

大數據中心申請與資料庫介紹

講者: 邱雯琪

2020/1/15

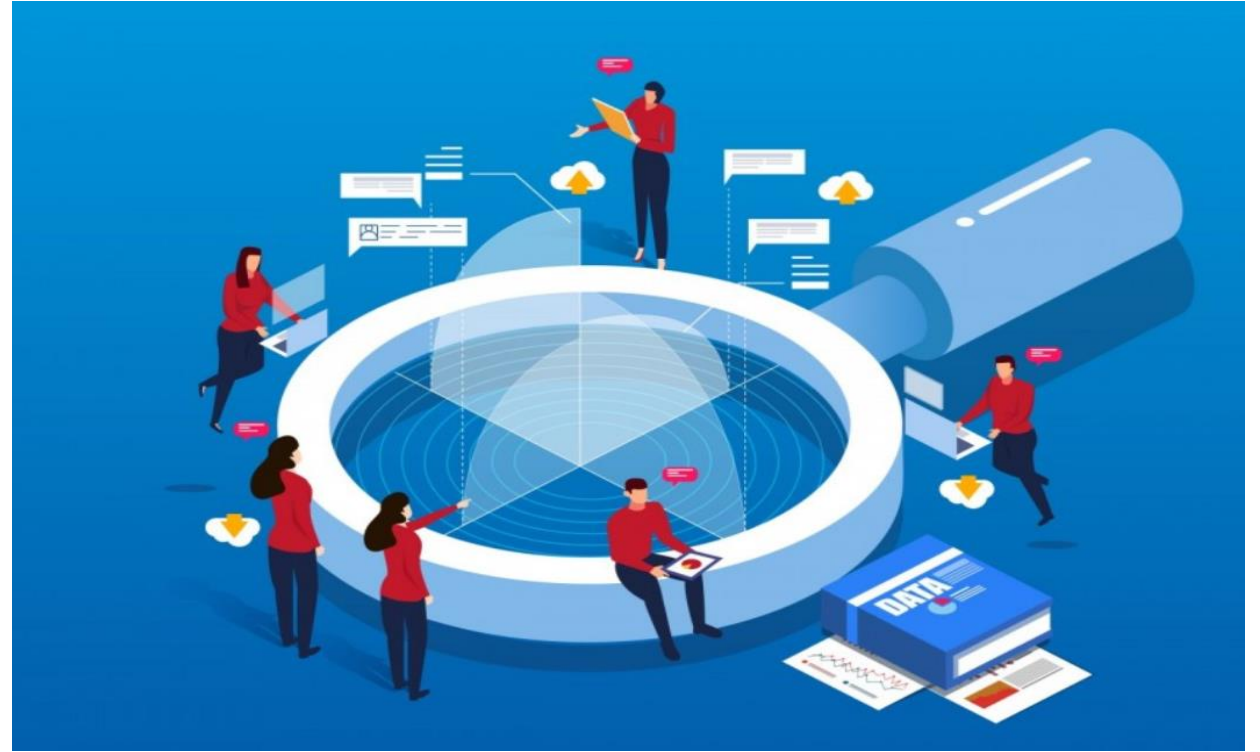


臺北榮民總醫院 大數據中心

Big Data Center

Outline

- 醫療大資料庫和軟體介紹
- 北榮大數據中心申請
- SQL新手初階小班課





1. 檔案放置PI處
2. 可慢慢想有空做
3. 需花錢購買



1. 資料更多元完整化
2. 須進加值中心跑資料
3. 資料庫不可以攜出
4. 需花錢購買



1. 在地化特色資料-臺北榮民總醫院病患過往就醫紀錄(包含生化抽血資料)、榮民資料
