

2022 國家級人體生物資料庫整合平台第二屆年會

邁向繁榮昌盛的整合平台

NBCT-towards prosperity and flourishing

建置整合平台醫療資訊共同模式之團隊作業模式

李修安 博士

國家衛生研究院 - 癌症研究所



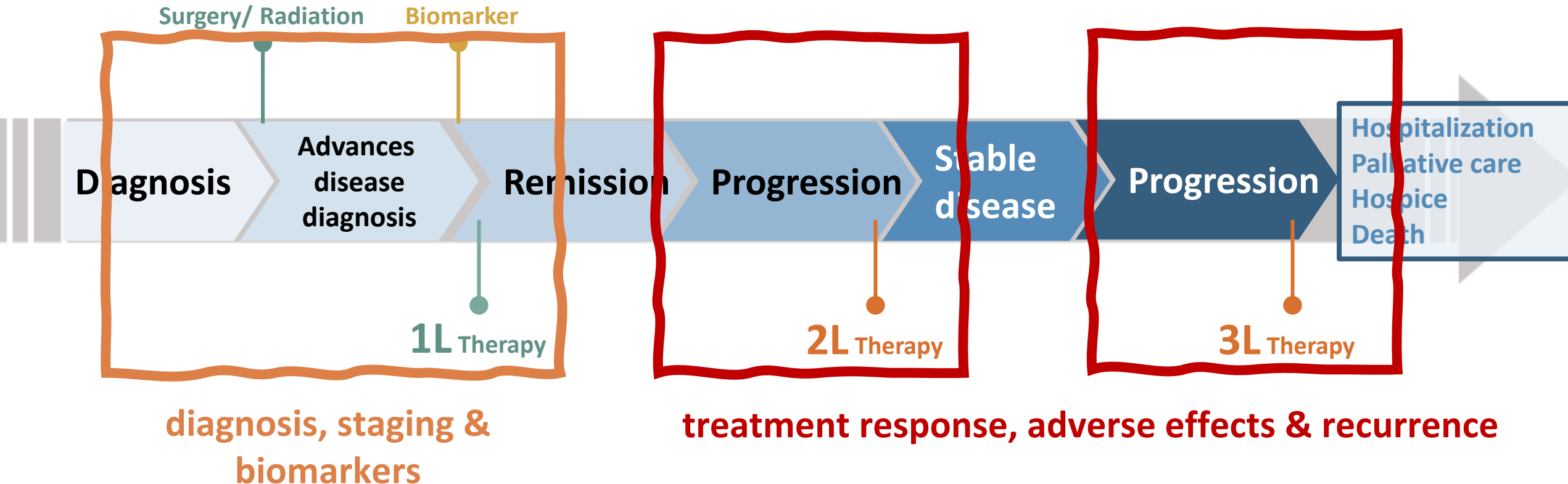
建構醫療大數據共同欄位模式 (NBCT CDM)

1

- 規劃資料欄位, 能夠分析病人疾病診斷、治療及預後追蹤的資訊
(to **assemble a model of common data elements for EHRs**)
- 促進資料的互通性
(to **facilitate data interoperability** for patient care and research)
- 維持足夠的資料品質
(to **maintain sufficient data quality**)

PATIENT'S JOURNEY

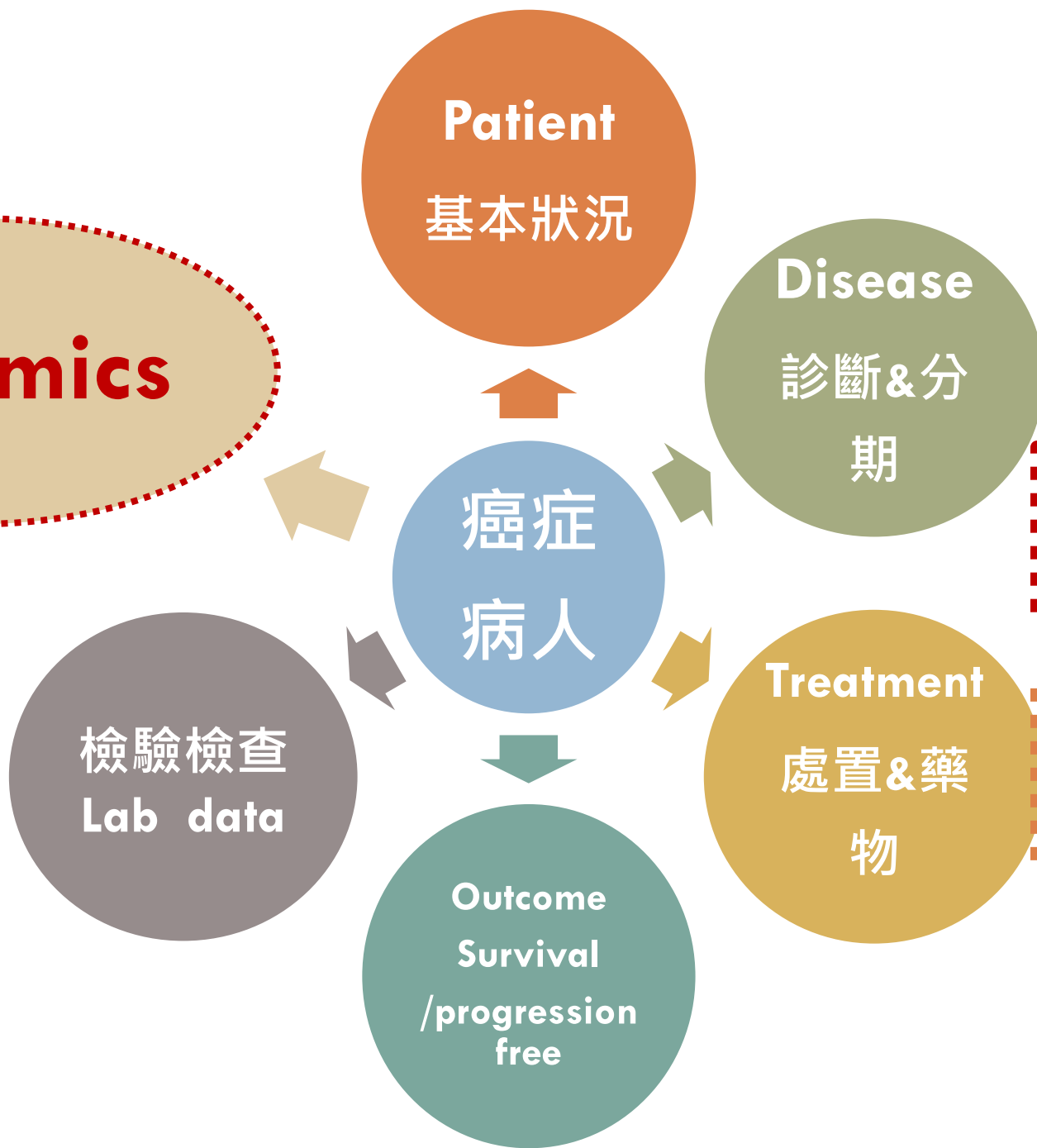
Need to have sufficient information
to find fit-for-use data



Data users look for certain part of data in patient journey since they are trying to develop medical products for various purposes. We start by identifying these factors first in the search filter.



透過直接從電子健康記錄(EHR)中匯總數據，無需臨床醫生以特定格式取得數據



臨床醫療資料 (EHR):

