



NBCT

「醫療基因大數據之運用」 座談會



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/>

醫療大數據的收集

這段時間密集被討論...

- 一種 **Real-world data (RWD)**
- 可以產生 **Real-world evidence (RWE)**
- 用來查驗登記 & 支持臨床試驗

醫療大數據 遇上 Biobank

3



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, ...)

4

當醫療大數據 & 基因數據的結合

科學研究可以更進一步到:

**treatment response, adverse effects &
tumor recurrence**

超前部屬! 避免 undesired outcomes 發生!

UK biobank

5

- UK Biobank aims to improve the **prevention, diagnosis and treatment** of a wide range of serious and life-threatening illnesses.
- It is following the health and well-being of 500,000 volunteer participants and provides health information, which does not identify them, to approved researchers in the UK and overseas, from **academia and industry**.

<https://www.ukbiobank.ac.uk/>

All of Us

6

- *All of Us* is working to **improve health care through research**.
- building a diverse database that can inform thousands of studies on a variety of health conditions. This creates more opportunities to:
 - Know the **risk factors** for certain diseases
 - Figure out which **treatments** work best for people of different backgrounds
 - Connect people with the **right clinical studies for their needs**
 - Learn how **technologies** can help us take steps to be healthier

Popgen biobank

7

- Popgen is used to research molecular and non-molecular risk factors for numerous diseases of civilization. A better understanding of the molecular and non-molecular basis of these diseases should contribute to improved **prevention, diagnosis and treatment** of these diseases in the long term.