2. 檔案放置PI處
3. 不用錢



健保資料庫資料來源模式



各家醫院每月申報檔案



健康資料加值應用協作中心

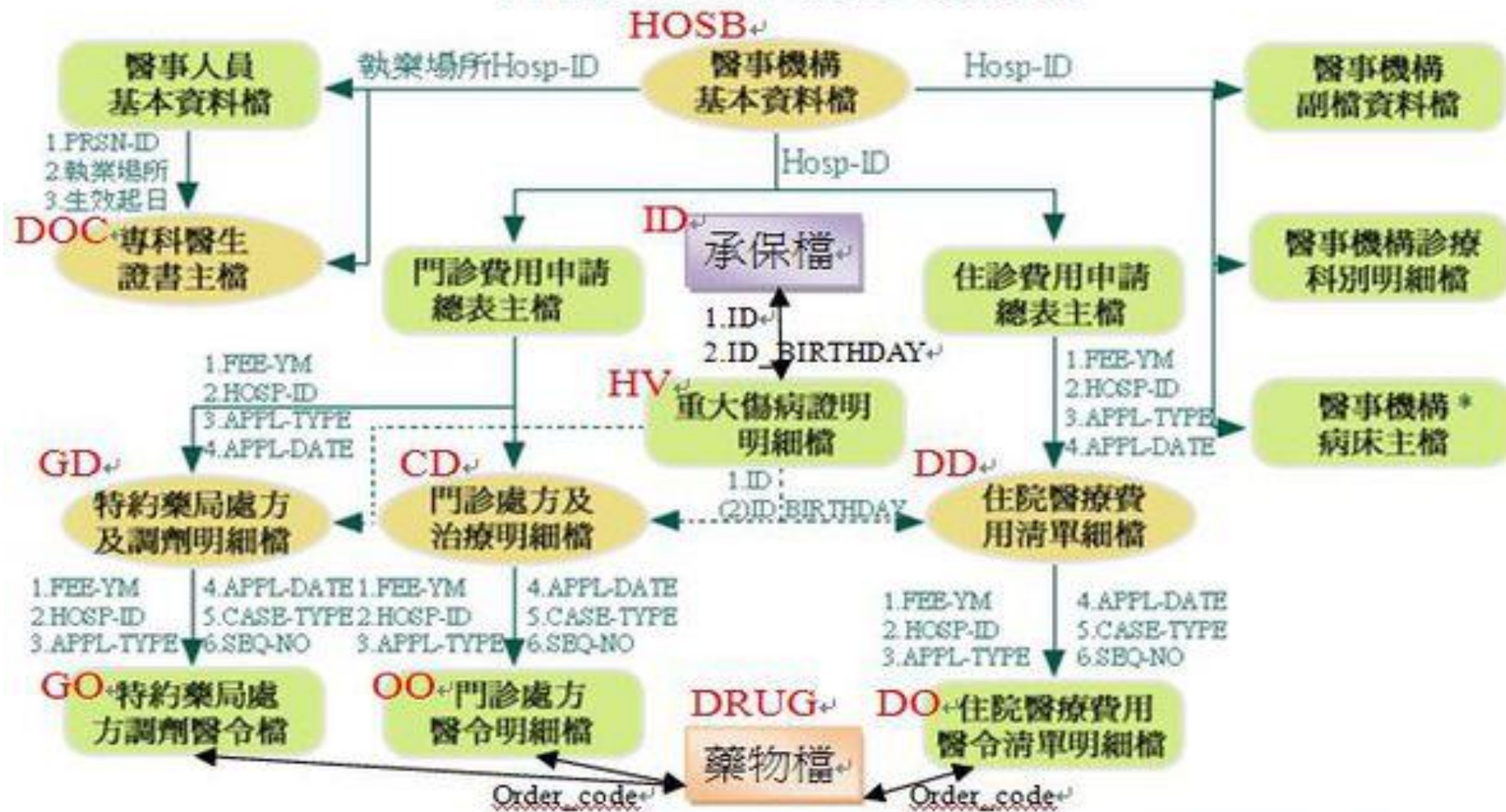
Health Information Application(CCHIA), Ministry of Health and Welfare

當初設計資料庫是為了申報處理給付而不是為了醫學研究.所以不一定符合所有醫療人員想要的資料

但我們真的需要那麼多資訊才能完成一篇研究嗎????

