



中央研究院  
資訊科學研究所

Institute of Information Science, Academia Sinica • Taipei, Taiwan, ROC

TR-IIS-08-002

## 生醫資訊整合資料中心可行性評估 之方法與芻議

李宜芳、李國彬、張傳雄、王學偉、許鈞南、楊永正



Mar. 17, 2008 || Technical Report No. TR-IIS-08-002

<http://www.iis.sinica.edu.tw/page/library/LIB/TechReport/tr2008/tr08.html>

# 生醫資訊整合資料中心可行性評估之方法與芻議

作者：李宜芳<sup>1,\*</sup>、李國彬<sup>2,\*</sup>、張傳雄<sup>2,3</sup>、王學偉<sup>2,4</sup>、

許鈞南<sup>1,&</sup>、楊永正<sup>2,3,&</sup>

<sup>1</sup> 中央研究院資訊科學研究所

<sup>2</sup> 國立陽明大學系統與合成生物學中心

<sup>3</sup> 國立陽明大學醫學資訊研究所

<sup>4</sup> 國立陽明大學微生物與免疫學研究所

\* 並列第一作者(equal contribution)

<sup>&</sup> 責任作者(corresponding author)

## 目 錄

1	緣起 .....	1
1.1	背景說明.....	1
1.2	國內外產學界現況與發展趨勢.....	2
1.3	未來環境預測與問題評析.....	4
1.4	本文目標.....	5
1.5	本文範圍.....	5
2	資訊整合中心規劃建議 .....	7
2.1	組織架構與營運模式.....	7
2.1.1	營運中心設立與佈局.....	7
2.1.2	帶動生醫研究之運作模式.....	9
2.1.3	營運模式分析規劃方法.....	9
2.1.4	財務規劃.....	11
3	資料庫取得、更新與品質保證機制 .....	11
3.1	實體集中.....	12
3.2	分散式整合.....	14
3.3	安全性規劃.....	16
3.4	資訊安全政策.....	17
3.5	匿名化與去識別化 (De-Identification) .....	20
3.6	解決方案初步評估.....	20
4	可行性研究 .....	22
4.1	技術方面之可行性研究.....	22
4.2	模擬簡易實作方案.....	23
4.3	提供各種生醫應用情境.....	23
5	生醫資訊整合資料中心未來遠程規劃 .....	24
5.1	資料來源.....	25
5.2	規劃方式.....	26
6	參考資料 .....	27

## 圖目錄

圖 1. PESTEL 分析模型 .....	7
圖 2 顯示營運中心組織架構規劃。 .....	8
圖 3 整合資料中心營運模式規劃範例.....	10
圖 4 . SWOT 分析 .....	11
圖 5. 集中式資料中心架構圖.....	13
圖 6. 資料倉儲架構示意圖.....	14
圖 7. 分散式資料中心架構圖.....	15
圖 8. Web Services 授權憑證.....	16
圖 9. 資安政策實作規劃.....	18
圖 10. 資安政策導入.....	18
圖 11. 資訊安全處理小組組織架構.....	19
圖 12. 應用情境範例.....	24

# 1 緣起

## 1.1 背景說明

行政院於民國九十四年四月宣佈啟動台灣生醫科技島計畫，將以五年時間投入 150 億元預算，內容包括建立國民健康資訊基礎建設整合建置計畫 (NHIP)、台灣人疾病及基因資料庫先期規劃 (Taiwan Biobank) 及臨床試驗研究體系三大重點。生醫科技島計畫最主要的內容，在於利用資訊科技的資料建置，以實證為基礎，探索生命現象中遺傳資訊、疾病、及環境的交互作用，並整合政府重要生醫資訊相關計畫。要達到此目的，須有統一的資料交換標準格式，以使得不同單位所產生的資料可以交流互享，以擴大研究數量和擴展研究深度，令研究更可信。資料整合也可促進生醫科技在藥品、醫療器材及生技產品研發的速度，帶動及加速生技醫療產業的發展。

台灣在過去幾十年建立的生物醫學和臨床研究基礎相當雄厚，大量的生醫研究資料散佈在醫學研究中心、大學或政府研究機關。例如從民國六十八年起就開始收集資料的癌症登記資料庫，將所有癌症病人的診斷，治療和追蹤等資料集中存放。另外，全民健康保險在民國八十四年開辦以來，也已經累積了十二年的住院醫療費用和門診處方等資料。未來再加上台灣生醫科技島中規劃的台灣人疾病及基因資料庫和電子病歷等國民健康資訊建設，對生醫研究學者而言，這些基礎醫學和健康保險等資料，即使放眼全世界，也是不可多得的珍貴研究材料。

然而，以目前這種研究資料存放在個別資料庫的保管方式，想要產生一加一大於二的效果是十分困難的。第一，研究學者無法快速得知到底哪些資料是存在的；第二，當他們確定資料存在之後，必須辛苦的從各個資料庫分別收集所需的資料；第三，最後，由於每個資料庫各自開發編碼和去識別化等技術，不但浪費有限的資源，而且從不同來源獲得的資料無法互相直接比對。本文在資訊技術層次的主要目的就是在解決上述三個問題。我們希望一個整合型的台灣生醫資料中心，可以提供一個“一次購足”(One-stop Shopping)的環境，讓生醫學者在本地資料中心方便的找到想要的研究資料。

另一方面，在營運上不但要滿足學術研究的需求，也要支援產業的發展。在財務上除了初期的建制外，在營運上希望中心能自給自足，避免造成政府的負擔，所以必須先規劃良好的資源推估和財務分析。最後由於生醫研究資料包含很多個人隱私方面的敏感資訊，對於其應用所涉及的倫理法律及社會影響，是不可忽視的重要課題。本文所討論的設立生醫資料中心，是希望未來將醫療病歷資料，族譜戶籍資料，個人生活習慣，癌症治療追蹤，甚至是基因型資料都整合在一起，讓研究者可以交叉比對，長期追蹤。這中間牽涉到個人隱私的深度和廣度都遠超過傳統的生醫研究。況且，個人資料集中化之後，經手的技術人員極多，電子化的檔案又容易複製流通，在隱私保密層次又增加許多複雜度。因此，若不能解決跨平台的醫療資訊交換，及其應用所衍生的倫理法律及社會問題，就算在技術與營運上可行，也不宜貿然推動，必須在規劃時預先參考。

## 1.2 國內外產學界現況與發展趨勢

### 美國國家電子臨床試驗及研究網路 (National Electronics Clinical Trials and Research Network)

2003 年美國國家衛生研究院 (NIH) 院長塞烏尼 (Elias Zerhouni) 宣佈對美國政府資助的醫學研究進行重整。塞烏尼說，這項名為 NIH 發展路徑圖 (Roadmap) 的計畫，將使美國醫學研究人員克服接觸及共享數據的障礙。這項稱為「美國國家電子臨床試驗及研究網路」(National Electronics Clinical Trials and Research, NECTAR) 的計畫，把龐大且分散的資料結合成一個巨大的資料庫，最終目的在於促進轉譯醫學 (Translational Medicine) 的發展。

從事傳統的臨床試驗時，研究人員常常需要自己重新建立一些原本就存在的資料，例如病人治療及診斷的數據。然而這些數據或資料很有可能已經被別人做過，而且放在一些醫院或政府機構的資料庫裡。對研究人員而言，問題是他們無從得知哪裡有哪些資料，所以不得不自己重複蒐集或建立這些資料。美國國家衛生研究所的 NECTAR 正好可以解決這類的問題，研究者因此可以有更多的時間來專注在研究內容上或者擴大原本非常受限的研究範圍。美國國家衛生研究所一共資助設立了七個國家級的生物醫學計算中心 (National Centers for Biomedical Computing)，範圍包含系統生物、影像分析、資料整合、疾病研究等等。其中 Informatics for Integrating Biology and the Bedside (I2B2) 是其中一個研究中心。它的任務在整合臨床試驗和基礎生物醫學研究所產生的資料，以加速研究者對大型複雜疾病的了解。

### 美國癌症生醫資訊格網 (caBIG) 目前發展現況

目前美國國家癌症研究所 (National Cancer Institute, NCI) 大力推展之癌症生醫資訊格網 (Cancer Biomedical Informatics Grid, caBIG) 計畫已邁入第三年，也已正式納入 NIH 的重要推動項目之下。目前世界各國的學術研究機構相當踴躍的參與 caBIG 社群各種相關標準之開發與應用，致力於加速生醫資料之標準化與交換。也由於 caBIG 是一個開放式的研究團體，擁有許多主流大學與研究機構的參與，短短三年間已開發了相當多的基礎建設工具與資料集。如 caGRID 就是採用 Globus 工具組開發的網格 (Grid) 環境，目前已經提供相關之應用程序介面 (Application Programming Interface, API) 與工具組來提供基本網格環境之建立與開發，此外還提供相當多有關字彙對應工具 (Vocabulary Mapping Tools)，如 caBIG LexGRID 即是 Mayo Clinic 所提供的字彙伺服器，提供控制字彙對應的服務。另外一個廣為使用的例子是癌症資料標準倉儲 (Cancer Data Standards Repository, caDSR) 的開發，目前 caDSR 收集了一萬五千多個生醫資料庫所需要的資料欄位名稱，並利用統一醫學語言系統 (Unified Medical Language System, UMLS) 的 Concept Unique Identified (CUI)，來標示資料欄位所屬的 Concept 名稱。因此，各個研究單位利用 caDSR 所提供的共同資料元件 (Common Data Elements)，將可以使用標準的欄位名稱，很快的與其他研究單位互相交換資料，或者將各自研究單位所擁有的區域性字彙對應 (Local Vocabulary Mapping) 到 caDSR，將可以把研究單位所擁有的資料模型連接到參

考模式 (reference model)，如此以來，所有與參考模式 (Reference Model) 有聯結的資料欄位都可以互相交換與對應。

### 國內生醫資訊

自從全民健康保險開辦以來，全台灣所有醫療院所的門診、急診以及住院資料，均統一申報至中央健康保險局，並建立一套完整的全民健康保險研究資料庫，簡稱健保資料庫 (National Health Insurance Research Database, NHIRD)，提供研究者學術研究之用。健保資料庫的內容，除了保險給付相關的金額資料外，更重要的，包含了病患的診斷、處置、用藥等等的處方與醫令資料，這些完整的就醫歷史資料，將有助於全民健康與醫療行為等相關議題的回溯性研究。

