

95-133-3301
MOTC-IOT-94-SDB004

道路交通事故相關資料整合系統 雛形建置研究(II)- 資料與系統擴充(2/2)



交通部運輸研究所

中華民國 95 年 9 月

95-133-3301
MOTC-IOT-94-SDB004

道路交通事故相關資料整合系統 雛形建置研究(II)- 資料與系統擴充(2/2)

著者：鄭銘章、董基良、陳菀蕙、蔡興國、黃維信、許峻嘉
陳一昌、張開國、賴靜慧

交通部運輸研究所

中華民國 95 年 9 月

國家圖書館出版品預行編目資料

道路交通事故相關資料整合系統雛形建置研究.
II, 資料與系統擴充(2/2) / 鄭銘章等著. --
初版. -- 臺北市 : 交通部運研所, 民95
面 ; 公分
參考書目:面
ISBN 978-986-00-6952-5(平裝)

1. 交通事故 - 資料庫

557.15029

95020456

道路交通事故相關資料整合系統雛形建置研究(II)-資料與系統擴充(2/2)
著者: 鄭銘章、董基良、陳苑蕙、蔡興國、黃維信、許峻嘉、陳一昌、
張開國、賴靜慧

出版機關: 交通部運輸研究所

地址: 臺北市敦化北路 240 號

網址: www.iot.gov.tw (中文版>圖書服務>本所出版品)

電話: (02)23496789

出版年月: 中華民國 95 年 9 月

印刷者: 良機事務機器有限公司

版(刷)次冊數: 初版一刷 130 冊

本書同時登載於交通部運輸研究所網站

定價: 200 元

展售處:

交通部運輸研究所運輸資訊組・電話: (02)23496880

國家書坊台視總店: 臺北市八德路 3 段 10 號 B1・電話: (02)25781515

五南文化廣場: 臺中市中山路 6 號・電話: (04)22260330

GPN: 1009502601

ISBN(10 碼): 986-00-6952-2 (平裝)

ISBN(13 碼): 978-986-00-6952-5 (平裝)

著作財產權人: 中華民國(代表機關: 交通部運輸研究所)

本著作保留所有權利, 欲利用本著作全部或部分內容者, 須徵求交通部運輸研究所書面授權。

交通部運輸研究所合作研究計畫出版品摘要表

出版品名稱：道路交通事故相關資料整合系統雛形建置研究(II) - 資料與系統擴充(2/2)			
國際標準書號(或叢刊號) ISBN 986-00-6952-2 (平裝)	政府出版品統一編號 1009502601	運輸研究所出版品編號 95-133-3301	計畫編號 94-SDB004
本所主辦單位：運輸安全組 主管：陳一昌 計畫主持人：陳一昌 研究人員：張開國、賴靜慧 聯絡電話：02-23496861 傳真號碼：02-25450429	合作研究單位：國立中央大學 計畫主持人：鄭銘章 研究人員：董基良、陳苑蕙、蔡興國、黃維信、許峻嘉 地址：桃園縣中壢市中大路 300 號 聯絡電話：03-4228550		研究期間 自 94 年 2 月 至 94 年 11 月
關鍵詞：資料庫連結、事故資料、交通安全			
摘要： <p>本計畫為 2 年期(2004-2005)計畫的第 2 年，今(2005)年計畫除持續更新道路交通事故相關資料整合系統與資料，以及改善「事故關聯資料分析與評估系統」外，也依據 2004 年的研究成果及建議，以 1999-2002 年的道路交通事故資料所連結出的資料為主要研究範圍，進行 Yes-chain 與 No-chain 資料的應用；包括：(1)建立供專業人士使用的網頁資料查詢系統；(2)利用本計畫資料的特性，探討連結率改善課題；(3)支援本所運輸安全組的其它合作計畫，如先進安全車等；(4)資料探勘技術之應用潛力初探。</p>			
出版日期	頁數	定價	本 出 版 品 取 得 方 式
95 年 9 月	184	200	凡屬機密性出版品均不對外公開。普通性出版品，公營、公益機關團體及學校可函洽本所免費贈閱；私人及私營機關團體可按定價價購。
機密等級： <input type="checkbox"/> 密 <input type="checkbox"/> 機密 <input type="checkbox"/> 極機密 <input type="checkbox"/> 絕對機密 (解密條件： <input type="checkbox"/> 年 <input type="checkbox"/> 月 <input type="checkbox"/> 日解密， <input type="checkbox"/> 公布後解密， <input type="checkbox"/> 附件抽存後解密， <input type="checkbox"/> 工作完成或會議終了時解密， <input type="checkbox"/> 另行檢討後辦理解密) <input checked="" type="checkbox"/> 普通			
備註：本研究之結論與建議不代表交通部之意見。			

**PUBLICATION ABSTRACTS OF RESEARCH PROJECTS
INSTITUTE OF TRANSPORTATION
MINISTRY OF TRANSPORTATION AND COMMUNICATIONS**

TITLE: Prototype System of Integrated Road Traffic Accident Data (II)-Data and System Expansion (2/2)			
ISBN(OR ISSN) ISBN 986-00-6952-2(pbk.)	GOVERNMENT PUBLICATIONS NUMBER 1009502601	IOT SERIAL NUMBER 95-133-3301	PROJECT NUMBER 94-SDB004
DIVISION: Safety Division DIVISION CHIEF: Chen, Isaac I. C. PRINCIPAL INVESTIGATOR: Chen, Isaac I. C. PROJECT STAFF: Chang, Kaikuo; Lai, Jing Huei PHONE: 886-2-23496861 FAX: 886-2-25450429			PROJECT PERIOD FROM Feb. 2005 TO Nov. 2005
RESEARCH AGENCY: National Central University PRINCIPAL INVESTIGATOR: Jeng, Ming-Jang PROJECT STAFF: Doong, Ji-Liang; Chen, Wan-Hui; Chai, Sin-Kuo; Huang, Wei-Shin; Hsu, Chun-Chia ADDRESS: No.300, Zhongda Rd., Zhongli City, Taoyuan County 32001, Taiwan, R.O.C. PHONE: 886-3-4228550			
KEY WORDS: Database Linkage, Accident Data, Road Safety			
ABSTRACT: <p style="margin: 10px 0;">This report summarizes the second year work of a two-year (2004-2005) project. In addition to updating data and improving the "Analysis and Evaluation System of Events Link Data", data during 1999-2002 from Yes-Chain and No-Chain is used in the application fields according to the results and suggestions from the study of 2004. The major applications are as follows. Firstly, a web query system for specialists is built. Secondly, a tailored method of improving linkage rate is examined. Thirdly, a set of required data is provided to support other projects, for example, "Advanced Safety Vehicles." Finally, the potential of applying data mining techniques is explored.</p>			
DATE OF PUBLICATION September 2006	NUMBER OF PAGES 184	PRICE 200	CLASSIFICATION <input type="checkbox"/> RESTRICTED <input type="checkbox"/> CONFIDENTIAL <input type="checkbox"/> SECRET <input type="checkbox"/> TOP SECRET <input checked="" type="checkbox"/> UNCLASSIFIED
The views expressed in this publication are not necessarily those of the Ministry of Transportation and Communications.			

目錄

第一章 緒論.....	1-1
1.1 計畫背景分析.....	1-1
1.2 研究範圍與對象.....	1-2
1.3 研究內容與工作項目.....	1-3
第二章 文獻回顧.....	2-1
2.1 機率連結理論與發展.....	2-1
2.2 線上資料查詢系統.....	2-6
2.2.1 國內網站.....	2-6
2.2.2 國外網站.....	2-9
2.3 資料探勘技術.....	2-21
2.4 損傷之長期結果指標.....	2-24
2.4.1 生活品質量表.....	2-24
2.4.2 美國 NHTSA 的生活功能指標.....	2-25
第三章 道路交通事故相關資料庫.....	3-1
3.1 道路交通事故資料庫.....	3-1
3.2 死因資料庫.....	3-5
3.3 車輛保險資料庫.....	3-6
3.3.1 88-91 年車輛保險資料.....	3-9
3.3.2 92-93 年車輛保險資料.....	3-10
3.4 健保資料庫.....	3-11
3.5 人車監理違規資料庫.....	3-12
3.5.1 資料整理.....	3-13
3.5.2 資料淨化.....	3-15
3.5.3 資料整併.....	3-19
第四章 道路交通事故相關資料整合系統之建置.....	4-1
4.1 資訊系統的建置.....	4-1
4.1.1 硬體系統.....	4-1
4.1.2 軟體系統.....	4-3

4.1.3 資訊安全系統.....	4-4
4.2 「資料維護整合系統」的開發.....	4-7
4.2.1 資料匯入.....	4-8
4.2.2 資料淨化.....	4-10
4.2.3 資料整併.....	4-12
4.2.4 其他資料處理功能.....	4-13
4.3 「事故關聯資料分析與評估系統」的開發.....	4-14
4.3.1 系統開發工具.....	4-14
4.3.2 資料編碼系統.....	4-15
4.3.3 系統功能及特色.....	4-17
4.4 「線上查詢系統」的開發.....	4-23
第五章 系統連結結果的應用.....	5-1
5.1 被淨化資料之連結探討.....	5-1
5.2 強制險資料連結探討.....	5-6
5.2.1 事故資料與強制險已決賠資料連結探討.....	5-7
5.2.2 強制險受害人資料連結之探討.....	5-11
5.2.3 事故資料、強制險已決賠資料與健保資料之連結探討.....	5-13
5.3 與先進安全車輛系統之結合研究.....	5-15
5.3.1 事故後第一次住院.....	5-15
5.3.2 事故後 30 天內死亡.....	5-16
5.4 資料探勘技術之未來應用潛力.....	5-18
5.4.1 CHAID 預測模式之探討.....	5-18
5.4.2 CART 預測模式之探討.....	5-22
5.4.3 倒傳遞類神經網路預測模式之探討.....	5-23
第六章 結論及未來建議.....	6-1
6.1 結論.....	6-1
6.2 未來建議.....	6-4
參考文獻.....	參-1
附錄 1、事故資料新舊欄位對照表.....	附-1
附錄 2、車輛保險賠案資料欄位說明.....	附-13

附錄 3、人車監理違規資料欄位說明.....	附-18
附錄 4、人車監理違規資料之專家訪談紀錄.....	附-19
附錄 5、期中審查意見處理情形.....	附-21
附錄 6、期末審查意見處理情形.....	附-28
附錄 7、工作會議紀錄.....	附-32
附錄 8、簡報資料.....	附-40

表目錄

表 2.1	機率連結軟體整理介紹.....	2-4
表 2.1	機率連結軟體整理介紹(續).....	2-5
表 2.2	資料探勘功能定義與其運用之技術.....	2-22
表 3.1	本計畫所使用的資料詳細列表.....	3-2
表 3.2	研究用各年度事故資料原始筆數.....	3-3
表 3.3	事故各年度資料淨化過程筆數刪減統計彙整表.....	3-4
表 3.4	研究用各年度死因資料原始筆數.....	3-6
表 3.5	死因各年度資料淨化過程筆數刪減統計彙整表.....	3-6
表 3.6	各年度車險原始資料筆數.....	3-7
表 3.7	強制險受害人各年度資料淨化過程筆數刪減統計彙整表.....	3-7
表 3.8	強制險各年度資料淨化過程筆數刪減統計彙整表.....	3-8
表 3.9	任意險各年度資料淨化過程筆數刪減統計彙整表.....	3-9
表 3.10	目前已取得之健保資料筆數.....	3-12
表 3.11	目前所取得人車監理違規資料範圍.....	3-13
表 3.12	人車監理違規資料各工作區資料筆數統計.....	3-14
表 3.13	人車監理違規資料淨化記錄表.....	3-17
表 3.13	人車監理違規資料淨化記錄表(續).....	3-18
表 3.14	違規唯一碼出現次數統計表.....	3-20
表 3.15	結案代碼分類整理.....	3-21
表 3.16	整併前後結案代碼筆數之變化.....	3-22
表 4.1	系統建置歷程說明.....	4-2
表 5.1	90 年 PO 與 DA 資料配合其他欄位之比較結果.....	5-2
表 5.2	身分證號同、出生年月日不同的資料比對分佈.....	5-3
表 5.3	身分證號同、出生年月日不同的資料主要形態統計.....	5-3
表 5.4	民國 90 年 PO 資料中身分證號空白其他欄位資料筆數統計	5-4
表 5.5	PO 中身分證號空白者，利用其他欄位連結民國 90 年 DA 資料之結果	5-5
表 5.6	PO 及 DA 身分證號檢查碼錯誤者，以身分證號前 9 碼重新連結的結果	5-6

表 5.7 PO_VI 身分證相同之連結	5-8
表 5.8 PO_VI 身分證號且事故日與出險日相同之連結	5-8
表 5.9 PO_VI 車牌號碼相同之連結	5-9
表 5.10 PO_VI 車牌號碼且事故日與出險日相同之連結	5-9
表 5.11 PO_VI 身分證號相同或車牌號碼相同之連結	5-10
表 5.12 PO_VI 身分證號或車牌號碼相同，但事故日與出險日相同之連結 ..	5-10
表 5.13 PO_VT 身分證相同之連結	5-11
表 5.14 PO_VT 身分證相同且事故日與出險日相同之連結	5-11
表 5.15 VI_VT 身分證號相同之連結	5-12
表 5.16 VI_VT 身分證號相同且事故日相同之連結	5-12
表 5.17 強制險已決賠與健保連結用資料筆數.....	5-13
表 5.18 強制險已決賠肇事人出險後就醫人數.....	5-14
表 5.19 強制險已決賠肇事人出險當天就醫人數.....	5-14
表 5.20 事故當事人事故後就醫人數.....	5-14
表 5.21 事故當事人事故當天就醫人數.....	5-14
表 5.22 小客車在不同撞擊部位的事故後第一次住院的醫療成本分析.....	5-16
表 5.23 為小客車駕駛發生死亡交通事故的主要受傷部位(AIS)次數分配	5-17
表 5.24 BPNN 模型	5-25
表 5.25 BPNN 模型中預測變數之重要性指標	5-26

圖目錄

圖 2-1	經濟統計資訊網路查詢系統功能畫面	2-6
圖 2-2	統計資訊網統計圖表	2-7
圖 2-3	衛生統計資訊網統計資料選單頁	2-8
圖 2-4	衛生統計資訊網統計資料結果列表頁	2-8
圖 2-5	警政署線上統計資料查詢畫面	2-9
圖 2-6	IRTAD 文字敘述性之報告資料.....	2-10
圖 2-7	IRTAD 圖形統計圖表式統計資料.....	2-10
圖 2-8	東京都警視廳交通事故發生統計圖	2-11
圖 2-9	設定查詢資料年份	2-11
圖 2-10	設定變數條件	2-12
圖 2-11	選擇資料年期及類別	2-12
圖 2-12	選擇分析變數項目	2-13
圖 2-13	顯示統計長條圖	2-13
圖 2-14	FMCSA 線上分析網站.....	2-14
圖 2-15	Crash Profiles Online 分項比較項目	2-14
圖 2-16	Crash Profiles Online 分項歷年資料比較圖.....	2-15
圖 2-17	查詢條件設定畫面	2-16
圖 2-18	查詢結果畫面	2-16
圖 2-19	制式統計表格資料	2-17
圖 2-20	查詢表格細分項設定	2-17
圖 2-21	CODES 研究主題資料下載畫面	2-18
圖 2-22	指定欄位查詢結果畫面	2-19
圖 2-23	WISQARS 查詢變數設定畫面	2-20
圖 2-24	WISQARS 查詢結果顯示畫面	2-20
圖 4-1	系統資訊架構圖	4-6
圖 4-2	資料庫維護管理系統主功能畫面	4-7
圖 4-3	事故資料庫處理選項功能畫面	4-8
圖 4-4	死因資料庫處理選項功能畫面	4-8

圖 4-5	事故資料匯入功能畫面	4-9
圖 4-6	純文字檔資料匯入功能畫面	4-9
圖 4-7	身分證號篩選功能畫面	4-10
圖 4-8	車牌號碼篩選功能畫面	4-11
圖 4-9	日期處理功能畫面	4-11
圖 4-10	資料孤兒處理畫面	4-12
圖 4-11	ICDMAP 資料格式轉換工具	4-13
圖 4-12	資料庫連結參數設定功能畫面	4-14
圖 4-13	資料連結設定層畫面	4-18
圖 4-14	事故與死因連結流程圖	4-19
圖 4-15	事故與車輛強制險連結流程圖	4-19
圖 4-16	事故與人車監理違規資料連結流程圖	4-20
圖 4-17	事故資料時間框及資料連結設定層	4-20
圖 4-18	門診與住院資料關聯層	4-21
圖 4-19	住院就醫特性連結流程圖	4-22
圖 4-20	西醫門診就醫特性連結流程圖	4-22
圖 4-21	輸出欄位設定功能畫面	4-23
圖 4-22	資料輸出功能畫面	4-23
圖 4-23	三層式架構示意圖	4-24
圖 4-24	變數選擇畫面	4-25
圖 4-25	線上查詢結果	4-26
圖 5-1	事故後第一次住院不同車種的撞擊部位次數百分比分佈	5-16
圖 5-2	不同車種的碰撞部位百分比分佈	5-17
圖 5-3	CHAID 決策樹 (Decision Tree)	5-21
圖 5-4	CART 決策樹 (Decision Tree)	5-22
圖 5-5	BPNN 傳遞模式	5-23
圖 5-6	BPNN 訓練模型	5-24
圖 5-7	激發函數圖	5-24

第一章 緒論

1.1 計畫背景分析

運輸是民眾每天生活中重要的一環，由交通所造成的各種問題，如：噪音、污染、車禍等，已被視為健康問題的一部份。若採用世界銀行以轉型期國家的國家生產毛額(Gross National Product, GNP)的 1.5%來估算該國的車禍成本，則台灣 2004 年的 GNP 為 3,167 億美元，每年因車禍所產生的成本約 47.5 億美元，折合新台幣約為 1,568 億元。由於車禍所產生的鉅額損失，以及車禍受傷而衍生的復健、慢性疾病、長期照護等相關長期問題，佔社會負擔中相當大的比例。故世界衛生組織(World Health Organization, WHO)在 2004 年便以交通安全為主題，結合各種國際活動，希望各國政府能重視這個議題。

WHO 在 2004 年終所出版的報告[20]中指出，要改善交通安全，減輕因車禍所產生的傷亡結果，基本車禍相關資料的收集是很重要的，但因車禍資料通常在不同處理階段由不同單位負責，如：車禍現場由警察處理、傷害的治療在醫院進行、傷害或車輛損失的賠償由保險公司負責等。因此，需利用資料連結的方式，來串連分散於各處的資料，連結後的結果對於預防及減輕車禍傷害是相當有幫助的。美國的 CODES[6,7]及 CIREN[15,16]便是串連發生車禍的車輛、醫療、環境等有關資料，提供給車廠、醫生、運輸等研究人員進行傷害研究，希望能由人、車、路及環境等層面，來進行車禍傷害的防治工作。

本所自 1999 年[22]開始嘗試將台灣與車禍有關的資料串連起來，包括：警察單位的道路交通事故資料、衛生單位的死因及健保資料、財政單位的車輛保險資料。2003 年本所再進一步建立連結這些車禍相關資料庫的雛形系統[23]，2004 年便以這些既有基礎，推動為期 2 年(2004-2005)的系統及資料擴充計畫。在 2004 年的計畫中[24]，除了證明以 ICDMap 軟體可快速將傷害資料轉成國際常用的 AIS(Abbreviated Injury Scale)及 IIS(Injury Impairment Scale)嚴重度指標外，在資料的應用層面上，除了支援本所內部其他的研究計畫外[25-27,28]，也提供所外其他學術研究人員的合外研究上[29]，證明資料確實可以支援本所內部、外部與運輸安全課題相關的研究。此外，2004 年時更進一步整合了人車監理違規資料，使預防車禍傷害的研究，具有由車禍時、車禍後拓展至車禍前的分析潛力。歷經

2003-2004 年 2 年的研究，目前已有多篇因應計畫要求的研究成果發表在研討會中[30-37]。

由 2004 年計畫執行的經驗中，也發現以確定性連結(Deterministic Linkage)方式雖可較確定所串連的資料同屬一人，但相對地亦使得某些資料庫的連結率，因連結鍵值的不全而降低，如：車輛保險資料與道路交通事故資料。綜整這些研究成果與經驗，2004 年的計畫中，將各式連結資料區分成 2 部分，即 Yes-chain(所有判斷條件均為 yes 的資料)與 No-chain(有任何一項判斷條件為 no 的資料)，詳細區分內容可參閱 2004 年報告書[24]。2004 年計畫並規劃下列建議：

1. 以網頁呈現成功連結資料的統計表，供專業應用查詢。
2. 改善確定性連結鍵值錯誤、空白資料，以及連結鍵值間不一致資料的連結率。
3. 繼續維持以資料支援安全研究的機制。
4. 更新「事故關聯資料分析與評估系統」。

1.2 研究範圍與對象

今(2005)年計畫除持續更新道路交通事故相關資料整合系統與資料，以及改善「事故關聯資料分析與評估系統」外，也依據 2004 年的研究成果及建議，以 1999-2002 年的道路交通事故資料所連結出的資料為主要研究範圍，依據不同的工作需求，採用合適之資料種類，進行 Yes-chain 與 No-chain 資料的應用；此部分的範圍與對象如下：

1. 參考國外之線上查詢系統，利用本計畫所產製的資料統計表，整合至本所(運輸安全組)的「運輸安全專業網站」供專業應用查詢使用。
2. 機率連結的部份，一方面評估國外現有軟體與應用之發展現況，以及其於國內資料之可行性，另一方面則使用在資料淨化過程中剔除之資料，探討資料再連結之情形。
3. 提供資料給本所其他合作計畫進行分析的部分，此以本所(運輸安全組)的研究計畫為優先。
4. 資料探勘技術之應用潛力初探。

1.3 研究內容與工作項目

本計畫為 2 年期(2004-2005)計畫的第 2 年(2005)，具體研究內容與工作項目如下：

1. 因應第 4 及第 5 章的研究需要，收集國內外文獻（包括：機率連結、資料探勘等）資料，以及網頁查詢系統資料。2005 年完成的文獻資料收集與分析請見第 2 章。
2. 採 2004 年配合各資料提供單位之作業方式要求，完成新年度資料的更新處理。2005 年完成之資料處理內容包括：(1)92 年及 93 年事故資料庫之彙整；(2)92 年死因資料彙整；(3)92 年及 93 年車輛理賠(任意險及強制險)資料及強制險受害人資料之彙整；(4)人車監理違規資料之重新整理工作。研究成果請見第 3 章。
3. 完成整合系統之軟硬體平台建置。研究成果請見第 4.1 節及 4.2 節。
4. 改善及擴充「事故關聯資料分析與評估系統」。2005 年的改善成果請見第 4.3 節之說明。
5. 利用 2004 年已成功連結的資料，完成線上專業查詢系統，並整合至本所(運輸安全組)「運輸安全專業網站」中。此部分之研究成果請見第 4.4 節。
6. 除評估國外機率性連結相關之軟體功能外，並另由淨化過程中刪除的資料及經淨化後而無法成功連結之資料，進行機率性連結的實作。2005 年完成的本計畫資料連結率探討請見第 5.1 節及 5.2 節。
7. 支援其他本所研究計畫所需之資料。其他計畫應用本計畫資料的成果摘要說明請見 5.3 節。
8. 完成資料探勘技術的應用潛力初探。研究成果請見第 5.4 節。
9. 完成 2 篇研討會論文之發表，請見參考文獻[36,37]。

第二章 文獻回顧

本章中針對今(2005)年計畫中的工作內容，包括：改善資料連結率、建立資料查詢網站、資料探勘技術應用等，回顧所蒐集的相關文獻資料與網站料查詢系統於下。

2.1 機率連結理論與發展

國內因有身分證號可做為每個人的唯一識別碼，因此在資料連結的處理上較國外資料連結系統方便，但是國內資料亦會因資料收集及鍵入等問題，而造成資料無法正確連結。對於部份無法連結，或是連結成功率低於預期之資料，將嘗試使用機率連結進行資料連結，以提高可供分析用之資料量。

記錄連結(Record Linkage)是由 Newcombe[14]首先提出記錄連結的定義，而由 Fellegi 及 Sunter[3]發展出機率連結的數學模式，並依資料連結情況分為確定性連結、機率連結及混合性連結三種連結模式。記錄連結一開始是被應用在追蹤家族病史以及醫院記錄的分析上，爾後已有相當多之應用系統被發展出來，如加拿大統計局的 GRLS(Generalized Record Linkage System)[60]、美國加州的自動死亡連結系統(GAMLIS)、美國交通事故資料連結系統(The Crash Outcome Data Evaluation System, CODES)、西澳的 WADLS(Western Australian Data Linkage System)[2]、維多利亞的就診資料連結系統(Victorian Admitted Episodes Dataset, VAED)[19]等，應用的領域也擴及至人口普查、農業調查、犯罪偵查、醫藥預防保健、疾病統計、商業活動調查等。

在國外研究使用的資料庫中，由於常缺乏個人的唯一辨識碼，若要串連同屬一個人的相關資料時，僅能使用替代欄位，如社會安全號碼(Social Security Number, SSN)、姓名、性別、生日、郵遞區號、職業……等。而在無法進行確定性連結時，僅能以機率性連結的方式處理。從 1959 年 Newcombe 提出記錄連結的觀念，到 Fellegi 及 Sunter 提出數學模式，至今歷經數十年的發展，也不斷有新的資料處理技術被提出來，但無論使用何種資料連結方法，在資料連結前都須先對資料內容有充份的了解，才能為資料連結提供最好的解決方案。在牛津大學的 ORLS(Oxford Record Linkage Study)系統[8]發展經驗中，資料錯誤的類型可概分為四大類：(1)謄寫或是替代性的錯誤(Transcription or Substitution Errors)，如

將手寫的 1 及 7、3 及 8 抄錄錯誤，錯誤率約有 86%。(2)位置錯誤(Transposition Errors)，如將第 2 碼及第 3 碼資料對調，此類的錯誤率約有 8%。(3)位移錯誤(Shift Errors)，如所有資料向右或向左平移，此類的錯誤率不到 1%。(4)無法歸類在前三項的其他的錯誤，約有 5%。Gill[5]區分出幾種常見的類型：(1)一生中鮮少做改變的資料，如姓氏、性別、出生日期等；(2)會隨著社會環境做改變，如住址、工作職業、婚姻狀態等；(3)與個人相關且較特殊的資料，如重大疾病種類及位置、藥物過敏疾狀等；(4)與家族相關的資料，如家族姓氏、生育排行等。

在資料連結過程中，若變數重複性高，則要去辨識是否為同一人的可能性則較為困難，因此在資料連結過程中，常需輔以連結權重值(Weights)的計算，以提高連結成功率，權重值也代表了兩筆記錄是否為同一筆的可能性。對於權重值的計算通常需參考資料實際上可能存在的問題，而對權重有基本的計算模式，如當兩筆記錄的出生年差一年或十年時，系統考量可能會因人為疏失而造成資料鍵入的錯誤，因此在系統權重的設定上，只做少幅度的減少。但若出生年差異在一個較特殊的差距時，則權重值會有明顯的調整，如出生年差異達七年時[5]。

從文獻收集整理中發現，許多機率連結系統發展過程中，常需藉由實證經驗來決定權重數值的範圍，而且初期在資料連結後，須再利用人工檢核的方式來確認資料連結的正確與否，進而設計出適合該資料的權重範圍及判斷門檻值。

在國外的研究資料中發現，在資料連結前會對資料做初步的分類處理，常見的處理方法有「語音發音分類法」、「檔案群組分類法」(File Blocking Items)及「名稱簡稱法」(Name Compress)等。而目前被發展出來的「語音發音分類法」包括「Soundex Code」、「SINGS Code」(Soundex, Initial Numericise Granick, Smith)、「NYSIIS」(New York State Information and Intelligence System)及牛津大學 ORLS(Oxford Record Linkage System)使用的「ONCA」(Oxford Name Compression Algorithm)，雖然部份系統發展過程中，對於部份相似的文字無法正確分類，但隨著技術的發展，目前對於文字發音的分類上也越來越準確，但目前的語音發音分類法還是較適合於英語語系的語言。「檔案群組分類法」是將檔案資料進行排序或區塊化(Blocked)的處理，以增加資料比對及搜尋的效率，但風險是當資料分類錯誤時，將無法成功比對。而「名稱簡稱法」則是為類似的名稱建立固定長度的唯一碼，可將同音字、拼音錯誤或鍵入錯誤的資料做適當的分類，以提高資料比對成功率。如 ORLS 系統在檔案完成區塊化後，及在進行資料排序

比對前，會為每一筆資料增加四個不同的變數，包括：(1)可用來辨識該筆資料的唯一碼；(2)用來記錄連結結果的配對記錄碼；(3)可用來記錄不同版本的編碼；(4)資料輸出入對應號碼[5]。

國外研究累積數十年的經驗，除有完整的理論及實務處理經驗外，並已發展出專屬的機率連結套裝軟體或資訊服務，本研究亦對文獻提及或是現有資訊市場上可找到之軟體或服務，依軟體的發展及資料處理特性整理如表 2.1 所示。但目前的軟體服務仍較偏重在英語語系之文字，對於雙位元之語言(如中文、韓文、日文等)則較不適合。在軟體或資訊服務的費用上，會隨著處理的資料量大小而異，基本上費用也不便宜。而在本計畫 2004 年的研究[24]中，曾使用 Linkage Wiz 做資料連結測試，測試結果發現判斷變數的選擇及權重值的設定都會影響結果，但受限於套裝軟體並無法進一步了解其權重及門檻值的設定方式。

綜上所述，國外不像台灣每個人都有專屬的身分證號，因此對於不同資料庫間的連結，僅能使用其他替代欄位進行連結，如姓名、社會安全號碼、性別、出生日期、死亡日期、姓氏、居住地址、郵遞區號、婚姻狀態……等，因此須要發展機率連結來解決資料連結的問題。此外，當選用的連結資料庫不同時，均需依資料特性挑選合適的欄位做為連結欄位，但當資料登錄不完整或不正確，或連結後權重設定不良時，都有可能造成後續資料連結失敗，此些問題都是在發展「機率連結」時所須考量的地方。

目前文獻回顧所收集到的系統包括有：美國的 CODES、CIREN、GAMLIS；加拿大的 GRIS；西澳的 WADLS、VAES；英國牛津大學的 ORIS 等。每一個系統均有蠻長的發展時間，所累積的資料處理經驗也相當豐富，因此在於資料連結時使用的權重值、門檻值均有其參考值。參考國外發展的經驗，對於本研究後續的資料連結分析，也提供某種程度的參考，如當使用生日輔助身分證號的連結時，可設定出生年份 ± 3 年應可視為合理的範圍。但以本計畫所取得之研究用資料，每個資料庫均有身分證號的欄位，在資料欄位變數特性上遠優於國外資料欄位，且國外所使用的機率連結軟體，較強調英語語系資料的處理，並不適用於雙位元的中文語系。因此現階段之研究應先求如何善用不完整或不正確的身分證號，如：只利用身份證號前九碼作為連結鍵，以及充分利用本計畫所取得的不同資料庫間的互補效益。

表2.1 機率連結軟體整理介紹

公司名稱及相關網址	產品名稱	產品特色	備註
Identity Systems http://www.identitysystems.com	Identity Search Server(ISS)	1. 利用 C 語言撰寫核心運作程式，包含 Connection Server, Search Server, Console Server。 2. 提供以 Java 撰寫之 Web 操作界面。 3. 支援資料庫包含 Oracle, DB2/UDB, SQL Server，或是透過 ODBC 可存取之資料庫。 4. 支援作業平台包括 AIX, HP/UX, Linux, Windows NT/2K/2003, Solaris, OS/390, z/OS 等。	售價從美金 45,000 元到 300,000 元以上不等
	SSA-NAME3	1. 可試用在固定或未固定格式之資料。 2. 使用 Java 所撰寫之操作環境，可單機執行或透過 TCP/IP 與 SSA-NAME3 系統處理程序連線。	
	CJK-Support LinkageWiz	支援 ISS, SSA-NAME3 在中文、日文及韓文等雙位元語系。 1. 利用 NYSIIS 及 SOUNDDEX 規則進行語音姓名配對。 2. 可自行針對不同欄位調整及建立權重。 3. 可建立 Nick Name 對照表。 4. 支援不同來源之資料庫，如：Text Files, Dbase, FoxPro, MS Excel, MS Access 及 SQL Server 等。 5. 提供人工檢視功能。 6. 作業環境 Windows 98/Me/NT 4.0/2000/XP。	
ChoiceMaker Technologies Inc. http://www.choicemaker.com/		1. 主要提供諮詢顧問、教育訓練及專案研究之服務。 2. 利用機器學習(machine learning)進行資料處理。	
GRLS(Generalized Record Linkage System) Statistics Canada http://www.statcan.ca		1. 由加拿大統計局使用 Fellegi-Sunter 理論發展而成，目前為第四版。 2. 提供 one-file(internal)及 two-file 的連結。 3. 為 client-server 架構，Unix Server 或 mainframe。 4. 需搭配 Oracle RDBMS 8.04 版本以上資料庫。 5. 提供多人共用服務。	由南美洲人民服務部所發展。售價從美金 600 元至 4,000 元不等。可線上下載試用版，但僅能處理 750 筆資料。
Strategic Matching, Inc http://www.linksolv.com/	LinkSolv	1. 使用 Access 資料庫做為資料儲存使用。 2. 提供線上討論區做為使用經驗交流。 3. Utah Crash Outcome Data Evaluation System(CODES)所使用。	