國衛院資料檔案之間關係

各檔案間串檔變項說明



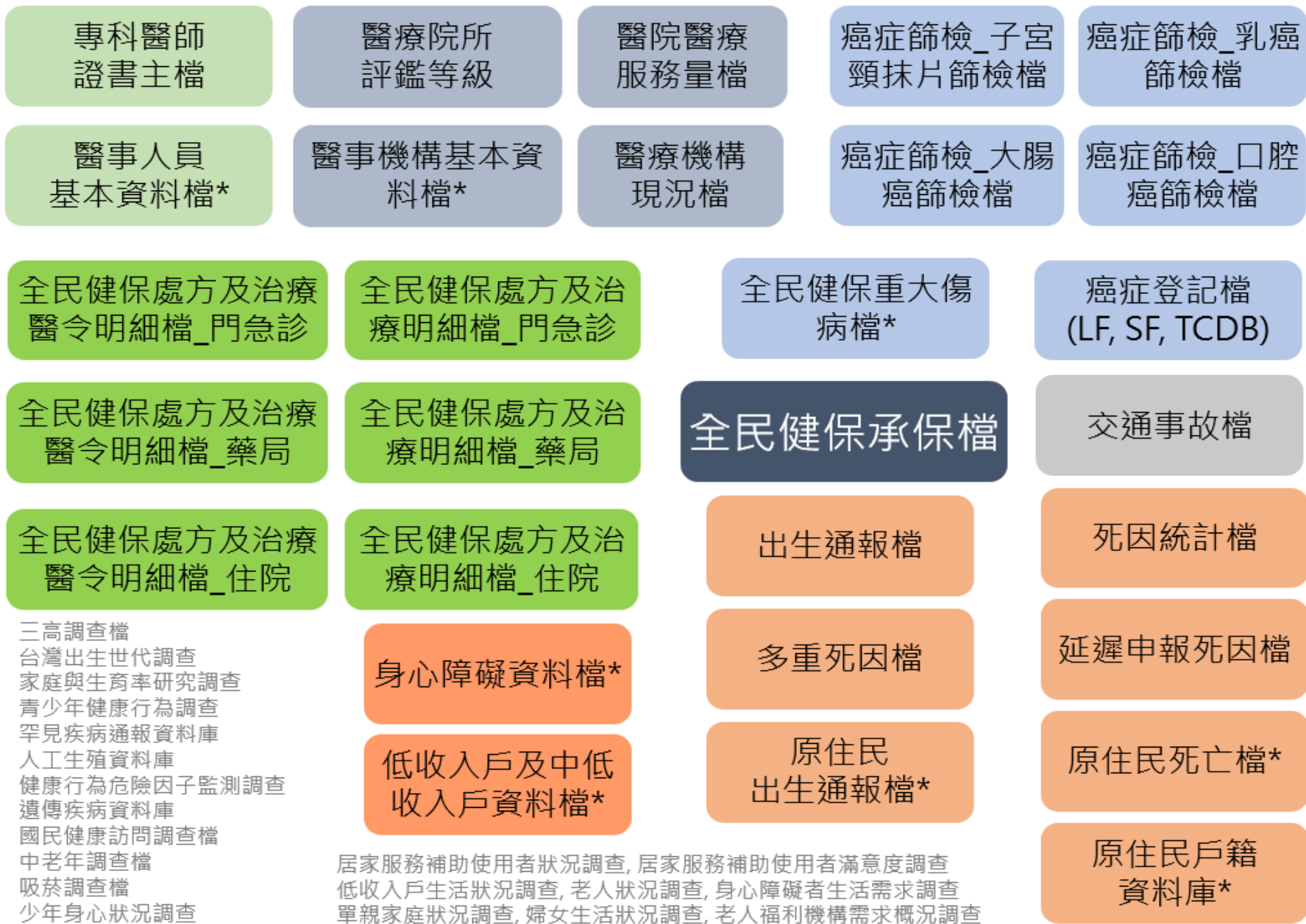
註:*須注意生效起訖日期

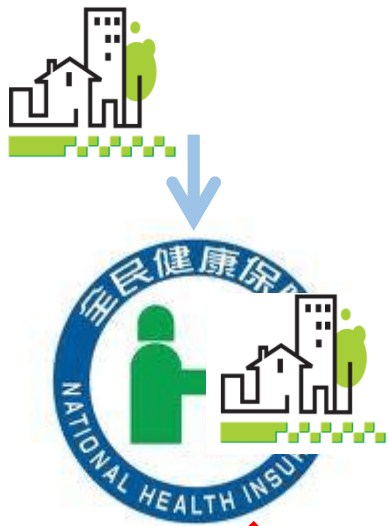
(2)可由ID+BIRTHDAY串檔

→ 各檔案間由所註明變項串檔可獲得對應資訊

→ 各檔案間可由所註明變項串檔,但未必獲得對應資料

現今衛服部整合政府各單位資料庫資料齊全

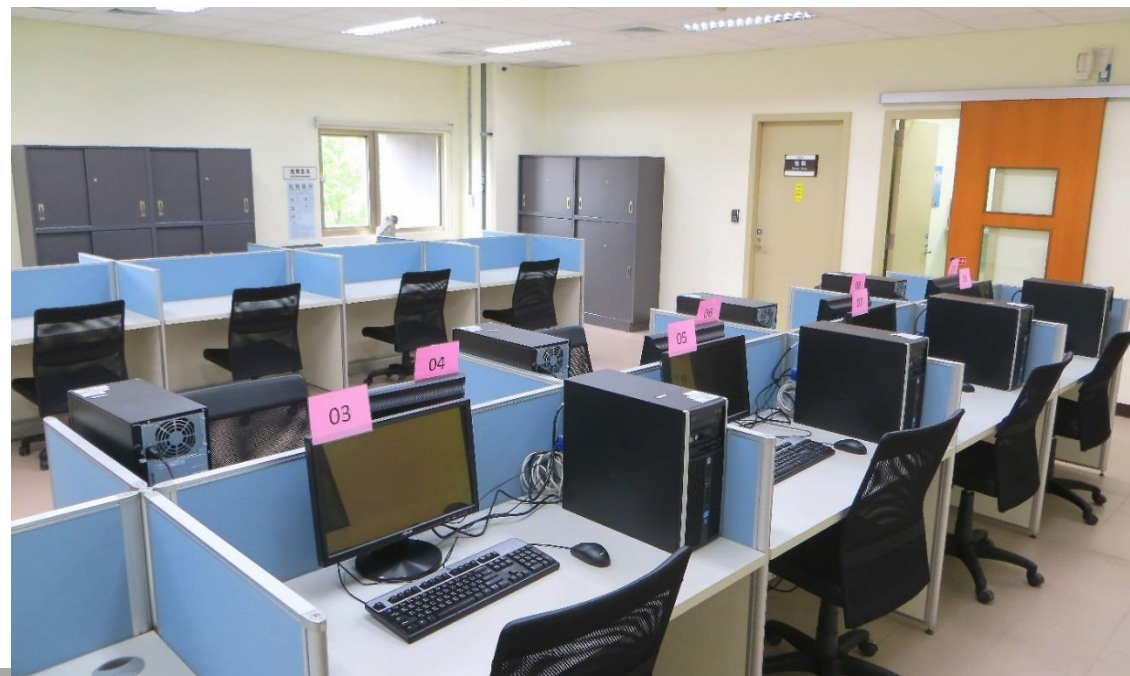