<https://www.epidemiologie.uni-kiel.de/biobanking/biobank-popgen>

Values for stakeholders

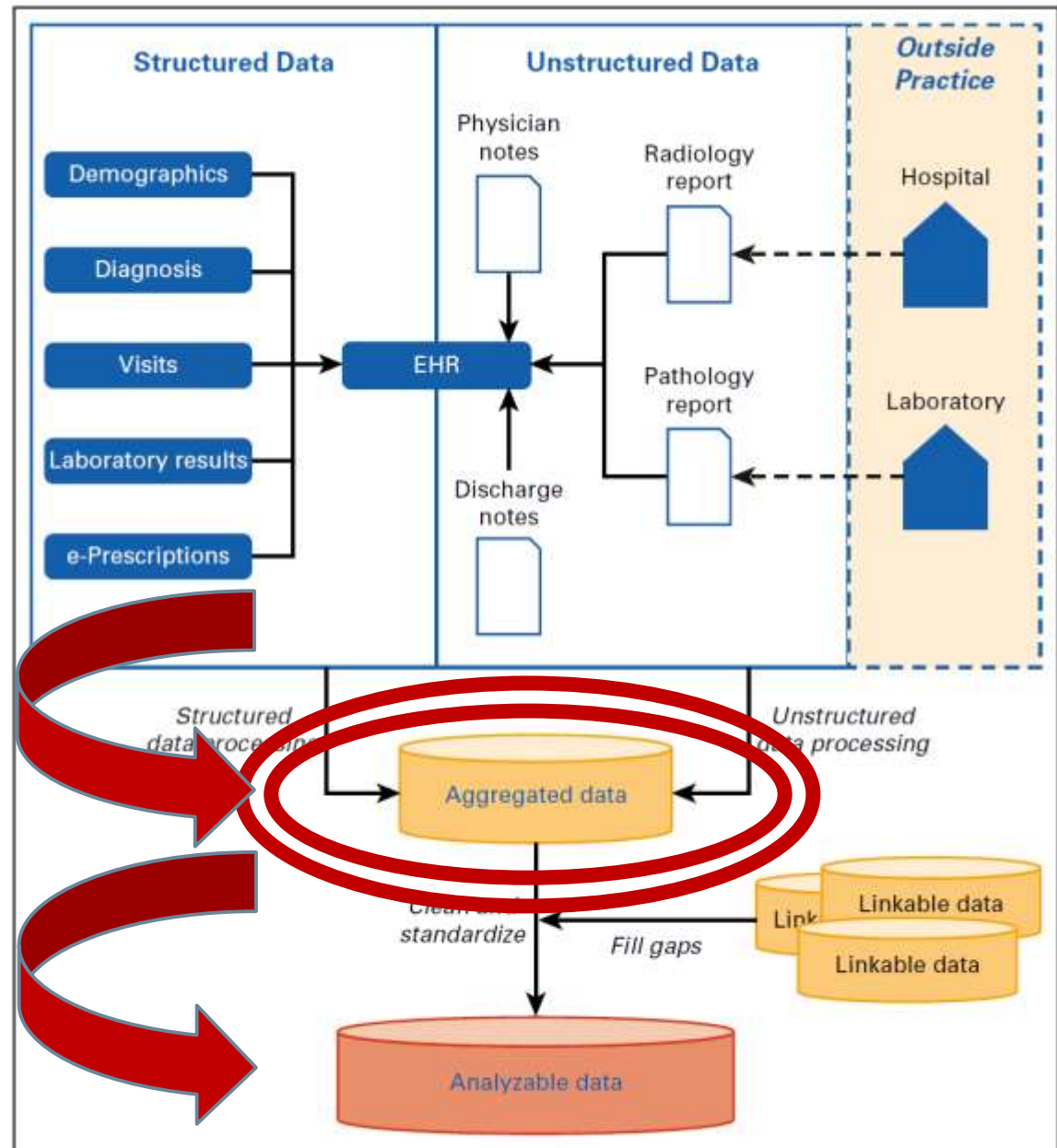
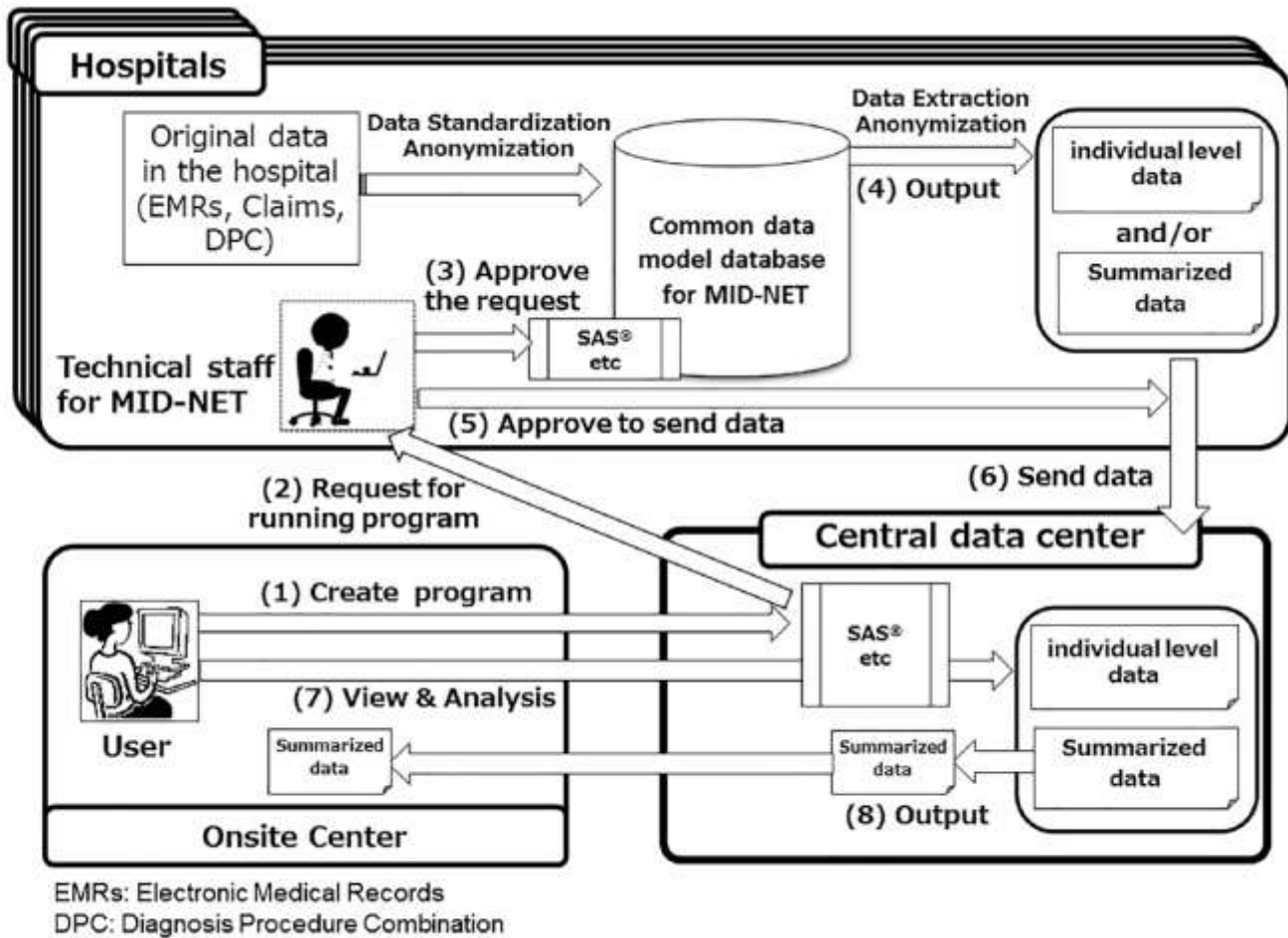
8