藉由健保資料庫中，完整的門診處方及治療明細檔與門診處方箋明細檔，可以長期研究臺灣藥物使用與診斷之間的關係，以及藥物交互作用與不適當用藥造成病患安全上之可能重要影響。另一方面，健保資料庫也可以用來研究疾病與疾病之間的關係，除了需要參考門診處方及治療明細外，尚需有住院醫療費用清單明細的輔助，方能反映出單一病患疾病診療及處置的過程，進而完整與全面性的探索國內疾病與環境的關係，以及可能的相關因子。

以國內而言，除了中央健保局委託國家衛生研究院保管並開放申請全民健保研究資料庫，以及衛生署推動的網路健康服務推動計畫外，很少有大型的生醫資料庫整合計畫。癌症登記資料庫是衛生署委託癌症登記工作小組進行癌症資料收集，戶政資料由內政部戶政資訊小組掌管，死因檔作業則由衛生署統計室負責。「台灣人疾病及基因資料庫」的資訊分項先期規劃是委由資策會創新應用服務研究所負責。在資策會的先期規劃中，的確有研究全民健保研究資料庫和台灣人疾病及基因資料庫的資料交換問題。不過由於台灣人疾病及基因資料庫尚未正式開始收集案例，相關倫理法律議題也還沒有完全釐清，這裡的資料交換還停留在規劃階段。NHIP 部份雖將推動病例電子化，但在倫理、法律與社會議題上仍有爭議。因此本計畫將 Biobank 與 NHIP 都列為遠期規劃。本文對於整合全民健保，癌症登記，死因和戶籍這四個資料庫作簡單的介紹。

由於健保資料庫囊括全體國人珍貴的醫療資料，中央健康保險局應學界之請託，在保障民眾隱私以及資料安全的前提下，委託給國家衛生研究院以全民健保的資料為基礎，建立一個以學術研究為目的之資訊資料庫，提供給學術單位及非營利機構之學者專家進行醫藥衛生相關研究，由中央健康保險局提供加密的資料檔案，經抽樣、分檔等步驟，建立新的資料庫供學術研究用。希望透過學界的研究分析，能對臺灣地區的醫療保健及公共衛生相關問題有所瞭解，進而提出更完善的建議供決策者參考。資料庫內容包含醫事機構檔、醫事人員檔、醫事病床檔、門診與住院申報檔。國家衛生院的網站有清楚的申請步驟和所需費用的說明。從其網站並可得知，九十五年度共有 201 件申請案，九十四年度有 256 件申請案。

癌症登記資料庫始於民國六十八年，衛生署建立 50 床以上醫院為申報單位之癌症登記系統。後來於八十五年委託『癌症登記工作小組』進行癌症資料收集。於九十二年公佈之癌症防制法，規定癌症預防治療機構應向中央主管機關所委託之學術研究機構，提報新發生之癌症各案與期別等相關診斷及治療資料，確立了

癌症登記的法源依據。資料內容包括申報機構，各案基本資料、診斷資料、治療方式及追蹤資料。目前由中華民國公共衛生學會癌症登記小組負責收集，國民健康局癌症組為保管單位。

民國九十四年公佈之戶籍法修正案，增列各檢察機關及醫療院所，於出具死亡相關證明後，必須通報當地政府。醫療院所應於七日內，以網路傳輸方式透過死亡通報資料自動傳遞平台通報。其資料格式包括死亡者姓名、身分證字號、地址、死亡時間、場所、種類、原因、發病與否、醫師姓名等。這些資料最後彙整於衛生署，形成全國死因資料檔，簡稱死因檔。目前死因檔資料由衛生署統計室保管。

自民國八十六年九月全國戶役政資訊系統連線，以電腦化作業取代人工作業後，全國各戶役政單位全面以電腦連線作業方式處理戶役政業務。全國性之戶籍資料庫可以提供查詢各種戶籍統計資料，例如當月戶數、出生、結婚數、每戶年齡組成等。目前保管單位為內政部戶政司戶籍作業科。目前戶政司規定公務機關及學術研究可以透過戶政資訊系統申請戶役政資料。戶籍資料庫對生醫研究的價值在於它提供研究者一個追查家族成員和個人遷徙的資料。例如對於遺傳性疾病的研究，家族成員的資料是非常重要的。

前文提到行政院於民國九十四年開始啟動台灣生醫科技島計畫。計畫中的第一個主要分項是國民健康資訊基礎建設整合建置計畫(NHIP)，是在資訊安全的前提下，建立電子病歷及醫療影像傳遞的標準格式，並建置遠距醫療的資訊共通平台，希望能透過網際網路來做好個人化健康管理。第二個分項是台灣人疾病及基因資料庫，目的在探索基因，疾病及環境的交互作用，希望能應用在個人化疾病診斷、治療與預防。此資料庫涉及科技對倫理法律及社會議題，而有較多社會疑慮。第三個分項屬於在臨床研究方面，衛生署已推動卓越臨床研究中心與專科卓越中心，分別設在臺大醫院、三軍總醫院、萬芳醫院和成大醫院。另外衛生署也在國家衛生研究院成立國家臨床試驗網絡，在生技中心成立轉譯醫學臨床研究中心。如何納入這些資訊，以強化連結基因基礎與致病機制的轉譯醫學研究，是本計畫中「未來規劃」部份的重點。

### 1.3 未來環境預測與問題評析

目前政府在生醫科技島計畫下，有國民健康資訊建設，台灣基因資料庫，臨床研究資訊管理系統等大型計畫。它們各自有自己的資訊系統，若無法交換、共享、甚至整合這些資訊，就無法產生協同增益(Synergy)的效果。此外，在國際上雖然也未達到整合的理想，但也逐漸朝此方向邁進，例如美國的CaBIG(Cancer Biomedical Information Grid)計畫。因此有發展共通平台，整合上述資訊的必要。

在有誘因與政府計畫的要求下，確實有資料共享之需求，作為專家們取得共識的基礎。如果政府機關有足夠公權力，可以要求各資料保管者，將其保管之資料以集中形式保管在整合資料交換中心。這樣的優點是使用者將可以較便捷的瀏覽或存取資料中心保管的資料，但缺點是資訊安全或隱私洩漏的可能會提高。如果我們採用分散式資料中心，要求資料保管者負責安全、隱私等問題，即使一個



資料庫產生資訊安全上的漏洞，也不會影響到其他的資料庫。當然分散式保管資料的共享資料中心在資料瀏覽和存取上必然不若集中式資料中心來的直接。在本文中，建議應請專家們達成共識找出最佳方案，並規劃出未來發展的路徑圖，供衛生署作為未來決定是否要進行建置的參考。

除了健保資料，癌症資料，戶籍和死因等資料庫外，由學者自發性參與的亞洲 Cohort 聯盟，或是由學會等推動疾病登錄系統，都可以成為學者們選擇病患參與臨床實驗的來源。雖然前述四項是本文規劃資料交換的目標，但是在未來卻需要考慮其他的資訊來源。例如在臨床試驗方面，在選取可能的病人參與臨床試驗後，臨床資料將存放於臨床資訊管理系統，而由病人樣品所衍生的功能基因體資訊，可能置於國家基因體計畫下的生物資訊核心設施，或由學者自行保管。這些資訊可能由不同的團隊產生，必須交換、共享才能瞭解致病機制，若能在平台上整合，並作資訊探採則是現在的夢想。不過這一切是否可行，均須通過倫理、法律與社會議題的考驗，不宜貿然推動。

## 1.4 本文目標

1. 規劃生醫資訊整合資料中心，集中整合現有生醫資料庫，以提供各資料庫間之資料比對服務，促使生醫領域資料得以有效的整合與利用。
2. 考慮生醫資訊整合相關的倫理、法律及社會影響議題（ELSI）問題，研擬出既可保障個人隱私又可促進生醫資訊發展的資訊平台。
3. 進行可行性研究及規劃未來新興生醫資料庫整合計畫。規劃生醫資訊整合資料中心之軟體硬體架構，提供各資料庫間資料比對服務能力，建立不同資料來源之資料交換標準。

## 1.5 本文範圍

本文主要討論如下：

1. 如何規劃生醫資訊整合資料中心之組織架構及營運模式。
2. 如何規劃下列四個資料庫取用及資料更新機制。
  - (1) 全民健康保險研究資料庫
  - (2) 癌症登記資料庫
  - (3) 死因資料庫
  - (4) 戶籍資料庫
3. 如何規劃生醫資訊整合資料中心的服務品質保證機制。
4. 以資訊平台健全生醫資訊整合相關倫理，法律，及社會影響議題。
5. 生醫資訊整合資料中心可行性研究及未來遠程願景規劃，未來新興生醫資訊資料庫整合計畫。

建議資訊整合中心應增的工作項目：

1. 設立專屬網站，推廣資訊共享對生醫和臨床研究的重要性，一方面及早讓民眾有表達意見的機會，另一方面也可透過資訊透明化，化解民眾的疑慮。
2. 邀請人權團體、原住民代表、消基會、台灣醫療改革基金會、立法委

- 員參加公開說明會，力求減少民眾對整合生醫資料所產生之疑慮。
3. 模擬簡易實作方案，呈現保護隱私的能力。
  4. 提供學者使用生醫整合資訊的情境，作為營運中心未來推展之參考。
  5. 開「資料庫品質保證專家會議」與「可行性評估專家會議」各一場。

## 2 資訊整合中心規劃建議

### 2.1 組織架構與營運模式

#### 2.1.1 營運中心設立與佈局

在初期規劃下，整合資料中心應以設立在公立學術研究機構為原則。原因如下：(1) 設立在學術機構可以獲得價格遠為低廉的軟體和硬體設備；(2) 存放在本資料中心的資料所有權應為全體國民，公立或學術機構較易於實現一個不以營利為目的之生醫資訊整合資料中心，預期將有利於降低社會大眾對個人資料不當使用之疑慮；(3) 評估設在公立學術機構，將有利於資料中心長遠並穩定的運作。但也應研究將營運中心設立在私人企業或私人學術單位的利弊。