各醫院每月申報檔案

健康資料加值應用協作中心

Health Information Application (CCHIA), Ministry of Health and Welfare



識別碼加密
實體隔離
刪除敏感資料
不能上網攜帶手機

衛生福利部
「健康資料」加值中心

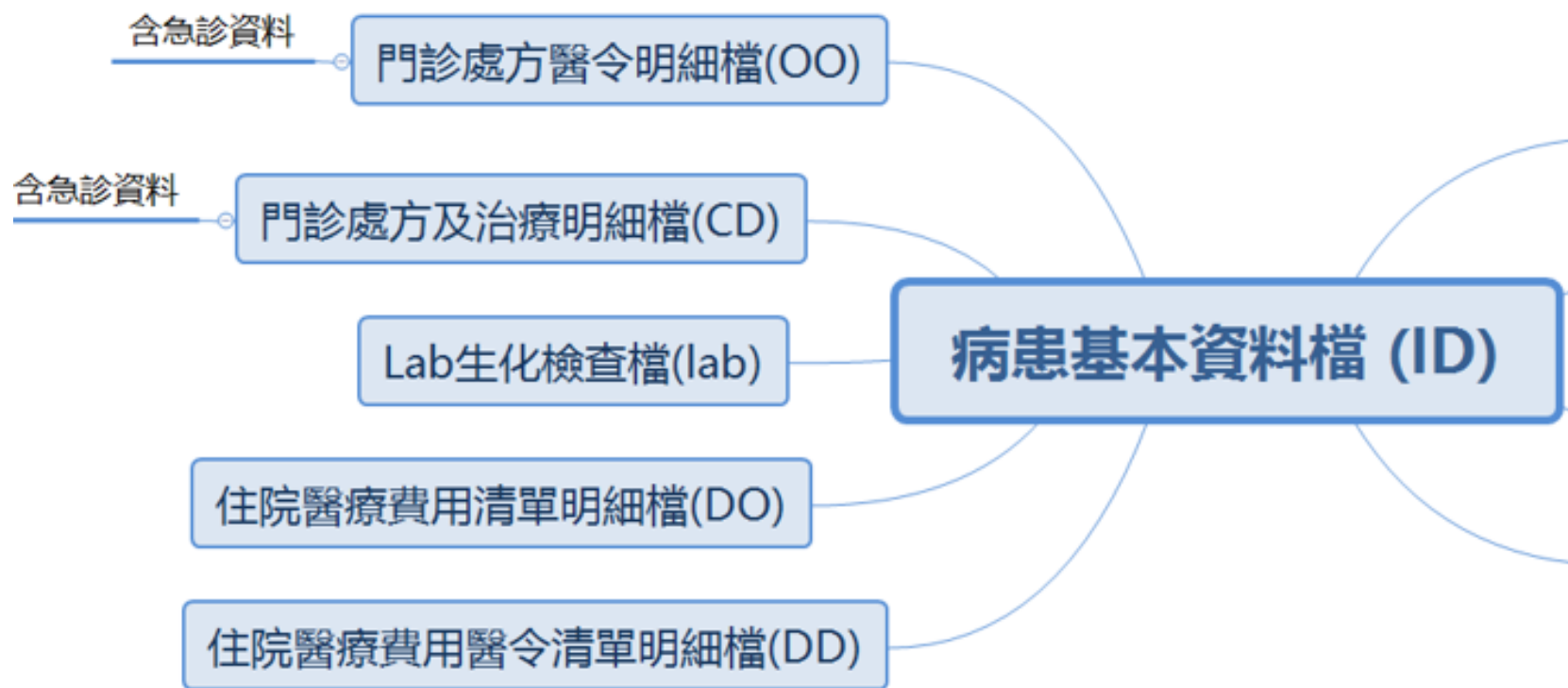




臺北榮民總醫院 大數據中心

Big Data Center

現在已完成可申請



譯碼簿說明

承保資料檔 (適用 99 年以後資料)