表2.1 機率連結軟體整理介紹(續)

公司名稱及相關網址	產品名稱	產品特色	備註
http://members.shaw.ca/andre.wajda/linkpro.html	LinkPro	<ol style="list-style-type: none"> 以 SAS 巨集撰寫而成，整合至 SAS 軟體中，無法單獨執行。 利用 NYSIIS 及 SOUNDEX 規則進行語音姓名配對。 以 \log_2 連結成功的次數做為權重計算。 支援資料格式：SAS Data Sets。 	
DataFlux Corporation http://www.dataflux.com/	dIPower Studio 4.3	<ol style="list-style-type: none"> SAS 的子公司 可透過 ODBC 與其他資料庫進行資料存取。 支援 Windows NT/2000 	
Ascential Software http://www.ascential.com/	Blue Fusion SDK Ascential ProfileStage, Ascential QualityStage, Ascential DataStage	使用者可利用 SDK 自行開發應用程式，但需配合 SAS 使用整合 AutoMatch 核心技術發展而成，目前已未單獨提供 AutoMatch 軟體。	目前已被 IBM 公司所併購。
Intech Solutions Pty. Ltd. http://www.intechsolutions.com.au	IQ Rapid Address	可由所輸入的部份地址資料，自資料庫中搜尋出正確、完的地址資料	
	IQ Standardiser	<ol style="list-style-type: none"> 利用語法分析去轉換、標準化及淨化資料，讓輸入之資料結構化。 可與澳洲的郵政地址檔案(Australia Postal Address File)結合，以修正使用者所輸入之地址資料。 	
Trillium Software http://www.trilliumsoftware.com	IQ Matcher	可處理重覆之資料，並提供資料機率性連結處理。	
SAS Institute Inc. http://support.sas.com/rnd/warehousing/cleanse/	Trillium Software	為一大型的資料淨化及資料連結軟體，較適用於商業用途(如 CRM, ERP, SCM 等)	
AccuMatch http://www.accumatch.com/	SAS Data Quality-Cleanse Software	SAS 新發表針對資料處理之軟體，適用於 SAS 8.2 版中。	
	Searchlight	<ol style="list-style-type: none"> Java-based 機率連結軟體 可利用姓名、地址、生日、電話號碼等做連結。 	
http://www.the-link-king.com/	The Link King	<ol style="list-style-type: none"> 由 MEDSTAT 所發展的演算法。 需與 SAS 8.2 版以上搭配使用，但在 9.0 版會有較好之使用效果。 利用 NYSIIS 及 SOUNDEX 規則進行語音姓名配對 	
http://nedinfo.nih.gov/docs/RLERSearch.htm	RLERSearch - Probabilistic record linking engine	供 Perl 使用的機率連結使用的函式庫。	

2.2 線上資料查詢系統

隨著網際網路所帶來的便利，越來越多的數據資料都會透過網路進行訊息發布，也包含了定期的統計資料(如年報、季報等)。在國內外也有相當多之單位會在特定時間進行統計資料的發布，本節將針對提供線上資料查詢的國內外網站進行介紹，並參考其設計內容以做為後續設計之參考。

2.2.1 國內網站

1. 中華民國經濟部全球資訊網[58]

主要提供經貿統計相關資料，包含「國家經濟統計」、「工商業生產統計」、「進出口貿易統計」等，此部份之資料多為靜態的報表資料下載(多為 HTML 或 Excel 格式)。而目前經濟部在「經濟統計資訊網路」中也提供動態的資料查詢，包含了「工業生產統計資訊」、「工商企業經營概況」、「業務統計」等。欲使用此網站功能，需先成為網站會員，且採取使用者付費之維運方式，初加入會員會提供 100 點基本點數，每完成一項查詢即進行點數扣抵。在查詢畫面中，已將各分項查詢內容做詳細之區分，使用者僅需勾選查詢項目即可，再設定查詢週期(如月或年)及查詢起迄期間(僅至月份)後，即可進行動態資料查詢，並以 HTML、Excel 及 Text 的檔案型式提供查詢結果顯示及下載檔案，查詢畫面如圖 2-1 所示。

圖 2-1 經濟統計資訊網路查詢系統功能畫面

2. 中華民國統計資訊網[61]

為主計處所設計之統計專門資訊網，包含「全國統計資料」及「資訊服務」兩大類，並設計「專業人士」及「一般民眾」兩種不同的使用者網頁，涵蓋之範圍相當廣泛，所提供的資料有概述性的文字性敘述，亦有統計報表，呈現方式以 HTML、Word 檔、Excel 檔、PDF 檔或圖表之方式為主，畫面如圖 2-2 所示。

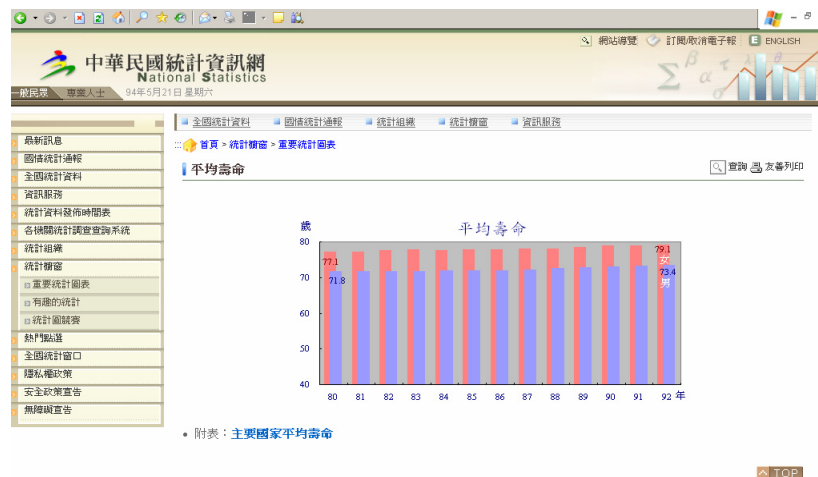


圖 2-2 統計資訊網統計圖表

3. 衛生資訊統計網[55]

為行政院衛生署統計室所建置的網頁，包含了「醫政統計」、「藥政統計」、「食品衛生統計」、「疾病管制統計」、「保健統計」……等統計資訊。藉由適當的分類，使用者可由網頁左邊選單中勾選所欲查詢的資料項目，再由展開選單中勾選欲查詢的資料年份(如圖 2-3 所示)，最後再以列表的方式，呈現不同的統計資料，並以 Excel 格式提供下載(如圖 2-4 所示)。



圖 2-3 衛生統計資訊網統計資料選單頁



圖 2-4 衛生統計資訊網統計資料結果列表頁

4. 內政部警政署全球資訊網[59]

所提供的統計資料均是以警政相關之資訊為主，除了一般常態性的統計資料外，如警政統計年報、刑案統計指標、警政統計通報等，亦提供專題式的分析資料，針對特定需求之主題進專案式的統計資料。資料的提供均以固定內容的方式提供(如 PDF、Word 或 Excel 格式)，畫面如圖 2-5 所示。



圖 2-5 警政署線上統計資料查詢畫面

國內已有越來越多之網站提供線上統計資料下載服務，僅大多侷限於定期固定格式之統計資料(如年報、季報、月報等)，或是針對特定研究主題所進行之專題研究。而所提供下載之資料格式大多為 HTML、PDF、Word 或 Excel 格式。整體而言，目前網站所提供之統計功能較欠缺由使用者自行設定查詢變數之線上統計分析功能。

2.2.2 國外網站

1. 國際道路交通及意外事故資料庫(IRTAD)[53]

提供了國際間不同國家在道路交通及意外事故的比較資料，目前所包含的國家有澳大利亞、奧地利、比利時、加拿大、捷克、丹麥、美國、西班牙、日本、韓國、義大利等共三十個國家的資料。其統計報告資料除了以文字敘述方式呈現外，亦會以圖形的方式做呈現，如圖 2-6、圖 2-7。

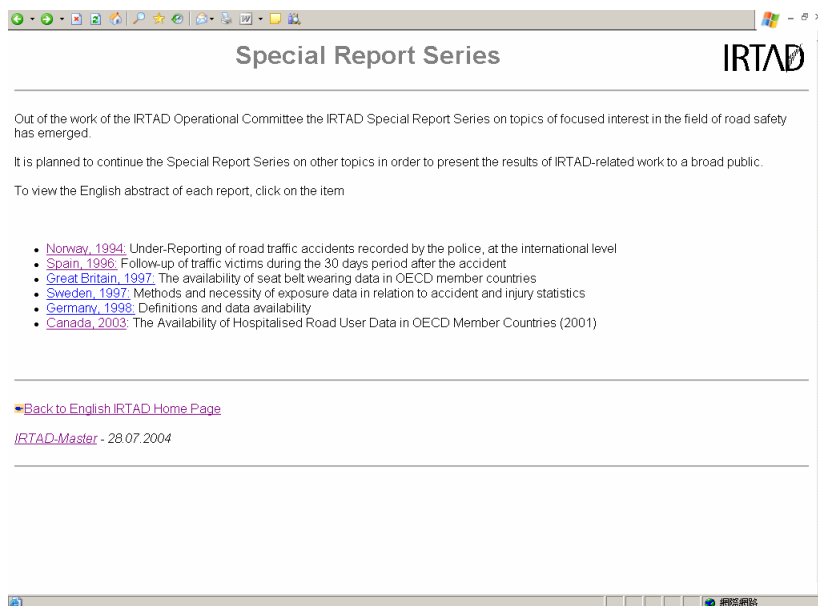


圖 2-6 IRTAD 文字敘述性之報告資料

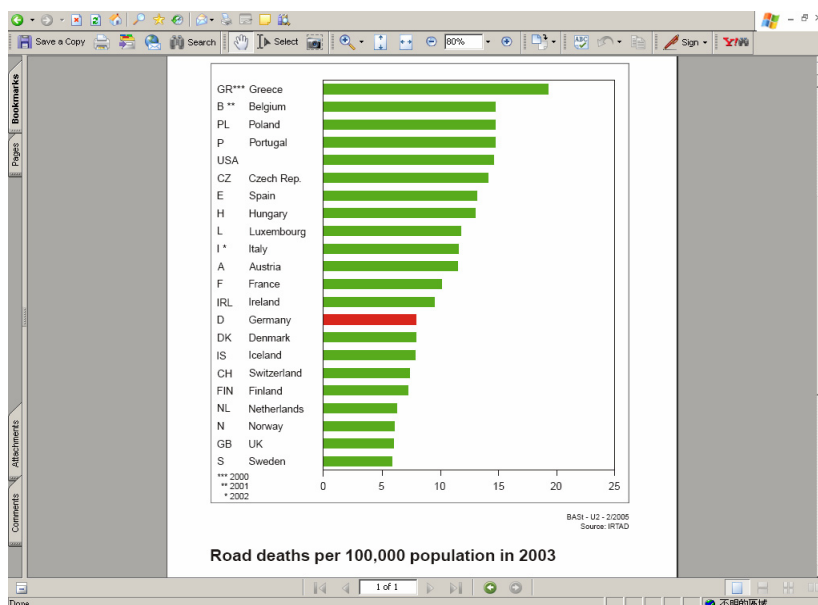


圖 2-7 IRTAD 圖形統計圖表式統計資料

2. 東京都警視廳[57]

日本東京都警視廳所提供的統計資料包含「犯罪發生狀況」、「交通量統計概要」及「交通事故發生狀況」等的統計資料；在「交通事故發生狀況」的統計資料中包含二天前與歷年的交通事故發生狀況統計，及利用電子地圖做為點選方式的查詢方式，利用電子地圖及統計圖表的方式呈現統計資料，如圖 2-8 所示。



圖 2-8 東京都警視廳交通事故發生統計圖

3. Fatality Analysis Reporting System (FARS)[56]

為 NHTSA 從 1975 年所開始進行的研究，1996 年開始建立線上查詢系統，可提供更直覺、更有效的嚴重交通事故查詢方式，使用者僅需在網頁上點選所需的資料庫及其資料欄位，再設定查詢方式即可在線上顯示統計結果。其特性包括：(1)提供次數(Frequency)統計(無提供平均數或變異數等)；(2)非資料庫串聯查詢(事故資料表變數)；(3)提供單筆資料輸出。相關查詢畫面如圖 2-9、圖 2-10 所示。

Main

- Home
- Did you know?
- Acronyms
- Terms

Reports

- Trends
- Crashes
- Vehicles
- People
- States

STEP 1. Choose a Year

The Fatality Analysis Reporting System (FARS) contains data on all vehicle crashes in the United States that occur on a public roadway and involve a fatality. This FARS Query System provides interactive public access to fatality data through this web interface.

If you would like additional training on using this query system, try some of the exercises below.

- [Univariate Tabulation Exercises](#)
- [Cross Tabulation Exercises](#)
- [Case Listing Exercises](#)

If you would like to see the final reports, click links below.

- [Univariate Tabulation Reports](#)
- [Cross Tabulation Reports](#)
- [Case Listing Reports](#)

Select a Year to Query:

圖 2-9 設定查詢資料年份

圖 2-10 設定變數條件

4. CARE Online Alabama[51]

主要提供阿拉巴馬州(Alabama)1999 年至 2003 年「事件特性資料」或「事故當事人資料」，包含了特定州別、事故型態或各種組合型態，可選擇不同的分析變數(如事故時間、高速公路、市區道路等)，其查詢特性包括：(1)次數(Frequency)統計(無提供平均數或變異數等)；(2)非資料庫串聯查詢(事故資料表變數)；(3)次數頻率及頻率百分比(含累加)；(4)事故頻率長條圖；(5)不同探討區域或範圍之次數比較(如不同郡別事故嚴重性比較)(6)交叉資料表(Cross Table)可自訂查詢結果的行(Row)及列(Column)變數。相關查詢畫面如圖 2-11、圖 2-12、圖 2-13 所示。此系統也提供其他州的查詢資料，如喬治亞州(Georgia)、佛羅里達州(Florida)、田納西州(Tennessee)、北卡羅萊納州(North Carolina)等。部份州區為確保資料保密及隱私問題，在使用線上分析系統時，需先取得登入的帳號及密碼。

HELP:
What is a data source? A data source is the set of data that CARE is set to process. A dataset contains every record for a year or a set of years. For example if you select 2002 Crash Data as your dataset you will be looking at the set of data containing every traffic crash record for the year 2002.

圖 2-11 選擇資料年期及類別

Selected Filter: **Bicycle**

☒ Suppress Nulls ☒ Suppress Zero-Valued Frequencies

From the listing below, select the variables on which to perform the frequency.
Hold down the shift key or the control key to select more than one variable.

Select variables:

- V005: YEAR
- V006: WEEK
- V007: TIME
- V008: DAY OF WEEK
- V009: RURAL OR URBAN
- V010: HIGHWAY CLASS
- V011: INTERSECTION
- V012: CNTRL ACCESS LOCATION
- V013: PRI CONTRIB CIRC
- V014: PRIMARY CONTRIB UNIT
- V015: FIRST HARMFUL EVENT
- V016: EVENT LOCATION

Next >>

圖 2-12 選擇分析變數項目

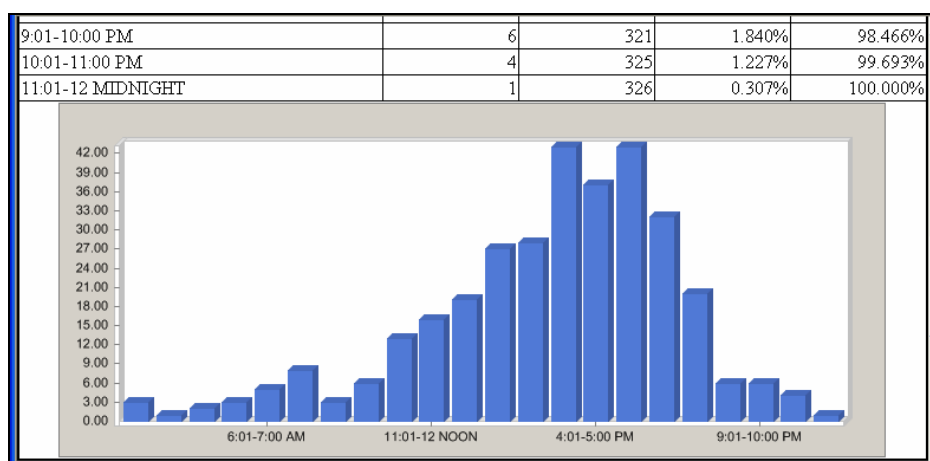


圖 2-13 顯示統計長條圖

5. Federal Motor Carrier Safety Administration (FMCSA)[49]

由佛羅里達州(Federal)運輸部門所建置的線上分析網站(如圖 2-14 所示)，進入 Crash Profiles Online 網站後，先由全美地圖中點選欲分析的州別，再由高反差的地圖中直接瀏覽該州總覽資訊，並以表格及圖表方式呈現。此外亦可由 State Profiles 中進行不同年度中各州與全國資訊的比較分析，系統並將比較項目區分為概要(Summary)、車輛(Vehicle)、駕駛者(Driver)、環境(Environment)、碰撞(Crash)…等項目。各項目中又有不同的細項比較，如駕駛者比較內容又區分為駕駛年齡、駕照種類、安全帶、飲酒…等，畫面如圖 2-15 所示。在分項的資料呈現在可選擇表格、圖表、敘述、歷年表格、歷年圖表共五種呈現方式，如圖 2-16 所示。

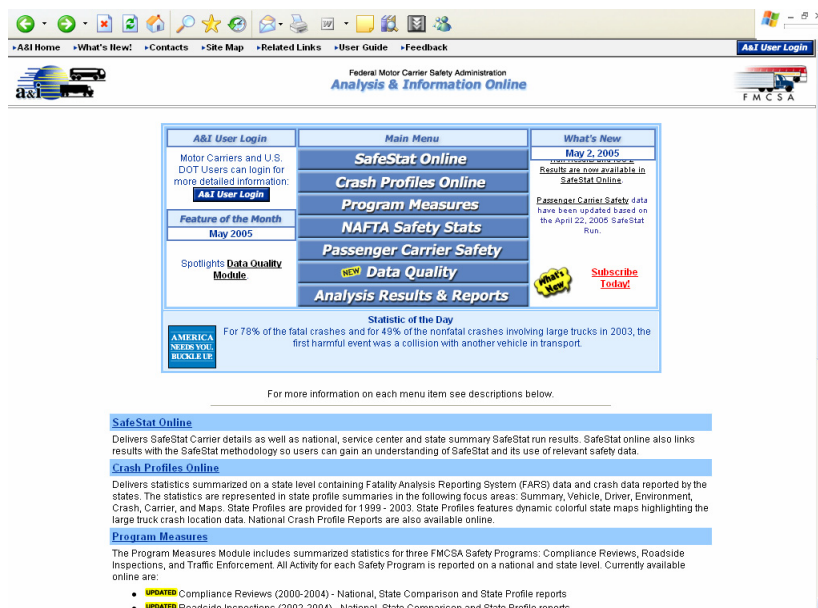


圖 2-14 FMCSA 線上分析網站

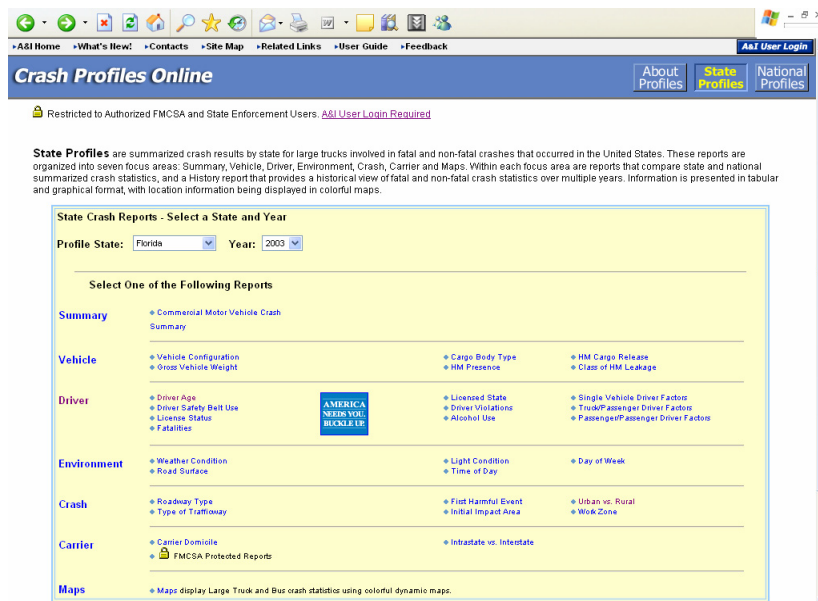


圖 2-15 Crash Profiles Online 分項比較項目

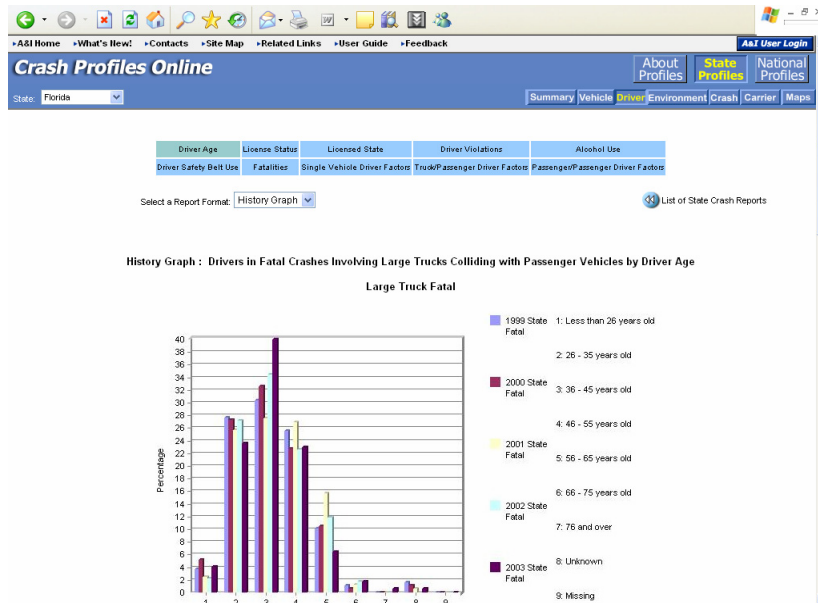


圖 2-16 Crash Profiles Online 分項歷年資料比較圖

6. 伊利諾州車禍事故資料庫 [50]

伊利諾州(Illinois)收集了 1994 至 2003 年交通事故警察的調查報告資料，在查詢畫面中共分為車禍特徵查詢、含摩托車及未含摩托車三大類。此三大類具有類似的查詢畫面，但查詢項目不盡相同，在車禍特徵查詢方面，包含了碰撞型態、事故嚴重度、光線、天氣、星期幾等，查詢畫面及結果如圖 2-17、圖 2-18 所示。此外，伊利諾州還建立了，死亡率統計、外傷登錄等資料庫查詢系統。其中死亡率統計資料庫是利用緊急醫療救援系統(Emergency Medical Service, EMS)的資料所建構而成，其中死亡分類的部份採用 ICD-9 及 ICD-10 兩種編碼。進入查詢表單後，分別設定不同的查詢條件，如年齡層、性別、居住地區等。在輸出條件上，可以設定以先前的某項查詢條件為主，並採取升冪或是降冪的方式處理，最後以柱狀圖及表格的方式呈現查詢結果。

DESCRIPTIVE STATISTICS OF AGGREGATE CRASH DATA
QUARTERLY DATA FOR 1994 - 2003
Illinois Department of Transportation
Division of Traffic Safety
Crash Report Database

CRASH CHARACTERISTICS

*Note: To select more than one specific category within each data element, click on a category and then hold down the **control key** while clicking on additional categories.*

For details regarding each data element in this database, [see data dictionary](#).

Query Criteria

Collision Type
Select All
Vehicle Overtum
Pedestrian/ Pedalcyclist
Animal
Fixed Object

Severity of Crash
Select All
Fatal Crash
Injury Crash
Property Damage

Lighting Conditions
Select All
Daylight
Dawn
Dusk
Darkness

Weather
Select All
Clear
Fog/Smoke/Haze
Rain
Severe Cross Wind

Day Of Week
Select All
Monday
Tuesday
Wednesday
Thursday

Time Period

County of Occurrence

圖 2-17 查詢條件設定畫面

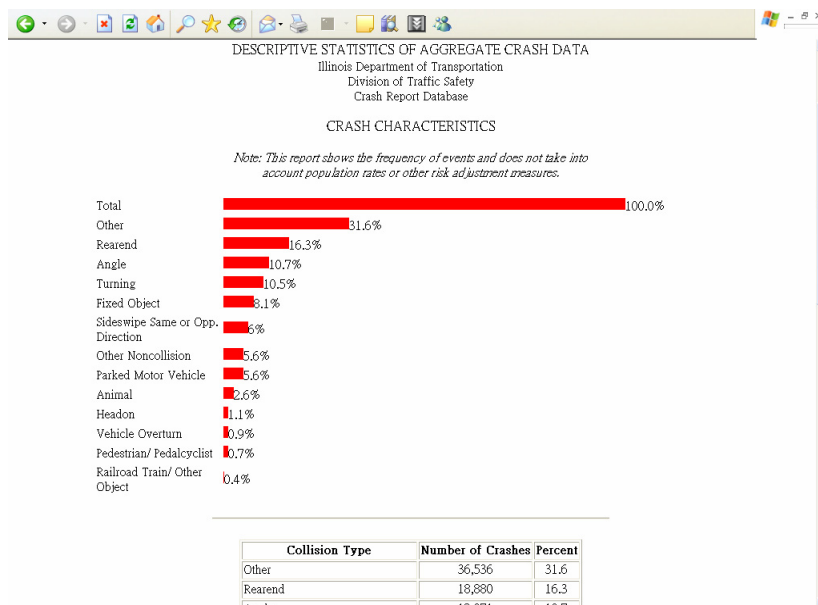


圖 2-18 查詢結果畫面

7. 加拿大統計局(Canadian Statistic)[60]

提供了兩種不同統計資料，在免費的查詢系統部份，將統計資料區分為不同的主題，如農業、人口統計、貿易、交通運輸等，在各分項裡面都會有其更細部的分類。以交通運輸為例，又可區分為空運、海運及一般公路運輸，而在進入細項之後，則會提供各式的統計表格(如圖 2-19 所示)。

Statistics Canada / Statistique Canada

www.statcan.ca Canadian Statistics

Search Canadian Statistics [Key Word(s)] Search

Related tables: Modes of transport

Motor vehicle registrations, by provinces and territories (Newfoundland and Labrador, Prince Edward Island, Nova Scotia, New Brunswick)

	Canada	N.L.	P.E.I.	N.S.	N.B.
Total vehicle registrations	25,100,296	416,773	92,579	645,525	586,586
Total road motor vehicle registrations	19,081,478	266,087	80,997	544,275	469,468
Vehicles weighing less than 4 500 kilograms	17,920,360	253,182	75,162	518,429	442,592
Vehicles weighing 4 500 kilograms to 14 999 kilograms	389,810	3,997	1,659	8,475	7,495
Vehicles weighing 15 000 kilograms or more	285,154	2,856	2,674	6,944	3,953
Buses	77,447	1,348	49	1,818	2,743
Motorcycles and mopeds	408,706	4,702	1,451	8,608	12,683
Trailers	4,492,733	30,107	9,695	49,684	75,713
Off-road ¹ , construction, farm vehicles	1,526,083	120,578	1,887	51,555	41,404

1. Off road vehicles include snowmobiles, dune buggies and amphibious vehicles.
Source: Statistics Canada, CANSIM table (for fee) 405-0004.
Last Modified: 2005-04-22

Find information related to this table (CANSIM table(s), Definitions, data sources and methods, The Daily publications, and related Canadian Statistics tables).

圖 2-19 制式統計表格資料

此外，為了提供彈性更大的資料查詢功能，加拿大統計局提供自訂表單輸出功能，此服務稱為 CANSIM(如圖 2-20 所示)。進入系統後可先選擇資料列表的方式，如依主題、文字搜尋、表格編號等。主題列表的內容則與先前介紹的分類相同，再進入所欲查詢的表格內容後，即再進行細項的選擇，設定完成後即可進行資料輸出。但此部份每次查詢需付 3 塊加幣的查詢費用。

Canada

Statistics Canada / Statistique Canada

Table 405-0004 - Road motor vehicles, registrations, annual (Number) (126 series)

Definitions, data sources and methods: Road Motor Vehicles - Registration - 2747

Categories: Modes of transport

View vector directory

View latest article from The Daily related to this table.
Display list of free Canadian Statistics tables related to this table.

Note: You must select at least one item from each one of the lists below.

