAO1, 自費: TOTFAO2)  
住院醫療申報檔案總表 (TOTFBE)  
住院醫療申報檔案醫令 (健保給付: TOTFBO1, 自費: TOTFBO2)  
檢驗檢查資料檔 (健保給付: LABH1, 自費: LABH2)  
癌症登記長表資料檔 (CRLF)  
癌症登記短表資料檔 (CRSF)  
死亡資料檔 (DEATH)  
癌症個案管理追蹤資料檔 (CASE)

B

## 主題資料庫:

- disease-specific database
- project-specific database

C

出庫  
提供  
研究

醫療機構 biobank

## 步驟一：決定匯入臨床資料的方案

方案一：資訊室提供 A框(上傳健保署&國健署資料檔形式)給自己醫院 biobank, 國衛院再協助轉成B框 & C框

# A框 資料說明

國家級人體生物資料庫整合平台

## 共同資料欄位 格式說明

國家衛生研究院

2020 v2

單元二：資料庫內容欄位說明

- 一、門診醫療申報檔案 (檔案代碼:TOTFA)
- 二、住院醫療申報檔案 (檔案代碼:TOTFB)
- 三、檢驗檢查資料檔 (檔案代碼:LAB)
- 四、癌症登記長表資料檔 (檔案代碼:CRLF)
- 五、癌症登記短表資料檔 (檔案代碼:CRSF)
- 六、死亡資料檔 (檔案代碼:DEATH)
- 七、癌症個案管理追蹤資料檔 (檔案代碼:CASE)

附錄

附件一：「特約醫事服務機構門診醫療費用點數申報格式及填表說明」(XML 檔案格式) (107.7.11)

附件二：門診申報資料上傳格式作業說明

附件三：「特約醫事服務機構住院醫療費用點數申報格式及填表說明」(XML 檔案格式) (107.7.11)

附件四：住院申報資料上傳格式作業說明

附件五：「特約醫事服務機構上傳檢驗（查）結果、人工關節植入物資料及出院病歷摘要格式說明

附件六：特約醫事服務機構上傳檢驗（查）結果之作業說明

附件七：中央健康保險署 CODE BOOK 譯碼簿 (2019/3/12 版本)

附件八：橘色標示附件一、三、五中需要擷取的欄位

附件九：精簡列表附件一、三、五中需要擷取的欄位

## 步驟一：決定匯入臨床資料的方案

方案二：資訊室依據B框的欄位定義，將資料整理成CSV資料檔形式給自己醫院 biobank, 國衛院再協助轉成 C框

# 電子病歷系統

## 共同資料欄位 (common data elements):

門診醫療申報檔案總表 (TOTFAE)  
門診醫療申報檔案醫令 (健保給付: TOTFAO1, 自費: TOTFAO2)  
住院醫療申報檔案總表 (TOTFBE)  
住院醫療申報檔案醫令 (健保給付: TOTFBO1, 自費: TOTFBO2)  
檢驗檢查資料檔 (健保給付: LABH1, 自費: LABH2)  
癌症登記長表資料檔 (CRLF)  
癌症登記短表資料檔 (CRSF)  
死亡資料檔 (DEATH)  
癌症個案管理追蹤資料檔 (CASE)

**B**

## 主題式資料庫:

- **disease-specific database**
- **project-specific database**

**C**

出庫  
提供  
研究

醫療機構 biobank

## 步驟二：

機構內盤點各項資料檔  
來源&負責單位

# B框: Common Data Model / Elements 檔案名稱

- 門診醫療申報檔案總表 (TOTFAE)
  - 門診醫療申報檔案健保給付醫令 (TOTFAO1)
  - 門診醫療申報檔案自費項目醫令 (TOTFAO2)
    - 自費項目以癌症藥品為主, 如附件:自費項目.pdf
  - 住院醫療申報檔案總表 (TOTFBE)
  - 住院醫療申報檔案健保給付醫令 (TOTFBO1)
  - 住院醫療申報檔案自費項目醫令 (TOTFBO2)
    - 自費項目以癌症藥品為主, 如附件:自費項目.pdf
  - 檢驗檢查資料檔格式(包含健保給付及自費項目) (LABH1 & LABH2)
- 癌症登記長表資料檔 (檔案代碼:CRLF)
  - 癌症登記短表資料檔 (檔案代碼:CRSF)
  - 死亡資料檔 (檔案代碼:DEATH)
  - 癌症個案管理追蹤資料檔 (檔案代碼:CASE)

門診醫療申報檔案總表格式  
(檔案名稱: TOTFAE, CSV 檔案格式)

# 欄位格式範例

符號	欄位 ID	資料名稱	長度	屬性	
*	t2	服務機構代號	10	X	衛生福利部編定之
*	t3	費用年月	5	X	第1、2、3碼為民國位者前補0。例如5
*	t5	申報類別	1	X	1:送核 2:補報
*	t6	申報日期	7	X	第1、2、3碼為民國位者前補0。例如5
*	d1	案件分類	2	X	一、詳註11、註19 二、不足位者前補