ID

 英文檔案名： ID 長度 107
 中文檔案名： 承保資料檔

序號	英文欄位	中文欄位	資料型態	長度	起始位置	迄末位置	資料描述
1	ID	個人身份證號	C	32	1	32	已轉碼，亦即保險對象
2	INS_ID	被保險人身份證號	C	32	33	64	已轉碼
3	INS_ID_TYPE	身份別	C	1	65	65	請參考帶碼說明B-77之「身份別說明」
4	INS_AMT	投保金額	N	6	66	71	一至三類被保險人月投保金額。未滿6位數前面補0，若為眷屬資料，投保點數為0。
5	ID_BIRTHDAY	出生年月	C	8	72	79	YYYYMM,左靠不足補空白
6	ID_SEX	性別	C	1	80	80	M：男性 F：女性 U：不詳
7	INS_RELATION	稱謂代號	C	1	81	81	1：配偶 2：父母 3：子女 4：祖父母 5：孫子女 6：外祖父母 7：外孫子女 8：曾祖父母 9：外曾祖父母 0：其他親屬 U：不詳
8	UNIT_INS_TYPE	單位屬性	C	3	82	84	請參考代碼說明B-74之「單位屬性代號」；若為眷屬資料，則為空白
9	REG_ZIP_CODE	郵遞區號	C	4	85	88	請參考代碼說明B-13之「地區代碼、名稱、分局及郵遞區號」
10	TX_CODE	異動別	C	2	89	90	01: 加保異動 02: 轉入異動



大數據中心

病患基本資料檔譯碼簿 ID

代號	序號	欄位名稱	中文名稱	資料型態	長度	備註
V	1	PHISTNUM	病歷號	C	36	已加密
V	2	PIDNO	身分證字號	C	36	已加密
V	3	PSEX	性別	C	1	1: 男; 2: 女
V	4	PBIRTHDT	出生年月	C	6	僅提供西元年月
V	5	PPATZIP	郵遞區號	C	3	
V	6	PFSTVDT	初診日期	C	8	西元年月日
V	7	PPBLOOD	血型	C	3	

釋例-大數據病患基本資料檔-I



中文名稱	病歷號	身分證字號	性別	出生年月	郵遞區號	初診日期	血型
欄位名稱	PHISTNUM	PIDNO	PSEX	PBIRTHDT	PPATZIP	PFSTVDT	PPBLOOD

中文名稱是為了讓我們好記得
但寫程式時要用欄位名稱(英文)來寫

釋例-大數據病患基本資料檔-II

PHISTNUM	PSEX	PBIRTHDT	PPATZIP	PFSTVDT	PPBLOOD	PLSTVDT	PDEADFLG
835079	2	194111	100	19660811		20170719	0
8363728	2	193203	111	19800331	O +	20190625	0
8438157	2	194209	112	19800419	A +	20190523	0
8478744	2	193505	232	19800507	O +	20190520	0
8478744	1	193505	232	19800507	O +	20190520	0
9582712	1	192804	111	19800527	O +	20190513	0

多筆資料累加後資料整理如上

Microsoft®
EXCEL

STATA

Microsoft®
SQL Server®

IBM
SPSS®

SAS®

R

1. 資料轉換

READ

Extraction

2. 變數處理 Processing

SELECT, FILTER
APPEND, UNION
JOIN, SORT

3. 結果分析 Analyzing

AGGREGATE
OUTPUT
STATISTICS

市面上常見軟體

軟體	Excel	R	SQL	SAS	STATA	SPSS
取得可近性	★	★	★	需花錢 取得	需花錢 取得	需花錢 取得
是否可用 在資料庫 整合	★	★	★	★	★	★
統計分析	★	★		★	★	★

你知道生活廣場的營業額多少嗎



不要小看了我們的實力

北榮大數據中心介紹



臺北榮民總醫院 大數據中心

Big Data Center

<https://wd.vghtpe.gov.tw/bdc/> 大數據中心獨立網址



臺北榮民總醫院 大數據中心

Big Data Center

[回醫療體系歡迎頁](#) | [科部首頁](#) | [網站導覽](#) | [English](#)



最新消息

單位簡介

申請流程

相關法規

部科宣導

相關連結

文件下載

諮詢管道

說明



108/4/18 (週四) 中午12:30 於醫科大樓一樓會議室
召開全院性大數據運用事宜說明討論會議
主持人：大數據中心 主任 陳曾基醫師 | 與會人數：120人

公告

[更多公告](#)

醫療新知

[更多醫療新知](#)