Geography (14 items)

Select all Deselect all View checklist

Canada
Newfoundland and Labrador
Prince Edward Island
Nova Scotia
New Brunswick

Type of vehicle (9 items)

Select all Deselect all View checklist

Total, vehicle registrations
Total, road motor vehicle registrations
Vehicles weighing less than 4,500 kilograms
Vehicles weighing 4,500 kilograms to 14,999 kilograms
Vehicles weighing 15,000 kilograms or more

Reference period:
From: 2004 To: 2004 (Annual data)

圖 2-20 查詢表格細分項設定

8. Crash Outcome Data Evaluation System(CODES)系統

CODES 則整合了交通意外事故資料庫、醫療資料庫、保險及車輛資料庫，將連結後之資料做統計分析，並透過網路進行跨州的研究，目前全美共有 27 州參與。其中威斯康辛州(Wisconsin)的 CODES 系統[52]將收集而來的資料，分成不同的研究主題，如兒童安全、急診室救護資訊、腳踏車碰撞事故、摩托車碰撞事故、行人事故等單位。在進入各別主題後，系統會先將資料依月份先做分類，並列出探討的主題統計資料，再以 PDF 方式呈現。系統畫面如圖 2-21 所示。

而威斯康辛州所提供的 CODES 查詢系統稱為 Cube Analyzer，在彈跳視窗中，由使用者先設定所欲查詢的資料種類(目前有提供三大類，碰撞事故相關資訊、碰撞與醫療住院系統整合之資訊、碰撞車輛相關資訊)。選好查詢資料庫後，畫面左邊會提供兩類的查詢方式，第一種是由系統所制定好的查詢條件，點選後系統即帶出查詢資料。第二種則是採自訂欄位的方式，可依序選定不同的查詢欄位，系統會做分項展開，當所選的欄位越多時，資料將被分解的更細，查詢結果畫面如圖 2-22 所示。

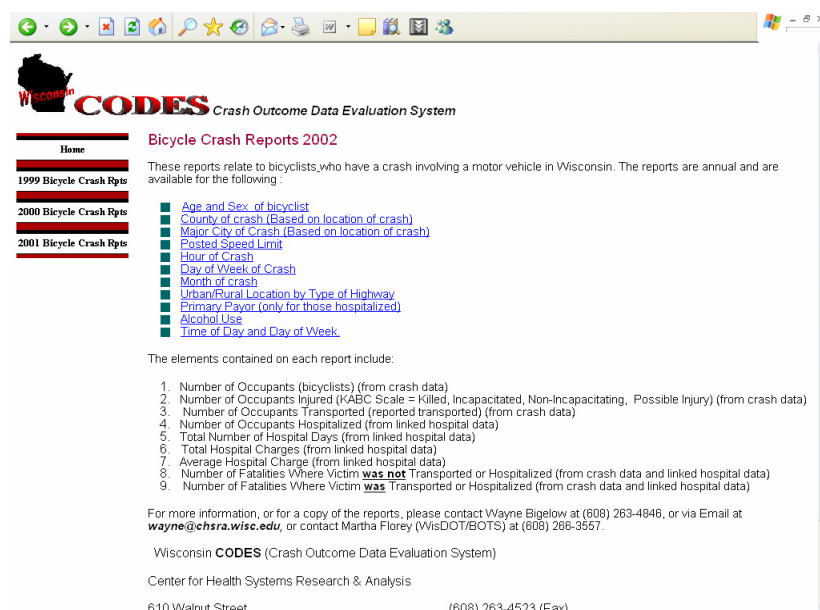


圖 2-21 CODES 研究主題資料下載畫面

</

圖 2-22 指定欄位查詢結果畫面

9. National Center for Injury Prevention and Control[54]

這是由美國疾病管制局(Centers for Disease Control and Prevention, CDC)裡的傷害及防治中心所製作的線上查詢系統，稱為 WISQARS(Web-based Injury Statistics Query and Reporting System)。其查詢資料共區分為致死(Fatal)及非致死(Nonfatal)兩大類，其查詢項目包括死亡傷害查詢(Fatal Injury Reports)、主要死因報告(Leading Causes of Death Reports)及生命損失年(years of Potential Life Lost, YPLL)等項。系統所提供查詢的資料年份從 1981 至 2002，其中在 1999 年時在死因的資料記錄上由原有的 ICD9 改為 ICD10，故以該年做為資料查詢年份分割點。

進入各查詢畫面後，可設定所需輸出的資料項目或變數範圍，其中在進階的變數設定上，可使用以四年為單位或採自訂範圍的方式做資料輸出(如圖 2-23)，並可選擇不同的年份資料做為比較基準。送交查詢後會在網頁中顯示基本的查詢結果(如圖 2-24)，而詳細的查詢資料可自行下載 CSV 檔觀看。

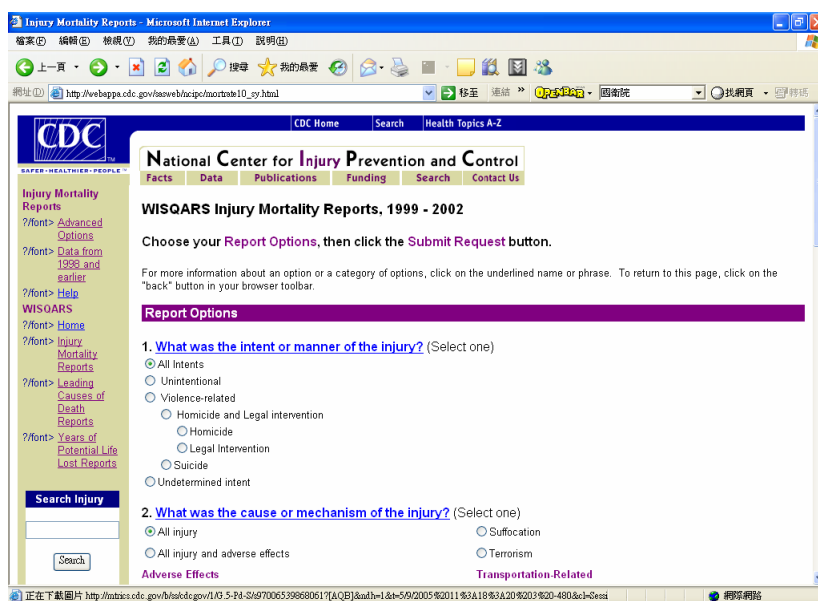


圖 2-23 WISQARS 查詢變數設定畫面

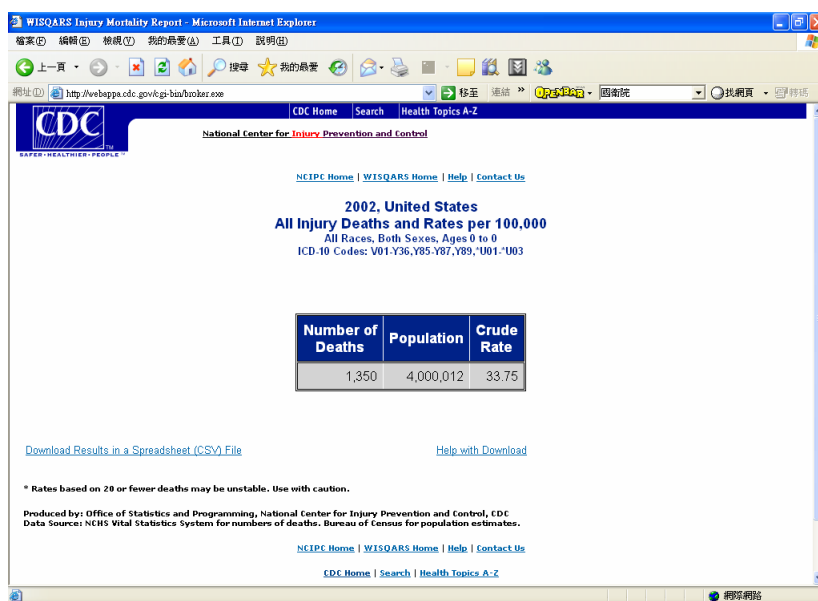


圖 2-24 WISQARS 查詢結果顯示畫面

由以上與車輛、健康、車禍等有關的國外網頁查詢系統可發現，由使用者自行選擇變數進行交叉統計的資料查詢方式，在國外堪稱常見。在與車禍有關的網頁資料查詢系統中，其可供查詢之變數的呈現方式，多數會以損傷結果、車禍事件、人、車、環境(路)幾個層面進行分類；這些網站雖均公開給一般大眾，但對於專業術語多數並未有特別的解釋，僅是著重於說明資料來源以及資料收集方式，故網頁資料查詢的設定對象，應是具有基本相關專業知識的特定大眾。

由所收集的國內外網頁資料查詢系統可知，制式統計表格的資料查詢在國內已有相當多的單位可提供此種資料流通管道，而由使用者自行在線上選擇變數，進行即時網頁資料查詢的技術也應屬成熟，故採用國外網頁的線上彈性查詢方式，應較符合本計畫設定以專業人士為使用對象的需求；再者，以損傷結果、車禍事件、人、車、環境(路)分類資料的作法，實係符合一般車禍分析的主要考量層面，故本計畫以此為原則，進行查詢變數的分類；而有關線上說明部份，也是在假設使用者具有基本相關專業知識的前提下，進行設計。

2.3 資料探勘技術

資料探勘(Data Mining)強調的是利用分類、關連性、序列分析、群集分析、機器自我學習及其它統計方法，從大量的資料中萃取出隱含的、不為人所知、可信而有效的知識，以做為決策輔助或預測的參考[40,18]。眾多學者對於資料探勘前的資料處理步驟，各有所不同，但主要可區分成四個大階段：(1)目標資料的選取，分析現有的模型，以確認資料探勘應用的領域，進一步設定此模型的目標及評估準則，並逐一考量影響此模型潛藏的因素。此過程將關係資料探勘的成功與否。(2)資料的準備，自龐大的資料庫系統中篩選出所欲分析的資料欄位，並轉換成符合資料探勘工具所需的資料格式。(3)資料探勘工作，為資料探勘的核心，藉由資料探勘演算法，找出隱藏在資料背後的規則、特性、樣式。(4)結果分析，解釋並評估前階段所產生的結果。

在資料探勘的分析功能上，由於每位學者所區分的角度不同，所提出的項數也有所不同，事實上為大同小異，皆概括了資料探勘功能之範圍。這些分析功能通常只能在某種限制下的環境達到有限度的功能，但是在我們所面臨的大部分問題，例：經濟、社會上等，幾乎都可以採用這些分析來解決。參考由 Berry 及 Linoff[1,41]所提出的六種分析功能分別是分類、推估、預測、關聯分組、群集化、描述及視覺化，而資料探勘的技術同樣依不同研究領域可以分為以下幾類[4]，相關內容整理如表 2.2：

1. 統計方法(Statistical Approaches)：資料探勘使用許多的統計工具，包含貝氏網路、迴歸分析、相關分析、群集分析，通常統計模式是經由訓練資料集來建立的，再從模式中找出規則及特徵。

表2.2 資料探勘功能定義與其運用之技術

主要技術	定義
分類 (Classification)	運用訓練組資料的特徵屬性來建立模式，以作為預測未來資料的分類。常用的技術有記憶基礎理解以及決策樹，其演算法有 CART、C4.5、CHAID、CN2 等。此外，類神經網路使用大量簡單而相連的人工神經元來模擬生物神經網路的能力，也由於類神經網路具有高速運算、記憶、學習與過濾雜訊、容錯等能力，因此運用於資料的分類及預測等問題[42]。
推估 (Estimation)	根據既有的連續性數值相關屬性資料，來求得某一屬性未知的值。例如，按照信用申請者之教育程度、行為別來推估其信用卡消費量。所使用的技術有相關分析、迴歸分析及類神經網路。
預測 (Prediction)	根據資料過去的屬性觀察值來推估該屬性未來的值以及趨勢。常用的技術有時間序列、迴歸分析、類神經網路、決策樹等，而模糊理論也常用於預測，應用模糊理論處理資料，可避免因資料的屬性分界而造成尖銳分隔，也可同時處理數值屬性及類別屬性的資料。另外，基因演算法與類神經網路結合也常用於預測分析上[1,41]。
關聯分組(Affinity Grouping)	找出事物彼此之間是否具有關聯性，所使用的技術有購物籃分析，相關的演算法有 Apriori、Partition、Sampling、Dynamic Hashing and Pruning、Dynamic Itemset Counting 等。
群集化(Clustering)	具有相同特徵的資料自動群集在一起，其方法有 K-means Method、Agglomeration Methods 等，相關演算法有 Partitioning Around Medoids、Clustering LARge Applications、Clustering Large Applications based on RANdomized Search。
描述及視覺化(description and visualization)	以不同角度或不同層次上簡單的描繪複雜的資料，例：基本統計量的描述等，透過這種方式來對資料有更多的了解與解釋。常用的技術有購物籃分析、連結分析、決策樹等。

2. 機器學習方法(Machine Learning Approaches)：找出一個最佳的模式來符合測試資料，類似統計方法，但機器學習方法是利用資料自動化學習過程，自動歸納出分類規則及建立模組。常見的方法有決策樹歸納及概念式分群。
3. 資料庫導向方法(Database-oriented Approaches)：資料庫導向方法並不像統計方法與機器學習方法在尋找最佳的模式，而是著重在處理現有的資料，是一種屬性導向歸納，反覆從大量的資料中找出共同的規則與模式。
4. 視覺探索(Visual Exploration)：將多維度的資料轉換呈現為點、線及區域的視覺化物件，讓使用者以動態顯示及探究有興趣的部分，來分析出資料的模式。
5. 其他方法：如約略集合(Rough Set)運用於資料的分類及分群，而類神經網路(Neural Network)則運用於資料的分類及預測。

資料探勘近年來在醫療方面的應用有逐漸受到重視的趨勢，如乳癌的檢查、動脈疾病的風險預測、心臟疾病的診斷及疾病追蹤等。澳大利亞健康保險委員會(Health Insurance Commission, HIC)與 IBM 合作引用資料探勘的分類技術，分析處方之各項資料，偵測發掘不合宜的處方，同時做財務報表的檢查與歷史性的資料比較，提供給醫師做參考。美國 IMS HEALTH 公司的服務包括有處方的研究、銷售業務的管理、市場研究、商業分析、預測技術系統和資訊服務，是目前是世界上健康照顧業資訊解答及諮詢的主要供給者，他們從所收集的巨量資料裡去探索、開採資訊，引用資料探勘的工具從事資料分析後，從以前資料的提供者，成為如今的策略資訊提供者，提供的是醫師用藥的特性行為，並且預測更多的用藥趨勢，疾病處置的行為模式及醫生與病患間的互動模式。

在國內則有林信忠[43]採用類神經網路、區別分析、邏輯回歸分析等技術，針對中醫、西醫、牙醫診所的資料，探討健保醫療費用異常虛報的要因，並建構模式以幫助健保局減少當醫療費用支出。洪哲倫[44]將醫師開藥的處方箋建立成資料庫，再利用資料探勘來找出資源浪費的原因。除此之外，在其它研究中還有依據長期病的病徵來找尋各病徵之間的關係，來幫助醫師做疾病的診斷。甚至還有一個較為先進的應用，就是基因序列的解析。蔣肇慶[45]其研究使用關連性規則，在一個病患的醫令資料倉儲裡，探究有相同疾病的分類碼 ICD-9-CM 的不同病患，找尋相關聯的醫令項目，作為各個病例的基本醫令組，成為適合我國本土化的 AP-DRGs。周賢昭[46]即以時間序列及神經內科中的腦中風病人為資料樣本。在一系列的流程歷史記錄中找出共同的時間相依模式，這個模式能告訴相關

人員如何安排治療活動，以及執行的時間週期。而且依病人診斷資料歸納出法則，可預測病人適合的醫療路徑。

由上述幾個國內外之實例顯現，在醫療資訊的領域裡，若是能善於利用資料探勘的技術與方法，並充分的發揮其資料隱藏性的分析、預測及模型的建立等，是可以作為決策者參考的重要決策分析與指標。

2.4 損傷之長期結果指標

在 2004 年的研究中，對於車禍後損傷的探討較偏重在短期醫療就醫特性，利用 ICDMAP 軟體獲得 AIS 指標，來評估損傷發生後的短時間內，對受害者生命整體的影響。但實際上因事故所造成的傷害，短期醫療雖可恢復大部份的身體生理機能，但也有可能需要長期的復健或是醫療照護，除了對於生活品質造成影響外，也會帶來莫大的社會、家庭及個人的經濟損失。基此，今年再進一步收集有關事故後損傷所造成的長期結果評估指標，以做為未來將損傷評估延伸至長期的參考。

2.4.1 生活品質量表

國際上對於生活品質常用量表做為評估工具，在國外的文獻中常被提及有關健康相關生活品質的量表有六個，包括 SIP(The Sickness Impact Profile)、NHP(Nottingham Health Profile)、QWB(The Quality of Well-Being Scale)、SF-36(The Short-Form-36 Health Profile)、EQ-5D(The European Quality of Life Scale)及 WHOQOL(The World Health Organization Quality of Life Questionnaire)。各量表的發展上，大都從小地區的資料調查發展至大範圍的資料收集(除 EQ-5D 及 WHOQOL 外)，其研究成員的組成主要包括了專業醫生、社會學家、心理學家、經濟學家等，會依量表所欲評估的對象，而挑選不同領域的專家學者。而量表發展的理由從傳統的死亡率、罹病率的測量，到病人日常生活功能的評量都有，其目的希望能了解受測者對於生活因生病所造成的行為能力喪失、社會情感的影響，及個人對於所處文化及價值體系的感受程度等。其評量的範疇相當廣泛，包括了生理健康、心理狀態、獨立程度、社會關係、個人信念、環境等面向。而量表資料的收集則採用由受測人員自填或由訪問人員填寫的方式進行，作答時

間從數分鐘到三、四十分鐘不等[38,39]。

由於量表資料收集的方式，需大量的研究調查人員協助，且所花費的時間相當長，並不適合於本計畫資料之特性。但未來若有其他研究學者，能先用本系統的資料內容，去擷取研究用的資料特徵，再配合問卷的施測，或許能進行更深入的研究。

2.4.2 美國NHTSA的生活功能指標(Functional Capacity Index, FCI)

由於生活品質量表的研究方法是採問卷方式進行，對於本計畫的大量既存電子資料檔的直接應用而言並不適合。而 NHTSA (National Highway Traffic Safety Administration)自 1994 年所開始發展的 FCI(Functional Capacity Index)指標[9]，除了具有可與較側重短期內對生命威脅影響效果的 AIS 評估系統相銜接外，更可進一步將損傷的評估系統擴及長期結果，且近期的文獻資料顯示，FCI 在評估生理的長期功能能力上，為一具有潛力的有效預測指標。

Schluter 等人[17]指出，流行病學的研究需要一種將異質個體群組化成相關的次集合的方法，而隨著損傷的長期結果日益受到重視，須將 AIS 所反應的「對生命的威脅」延伸至能預測未來功能結果(Functional Outcome)的能力喪失(Disability)指標。FCI 便是以此觀念所發展出來的指標，即依據「對功能能力的威脅」來分類損傷的方法。MacKenzie 等人[10]則利用 American Medical Association(AMA)所公布的「Guides To Physical Impairment」方法，量測受傷者的下肢功能喪失百分比，評估 FCI、IIS 及 AIS 等三項指標在估測損傷之長期結果上的表現能力。其中 IIS 是 AAAM 為了評估損傷的長期結果，藉由 35 位不同專長的物理師，定義 0~6 共 7 個等級來描述能力喪失程度(Impairment Level)對功能(Function)的影響。而 FCI 則是以 10 種日常生活功能進行分類(Dimension)，每一種功能各有 A~G 共 7 個等級，再配合 13 位臨床專家及 101 位一般民眾、能力喪失者及相關社團人員進行實驗。先藉由生活品質問卷量表評估生活功能，再由專家評估受測者的生活功能，而發展出 FCI 指標。而為驗證 FCI 之預測能力，研究學者便採用實證的方式，利用現有的生活品質指標、IIS 或 AIS 進行比對。

MacKenzie 等人[12]以「檢視 FCI 的分佈特性」、「FCI 與其它廣泛應用的指標間的相關性」、「FCI 分辨不同型態及嚴重度傷者的能力」等三種方式，來驗證 FCI。研究中利用電話訪問，針對 12 個創傷中心的 1,240 位頭部受鈍傷的損

傷者，進行損傷後 1 年時的生活狀況調查，其中 656 位病患同時再以郵寄問卷方式，完成 SIP 的調查。研究結果指出，FCI 分數與 SIP、SF-36 中的生理健康類分數間的相關性高，而 FCI 顯示依據頭部創傷及其嚴重度來分辨病患時結果較佳。該研究也指出，FCI 的預測結果還需更進一步做測試。

MacKenzie 等人[11]在「The Functional Capacity Index」的簡報資料中指出，在 FCI ver.2 中，進一步將 FCI 等級指派至超過 2,000 個以上的 AIS 損傷描述(AIS 2005 版本)。FCI 中所強調的功能能力(Capacity)僅指生理、認知功能，不包括精神社會(Psychosocial)功能。FCI 與 SIP、SF-36 二者間的生理健康指標間相關性高，而與 SF-6D 的整體相關性也高，在無下肢或頭部損傷、下肢損傷但無頭部損傷等二種狀況時，FCI 與 SF-6D 的指標值則極為接近。Schluter 等人[17]利用澳洲昆士蘭的創傷登記系統，進行 12 個月的電話追蹤調查，以所量測的 FCI 來驗證原始預測的 FCI。研究結果顯示，單一傷處與多重傷處的病患，其預測的 FCI 有明顯差異，有多個 FCI 值的病患，較可能出現所預測的功能喪失現象。在最好的 Max FCI 預測模式中顯示，頭部損傷及多重傷處中包含下肢損傷等二種狀況，明顯與 FCI 觀測值相關，頭部損傷者會有較高 FCI 的可能性為頭部未損傷者的 2.2 倍，而多重傷處中下肢有損傷者會有較高 FCI 的可能性，也為多重傷處中沒有下肢損傷者的 2.2 倍。

由文獻回顧中的確可發現，FCI 除可與 AIS 有效的銜接外，並可應用在長期結果的預測評估上，若配合國外發展的 FCI 軟體，應具有可透過電腦自動處理本計畫中大量醫療資料的可能性。FCI 分析時使用的 AIS 資料，亦可由 2004 年研究計畫所採購的 ICDMAP 軟體利用 ICD-9 編碼計算後取得，故以 FCI 做為長期結果的評估指標，應具可行性。

第三章 道路交通事故相關資料庫

本計畫所使用之道路交通事故相關資料庫，主要有道路交通事故資料庫、死因資料庫、車輛保險(強制險及任意險)資料庫、健保資料庫及人車監理違規資料庫。目前所使用的資料範圍從 88 年至 93 年不等，視各資料庫所取得之內容而定。在 2003 年及 2004 年所進行的研究計畫中，已建立各資料庫之淨化及處理步驟，但本年度所取得的事故與車輛保險資料因資料格式與以往不同，因此對於資料的處理方式亦有做調整。另外，人車監理違規資料的部份，經專家訪談後重新調整資料整併之原則。本章中將對於既有的資料處理方式做簡要的說明(詳細的處理方式請見 2004 年之報告)，並補充說明新增的資料處理方式，以期能對資料的處理有更完整的認識。

本計畫迄今所取得之各類資料庫，其主管機關、資料匯集單位、資料取得年份、資料格式、資料筆數等資料，整理如表 3.1。而不同資料庫之詳細資料內容，可見各節中資料庫之說明。

3.1 道路交通事故資料庫

事故資料庫是登載事故發生後，警察現場處理時針對人、事及物所記錄之內容。在兩種情形下事故資料不會登載在本資料庫中：(1)事故發生後並未報警；(2)屬 A3 類事故，僅有財損並未有人員受傷。本計畫主要是使用事故後有人員受傷或死亡之 A1 類及 A2 類資料為主。目前使用之事故資料庫是由內政部警政署交通組所提供，目前已取得之資料年度為 74 年至 93 年，但本計畫僅使用 88 年後之事故資料，目前研究使用之事故資料原始筆數整理如表 3.2，經淨化處理後之資料筆數如表 3.3。

事故資料目前是以 Access 資料庫之型態儲存，本計畫所使用的資料年份中，88 年至 91 年間是將事故資料分別記錄於人、事及地三個資料表中，但 92 年起所使用之事故資料表，則簡化為基本事故資料與事故當事人兩個，且記錄的資料欄位及變數亦有做修改。為能順利進行事故資料的匯入處理，參考「易肇事地點改善作業手冊之教育訓練計畫」[21]所完成的「事故資料新舊欄位對照表」(如附錄一)，資料整併時是以新的資料庫格式為主，把舊有的資料庫內容透過對照表，將資料欄位進行轉換，但並不轉換資料數值，同時每筆資料中加註其所屬資料定義年份，以作為辨識資料定義年度的依據。

表3.1 本計畫所使用的資料詳細列表

主管機關	資料匯集單位	資料庫名稱	資料年份	取得方式	資料格式	檔案大小(mb)	資料筆數
警政單位	內政部警政署交通組	道路交通事故資料	88-93	原始資料	Access mdb 檔	477	1,104,527
		健康保險資料	西醫門診	88-92	進駐健保局 數據中心作 業後取得。 (註二)	Access mdb 檔	10,720
中醫門診	88-92		1,120	7,497,976			
牙醫門診	88-92		1,155	5,906,312			
住院門診	88-92		318	5,014,067			
藥局	90-92		955	1,288,847			
衛生單位	中央健康保險局	死因資料	88-92	原始資料	純文字檔	48	632,953
		強制險已決賠資料	88-93	原始資料	純文字檔	164	727,583
			92-93			41	416,660
			88-93			620	3,260,003
財政單位	財團法人保險事業發展中心	任意險已決賠資料			純文字檔	3,144	15,478,337
		人車監理違規資料	(註一)			(註三)	18,762
交通單位	中華電信數據分公司	總計					

註一：各工作區所取得之資料時間範圍不盡相同，詳細資料請見表 3.11。

註二：由健保局代為進行資料連結，攜出之資料不包含個人隱私之可辨識內容。

註三：由中華電信數據分公司依本計畫所提供之資料進行連結。

表3.2 研究用各年度事故資料原始筆數

		資料表名稱					
年 度	資 料 類 別	KAM01 (地)	KAM02 (事)	KAM03 (人)	KAM00 (基本事故 資料)	KAM01 (事故當事 人)	合 計 (註)
88	A1	2,487	2,487	6,751	-	-	57,215
	A2	21,847	21,847	50,464	-	-	
89	A1	3,207	3,207	8,297	-	-	123,604
	A2	49,739	49,739	115,307	-	-	
90	A1	3,142	3,142	8,018	-	-	147,258
	A2	60,702	60,702	139,240	-	-	
91	A1	2,725	2,725	7,077	-	-	190,175
	A2	80,626	80,626	183,098	-	-	
92	A1	-	-	-	2,560	6,691	274,034
	A2	-	-	-	117,651	267,343	
93	A1	-	-	-	2,502	6,453	312,241
	A2	-	-	-	134,718	305,788	

註：合計之資料為不同資料表連結後之筆數。

表3.3 事故各年度資料淨化過程筆數刪減統計彙整表

		88 年	89 年	90 年	91 年	92 年	93 年
原始資料總筆數		57,215	123,604	147,258	190,175	274,034	312,241
資料孤兒		9	16	13	2	0	0
身分證號淨化	資料空白	7,014	12,878	14,103	15,672	19,799	22,118
	長度不合	66	241	311	366	601	474
	第 1 碼錯誤	2	94	119	191	244	139
	第 2 碼錯誤	1	4	22	71	407	962
	3-10 碼錯誤	0	0	2	0	0	1
	編碼錯誤	189	373	433	141	816	1,210
車牌號碼淨化	空白	3,363	8,154	10,290	12,307	16,834	18,918
	長度小於 5	18	110	27	69	180	66
	特殊符號	*	0	0	0	0	0
		?	0	0	2	1	0
		.	0	15	2	0	0
		@	0	0	0	0	0
		`	1	0	0	4	0
	#	0	0	0	0	0	0
日期淨化		0	0	0	0	0	0
剩餘筆數		46,552	101,719	121,934	161,351	235,153	268,353
刪除百分比(%)		18.64	17.71	17.20	15.16	14.19	14.06
合計筆數		935,062					

事故資料使用前所需的處理步驟及內容簡要說明如下：

1. 事故流水號

一筆事故資料會依資料內容分別記載在三個不同的資料表中，為能清楚了解每一事故當事人所處之事故發生情形，因此必須透過一個唯一碼，將三個資料表關聯成一個完整的資料表。在原事故資料中所記載的事故流水號(六碼)，會因新年度而重新編碼，甚至同一年度會有重複使用之情形，因此無法做為資料連結使用。在本計畫是使用事故流水號及事故發生年月(共十碼)，產生新的事故流水號。

2. 資料孤兒

在資料鍵入過程中有可能因為重複輸入或其他原因，而造成某一資料表中資料重複存在之情形。當利用新的事故流水號進行資料表連結過程中，會有無法連結之情形，此即為資料孤兒。此資料無法進行更進一步之使用，因此必須將其剔除。

3. 身分證號及車牌號碼淨化

依先前研究所建立之淨化原則，將有問題之身分證號及車牌號碼資料剔除，以供後續進行確定性連結使用。

4. 事故唯一發生人次

在本計畫中主要是由事故資料做為出發點，連結不同性質資料庫，再依連結後之資料進行分析研究。但在事故資料庫中曾有同一人在不同時間發生事故，若以此人資料進行連結時，可能會發生重複連結之情形，而不易正確辨識資料連結時間點是否正確，因此在研究中只使用僅發生一次事故之資料。

3.2 死因資料庫

死因資料是由行政院衛生署所提供，本計畫使用之各年度死因資料原始筆數整理如表 3.4，經淨化處理後之資料筆數如表 3.5。目前提供的死因資料是為純文字檔，且新年度的資料格式並未有改變，因此原有的維護系統中即可進行資料的轉換工作。目前死因資料使用前，所需的處理步驟及內容簡要說明如下：

1. 身分證號淨化

所取得之死因資料中，身分證號存有編碼錯誤的問題，但 90 年後之資料已較以往改善。

2. 身分證號重複

在我國原則上每個人的身分證號應是唯一的，但前期研究在死因資料庫中曾發現同一身分證號卻有兩個不同的死亡日期。為避免造成後續資料連結之問題，

因此將研究用期間內有重複之身分證號剔除。

表3.4 研究用各年度死因資料原始筆數

年度	筆數
88	124,991
89	124,481
90	126,667
91	126,936
92	129,878
總計	632,953

表3.5 死因各年度資料淨化過程筆數刪減統計彙整表

		88 年	89 年	90 年	91 年	92 年
原始資料總筆數		124,991	124,481	126,667	126,936	129,878
身分證號淨化	資料空白	0	0	0	0	0
	長度不合	0	0	0	0	0
	第 1 碼錯誤	0	0	0	0	0
	第 2 碼錯誤	0	0	0	0	0
	3-10 碼錯誤	3	4	0	0	0
	編碼錯誤	3,017	3,118	138	56	226
日期淨化		0	0	0	0	0
剩餘筆數		121,971	121,359	126,529	126,880	129,652
刪除百分比(%)		2.42	2.51	0.11	0.04	0.17
剩餘筆數		626,391				

3.3 車輛保險資料庫

車輛保險資料是由財團法人保險事業發展中心所提供的已決賠資料，其資料格式為純文字檔，可透過本計畫所發展之資料維護整合系統進行資料匯入及資料淨化等處理工作。在 2003 年的研究中嘗試使用強制險已決賠資料與事故資料進行連結，但成功連結之資料比例不如預期。在今年研究進行期間，獲知另有登載

所有受害人資料的理賠記錄，此應可提高資料連結之成功率。今年取得之 92、93 年新格式的車輛保險已決賠及強制險受害人資料（格式如附錄二）的匯入情形說明於下，而有關資料連結率的結果探討則請見 5.2 節。

本計畫取得之車輛保險資料年度從 88 年至 93 年，各年度車險之資料原始筆數整理如表 3.6，淨化後之資料筆數如表 3.7、表 3.8、表 3.9。

表3.6 各年度車險原始資料筆數

年度	強制險		任意險
	車險賠案	車險受害人賠案	車險賠案
88	78,587	-	507,683
89	117,699	-	530,267
90	126,067	-	543,822
91	136,028	-	576,848
92	127,661	195,639	524,778
93	141,541	221,021	576,605
總計	727,583	416,660	3,260,003

表3.7 強制險受害人各年度資料淨化過程筆數刪減統計彙整表

		92 年	93 年
原始資料總筆數		195,639	221,021
身分證號淨化	資料空白	0	0
	長度不合	1,163	1,137
	第 1 碼錯誤	317	390
	第 2 碼錯誤	417	761
	3-10 碼錯誤	1	12
	編碼錯誤	872	959
剩餘筆數		192,869	217,762
刪除百分比(%)		1.42	1.47
合計筆數		410,631	

表3.8 強制險各年度資料淨化過程筆數刪減統計彙整表

			88 年	89 年	90 年	91 年	92 年	93 年
原始資料總筆數			78,587	117,699	126,067	136,028	127,661	141,541
身分證號淨化	資料空白		1	0	33	32	11	9
	長度不合		136	189	254	263	242	267
	第 1 碼錯誤		44	109	107	142	109	101
	第 2 碼錯誤		0	3	47	40	54	173
	3-10 碼錯誤		2	2	0	2	0	2
	編碼錯誤		550	896	794	796	625	618
車牌號碼淨化	空白		626	936	1,249	1,581	1,888	2,157
	長度小於 5		21	22	99	104	109	134
	特殊符號	*	2	11	21	17	16	28
		?	2	0	0	0	0	0
		.	1	0	3	3	8	14
		@	0	0	0	0	1	2
		`	1	4	2	1	1	2
		#	0	0	0	0	0	0
日期淨化			4	3	2	0	0	0
剩餘筆數			77,197	115,524	123,456	133,047	124,597	138,034
刪除百分比(%)			1.77	1.85	2.07	2.19	2.40	2.48
合計筆數			711,855					

表3.9 任意險各年度資料淨化過程筆數刪減統計彙整表

			88 年	89 年	90 年	91 年	92 年	93 年
原始資料總筆數			507,683	530,267	543,822	576,848	524,778	576,605
身分證號淨化	資料空白		3	4	0	1	4	4
	長度不合		1,220	1,228	1,174	1,265	1,243	957
	第 1 碼錯誤		504	556	640	810	522	491
	第 2 碼錯誤		326	39	79	6,476	8,906	10,976
	3-10 碼錯誤		3	3	1	11	1	6
	編碼錯誤		4,952	4,962	4,188	4,105	3,229	3,709
車牌號碼淨化	空白		5,942	5,054	4,776	7,221	7,809	9,725
	長度小於 5		550	420	213	306	255	261
	特殊符號	*	145	155	145	140	399	1,186
		?	0	0	2	2	0	1
		.	10	43	318	223	256	785
		@	3	1	1	0	14	40
		`	5	5	10	4	7	8
		#	2	0	2	2	2	9
日期淨化		7	4	2	4	1	2	
剩餘筆數		494,011	517,793	532,271	556,278	502,130	548,445	
刪除百分比(%)		2.69	2.35	2.12	3.57	4.32	4.88	
合計筆數		3,150,928						