檢驗檢查資料檔格式(包含健保給付及自費項目)  
(檔案名稱: LABH, CSV 檔案格式)

欄位 ID	資料名稱	長度	屬性	中文名稱/資料說明
h1	報告類別	1	X	一、報告類別代碼：1：檢體檢驗報告；2：影像報告；3：病理報告；4：人工關節植入物資料。 二、報告類別代碼「1」屬於「非報告型資料」；報告類別代碼「2」及「3」屬於「報告型資料」。 三、本欄為必填欄位。

表7-1、癌症個案管理追蹤資料檔 (檔案名稱: CASE)

[每個個案有多筆追蹤資料]

欄位 ID	資料名稱	長度	屬性	中文名稱/資料說明
MRN	病歷號碼	10	X	
m2	輸入日期	8	X	輸入資料日期, 西元年月日 YYYYMMDD
m3	追蹤日期	8	X	西元年月日 YYYYMMDD
m4	追蹤方式	1	X	1：門診訪視 2：住院訪視 3：電訪 4：其他 5：病歷查閱

表4-1、癌症登記長表資料檔 (檔案名稱: CRLF)

序號	欄位名稱	英文欄位名稱
1.1	申報醫院代碼	Reporting Hospital Code
1.2	病歷號碼	Medical Record Number
1.3	姓名 (此欄位不需要)	Name
1.4	身分證統一編號	ID Number

一、同健保醫療費用點數「申報日期」。  
二、第1、2、3碼為民國年份，不足位者前補0。

# 藍色上傳健保署資料, 紫色上傳國健署, 紅色自費項目

- 門診醫療申報檔案總表 (TOTFAE)
- 門診醫療申報檔案健保給付醫令 (TOTFAO1)
- 門診醫療申報檔案自費項目醫令 (TOTFAO2)
  - 自費項目以癌症藥品為主, 如附件:自費項目.pdf
- 住院醫療申報檔案總表 (TOTFBE)
- 住院醫療申報檔案健保給付醫令 (TOTFBO1)
- 住院醫療申報檔案自費項目醫令 (TOTFBO2)
  - 自費項目以癌症藥品為主, 如附件:自費項目.pdf
- 檢驗檢查資料檔格式(包含健保給付及自費項目) (LABH1 & LABH2)
  - 癌症登記長表資料檔 (檔案代碼:CRLF)
  - 癌症登記短表資料檔 (檔案代碼:CRSF)
  - 死亡資料檔 (檔案代碼:DEATH)
  - 癌症個案管理追蹤資料檔 (檔案代碼:CASE)

## 步驟三

- 測試先以10個病人為例(同一癌症為佳), 依照CDM擷取出對應資料, 平台辦公室協助進行欄位比對, 確定內容符合
- 可能會往來數次, 直到項目皆符合
- 自費項目編碼整理

## 步驟四

依照共同資料欄位, 擷取biobank病人對應的臨床資料

依照 Biobank 管理規則, 取得 biobank 收案病人之 chart ID & biobank ID 的對應個案編碼名單

biobank ID	chart ID
:	:
:	:
:	:
:	:

醫院資訊室使用 chart ID 串聯電子病歷資料, 進 biobank 前, 僅留下 biobank ID 進入 biobank

門診醫療申報檔案總表 (TOTFAE), 門診醫療申報檔案醫令 (TOTFAO), 住院醫療申報檔案總表 (TOTFBE), 住院醫療申報檔案醫令 (TOTFBO), 檢驗檢查資料檔基本資料 (LABH), 癌症登記長表資料檔 (CRLF), 癌症登記短表資料檔 (CRSF) 死亡資料檔 (DEATH), 癌症個案管理追蹤資料檔 (CASE)

主題資料庫: disease-specific, project-specific databases

出庫

醫療機構的 biobank, 使用 biobank ID 處理出庫申請案

- **步驟五：資料除錯**
- 步驟六：測試執行主題資料檔程式
- 步驟七：驗證主題資料檔

24

## 步驟六 example - 以乳癌為例



# Example: 整理乳癌主題資料 (273欄位)

## 共同資料欄位

- 癌症登記長表資料檔 (檔案代碼:CRLF)
- 癌症登記短表資料檔 (檔案代碼:CRSF)
- 門診醫療申報檔案總表 (TOTFAE)
- 門診醫療申報檔案健保給付醫令 (TOTFAO1)
- 門診醫療申報檔案自費項目醫令 (TOTFAO2)
- 住院醫療申報檔案總表 (TOTFBE)
- 住院醫療申報檔案健保給付醫令 (TOTFBO1)
- 住院醫療申報檔案自費項目醫令 (TOTFBO2)
- 檢驗檢查資料檔格式(包含健保給付及自費項目) (LABH1 & LABH2)
- 死亡資料檔 (檔案代碼:DEATH)
- 癌症個案管理追蹤資料檔 (檔案代碼:CASE)