關於本會

案件審查

受試者

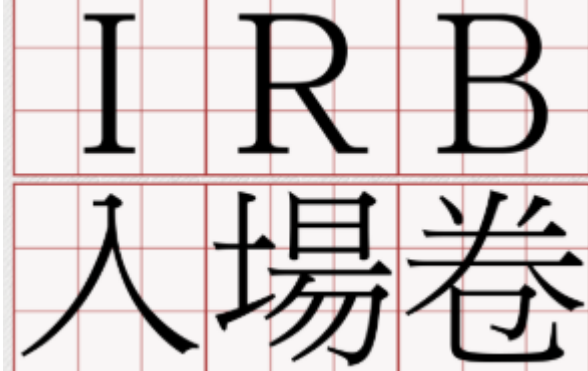
相關法規

文件下載

PTMS專區

相關連結

聯絡我們



PTMS專區

[首頁](#) > [PTMS專區](#) > PTMS

PTMS

PTMS

COI操作手冊

[置頂] 108年度 **必修課程行程表** (請點此進入查詢詳細報名網址與課程資訊)

人體試驗委員會 最新消息 5170

2019必修課程時程表.pdf

A- A A+

108年度全年必修課程行程如附檔，請參考。

課程報名請至全院開課系統，網址：<https://edu.vghtpe.gov.tw/> (課程查詢:請輸入「課程代碼」或「人體」查詢)

1. PTMS帳號申請
2. 上必修課程

A- A A+

申請計畫重要必讀：

** 請熟讀 [申請IRB及資料注意事項](#) **，在計畫中一定要寫清楚所要申請的資料，IRB核准什麼資訊室依此給予資料（很重要喔）。

[IRB 臨床試驗計畫中文計畫書摘要\(使用健保資料庫適用\)](#) - 撰寫範例 請參照。

** 請閱讀 [譯碼簿](#) 及 [申檔說明](#) **

申請流程如下： ****目前簽核需以紙本簽核**** *****說明1- [點下去皆有範例](#)；說明2-[點下去可下載表格](#)，請務必先閱讀，謝謝*****

1. 請在本院3270系統中申請 ● CHANGE 『[資訊應用系統變更申請](#)』（請按），操作方式請參考（請按）。資料填好後，務必請申請人及單位主管核章。

2. 填寫 [大數據中心申請書](#)（內頁需簽名）+ [聲明書](#)（請按）+ IRB通過之 [同意臨床試驗證書](#)（含計畫書）+ 臨床研究受試者保護中心核發之 [同意執行證明書](#) 等文件。

. 請備妥前述第1、2項資料，以公文傳送或親送資訊室編號後，再依序會辦IRB、大數據中心、病歷室、醫企部等各單位。經主任秘書核可後，資訊室才能給予目前『已開放申請範圍』之數據資料。

* 申請流程可洽 大數據中心 楊醫檢師 分機：4065

入場券

如申請欄位超出譯碼簿目前給予的內容.請先和資訊是聯繫是否有資料.在寫IRB

大數據中心如何幫助您拿到IRB入場卷

- 目前中心有提供標準資料檔萬人抽樣檔,歡迎PI申請使用
- 未來將設計出-IRB 臨床試驗計畫中文計畫書摘要範本(大數據中心資料庫適用)