第一階段

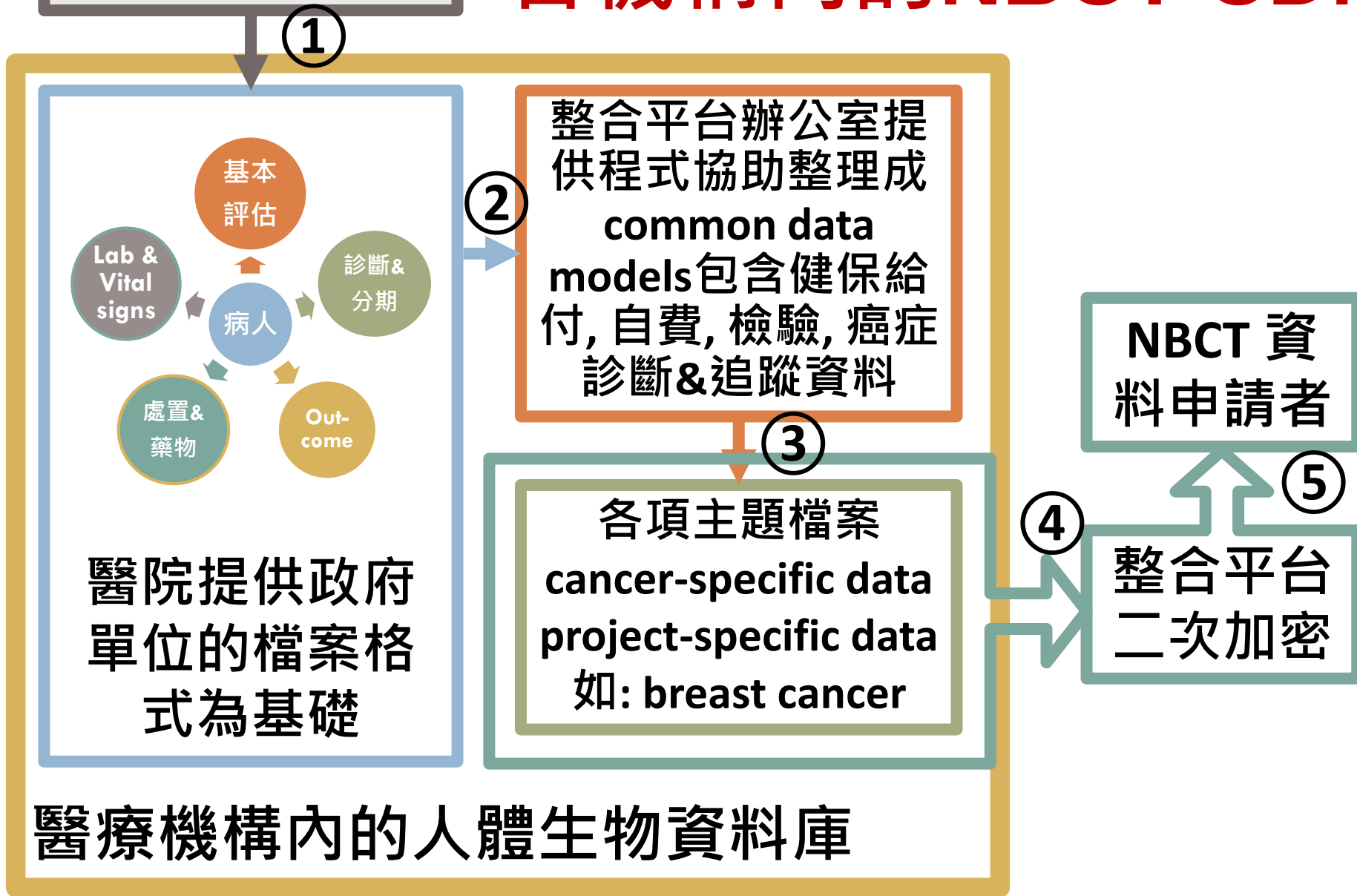
- 結構式資料

後續

- 非結構式資料
 - 文字報告
 - 影像

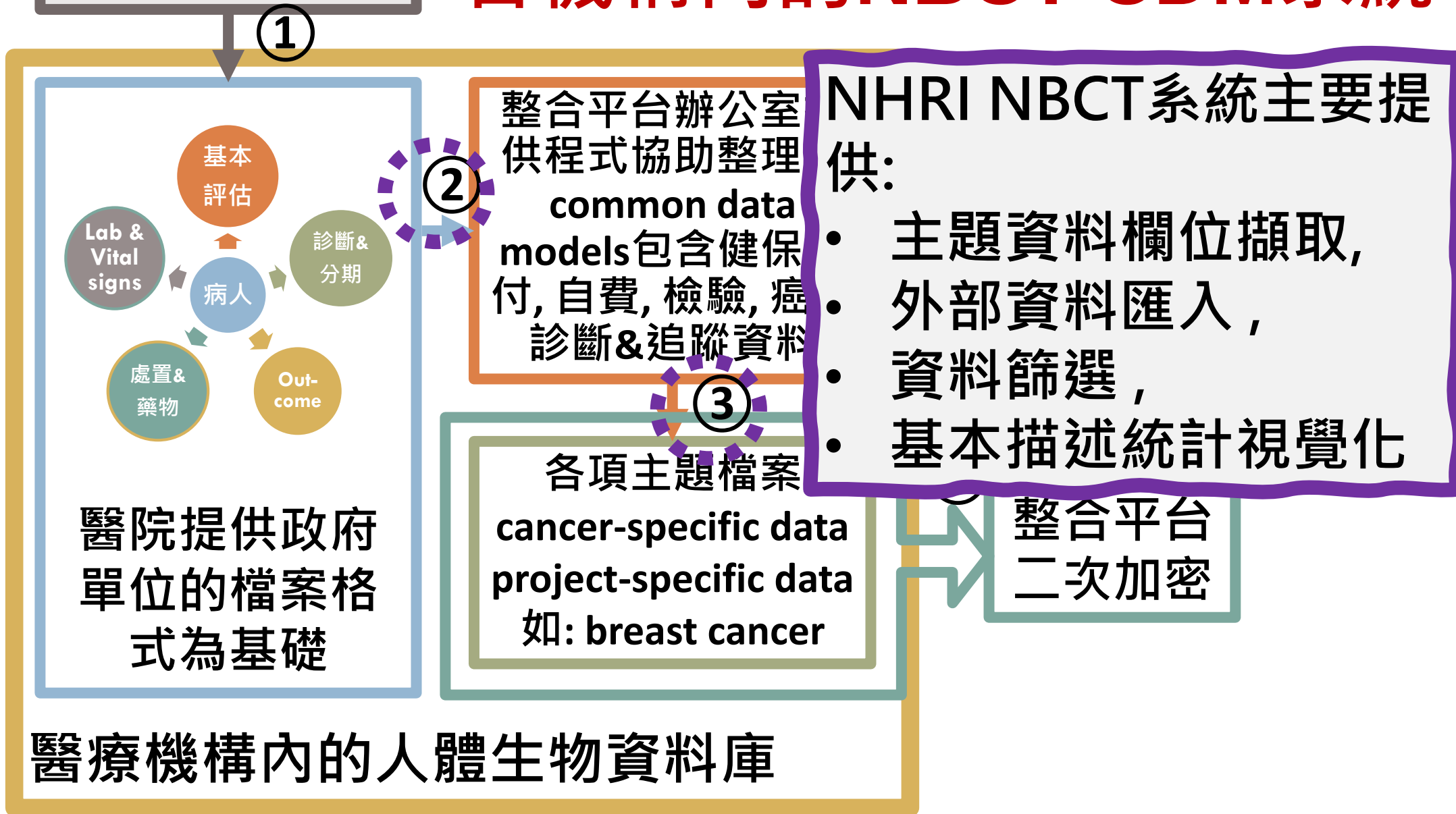
電子健康資料

各機構內的NBCT CDM系統



電子健康資料

各機構內的NBCT CDM系統

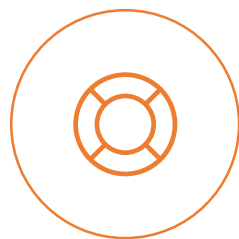


醫療資訊共同模式之價值與目標



永續經營

建立整合平台共同模式，輔助臨床健康資料整合，並建立永續發展架構。



品質管理

資料內容的品質管理、整合需求，建構優良的資料儲存與檢索架構，提供使用者搜尋。