建議規劃數種營運中心設立的模式，比較個別設立模式（包括私人企業）之間的優缺點，提供政府最後決策的參考。

- 營運中心設立在財團法人組織之下，國內方面，公設財團法人的成功範例是國家衛生研究院負責的全民健保研究資料庫；私人方面有財團法人醫學研究倫理基金會設立的聯合人體試驗委員會。以整合生醫資料中心而言，財團法人可能的選擇有國家實驗研究院的國家高速網路與計算中心、國家衛生研究院等。
- 營運中心設立在社團法人組織之下，可能的選擇有台灣醫學資訊學會、台灣生物資訊學會等。
- 營運中心設立在國立大學或中央研究院。
- 營運中心設立在衛生署內部。
- 營運中心設立在私人企業，國外現有的例子是美國的 United BioSource Corp.，旗下有癌症資料庫（Oncology Electronic Medical Records）和病人資料庫（Individual Patient Databases）等生醫資料庫。

建議營運中心設立的分析及規劃宜採用 PESTEL 方法。PESTEL（Political, Economic, Social, Technological, Environmental, Legal）是一個在策略管理（Strategic Management）上常用的一個分析模式。此模式通常用六個外在環境的影響因子來分析某一個營運中心、組織或營運模式是否會成功。

圖 1 為 PESTEL 方法的示意圖。

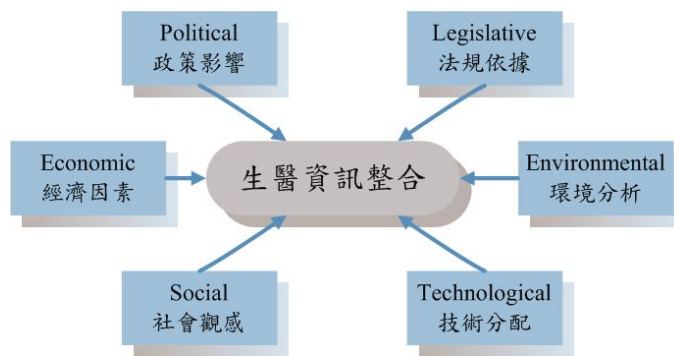


圖 1. PESTEL 分析模型

- 以政治環境而言，營運中心設立在政府機關可以方便利用國家公權力來配合行政院對台灣生醫科技島的長遠規劃。
- 以經濟層面而言，營運中心設立在國家電腦中心或許是較適當的選擇，因為硬體環境和人力資源的考量。
- 從社會層面來講，我們要考慮營運中心設在哪裡最容易為社會一般大眾接受，因為營運中心掌握了牽涉個人隱私的大量資料，或許政府機關和國立大學會產生較少爭議。
- 以技術層次來看，我們要考慮哪些機構擁有設立生醫資訊共享資料中心的能力，這中間包括領導團隊的研發經歷，執行人員的教育訓練背景，以及他們對保障資訊安全的能力。
- 最後，環境以及法律的考量，資料中心設立在財團法人或私人企業是否違法、設立在政府機構是否會限制資料中心在經費運用以及人才招募上的運作等等。

建議應用 PESTEL 方法來評估資料中心資料庫系統的規劃。當然最後資料中心的設立模式以及資料分享的營運模式都可在召開的專家會議中以德爾菲法 (Delphi Method) 或其他類似方法由專家小組得到一個合理且可靠的結論。

圖 2 描述營運中心的組織規劃架構。營運中心組織架構力求精簡，各組專業分工，包含諮詢委員會、審查委員會、資訊安全委員會、系統維護部門以及行政管理辦公室。諮詢委員會由生醫或資訊領域的專家組成，為營運中心最高指導單位，負責監督中心資料庫的合理使用。審查委員會進行生醫研究申請案之審查業務，審查委員會聘請生醫、資訊和 ELSI 等各界專家學者，對於研究案進行獨立不受干擾之審查。資訊安全委員會負有監督營運中心資訊安全的任務，並確保營運中心符合行政院訂定的資訊安全管理規範。目前外界對現行生醫資料使用的資訊安全機制頗有疑慮，例如消基會認為資訊安全須由民間團體或外界資訊安全專家加以評估，不能僅由資料管理單位自認為安全就將資料公開。資訊整合中心的資訊安全監督委員會，在規劃中可以在審查委員會和行政管理辦公室之上獨立進行資訊安全稽核；此外，也應由外部的資訊安全委員會來做不定期的檢察 (參閱 3.4 與圖 11)。系統維護部門負責軟硬體和資料庫維護等業務。行政管理辦公室負責網站維護，客戶服務及中心推廣等業務。

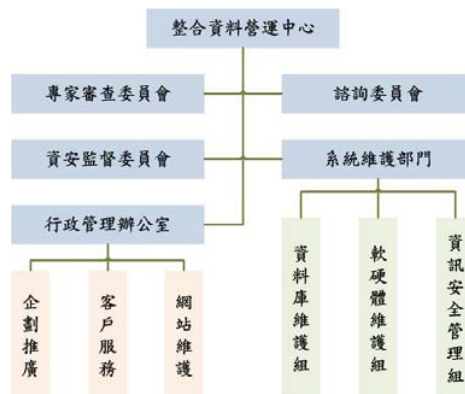


圖 2 顯示營運中心組織架構規劃。

### 2.1.2 帶動生醫研究之運作模式

生醫科技島計畫的主要內容，在於利用資訊科技的資料建置，以實證為基礎，探索生命現象中遺傳資訊、疾病及環境的交互作用，並整合政府重要生物資訊相關計畫。本文所規劃的整合資料中心，其目的在為生醫研究學者解決資料共享的障礙，以加速轉譯醫學的發展。以下我們簡單列舉數種生醫研究的運作模式：

- 衛生福利健保政策相關的研究：例如國立陽明大學衛生福利研究所吳肖琪教授於民國九十年應中央健保局委託，從事健保按例計酬對醫療品質的相關性研究。這類研究需要互相參照健保資料庫和衛生署死因資料。在當時又由衛生署資訊室直接提供可整合的資料，但在資料種類不斷增加的情形下，衛生署的人力勢必無法經常性地提供這種服務。本文所建議的資料中心提供了一個授權單一窗口，研究人員可以很方便的找到他們需要的生醫研究資料。
- 癌症醫療資源使用：癌症是國內十大死因的第一名，關於癌症的醫療資源使用，自然變成極端重要。假設研究人員計畫分析國內癌症病人的醫療利用，探討不同癌症病人在病人特質、醫院特質、醫療服務類型、治療處置與癌症治療分期之醫療資源使用情形（例如次數、日數及費用）。在這個例子中研究人員可能需要癌症登記資料檔，全民健保資料檔，甚至戶籍資料來追蹤患者家族患病史。就像前文提到的，美國國家衛生研究院成立所謂 NECTAR 計畫就是想要整合這類大型臨床研究牽涉到的分散且資料量很大的資料庫，本資料中心設計將健保資料庫、癌症登記資料庫、死因資料庫和戶籍資料庫整合在一個平台，這將有助研究者從事大型臨床研究。
- 整合基因資料和臨床研究：近年來基因體的研究已經快速累積許多人類基因型和疾病發生進展的相關研究。例如人類 MDM2 基因上的 SNP309 已經被證實和乳癌的發生有關。研究更進一步發現，在女性患者身上，有無動情激素接受體（Estrogen Receptor）對乳癌的發展有明顯相關。今天假設一個研究者想從流行病學分析角度，研究服用賀爾蒙藥物對於癌症發生率的關聯性，他必須利用癌症資料庫和健保資料庫的資料。如果研究者想研究動情激素接受體陽性和陰性的女性的癌症發生年齡，他可能需要透過戶籍資料或台灣生物資料庫來取得實驗對象的血液檢體來做基因分型，若沒有一個整合型生醫資料交換平台，這一類的生醫研究將非常難以進行。

### 2.1.3 營運模式分析規劃方法

規劃營運模式時至少應研究以下三種營運模式：

1. 衍生自國家衛生研究院保管全民健保資料庫的成功經驗，第一種營運模式是資料統一由資料中心保管，欲使用本中心資料的研究者，必須透過線上申請程序，詳述研究目標、研究方法、研究範圍、所需資料種類、有無倫理法律爭議等資料，經本中心聘任的專家諮詢審查小組委員審查通過後，中心工作人員會將相關資料從資料庫中擷取出來，

以光碟或是網路傳送的方式交付研究人員。這裡所提到的資料是集體資料，研究人員無法辨識當中的個體資料。圖 3 表示此種營運模式，資料供給端和消費端並不會直接接觸，而是透過整合資料營運中心來產生互動。中心網站會公佈存放在本中心資料庫內的資料種類、資料範圍，以及詳細的申請手續、資料收費標準，研究人員在此種營運模式之下無法直接存取資料庫內容。從應用情境 (Scenario) 來說，研究者可能採用這樣的研究流程：(1) 首先產生計畫構想，再於整合資料中心網頁上瀏覽中心保管的資料資訊；(2) 符合開放申請帳號的研究者可在網站上申請帳號；(3) 有帳號後，研究者可以線上填寫資料，簡單敘述研究目的、範圍、方法，並精確描述擬申請下載的生醫研究資料；(4) 本中心內部的專家審查委員會將會初步審查是否接受該研究者的申請；(5) 若申請通過，資料中心會和申請者確認費用，一旦繳費完成，申請者將能以光碟或網路傳輸的方式獲得資料；(6) 研究完成後，如果使用者就職於國內公私立大專院校、政府所屬業務相關單位、或非營利研究機構，並將研究成果發表於學術期刊或學術研討會，使用者必須將研究成果回報整合資料中心，不可擅自利用原資料從事另外未經申請的研究。至於非學術界使用者是否須回報研究成果，應在資訊中心進行建置時另行規劃。