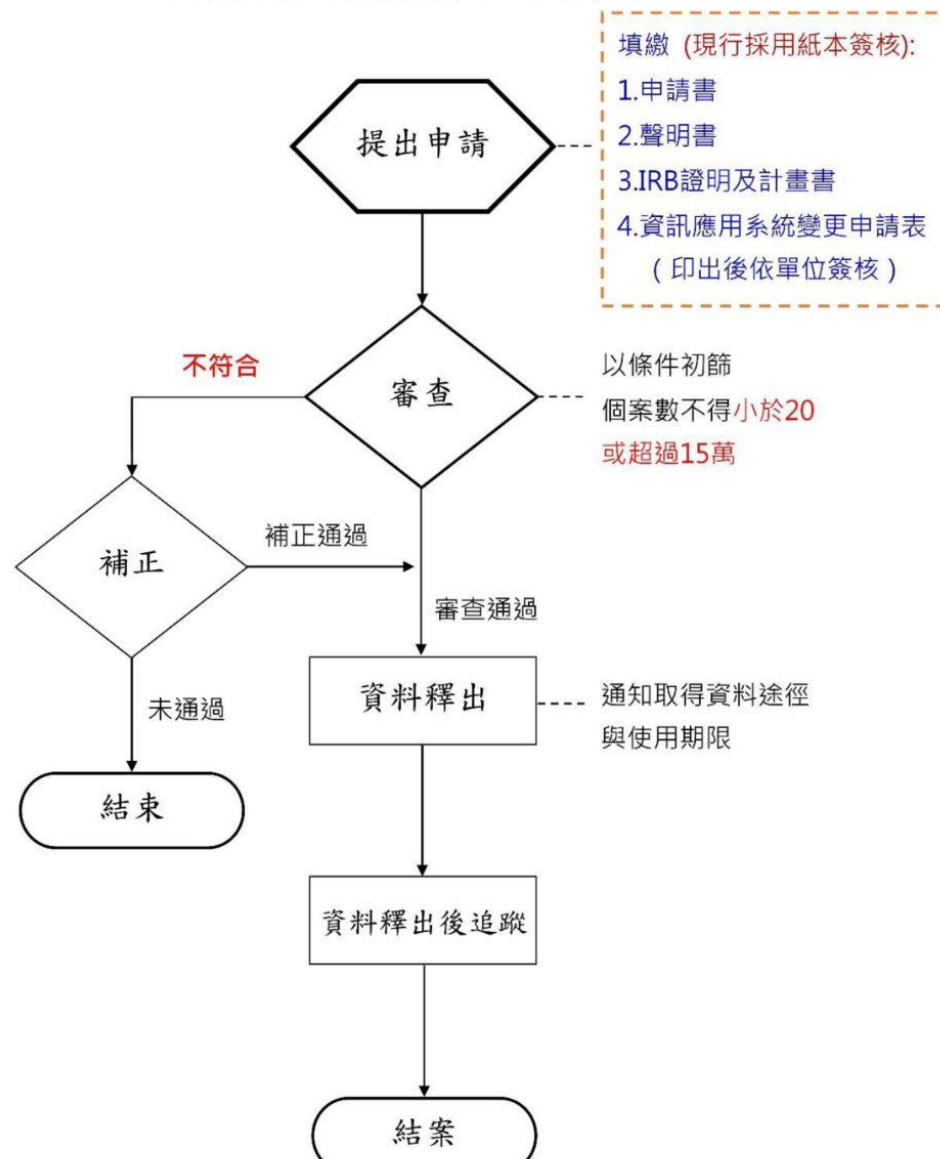
文件下載

A- A A+

1. [申請IRB及資料注意事項](#)
2. [檔案1-北榮大數據中心 流程說明及資料申請單 20191105版](#)
3. [檔案2-北榮大數據中心譯碼簿20190901版](#)
4. [串檔說明](#)
5. 研究技巧訓練課程：[108.10.29~108.11.6 R課程教學講義-更新檔](#) (108.11.26 修正)

- IRB說明
- 申請流程
- 譯碼簿
- 串檔說明
- 課程資料

臺北榮民總醫院大數據中心資料申請流程



特種資料宣導-保護個資很重要

病歷屬於特種個資，個資應用有法律規範，
尤其是有關去識別化的實務，醫療機構更應注意。

註1：本中心提供之檔案為主題歸人檔(CSV 格式)，申請者需提供定義(可以附件提供)，凡申請資料期間符合定義之病患，均屬本申請案的個案，依照篩選個案條件歸人後，串聯該病患於本院之其他資料。本中心得請申請者修改篩選條件。

註2：上述資料檔有關個人資訊欄位已經去識別或模糊化處理。

註3：請依照上述範例填寫需求，若上述欄位不夠填寫，可以 Excel 檔填寫後同時檢附。

註4：若有特別的篩選條件，也請補充說明在上方表格。

註5：疾病診斷請填寫完整，例如 250 糖尿病、250.0 未提及併發症之糖尿病、250.02 第二型(非胰島素依賴型，成人型)或未明示型 糖尿病，未提及併發症，無法控制...等。

大數據中心申請單上有明確說明
請大家保護手中資料庫

特種資料宣導-保護個資很重要

特種資料宣導-保護個資很重要

大數據中心目前釋出資料的時間

- 門住院資料檔申請年限資料: 2011/1/1 ~2018/6/30
- 檢驗資料(目前開放2017/1/1~2018/6/30)-無影像檔
- 數據資料會定期更新日期半年或一年視資料複雜情況而定.
- 因清資料擴日耗時無法天天周周月月更新 還請見諒

其他注意事項

- <https://wd.vghtpe.gov.tw/bdc/>
大數據中心獨立網站
- 本中心負責行政申請流程、數據教學、推廣、諮詢、清資料庫
研究者負責保管好資料、設計好研究、寫好論文產出