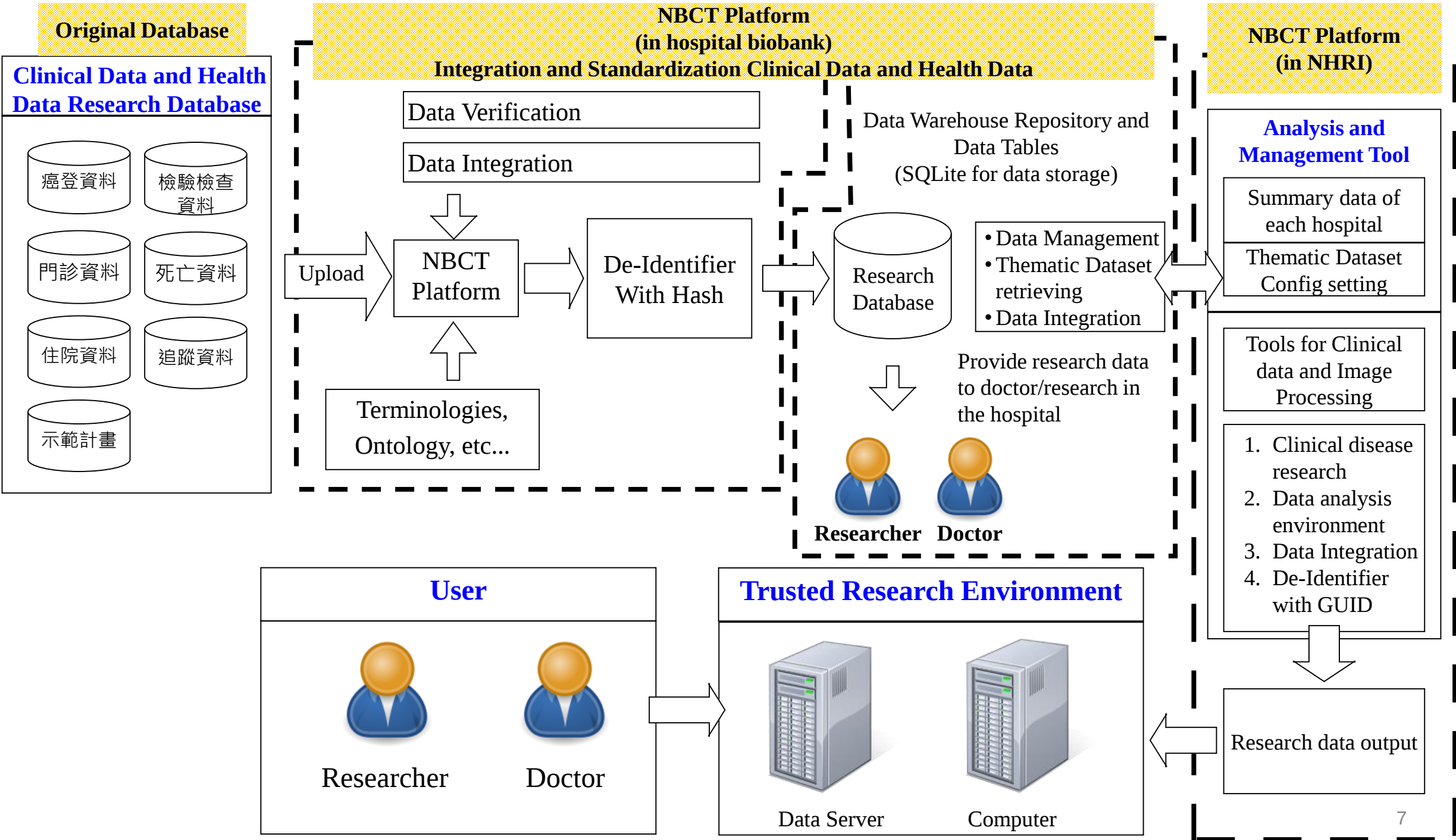
互操作性

資料具有共通模式、內涵，以利需求單位整合、整理與使用，並具備國際互通性。

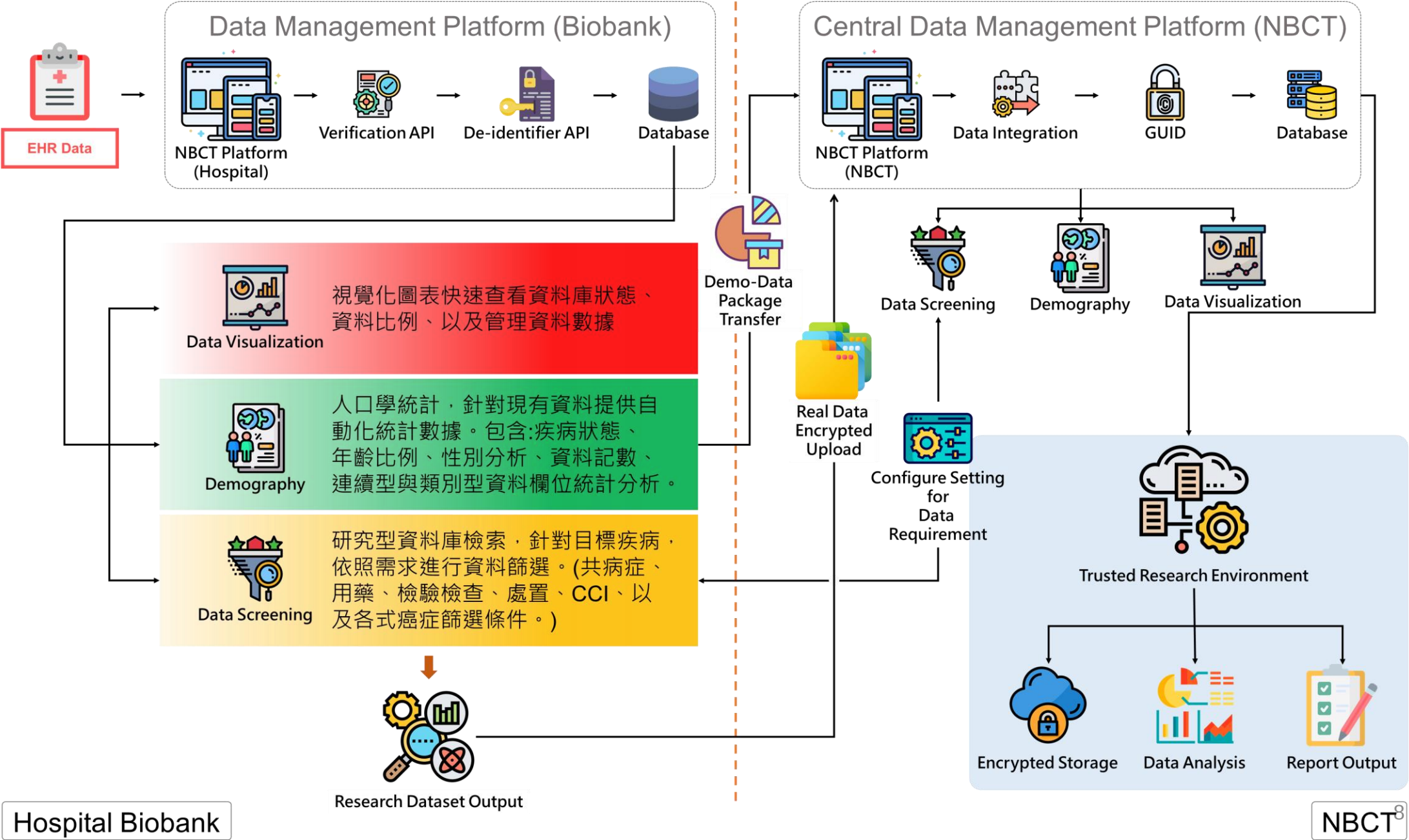


資料加值

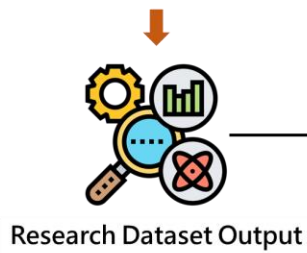
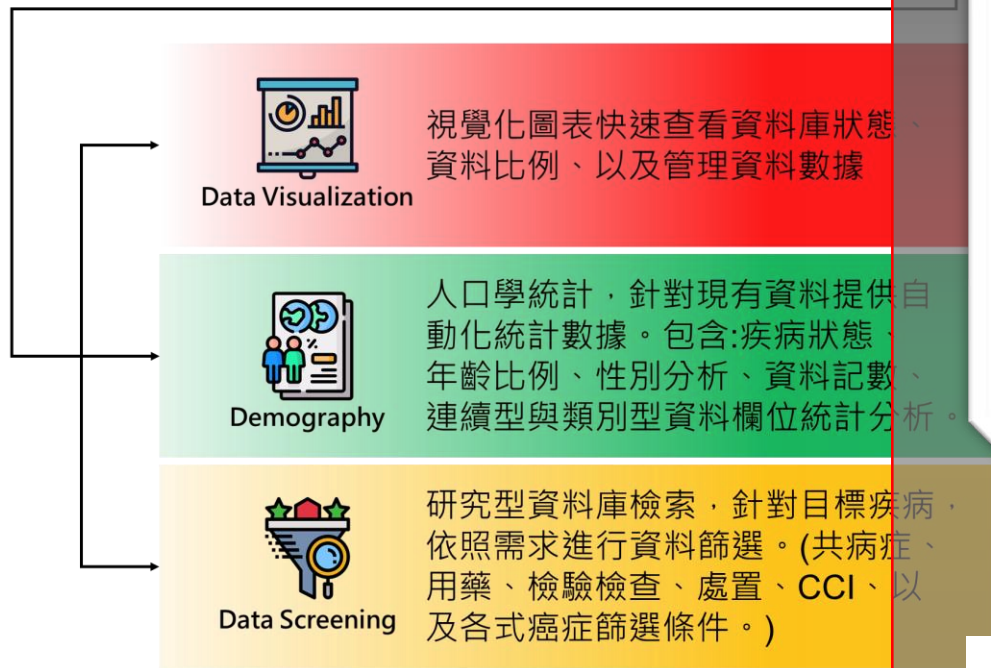
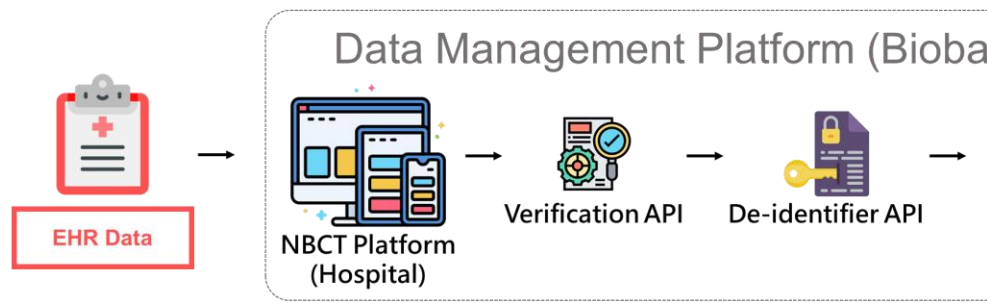
輔助醫療院所資料有意義的加值使用，提供檢索、篩選機制，建構主題式資料庫。