### 營運模式範例

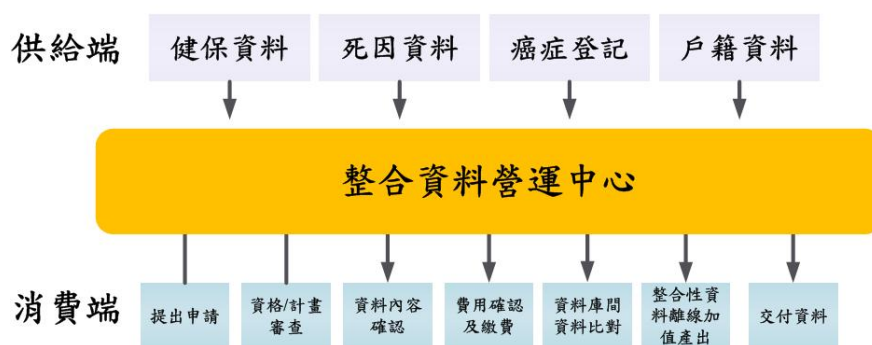


圖 3 整合資料中心營運模式規劃範例

- 第二種營運模式類似國外的 CRO (Contract Research Organization)，由本中心專任並具統計或生醫研究背景的研究人力，替欲使用本中心資料的學者規劃並從事臨床相關研究。特點在研究學者將其臨床或生物醫學相關研究直接委託本中心設計研究方法和從事數據分析。
- 第三種營運模式是最為開放的模式，研究學者取得帳號以及相關使用資料庫權利之後，將可以在本中心專屬網站上直接瀏覽健保、戶籍或死因等資料。本中心並不提供大筆資料下載服務，研究者無法批次下載大量的資料。如果他們有大量資料存取的需求，可以通過上述第一種營運模式來申請。

在執行規劃時，應可以採取強弱機危綜合分析法 (Strengths, Weaknesses, Opportunities and Threats, SWOT Analysis) 來考量各種營運模式的優勢、劣勢、



機會與威脅。以上述第一種營運模式來說，圖 4 列出了可能的 SWOT 分析。

STRENGTHS	WEAKNESSES
<ul style="list-style-type: none"> <li>資訊安全較易實現</li> <li>專家審查委員 (IRB) 可以幫助審核研究範圍和資料運用</li> <li>整合資料中心不需介入申請者的研究，節省中心人力</li> </ul>	<ul style="list-style-type: none"> <li>資料本體隱藏在資料中心，對申請者而言較為不便</li> <li>申請手續較為冗長</li> </ul>
<ul style="list-style-type: none"> <li>採用類似現行健保資料庫開放方式，部分研究者已熟悉此營運模式，有利於資料中心日後的推廣</li> </ul>	<ul style="list-style-type: none"> <li>如果專家審查委員會對於生醫資料共享採用相對嚴格的標準，將影響學者使用資料中心的意願</li> </ul>
OPPORTUNITIES	THREATS

圖 4 . SWOT 分析

#### 2.1.4 財務規劃

討論不同經營模式，會有不同的經費需求。財務規劃的目標是為了減輕政府的財務負擔，希望在初期的投資後，逐漸可做到自給自足。因此必須考慮到設備的折舊更新及各種運作成本(包括：人事成本、ISO 認證費用、系統維修、水電、辦公室租用、電腦房租用、推廣費用、委員審查費及交通費、工作人員用電腦、軟硬體等費用)，以便計算該如何收費。在另一方面，收費可能又會讓使用者卻步，因此必須有好的宣傳策略，並有成功的案例。對於規劃後所提出的營運模式針對其營運能力、盈利能力和發展能力做財務綜合分析，將各方面財務分析納入一個整體之中，全面地對經營狀況、財務狀況進行解剖與分析。綜合的分析法可採用杜邦財務分析法。

杜邦財務分析體系 (The Du Pont System) 是利用各財務指標間內在關係，對營運單位綜合經營理財及經濟效益進行分析評價的方法，最初由美國杜邦公司創立並成功的運用而得名。杜邦財務分析體系的核心是淨資產收益率，其他各項指標均圍繞著這一指標，通過研究彼此間的依存關係，可以展示事業體的獲利能力及其前後因果。

通過對杜邦體系自上而下的分析，不僅可以瞭解營運規劃的各項財務指標間的結構關係、查明各項指標變動的影響因素，而且還可以最佳化經營理財狀況，針對定價策略、節約成本、優化資金配置、加強資產周轉以及確立風險意識來提高經營效益。

### 3 資料庫取得、更新與品質保證機制

資料庫取得及資料更新機制，除了考慮資料安全外，最重要的是必須和 ELSI 機制有配套措施。要能整合科學研究者的共識，發展各種有價值之研究設計與社會共識獲得的模式。ELSI 要和資訊平台建構時的安全、保密與管理機制，以及

相關的科學運用模型，充分合作，以協助社會共識之建立及符合倫理要求之告知同意（Informed Consent）程序的進行。

集中式資料存取架構，將「生醫資訊整合資料中心」所整合的不同類型資料，分別由不同資料擁有單位完整複製一份進生醫資訊整合資料庫中，這種作法可以使生醫資訊整合資料庫內部系統的執行相較於分散式的整合處理更有效率，然分散式架構的優勢在於能夠迅捷地協同各種不同組織單位的運作，在各自發展與整合中取得平衡。不論何種架構在規劃上都必須考慮營運時的儲存空間需求，不同資料提供單位間的互相信任問題，以及未來資料更新問題。同時應確保資料庫內部倫理機制建立，與其他資料庫（包括醫療紀錄資料庫）關係建立；例如，資料庫間進行交叉比對程序設計，個人隱私權如何維護等等。同時，也應注重如何建立資料庫對外的媒體網路，及如何做到跨國際的合作管道。完整「生醫資訊整合資料中心」系統架構應透過相關委員會等專家小組成員與各單位訪談後，提由專家會議討論，並彙整成完整建議報告書供主管機關決策參考。

### 3.1 實體集中

在集中式架構中，所有的資訊技術（Information Technology, IT）功能、應用軟體的開發和維護，及 IT 的整體運營情況都將直接彙整在中心，所有的 IT 資產，包括軟硬體、人員及技術都由中心直接掌控。從表面看來，集中式的 IT 組織架構無疑是最具吸引力的管理方式，它有許多顯而易見的優勢：軟硬體的配置與控制簡單，且集中式運營也可以帶來實實在在的規模效益，如集中式的資料中心能夠縮減支出 10%~15%(參考 CIO 雜誌，<http://www.cio.com/>)；集中式的管理還可以簡化人員配備，如作業系統支援小組；統一的 IT 結構和標準還可以解決軟體的相容問題和相關費用。總結起來，節約成本和易於管理是兩個最重要的原因，而整合後實體資料庫也可與國際其他資料庫合作。唯以不同的成本計算方式，可能會得到不同的結論，因此建議採開放態度，待所有因素一併考量後再做出最終的建議。

集中式架構最廣泛的應用就是資料倉儲系統，資料倉儲起源於為了解決作業資料在各方匯集後不利於分析的缺點，這個概念始於 1984 到 1988 期間，並以不同的形式在不同的組織發展。在 1988 年 Barry Devlin 與 Paul Murphy 兩位學者首次發表文章定義資料倉儲的架構，他們建議針對組織中各種專業的使用者，另外建立一個唯讀（Read Only）性的資料庫，將作業性資料庫、檔案或歷史資料等整合進來，並提供整合性工具與資料目錄說明，以便進行分析工作。也就是從各種不同的資料來源中獲取資料，以單一、完整及一致性的原則來加以儲存，讓使用者有效地瞭解並使用這些資料。

以集中式資料庫機制來規劃台灣生醫資訊整合資料中心，將現行已建置之癌症登記資料庫、戶籍資料庫、健保資料庫及死因檔等，彙集整合於資料中心，亦可併入憑證模組以提高資料交換時，在技術上與法律(電子簽章法)上的層層保護。集中式資料中心運作可參考圖 5。



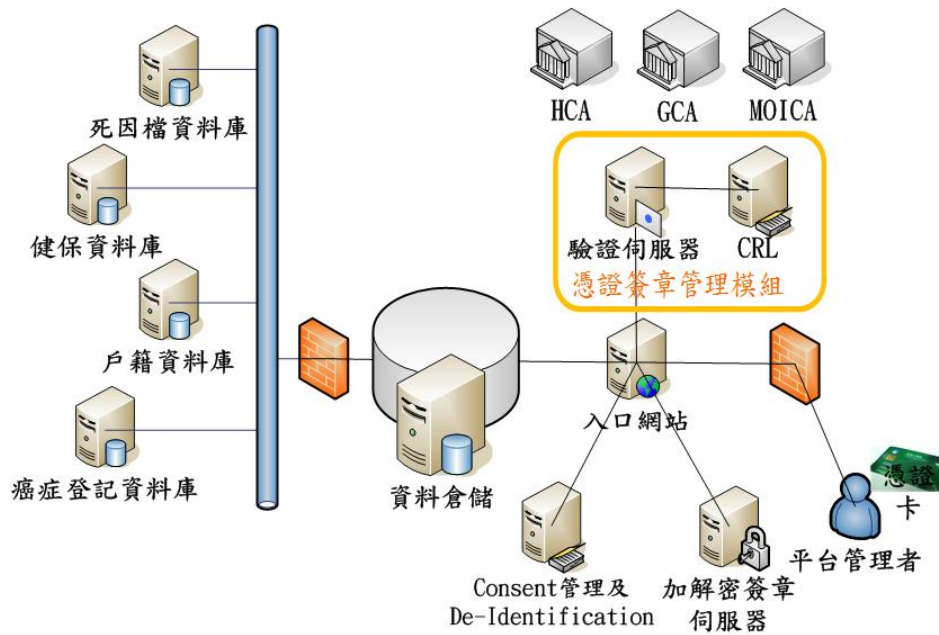


圖 5. 集中式資料中心架構圖

集中式架構在資料中心建置一個資料倉儲，將四個資料庫的資料定期做更新匯入，做為集中統籌的管理與資料對應。當申請者在平台提出資料要求，平台管理者可直接在中心的倉儲中取得資料並且回應給申請者，資料申請流程如下：

1. 申請者提出申請。
2. 平台管理者經過資訊中心憑證驗章確定其管理權限與身份。
3. 平台管理者透過同意書(Consent)管理正確地查詢可用於研究之個案名單後，直接在資料倉儲中取資料。
4. 平台管理者利用平台操作取出資料並經過加密與簽章，利用光碟或其他儲存裝備送交至申請者手中，完成資料交付。