心有多廣，路就有多寬
研究路上總是孤單寂寞的
但大數據中心陪著你

哈囉哈囉本次是- SQL初階新手研究小班教學課

- SQL是甚麼-我會對它有興趣嗎
- 小班課時間以及如何報名
- 小班課我應該要盡什麼責任

SQL長這樣-----推程式碼

```
檔案(F) 編輯(E) 檢視(V) 專案(P) 偵錯(D) 工具(T) 視窗(W) 說明(H)
新增查詢(N) 泛型偵錯工具
master 執行(X) 偵錯(D)
物件總管 20180302.參考答案 (...UL\Angela (52))
DESKTOP-VSOPPUL\SQL (SQL Serve
  資料庫
  安全性
  伺服器物件
  複寫
  PolyBase
  Always On 高可用性
  管理
  Integration Services 目錄
  SQL Server Agent (已停用Agent X
  XEvent 分析工具
4. 30-60歲男生對於3C 接受度高，
   20-40 歲女性 3 C 接受度高，請問這個族群市場成長率 ??
*/
-- 2008
select * from h_nhi_opdte97 where CURE_ITEM_NO1='E4'; -- 2231
-- 2009
select * from h_nhi_opdte98 where CURE_ITEM_NO1='E4'; -- 2530

select (2530-2231.0)/2231 * 100.0 --- 成長 13.4%
----- 1: 男性
-- 2008
select * from h_nhi_opdte97 where CURE_ITEM_NO1='E4' and ID_S='1'; -- 1025
-- 2009
select * from h_nhi_opdte98 where CURE_ITEM_NO1='E4' and ID_S='1'; -- 1143

select (1143-1025.0)/1025 * 100.0 --- 男生成長 11.6%
----- 2: 女性
-- 2008
select * from h_nhi_opdte97 where CURE_ITEM_NO1='E4' and ID_S='2'; -- 1196
-- 2009
select * from h_nhi_opdte98 where CURE_ITEM_NO1='E4' and ID_S='2'; -- 1376

select (1376-1196.0)/1196 * 100.0 --- 女生成長 15.1%
----- 1: 男性 30-60 歲
-- 2008 30-60 歲 --> 生日 1948/1 ~ 1978/12
select * from h_nhi_opdte97 where CURE_ITEM_NO1='E4' and ID_S='1'
and
birth_ym between '194801' and '197812'; -- 412
-- 2009 30-60 歲 --> 生日 1949/1 ~ 1979/12
select * from h_nhi_opdte98 where CURE_ITEM_NO1='E4' and ID_S='1'
and
birth_ym between '194901' and '197912'; -- 482
select (482-412.0)/412 * 100.0; --- 男性 30-60 歲成長 17.0%
----- 2: 女性 20-40 歲
```

本課程僅教程式語法
無統計方法、論文寫作

SQL課程上課表

時間	課程	老師/助教
2/18 星期二 晚上18:00-20:30 致德樓第五會議室	Setup SQL at your NB -資料處理SQL系統安裝 Data management: (1). Data importing 大型健康資料觀察性研究資料處理－(1)資料匯入	陳育群/ 邱雯琪
2/25 星期二 晚上18:00-20:30 致德樓第五會議室	Data management: (2). Data transformation 大型健康資料觀察性研究資料處理－(2)變數處理	陳育群/ 邱雯琪
3/3 星期二 晚上18:00-20:30 致德樓第五會議室	Data management: (3). Record linkage 大型健康資料觀察性研究資料處理－(3)資料串連	陳育群/ 邱雯琪
3/10 星期二 晚上18:00-20:30 致德樓第五會議室	Data management: (4). Sub-query 大型健康資料觀察性研究資料處理－(4)子查詢	陳育群/ 邱雯琪
3/17 星期二 晚上18:00-20:30 致德樓第五會議室	Data management: (5). Data matching 大型健康資料觀察性研究資料處理－(5)資料配對	陳育群/ 邱雯琪
3/24 星期二 晚上18:00-20:30 致德樓第五會議室	From Data to Result(6): 大型健康資料觀察性研究統計分析- 論文解析與實作 (6)	陳育群/ 邱雯琪

遊戲規則

- 108/11/29上過大數據中心開立SQL介紹課程(必)
- 年輕初學研究人員優先(選)

如何報名

★ 今天演講後請上 下列網址報名

<https://forms.gle/YzEqshqRhwRd2Wgf6>



如果沒報到.....

如果報名成功1/20前將有媚兒通知您.
沒收到很遺憾代表本次無報名成功

- SQL本課程秉持以學生和用戶為中心原則.以後是否會持續開立小班教學或變更授課方式.將會依照本次學生回饋而制定.
- 大數據中心陸陸續續會開立課程.請密切注意公告

報上的同仁注意事項

- 請攜帶筆電-每堂都需要筆電
- 每堂課均需參與.不然會跟不上課程!除非你可以找到同學願意教你.讓你跟上進度唷.**(第一堂沒來那後面就真的抱歉了)**
- 每堂都會有課後作業要繳交



這不是必修課.但是你投資你自己的課程!

謝謝聆聽

如有相關開課或課程建議 請洽大數據中心
邱小姐 chiuangelala@gmail.com;分機:1244
楊醫檢師 分機:4065

本次Q&A

- **建議:希望開立專門解釋驛碼簿和資料庫說明課程**

回答:OK.後續將會開立

- **問題:對基礎研究有興趣**

回答:最近將有開立基礎研究相關課程.已在開課系統公告

- **問題:PSPP、SAS EG、google analysis**

回答:PSPP類似SPSS簡化免費版.SAS EG為SAS系列套組.

- **希望開立機械設備狀態分析課程**

回答:這大數據中心無法幫助您,可上開課系統尋找看看

- **問題:想知道大數據有提供那些資料**

回答:請上大數據中心網站查詢