整合平台醫療資訊共同模式

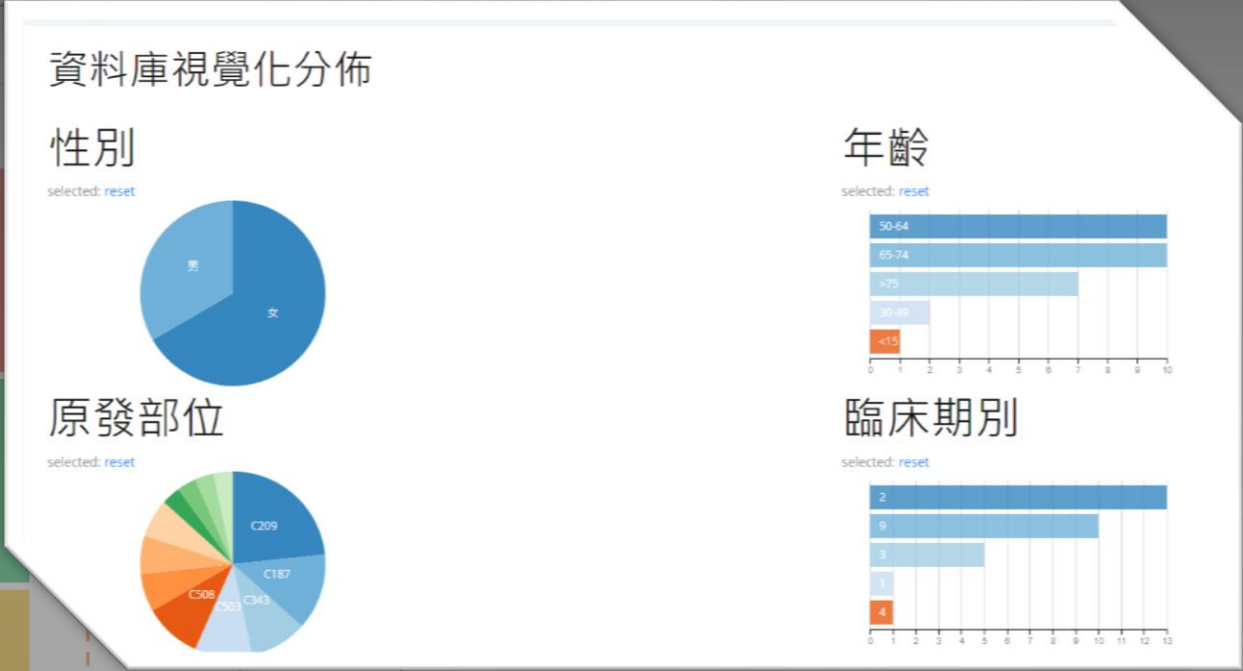


整合平台醫療資訊共同模式

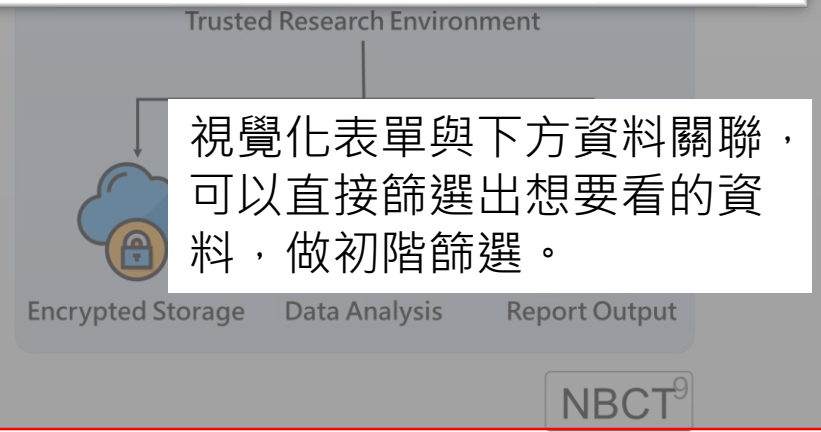


Hospital Biobank

資料視覺化

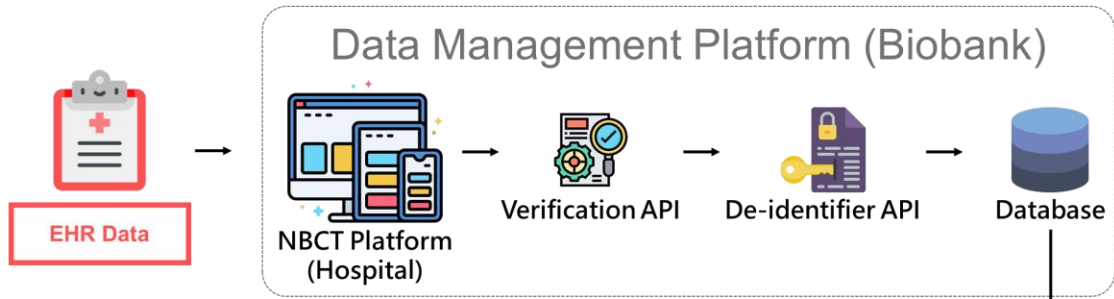


- 4 項目:
1. 性別
 2. 年齡
 3. 原發部位
 4. 臨床期別



視覺化表單與下方資料關聯，可以直接篩選出想要看的資料，做初階篩選。

整合平台醫療資訊共同模式



Data Visualization

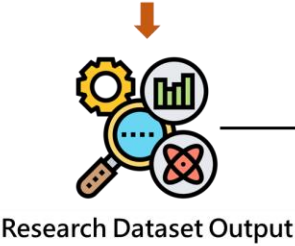
視覺化圖表快速查看資料庫狀態、資料比例、以及管理資料數據

Demography

人口學統計，針對現有資料提供自動化統計數據。包含:疾病狀態、年齡比例、性別分析、資料記數、連續型與類別型資料欄位統計分析。

Data Screening

研究型資料庫檢索，針對目標疾病，依照需求進行資料篩選。(共病症、用藥、檢驗檢查、處置、CCI、以及各式癌症篩選條件。)



Hospital Biobank

統計

Summary Report?

顯示全部資料 download

ssf1 查看

Show 10 entries

CASE_total_counts	CRLF_total_counts	CRSF_total_counts	DEATH_tot
0	1	1	0
0	1	1	0
0	1	1	0
1	1	1	0
1	1	1	0
1	1	1	0
1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	0

Showing 1 to 10 of 30 entries

cstage 細項計數

selected: reset

download

Show 10 entries

Search:

Type	Value_count
02A	1 (3.33%)
000	2 (6.67%)
001	6 (20.0%)

表單資訊匯總：
不同類型的表單資料量統計。
個別數據集、連續數據、類別數據和交叉表的描述表。

NBCT 10

統計

EHR Data

Data Management Platform (Biobank)



Central Data Management Platform (NBCT)

CRLF

送出

Female	Male	Total	Table_line	Newest_date	Oldest_date
20	10	30	30	2022-09-29 16:29:49.1451804	2022-09-29 16:29:49.1451804

連續型資料描述

Description of Continuous Data?

download

Show 10 entries

Search:

Index	Mean	Std	Min	25%	50%	75%	Max	Count
age	64.0333333333	16.7712466037	0	55.5	70	74	83	30

Showing 1 to 1 of 1 entries

Previous 1 Next

類別型資料描述

Description of Category Data?

download

Show 10 entries

Search:

Index	Count	Missing	Unique	Top	Freq
pni	30	0	5	1	9
size	30	0	9	999	21
dmconf	30	0	30	20170211	1
confirm	30	0	3	1	27
grade_p	30	0	5	2	13
grade_c	30	0	5	2	13

整合平台醫療資訊共同模式

統計

查詢紀錄

Cross Data

download

Show 10 entries

Search:

Site	Male	Female	Total
C209	3(42.9%)	4(57.1%)	7
C187	3(75.0%)	1(25.0%)	4
C508	0(0.0%)	3(100.0%)	3
C503	0(0.0%)	3(100.0%)	3
C343	2(66.7%)	1(33.3%)	3
C349	0(0.0%)	2(100.0%)	2
C341	2(100.0%)	0(0.0%)	2
C186	0(0.0%)	2(100.0%)	2
C502	0(0.0%)	1(100.0%)	1
C501	0(0.0%)	1(100.0%)	1

Showing 1 to 10 of 12 entries

Previous 1 2 Next

Hospital Biobank

Research Dataset Output

Encrypted Storage

Data Analysis

Report Output

NBC¹²

整合平台醫療資訊共同模式



Data Management Platform (Biobank)



Data Visualization

視覺化圖表快速查看資料庫狀態、資料比例、以及管理資料數據



Demography

人口學統計，針對現有資料提供自動化統計數據。包含:疾病狀態、年齡比例、性別分析、資料記數、連續型與類別型資料欄位統計分析。



Data Screening

研究型資料庫檢索，針對目標疾病，依照需求進行資料篩選。(共病症、用藥、檢驗檢查、處置、CCI、以及各式癌症篩選條件。)



Research Dataset Output

Hospital Biobank

資料篩選

癌症個案條件篩選

非癌症疾病

選擇檔案 未選擇任何檔案 匯入 清空欄位 送出資料

主題資料庫(參考)

口腔癌 大腸癌 食道癌 卵巢癌 肝癌 乳癌 肺癌 胃癌 胰臟癌 頭頸癌

癌症名稱

請輸入疾病名稱 請輸入疾病名稱

* 請輸入ICD-O-3代碼(以逗號分隔) 請輸入ICD-O-3代碼(以逗號分隔)