實體集中式資料倉儲的資料同步問題(如圖 6. 資料倉儲架構示意圖)，學術研究上稱為視界維護 (View Maintenance) 問題，視界維護問題是指當資料來源有更新動作時，如何讓資料倉儲中的資料同步更新。由於資料倉儲中的資料並非各資料來源資料的副本而已，而是經過記錄選擇 (Select)、欄位選擇 (Project)、比對連結 (Join)、刪除重複 (Removal of Duplicates)，還有加密 (Encryption) 等動作後所得到的新資料表格，這新資料表格稱為視界 (View)。因此視界的更新維護便是資料倉儲實做上的重要議題。

另一個議題是更新動作的頻率問題。由於本文中討論的四個資料庫整合議題並不會有大量的資料查詢需求，而且資料是否為最新對生醫研究而言並不重要，因此只要定期更新即可。

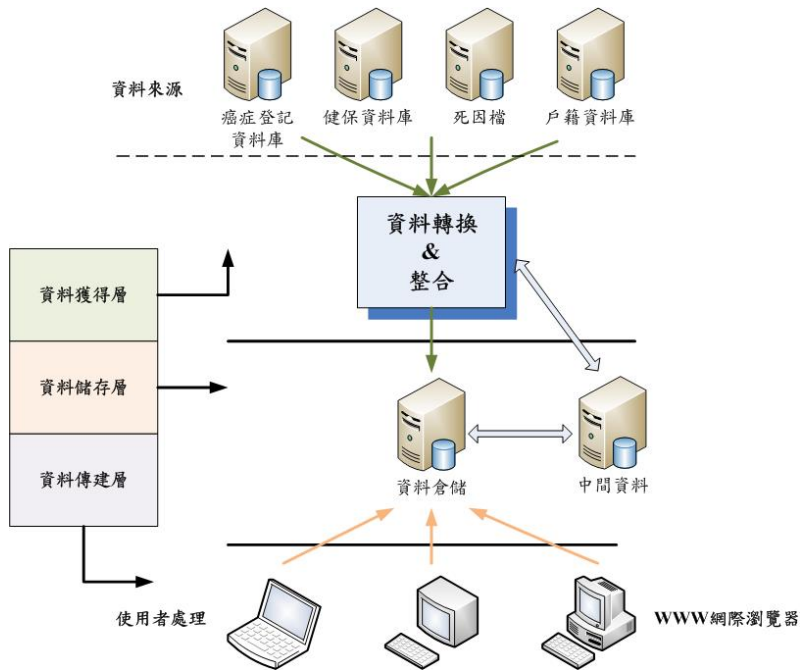


圖 6. 資料倉儲架構示意圖

### 3.2 分散式整合

產業的利基已不再是孤軍奮鬥的時代，想要有效利用龐大的生物醫療資源勢必要整合各單位組織內部與外部資源，在一片資訊化聲浪中，各研究機構以各自的平台建構自己的系統，形成資源多樣性且豐富的環境，在這樣的環境結構下，不同的系統建立、儲存不同的重要資料，因而資料來源多樣且複雜。為了整合而制式化的統一系統在實作上是非常艱難且高成本的，除此之外，統一化做法受限制的不只是各個資訊系統，也阻礙了各機構自由發展的創意。一般而言，若要整合異質系統的最好方法就是讓這些系統一起運作，所謂的一起運作是指，系統本身遵循著標準，提供一些介面或處理機制與外部溝通。分散式系統架構能有效的達成此一需求，建立溝通介面，這個介面越中立越好，而且必須由大家共同配合遵守，以避免因為單方面系統、程式內部的調整而衝擊到大家的溝通。規劃虛擬化儲存的分散式架構可將多種不同格式資料來源合併而呈現出強大的能力，以實現虛擬資料庫的概念。和其它資料庫模型相比，其優點有：(1) 資料存取透明性、(2) 異質整合能力、(3) 高度包容力（容許資料多樣性）、(4) 強化架構延伸力與開放性、(5) 資料來源自主與(6) 效能最佳化等獨特優勢。因此分散式的鬆散藕合非常適合應用於各自獨立發展卻也能相互整合創造利基的生物醫療研究。

本文欲整合四種既有之資料庫，癌症登記資料庫、戶籍資料庫、健保資料庫及死因檔，各有其所屬的管轄單位也各自有其使命與資訊管理系統。分散式邏輯整合的規劃上重點在於建立中立的介面，在技術上採用 HL7(Health Level Seven) 醫療資訊標準做為不同系統間的開放性標準溝通協定，標準的中介協定有助於提升再利用率，利用中間層（Middle-tier）並且使用元件化的模型連接資料中心與各分散節點，使資料整合中心成為一個虛擬的資訊管理中心，分散式資料中心的

運作可參考圖 7。資訊管理中心成為聯繫遠端不同組織單位間的平台，並提供完整的協定參照以協助各個單位建立標準的資料互通介面，擁有架構快速延伸的特性，可以使不同的資料或服務的提供者透過資訊管理中心的運作機制與技術支援在最短的時間內加入平台提供服務。

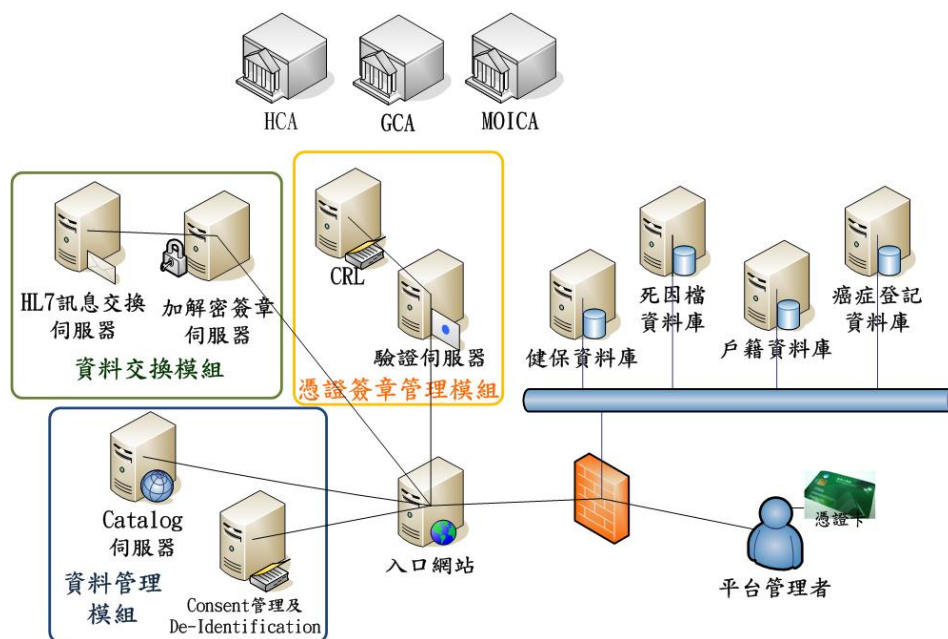


圖 7. 分散式資料中心架構圖

分散式的資料中心強調不儲存實體資料，中心沒有資料的備份只有資料間的對應表，以即時(Real Time)的方式傳遞最新的訊息，在憑證的部分後端即時連接的資料庫以伺服器憑證來相互驗證各個資料庫，以確保資料的正確來源。

分散式架構下，所有的資料交換發生在提出資料需求當下，資料中心在審查過申請者資料並且通過之後，平台管理者以即時的方式透過平台的資訊系統，將申請者的資料需求進行分解處理，透過 HL7 訊息交換閘道，分送訊息請求(Message Request)至各個資料庫，並由各個資料庫各自回應資料並做加密簽章處理，中心收到資料後再將資料以離線儲存設備寄送給申請者。資訊傳遞的流程如下：

1. 申請者提出申請。
2. 平台管理者經過資訊中心透過憑證驗章確定身份，管理者使用平台資訊系統查詢並取出資料；資料管理模組根據申請者所提出的研究資料要求，透過機制組合為查詢語法。
3. 查詢語法送交資料交換模組，轉為 HL7 訊息 (Message)，再經過加解密與簽章，分別傳送至分散的四個資料庫查詢。
4. 返回的資料由四個資料庫端伺服器做加密簽章，透過 SSL 通道送達資料中心。
5. 經過驗章與解密，將 HL7 訊息接收後，傳遞回平台管理者，並由其利用光碟或其他儲存裝備送交至申請者手中，完成資料交付。

資料更新機制的部分由目錄(Catalog)伺服器來管理詮釋資料(Meta Data)，所有的資料庫異動統一在中心的資料表裡記錄。而同意書(Consent)管理伺服器用來管理參與研究或退出研究的個案，當申請資料的要求進入中心後，先確定此筆個案資料是研究的參與者無誤後才進行資料的查詢。

### 3.3 安全性規劃

安全性的規劃上集中式採用標準的公開金鑰基礎架構(Public Key Infrastructure, PKI)為安全性驗證架構，在分散式的架構驗證上網路服務(Web Services)也訂定出標準的操作守則。2004年Oasis-Open組織發表了WS-Security 1.0標準，提供Web Services在溝通協定上重要的安全技術方針，2006年更進一步發表版本1.1稱之為WSS(Windows Sharepoint Services)。WSS在系統運作參與者的鑑別(Authentication)、數位簽章(Digital Signature)以及XML資料加密(Encryption)上做了明確的指示規範，以確保資訊在網路傳輸過程中不易遭受偽造、非法存取、竄改、竊取或截聽該內容，能夠鑑別交易雙方的身分，並確保網路中所傳輸之資料的一致性(Integrity)以及為防止非法使用者得知已保護之敏感資料的內容，確保對數據使用加密方法所得的私密性(Confidentiality)。



圖 8. Web Services 授權憑證

完整的建置參考圖 8，經由訊息送出端的簽章、加密與接收端的驗證與解密完成一次訊息的傳遞，說明如下。

Web services 送出訊息端：

1. 首先由伺服器端傳送驗證訊息，並將伺服器端之憑證送出，此憑證為符合 X.509 規格的單一憑證或憑證鏈，並使用其私鑰做簽章的動作。
2. 以公鑰為訊息做加密，確保只有接受者能夠存取這個訊息。