3.3.1 88-91年車輛保險資料

在 2003 年的研究計畫中，針對資料的使用已建立基本處理原則，整理如下：

1. 身分證號

在車輛保險資料中有「肇事駕駛人」及「被保險人」兩欄的身分證號，其中「被保險人」記載的是車輛所有人，因所有人可能為法人或是自然人，故不進行資料淨化，但可做為車輛所有人種類之辨識用。因此資料淨化過程中僅淨化「肇事駕駛人」欄位之資料，並做為與其他資料庫連結用之欄位。

2. 車牌號碼

因車牌號碼並無明確之編碼或檢核原則，基本上僅剔除空白、長度不足及含有特殊字元之資料。該欄位是除了身分證號外，可進行確定性連結的欄位。在 5.2 節的內容中將探討使用車牌號碼連結事故資料之情形。

3. 日期

在車輛保險已決賠資料中與日期相關之欄位包括賠案受理日、結案日期、肇事駕駛人出生年、被保險人出生年、出險日期、理賠/追償日期等。在使用上主要是以出險日期為主，理論上出險日應為事故發生日，可做為判斷是否為事故後之出險記錄使用。

在 2004 年的研究中為測試「事故關聯資料分析與評估系統」輸出之資料能否提供車輛保險相關之研究學者使用，並將使用經驗回饋做為資料連結系統之修改建議。在資料處理過程中，所面臨最大之挑戰為同一事故多次理賠記錄之整併工作，在整併過程中發現同一理賠案件中，相同欄位卻有不同之記載內容，致需進行許多整併判斷，例如：性別資料之整併，是參考資料所登載之身分證號，使用身分證號第二碼做為性別資料等。資料整併之問題，在今年取得之 92-93 年資料中，已由保發中心先進行整併，將可避少後續資料整併時之問題。

3.3.2 92-93年車輛保險資料

在本年度所取得之車輛保險資料主要包括「強制車險賠案賠款資料」、「強制車險受害人賠款資料」及「任意險賠案賠款資料」，賠案賠款資料採一筆賠案一筆記錄之方式處理，詳細之資料表格式請見附錄二。在資料的匯入處理上，因為其資料格式與以往不同，必須重新修改資料匯入功能，因此在使用整合系統的「資料維護整合系統」功能時，須特別留意資料年度。較特別之資料處理內容說明如下：

1. 資料格式

雖資料仍為純文字檔格式，但新取得之資料第一行包含了欄位名稱，因此在使用前須先加以剔除。另外，先前的格式是採固定長度之儲存方式，即便是空白資料仍補上空白鍵，使得每一筆資料的長度均相同。資料轉入時僅需留意長度及中文字雙位元之問題即可，目前轉換程式已具備中文字雙位元長度處理之功能。

新取得之車輛保險資料中，不同欄位間之資料是使用逗號(,)隔開，部份欄位之資料還使用雙引號(")包覆，且每筆資料之長度亦不固定，因此在處理時須特別留意。目前的程式功能中已能處理新資料格式之轉入。

2. 異常資料內容

在進行逐筆資料轉入時，是使用逗號(,)做為欄位分隔符號，但實際轉換過程中發現，資料欄位中有可能包含逗號，會造成資料轉入之問題，須事先使用人工之方式進行處理。

3.4 健保資料庫

今年因為健保局內部組織調整，致本計畫無法依 2004 年的進駐作業方式，繼續更新資料。故以下僅摘錄以往資料處理結果及內容做為參考。

本計畫取得之 88 年至 92 年各類健保資料筆數整理如表 3.10。主要的處理步驟及內容簡述如下：

1. 就醫間隔及就醫順序

為能進行就醫特性之分析，因此利用相鄰兩次的就醫日期，計算就醫間隔。而就醫順序則是以事故發生日與就醫日之間隔，配合健保卡就醫序號等做計算而得。

2. 住院資料整併

在醫療院所向健保局申報住院資料時，有可能因連續(長期)或申報日期的關係，將同一次住院資料分成多次申報，為能更清楚住院之情形，因此須將此類連續住院資料予以合併。健保局對於住院資料的計算是採「算進不算出」之方式，因此連續住院者理論上會填寫相同之入院日，但住院期中出院日期保持空白，直至真正出院為主，因此要將此類資料予以合併。資料處理過程曾發現，同一病人在同一家醫院同一天出院、同一天入院，理應為同一次住院，但實務上仍有非連續住院之情形，故此類情形不納入連續住院之整併。

3. 資料合併

在連續住院之資料整併部份，會將住院期間的醫療費用、病床使用天數等進行加總處理，但部份資料則保留第一筆及最後一筆記錄，如診斷碼等。

表3.10 目前已取得之健保資料筆數

資料庫	資料年份	原始資料筆數	日期淨化	剩餘資料筆數	資料整併	整併後資料筆數
西醫門診	88 年	1,933,222	383	1,932,839	-	1,932,839
	89 年	3,225,852	176	3,225,676	-	3,225,676
	90 年	20,110,822	10	20,110,812	-	20,110,812
	91 年	17,031,395	10	17,031,385	-	17,031,385
	92 年	15,265,138	19	15,265,119	-	15,265,119
中醫門診	88-92 年	7,498,034	58	7,497,976	-	7,497,976
牙醫門診	88-92 年	5,906,457	145	5,906,312	-	5,906,312
藥局門診	88-92 年	5,014,072	5	5,014,067	-	5,014,067
住院	88-92 年	1,355,754	0	1,355,754	66,907	1,582,326

3.5 人車監理違規資料庫

違規資料是由中華電信數據分公司所負責維護，本計畫在 2004 年研究中，以提供身分證號與車牌號碼委請中華電信數據分公司代為勾稽資料的方式，取得研究所需資料，於進行初步處理後，使用小範圍的處理後資料確認資料應用的可行性。今年除了透過訪談中華電信數據分公司負責處理違規資料之人員，以更清楚地瞭解人車監理違規資料實際處理之情形，及對於資料淨化與整併建立有更明確之處理方法外，並加入後續取得之台北市最近一次歸檔前的人車監理違規資料，重新整理本計畫的人車監理違規資料。

本計畫取得之資料範圍如表 3.11 所示。至於新年度車禍涉案者的人車監理違規資料，因勾稽作業進行時程較晚，故納入未來處理。目前國內人車監理違規資料劃分為台北市、高雄市、台北區、新竹區、台中區、嘉義區和高雄區等共七個區域，各區所包含之管轄範圍如表 3.11 所示。而人車監理違規資料分別由各區管理資料取得，但由於各區歸檔時間不盡相同，可取得之資料僅能由最近歸檔

日至取資料日為止，因此各分區所取得之資料時間範圍也不盡相同。

表3.11 目前所取得人車監理違規資料範圍

區域屬別	區域範圍	歸檔時間	取得資料時間範圍
台北市	台北市(共取得兩次)	91/09/01	91/09/01~93/05/01
高雄市	高雄市	88/01/01	88/01/01~93/05/01
台北區	花蓮縣、宜蘭縣、基隆市	90/09/01	90/09/01~93/05/01
新竹區	桃園縣、新竹縣、苗栗縣	90/01/01	90/01/01~93/05/01
台中區	台中縣、彰化縣、南投縣	90/01/01	90/01/01~93/05/01
嘉義區	雲林縣、嘉義縣、台南縣	91/01/01	91/01/01~93/05/01
高雄區	高雄縣、屏東縣	90/09/01	90/09/01~93/05/01

3.5.1 資料整理

本計畫所使用之人車監理違規資料，係由提供給中華電信數據分公司之「身分證號」及「車牌號碼」所連結之資料，連結後之資料分別稱為「人違規資料」與「車違規資料」。

1. 資料匯入

中華電信數據分公司所提供之人車監理違規資料為逐筆之純文字檔資料(資料格式如附錄三)，在使用此資料前須先利用程式將資料轉入資料庫中。在進行資料轉入前，須先確認文字檔之格式為 PC 或 Unix 系統之文字格式，若為 Unix 系統之文字檔格式，則需先將其轉換成 PC 格式，否則將造成程式在拆解字串時發生處理錯誤之情形。另外，在資料轉入後需核對資料筆數是否一致，在實際轉檔過程中，曾發現因文字檔中隱藏部份不可視字元，而造成字串拆解錯誤，因此轉換時需先予以替掉此不可視字元。

2. 工作區代碼

在資料來源工作區的記錄上，為保有後續資料的擴充性及方便回溯合併前的來源資料檔，因此將對資料來源重新做編碼記錄。目前所規劃之代碼共分為三碼，第一碼為工作區代碼，由 A~G；第二碼 1 表人違規，2 表車違規；第三碼表

資料取得之順序。如代碼為 A11 表台北市人違規第一次取得之資料，A22 代表台北市車違規第二次取得之資料，各分區之代碼及資料筆數整理如表 3.12 所示。

3. 違規唯一碼

由於警察登記的違規單號並非唯一值，因此使用「違規資料單號」(TKT_NO)與「車牌號碼」(PLT_NO)產生「違規唯一碼」，以此做為違規唯一辨識值。

4. 大量資料之處理方式

目前所取得之人車監理違規資料共約有一千五百餘萬筆，為縮短資料轉入資料庫、資料計算及資料查詢之速度，在進行資料匯入時，可採用多台伺服器同時轉入處理，待完成資料匯入後再透過網路將資料彙整至一部較高等級之伺服器中。而為保留後續計算各分區及歷次取得資料之淨化處理情形，因此使用先前定義好之工作區代碼做為資料表格名稱。

後續在進行跨表格查詢時，透過檢視表(VIEW)(或稱為虛擬表格)的功能，將不同用途之資料表建立共用的虛擬表格，可減少資料重複放置之問題，也可以縮短資料庫執行運算處理之時間。為加快資料查詢之速度，過程中除適當的為資料表建立索引外，亦視需求將資料表拆解成數十個資料量較少之資料表，由目前的使用經驗當資料表內容約為十萬筆資料時，可獲得不錯之處理速度，亦可避免產生過多次資料表之問題。因此本計畫會視需求彈性拆解資料表，並新增一個流水號欄位做為主索引。

表3.12 人車監理違規資料各工作區資料筆數統計

區域屬別	代碼	取得順序	人違規資料筆數(1)	車違規資料筆數(2)
台北市	A	1	477,682	2,189,144
		2	99,166	1,364,495
高雄市	B	1	370,386	1,076,969
台北區	C	1	492,194	1,375,307
新竹區	D	1	566,935	1,738,585
台中區	E	1	614,480	1,382,240
嘉義區	F	1	655,557	1,584,873
高雄區	G	1	410,425	1,079,899
總計			3,686,825	11,791,512

3.5.2 資料淨化

在完成資料轉檔工作後，為確保後續使用資料之正確性，因此需先針對重要欄位(如單號、車牌號碼及結案代碼等)進行淨化，以剔除部份不合理或有問題之資料，被剔除的資料則保留在未來進行探討。本計畫透過訪談中華電信數據分公司之人員所釐清的問題整理如下。

1. 罰則或結案代碼不一致之情形

人車監理違規資料是由七個工作區各自獨立運作的一種分散式動態資料，各依該區監理單位的違規罰單處理狀況來登錄資料，例如：對於處罰內容的認知不同時，便會修改違規罰單內容，或是在不同的處理階段，便會動態登錄違規罰單狀態等，凡此均會造成同一筆資料在不同工作區有不同的記錄內容，包括：同一筆違規，卻有兩個不同的罰則及結案代碼。

2. 資料轉檔所造成資料不一致之情形

不同工作區彼此間在進行資料轉檔時，是以部份欄位資料進行轉換，其餘欄位再由各工作區原有之資料進行回補，因而造成不同工作區中同一筆違規記錄出現部份資料不相同之情形。

3. 申覆所造成資料不一致之情形

違規受罰人若覺得罰單內容有誤，可向監理單位提出申覆，如車子違規被照相舉發，但只要車主澄清駕駛者是他人，填寫切結書後，即可辦理罰單移轉。因此有可能因申覆的結果，而改變了違規罰單之內容。原則上各工作區在進行違規罰單資料異動時，同一工作區內僅會保留一筆違規罰單記錄，會保留最新的異動資料，包括異動時間及處理人員代碼等。資料異動同時會去更新其他工作區之資料，以盡可能維持各工作區間資料的同步。

4. 行政相關作業所造成資料不一致之情形

違規罰單可能因行政作業之處理，如該違規資料與該工作區之資料管理已無關係時，便將資料移至其他工作區或是移交原告發單位，並變更資料狀態。但實務上有可能案件送回原告發單位後，警察還是認為要處罰，所以同一案件會再送回監理單位重新建檔，因此產生資料不一致或是資料重複之情形。

另外，法規條文會依現實情況而有所編修，對此中華電信相關資料處理人員的建議是保留到「條」的部份即可。除法規的編修外，隨著時代的變化，對於罰款繳納的方式也有不同，據了解現在可對罰款採分期付款的方式繳納，而此便會

產生結案或未結案兩種記錄形式。

5. 資訊系統所造成資料不一致之情形

早期因電腦軟硬體平台不若現在完整，因此較容易發生資料重複鍵入及資料鍵入錯誤之情形，因而容易產生資料不一致之情形。目前為求一致化的作業，對於單號的編碼已著手進行全國一致化的處理。且為維持各工作區資料的一致性，每隔一段時間各工作區間的資料便會進行交叉比對，將資料內容一致化。

本計畫依專家訪談結果所訂定之淨化原則整理如下：

1. 剔除欄位空白資料

為確定每件違規事件的唯一辨識值，使用「違規資料單號」(TKT_NO)與「車牌號碼」(PLT_NO)產生「違規唯一碼」，因此剔除單號及車牌號碼空白之資料。另外，結案代碼亦為記錄違規內容之重要資訊，若有空白者亦一併剔除。

2. 剔除結案代碼為「刪除」“E”者

結案代號欄位之代碼為“E”者為輸入錯誤之資料，因此刪除此類資料。

3. 剔除結案代碼為「移出」“C”或「文移」“M”者

此兩類代碼表該筆資料與該工作區之資料管理已無關係，如將資料移至其他工作區或是移交原告發單位。但實務上有可能案件送回原告發單位後，警察還是認為要處罰，所以同一案件會再送回監理單位重新建檔。對於此類的資料，除「結案代碼」外，其他欄位進行整併。

4. 剔除結案代碼為「免罰」“F”或「先免」“Y”者

此類資料有可能送交裁決後不處罰，所以無違規之問題，因此與此類結案代碼相同違規唯一碼之資料均予以剔除。

經前述條件淨化後之資料筆數記錄如表 3.13 所示，目前淨化後之資料筆數共有 15,016,600 筆。

表3.13 人車監理違規資料淨化記錄表

監理區處		原始資料筆數		剔除空白欄位								剔除結案代號為 E		剔除結案代號為 C	
				單號		車牌號碼		結案代號							
名稱	代碼	人違規	車違規	人違規	車違規	人違規	車違規	人違規	車違規	人違規	車違規	人違規	車違規	人違規	車違規
台北市	A	477,682	2,189,144	0	0	143	0	1	1	1,617	8,147	1,972	15,676		
		99,166	1,364,495	0	0	114	0	0	0	221	2,086	829	8,981		
高雄市	B	370,386	1,076,969	0	0	1,413	0	0	0	1,087	7,564	5,610	15,397		
台北區	C	492,194	1,375,307	0	0	431	0	0	1	1,414	7,427	1,758	1,809		
新竹區	D	566,935	1,738,585	0	0	19	0	0	1	1,586	10,378	5,393	10,814		
台中區	E	614,480	1,382,240	0	1	208	0	0	53	1,828	11,507	1,306	4,383		
嘉義區	F	655,557	1,584,873	0	0	202	0	2	5	1,617	14,321	4,355	9,787		
高雄區	G	410,425	1,079,899	0	1	4,585	0	1	11	7,209	17,973	10,692	19,202		
總計		3,686,825	11,791,512	0	2	7,115	0	4	72	16,579	79,403	31,915	86,049		

表3.13 人車監理違規資料淨化記錄表(續)

監理區處		剔除結案代號為 M		剔除結案代號 F 的同一違規記錄		剔除結案代號 Y 的同一違規記錄		剩餘資料筆數	
名稱	代碼	人違規	車違規	人違規	車違規	人違規	車違規	人違規	車違規
台北市	A	43	73	2,447	41,624	43	369	471,416	2,123,254
		2	16	402	9,431	0	0	97,598	1,343,981
高雄市	B	274	7,741	7,057	25,023	2	12	354,943	1,021,232
台北區	C	517	6,566	2,246	11,261	12	252	485,816	1,347,991
新竹區	D	1,341	6,077	7,237	22,635	1	56	551,356 ^註	1,688,624
台中區	E	388	5,178	2,462	8,587	2	9	608,286	1,352,522
嘉義區	F	679	10,612	4,533	27,564	2	111	644,167	1,522,473
高雄區	G	442	14,328	1,523	11,359	2	54	385,971	1,016,971
總計		3,686	50,591	27,907	157,484	64	863	3,599,553	11,417,047

註：D11 中有兩筆車牌資料含「，」，會造成 SQL 指令處理之問題，故先將其記錄移至 MVD_ERR_PLT_NO 中，剩餘資料再扣 2。

3.5.3 資料整併

本計畫所使用的違規資料不但是來自七個不同工作區的動態資料，而且是分別使用身分證號與車牌號碼所連結出來之資料，因此資料內容上必會存有違規唯一碼重複之問題，致需要將同樣違規唯一碼之資料進行合併。在彙整七個工作區共 16 個資料檔後，利用違規唯一碼做為計算是否為同一筆資料。由表 3.14 之統計結果可知，違規唯一碼出現一次之百分比已達 81.44%，若含違規唯一碼出現兩次時，則已高達 98.38%。因此若在尚未能確認資料合併處理原則時，可單就僅出現一次之違規資料進行連結分析。若要進行資料合併時，可先就違規唯一碼出現兩次者進行合併，以減少合併時所需之處理判斷原則。

依據中華電信數據分公司之資料處理經驗為，合併時結案代碼為「已結」狀態所代表的違規處理完成度高於「部結」，而「部結」高於「未結」；資料完整性的部份，以結案代碼為「結案」「B」、「郵撥」「Q」、「吊扣」「H」及「併案」「G」所記載之違規資料最為正確，當與其他資料進行合併時，以此四個結案代碼為最優先，當有空白資料時再用其他資料回補；若結案代碼非「結案」「B」、「郵撥」「Q」、「吊扣」「H」及「併案」「G」時，則以最新之結案狀態為主，仍維持「已結>部結>未結」之原則。對於結案狀態之分類整理如表 3.15。

本計畫的違規資料整併是以違規唯一碼出現兩次之資料為主，而為了解整併過程中結案代碼的變化情形，茲就有出現之結案代碼進行將整併前與整併後筆數之變化整理如表 3.16。由表中可發現，整併前代碼較多之前五名依序為，結案(B)、郵撥(Q)、銀代(P)、先結(J)及吊扣(H)。而整併後代碼較多之前五名依序為：結案(B)、郵撥(Q)、先結(J)、銀代(P)及吊扣(H)，排名上僅有先結與銀代對調，且此前五名所佔之資料筆數已達 96%以上。

表3.14 違規唯一碼出現次數統計表

違規唯一碼 重複次數	筆數	影響資料 筆數	累計影響 資料筆數	影響筆數百 分比(%)	累積影響筆 數百分比(%)
1	12,229,848	12,229,848	12,229,848	81.44	81.44
2	1,271,434	2,542,868	14,772,716	16.93	98.38
3	61,044	183,132	14,955,848	1.22	99.60
4	13,153	52,612	15,008,460	0.35	99.95
5	504	2,520	15,010,980	0.02	99.96
6	722	4,332	15,015,312	0.03	99.99
7	64	448	15,015,760	0.00	99.99
8	12	96	15,015,856	0.00	100.00
9	5	45	15,015,901	0.00	100.00
10	21	210	15,016,111	0.00	100.00
11	7	77	15,016,188	0.00	100.00
12	21	252	15,016,440	0.00	100.00
13	1	13	15,016,453	0.00	100.00
14	4	56	15,016,509	0.00	100.00
15	1	15	15,016,524	0.00	100.00
18	2	36	15,016,560	0.00	100.00
19	1	19	15,016,579	0.00	100.00
21	1	21	15,016,600	0.00	100.00
總計	13,576,845	15,016,600	-	100.00	-

表3.15 結案代碼分類整理

結案分類	代碼名稱	代碼	代碼名稱	代碼
未結	列管	A	申訴	U
	單退	S	異議	V
	債憑	R	裁退	T
部結	部結	K	先部	2
已結	結案	B	先免	Y
	郵撥	Q	吊銷	I
	銀代	P	溢繳	7
	先結	J	先扣	1
	免罰	F	拍賣	N
	移出	C	退款	5
	吊扣	H	先併	Z
	先郵	X	移入	D
	文移	M	應移	8
	併案	G	強執	f
	先代	3	吊緩	h
	先他	4	他收	p
	通信	q	先通	x

表3.16 整併前後結案代碼筆數之變化

結案分類	代碼名稱	代碼	整併前		整併後	
			筆數	百分比(%)	筆數	百分比(%)
未結	列管	A	48,297	1.90	23,512	1.85
	單退	S	4,121	0.16	2,019	0.16
	債憑	R	245	0.01	122	0.01
	申訴	U	2	0.00	1	0.00
部結	部結	K	963	0.04	443	0.03
	先部	2	99	0.00	18	0.00
已結	結案	B	833,457	32.78	428,223	33.68
	郵撥	Q	802,691	31.57	402,699	31.67
	銀代	P	431,869	16.98	98,622	7.76
	先結	J	241,572	9.50	232,459	18.28
	吊扣	H	145,382	5.72	73,439	5.78
	先郵	X	15,950	0.63	2,547	0.20
	併案	G	595	0.02	302	0.02
	先代	3	1,680	0.07	251	0.02
	先他	4	2,392	0.09	19	0.00
	吊銷	I	1,749	0.07	876	0.07
	溢繳	7	17	0.00	8	0.00
	先扣	1	31	0.00	12	0.00
	拍賣	N	10	0.00	5	0.00
	退款	5	8	0.00	4	0.00
	先併	Z	102	0.00	51	0.00
	移入	D	13	0.00	6	0.00
	應移	8	1	0.00	0	0.00
	積註	L	11,622	0.46	5,796	0.46
總計			2,542,868	100.00	1,271,434	100.00

第四章 道路交通事故相關資料整合系統之建置

4.1 資訊系統的建置

本計畫自 2003 年開始執行至今三年期間，已歷經三個不同的研究階段，各階段之系統建置內容整理如表 4.1。第一階段著重於各資料內容的了解及連結測試，並依連結結果進行不同應用主題之研究；第二階段以上一階段之研究經驗為基礎，為建立標準化的資料處理流程，及因應更大量的新增資料，嘗試建立標準的資料處理及連結流程；第三階段則延續前二年之成果，除改善現有系統之問題及功能擴充外，並參考國外學術研究單位所建置之線上查詢系統，將經標準流程處理後之資料，在確保個人隱私之前提供，提供線上查詢服務。系統歷經三階段的開發，在軟硬體系統的建置上，也必須隨之調整。尤其是在第三階段所提供的線上查詢，如何做好資訊安全亦是相當重要之資訊工程。

計畫執行初期，對於資訊系統的建置，從軟硬體的建置成本、擴充性、穩定性、相關資源及後續維護等角度進行評估。大型主機雖然其穩定性高，但是其軟硬體建置成本相當昂貴，所搭配的 Unix 作業系統，非一般人所熟悉，須具有專業資訊背景者才能操作，容易造成後續系統維護之問題。而採 PC Base 的 Linux 系統，雖然其穩定性佳、系統建置費用便宜，但應用系統的開發不若視窗軟體的便利及友善，及系統維護上亦非一般人員即可進行。因此決定採用 PC Base 的硬體系統及視窗(Windows)開發工具為主。本節中將就不同年度所建置的資訊系統進行說明介紹。

4.1.1 硬體系統

依各年度的工作重點及需求，對於硬體系統的建置可區分為三大階段，各階段內容分述如下：

第一階段：雛型系統建置

在不完全了解研究用資料內容，及資料處理過程中所可能產生的資料膨脹情形下，先採購低階伺服器，主要做為資料淨化、資料連結測試使用。在研究過程中，該設備亦曾做為進駐健保局時資料處理使用。

表4.1 系統建置歷程說明

	2003 年 雛型系統建置期	2004 年 資料維護系統建置期	2005 年 線上查詢系統建置期
系統設計概念	採單一伺服器架構，供資料儲存、資料淨化及資料連結使用。	增加專屬資料連結及處理伺服器，方便進駐資料提供單位，進行資料處理及連結使用。	採三層式架構建置線上查詢系統，供線上查詢所開放之資料內容。
硬體系統建置	添購低階伺服器	添購中階伺服器	1. 添購超薄型網站伺服器 2. 添購磁碟陣列資料備份系統 3. 添購不斷電系統
軟體設計重點	1. 彙整各來源資料檔至資料庫中 2. 開發資料淨化之功能模組 3. 開發健保資料整併模組 4. 提供連結後資料進行研究分析使用	1. 修改前一年度之程式功能模組，改善資料處理效能，擴充為資料維護整合系統。 2. 設計資料連結雛型系統，制定各資料連結之標準流程。 3. 提供其他領域專家學者研究用資料。	1. 擴充資料維護整合系統功能。 2. 修改資料連結雛型系統。 3. 開發 Web 線上查詢系統功能，以樞紐分析之方式呈現統計資料。
系統開發工具	Visual Basic、SQL Server、ADO	Visual Basic、SQL Server、ADO	Visual Basic、SQL Server、ADO、HTML、ASP

在完成該年度之研究後，發現資料本身所佔之容量並不大，但為特定研究用途所建立的資料檔，會因為運算、連結等處理過程，而讓資料檔快速成長，因此需要更大容量的硬碟空間供資料儲存使用。另外，在進駐健保局進行資料處理期間發現，所欲處理的健保資料量相當龐大，現有電腦之負擔已相當沈重，往後若要進行更多年度資料的處理，應準備更高等級的電腦配備。

第二階段：資料維護系統建置

參考第一階段的使用經驗，添購一台中階伺服器，專供進駐其他資料提供單位進行資料處理使用。該伺服器器具備雙中央處理器、高容量之硬碟與記憶體及磁碟陣列等功能，對於大量資料之處理運算，已具備較佳之處理效能。

不同資料庫連結後之資料，係直接存放在硬碟中，雖然使用上相當便利，但卻存在硬碟可能毀損之風險。為方便資料的即時取用及安全保護，規劃添購磁碟陣列及不斷電系統，以對研究用資料進行更完善之防護。

第三階段：線上查詢系統建置

在第三階段的系統建置過程中，主要添購之硬體設備包括：磁碟陣列、不斷電系統及外接式備份系統。而為配合本年度研究計畫所規劃開發的線上查詢系統，再添購一台專屬網頁伺服器，並與運輸安全專業網站結合，供特定會員線上查詢資料使用。為確保資料的安全，該網頁伺服器與其他伺服器採取實體隔離方式，並不直接進行網路連線，僅讀取由其他伺服器使用光碟片所儲存之資料。

4.1.2 軟體系統

本計畫軟體系統之建立主要是以微軟的視窗環境為主，包括伺服器作業系統(Windows 2003 Server)、資料庫系統(MS SQL Server)、應用系統開發工具(Visual Basic)、資料庫存取函式庫(ActiveX Data Object, ADO)、網頁互動式開發工具(Active Server Page, ASP)等，各軟體的使用情況說明如下。

1. Windows 2003 Server

電腦運作的作業系統，即應用系統的工作環境。包括 MS SQL Server 及本計畫所開發的應用程式，均架構在其上執行。

2. MS SQL Server

屬於中大型資料系統，在本計畫中主要做為各資料儲存、連結及查詢使用。系統開發過程中，會視需求使用部份資料庫專有的操作功能，如檢視表(View)，以方便系統開發。

3. Visual Basic

為資料維護整合系統及資料連結系統最主要的開發工具，藉由視窗化的物件可設計出友善的使用者界面，而物件模組化的程式碼，則方便程式功能的擴充。

4. ADO

此為微軟所提供與資料庫連結的函式庫，支援單機及網路資料庫存取功能，可彈性修改資料庫相關連結參數，變更資料庫連結位置。藉由內容的資料庫存取功能，配合 Visual Basic 或 ASP 可開發與資料庫連結之互動式系統。

5. ASP

主要做為開發線上查詢系統使用，收集使用者的查詢變數，再送交資料庫系統進行查詢，將查詢後的結果利用標準 HTML 做資料呈現，即可完成互動式線上查詢系統之開發。

4.1.3 資訊安全系統

資訊安全的概念，是泛指與電腦相關領域的安全措施，根據電腦的實體組成，以及電腦使用上的因素，資訊安全又可區為下列幾種[47]，本節將針對此三部份，說明本計畫所採取的資訊安全防護措施。

1. 硬體安全：電腦硬體環境控制、機房管理、硬體設備使用安全。
2. 軟體安全：系統安全、應用程式安全、個人資料安全。
3. 個人安全：人身安全、個人隱私安全、網路通訊安全。

在硬體安全的部份，為確保電腦系統不會因為外力(如跳電)的影響，而造成電腦無法正常運作，甚至資料的毀損。因此添購不斷電系統，避免因突發的跳電而影響電腦正常的運作。當不斷電系統電源即將耗盡前，會主動觸發作業系統進行正常關機程序，以避免造成作業系統的毀損。在資料儲存部份，也添購了磁碟陣列，並使用 RAID 5 的儲存方式，將同位元的資料存放在不同的硬碟上，當有資料毀損時，只要更新磁碟，即可透過同位元的演算法，修復毀損的資料。而其輪轉同位元的處理方式，也具備不錯的讀寫效能[48]。

本計畫所使用的電腦系統，除提供線上查詢之伺服器外，其餘伺服器均不進行網路連線，以做到實體隔離的防護工作。但為確保伺服器在讀取外來資料時的安全，在軟體安全的部份仍維持一定程度的防護措施。首先在安裝作業系統及應用程式安裝時，除安裝合法授權軟體外，並會下載最新版本的修護程式，以進行軟體系統的防護。另外，系統亦安裝防毒軟體，以掃描新存入之資料。而經由資料連結系統輸出並提供做線上查詢使用之資料，採實體隔離之方式，以杜絕任何可能來自於外界之危害。且電腦系統的使用僅授權維護管理人員，以避免其他人有接觸之機會。

由於本計畫所取得之資料中，包含與個人隱私相關之資料，如姓名、身分證號、出生日期、地址、車牌號碼等，若有不慎將造成嚴重個人隱私資料外洩之問題。在與健保局合作的過程中，為能確保個人就醫資料不外洩，對於攜出之資料中不能含有可辨識個人資料之欄位，或是有從其他資料庫回推就醫者身份之情形。因此對於攜出之資料須事先處理，包括減少攜出的資料欄位數目、事先進行資料的整併及資料欄位的重新計算等。

另外，後續研究用之資料若欲透過網路進行資料傳遞，提供給所需要之研究學者，在資料傳遞的安全防護上，首先可使用自然人憑證做為第一道的身份確認，完成身份確認後，在資料傳輸過程中可再結合 SSL（Secure Socket Layer）加密，以確保資料在傳輸過程中的安全。