ICD-O-3建議使用的代碼

Trusted Research Environment

癌症條件

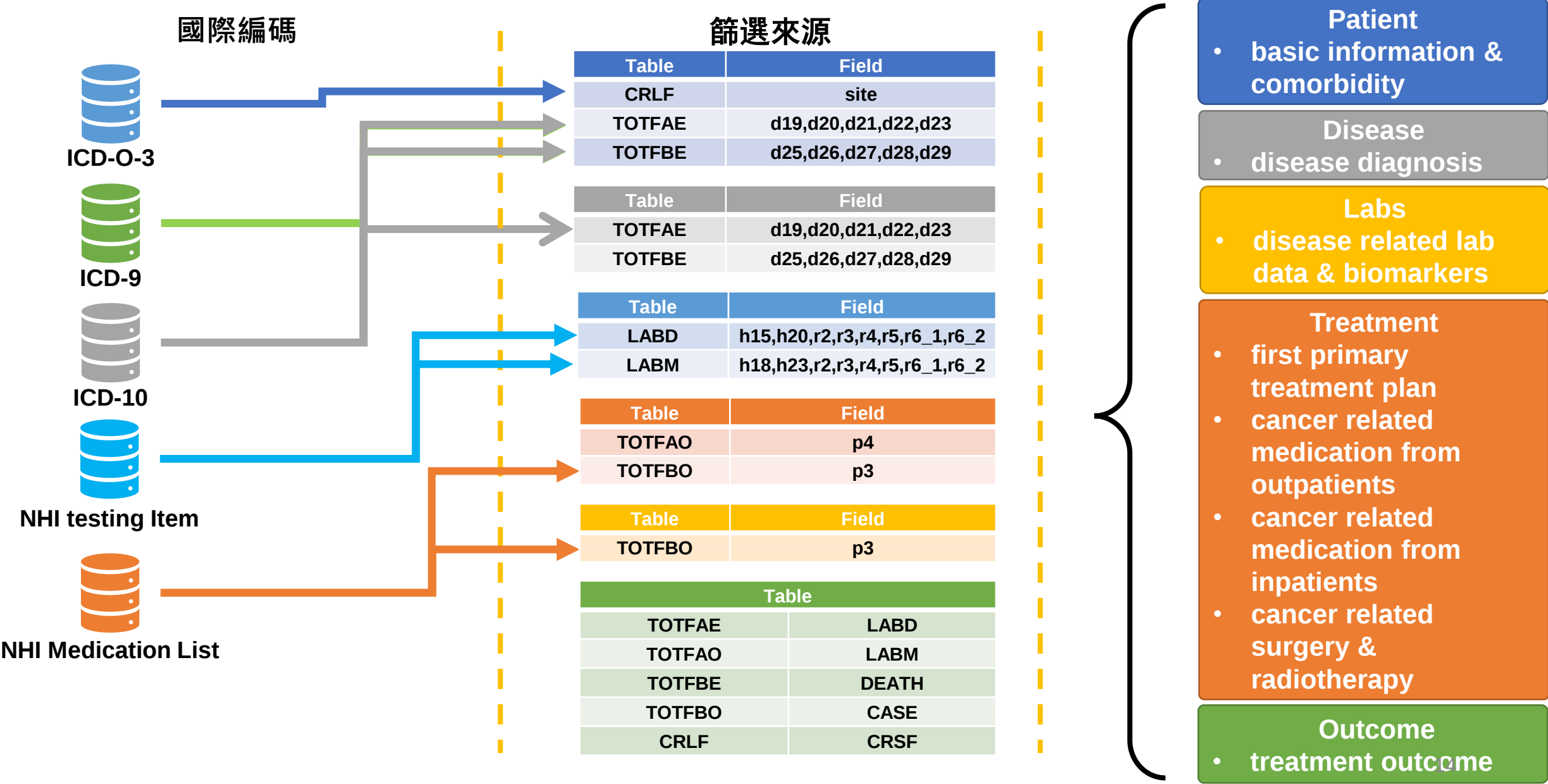
組織型態(hist) 請輸入ICD-O-3.2編碼(以逗號分隔)

腫瘤分期-臨床(cstage) 請輸入臨床期別

腫瘤分期-病理(pstage) 請輸入病理期別

首次治療前生活功能狀態評估(pst) 請填入三碼(以逗號分隔)

資料篩選與通用資料使用



資料篩選



選擇檔案 未選擇任何檔案

匯入

清空欄位

送出資料

主題資料檔(參考)

口腔癌

大腸癌

食道癌

卵巢癌

肝癌

乳癌

肺癌

胃癌

胰臟癌

頭頸癌

癌症名稱

請輸入疾病名稱 ?

乳

* 請輸入ICD-O-3代碼(以逗號分隔) ?

請輸入ICD-O-3代碼(以逗號分隔)

ICD-O-3建議使用的代碼

C50--乳房惡性腫瘤

C500--乳頭和網眼狀空間

C501--乳房中央部分

C502--乳房上內象限

C503--乳房下內象限

C504--乳房上外象限

C505--乳房下外象限

C506--乳房腋尾部

C508--乳房交搭跨越的惡性腫瘤

C509--未明確的乳房

資料篩選



主題資料檔(參考)

- 失智症
- 高血壓
- 高血糖
- 高血脂
- 腎病
- 肝病
- 腦血管病
- 心臟病
- 肺炎

疾病名稱 [?]

請選擇主表

門診和住院

門診次數

≥次數

住院次數

≥次數

請輸入疾病名稱 [?]

糖

請輸入ICD-9(以逗號分隔) [?]

請輸入ICD9(以逗號分隔)

請輸入ICD-10(以逗號分隔) [?]

請輸入ICD10(以逗號分隔)

ICD9建議使用的代碼

- 250--糖尿病
- 25000--第二型或未明示型糖尿病，未提及併發症（非胰島素依
- 25001--第一型糖尿病，未提及併發症（胰島素依賴型，幼年型
- 25002--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
- 25003--第一型(胰島素依賴型，幼年型)糖尿病，未提及併發症，無法控制

ICD10建議使用的代碼

- D550--葡萄糖 - 6 - 磷酸脫氫酶[G6PD]缺乏所致的貧血
- D552--糖分解酵素疾患所致的貧血
- E08--起因於潛在病的糖尿病
- E080--起因於潛在病的糖尿病，伴有高滲透壓
- E0800--起因於潛在病的糖尿病，伴有高滲透壓，未伴有非酮病之高血糖-高滲透壓的昏迷
- E0801--起因於潛在病的糖尿病，伴有高滲透壓，伴有昏迷
- E081--起因於潛在病的糖尿病，伴有酮酸中毒

資料篩選

基礎資訊:

- 性別
- 年齡

癌症疾病資訊:

- 組織學
- 臨床階段
- 病理階段
- 治療前生活狀態評估
- 特定疾病因子

用藥:

- 期間
- 藥物資訊(by ATC or NHI Code)

手術:

- 期間
- 手術資訊(by ICD-PCS Code)

檢驗檢查:

- 期間
- 檢驗資訊(by NHI Code)

共病:

- 追蹤時間
- ICD-9/10

併發症:

- 追蹤時間
- ICD-9/10

檢驗檢查報告:

- 期間
- (by NHI Code)

手術報告:

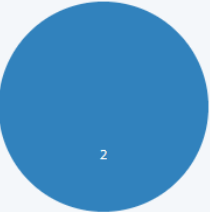
- 期間
- (by NHI Code)

整合平台醫療資訊共同模式

資料篩選

Sex

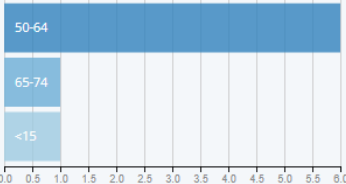
selected: [reset](#)



Category	Count
Sex	2

Age

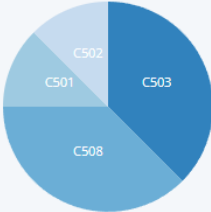
selected: [reset](#)



Age Group	Count
50-64	6
65-74	1
<15	1

Site

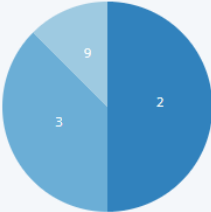
selected: [reset](#)



Site	Count
C501	1
C502	1
C503	2
C508	3

Grade_P

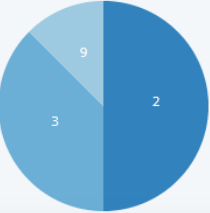
selected: [reset](#)



Grade_P	Count
2	4
3	4
9	1

Grade_C

selected: [reset](#)



Grade_C	Count
2	4
3	4
9	1

人口描述表

Description of Result

[download](#)

Show entries

Search:

Index	Count	Missing	Unique	Top	Freq
age_group	8	0	3	50-64	6
grade_p	8	0	3	2	4
grade_c	8	0	3	2	4
site	8	0	4	C508	3
sex	8	0	1	2	8

Showing 1 to 5 of 5 entries

[Previous](#) [1](#) [Next](#)



Report Output

整合平台醫療資訊共同模式

資料篩選

癌症資訊

全部下載

Sex

selected: reset

Grade

selected: reset

人口描述表

Describe

download

Show 10

Count

8

8

8

8

8

Showing 1 to 5 of 5 entries

請輸入癌症部位特定因子

癌症部位特定因子(ssf8) ①

請輸入癌症部位特定因子

癌症部位特定因子(ssf9) ①

請輸入癌症部位特定因子

癌症部位特定因子(ssf10) ①

請輸入癌症部位特定因子

基本資料

性別

請輸入性別

年齡

請輸入年齡

使用藥物 ①

請輸入追蹤起始時間(僅病前) ①

請輸入開始時間(月)

請輸入結束時間

請輸入結束時間(月)

搜尋

相關藥物 ①

請輸入藥物名稱

請輸入藥物代碼 ①

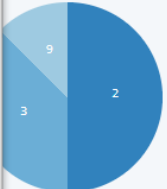
請輸入藥物代碼

CYCLOPHOSPHAMIDE 【L01AA01】 × METHOTREXATE 【L01BA01】 × FLUOROURACIL 【L01BC02】 × GEMCITABINE 【L01BC05】 × CAPECITABINE 【L01BC06】 × TEGAFUR 100MG + URACIL 224MG 【L01BC53】 × VINORELBINE 【L01CA04】 × PACLITAXEL 【L01CD01】 × DOCETAXEL 【L01CD02】 × DOXORUBICIN 【L01DB01】 × EPIRUBICIN 【L01DB03】 × CISPLATIN 【L01XA01】 × CARBOPLATIN 【L01XA02】 × TRASTUZUMAB 【L01XC03】 × BEVACIZUMAB 【L01XC07】 × ERIBULIN 【L01XX41】 × MEGESTROL 【L02AB01】 × LEUPRORELIN DEPOT 【L02AE02】 × GOSERELIN ACEATE 【L02AE03】 × TAMOXIFEN 【L02BA01】 × TOREMIFENE 【L02BA02】 × ANASTROZOLE 【L02BG03】 × LETROZOLE 【L02BG04】 × EXEMESTANE 【L02BG06】 ×