- **Accurate, timely diagnosis**
- **More certain and more favorable treatment outcomes**
- **Greater confidence in health decisions**
- **Improve quality of life**
- **Financial efficiency, cost-effectiveness**
- **Identify sub-groups of patients**

醫療大數據的收集

9

- 需要和醫療院所合作, 盡力減少醫院端的負擔, 使用現成上傳政府機構資料
- 需要**sufficient**: 包含診斷, 治療過程及治療追蹤的資料, 需要包含自費藥品及檢驗資料
- 需要達到國內的 interoperability (common data models/elements)
- 需要有國際 interoperability (FHIR HL7)
- 需要能夠維持足夠的資料品質
- 提升以實證資料分析為基礎的醫療照護品質



Outline of the MID-NET® system and the process of data extraction, transfer, and analysis

MID-NET: Yamaguchi et al.,
Pharmacoepidemiol Drug Saf 2019.

Fig 1. Process for converting **Flatiron:** Abernethy, et al., J Oncol Pract 2017

Example: 整理肺癌病人資料

12

假設某一機構biobank有2866個病人的檢體, 需要建置肺癌(lung cancer)病人的資料檔

Common data model

名稱

crlf101.sas7bdat

crsf101.sas7bdat

death103.sas7bdat

totfae103.sas7bdat

totfao103.sas7bdat

totfbe103.sas7bdat

totfbo103.sas7bdat

130個肺癌病人之基本診斷: 生日, 性別, 身高, 體重, 是否復發, 最初就診日, 診斷日, 確診方式, cell type, ECOG, EGFR

130個肺癌病人之存活狀況: 是否死亡, 死亡原因 (肺癌, 癌症, all cause)







130個肺癌病人診斷前之疾病: congestive heart failure, cerebrovascular disease, chronic pulmonary disease, peptic ulcer disease, diabetes

130個肺癌病人之癌症治療藥品的使用: 任一種癌症藥品, 任一種標靶藥品, 任一種化療藥品, gefitinib, erlotinib, afatinib; 開始日 & 結束日

CDM 中同一人的
不同性質資料存
放在不同的資料
檔

Common data model

名稱

-  crlf101.sas7bdat
-  crsf101.sas7bdat
-  death103.sas7bdat
-  totfae103.sas7bdat
-  totfao103.sas7bdat
-  totfbe103.sas7bdat
-  totfbo103.sas7bdat

VIEWTABLE: Inp.Crlf101

	ID	SEX	BIRTH_Y	RESID	DIAGAGE	SEQUENCE	CLASS	CONT_DT	DIAG_DT	CASITE	LATERAL	HIST	BEHAVIOR	GRADE	CONFIR
1		2	1947	3102	061	03	1	20090325	20090325	C182	0	8140	3	9	1
2		2	1947	3102	061	02	2	20090317	20090309	C187	0	8140	3	9	1
3										C189	0	8010	3	9	7
4										C209	0	8140	3	4	1
5										C343	1	8140	3	9	1
6										C343	1	8140	3	9	1
7		1	1946	3123	068	01	1	20150820	20150820	C220	0	8170	3	3	2
8		2	1946	3104	063	01	2	20090513	20090418	C502	1	8500	2	2	1