本計畫最後所完成的資訊系統架構如圖 4-1 所示。

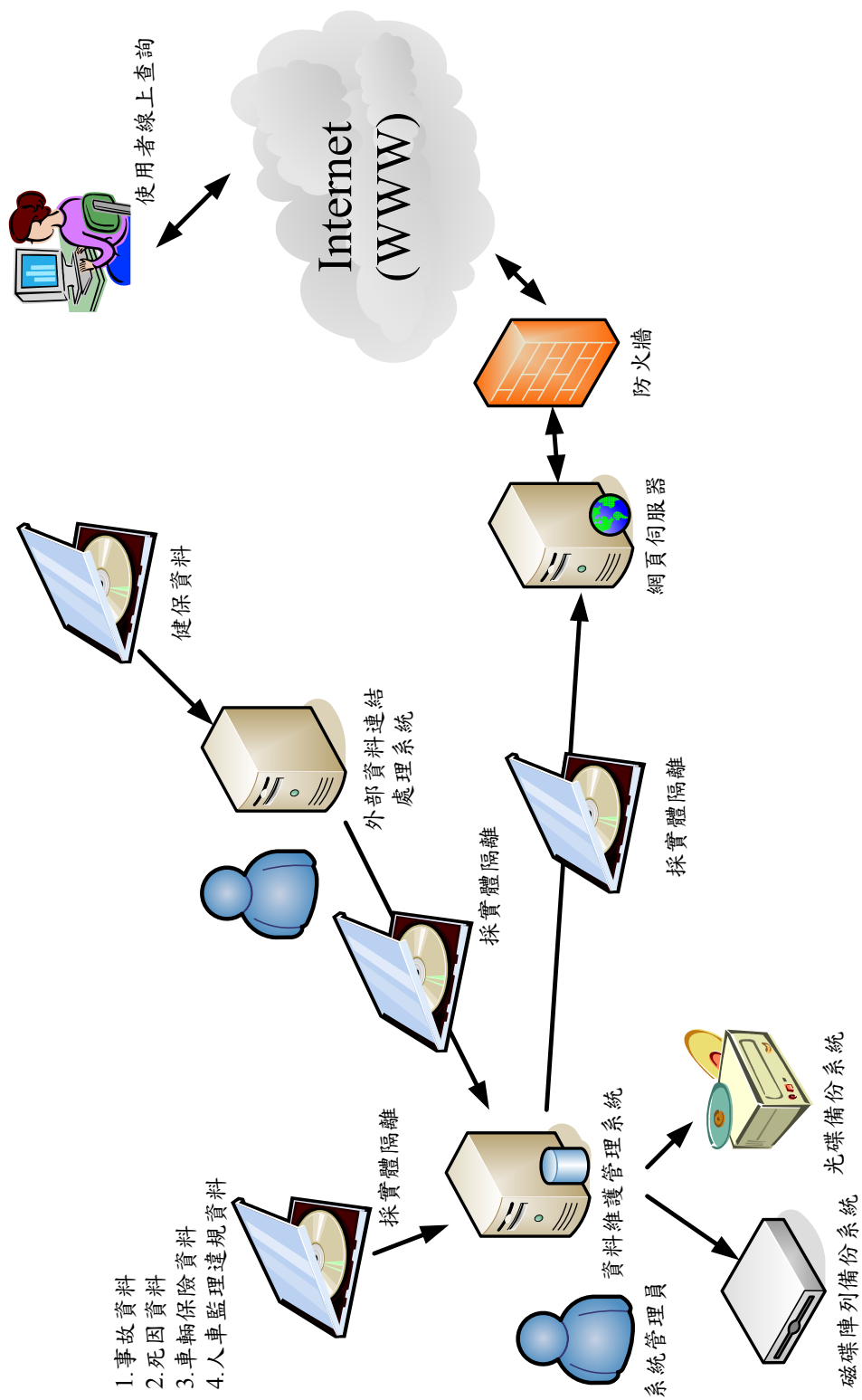


圖 4-1 系統資訊架構圖

4.2 「資料維護整合系統」的開發

在 2004 年的研究中，定義了資料匯入、資料淨化、資料連結等處理原則，而為確保各項資料的處理均能在一致的準則下進行維護，也方便管理者後續能對資料自行維護，因此開發此「資料維護整合系統」，整合及改善計畫執行過程中，對於各項資料處理所撰寫的獨立資料處理功能。

在開發過程中，陸續進行不同思維下的系統整合，譬如在身分證號的淨化部份，曾考量在各原始資料庫中均得進行此處理步驟，因此希望能在單一程式畫面便能進行不同資料庫的身分證號淨化處理工作。但在處理過程中發現，雖然資料淨化的處理原則相同，但不同的資料庫中其資料欄位不盡相同，尤其是事故資料又分屬在三個不同的資料表中(92 年起已整合成兩個資料表)，因此要在一個功能畫面中整合所有資料庫的處理工作，不但會增加系統開發的難度，對於使用便利性亦是一項考驗。再加上每個資料庫的處理步驟不盡相同，資料維護者將有可能忽略過程中的某一步驟，而造成操作上的失誤。因此以模組化的設計觀念，將共通的資料處理功能獨立成特定的功能模組，依不同的資料庫需求，設計專屬的資料維護選單，以方便資料維護者進行資料的匯入及淨化處理，並保留系統的擴充功能。未來各擴充功能可先單獨開發測試，完成後再陸續整合至維護系統中。

以下就各系統中的主要功能做介紹，資料維護整合系統主要功能選單畫面如圖 4-2 至圖 4-4 所示。對於資料維護整合系統的介紹，仍將相同或類似之功能一起介紹，但藉由不同操作畫面之說明，以了解其差異。



圖 4-2 資料庫維護管理系統主功能畫面



圖 4-3 事故資料庫處理選項功能畫面



圖 4-4 死因資料庫處理選項功能畫面

4.2.1 資料匯入

本計畫所取得之資料均來自不同的單位，其提供的資料格式主要可區分為兩大類：mdb 資料檔(道路交通事故資料庫及健保資料庫)與純文字檔(除道路交通事故資料庫及健保資料庫外均屬之)。雖同是文字檔格式，但因資料欄位不盡相同，因此處理的方式亦有所差異。經目前模組化之設計後，管理者只要修改資料欄位格式表，系統即可進行資料匯入工作。目前完成的資料匯入功能說明如下。

1. mdb 資料檔

在警政署所提供之事故資料中，是以 Access mdb 檔做記錄，其內同時包含不同年度的事故資料，因此在轉換時可由管理者指定欲轉入的資料年度，以記錄各年度之原始資料筆數。在所取得之資料中，92 年起已將原有人、事、地三個資料表，整合成兩個資料表，且資料欄位、變數定義亦做部份改變。匯入程式中

具備同時轉換不同資料格式之功能，而變數定義部份則保留原有的資料內容，不做任何變更，但會加註每筆資料所屬之資料定義年份，功能畫面如圖 4-5 所示。



圖 4-5 事故資料匯入功能畫面

2. 純文字檔

純文字檔是最主要的資料來源格式，在進行此類資料匯入時，須先了解不同資料檔中的資料格式，再依資料格式訂定不同的資料轉入參數檔。系統在進行資料匯入時，須先載入參數檔、產生資料表、載入欲匯入的資料檔、選定資料年度，最後即可進行資料轉入功能。畫面功能如圖 4-6 所示。

另外，車輛保險資料部份，92 年後的資料格式已有所修改，因此在進行資料轉換前，必須選取正確的資料格式年度。



圖 4-6 純文字檔資料匯入功能畫面

4.2.2 資料淨化

資料淨化之工作主要在於將不合理，且會影響連結結果之資料予以剔除，剔除後之資料暫不納入研究中，待後續可另外進行探討。主要的資料淨化項目包括：身分證號淨化、車牌號碼淨化、資料孤兒處理等。各項功能分述如下。

1. 身分證號淨化

在本計畫所取得之資料庫中，只要資料中含有身分證號便需進行此篩選處理，車輛保險資料中的被保險人資料則除外；因車輛所有人的身份除為自然人外，亦有可能為法人，故被保險人身分證號欄位有可能登記為統一編號，因此不適用此篩選原則。但後續在處理車輛所有人別時，則必須依此欄位資料進行判斷。

在進行資料篩選時，可選擇單一或是多篩選原則，並可指定被剔除資料的輸出欄位，此輸出資料後續可使用機率性連結的處理概念，做為改善資料連結成功率的研究資料。對於淨化之資料，系統提供三種不同的資料顯示方式：(1)螢幕顯示資料筆數，但並不直接刪除；(2)轉存至其他資料表；(3)直接刪除。淨化過程中，使用者可依需求選擇適當之資料處理方式。身分證號淨化功能畫面如圖 4-7 所示。

圖 4-7 身分證號篩選功能畫面

2. 車牌號碼淨化

車牌號碼並不若身分證號有其特定之編碼原則，故僅提供資料空白、長度小於 5 及特殊符號的篩選(如圖 4-8 所示)。其中車牌號碼所可能包含的特殊符號相當多，資料維護者可自行輸入欲篩選的符號，再由程式進行資料處理。

圖 4-8 車牌號碼篩選功能畫面

3. 日期處理

在不同來源資料庫中，對於日期資料之記載方式不盡相同，有存放在單一日期格式之欄位，亦有將年、月、日分別存放至三個不同資料欄位中。為方便後續資料使用，因此須將其合併至一個日期欄位中。軟體功能畫面如圖 4-9 所示。

圖 4-9 日期處理功能畫面

4. 資料孤兒處理

該功能主要是提供道路交通事故資料中，連結分散在不同資料表中的同一起事故資料。在進行資料內部連結時，有可能因資料記載之問題，而造成資料無法連結之現象，此情形稱之為資料孤兒。在執行此一功能時，須針對欲連結的資料產生事故唯一碼(由事故流水號+事故發生年月，共 10 碼)，並選擇資料處理的方式，如直接刪除、產生備份及畫面顯示資料筆數等。功能畫面如圖 4-10 所示。

圖 4-10 資料孤兒處理畫面

5. 身分證號重覆

在事故資料中有可能同一個人在不同時間內發生兩次以上的事故，當與其他資料進行連結時，有可能會有兩筆不同的連結記錄，因而影響資料的分析。如事故與死因資料連結時，同一個人有可能連結出兩筆死亡記錄，因而影響死亡距事故天數的統計結果。因此將身分證號重覆之資料分離出來，供未來研究。

4.2.3 資料整併

在進行資料匯入及資料淨化時，有可能是採逐年處理之方式進行。但在後續在進行資料連結時，卻有可能是跨年度之資料直接進行連結，因此須進行資料整併之工作。資料整併主要可分為三大類，分述如下：

1. 歷年度單筆資料之整併

主要是用於合併單一年度之資料，適用之對象包括道路交通事故資料庫、

死因資料庫、車輛保險資料庫。

2. 健保住院資料之整併

依在 3.4 節中說明的健保住院資料整併處理原則，撰寫至程式功能中。

3. 人車監理違規資料之整併

依在 3.5 節中說明的人車監理違規資料整併原則，撰寫至程式功能中。

4.2.4 其他資料處理功能

1. ICDMAP 格式轉檔功能

在 2004 年所進行之研究中，已引入 ICDMAP 軟體，利用健保資料庫中的 ICD 碼，進行受傷嚴重度分析。在使用 ICDMAP 進行資料計算前，先須由資料庫中輸出各筆就醫資料的年齡及 ICD 診斷碼，再透過此轉檔程式，轉換成符合 ICDMAP 之資料格式，最後再將檔案交由 ICDMAP 進行批次處理，軟體功能畫面如圖 4-11 所示。



圖 4-11 ICDMAP 資料格式轉換工具

2. 系統參數設定

為方便使用者切換連結的資料庫，因此系統亦提供資料庫連結參數的設定，相關功能畫面如圖 4-12 所示。



圖 4-12 資料庫連結參數設定功能畫面

4.3 「事故關聯資料分析與評估系統」的開發

設計「事故關聯資料分析與評估系統」是為了能對於各資料間的連結能有一致性的處理方法，並盡可能滿足多數研究之需求，詳細說明請見 2004 年的研究計畫成果。

在本年度的研究計畫中，著重系統功能及操作流程的改善，調整資料的連結流程並進行連結資料結果正確性的細部測試。此系統藉由圖形化界面的呈現方式，讓使用者清楚了解資料連結的流程及判斷的準則，對於連結後的結果可自行設定資料輸出欄位；輸出後的資料，在確保資料存取的安全，僅能透過光碟片將資料轉存至線上查詢伺服器中，供線上查詢使用。

對於本系統的資料擴充性方面，系統將不同資料庫來源的資料存放在不同的資料表中，各資料表間再透過關聯性欄位即可進行資料的連結。為加快資料的存取速度，在資料庫系統中會依資料內容進行分割，再存放至不同的資料表中，透過檢視表(View)或稱為虛擬資料表格的處理方式，便可關聯出不同查詢需求下所需的資料內容。為確保新資料與本系統的相容性問題，新取得之資料需先透過前述「資料維護管理系統」的資料匯入、資料淨化等過程的處理，才可匯入本系統中。

本系統開發工具、檔案編碼原則及系統功能特色等分述如下。

4.3.1 系統開發工具

在系統規劃之初，即強調圖形化及友善的使用者界面，並考量本計畫資料

庫系統是以微軟的 SQL Server 為主，為求能讓系統達到最好的匹配性、穩定性及延續性，因此採用 Visual Basic 設計使用者界面。與資料庫的連結溝通上，則透過資料庫函式庫 ADO(ActiveX Data Object)與 SQL Server 進行連結。

在本計畫取得之資料已超過六千萬筆，為能讓資料庫系統發揮更好之資料連結處理效能。因此系統開發過程中，對於 SQL Server 資料庫亦進行細部效能的改善，主要的改善工作包括：

1. 建立主索引鍵，以有效增進資料連結時之效率。
2. 將資料依使用之需求，進行資料的切割。因為單一資料表中，當資料筆數過大時，將會影響資料搜尋效能。因此將資料依與事故日之前後順序、身分群組進行資料切割。
3. 建立檢視表依資料使用需求，將分割後之資料再組合成完整之資料表格，此作法對於資料表的管理相當有彈性，亦可應用在擴充新年度資料時，可減少程式碼之修改需求。

4.3.2 資料編碼系統

在本系統中，每個連結畫面中均有相當多之資料輸出點，在不同資料連結層中，如何有效管理輸出之資料，並了解輸出資料彼此間之關聯性，是一項很重要之管理課程。為配合系統程式化的發展，在 2004 年所設計的系統中，已制定一套二層式編碼原則，希望能在編碼名稱中可分辨出不同的篩選條件，以方便管理。在原來的設計過程中，為避免編碼長度過長之問題，在進入第二層編碼前會先濃縮第一層編碼內容。此作法雖可有效控制編碼長度，但無形中卻降低了可讀性。

因此在本年度的研究中，以 2004 年之編碼原則為基礎進行修改，制定一套 21 碼之編碼原則，對於尚未連結之變數項目則以「0」做代替。基本原則說明如下：

1. 建立各資料庫之代碼：
 - (1) 事故資料庫為 P。
 - (2) 強制險為 V。
 - (3) 任意險為 A。
 - (4) 死因資料庫為 D。

- (5) 人車監理違規為 U。
 - (6) 健保住院為 I。
 - (7) 健保門診為 O。
2. 第 1 至 6 碼為事故資料應用連結層，主要是設定事故資料庫與何種資料庫進行連結，並設定相關之連結變數與時間框。
 3. 第 7 至 12 碼為違規行為與就醫資料連結層，主要設定連結資料與事故發生日前後之關係，及其基本的連結變數設定。
 4. 第 13 至 21 碼為就醫特性資料連結層，主要進行門診及住院就醫行為之探討。
- 各碼之詳細說明如下：

完整編碼：P□-□□□-□-□□□□□□-□□□□□□□□□□

第 1 碼：事故資料庫

第 2 碼：所連結的資料庫代碼

第 3 碼：身分證號連結條件(事故資料與強制險、任意險及人車監理違規連結時，即限定事故資料庫中之事故當事人為駕駛人。強制險與任意險為肇事駕駛人。)

0：不須考慮；

1：相同；2：不同。

第 4 碼：車牌號碼連結條件(D 用死因碼限定為 E810-E819，U 用違規種類，可知是否為當時肇事的違規)

0：不須考慮；

1：相同；2：不同。

第 5 碼：日期連結條件(P 用事故日，V 及 A 用出險日，U 用違規日，D 用事故後 30 日內死亡)

0：不須考慮；

1：相同；2：不同。

第 6 碼：連結成功為 0，未連結成功在原來資料庫後加 A(P 資料庫)，B(其他資

料庫)。

第 7 碼：事故前後違規或就醫記錄

0：不考慮

1：事故前；2：事故當天；3：事故後

第 8-10 碼：距事故日的天數，如 000 表當天，030 表三十天內(含)，365 表一年內(含)，999 表一年以上，其他選項日期在 365 天以內。

第 11-12 碼：所連結的健保資料庫代碼如下，若是違規資料庫則代碼為 99(違規行為探討)

第 13 碼：有無轉院記錄，「0」表不須考慮；「1」表示有轉院；「2」表示無轉院。

第 14 碼：是否為長期住院，「0」表不須考慮；「1」表示長期住院；「2」表示非長期住院。

第 15-17 碼：長期住院天數，若第 14 碼為「0」或「2」，則記錄「000」，否則記錄判斷天數。

第 18 碼：「0」表不考慮；「1」表 5 天內有就醫記錄；「2」表 5 天內無就醫記錄。

第 19 碼：就醫科別，「0」不須考慮；「1」10 科內；「2」10 科外。

第 20-21 碼：就醫時間框的類別

00：不考慮；01：門診，30 天內 7 次門診

02：住院，45 天內

03：住院加回診，45 天內出院後有 1~3 次門診

4.3.3 系統功能及特色

系統所應規劃的基本功能如下：

1. 提供圖形化及友善的使用者界面，方便操作人員進行系統操作。
2. 資料連結過程中，能準確掌握資料筆數變化，供資料連結正確性之檢核。
3. 可自行設定資料下載欄位。
4. 連結後之資料，可依設定之欄位進行資料輸出。
5. 能由使用者自行設定不同時間範圍之資料進行連結。

期能滿足大部份之應用研究需求。

系統依選用的連結資料庫及篩選變數設定的內容，將系統功能區分為五大層，各層的功能、特色及相關畫面說明如下。

1. 第一層：資料連結設定層

本系統的設計主要是以事故資料為出發點，選擇要探討連結結果的其他資料庫。在此層中，使用者可自行設定欲分析的事故資料庫時間範圍及欲連結的資料庫，本計畫提供的事故資料從 88 年至 91 年。程式畫面如圖 4-13 所示。

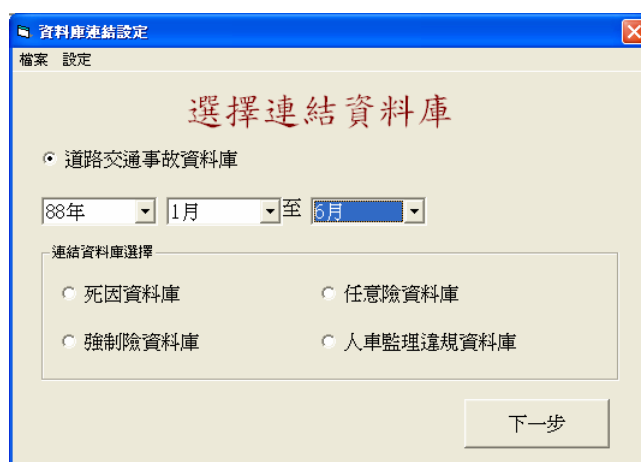


圖 4-13 資料連結設定層畫面

2. 第二層：事故與單一資料庫連結層

在完成上一層的設定後，即進入本層功能畫面中。在本連結層中可產生探討事故與單一資料庫間連結結果特性的資料。以死因資料庫為例(如圖 4-14 所示)，進入此畫面後，第一步驟即決定以身分證號做為確定性連結，可連結之資料隨後再使用死因碼及死亡日距事故天數做連結變數(可由使用者彈性調整日期間隔)，過程中可在各種連結進行資料的輸出(輸出功能將於第 6 點中詳述)，或是選擇「往下一步驟」進入「事故資料時間框及資料連結設定層」，進行下一階段的資料連結設定。

在事故與車輛強制險已決賠資料的連結流程中，主要功能與死因資料庫之連結類似，只是將死因碼改為使用車牌號碼篩選、死亡距事故日改為事故日與出險日是否同一天，如圖 4-15 所示，過程中也可隨時選擇檔案輸出或是進行下一步驟。事故與人車監理違規資料的連結流程則如圖 4-16。

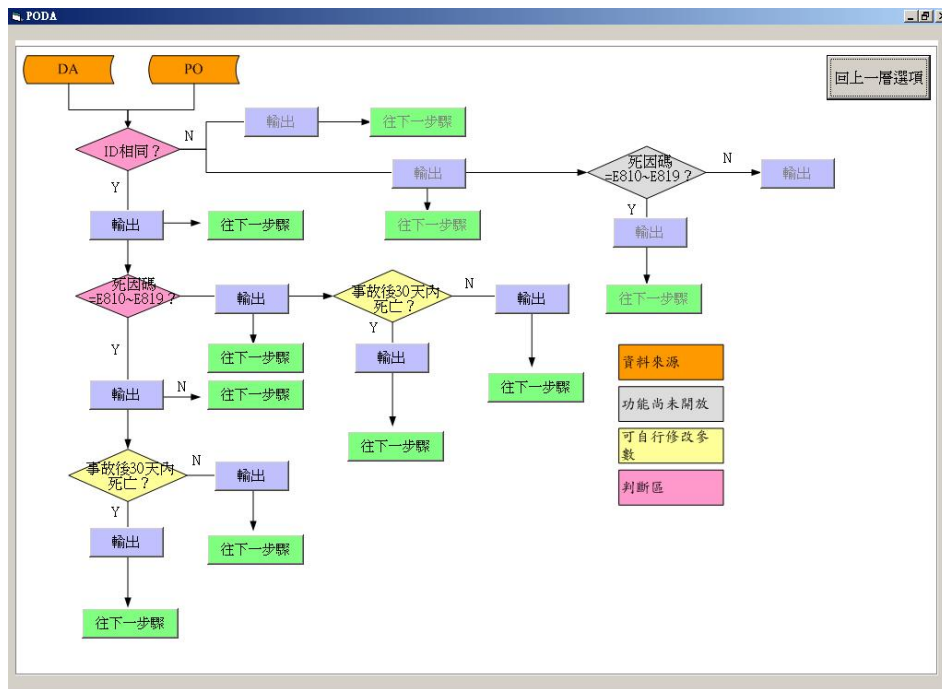


圖 4-14 事故與死因連結流程圖

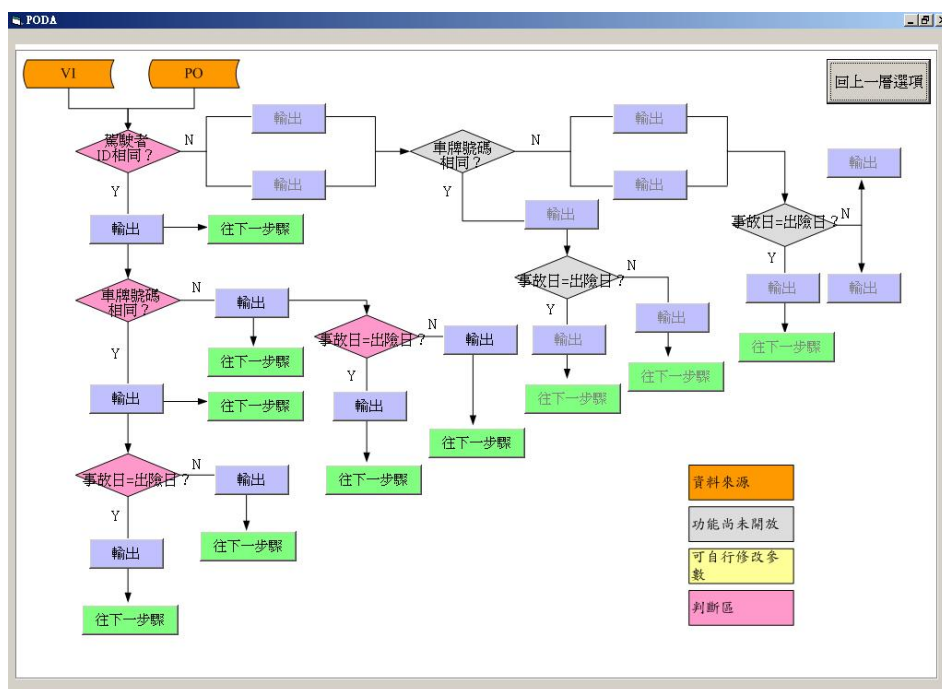


圖 4-15 事故與車輛強制險連結流程圖

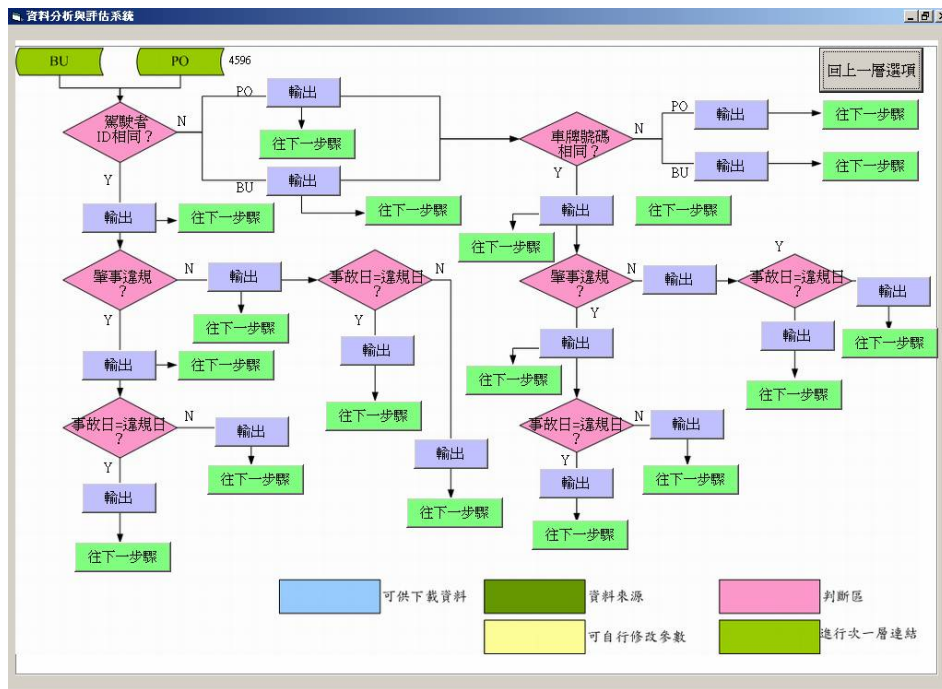


圖 4-16 事故與人車監理違規資料連結流程圖

3. 第三層：事故資料時間框及資料連結設定層

本層之功用主要用於設定連結資料與事故發生日之先後順序關係，及選擇距離事故發生日幾天內之資料。在健保資料的勾選方面，使用者可單獨選擇住院資料或是西醫門診資料進行連結，亦可同時勾選兩類健保資料，功能畫面如圖 4-17 所示；人車監理違規資料則直接勾選。當單獨勾選一項時，系統將跳過下一層之連結功能，直接進入「門診或住院就醫特性連結層」。

圖 4-17 事故資料時間框及資料連結設定層

4. 第四層：門診與住院資料關聯層

當使用者在前一層中同時勾選住院及西醫門診資料後，將進入此連結功能頁，如圖 4-18 所示。在此功能畫面中，將可篩選出純住院、純門診及同時有門診及住院之資料。在確定分析資料類別後即可進入下一層的連結功能畫面。

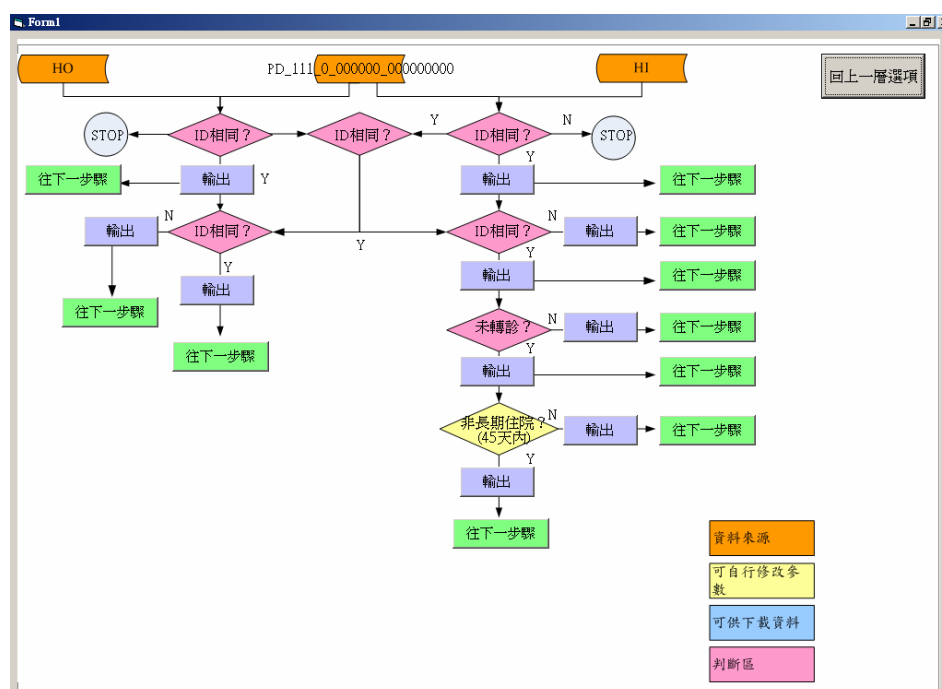


圖 4-18 門診與住院資料關聯層

5. 第五層：門診或住院就醫特性連結層

此功能層是將本計畫所定義之事故後住院就醫條件設計成不同的判斷準則，如圖 4-19 所示，供未來進一步探討；西醫門診之就醫特性如圖 4-20 所示。

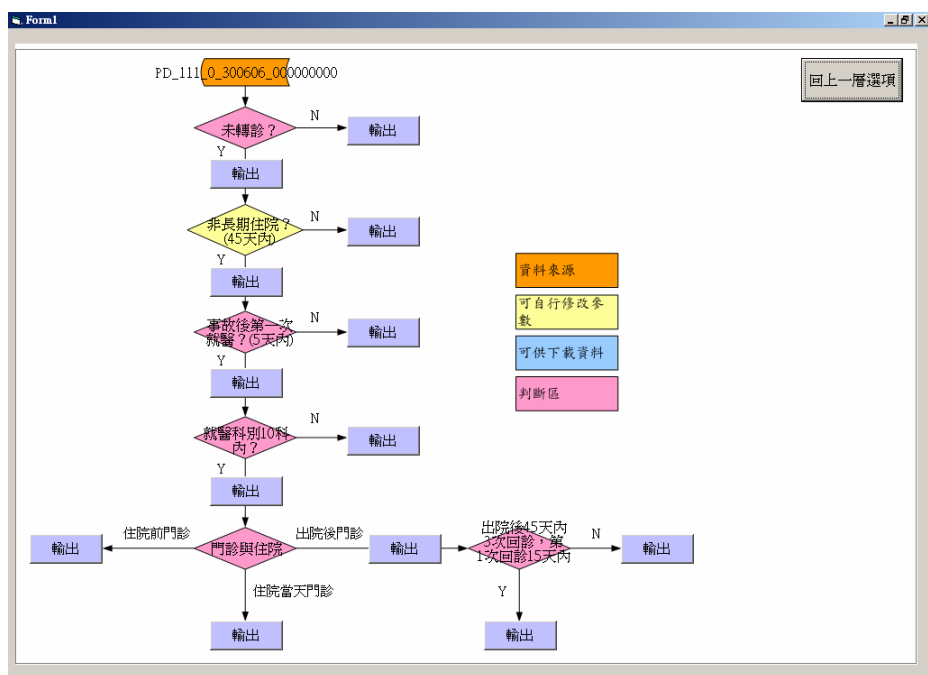


圖 4-19 住院就醫特性連結流程圖

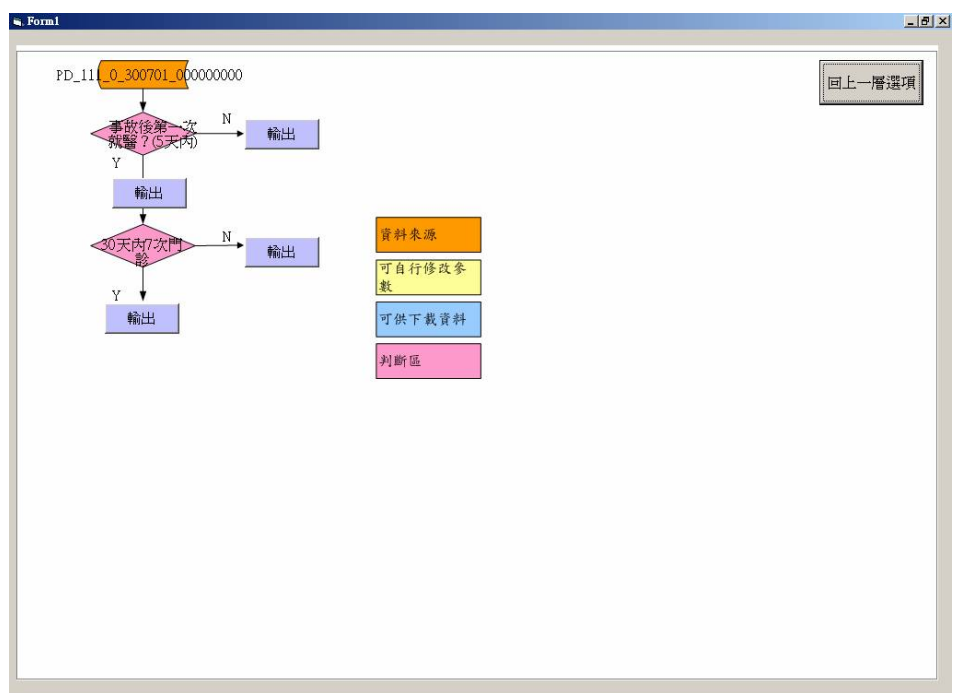


圖 4-20 西醫門診就醫特性連結流程圖

6. 其他

本系統除具備前述五項之連結判斷功能外，系統還具備各資料庫輸出欄位的設定功能，由系統管理者事先設定各資料庫中所允許輸出之欄位，爾後在連結

功能畫面中進行資料輸出時，將僅能輸出系統允許之欄位，功能如圖 4-21 所示。

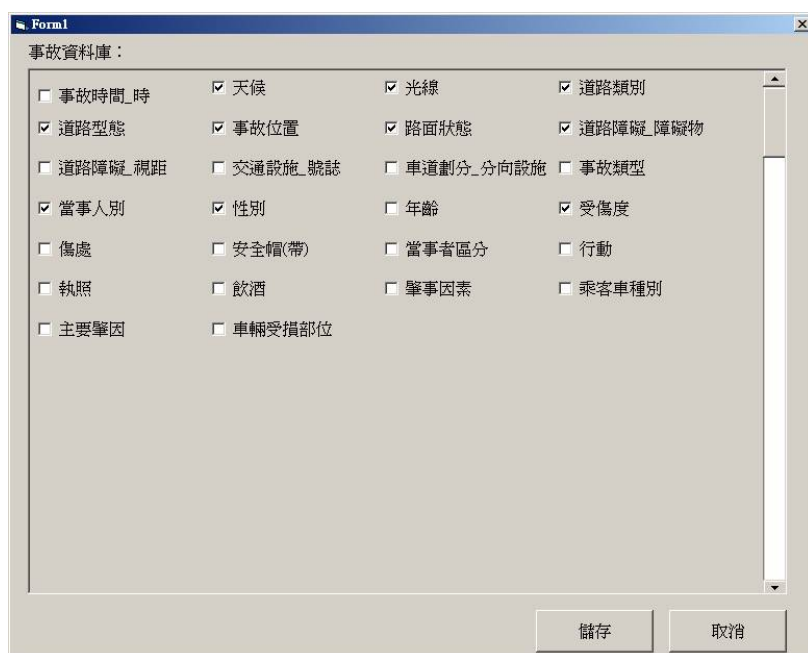
A screenshot of a software window titled 'Form1' with a subtitle '事故資料庫：'. The window contains a grid of 20 checkboxes for selecting data fields. The fields are arranged in four columns and five rows. The first column includes '事故時間_時', '道路型態', '道路障礙_視距', '當事人別', '傷處', '執照', and '主要肇因'. The second column includes '天氣', '事故位置', '交通設施_號誌', '性別', '安全帽(帶)', '飲酒', and '車輛受損部位'. The third column includes '光線', '路面狀態', '車道劃分_分向設施', '年齡', '當事者區分', '肇事因素', and an empty checkbox. The fourth column includes '道路類別', '道路障礙_障礙物', '事故類型', '受傷度', '行動', '乘客車種別', and an empty checkbox. At the bottom right, there are two buttons: '儲存' (Save) and '取消' (Cancel).