藥物建議使用的代碼

觀看設定檔

設定檔下載



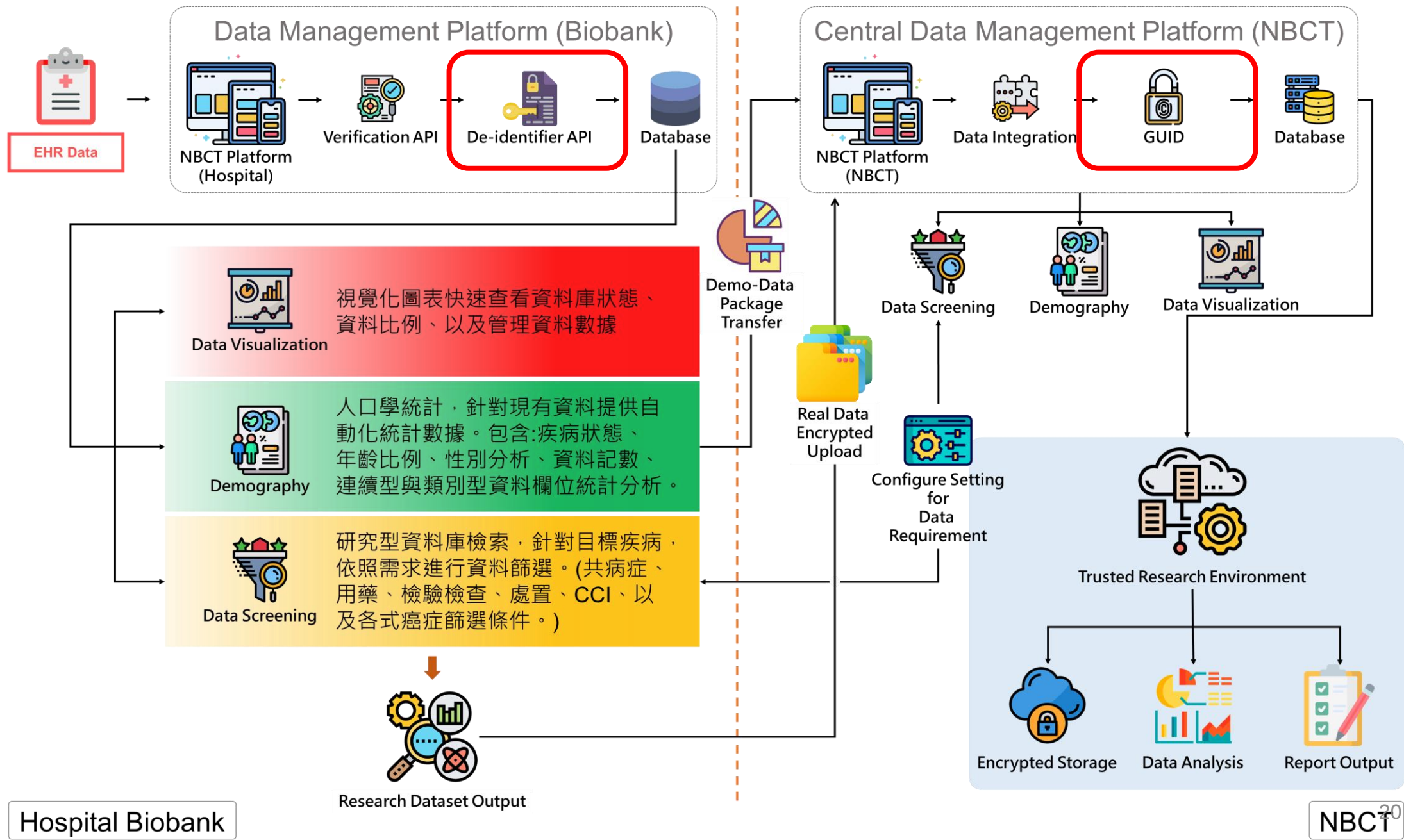
Search:

Previous

1

Next

整合平台醫療資訊共同模式



資料收集及去識別化流程



醫院擷取共同欄位所需要的所有檔案資料

- 彙整臨床資料的方案
- ① 方案A：
 - ② 方案B：

共同欄位資料進入轉換程式，形成hash

- ① 程式系統安裝於醫院內
- ② 包含未加密之身分證統一編號、生日完整資料轉換產出具hash之共同欄位資料
- ③ 身分證及生日與[hash ID]對應表存於各醫院內(mapping table 1)

由hash形成GUID

- ① 醫院提供去識別化後的完整共同欄位資料至整合平台
- ② [hash ID]生成[GUID]由整合平台辦公室完成
- ③ [hash ID]與[GUID]之對應表由整合平台留存(mapping table 2)

釋出時僅提供 GUID與共同欄位資料

- ① 後續出庫時，整合平台僅提供包含GUID的欄位資料給申請者
- ② 同一病人的就醫資料在跨機構合作之下，得以歸納在一起並兼顧去識別化



去識別化作業



患者	身分證統一編號	生日	共同欄位
A	A100000001	0541123	XXX
B	A200000002	0771102	XXX
C	A100000003	0440228	XXX

患者	身分證統一編號	生日
A	A100000001	0541123
B	A200000002	0771102
C	A100000003	0440228

hash轉換系統
安裝於醫院內



患者	Hash_ID
A	0DXJW23JDC
B	18FJ40C0D3
C	58VJ39DJ39

患者	Hash_ID	共同欄位
A	0DXJW23JDC	XXX
B	18FJ40C0D3	XXX
C	58VJ39DJ39	XXX

患者	Hash_ID	共同欄位
A	0DXJW23JDC	XXX
B	18FJ40C0D3	XXX
C	58VJ39DJ39	XXX



患者	Hash_ID
A	0DXJW23JDC
B	18FJ40C0D3
C	58VJ39DJ39

生成GUID由
整合平台完成



患者	GUID
A	1485c411-e06d
B	475fcc1d-b422
C	19ad9a8d-1240

患者	GUID	共同欄位
A	1485c411-e06d	XXX
B	475fcc1d-b422	XXX
C	19ad9a8d-1240	XXX

同一病人的就醫資料在跨機構合作之下，得以歸納在一起並兼顧去識別化



	身分證統一編號	生日
Patient A	A123456789	0541123
Patient B	B239384769	0771102
Patient C	F122398456	0440228



	身分證統一編號	生日
Patient A	A123456789	0541123
Patient D	A191059283	0830523
Patient G	P241833439	0590811



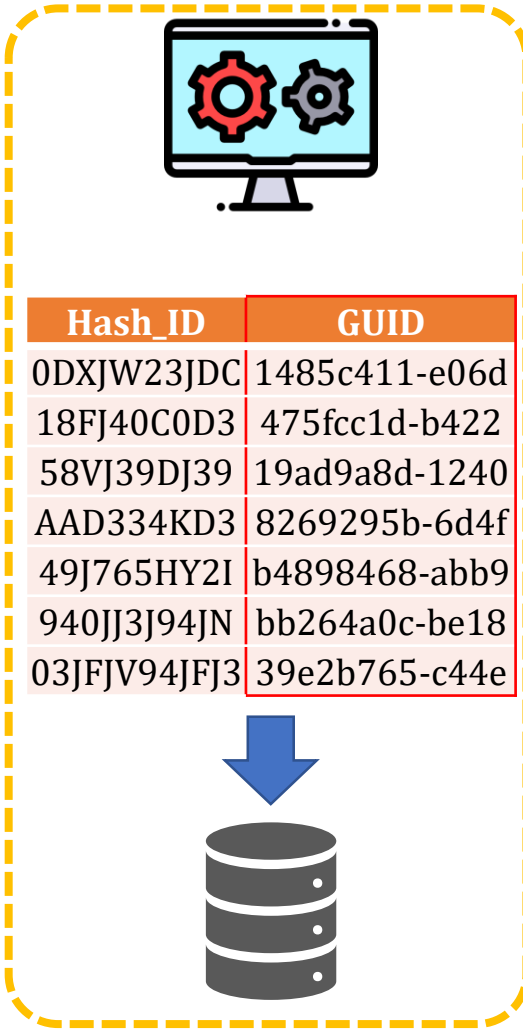
	身分證統一編號	生日
Patient E	J272627556	0820111
Patient F	Y152786348	0630228
Patient C	F122398456	0440228



身分證統一編號	生日	Hash_ID
A123456789	0541123	0DXJW23JDC
B239384769	0771102	18FJ40C0D3
F122398456	0440228	58VJ39DJ39

身分證統一編號	生日	Hash_ID
A123456789	0541123	0DXJW23JDC
A191059283	0830523	AAD334KD3
P241833439	0590811	03JFJV94JFJ3

身分證統一編號	生日	Hash_ID
J272627556	0820111	49J765HY2I
Y152786348	0630228	940JJ3J94JN
F122398456	0440228	58VJ39DJ39