癌登資料每人每個癌症一筆

VIEWTABLE: Inp.Death103

	ID	ID_ROC	SEX	COUNTY	BIRTH_YM	D_DATE	D_LOCA_CODE	D_PLACE	D_TYPE	M
1							207	1	1	2
2							118	1	1	2
3							102	1	1	2
4		0	1	5101	195012	20140925	5105	4	1	2
5		0	2	0119	195805	20140117	0116	1	1	2
6		0	1	0101	194010	20140224	0116	1	1	2

死亡資料每人最多一筆

VIEWTABLE: Inp.Totfae103

	ID	HOSP_ID	FEE_YM	APPL_TYPE	APPL_DATE	CASE_TYPE	SEQ_NO	new_func_date	PRSN_ID	PHAR_ID	C
25		000182551	201401	1	00069898	09	18365	19753	@]@)*#*";<]")&)&@)~)	12
26		005268930	201401	1	00006250	01	48	19753	!%+!;*+/_	^^(;(&%([12
27								19789	<%<V]V^<[[][(;*;!/[&	12
28								19761	@]@)*#*";<	&%&%<(!*<:	12
29								19769	@]@)*#*";<]#)#((@	12
30		000182551	201402	1	00068888	09	5607	19784	=#=#+>[!/#]#)#((@	12
31		005268930	201403	1	00006937	01	620	19791	_%_<(>##[#	^^(;(&%([12
32		000172117	201404	1	00023320	09	64064	19815	[%]"& ->_	^^(;(&%([12
33		000182551	201404	1	00068858	04	13607	19862	<%<V]V^<[&%&=#)_[[-[12
34		000182551	201404	1	00068858	09	19213	19828	@]@)*#*";<	&%&%<(!*<:	12

就診資料每人多筆

VIEWTABLE: Inp.Totfao103

	ID	HOSP_ID	FEE_YM	APPL_TYPE	APPL_DATE	CASE_TYPE	SEQ_NO	ORDER_TYPE	DRUG_NO	DRUG_USE
289		005268930	201401	1	00006250	01	48	9	05203C	0
290		00069898	201401	1	00006250	01	48	1	MA1	0
291								1	A032282421	1 TID
292								1	A043589421	1 QDHS
293								1	A004068429	1 Q2H
294								0	00110C	0
295										
296										
297		005268930	201401	1	00006250	01	48			
298		005268930	201401	1	00006250	01	48			
299		005268930	201401	1	00006250	01	48			
300		005268930	201401	1	00006250	01	48			
301		005268930	201401	1	00006250	01	48			
302		005268930	201401	1	00006250	01	48			

醫療處置及藥品處方每筆就診
資料對應多筆

Lung_cohort.csv (patient-level data)

14

The image shows a screenshot of a CSV data file named 'lung_cohort'. The data is organized into columns, which are color-coded to represent different categories of information. The columns are: orange (basic diagnosis), blue (comorbidities), yellow (drug use), and green (survival status). The rows represent individual patients. The file name 'lung_cohort' is visible in the bottom left corner of the spreadsheet interface.

基本診斷: 年齡, 性別, 身高, 體重, 是否復發, 最初就診日, 診斷日, 確診方式, cell type, ECOG, EGFR

藥品的使用: 任一種癌症藥品, 任一種標靶藥品, 任一種化療藥品, gefitinib, erlotinib, afatinib; 開始日 & 結束日

共病症: congestive heart failure, cerebrovascular disease, chronic pulmonary disease, peptic ulcer disease, diabetes

存活狀況: 是否死亡, 死亡原因 (肺癌, 癌症, all cause)

Keep in mind

15

- 檢體&資料提供者的充分知情同意
- 個人隱私資料的保護
- 資通安全的維護
- 使用者的快速與方便使用



Thank You
for Your Attention!

Comments or Questions?

Email: yhyang@nhri.edu.tw