圖 4-21 輸出欄位設定功能畫面

在各功能層不同步驟中進行資料輸出前，系統會先告知資料筆數，使用者亦可由此功能進行連結資料是否正確之檢核工作，在選擇輸出資料庫時，系統會自行判斷在該步驟中所能下載的資料庫內容。如在「事故與單一資料庫應用連結層」中，並無法輸出健保資料，功能畫面如圖 4-22 所示。

A screenshot of a software window titled 'PD_100_0_000000_000000000'. The window contains a text box labeled '資料筆數：' with the value '592'. Below this are several checkboxes for selecting data sources: '事故資料庫' (checked), '死因資料庫' (checked), '強制險資料庫' (unchecked), '任意險資料庫' (unchecked), '健保門診' (unchecked), and '健保住院' (unchecked). At the bottom, there are two buttons: '資料下載' (Download Data) and '取消下載' (Cancel Download). Below the buttons is a progress bar.

圖 4-22 資料輸出功能畫面

4.4 「線上查詢系統」的開發

為讓本計畫所產出之資料做更廣泛的應用，已能利用資料維護整合系統所

輸出之資料為基礎，挑選部份資料欄位做為查詢及輸出變數，在確保個人隱私之前提下，提供線上資料查詢服務的功能；惟功能的開放運作，則仍待與相關單位研處。

本系統的設計主要是參考威斯康辛州在 CODES 計畫中所提供的 Cube Analyzer 線上查詢系統功能，在該系統中將資料區分為三大類，包括與碰撞事故相關的資訊、碰撞與醫療住院系統整合之資訊及碰撞車輛相關之資訊。使用者可依需求自行新增、移除分析欄位，或依需求彈性調整欄位排列順序，系統會依選擇的變數進行資料的計算處理，最後在網頁中顯示計算結果，若有需要亦可下載 CSV 檔進行進一步的分析，其功能頗類似 Excel 功能中的樞紐分析。

本系統在設計時，亦採用與 Cube Analyzer 系統相同的 ASP 技術，使用 SQL Server 做為資料庫儲存系統，並採用三層式架構(如圖 4-23)進行系統建置。在第一層查詢端的部份除以 HTML 做為畫面顯示外，並使用 Javascript 進行變數選項的挑選處理。在完成變數選擇後，將查詢資料送交網頁伺服器，透過網頁伺服器與後端的資料庫系統進行連線查詢功能，計算後的結果再送交網頁伺服器以產生對應的 HTML 網頁，最後顯示於查詢者端。

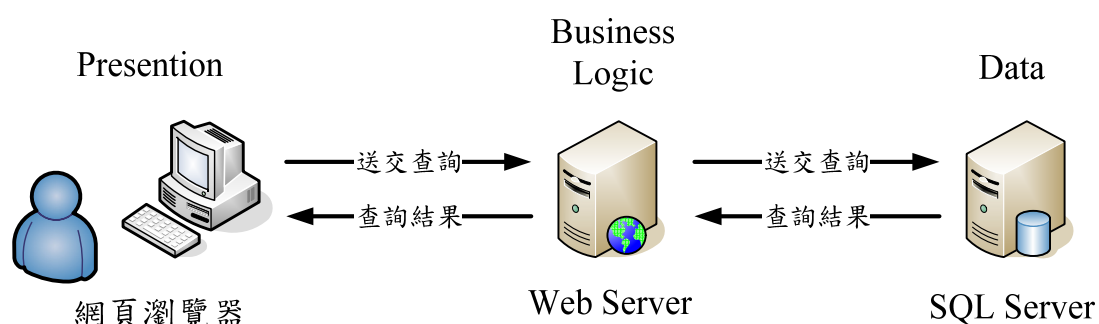


圖 4-23 三層式架構示意圖

本系統在資料的查詢部份，目前將連結資料區分為三大類：(1)事故資料與死因資料；(2)事故資料與門診資料；(3)事故資料與住院資料。使用者在進行查詢前，必須先選擇所欲查詢的資料種類。在進入變數項選擇頁時，系統則參考多數國外車禍資料網頁查詢系統的查詢變數分類方式，將變數依其內容區分為人、車(事件)及環境(路)三大類，使用者可自行挑選 X 軸或是 Y 軸的變數(如圖 4-24 所示)，在統計結果的顯示部份，在事故與死因資料中僅能顯示「人數」，但在

其他兩類資料中，則可選擇顯示「人數」或是「金額」。在送交查詢後，系統經與資料庫進行連線後，便可顯示計算結果(如圖 4-25 所示)。此外，各變數的原始資料中僅登載代碼，使用者不易自代碼中了解其代表的意義，因此系統在做最後查詢結果輸出時，會將所有變數的代碼轉換成實際之代表意義，以增加查詢結果的可讀性。

在系統設計過程中，為加快資料查詢處理速度，減少資料查詢及顯示往返的次數，因此在設計上採用一次查詢多次使用的方式處理。先收集所有的變數資料，一次提交資料庫做查詢，查詢後的結果將回傳至網頁伺服器中，再供不同的變數組合查詢結果使用。經測試後發現，採用此種資料處理方式，可有效改善查詢效率，並可處理多變數之資料篩選工作。

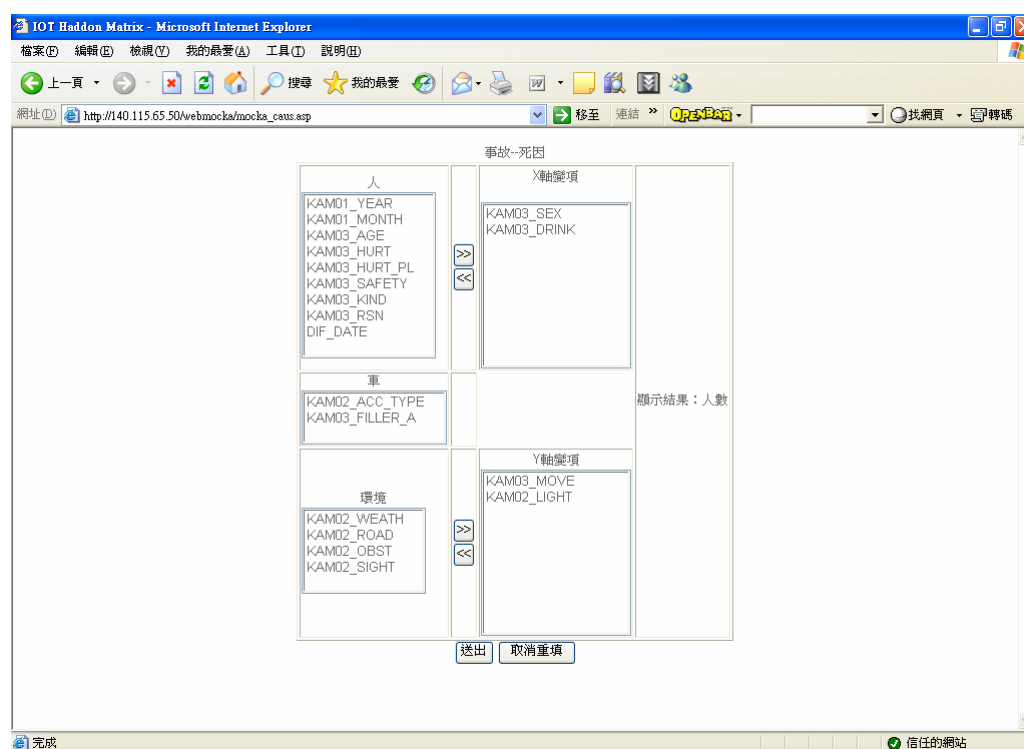


圖 4-24 變數選擇畫面

IOT Haddon Matrix - Microsoft Internet Explorer

檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)

← 上一頁 → 搜尋 我的最愛

網址(A) http://192.168.0.109/webmoc 移至 OPENBAR 找網頁 轉碼

KAM02_WEATH	KAM02_ROAD	KAM03_SEX, KAM03_HURT					
		男			女		
		死亡	受傷	未受傷	死亡	受傷	未受傷
暴雨	國道	0	0	1,415	0	54,873	0
	省道	0	0	0	0	0	0
	縣道	0	0	0	0	0	0
	鄉道	0	0	0	0	0	0
	市區道路	0	9,244	65,586	0	0	0
	村里道路	0	0	14,773	0	430	0
	專用道路	0	0	0	0	0	0
	其他	0	0	0	0	0	0
	國道	0	0	0	0	0	0
強風	省道	0	24,206	66,910	3,846	0	0
	縣道	0	60,486	0	0	5,334	0
	鄉道	0	8,083	7,107	0	0	0
	市區道路	0	2,782	9,961	0	0	1,523
	村里道路	0	23,221	4,151	0	67,027	0
	專用道路	0	0	0	0	0	0
	其他	0	377	0	0	10,182	0
	國道	0	0	0	0	0	0
	省道	0	2,122	2,542	0	37,279	0
風沙	縣道	0	0	675	0	0	0
	鄉道	0	0	0	0	0	0
	市區道路	5,202	12,562	11,057	0	10,822	0
	村里道路	0	28,916	0	0	0	0
	專用道路	0	0	0	0	0	0
	其他	0	0	0	0	0	0
	國道	0	0	0	0	0	0

完成 網際網路

圖 4-25 線上查詢結果

第五章 系統連結結果的應用

在 2003 年及 2004 年的研究報告所進行的多項研究主題探討，均是使用身分證號所做的確定性連結。由表 3.3、表 3.5 及表 3.7-表 3.9 所整理的資料淨化筆數中已刪除了部份身分證號、車牌號碼及日期資料不齊全或是不正確的資料，每個資料庫中被刪除的資料百分比，亦因各資料庫建置狀況的不同而有所差異。為能更了解資料的連結情形，本章中將針對連結用資料進行連結探討，也包括了應用資料探勘在連結資料的分析上，目前探討的內包括：

1. 事故資料及死因資料的連結，回補淨化過程中被刪除的資料，以了解資料再連結之情形，詳細內容如 5.1 節所述。
2. 強制險已決賠資料與事故資料的連結，分別使用身分證號、車牌號碼及日期，在不同連結條件下資料的連結情形，詳細內容如 5.2.1 節所述。
3. 強制險受害人資料的連結，分別使用身分證號及日期(事故日與出險日)，在不同連結條件下資料的連結情形，詳細內容如 5.2.2 節所述。
4. 強制險已決賠資料中的肇事駕駛人及事故資料的事故當事人，分別使用身分證號及日期，在不同連結條件下資料的連結情形，詳細內容如 5.2.3 節所述。
5. 與其他本所研究計畫之合作，詳細內容如 5.3 節所述。
6. 應用資料探勘在預測模式之探討，詳細內容如 5.4 節所述。

5.1 被淨化資料之連結探討

參考 2.1 節國外在做機率連結的研究經驗，可得到幾項可供參考的資訊：

1. 常用的連結欄位為姓名、性別、出生日期、死亡日期、社會安全號碼等，並視系統資料特性加入如居住地址、郵遞區號等空間地理資訊之資料，或是住宅遷出入日期、出入院日期等與事件發生相關之日期資料。但在分析本計畫所取得之資料欄位後發現，共通且可用之欄位較少，僅有性別、出生日期等。
2. 在使用性別做為連結欄位的判斷時，權重值通常會採取全對或全錯的計算，因為資料範圍明確，且資料輸入時有明顯之差異，因此便會給予如此之權重。
3. 雖然部份系統仍會視資料內容，給予細部的調整或容許誤差範圍，但在關於年份的資料比對處理上，一般允許的誤差時間為 ± 3 年。
4. 由國外發展經驗得知，在資料連結後常需再藉由人工檢視的方式，對連結後

的資料做更進一步的確認，並調整出該資料系統較適用的連結變數權重值。

在以身分證號相同便具有高度可能為同一人的前提下，來觀察國外將出生年份容許誤差設定為 ± 3 年，而做為可能是同一人的連結門檻條件，是否適合台灣的資料？因此以身分證號連結民國 90 年事故與死因資料，可成功連結 4,274 筆記錄，為確定連結後的資料的確為同一筆資料，因此再使用性別、出生年、出生月及出生日進行第二層的比對工作。比對後發現，即便身分證號相同，但再以其其他欄位做比對時，仍約有 5% 的資料至少有一個欄位資料不相同，也就是說約有 95% 的資料是完全相同的。資料比較結果如表 5.1

進一步使用性別相同，但在出生年、月、日有一欄位資料不同的 209 筆資料做進一步的測試。由表 5.2 資料的分佈可知，多數是集中在出生年 ± 3 年的範圍內(約佔有 80%)，而此區域內主要的資料型態統計如表 5.3 所示。另外由表 5.2 發現到，隨著差異年數的增加，可成功連結的資料筆數將隨之遞減，資料呈現的內容與文獻回顧所獲得之資訊一致，故出生年差異在 3 年內的資料，應是可參考的數據。

表5.1 90年PO與DA資料配合其他欄位之比較結果

性別	出生年	出生月	出生日	資料筆數	百分比(%)
相同	相同	相同	相同	4,060	94.99
			不同	100	2.34
		不同	相同	25	0.58
			不同	10	0.23
	不同	相同	相同	53	1.24
			不同	5	0.12
		不同	相同	2	0.05
			不同	14	0.33
不同	相同	相同	相同	5	0.12
總計				4,274	100.00

表5.2 身分證號同、出生年月日不同的資料比對分佈

出生 年差 異年 數	出生 月差 異月 數	出生日差異日數												
		0	1	2	3	4	5	6	7	8	9	10	11~28	總計
0	0	4,060	10	12	7	4	5	5	1	3	3	30	20	4,160
	1	7	1									1	1	10
	2	3		1		1				1			1	7
	3	3											0	3
	4	4											0	4
	5	2											0	2
	8	1			1		1						1	4
	10	5											0	5
1	0	11											0	11
	10										1		0	1
2	0	11											0	11
	4												1	1
3	0	6	1										0	7
4~73	0~12	27	3	2	3	0	0	0	0	1	0	1	6	43
總計		4,140	15	15	11	5	6	5	1	5	4	32	30	4,269

表5.3 身分證號同、出生年月日不同的資料主要形態統計

差異範圍	資料筆數	百分比(%)
出生年±3	166	0.794
出生年±3 & 出生月=0	129	0.617
出生年±3 & 出生日≤10	142	0.679
出生年、月、日有一不同者	209	1.000

在資料淨化過程中由表 3.3 及表 3.5 中可發現，事故資料中被刪除的資料多為身分證號空白者，而死因資料中則是檢查碼的錯誤。因此本小節將嘗試將此兩部份之資料透過其他欄位，嘗試改善連結情形。

1. PO 中身分證號空白的資料

以民國 90 年警察所登錄的事故資料中，身分證號空白的資料為測試資料。其中身分證號空白的有 14,103 筆，此部份資料在性別、出生年、出生月、出生日四項資料情形如表 5.4 所示，其中出生年、月、日的資料此三項資料中，並沒有出現只有單一 1 項或 2 項空白的現象。由表 5.4 中身分證號空白的 14,103 筆記錄，扣除出生年、出生月、出生日空白的 11,013(3,936+7,077)筆記錄後，剩餘 3,090 筆資料，其中，有出生年、出生月、出生日，但性別空白者有 12 筆，而出生年、出生月、出生日及性別均有值者有 3,078 筆。

表5.4 民國90年PO資料中身分證號空白其他欄位資料筆數統計

性別	出生年、月、日	資料筆數	百分比(%)
空白	空白	3,936	27.91
	有效值	12	0.09
有效值	空白	7,077	50.18
	有效值	3,078	21.83
總計		14,103	100.00

利用出生年、出生月、出生日及性別均非空白的 3,078 筆資料，以出生年、出生月、出生日及性別此 4 項變數為連結鍵，重新連結民國 90 年當年的死因資料(DA)後，共連結出 3,762 筆資料(表 5.5)，經人工檢查可發現有許多是一筆警察記錄(PO)連結到多筆死因記錄(DA)的結果，亦即可能會有不同人但出生年、出生月、出生日及性別相同者的資料，被連結在一起，也可能會有同一個人但發生多次車禍的資料，被重複連結。其中有 695 筆記錄係為一筆警察資料連到一筆死因資料的結果，佔出生年、出生月、出生日及性別均有值者的 3,078 筆記錄中的 22.58%，為所有身分證號空白記錄的 4.93%(695/14,103=0.0493)。

表5.5 PO中身分證號空白者，利用其他欄位連結民國90年DA資料之結果

每組一對多連結資料中的記錄筆數	次數	總計	每組一對多連結資料中的記錄筆數	次數	總計
1	695	695	10	5	50
2	295	590	11	5	55
3	164	492	12	4	48
4	117	468	13	4	52
5	62	310	14	5	70
6	54	324	15	1	15
7	30	210	16	2	32
8	26	208	35	1	35
9	12	108	總計		3,762

使用出生年、出生月、出生日及性別等 4 個變數所組成的組合變數，來辨識同一個人，上述資料顯示可能產生較大的誤差，故後續可再加入死亡日期在事故後 30 天內，及 E-Code 在 E810-819 的判斷，應可有助於降低誤差的發生。

2. 身分證號檢查碼錯誤的資料

身分證號檢查碼錯誤的資料，改以身分證號前 9 碼進行連結後，可重新連回在資料淨化過程中被刪除的資料筆數約有 2-4%，如表 5.6 所示。因死因資料中被刪除的資料大多是因檢查碼錯誤所造成的，因此可連結回補的資料應較多。這些可以重新連到的資料中，並無性別不同者，E810-819 的記錄中，出生年、月、日相同的比例高，約佔 85%以上，而非 E810-819 的記錄中，出生年、月、日均不同的比例高，約佔 80%以上。故用身分證號前 9 碼同的連結，再輔以出生年日期相同，且 E-Code 在 E810-819 的範圍內時，可能有 85%的可能性是為同一人，因此可考量做為另一參考的連結條件。

表5.6 PO及DA身分證號檢查碼錯誤者，以身分證號前9碼重新連結的結果

身分證號檢查碼錯誤的資料庫	年份 (民國)	檢查碼錯誤的資料筆數	以前 9 碼重新連結的結果	
		(a)	資料筆數(b)	(b)/(a) 之百分比(%)
DA	88	3,017	42	1.39
	89	3,118	56	1.80
	90	138	3	2.17
	91	56	2	3.57
PO	88	189	4	2.12
	89	373	7	1.88
	90	433	7	1.62
	91	141	3	2.13

5.2 強制險資料連結探討

在 2004 的研究中，使用 88 年至 91 年事故資料(事故當事人)與強制險已決賠資料(肇事駕駛人)，曾初步使用身分證號進行連結，但並未詳細的使用各年度的資料及不同連結條件進行連結探討，包含結合使用車牌號碼及日期做為判斷條件之連結。為了解已取得資料間的連結情形，將就下列幾種資料組合，使用各資料間可連結的欄位，如身分證號、車牌號碼或日期進行連結。

1. 事故資料與強制險已決賠資料
2. 事故資料與強制險受害人資料
3. 強制險已決賠資料與強制險受害人資料
4. 強制險已決賠資料與健保資料
5. 事故資料與健保資料

在事故資料或強制險資料中存在同一人重複發生事故或是同一肇事者重複理賠之資料，兩種資料在連結時其資料連結模型是屬於「一對多」之情形，為方便資料之統計，因此將每一次的事件或理賠均視為獨立的資料(給予每一筆資料

新的唯一碼)，後續在做資料統計時，僅須計算不重複的唯一碼的筆數，即可計算出連結後資料佔連結用資料的百分比。

另外，本年度新取得的強制險受害人資料，為強制險理賠時所記載同一次理賠中的受害人資料，在資料連結前，同樣給予每一筆受害人新的唯一碼，連結後再統計唯一碼的資料，即可計算可成功連結之資料筆數及百分比。

5.2.1 事故資料與強制險已決賠資料連結探討

本節主要探討事故資料與強制險已決賠資料之連結(簡稱 PO_VI 連結)情形，連結用之資料是使用表 3.3 與表 3.5 已完成淨化後 88 年至 93 年之資料。分別就下列六種連結情形進行探討。

1. 身分證號相同
2. 身分證號相同且事故日與出險日相同
3. 車牌號碼相同
4. 車牌號碼相同且事故日與出險日相同
5. 身分證號相同或車牌號碼相同
6. 身分證號相同或車牌號碼相同，且事故日與出險日相同

連結後之結果整理於表 5.7 至表 5.12 中所列之資料筆數，其結果綜合說明如下：

1. 在同一種連結條件下，不同年度的事故資料可連結率均相差不遠，並沒有很明顯的變化趨勢。
2. 在不同連結條件下，不同年度的強制險已決賠資料，可連結率有逐年增加之趨勢。
3. 綜合各種連結條件，事故資料平均可連結率約為 15%、強制險已決賠資料約為 22.5%。
4. 事故資料與強制險已決賠資料，其可供連結判斷之欄位主要有事故當事人身分證號與肇事駕駛身分證號、事故車輛車牌號碼與肇事車輛車牌號碼、事故日與出險日，表 5.12 之連結結果應是目前所取得資料下，兩個資料庫所可能連結到的最大範圍。

表5.7 PO_VI身分證相同之連結

	事故資料			強制險已決賠資料		
	連結用資料	可連結資料筆數	可連結百分比(%)	連結用資料	可連結資料筆數	可連結百分比(%)
88 年	46,552	6,841	14.70	77,197	9,423	12.21
89 年	101,719	14,370	14.13	115,524	20,578	17.81
90 年	121,934	18,412	15.10	123,456	25,362	20.54
91 年	161,351	24,438	15.15	133,047	32,186	24.19
92 年	235,153	33,698	14.33	124,597	40,627	32.61
93 年	268,353	39,630	14.77	138,034	48,032	34.80

表5.8 PO_VI身分證號且事故日與出險日相同之連結

	事故資料			強制險已決賠資料		
	連結用資料	可連結資料筆數	可連結百分比(%)	連結用資料	可連結資料筆數	可連結百分比(%)
88 年	46,552	6,211	13.34	77,197	8,508	11.02
89 年	101,719	12,783	12.57	115,524	18,314	15.85
90 年	121,934	16,517	13.55	123,456	22,678	18.37
91 年	161,351	21,749	13.48	133,047	28,734	21.60
92 年	235,153	30,075	12.79	124,597	36,752	29.50
93 年	268,353	35,318	13.16	138,034	43,504	31.52

表5.9 PO_VI車牌號碼相同之連結

	事故資料			強制險已決賠資料		
	連結用資料	可連結資料筆數	可連結百分比(%)	連結用資料	可連結資料筆數	可連結百分比(%)
88 年	46,552	7,434	15.97	77,197	8,879	11.50
89 年	101,719	15,467	15.21	115,524	19,643	17.00
90 年	121,934	19,367	15.88	123,456	24,054	19.48
91 年	161,351	25,892	16.05	133,047	31,046	23.33
92 年	235,153	35,330	15.02	124,597	38,993	31.30
93 年	268,353	41,347	15.41	138,034	46,394	33.61

表5.10 PO_VI車牌號碼且事故日與出險日相同之連結

	事故資料			強制險已決賠資料		
	連結用資料	可連結資料筆數	可連結百分比(%)	連結用資料	可連結資料筆數	可連結百分比(%)
88 年	46,552	6,880	14.78	77,197	8,134	10.54
89 年	101,719	14,018	13.78	115,524	17,731	15.35
90 年	121,934	17,715	14.53	123,456	21,911	17.75
91 年	161,351	23,516	14.57	133,047	28,184	21.18
92 年	235,153	32,176	13.68	124,597	35,829	28.76
93 年	268,353	37,758	14.07	138,034	42,805	31.01

表5.11 PO_VI身分證號相同或車牌號碼相同之連結

	事故資料			強制險已決賠資料		
	連結用資料	可連結資料筆數	可連結百分比(%)	連結用資料	可連結資料筆數	可連結百分比(%)
88 年	46,552	8,131	17.47	77,197	9,859	12.77
89 年	101,719	16,888	16.60	115,524	21,623	18.72
90 年	121,934	21,179	17.37	123,456	26,541	21.50
91 年	161,351	28,027	17.37	133,047	33,728	25.35
92 年	235,153	38,250	16.27	124,597	42,163	33.84
93 年	268,353	44,645	16.64	138,034	49,852	36.12

表5.12 PO_VI身分證號或車牌號碼相同，但事故日與出險日相同之連結

	事故資料			強制險已決賠資料		
	連結用資料	可連結資料筆數	可連結百分比(%)	連結用資料	可連結資料筆數	可連結百分比(%)
88 年	46,552	7,341	15.77	77,197	8,764	11.35
89 年	101,719	14,833	14.58	115,524	18,854	16.32
90 年	121,934	18,785	15.41	123,456	23,330	18.90
91 年	161,351	24,607	15.25	133,047	29,550	22.21
92 年	235,153	33,594	14.29	124,597	37,484	30.08
93 年	268,353	39,176	14.60	138,034	44,427	32.19

5.2.2 強制險受害人資料連結之探討

在事故資料中所登記的是所有的事故當事人，而強制險已決賠資料中僅有肇事駕駛人是事故當事人。在 5.2.1 節的連結探討中發現，事故資料中可連結之資料筆數似乎不如預期，是否與強制險已決賠資料中登記之資料有關？若能取得強制險已決賠資料中其他的受害人資料，是否能提高資料連結率？因此本節將就強制險受害人資料，分別與事故資料與強制險已決賠資料進行連結探討。

事故資料與強制險受害人資料的連結(簡稱 PO_VT 連結)，連結用之資料是使用表 3.3 與表 3.8 已完成淨化後 92 年及 93 年之資料。因此兩資料庫間僅有身分證號及日期為共同可連結之欄位，因此將分別就下述兩種方式進行連結探討。

1. 身分證號相同
2. 身分證號相同且事故日與出險日相同。

連結結果顯示，表 5.13 並未比表 5.7 為佳；表 5.13 及表 5.14 中強制險受害人資料可連結成功率約為 30%左右。

表5.13 PO_VT身分證相同之連結

	事故資料			強制險受害人資料		
	連結用資料	可連結資料筆數	可連結百分比(%)	連結用資料	可連結資料筆數	可連結百分比(%)
92 年	235,153	31,365	13.34	192,869	55,948	29.01
93 年	268,353	38,121	14.21	217,762	68,818	31.60

表5.14 PO_VT身分證相同且事故日與出險日相同之連結

	事故資料			強制險受害人資料		
	連結用資料	可連結資料筆數	可連結百分比(%)	連結用資料	可連結資料筆數	可連結百分比(%)
92 年	235,153	28,084	11.94	192,869	51,041	26.46
93 年	268,353	34,169	12.73	217,762	63,114	28.98

在事故資料與強制險受害人資料的連結結果中顯示，事故資料可連結資料並不如與強制險已決賠資料之連結成功率，是否因強制險已決賠資料與強制險受害人資料中仍有未釐清之關係，此處將就此兩資料庫（表 3.7 及表 3.8 之資料）進行連結探討(簡稱 VI_VT 連結)。分析兩者資料庫間之欄位內容時，僅有身分證號及出險日期為可僅連結判斷之欄位，因此將就下述兩種條件進行連結。

1. 身分證號相同
2. 身分證號相同且事故日相同

表 5.15 及表 5.16 為連結後之結果，由資料表中發現兩者間之可連結百分比亦不如預期中高；強制險已決賠資料中僅記錄肇事駕駛人之資料，若駕駛人亦屬於受害人，理當具有較高之連結成功率。

表5.15 VI_VT身分證號相同之連結

	強制險已決賠資料			強制險受害人資料		
	連結用資料	可連結資料筆數	可連結百分比(%)	連結用資料	可連結資料筆數	可連結百分比(%)
92 年	124,597	43,926	35.25	192,869	54,672	28.35
93 年	138,034	48,985	35.49	217,762	61,349	28.17

表5.16 VI_VT身分證號相同且事故日相同之連結

	強制險已決賠資料			強制險受害人資料		
	連結用資料	可連結資料筆數	可連結百分比(%)	連結用資料	可連結資料筆數	可連結百分比(%)
92 年	124,597	42,937	34.46	192,869	53,109	27.54
93 年	138,034	47,935	34.73	217,762	59,533	27.34

5.2.3 事故資料、強制險已決賠資料與健保資料之連結探討

在強制險理賠中就醫資料是一項相當重要之理賠依據，且事故後若有人員受傷，亦應該會有就醫記錄。即便事故後私下和解，而未有事故記錄，在理賠時只要檢具警察之備案資料及就醫資料便可進行理賠，因此本節將嘗試從就醫記錄中探討其間可能之關係。連結用之資料為 90 年至 91 年的事故資料（如表 3.3）、強制險已決賠資料（表 3.7）及健保西醫門診與住院資料（如表 3.10）。

在強制險已決賠資料中，因存有同一次事故多次理賠之情形，而存在身分證號重複之情形，當然也有可能是在不同日期所辦理之出險記錄，為更清楚資料連結之情形，因此對於多筆身分證號且出險日均相同之資料，則視為同一筆記錄。而事故資料則視每一筆資料為獨立事件，連結用資料整理後如表 5.17。