同一位患者在不同機構，會形成同樣的GUID

資料對照表儲存機制

Mapping table 1 :
< 患者身分與hash ID對照表 > 存於各醫院內



患者	身分證統一編號	生日	Hash_ID
A	A100000001	0541123	0DXJW23JDC
B	A200000002	0771102	18FJ40C0D3
C	A100000003	0440228	58VJ39DJ39



患者	身分證統一編號	生日	Hash_ID
A	A100000001	0541123	0DXJW23JDC
D	A200000004	0830523	AAD334KD3
G	A200000007	0590811	03JFJV94JFJ3



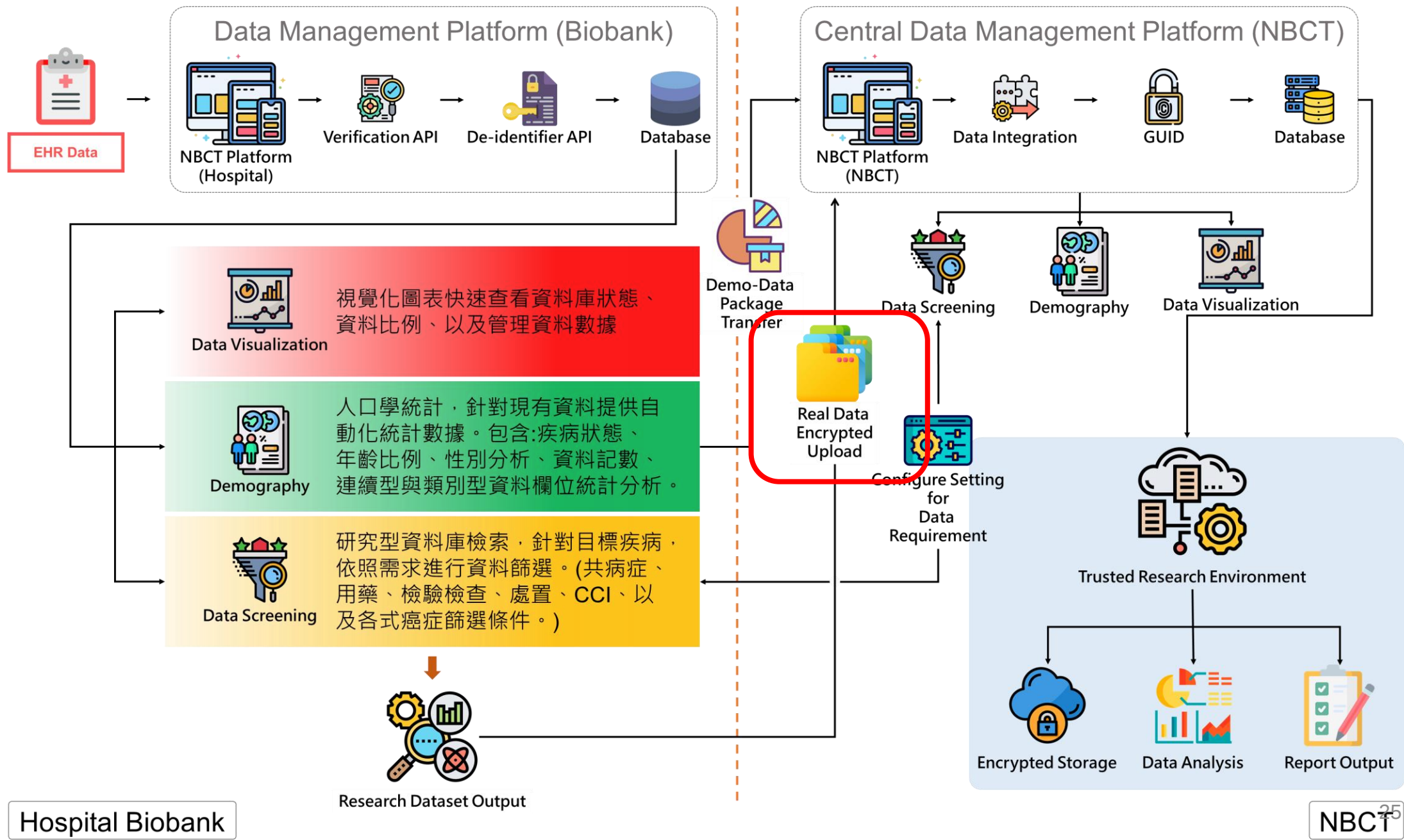
患者	身分證統一編號	生日	Hash_ID
E	A100000005	0820111	49J765HY2I
F	A200000006	0630228	940JJ3J94JN
C	A100000003	0440228	58VJ39DJ39

Mapping table 2 :
[hash ID]與[GUID]之對應表由整合平台留存

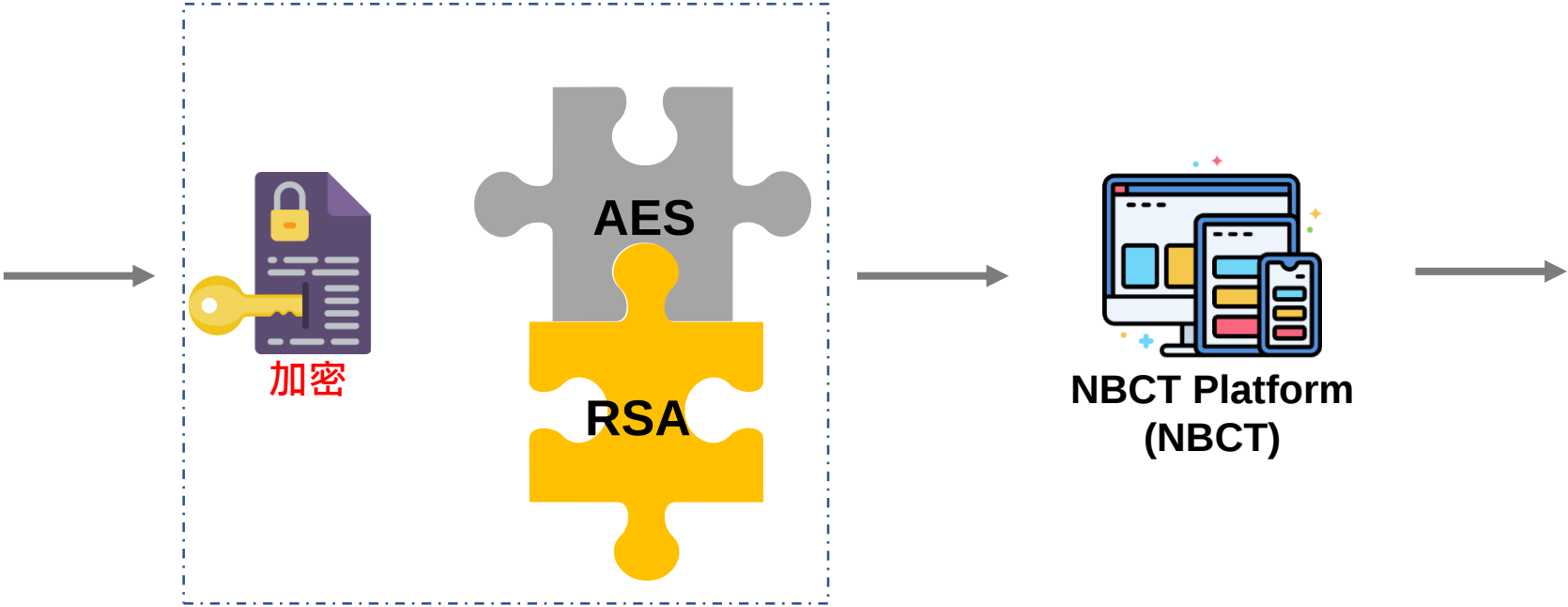


患者	Hash ID	GUID
A	0DXJW23JDC	1485c411-e06d
B	18FJ40C0D3	475fcc1d-b422
C	58VJ39DJ39	19ad9a8d-1240
D	AAD334KD3	8269295b-6d4f
E	49J765HY2I	b4898468-abb9
F	940JJ3J94JN	bb264a0c-be18
G	03JFJV94JFJ3	39e2b765-c44e