Web services 訊息接收端：

1. 以訊息送出者的公鑰來鑑別訊息是否確為其交握的對象並且驗證資料的一致性。



2. 以其擁有的符合 X.509 之私鑰對訊息做解密動作。

保護個人身份識別資料，可以採用標準的公開金鑰基礎架構（PKI）進行加解密的作業，但必須嚴密保護私鑰（Private Key）。本文將就各種不同的保護措施進行評比分析，做為未來建置案需求規格之依據。

私鑰的保護措施可以依其保護金鑰的嚴密程度來區分。最鬆散的保護方式是儲存在網路上的電腦。這樣的方式會有遭人入侵竊取的危險。比較嚴密的方式是將金鑰本身再以 PKI 的方式加密。若要求最嚴密的保護方式，則可以考慮使用硬體鎖保護。要竊取金鑰必須先竊取硬體鎖才行。但無論任何硬體鎖，都有失效的可能，因此必須考慮回復（Key Recovery）的問題，也就是在這種狀況下仍能取回金鑰的功能。基本上，採用硬體鎖時，作業流程是一旦產生了一組金鑰，便立刻將金鑰儲存在某一台電腦上，然後立刻加密上鎖。這台電腦必須永遠保持離線。為了防止硬體鎖失效，可以將金鑰以同樣的方式複製在好幾台異地備援用的電腦上。

原則上，個人身份識別資料加密後，除非資料毀損，非修復原始資料否則無法持續進行中的研究，否則不應該有解密的需求。因此也可以考慮加密後立刻清除私鑰。但若允許資料在加密釋出給研究人員使用後，個人仍可以要求被排除在該研究以外，則仍有必要做到解密、查出資料中是否含有要求排除的個人的資料，然後刪除關於此人的資料的動作。這樣一來問題會變得複雜。但若真有這類的需求的話，仍可透過善加保護的金鑰機制解決。

另外，在現行的加解密與憑證核發的機制上，目前各單位皆有其自己提供的憑證與加密演算法，一方面造成系統重覆建置的資源浪費，也造成在資料共享上的無法統一，因而本文建議應評估各種現行機制，朝向健全統一的機制作規劃，未來在建置中應聽取生醫研究人員及資訊安全專家的意見後決定。至於系統傳輸、運算過程中所產生的憑證與金鑰，應交由哪一個組織或單位來負責發放與保管，建議應交由具有公權力之政府單位主管機關保管，未來也將與營運模式及 ELSI 方面的問題同時考慮，並交由專家學者討論，方能提出可行之規劃。

### 3.4 資訊安全政策

另外，亦應規劃一個資料中心的整體資訊安全政策（圖 9）。整體資訊安全流程可以考慮採用 ISO 17799/27001 的標準認證或 COBIT 標準（Control Objective for Information and Related Technology），以確認資料的安全 OCOBIT 係由 IT 治理協會（IT Governance Institute, ITGI）所發行，是關於 IT 安全和控管實務的標準，對管理階層、使用者以及資訊系統稽核、控管和安全的從業人員提供

ISO 17799/27001 則為針對資訊安全架構所制訂的國際標準。此標準源自 BS 7799，是以風險為核心考量訂定的保護措施與制度，目標是確保制度的落實與持續改進。透過從風險的角度來思考，可決定到底該要採用什麼樣的資安措施，並將適當的資源投入在適當的問題上。ISO 17799/27001 的導入方法如圖 10 所示。

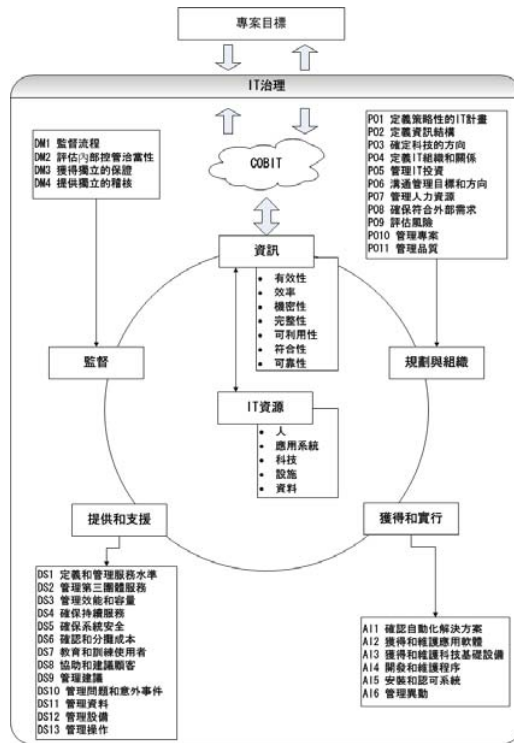


圖 9. 資安政策實作規劃 (來源:參考資料 9)

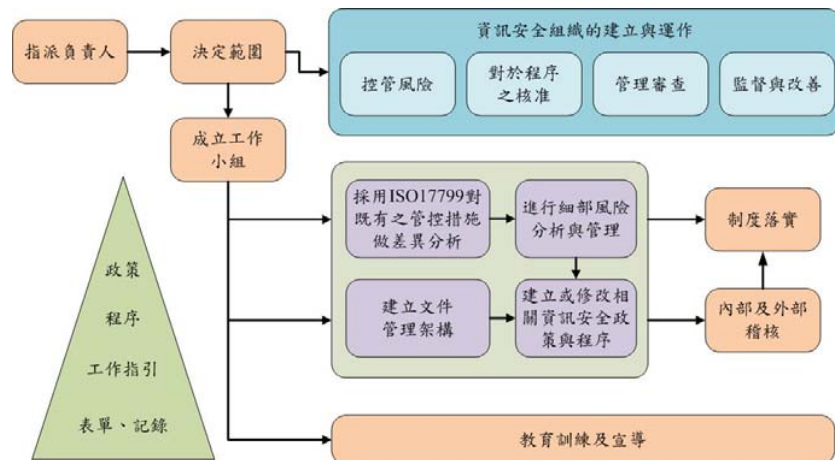


圖 10. 資安政策導入

但導入並取得通過 ISO 認證，雖然並不代表資料中心的資訊安全就萬無一失，但是通過 ISO 17799/27001 應該至少完成以下的規劃目標：

- 在通過認證的範圍內，資料中心有定義相關的資訊安全政策。
- 資料中心曾經考慮過在範圍內可能的資安風險，並且會採用合適的管控措施去管控。
- 資料中心將會針對資安的政策及相關程序做定期的檢討，並且對發生的缺點做出追蹤。

就範圍的選擇上，可以選機房或是和外界接觸最小的範圍，或是以維運為

主。同時必須確定幾個重要的項目都有做到：

- 資訊安全小組，確定該範圍的負責人在小組之內，並且確定每次的運作都有留下會議紀錄。
- 確定目前的程序有考慮到 ISO17799 的控制項目，並且產出適用性聲明。
- 檢查記錄的完整性，確認說、寫、作一致（有完整之管理循環及三個月執行紀錄）。
- 確保稽核及執行人員有足夠的能力，並且有訓練計劃。
- 有執行教育訓練及宣導，並且有相關的記錄。
- 有合適的文件管理程序，確定文件及記錄有經過合適人員的審查。
- 確保有針對相關程序的內部稽核程序，並且著重缺失的改善與追蹤。

ISO 17799/27001 和軟體工程專案控管標準能力成熟度整合模式 (Capability Maturity Model Integration, CMMI) 不同。ISO 17799/27001 沒有深度或等級的概念。未來生醫資料中心應該揭露自己的安全政策，並說明自己是採用什麼樣的手段來確保安全，以便外部的社會大眾能夠瞭解。對於申請資料的研究單位而言，生醫資料中心可以要求各研究單位提供他們的安全政策與防護手段(如提供適用性聲明)，甚至反應在雙方的契約上。而通過認證，表示有一個獨立的第三方去對對方是否遵守安全政策進行審查或稽核。

資訊安全之維護可以在營運單位中規劃設置一個資訊安全處理小組，此小組直屬於營運單位內部的資安稽核小組，下轄各子單位的資訊安全執行小組。除了內部的資安稽核小組外，可以在營運單位外部，另由專家學者社會人士等組成外部稽核單位予以監督。資訊安全處理小組則負責下列事項：

- 設備資產管理
- 網路安全管理
- 實體環境管理
- 系統管理與存取控制
- 資訊安全稽核
- 人員管理與教育訓練

資訊安全處理小組組織架構如圖 11 所示。

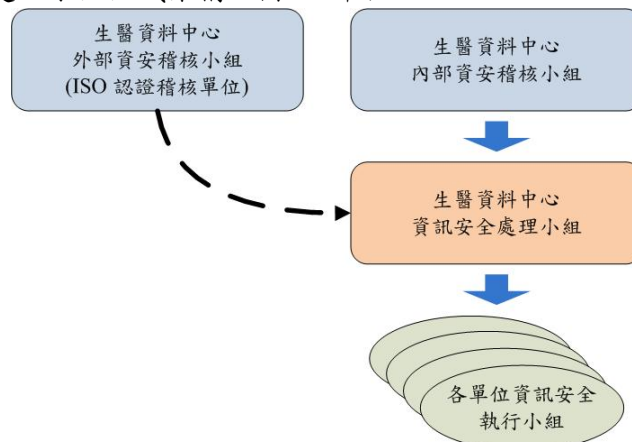


圖 11. 資訊安全處理小組組織架構

### 3.5 匿名化與去識別化 (De-Identification)

去識別化是一種移除或改變資料的程序。在生醫的記錄中有些資料關於識別病患的身份。去識別化的技術在於可以讓研究者或非臨床應用的人能夠在不犧牲病患隱私權的狀況下使用這些醫療資料。去識別化的技術在生物醫療研究中扮演關鍵地位，尤其在 HIPAA (Health Insurance Portability and Accountability Act) 法案通過後更顯得急迫性的需要。去識別化遵循 HIPAA 的規範將樣本資訊隱私部分的關聯去除。