表 5.18 為強制險已決賠資料中肇事駕駛者在事故發生後之就醫人數，由表中可發現，出險後有就醫的比例相當高。但若再分析出險後當天的就醫資料，由表 5.19 則可發現可連結之資料已明顯降低；後續若能使用受害人資料與健保資料進行連結探討，應可更了解出險後受害人之就醫情形，以搭配解釋目前之連結結果。

表 5.20 為事故發後事故當事人之就醫人數，由表中可發現就醫之比例相當高，即便只觀察事故當天之就醫(如表 5.21)，就醫人數佔事故當事人之比例亦相當高。

表5.17 強制險已決賠與健保連結用資料筆數

	強制險已決賠資料		事故資料	健保資料	
	原始資料筆數	篩選後資料筆數		西醫門診	住院
90 年	121,934	91,990	121,934	20,110,812	1,285,117
91 年	161,351	101,463	161,351	17,031,385	

表5.18 強制險已決賠肇事人出險後就醫人數

	強制險資料	西醫門診		住院資料	
	連結用資料 筆數	可連結資料 筆數	可連結百分 比(%)	可連結資料 筆數	可連結百分 比(%)
90 年	91,990	77,199	83.92	28,413	30.89
91 年	101,463	86,387	85.14	30,626	30.18

表5.19 強制險已決賠肇事人出險當天就醫人數

	強制險資料	西醫門診		住院資料	
	連結用資料 筆數	可連結資料 筆數	可連結百分 比(%)	可連結資料 筆數	可連結百分 比(%)
90 年	91,990	11,967	13.01	9,781	10.63
91 年	101,463	13,938	13.74	15,116	14.90

表5.20 事故當事人事故後就醫人數

	事故資料	西醫門診		住院資料	
	連結用資料 筆數	可連結資料 筆數	可連結百分 比(%)	可連結資料 筆數	可連結百分 比(%)
90 年	121,934	97,206	79.72	38,680	31.72
91 年	161,351	128,523	79.65	41,497	25.72

表5.21 事故當事人事故當天就醫人數

	事故資料	西醫門診		住院資料	
	連結用資料 筆數	可連結資料 筆數	可連結百分 比(%)	可連結資料 筆數	可連結百分 比(%)
90 年	121,934	49,781	40.83	18,134	14.87
91 年	161,351	66,077	40.95	22,273	13.80

5.3 與先進安全車輛系統之結合研究

在本年度之研究範圍中，包含提供研究用資料供其他本所合作計畫進行分析。因此在研究過程中由本計畫提供 88 年至 90 年的道路交通事故資料、死因資料及健保資料（門診及住院），供先進安全車輛系統發展之推動與研究使用[26]，該研究針對(1)事故後第一次住院；(2)事故後 30 天內死亡兩部份進行探討，本節中僅摘錄部份研究成果，詳細之應用成果請詳見該研究報告之內容。

5.3.1 事故後第一次住院

圖 5-1 為事故後第一次住院不同車種的撞擊部位次數百分比分佈。由圖 5-1 的數據顯示，不論是大客車、小客車、大貨車與大貨車，主要的撞擊部位均集中在「前保險桿」、「車前」、「左前」以及「右前」等 4 個部位。若是由個別的车種來看，以小客車而言，「前保險桿」的比例最高，其次為「左前」，「右前」次之；以小貨車而言，「左前」的比例最高，其次為「前保險桿」，「車前」及「右前」次之；大貨車則是以「前保險桿」、「車前」及「左前」的比例較高；大客車的部分由於樣本數較少，但整體而言，仍以車輛前方部位的比例較高。此外，撞擊部位未記錄或是無法辨識的比例偏高。

表 5.22 為小客車在不同撞擊部位(包括「前保險桿」、「車前」、「右前」、「左前」、「全毀」以及「未記錄或無法辨認」)於事故後第一次住院的醫療成本分析。由表 5.22 的數據顯示，事故後第一次住院的醫療成本，以小客車全毀為最高，而車輛前方不同撞擊部位的平均醫療成本並無明顯差異。

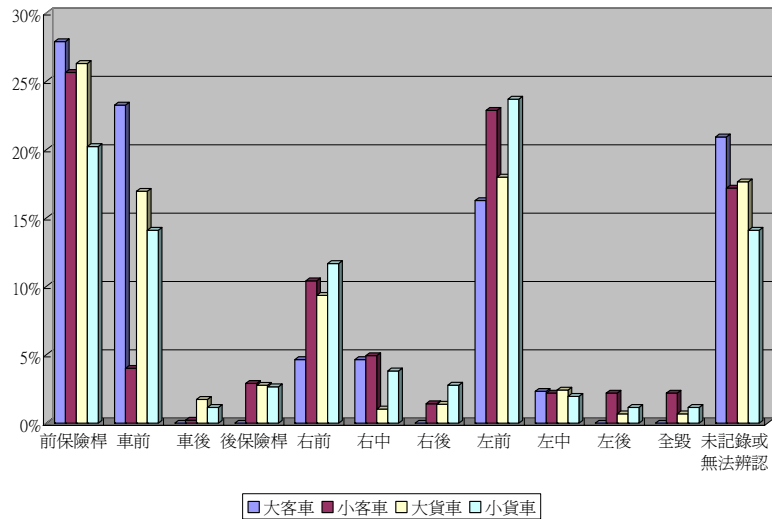


圖 5-1 事故後第一次住院不同車種的撞擊部位次數百分比分佈

表5.22 小客車在不同撞擊部位的事故後第一次住院的醫療成本分析

車種 撞擊部位	小客車		
	平均數	標準差	樣本數
前保險桿	41,022	126,543	722
車前	41,724	78,393	113
右前	39,690	101,494	293
左前	39,224	93,145	644
全毀	50,917	75,889	62
未記錄或無法辨認	39,224	85,142	484

5.3.2 事故後30天內死亡

不同車種的碰撞部位次數百分比分佈如圖 5-2 所示。由圖 5-2 資料顯示，小客車的碰撞部位以「前保險桿」與「左前」為主；小貨車與大貨車的碰撞部位則是以「前保險桿」、「車前」與「左前」為主。

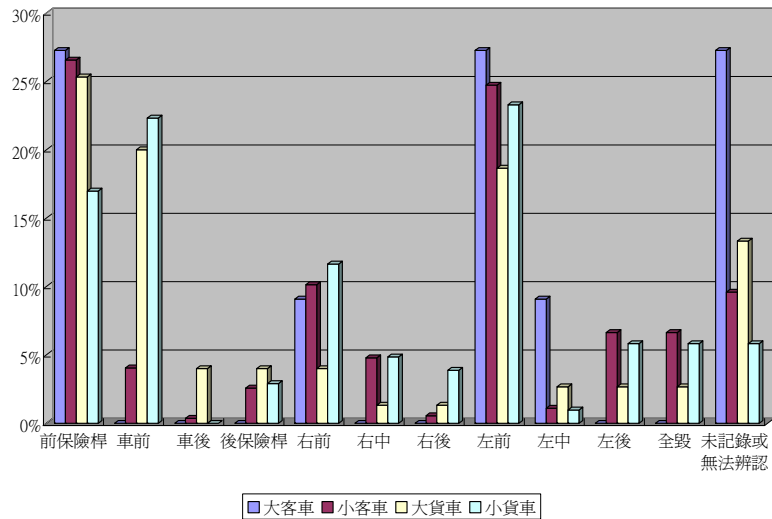


圖 5-2 不同車種的碰撞部位百分比分佈

表 5.23 為小客車發生死亡交通事故，撞擊部位在「前保險桿」與「左前」時，駕駛人的主要受傷部位（AIS 受傷部位）次數分配，由表 5.23 的資料發現，主要受傷部位以「頭部」最高，其次則是「胸膛」。

表5.23 為小客車駕駛發生死亡交通事故的主要受傷部位(AIS)次數分配

撞擊部位 主要受傷部位(AIS)	前保險桿	左前	合計
無效碼	6	9	15
頭部	94	82	176
臉部	3	1	4
頸部	2	0	2
胸膛	28	30	58
腹部及骨盆內	5	5	10
脊椎	1	0	1
表皮、燒傷及其他創傷	1	2	3
無法判斷	4	5	9
合計	144	134	278

5.4 資料探勘技術之未來應用潛力

本計畫對於資料探勘技術之應用研究，雖非本計畫今年預定工作內容，惟為瞭解此技術的未來應用潛力，故於今年計畫中進行可行性之探討，未來則待其他研究進行更深入之探討。在本節中將以資料探勘方法裡的 CHAID、CART 及倒傳遞類神經網路(Back Propagation Neural Networks)嘗試建立影響高醫療費用的機車事故預測模式，來探討影響高醫療費用的機車事故原因。

CHAID 及 CART 是 SPSS 的附掛軟體 AnswerTree 中所提供二項資料探勘工具，因本計畫在去年(93)計畫中已購買 SPSS 及 AnswerTree，故先利用此工具探勘資料。而 SPSS 公司另出版一套資料探勘的專業軟體 Clementine，能進行除了分類樹以外的分析，本研究團隊便利用此軟體中的倒傳遞類神經網路工具，再進行一項資料探勘技術的應用測試。

研究用資料是使用 90 年道路交通事故資料庫與 90 年全民健保資料庫的連結結果，並挑選機車騎士事故後第一次住院者，共 21,691 筆資料。預測的變項包括速限、天候、光線、號誌動作、道路型態、性別、年齡、駕駛執照資格、安全帽、事故發生時間、事故類型、道路種類、路面狀態、當事者區分、事故位置及肇事因素等共 16 項變項。至於因變數醫療費用高低的定義，則採平均值為界線，大於等於平均值者為高，小於平均值則屬低。

5.4.1 CHAID 預測模式之探討

在 CHAID 預測模式的探討上，為了證實所選擇的 16 個變項與醫療成本間具有相互影響的能力。因此，利用了資料探勘中的其中一項分類工具，決策樹(Decision Tree, DT)來判斷這 16 個變項是否具有其影響力(決策流程圖如圖 5-3 所示)，而其所呈現的結果在這 16 個變項中正確分類的機率為 75%(1-0.2477)。雖然一開始所選入的變項是 16 個，但經由卡方檢定之刪除與合併後，只剩下性別、年齡、事故類型、駕駛執照資格、安全帽、時間、路面狀態、道路種類等 8 個變項。部份變項內資料也重新進行分類，分類結果如下：

1. 年齡，由原先設定的七個等級，合併成(1)18~34 歲；(2) 17 歲以下及 35~54 歲；(3)54 歲三大類。
2. 事故類型，將原先三類的事故類型合併成(1)車與車；(2)人與汽(機)車、汽(機)車本身二大類。

3. 時間，由原來 24 個小時的分類，合併成(1)0~6；(2)6~23 二類。
4. 道路種類，由原來六類，合併成(1)市區道路、村里道路、縣道；(2)鄉道、省道、其他二大類。
5. 路面狀態，由原為三類，合併成(1)乾燥、冰雪、油滑、泥濘；(2)濕潤二大類。
而經由其分類後的結果可以從中發現一些規則：
 - (1) 當性別為男性，年齡在 17、35~54 歲內，所產生之醫療費用 69.73%為低費用、30.27%為高費用。
 - (2) 當性別為女性，年齡在 35~54 歲內，所產生之醫療費用 77.88%為低費用、22.12%為高費用。
 - (3) 當性別為女性，年齡在 55 歲以上，所產生之醫療費用 72.7%為低費用、27.3%為高費用。
 - (4) 當性別為男性，年齡在 55 歲以上，有適當駕照者所產生之醫療費用 70.45%為低費用、29.55%為高費用。
 - (5) 當性別為男性，年齡在 55 歲以上，無適當駕照者所產生之醫療費用 61.13%為低費用、38.87%為高費用。
 - (6) 當性別為女性，年齡在 0~34 歲內，事故時間為 6~11、14~20 所產生之醫療費用 86.42%為低費用、13.58%為高費用。
 - (7) 當性別為女性，年齡在 12~34 歲內，事故時間為 0~6 所產生之醫療費用 76.53%為低費用、23.47%為高費用。
 - (8) 當性別為男性，年齡在 18~34 歲內，事故類型為汽(機)車本身，未戴安全帽，所產生之醫療費用 33.33%為低費用、66.67%為高費用。
 - (9) 當性別為女性，年齡在 12~34 歲內，事故時間為 11~14、21~23，有適當駕照者所產生之醫療費用 84.08%為低費用、15.92%為高費用。
 - (10) 當性別為女性，年齡在 12~34 歲內，事故時間為 11~14、21~23，無適當駕照者所產生之醫療費用 83.87%為低費用、16.13%為高費用。
 - (11) 當性別為男性，年齡在 18~34 歲內，事故類型為車與車或人與汽(機)車，資格為有適當駕照，事故時間為 6 點以後所產生之醫療費用 78.19%為低費用、21.81%為高費用。
 - (12) 當性別為男性，年齡在 18~34 歲內，事故類型為車與車或人與汽(機)車，資格為有適當駕照，事故時間為 0~6 點之間所產生之醫療費用 80.11%為低費

用、19.89%為高費用。

(13) 當性別為男性，年齡在 18~34 歲內，事故類型為車與車或人與汽(機)車，資格為無適當駕照，道路種類為市區道路、村里道路或是縣道，所產生之醫療費用 69.38%為低費用、30.62%為高費用。

(14) 當性別為男性，年齡在 18~34 歲內，事故類型為車與車或人與汽(機)車，資格為無適當駕照，道路種類為省道、鄉道、其他，所產生之醫療費用 62.71%為低費用、37.29%為高費用。

(15) 當性別為男性，年齡在 18~34 歲內，事故類型為汽(機)車本身，有戴安全帽，路面狀態為乾燥、冰雪、油滑、泥濘所產生之醫療費用 65%為低費用、35%為高費用。

(16) 當性別為男性，年齡在 18~34 歲內，事故類型為汽(機)車本身，有戴安全帽，路面狀態為濕潤所產生之醫療費用 62.5%為低費用、37.5%為高費用。

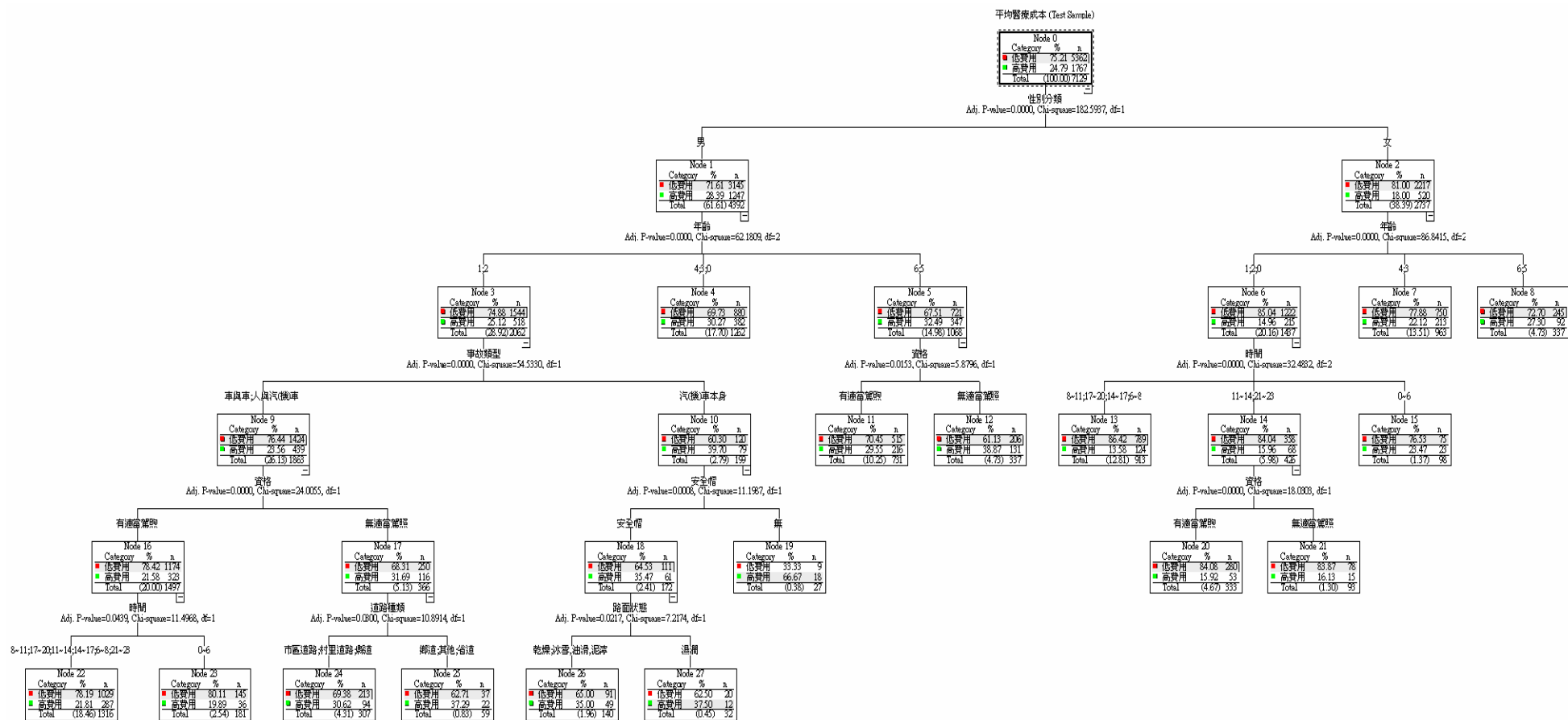


圖 5-3 CHAID 決策樹 (Decision Tree)

5.4.2 CART 預測模式之探討

CART 演算法是一個二元的演算法，應用於連續型的資料型態，每次分割將資料分割為兩個子集合。在預測變項的選擇上，則由原來的 16 個變項，挑選其中的性別、年齡、安全帽及事故類型進行測試。資料的分析是使用 AnswerTree 裡面的 CART 是以 Gini 作為選擇分割條件的依據(分析流程圖如圖 5-4 所示)，而從中得到以下之規則：

1. 當性別為女性，所產生之醫療費用 81%為低費用、18%為高費用。
2. 當性別為男性，年齡在 35 歲以上所產生之醫療費用 68.34%為低費用 31.66%為高費用。
3. 當性別為男性，年齡在 12~34 歲，事故類型為車與車或人與汽(機)車，所產生之醫療費用 76.05%為低費用、23.95%為高費用。
4. 當性別為男性，年齡在 12~34 歲，事故類型為汽(機)車本身，有戴安全帽所產生之醫療費用 64.29%為低費用、35.71%為高費用。
5. 當性別為男性，年齡在 12~34 歲，事故類型為汽(機)車本身，未戴安全帽所產生之醫療費用 32.14%為低費用、67.86%為高費用。

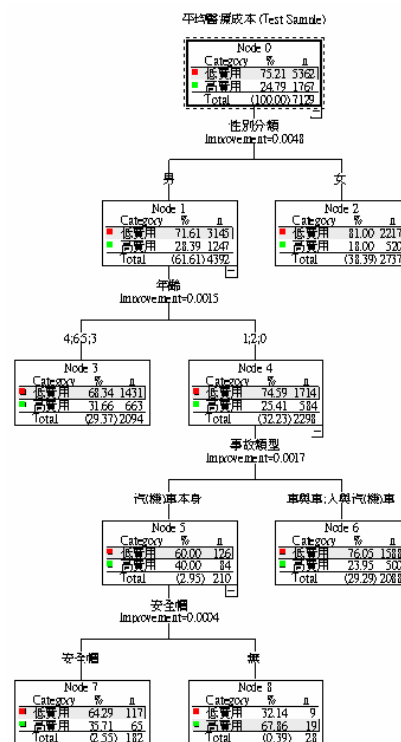


圖 5-4 CART 決策樹 (Decision Tree)

5.4.3 倒傳遞類神經網路預測模式之探討

逆傳遞類神經網路(BPNN)屬於監督式學習(Supervised Learning)網路，適合預測(Prediction)或分類(Classification)等應用。BPNN 由許多單層(Layer)網路所連結，每一層網路由數個神經元 (Neuron) 或稱節點 (Node)組成。在網路中每一個神經元的輸出，都乘上其相對應的加權連結值(Connection Weights)加總後，再透過激發函數(Activation Function)的計算產生輸出訊號(如圖 5-5)。

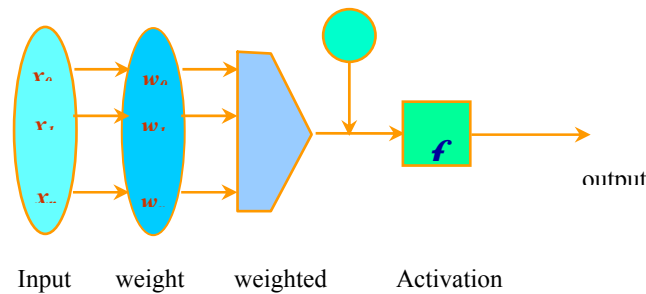


圖 5-5 BPNN 傳遞模式

BPNN 的學習或訓練分為二階段，即輸入傳遞階段與權重訓練階段。輸入傳遞階段以前授(Feed Forward)方式將訓練樣本(Training Example)，自輸入層經由隱藏層傳至輸出層，並計算輸出。權重訓練階段則以逆傳遞(Back Propagation)的方式，將輸出誤差反向傳回各層直到輸入層，以調整各連接權重，以減少輸出誤差。BPNN 使用訓練樣本組，經由多次的訓練迭代(Epoch)而達到最小輸出誤差時，即告訓練完成。

本研究所採用的 BPNN 訓練模型如圖 5-6 所示，由輸入層、隱藏層、與輸出層共三層神經元網路所組成。其架構與演算法說明如下：

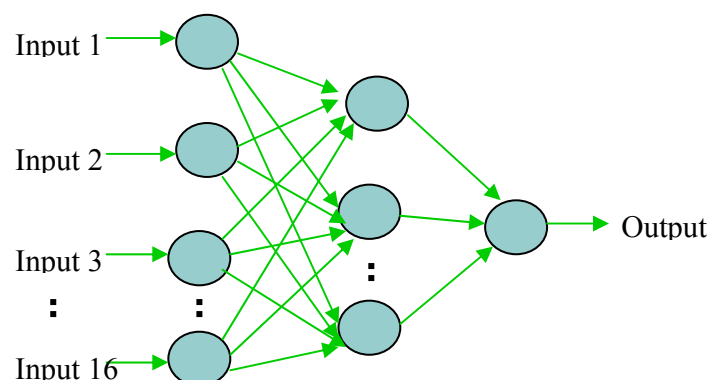


圖 5-6 BPNN 訓練模型

1. 輸入層(Input layer)

輸入(預測)變數為發生時間、速限、天候、光線、道路種類、道路型態、事故位置、事故類型、路面狀態、號誌動作、性別、年齡、安全帽、當事者、資格、肇事因素等 16 項，亦即輸入向量(Input Vector) $X = \{x_i, i = 1, 2, 3, \dots, 16\}$

2. 隱藏層(Hidden layer)

隱藏層共 30 個神經元(Neuron)，每一個神經元($j = 1, 2, 3, \dots, 30$)在第 p 個迭代的輸出 y_j 如下：

$$y_j(p) = f \left[\sum_{i=1}^{16} x_i(p) \cdot w_{ij}(p) - \theta_j \right]$$

其中使用 sigmoid 為 activation function (f)，如圖 5-7 所示； w_{ij} 為與輸入層之第 i 個神經元之連結權重， x_i 為其輸出， θ_j 為 bias。

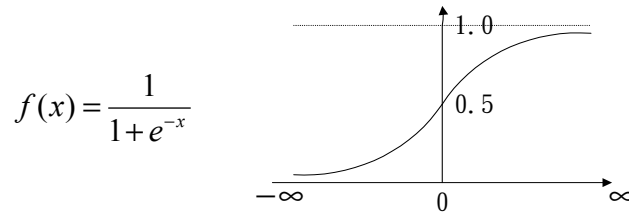


圖 5-7 激發函數圖

3. 輸出層(Output layer)

以平均醫療費用之高低為預測之目標變數，在第 p 個疊代的輸出 $y(p)$ 如下：

$$y(p) = \text{sigmoid} \left[\sum_{j=1}^{30} x_j(p) \cdot w_j(p) - \theta \right]$$

其中 w_j 為與隱藏層之第 j 個神經元之連結權重， θ_j 為 bias。輸出之誤差梯度(Error gradient, δ)為：

$$\delta(p) = y(p) \cdot [1 - y(p)] \cdot e(p)$$

其中 e 為輸出誤差。

4. 權重訓練(Weight training)

輸出的誤差梯度將以逆傳遞(Back propagation)的方式，自輸出至數入方向，層層傳回。每一個連結權重依照下列公式更新：

$$\Delta w_j(p) = \beta \cdot \Delta w_j(p-1) + \alpha \cdot y_j(p) \cdot \delta(p)$$
$$w_j(p+1) = w_j(p) + \Delta w_j(p)$$

其中 α 與 β 分別為學習速率(Learning rate)與動量(Momentum)

5. 參數設定

本研究使用 SPSS Clementine 中之類神經網路模組為發展工具，隨意選取 dataset 中之 70% 資料作為訓練樣本(Training examples)，其餘 30% 作為測試樣本(Testing examples)以防止過度學習(Overfitting)，學習速率 α 設定為 0.01~0.3，動量 $\beta=0.9$ 。

6. 模型建構結果

訓練完成之 BPNN 模型如表 5.24，任何交通事故資料均以其登錄之發生時間、速限、天候、光線、道路種類、道路型態、事故位置、事故類型、路面狀態、號誌動作、性別、年齡、安全帽、當事者、資格、肇事因素欄位做為模型之輸入，模型之輸出能以 88.35% 之準確率預測其醫療費用的高低。

表5.24 BPNN模型

Hidden Layer	30 neurons
Output Layer	1 neuron
Estimated Accuracy	79.13%

經由訓練逆傳遞類神經網路模型，評估各項預測變數之重要性歸類如表 5.25，重要性指標值愈接近 1 愈重要，代表其在此預測模型中對輸出的影響強度。例如，道路種類的重要性最高(0.839)表示此變數內之類別的改變對輸出(醫療費用之高低)的影響，較其他變數類別的改變為大。

表5.25 BPNN模型中預測變數之重要性指標

預測變數	重要性指標
道路種類	0.839
肇事因素	0.723
事故位置	0.718
年齡	0.712
道路型態	0.663
時間	0.656
天候	0.588
光線	0.555
路面狀態	0.503
事故類型	0.481
性別	0.415
號誌動作	0.305
當事者	0.296
安全帽	0.272
道路種類	0.269
資格	0.249

第六章 結論及未來建議

6.1 結論

1. 系統的規劃與建置

本計畫從 2003 年開始至今，已歷經 3 個不同階段的發展，各階段之發展重點如下：

- (1) 雛形系統建置期：著重於資料連結可行性之探討，並利用連結後之資料進行應用研究，了解連結後資料之可用性。
- (2) 資料維護系統建置期：累積前期的資料使用經驗做為系統修改依據，並擴大研究用資料範圍，再納入其他不同領域專家之使用經驗，以做為系統修改之參考，最後並規劃設計一致化的資料連結處理流程。
- (3) 線上查詢系統建置期：為擴大資料使用對象，採 3 層式架構建立線上查詢系統，並使用實體隔離之保護，以確保資料隱私之安全。

在系統設備的建置上，也逐年添購單機伺服器、資料連結用專用伺服器及網頁伺服器，並使用磁碟陣列做為資料儲存，以確保資料使用之安全及增加資料儲存容量。在資訊軟體的建置上，除評估系統建置與系統開發的便利性外，亦考慮後續維護的問題，因此使用微軟的作業系統、資料庫軟體及開發工具，建置本計畫之資訊系統。

2. 「道路交通事故相關資料維護整合系統」

以 2004 年所開發之程式為基礎，並依新取得資料之內容及資料處理之功能需求，修改系統功能，主要完成的內容包括：

- (1) 完成新年度資料的轉換工作，包括 92 年及 93 年道路交通事故資料庫、92 年死因資料庫、92 年及 93 年車輛保險(強制險及任意險)資料庫與強制險受害人資料庫。
- (2) 完成新舊道路交通事故資料格式之轉換，並以新年度資料格式為基礎，採向下相容之處理方式，將舊有資料轉換至新資料格式中。每筆資料均加註所屬的資料年度，以做為後續資料引用時辨識之用。
- (3) 新格式之車輛保險(強制險及任意險)資料，則重新撰寫資料轉換功能，並整併於現有之程式系統中。

- (4) 新取得之強制險受害人資料，亦整併於強制險的資料處理模組中。
- (5) 人車監理違規資料的整併原則，亦整併於系統功能中。
- (6) 系統具有擴充性，可依實際需求新增外掛功能模組，如 ICDMap 計算所需之資料格式轉換，亦採外掛模組之方式整併於現有系統中。
- (7) 系統的操作設計是以各資料庫之需求操作為主，讓維護者清楚掌握各資料庫的操作內容及流程。

3. 「事故關聯資料分析與評估系統」

在 2004 年之研究中，已針對不同資料庫間之特性及關聯，設計「事故關聯資料分析與評估雛形系統」，本年度以此為基礎，依測試結果進行系統功能的修改，主要完成的內容包括：

- (1) 整併重覆之資料輸出功能，修改資料連結流程圖。
- (2) 改善原有的資料連結編碼原則，增加編碼的可讀性，以了解連結過程之變數設定。
- (3) 圖形化的使用界面，使用者可清楚了解資料連結流程及變數設定內容，並可立即檢核連結後資料筆數，以確認連結的正確性。
- (4) 改善系統資料連結處理效能，縮短資料處理時間。

4. 「線上查詢系統」

為能讓「事故關聯資料分析與評估系統」的研究成果擴及至一般人員使用，並增加本計畫成果之應用影響力，因此規劃建置線上查詢系統，目前主要的完成內容包括：

- (1) 收集國內外提供線上查詢功能之網站，參考其資料查詢內容、變數查詢、資料呈現及其他特殊應用之功能，以做為線上查詢系統之規劃參考。
- (2) 主要參考威斯康辛州 CODES 系統所建立的 Cube Analyzer，以提供便利的變數選用功能及資料呈現方式，並參考 Haddon Matrix 之定義，將查詢變數分類為「人」、「車」及「環境」三大類。
- (3) 提供「事故與死因」、「事故與門診」及「事故與住院」三大類的資料查詢。
- (4) 提供類似樞紐分析之功能，由使用者自訂查詢變數及其顯示順序。

- (5) 將變數代碼轉換成實際記錄資料，以增加使用時的可讀性。
- (6) 可由使用者自行下載資料，進行後續的分析研究。

5. 資料連結與機率連結

為改善資料的連結成功率，及「no-chain」部分資料連結的可行性，以評估使用機率連結之可行性，目前之研究成果包括：

- (1) 在所評估的國外機率連結相關軟體及資訊服務中發現，除無法真正了解資料連結過程中的權重值設定原則外，且軟體對於字串資料的處理亦較偏重於英語語系資料，並不適用於雙位元的中文語系，且軟體的售價不便宜，與所能處理的資料量有關，因此並不適合直接引入國內。
- (2) 參考國外進行機率連結處理之經驗，已嘗試使用不同的方式進行資料連結測試，包括：(i)使用身分證號做確定性連結，但性別、出生年月日不同之資料，做門檻值測試；(ii)事故資料中身分證號空白，但有性別、出生年月日之資料；及(iii)死因資料中身分證號檢查碼錯誤，採用前九碼進行連結比對。測試後發現的確可以連結到部分應屬於同一筆之資料。
- (3) 道路交通事故資料與車輛保險資料的連結，依目前取得的資料內容，應可使用身分證號及車牌號碼，在事故日與出險日相同的條件下，所可能得到的最大聯集結果為可成功連結之資料。
- (4) 目前使用身分證號、車牌號碼或是日期所完成的確定性連結，連結後的資料筆數已相當多，且可進行研究探討的議題亦已相當豐富，建議先以確定性連結為主。

6. 長期照護

在 2003 年的研究中，利用事故與健保資料，探討了事故後的就醫特性，並界定出短期醫療(30 天內)的就醫時間框、就醫次數及就醫科別等。但並未對事故後所造成的長期性傷害進行探討，且無明確之時間框界定。目前的探討結果包括：