各藥品有無使用, 給藥日數, 開始日期, 結束日期			
Doxorubicin	Gemcitabine	TDM1	Anastrozole
LiposomalDoxorubicin	Carboplatin	Lapatinib	Exemestane
Epirubicin	UracilTegafur	Everolimus	Goserelin
Docetaxel	Capecitabine	Ribociclib	Leuprorelin
Paclitaxel	Vinorelbine	Palbociclib	Fulvestrant
Fluorouracil	Eribulin	Tamoxifen	Denosumab
Cyclophosphamide	Trastuzumab	Toremifene	Zoledronate
Methotrexate	Pertuzumab	Letrozole	

最靠近診斷日期檢驗檢驗結果			
白血球計數	平均紅血球體積	Creatinine_Blood	血糖
白血球計數日期	平均紅血球體積檢驗日期	Creatinine_Urine	血糖檢驗日期
Seg	紅血球分布寬度	Creatinine檢驗日期	三酸甘油酯
Seg 日期	紅血球分布寬度檢驗日期	AST	三酸甘油酯檢驗日期
紅血球計數	血小板計數	AST檢驗日期	白蛋白
紅血球計數日期	血小板計數日期	ALT	白蛋白檢驗日期
血紅素	凝血酶原時間	ALT檢驗日期	GGT
血紅素檢驗日期	Prothrombin_INR	ALP	GGT檢驗日期
血球比容值測定	凝血酶原時間檢驗日期	ALP檢驗日期	總膽固醇
血球比容值測定日期			總膽固醇檢驗日期

若需要可以再設定第二次治療的檢驗結果

# 乳癌主題資料檔的欄位擷取規則

27

1	欄位中文名稱	欄位英文名稱	資料說明定義	資料來源
138	共病症_moderate or severe liver disease	moderate or severe liver disease	癌症診斷日前一年出現2次門診診斷或1次住院診斷時定義為有該共病症。 ICD-9: 572.2~572.8, 456.0~456.2 ; ICD-10: I85.0, I85.9, I86.4, I98.2, K70.4, K71.1, K72.1, K72.9, K76.5, K76.6, K76.7	門診醫療申報檔案總表(TOTFAE): 主診斷代碼、次診斷代碼 住院醫療申報檔案總表(TOTFBE): 主診斷代碼、次診斷代碼
139	共病症_Metastatic solid tumor	Metastatic solid tumor	癌症診斷日前一年出現2次門診診斷或1次住院診斷時定義為有該共病症。 ICD-9: 196~199 ; ICD-10: C77~C80	門診醫療申報檔案總表(TOTFAE): 主診斷代碼、次診斷代碼 住院醫療申報檔案總表(TOTFBE): 主診斷代碼、次診斷代碼
140	共病症_AIDS/HIV	AIDS/HIV	癌症診斷日前一年出現2次門診診斷或1次住院診斷時定義為有該共病症。 ICD-9: 042~044 ; ICD-10: B20~B22, B24	門診醫療申報檔案總表(TOTFAE): 主診斷代碼、次診斷代碼 住院醫療申報檔案總表(TOTFBE): 主診斷代碼、次診斷代碼
141	白血球計數	WBC_count	確診時WBC count ( $10^3/uL$ )。最接近癌症診斷日的檢驗結果(診斷日前一個月)。08011C 全套血液檢查之WBC項目	檢驗檢查資料檔基本資料 (LABH): 醫令代碼, 檢驗項目名稱
142	白血球計數日期	Date of WBC count	YYYYMMDD	檢驗檢查資料檔基本資料 (LABH): 採檢/實際檢查/手術日期
143	Seg	Seg	確診時Seg(%)。最接近癌症診斷日的檢驗結果(診斷日前一個月)。08013C 白血球分類計數之Seg項目	檢驗檢查資料檔基本資料 (LABH): 醫令代碼, 檢驗項目名稱
144	Seg 日期	Date of Seg	YYYYMMDD	檢驗檢查資料檔基本資料 (LABH): 採檢/實際檢查/手術日期
	紅血球計數	RBC_count	確診時RBC count ( $10^6/uL$ )。最接近癌症診斷日的檢驗結果(診斷日前一個月)。08011C 全套血液檢查之RBC項目	檢驗檢查資料檔基本資料 (LABH): 醫令代碼, 檢驗項目名稱

**SAS 或是 other free software**

- 步驟五：資料除錯
- 步驟六：測試執行主題資料檔程式
- 步驟七：驗證(validate)主題資料檔



**謝謝!**

**Questions & Comments?**