除了直接的病患資訊 (姓名、住址、電話等等) 與編碼之外，去識別化也決定於資料量的大小，去識別化模組以隨機的方式重新組合臨床事件以避免因為資料量小而遭遇的隱私問題。在文字資料 (Free-text) 的部分，則以特定的圖樣 (Pattern) 與辭典比對 (Dictionary Matching) 來處理隱藏在文字資料中的個人隱私資訊。

雖然經過去識別化處理，仍不能排除有可能不需經過解密，卻能回溯推得個人的資料的可能性。尤其是有心人士若要擷取某些特定人士的資料，他們可以透過許多種方式由加密過的資料中直接回溯。目前『有心人士』搜尋特定人士資料的需求頗高，因此必須嚴加防範。例如八卦媒體記者可能會利用此一資料庫對政商界名人，或知名演藝人員進行資料採擷，選舉的競爭對手之間可能會利用本資料庫搜尋對手的健康弱點，一般民眾會透過徵信社利用本資料庫，監視配偶的忠貞度。因此必須針對這些情形加以預防。我們可以針對這些資料採擷的『攻擊』動作歸納為三種方式：

- 利用精密的查詢設計，透過統計上的單一或獨特性，使查詢結果直接暴露特定的個人資料。
- 有心人士可以試圖以分批查詢方式，拼湊出具有獨特性特徵的特定個人資料。
- 有心人士還可以試圖將查詢結果與外來的其他資訊來源拼湊出個人資料。外來的資訊來源如銀行往來、刷卡資料，電話通聯記錄等。

為了避免這類的情形，在規劃去識別化的方法與流程時，可針對以上的情形，針對不同的方法與流程，建置原型系統 (Prototype System) 加以模擬詳見本文 4.2 章節之“模擬簡易實作方案”。

需要注意的是在分散式的運作中，由於各單位的獨立運作會產生不同的資料識別方法，需要考慮比對資料與去識別化之間的平衡成為資訊安全議題中需要規劃考慮的重點。

### 3.6 解決方案初步評估

任何策略都必須建立在對本身與相關環境通徹瞭解的基礎上，建議在規劃上利用管理學的分析模型來幫助我們探討策略的可行性。由於生醫資訊整合資料中心在資訊的收集上涉及眾多議題，不僅僅是技術上的整合能力，還包括在倫理、法律以及社會影響上種種討論，因此在未來的規劃上我們採用 PESTEL 分析模型、結合 SWOT 分析的概念，以評斷採行之策略在各個面向上的優劣。PESTEL



分析模型又稱大環境分析，是分析巨集觀環境的有效工具，不僅能夠分析外部環境，而且能夠識別一切對策略有衝擊作用的力量。它是調查組織外部影響因素的方法，其每一個字母代表一個因素，可以分為六大因素，未來規劃應以此六大因素做為分析的主軸：

- 政治因素 (Political)：

是指對組織經營活動具有實際與潛在影響的政治力量和有關的政策等因素。目前政府的政策正積極推動生醫科技島計畫，此一政策的規劃陸續確立也讓未來在生醫資訊整合資料中心的解決方案規劃上可以有所遵循。

- 經濟因素 (Economic)：

是指組織內外部的經濟結構、產業佈局、資源狀況等。如同台灣生物資料庫 (Taiwan Biobank)、國民健康資訊建設計畫 (NHIP)、卓越臨床實驗與基因體計畫的生醫核心設施計畫，國內有許多產學合作計畫的推動，在外部經濟效益上呈現多樣化共榮的佈局，此一現象使得在整體架構上採用分散式架構略佔優勢，原因在於分散架構的快速整合異質系統與彈性的擴充能力在外部的經濟因素較能體現其效益。

- 社會因素 (Social)：

是指策略導致於在社會中的歷史發展、文化傳統、價值觀念等因素。集中處理生醫資料與分散管理在社會觀感中有著不同的意義，集中式的處理相當程度上意味著資訊管理者的權威與公信力，而分散式的管理強調分權獨立的資料協同運作，然而卻未必能夠滿足資料共享的期待，其間的取捨端看社會觀念的演進與改變。

- 技術因素 (Technological)：

技術要素不僅僅包括那些引起革命性變化的發明，還包括與策略本身有關的新技術的出現和發展趨勢以及應用前景。在技術上集中式的管理有其必然的優勢，相較於分散式系統對於安全性維護機制以及有效驗證的複雜策略，單純的集中管理在技術上是相對的穩固且節約成本。

- 環境因素 (Environmental)：

評估一個策略的活動、營運或服務中能與環境發生相互作用的要素。在建置的環境上分散式的架構有相當程度的困難存在於資料的交流通上，其中透過虛擬管理中心的控制而形成複雜的訊息繞送，在規劃上值得特別注意。

- 法律因素 (Legal)：

包括組織內外部的法律、法規和公民法律意識所組成的綜合系統。以目前的立法而言在訂立資料的保存、管理與使用上可以參見電腦處理個人資料保護法，其中第四條第五款與第十三條第三款，分別規定了依當事人請求與特定目的消失或屆期已滿的處理方式，此項要求在分散式的處理中有其複雜的規則需要訂立；而類似於檢體收集的資料可以參見研究用人體檢體採集與使用注意事項第十二條的規定去識別化與去連

結，資料的比對問題與法規裡的去識別如何在集中式與分散式的技術上取得平衡也是在資訊整合中心執行時討論的重心。

電子簽章法於 2002 年 4 月正式施行，是為國內推動電子化平台的重要法源依據，目的在於使符合程序做成的電子文件、電子簽章能夠取代實體書面與本人辦理簽名蓋章，使得網際網路上的資金要求與傳輸動作能夠被賦予效力。

## 4 可行性研究

本文中針對技術面的可行性評估方法做說明。技術面的評估可由所規劃之系統的能力、效率等方面客觀的評估，在專家會議中也比較容易取得共識。事實上在 ELSI 方面也有許多問題，但由於沒有客觀的標準，有時是依決策者的效益評估，來決定是否推動，可能一面推動，一面解決問題，因此未來實際規劃應著重政策分析。

### 4.1 技術方面之可行性研究

生醫資訊資料中心規劃連繫個人資料與醫療資料之平台，並利用網路科技傳輸與分享資訊，平台中的醫療資訊傳輸通道與安全機制為關鍵核心技術，集中式的資料中心所應用的資料倉儲技術與分散式資料中心架構的即時服務則為不同解決方案中的平台基礎架構。

網際網路上的電子資料傳輸首重交易者的權益與交易的資料內容能夠獲得保障，使得在網路中建立的交流機制能夠充份被信賴，因此，本文規劃運用公開金鑰基礎建設(Public Key Infrastructure；PKI)為平台的基礎，在分散式中心的規劃上更使用伺服器憑證配合 HL7 做醫療訊息的傳遞。電子化交易在商務平台上行之有年，許多的電子商務機制正蓬勃發展，使得相關技術達成成熟階段，政府也大量將 PKI 技術應用在 E 政府平台上提供許多便民服務，並且扶植了許多廠商進行國內的推廣活動，另外在醫療方面亦有衛生署的醫療憑證管理中心(HCA)，為機密性極高的醫療資訊提供較安全的防護。伺服器憑證的成功案例有行政院衛生署疾病管制局在 95 年推動的防疫物資管理資訊系統，讓各家醫院能夠透過伺服器憑證的驗證將資料做自動上傳申報。HL7 除了在國外有許多大型的軟體做為應用之外，在國內的推廣更有著多年的實作經驗，行政院衛生署在過去幾年大力的促成醫療院所間使用 HL7 交換資訊，促成許多醫療相關機構完成 HL7 溝通平台基礎建設，96 年度更有醫療資訊標準推動計畫-HL7 訊息驗證 / LOINC 試辦醫院，因而培育了許多國內 HL7 的人才。

在資訊平台的架構上，集中式的資料倉儲在商業界被廣泛應用於銷售與客戶資料分析，過去已經有許多研究成果並且已經內建到資料庫管理系統產品中，市場上銷售資料倉儲解決方案的公司有 ORACLE、IBM 與 Microsoft 等。而分散式的架構，在資訊應用的歷史上早已存在數十年，目前 IT 大廠如 IBM、Microsoft 與自由軟體 Red Hat 推出的平台裡都紛紛強調 SOA(Service-oriented architecture) 的架構與精神，並提供許多工具與軟硬體支援套件，可以加速開發與縮短即時交易的時間。在資訊的加解密部分，技術上除演算法持續創新之外，資訊設備也不

斷推陳出新，為了因應更大量的加解密運算，有許多現有的硬體加密加速器產品可以使用，使得我們能夠克服加解密技術在分散式資料中心裡即時做龐大資料運算的需求障礙。

本資料中心所規劃使用的技術皆應已成熟且完善，各項技術皆有實作前例，除了評估不同方案在技術上的成熟穩定外，也要根據軟硬體建置成本、維運成本、系統穩定度、系統擴充性、資訊安全性等等來做為評估的依據。

## 4.2 模擬簡易實作方案

規劃生醫資訊整合中心除了應提出可行方法的質性評估之外，對於整個營運方式的實際執行時的可行性量化支持也相當重要，因此在規劃進行中應一併提出模擬實作，來驗證規劃系統的可行性。主要在驗證查詢結果可以符合正確性，且能偵測出 ELSI 議題中合理與不合理查詢，並做出相對應的反應動作，當使用者的查詢或動作的執行可能造成爭議，系統也可以正確無誤的拒絕使用者的不合理請求。工作目標如下：

- 驗證規劃系統能完成四大資料庫的比對與連結
- 驗證規劃系統能確保查詢結果的正確性
- 驗證規劃系統的查詢結果無法回溯個人識別資訊，及其他 ELSI 專家所提出的需求

具體的做法上，應設計至少 14 題模擬查詢需求：

- 7 題是『ELSI 合理』的研究用查詢需求
  - ◆ 6 題是連結 2 個資料庫的查詢。
  - ◆ 1 題連結戶政與另外 2 個資料庫。
- 7 題是『ELSI 不合理』的查詢需求
  - ◆ 3 題是查詢結果會產生單一 (Unique) 的特徵個人資料。
  - ◆ 2 題是模擬有心人士試圖以分批查詢方式，拼湊出單一的特徵的個人資料。
  - ◆ 2 題是模擬有心人士試圖將查詢結果與外來的其他資訊來源拼湊出個人資料。