整合平台醫療資訊共同模式



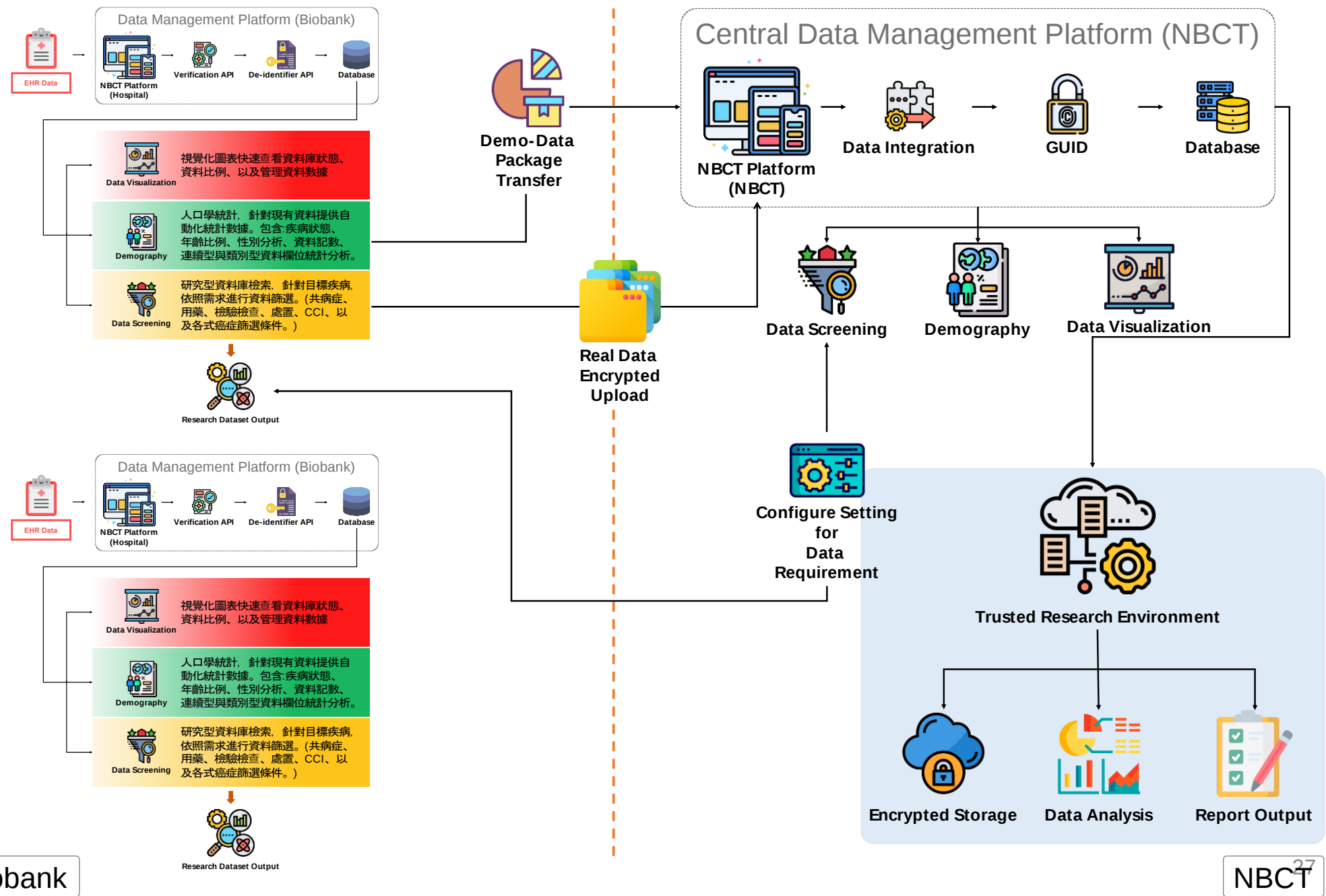
資料加密流程



AES
優點：適合加密大數據，加密速度快。
缺點：只有一把鑰匙，數據安全性低。

RSA
優點：兩把鑰匙，常用於隱私數據傳輸，數據安全性高。
缺點：加密數據量有限，加密速度慢。

整合平台醫療資訊共同模式



與各醫院討論機構內CDM的建置方案

醫院甲電子數據

數據清洗

數據整合

整合後數據

數據倉庫

A-specific database

B-specific database

project-specific database

數據分析平台

數據源

數據集成

數據分析報告

醫院乙電子數據

數據清洗

數據整合

整合後數據

數據倉庫

B-specific database

A-specific database

project-specific database

數據分析平台

數據源

數據集成

數據分析報告

方案一：資訊室提供 A 框的資料檔式給自己醫院 biobank，因需，再另協助 biobank 及 c 框

方案二：資訊室透過 CSV 檔式與 biobank 交換資料，biobank 再將資料檔式給自己醫院 biobank，需端再協助轉換成 C



針對測試檔案，核對欄位是否符合格式

測試檔案

欄位名稱	欄位值
姓名	張三
年齡	25
性別	男
職業	工程師
學歷	大學
婚姻狀況	未婚
收入	5000
住址	台北市
電話	123456789
電子郵件	abc@def.com

測試結果

欄位名稱	欄位值
姓名	符合格式
年齡	符合格式
性別	符合格式
職業	符合格式
學歷	符合格式
婚姻狀況	符合格式
收入	符合格式
住址	符合格式
電話	符合格式
電子郵件	符合格式

針對測試檔案，核對欄位是否符合格式

系統主畫面

- 系統資源資訊
- 系統管理與維護
- 門診資料轉檔與驗證

Biobank
個案

國際人類學生物學與農藝中心

提供高品質學術訓練 多元的職業機會

課程

人類學系

生物學系

農藝學系

環境與自然資源學系

社會工作學系

社會學系

心理學系

統計學系

教育學系

法律系

經濟學系

政治學系

國際關係系

國際貿易系

國際法系

國際經濟系

國際政治系

國際法律系

國際經濟法系

國際政治法系

國際關係法系

國際貿易法系

國際法學系

國際經濟學系

國際政治學系

國際法律學系

國際經濟法學系

國際政治法學系

國際關係法學系

國際貿易法學系

國際法學系

國際經濟學系

國際政治學系

國際法律學系

國際經濟法學系

國際政治法學系

國際關係法學系

國際貿易法學系

大學

國立中央大學

國立清華大學

國立台灣大學

國立交通大學

國立成功大學

國立陽明交通大學

國立政治大學

國立東華大學

國立高雄大學

國立高雄師範大學

國立高雄第一科技大學

國立高雄第二科技大學

國立高雄第三科技大學

國立高雄第四科技大學

國立高雄第五科技大學

國立高雄第六科技大學

國立高雄第七科技大學

國立高雄第八科技大學

國立高雄第九科技大學

國立高雄第十科技大學

國立高雄第十一科技大學

國立高雄第十二科技大學

國立高雄第十三科技大學

國立高雄第十四科技大學

國立高雄第十五科技大學

國立高雄第十六科技大學

國立高雄第十七科技大學

國立高雄第十八科技大學

國立高雄第十九科技大學

國立高雄第二十科技大學

國立高雄第二十一科技大學

國立高雄第二十二科技大學

國立高雄第二十三科技大學

國立高雄第二十四科技大學

國立高雄第二十五科技大學

國立高雄第二十六科技大學

國立高雄第二十七科技大學

國立高雄第二十八科技大學

國立高雄第二十九科技大學

國立高雄第三十科技大學

國立高雄第三十一科技大學

國立高雄第三十二科技大學

國立高雄第三十三科技大學

國立高雄第三十四科技大學

國立高雄第三十五科技大學

國立高雄第三十六科技大學

國立高雄第三十七科技大學

國立高雄第三十八科技大學

國立高雄第三十九科技大學

國立高雄第四十科技大學

國立高雄第四十一科技大學

國立高雄第四十二科技大學

國立高雄第四十三科技大學

國立高雄第四十四科技大學

國立高雄第四十五科技大學

國立高雄第四十六科技大學

國立高雄第四十七科技大學

國立高雄第四十八科技大學

國立高雄第四十九科技大學

國立高雄第五十科技大學

國立高雄第五十一科技大學

國立高雄第五十二科技大學

國立高雄第五十三科技大學

國立高雄第五十四科技大學

國立高雄第五十五科技大學

國立高雄第五十六科技大學

國立高雄第五十七科技大學

國立高雄第五十八科技大學

國立高雄第五十九科技大學

國立高雄第六十科技大學

國立高雄第六十一科技大學

國立高雄第六十二科技大學

國立高雄第六十三科技大學

國立高雄第六十四科技大學

國立高雄第六十五科技大學

國立高雄第六十六科技大學

國立高雄第六十七科技大學

國立高雄第六十八科技大學

國立高雄第六十九科技大學

國立高雄第七十科技大學

國立高雄第七十一科技大學

國立高雄第七十二科技大學

國立高雄第七十三科技大學

國立高雄第七十四科技大學

國立高雄第七十五科技大學

國立高雄第七十六科技大學

國立高雄第七十七科技大學

國立高雄第七十八科技大學

國立高雄第七十九科技大學

國立高雄第八十科技大學

國立高雄第八十一科技大學

國立高雄第八十二科技大學

國立高雄第八十三科技大學

國立高雄第八十四科技大學

國立高雄第八十五科技大學

國立高雄第八十六科技大學

國立高雄第八十七科技大學

國立高雄第八十八科技大學

國立高雄第八十九科技大學

國立高雄第九十科技大學

國立高雄第九十一科技大學

國立高雄第九十二科技大學

國立高雄第九十三科技大學

國立高雄第九十四科技大學

國立高雄第九十五科技大學

國立高雄第九十六科技大學

國立高雄第九十七科技大學

國立高雄第九十八科技大學

國立高雄第九十九科技大學

國立高雄第一百科技大學

The screenshot shows the Tableau Desktop interface with a dashboard titled '資料分析環境の分布' (Distribution of Data Analysis Environment). The dashboard is divided into three sections:

- Gender (性別):** A pie chart showing the distribution of gender. The legend indicates 'Male' (男性) and 'Female' (女性).
- Age (年齢):** A horizontal bar chart showing the distribution of age. The legend indicates 'Age Group' (年齢グループ) with categories like '18-24', '25-34', etc.
- Marital Status (婚姻状況):** A pie chart showing the distribution of marital status. The legend indicates 'Marital Status' (婚姻状況) with categories like 'Single' (未婚), 'Married' (既婚), etc.

The bottom of the dashboard shows the 'Columns' shelf with the 'Gender' field, and the 'Rows' shelf with the 'Age' field. The 'Marital Status' field is also visible in the 'Columns' shelf.

截至2022/11/21

後續

29

- 持續與各醫院合作建置NBCT CDM&系統
- 機構內NBCT CDM系統功能持續更新



Thanks!

Questions & Comments?

李修安

hsiuanglee@nhri.edu.tw

Billy72325@gmail.com