## 4.3 提供各種生醫應用情境

以下用一個癌症的例子，更具體的描述研究學者如何利用整合資料中心，來完成從前必須花很多時間的研究。在未來，推廣單位應提供其他的例子，協助營運中心宣傳，以增加其使用率。

2006 年，美國學者 Arnold Levine 發現一個人類 MDM2 基因上的單核苷酸變異 (SNP) 和數種癌症的發生有關。這一個單核苷酸變異是發生在 MDM2 基因上的 309 位置，所以被簡稱為 SNP309。圖 12 是目前學者們相信 SNP309 和癌症形成的關係。

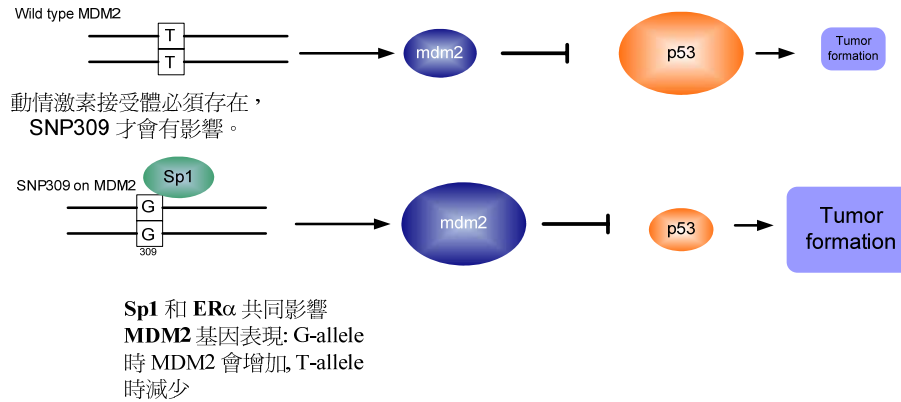


圖 12. 應用情境範例

在動情激素接受體（Estrogen Receptor）反應路徑的病人身上，如果她 SNP309 的基因型是 G/G 的話，蛋白質 Sp1 會比較容易結合到 MDM2 基因上，進而造成比較多量的 mdm2 蛋白質的生成。過量的 MDM2 蛋白質會減少癌症抑制基因 p53 的表現。而當細胞受到外來刺激的時候，不足的 p53 癌症抑制蛋白質容易讓腫瘤提早發生。所以簡單來講，在動情激素接受體反應路徑被活化的人身上，G/G 基因型容易造成腫瘤的形成。另外，文獻記載對於停經後以動情激素治療停經症候群的婦女，如果她們的 SNP309 是 G/G 型，她們有較高的機會得到癌症。（Cancer Res. 2006, 66, 5104）。

如果想要研究國內以賀爾蒙治療停經症候群的婦女的得到癌症的機會是不是大於沒有以賀爾蒙治療的婦女，我們可以這樣著手。首先，從全民健保資料庫裡取出 50 歲以上並有賀爾蒙治療病史的婦女。然後從癌症登記資料庫裡找出有她們之中有多少得到乳癌，從衛生署死因檔裡找出有多少已經因此癌症而死亡。這是相當直接的流行病學研究。進入到分子醫學的層次，按照我們的假設，並不是所有的以動情激素治療的人都容易得到癌症，只有那些 SNP309 的基因型是 G/G 的人才會有較高的機會得癌症。當然沒有基因分型的實驗資料，我們不可能知道哪些人擁有 G/G 的基因型，哪些人是 G/T，哪些人是 T/T。不過我們可以做的是，從內政部戶籍資料庫裡擷取這些案例的直系親屬或姊妹，分析這些人在健保資料和癌症登記資料中得到癌症的比例。對那些被賀爾蒙治療而又得到癌症的婦女，她們的 SNP309 基因型比較有可能是 G/G，然後我們就可以看她們的母親或姊妹，在還沒有停經的階段，是不是有較高的可能性得到癌症。如果是的話，表示 G/G 的基因型可能真的對癌症發病年齡有正向相關。以這個例子來說，我們可以明顯看出生醫資料庫共享對複雜疾病研究的價值。

如果未來我們能更進一步把台灣人疾病及基因資料庫也整合進入資料中心，研究的廣度就更大了。結合臨床資料與基因研究，我們可以把台灣人疾病及基因資料庫的基因型資料拿來，結合內政部戶籍資料庫，看看 SNP309 的 G/G 基因型到底對於乳癌（或其他文獻已有記載的癌症，例如肝癌）的發病有無直接的影響。

## 5 生醫資訊整合資料中心未來遠程規劃

科學園區的成功，曾是我國經濟迅速發展的重要關鍵，它主要是協助產業在研發的最後階段，將研究知識轉變成商品或技術，所以是一種「溫室養成模式」園區。對於技術性高的產業，需要吸引研究型人才加入，所以衍生出第二種將研究成果產業化的趨勢，可稱為「校園創業模式」。在發展生醫產業時，則需與醫學中心結合，提供「轉譯醫學應用」、「產品驗證」與「臨床試驗」的介面與平台，所以產生「認證服務模式」的生醫園區觀念。資訊整合的力量在各種產業已逐漸浮現，雖然紅極一時的電子商務因為推動的時間不恰當而泡沫化，但跨國性的公司，必須靠做好「知識管理」，才有競爭力。同樣的，台灣的生醫產業想要站穩腳步，走上國際，就必須先建立「中樞神經系統」，才能因應瞬息萬變的市場需求，抓住商機。

日本理化學研究所(RIKEN)在橫濱的生醫科學園區內新成立發育生物學研究所，在它旁邊就有生物醫學研究與創新研究所(Institute of Biomedical Research and Innovation, IBRI)，有助於將發育生物學研究應用到生物醫學上，更有「轉譯資訊中心(Translational Research Information Center)」，有助於將醫學研究落實到臨床實驗上。此外更有國際商務中心(International Business Center)在側，可將成果推廣。這種規劃讓每一個研究所只需要做好自己的本業，然後與其他研究所保持互動，即可產生協同增益(synergy)的效果。

許多跨國性的大型製藥公司已逐漸轉型為整合性製藥公司 (FIPCO – Fully Integrated Pharmaceutical Company)。其特色是將研發業務是外包給小型的生技公司、大學或醫院，在 1995 年外包比率約 25%，這個比率正在逐年增加，到 2005 年已高達近 50%左右。資訊系統在整合的過程中扮演著重要的角色。我國若要做藥物研發，也必須有好的資訊系統，才具有國際競爭力。

以臺灣有限的資源，若能將重點放在如何解讀與應用這些資訊，或可收迎頭趕上之效。將「基因型」數據，和「表現型」數據結合，將有助於將基因體醫學研究的成果，轉譯為臨床上可用的治療方法或藥物。近年來開發的新藥數目已有減少的趨勢，因此如何開發舊藥的新用途，將更為重要。此外，為充份發揮藥效，也逐漸要步入「個人化藥物」的境地。因此如何充份利用資訊，協助開發舊藥的新用途，或找到治病的藥物，將是最具競爭力的策略。在另一方面，發展「預防性醫學」，則可降低健保的支出，亦是資訊可以增加競爭力的另一方式。

## 5.1 資料來源

目前基因體國家型計畫與生醫科技島計畫已產生，或將產生大量的資訊，若能將這些資訊整合在一起，國內的生醫研發人員與生醫產業將獲得一個威力超強的武器，讓我們的國際競爭力迅速提昇，而達到「迎頭趕上」的效果。在願景規劃上，可以針對下列資料來源作訪談，整理出可供分享的生醫資料，格式標準化需求、可能的整合應用情境與整合計畫。在時程上應以台灣生物資料庫為優先整合的對象。

1. 臨床試驗中心
2. 台灣生物基因庫

3. 臨床試驗合作網路
4. 國民健康資訊計畫
5. 基因體醫學國家型計畫之生物資訊核心

## 5.2 規劃方式

生醫資訊整合中心規劃之計畫主持人層級應該與各單位取得合作同意書，並瞭解合作之方向，然後由計畫主持人與執行專案之專案經理合擬訓練計畫，並分三組人員去瞭解各單位可提供的生醫資料，資料的儲存方式，與格式標準化需求。此外也應瞭解各資料提供者希望由整合資料中心得到之回饋。在訪談完各單位後，主持人應邀請少數願意對本計畫付出比較多心力的專家一起規劃。其結論應成為在專家會議時，專案經理之報告內容。在資訊整合中心規劃開始後召開長程遠景規劃會議，廣徵學者意見，再據其結論，撰寫規劃書。

## 6 參考資料

1. Bond, G.L., Hirshfield, K.M., Kirchoff, T., Alexe, G., Bond, E.E., Robins, H., Bartel, F., Taubert, H., Wuerl, P., Hait, W., Toppmeyer, D., Offit, K., and Levine, A.J. (2006) MDM2 SNP309 accelerates tumor formation in a gender-specific and hormone-dependent manner. *Cancer Res.* 66:5104-10.
2. Bond, G.L. and Levine, A.J. (2007) A single nucleotide polymorphism in the p53 pathway interacts with gender, environmental stresses and tumor genetics to influence cancer in humans. *Oncogene.* 26:1317-23.
3. caBIG fact sheet, National Cancer Institute, USA, <http://biospecimens.cancer.gov/biospecimen/initiatives/caBIG/fact/>
4. Calder, A. and Watkins, S. (2006) *International IT Governance: An Executive Guide to ISO 17799/ISO 27001* (London, UK: Kogan Page Ltd)
5. CIO Magazine, International Data Group, <http://www.cio.com/>
6. PESTEL Analysis, <http://www.improvementnetwork.gov.uk/>
7. Project Management Institute. (2004) *A Guide to the Project Management Body of Knowledge (PMBOK Guide)* (Project Management Institute, Inc.)
8. 經濟部商業司 (2005) 2005 台灣 PKI 年鑑 (台灣: 經濟部)
9. 國際電腦稽核協會 (ISACA) (2003) *COBIT 資訊及相關技術的管理控制與稽核* (台灣: 中華民國電腦稽核協會)
10. 全民健保研究資料庫, 國家衛生研究院, <http://www.nhri.org.tw/nhird/index.php/>