- (1) 國內外研究對於長期結果並未有明確之時間框定義，且研究多採用評量問卷之方式進行生活品質的調查，但此研究方法並不適合本計畫之進行。

- (2) 由美國 NHTSA 所發展的 FCI 指標，其特色是可以側重短期內對生命威脅影響效果的 AIS(可使用 ICDMap 計算)評估系統銜接外，並可擴展至長期之醫療評估，進而計算長期生活改變後的影響，較適合本計畫所取得之資料。

7. 資料探勘技術之應用

資料探勘是目前頗受歡迎的研究工具，在文獻資料中亦發現有研究應用此技術在醫療資料之分析處理上。本年度研究期間亦針對使用現有資料，利用資料探勘進行應用研究之探討，目前已有初步之成果。但因應用研究並非本計畫今年度之工作內容，故本計畫並不再針對資料探勘技術的應用進行探討與分析，僅著重於可行性之探討及測試。

8. 研討會論文發表

已因應計畫要求，完成並發表兩篇研討會論文於「94 年道路交通安全與執法研討會」及「MIST2005 國際醫療資訊研討會」。兩篇論文名稱為「機車交通事故之醫療費用預測模型研究」[36]及「道路交通事故相關資料整合系統研究」[37]。

6.2 未來建議

歷經 3 年左右之發展，對於各項資料特色及使用均已具備豐富的使用經驗，而各項系統亦已建立整體性的功能架構，並保留系統的擴充彈性，未來將可依需求彈性增加系統功能。在連結後資料應用研究方面，3 年來已發表 6 篇研討會論文，由與會者的回饋亦可了解，本計畫已突破以往概念性或是區域性小範圍的研究內容，而是一個全國具有代表性的資料。本計畫除曾結合不同領域專家學者進行研究探討外，目前亦有國民健康局相關之研究計畫，與本計畫聯繫希望能有更進一步的合作，此乃表示本計畫所建立之系統，已慢慢發揮其效益。對於系統未來之發展，應可朝研究用資料取得、資料系統維護及連結資料應用三方面推動：

1. 研究用資料取得

為使本計畫所建立的系統，能持續發揮其效益，如何維繫資料持續性的取

得是一項重要的工作。目前道路交通事故資料庫及死因資料庫，在資料的取得及使用上已有固定模式，且在資料的處理及連結上，亦已建立基本之處理模式，後續僅需再持續更新即可。

車輛保險資料與事故資料的連結上，從目前所取得的資料來看，似乎仍無法找到好的連結成功率，是資料內容的原因？亦或是實務作業上所造成的結果，應該是可以再繼續深入探討的議題。健保資料庫是本計畫中相當重要的一項資料內容，但本年度研究執行期間，因健保局內部編制之調整，而無法取得新年度之資料，實為可惜。

人車監理違規資料庫在不同工作區所保留的資料範圍不甚相同，且資料清檔作業時間亦不固定，資料庫內容隨時在新增資料或是變更資料狀態，是處在一種動態改變的情形。而人車監理違規資料的連結，是由本計畫所提供的身分證號及車牌號碼進行連結，但本計畫所取得的資料有其固定的作業時間間隔，因此無法即時的提供做資料連結，因此僅能以某一時間點為基準，進行資料連結，再清除重複的資料。建立標準化的資料處理及整併作業流程，亦是後續可努力嘗試之方向。

2. 資料系統維護

目前開發建立的 3 項資訊系統中，其開發過程均採模組化的開發，具有擴充的彈性，可依實際需求做彈性修改。但要維護大量及新增的資料內容時，為方便系統的執行及運作，仍有部分工作有待改善：

- (1) 當要處理大量之資料時，部分工作目前仍須配合 SQL 指令，利用 SQL Server 所提供的執行環境進行資料的操作，未來應可就資料庫的維護及效能調校再進行更深入的探究。
- (2) 若有新的資料格式匯入時，會因資料欄位名稱的改變，而變更資料處理及資料欄位之設定，未來應可朝參數化及一致化的欄位名稱進行改善。

3. 連結資料應用

在研究過程中發現，目前所取得之資料不論是單一資料庫或是連結後之資料，其可進行研究探討之議題相當豐富，即便是無法連結之資料亦存在探討的價值。為讓本計畫所建立之資料，發揮長遠及持續性的影響，應可朝下述 2 方向努

力。將來各項研究成果可透過網站提供線上查詢或是下載服務，成為一個持續性、常態性之服務。

(1) 特定主題及常態性之研究

參考國內外現在所關切的研究議題，每年或是每季進行特定主題之研究成果發表，除可配合政策之發展進行研究議題之探討外，甚至可供未來政策推動之參考，以發揮更深遠之影響。

(2) 與其他單位之研究者之合作

不同單位或是研究者尚有其多樣化的研究需求，應可嘗試結合國內各相關單位，一起合作探討其所關切之議題，甚至與原始資料提供單位建立資源共享之機制，擴大資料之應用價值。

參考文獻

1. Berry, M. J. A., and Linoff, S. G., Data Mining Techniques: for Marketing, Sales, and Customer Support, New York, John Wiley & Sons Inc., 1997.
2. Carol, G., Diana, R., and John, B., Inside the Western Australian Data Linkage System. Symposium on Health Data Linkage. 20-21 March 2002 Tusculum House Auditorium Potts Point, Sydney, NSW.
3. Fellegi, I., and Sunter, A., A theory for Record Linkage, Journal of the American Statistical Association, Vol. 64, pp.1183-1210, 1969.
4. Fu, Y., Data mining task, technique and applications, IEEE POTENTIALS, Oct./Nov. 1997.
5. Gill, L. E., and Baldwin, J. A., Methods and technology of record linkage: some practical considerations, Textbook of medical record linkage, Oxford University press, 1987.
6. Johnson, S. W., and Walker, J., The Crash Outcome Data Evaluation System (CODES), NHTSA Technical Report, DOT HS 808 338, January 1996.
7. Johnson, S. W., Walker, J., and Utter, D., Research Note: Crash Outcome Data Evaluation System (CODES) Project Safety Belts and Helmet Analysis, Washington, DC: Department of Transportation, National Highway Traffic Safety Administration, February 1996.
8. Leicester, E. G., OX-Link: The Oxford medical Record Linkage System, Proceedings of an International Workshop and Exposition, March 20-21 1997, Arlington, VA.
9. MacKenzie, E., Damiano, A., Ditunno, J., Luchter, S., and Miller, T., Development of the Functional Capacity Index (FCI), National Highway Traffic Safety Administration, DOT HS 808 160, July 1994.
10. MacKenzie, E., Damiano, A., Miller, T., and Luchter, S., The development of the Functional Capacity Index, Journal Trauma. Vol. 41, pp. 799-807, 1996.
11. MacKenzie, E., Maria, S., Miller, T., Damiano, A., Luchter, S., and Gotschall, C., The Presentation on The Functional Capacity Index, Center for Injury Research & Policy, Johns Hopkins Bloomberg School of Public Health, 2002.
12. MacKenzie, E., Sacco, W., Luchter, S., Ditunno, J., Staz, C., Gruen, G., Marion, D., and Schwab, W., Validating the Functional Capacity Index as a measure of outcome following blunt multiple trauma, Qual Life Res. Vol. 11, pp. 797-808,

- 2002.
13. McCarthy, M., and MacKenzie, E., Predicting Ambulatory Function Following Lower Extremity Trauma Using the Functional Capacity Index, *Accident Analysis & Prevention*, Vol. 33, pp. 821-831, 2001.
 14. Newcombe, H. B., Kennedy, J. M., and Axford, S. J., Automatic Linkage of Vital Records, *Science*, Vol. 130, pp. 954-959, 1959.
 15. NHTSA and CIREN Center Staffs, NHTSA Crash Injury Research and Engineering Network (CIREN) Program Report, National Highway Traffic Safety Administration, DOT HS 809 377, December 2001.
 16. NHTSA and CIREN Center Staffs, NHTSA Crash Injury Research and Engineering Network (CIREN) Program Report, National Highway Traffic Safety Administration, DOT HS 809 564, October 2003.
 17. Schluter, P. J., Neale, R., Deborah, S., Luchter, S., and McClure, R. J., Validating the Functional Capacity Index: A Comparison of Predicted versus Observed Total Body Scores, *Journal Trauma*. Vol. 58, pp. 259–263, 2005.
 18. Sung, H. H., and Sang, C. P., Application of Data Mining Tools to Hotel Data Mart on the Intranet for Database Marketing, *Expert Systems With Applications*, Vol. 15, pp. 1-31, 1998.
 19. Vijaya, S., Toni, M. H., Michael, A., and Ric, M., Linkage of the Victorian Admitted Episodes Dataset, Symposium on Health Data Linkage, 20-21 March 2002 Tusculum House Auditorium Potts Point, Sydney, NSW.
 20. World Health Organization, World report on road traffic injury prevention, World Health Organization, 2004.
 21. 交通部運輸研究所，易肇事地點改善作業手冊之教育訓練計畫，2003。
 22. 交通部運輸研究所，道路交通事故相關資料連線系統分析與設計之研究，2000 年。
 23. 交通部運輸研究所，道路交通事故相關資料整合系統雛形建置研究(I)-基本雛形環境之建置，2004。
 24. 交通部運輸研究所，道路交通事故相關資料整合系統雛形建置研究(II)－資料與系統擴充(1/2)，2005。
 25. 交通部運輸研究所，先進安全車輛系統發展之推動與研究(I)，2004。
 26. 交通部運輸研究所，先進安全車輛系統發展之推動與研究(II)，2006。(初稿)

27. 交通部運輸研究所，應用駕駛模擬器開發智慧型運輸系統實驗平台之軟硬體規劃設計(2/4)，2005。
28. 交通部運輸研究所，高齡者交通事故之短期醫療支出研究，2005。
29. 交通部運輸研究所，道路交通事故整合資料庫之應用研究 2004。(初稿)
30. 黃維信、鄭銘章、林豐福，道路交通事故傷害之門診醫療分析探討，2003年健康與管理學術研討會，2003。
31. 陳苑蕙、許峻嘉、黃維信、賴靜慧，高齡者道路交通事故受傷嚴重性與醫療成本分析，道路安全與交通事故傷害研討會，2004。
32. 馬惠明、曾詩雯、賴靜慧，利用全民健保資料分析機動車交通事故傷害之傷害嚴重度—初步探討，道路安全與交通事故傷害研討會，2004。
33. 林豐福、賴靜慧、黃維信，道路交通事故整合資料庫與其應用之介紹，道路安全與交通事故傷害研討會，2004。
34. 陳苑蕙、黃維信、賴靜慧，駕駛者住院受傷嚴重性與住院成本之探討，道路交通安全與執法國際研討會，2004。
35. 黃維信、賴靜慧，肇事駕駛者交通事故與違規記錄關連性之探討，道路交通安全與執法國際研討會，2004。
36. 蔡興國、廖彥琪、黃維信、朱薇薇，機車交通事故之醫療費用預測模型研究，94年道路交通安全與執法研討會，2005/9/29，pp. 209-220，2005。
37. 黃維信、賴靜慧、董基良、鄭銘章，道路交通事故相關資料整合系統研究，MIST2005 國際醫療資訊研討會。
38. 蔡宜蓉，家庭照顧者的自我檢視--從活動健康與生活品質的面向看，http://www.familycare.org.tw/letter38_2.htm，2005 年 6 月查詢。
39. 姚開屏，簡介與評論常用的一般性健康相關生活品質量表兼談對未來研究的建議，中國測驗學會測驗年刊，第四十七卷，第二期，第 111-138 頁，2000 年。
40. 林傑斌、劉明德，資料採掘與 OLAP 理論與實務，文魁資訊出版，2002 年。
41. Berry, M. J. A., and Linoff, S. G., 著，彭文正譯，資料採礦－顧客關係管理暨電子行銷之應用，數博網資訊股份有限公司，台北，2001。
42. 曾義光，掌握潛在顧客－以網際探勘(Web Mining)幫助決策判斷，網際先鋒，pp. 32-36，2000。

43. 林信忠，資料發掘技術應用于健保醫療費用稽核之研究，元智大學管理研究所碩士論文，1998。
44. 洪哲倫，探討基植於時間序列上的群組化技術，逢甲大學資訊工程學系碩士論文，2000。
45. 蔣肇慶，代理人程式在門診醫療資訊系統的應用，國立中央大學資訊管理學系研究所碩士論文，1997。
46. 周賢昭，以資料礦探技術發展臨床路徑之研究，國立中山大學資訊管理學系碩士論文，1998。
47. 張博竣編著，資訊安全管理實務，松崗出版社，台北，2004。
48. 凌威科技，RAID5 磁碟陣列架構原理與介紹，
http://www.linwei.com.tw/raid_2.html，2005 年 6 月查詢。
49. <http://ai.fmcsa.dot.gov/mcspa.asp>，2005 年 6 月查詢。
50. http://app.idph.state.il.us/emsrpt/to_crash-forms.asp，2005 年 6 月查詢。
51. <http://care.cs.ua.edu/onlineanalysis.aspx>，2005 年 6 月查詢。
52. <http://linear.chsra.wisc.edu/CODES/Default.htm>，2005 年 6 月查詢。
53. <http://www.bast.de/irtad/>，2005 年 6 月查詢。
54. <http://www.cdc.gov/ncipc/wisqars/>，2005 年 9 月查詢。
55. <http://www.doh.gov.tw/statistic/index.htm>，2005 年 6 月查詢。
56. <http://www-fars.nhtsa.dot.gov/main.cfm>，2005 年 2 月查詢。
57. <http://www.keishicho.metro.tokyo.jp/>，2005 年 6 月查詢。
58. <http://www.moea.gov.tw/>，2005 年 6 月查詢。
59. <http://www.npa.gov.tw>，2005 年 6 月查詢。
60. <http://www.statcan.ca/>，2005 年 6 月查詢。
61. <http://www.stat.gov.tw>，2005 年 6 月查詢。

附錄 1、事故資料新舊欄位對照表

基本事故資料(整合資料格式)欄位說明 KAM00

整合資料檔案格式				舊資料檔案格式				整合格式備註	舊格式備註
序號	欄	項 稱	英文欄位名稱	型態	長度	中文欄位說明	英文欄位名稱(表格)	型態	長度
1	主鍵		KAM00_MN_KEY	文字	16	主鍵	KAM01_MN_KEY	文字	16
2	事故年月		KAM00_YM	文字	5	發生時間_年 發生時間_月	KAM01_YEAR(01) KAM01_MONTH(01)	int	4
3	總編號		KAM00_NO1	文字	5	類別 總編號	KAM01_NO_TYPE(01) KAM01_NO_SEQ(01)	char	1
4	處理單位代碼		KAM00_POL_NO	文字	4	處理警局代碼	KAM01_POL_NO(01)	char	4
5	處理編號		KAM00_NO2	文字	4	處理編號	KAM01_NO2(01)	char	4
6	發生時間	年	KAM00_YEAR	數字	4	發生時間_年	KAM01_YEAR(01)	int	4
7		月	KAM00_MONTH	數字	2	發生時間_月	KAM01_MONTH(01)	int	4
8		日	KAM00_DATE	數字	2	發生時間_日	KAM01_DATE(01)	int	4
9		時	KAM00_HOUR	數字	2	發生時間_時	KAM01_HOUR(01)	int	4
10		分	KAM00_MINT	數字	2	發生時間_分	KAM01_MINT(01)	int	4

整合資料檔案格式					舊資料檔案格式					整合格式備註	舊格式備註
序號	欄	項	名稱	英文欄位名稱	型態	長度	中文欄位說明	英文欄位名稱(表格)	型態	長度	
11			星期	KAM00_DAY	數字	2	發生時間_星期	KAM01_DAY(01)	int	4	由萬年曆產生
12	發生地點(街道)-縣市			KAM00_CITY	文字	6	縣市	KAM01_CITY(01)	varchar	6	
13	(1)發生地點-市區鄉鎮			KAM00_TOWN	文字	20	市區鄉鎮	KAM01_TOWN(01)	varchar	20	
14	(1)發生地點-村里			KAM00_LI	文字	10	地址	KAM01_ADD(01)	varchar	60	將舊格式地址資料以資料剖析方式填入新格式中
15	(1)發生地點-鄰			KAM00_LIN	文字	4					
16	(1)發生地點-街路			KAM00_STREET1	文字	16					
17	(1)發生地點-街路段			KAM00_SEC1	文字	4					
18	(1)交叉路			KAM00_STREET2	文字	16					
19	(1)發生地點-街路段			KAM00_SEC2	文字	4					
20	(1)發生地點-地點			KAM00_POS	文字	60					
21	(2)發生地點(路線里程)-公路			KAM00_HWY	文字	10	發生地點(公路)-線	KAM01_HWY(01)	char	10	
22	(2)發生地點-公里			KAM00_HWY_KM	文字	6	發生地點(公路)-公里	KAM01_HWY_KM(01)	char	6	
23	(2)發生地點-公尺			KAM00_HWY_M	文字	6	發生地點(公路)-公尺	KAM01_HWY_M(01)	char	6	

整合資料檔案格式				舊資料檔案格式				整合格式備註	舊格式備註
序號	欄 項 稱	英文欄位名稱	型態	長度	中文欄位說明	英文欄位名稱(表格)	型態	長度	
24	(2)發生地點-車道 方向	KAM00_WAY	文字	2	發生地點(公 路)-方向車道	KAM01_WAY(01)	char	2	
25	(3)發生地點(平交 道)-平交道	KAM00_TRN	文字	10	發生地點(平 交道)-線	KAM01_TRN(01)	char	10	
26	(3)發生地點-公里	KAM00_TRN_KM	文字	6	發生地點(平 交道)-公里	KAM01_TRN_KM(01)	char	6	
27	(3)發生地點-公尺	KAM00_TRN_M	文字	6	發生地點(平 交道)-公尺	KAM01_TRN_M(01)	char	6	
28	(1)發生地點-平交 道名稱	KAM00_INTRSEC	文字	10	發生地點(平 交道)-平交道	KAM01_INTRSEC(01)	char	10	
29	(1)死亡人數	KAM00_DEATH	數字	3	死亡人數	KAM01_DEATH(02)	int	4	
30	(2)受傷人數	KAM00_INJURE	數字	3	受傷人數	KAM01_INJURE(02)	int	4	
31	天候	KAM00_WEATHE R	文字	2	天候	KAM01_WEATH(02)	int	4	
32	光線	KAM00_LIGHT	文字	2	光線	KAM01_LIGHT(02)	int	4	
33	道路類別	KAM00_ROAD	文字	2	道路類別	KAM01_ROAD(02)	int	4	
34	速限	KAM00_SPEED	數字	3	速限	KAM01_SPEED(02)	int	4	
35	道路型態	KAM00_ROADTP	文字	2	道路型態	KAM01_ROADTP(02)	char	1	代碼需轉換，詳 見代碼轉換表。 新增 18 隧道(地 下道)

整合資料檔案格式					舊資料檔案格式					整合格式備註	舊格式備註
序號	欄	項 稱	英文欄位名稱	型態	長度	中文欄位說明	英文欄位名稱(表格)	型態	長度		
36	事故位置		KAM00_ACC_PLA CE	文字	2	事故位置	KAM01_ACC_PLACE(02)	char	1		代碼需轉換，詳 見代碼轉換表。 新增22 高速公路 路肩
37	(1)路面狀況-路面 鋪裝		KAM00_RD_PAV	文字	1	-	-	-	-		
38	(2)路面狀況-路面 狀態		KAM00_RD_CON	文字	1	路面狀態	KAM01_RD_CON(02)	char	1		
39	(3)路面狀況-路面 缺陷		KAM00_RD_FLW	文字	1	路面缺陷	KAM01_RD_FLW(02)	char	1		
40	(1)道路障礙-障礙 物		KAM00_OBST	文字	1	障礙	KAM01_OBST(02)	char	1		
41	(2)道路障礙-視距		KAM00_SIGHT	文字	1	視距	KAM01_SIGHT(02)	char	1		
42	(1)號誌-號誌種類		KAM00_SG_TYPE	文字	1	號誌種類	KAM01_SG_TYPE(02)	char	1		
43	(1)號誌-號誌動作		KAM00_SG_ACT	文字	1	號誌動作	KAM01_SG_ACT(02)	char	1		
44	車道劃分設施-分向 設施		KAM00_LANE	文字	2	車道劃分設施-快 -分向設施	KAM01_LANE(02)	char	1		代碼需轉換，詳 見代碼轉換表。
45	(1)分道設施-快車 道間		KAM00_LN_FAST	文字	1	分道設施-快 車道間	KAM01_LN_FAST(02)	char	1		
46	(2)分道設施-快慢 車道間		KAM00_LN_SLW	文字	1	分道設施-快 慢車道間	KAM01_LN_SLW(02)	char	1		
47	(3)分道設施-路面 邊緣		KAM00_LN_BDR	文字	1	分道設施-路 面邊緣	KAM01_LN_BDR(02)	char	1		代碼需轉換，詳 見代碼轉換表。

整合資料檔案格式				舊資料檔案格式				整合格式備註	舊格式備註
序號	欄 項 稱	英文欄位名稱	型態	長度	中文欄位說明	英文欄位名稱(表格)	型態	長度	
48	事故類型及型態	KAM00_ACC_TYP E	文字	2	事故類型	KAM01_ACC_TYPE(0 2)	char	1	代碼需轉換，詳 見代碼轉換表。
49	主要肇事因素	KAM00_MN_RSN	文字	2	主要肇事因素	KAM01_MN_RSN(01)			
50	資料輸入時間	KAM00_INPUT_T M	日期/ 時間		-	-	-	-	
51	流水號(舊)	KAM00_SEQNO	文字	6	流水號	KAM01_SEQNO(01)	char	6	
52	卡別(舊)	KAM00_20PSLNO	文字	2	卡別	KAM01_20PSLNO(01)	char	2	
53	警察局名稱(舊)	KAM00_POL_DP	文字	20	警察局名稱	KAM01_POL_DP(01)	char	20	
54	分局名稱(舊)	KAM00_POL_BR	文字	16	分局名稱	KAM01_POL_BR(01)	char	16	
55	專號案	KAM00_NO3	文字	2	專號案	KAM01_NO3(01)	char	2	新的由 KAM00_NO1 判 斷
56	巡邏否(舊)	KAM00_PATH	文字	1	巡邏否	KAM01_PATH(01)	char	1	
57	地址(舊)	KAM00_ADD	文字	60	地址	KAM01_ADD(01)	varchar ar	60	
58	道路代碼	KAM00_HWY_CO DE	文字	5	道路代碼	KAM01_HWY_CODE(01)	char	5	新的需由 KAM00_HWY 及 KAM00_RAOD 進行轉換
59	車輛損失(舊)	KAM00_CAR_LOS T	文字	1	車輛損失	KAM01_CAR_LOST(0 1)	char	1	

整合資料檔案格式							舊資料檔案格式				整合格式備註	舊格式備註
序號	欄	項	名稱	英文欄位名稱	型態	長度	中文欄位說明	英文欄位名稱(表格)	型態	長度		
60	(舊)			KAM00_CARDAMG	數字	2		KAM01_CARDAMG(01)	int	2		
61	(舊)			KAM00_MEDICARE	數字	2		KAM01_MEDICARE(01)	int	2		
62	(舊)			KAM00_OTHER	數字	2		KAM01_OTHER(01)	int	2		
63	快車道數(舊)			KAM00_FASTRD	數字	4	快車道數	KAM01_FASTRD(02)	int	4		
64	特定場所(舊)			KAM00_SPEC_PL	文字	1	特定場所	KAM01_SPEC_PL(02)	char	1		
65	路面狀況-慢車道鋪面有無(舊)			KAM00_RD_SL	文字	1	路面狀況-慢車道鋪面有無	KAM01_RD_SL(02)	char	1		
66	快車道鋪裝(舊)			KAM00_RD_FST	文字	1	快車道鋪裝	KAM01_RD_FST(02)	char	1		
67	標誌(舊)			KAM00_MARK	文字	1	標誌	KAM01_MARK(02)	char	1		
68	未滿14歲兒童事故-上下學(舊)			KAM00_KID_SCH	數字	4	未滿14歲兒童學童事故-上下	KAM01_KID_SCH(02)	int	4		
69	未滿14歲兒童事故-旅遊中(舊)			KAM00_KID_TRV	數字	4	未滿14歲兒童中童事故-旅遊	KAM01_KID_TRV(02)	int	4		
70	未滿14歲兒童事故-其他(舊)			KAM00_KID_OTH	數字	4	未滿14歲兒童童事故-其他	KAM01_KID_OTH(02)	int	4		
71	未滿14歲兒童事故-保護者在(舊)			KAM00_KID_GUA	數字	4	未滿14歲兒童者在童事故-保護	KAM01_KID_GUAD(02)	int	4		

整合資料檔案格式					舊資料檔案格式				整合格式備註	舊格式備註
序號	欄	項	名稱	英文欄位名稱	型態	長度	中文欄位說明	英文欄位名稱(表格)	型態	長度
72	未滿14歲兒童事故-獨自一人(舊)			KAM00_KID_SELF	數字	4	未滿14歲兒童事故-獨自一人	KAM01_KID_SELF(02)	int	4
73	未滿14歲兒童事故-有友伴(舊)			KAM00_KID_FRD	數字	4	未滿14歲兒童事故-有友伴	KAM01_KID_FRD(02)	int	4
74	(舊)			KAM00_FILLER_B	文字	44		KAM01_FILLER_B(01)	varchar	44

當事人事故資料(整合資料格式)欄位說明 KAM01

整合資料檔案格式				舊資料檔案格式				整合格式備註	舊格式備註
序號	欄 項 名 稱	英文欄位名稱	型態	長度	中文欄位說明	英文欄位名稱(表格)	型態	長度	
1	主鍵	KAM01_MN_KEY	文字	16	主鍵	KAM01_MN_KEY	文字	16	新資料主鍵由 KAM01_YM, KAM01_NO1,KA M01_POL_NO 三 欄組成
2	事故年月	KAM01_YM	文字	5	發生時間_年	KAM03_ EVENT_YEAR(03)	int	4	
3	總編號	KAM01_NO1	數字	5	發生時間_月	KAM03_EVENT_MONT H(03)	int	4	
4	發生時間_年	KAM01_EVENT_YEAR	數字	4	類別	KAM01_NO_TYPE(01)	char	1	
5	發生時間_月	KAM01_EVENT_MONTH	數字	2	序號	KAM01_NO_SEQ(01)	char	4	
6	處理單位代號	KAM01_POL_NO	文字	4	發生時間_年	KAM03_ EVENT_YEAR(03)	int	4	
7	第幾當事者	KAM01_PERSON	文字	2	發生時間_月	KAM03_EVENT_MONT H(03)	int	4	
8	中文姓名	KAM01_CNAME	文字	10	處理警局代碼	KAM01_POL_NO(01)	char	4	
9	英文姓名	KAM01_ENAME	文字	30	當事人別	KAM03_PERSON(03)	char	2	
					當事人中文名	KAM03_CNAME(03)	char	10	
					當事人英文名	KAM03_ENAME(03)	char	20	

整合資料檔案格式					舊資料檔案格式					整合格式備註	舊格式備註
序號	欄 項 名 稱	英文欄位名稱	型態	長度	中文欄位說明	英文欄位名稱(表格)	型態	長度			
10	屬(性)別	KAM01_SEX	文字	1	性別	KAM03_SEX(03)	char	1			
11	身分證字號	KAM01_ID	文字	10	身分證字號	KAM03_ID(03)	char	10			
12	出生日期民國前/國	KAM01_ERA	文字	1	出生日期民國前/國	KAM03_ERA(03)	char	1	0:前	1:國	代碼需轉換，詳見代碼轉換表。
13	出生日期	年 KAM01_YEAR	數字	4	出生日期_年	KAM03_YEAR(03)	int	4	民國		
14		月 KAM01_MONTH	數字	2	出生日期_月	KAM03_MONTH(03)	int	4			
15		日 KAM01_DATE	數字	2	出生日期_日	KAM03_DATE(03)	int	4			
16	當事人住址	KAM01_ADD	文字	86	當事人住址	KAM03_ADD(03)	char	86			
17	當事人電話	KAM01_TEL	文字	20	-	-	-	-			
18	受傷程度	KAM01_HURT	文字	1	受傷程度	KAM03_HURT(03)	char	1			
19	主要傷處	KAM01_HURT_PL	文字	2	主要傷處	KAM03_HURT_PL(03)	char	1			代碼需轉換，詳見代碼轉換表。
20	保護裝備	KAM01_SAFETY	文字	1	保護裝備	KAM03_SAFETY(03)	char	1			代碼需轉換，詳見代碼轉換表。新增 5 無
21	行動電話	KAM01_MPHONE	文字	1	-	-	-	-			

整合資料檔案格式					舊資料檔案格式				整合格式備註	舊格式備註
序號	欄 項 名 稱	英文欄位名稱	型態	長度	中文欄位說明	英文欄位名稱(表格)	型態	長度		
22	當事者區分 (類別)	KAM01_KIND	文字	2	當事者區分 (類別)	KAM03_KIND(03)	char	2		
23	臨時牌照	KAM01_AD_PLT_NO	文字	2	-	-	-	-		
24	當事人車輛牌 照號碼	KAM01_PLT_NO	文字	10	車牌號碼(使 用)	KAM03_PLT_TP(03)	char	2		
25	車輛用途	KAM01_CAR_USE	文字	1	車牌號碼	KAM03_PLT_NO(03)	char	8		
26	當事者行動狀 態	KAM01_MOVE	文字	2	砂石車	KAM03_TRUCK(03)	char	1		
					當事者行動狀 態	KAM03_MOVE(03)	char	1		

整合資料檔案格式					舊資料檔案格式				整合格式備註	舊格式備註
序號	欄 項 名 稱	英文欄位名稱	型態	長度	中文欄位說明	英文欄位名稱(表格)	型態	長度		
27	駕駛資格情形	KAM01_QUAL	文字	1	駕駛資格情形	KAM03_QUAL(03)	char	1		代碼需轉換，詳見代碼轉換表。新增 9 駕照吊扣或吊銷
28	駕駛執照種類	KAM01_DR_LICE	文字	2	駕駛執照種類	KAM03_DR_LICE(03)	char	1		代碼需轉換，詳見代碼轉換表。
29	飲酒情形	KAM01_DRINK	文字	1	-	-	-	-		
30	車輛撞擊部位-最初	KAM01_IN_CARPRT	文字	2	-	-	-	-		
31	車輛撞擊部位-其他	KAM01_OT_CARPRT	文字	2	-	-	-	-		
32	肇因研判	KAM01_RSN	文字	2	肇事因素	KAM03_RSN(03)	char	2		代碼需轉換，詳見代碼轉換表。
33	主要肇因研判	KAM01_MN_RSN	文字	2	主要肇事因素	KAM03_MN_RSN(03)	char	2		代碼需轉換，詳見代碼轉換表。
34	肇事逃逸	KAM01_ACC_RUN	文字	1	-	-	-	-	1:否,2:是	
35	職業	KAM01_OCCU	文字	2	-	-	-	-		
36	旅次目的	KAM01_TRIPURP	文字	1	-	-	-	-		
37	國籍	KAM01_NATION	文字	1	-	-	-	-	0:本國 1:外國	由 KAM03_OCCU 判別

整合資料檔案格式					舊資料檔案格式					整合格式備註	舊格式備註
序號	欄 項 名 稱	英文欄位名稱	型態	長度	中文欄位說明	英文欄位名稱(表格)	型態	長度			
38	備註	KAM01_ENDORSE	文字	4	備註人車圖號	KAM03_GRAPHIC(03)	char	4			
39	車種	KAM01_FILLER_A	文字	2	車種	KAM03_FILLER_A(03)	char	2			
40	輸入時間	KAM01_INPUT_TM	日期/ 時間		-	-	-	-			
41	UID(舊)	KAM01_UID	文字	8	UID	KAM03_UID(03)	char	8			
42	流水號(舊)	KAM01_SEQNO	文字	6	流水號	KAM03_SEQNO(03)	char	6			
43	年齡	KAM01_AGE	數字	3	年齡	KAM03_AGE(03)	int	3			
44	教育(舊)	KAM01_EDU	文字	1	教育	KAM03_EDU(03)	char	1			
45	飲酒情形(舊)	KAM01_O_DRINK	文字	1	飲酒情形	KAM03_DRINK(03)	char	1			
46	車輛受損部位 (舊)	KAM01_O_CARPRT	文字	6	車輛受損部位	KAM03_CARPRT(03)	char	6			
47	行業(舊)	KAM01_O_OCCU	文字	1	行業	KAM03_OCCU(03)	char	1			
48	(舊)	KAM01_FILLER_B	文字	2		KAM03_FILLER_B(03)	char	2			

附錄 2、車輛保險賠案資料欄位說明

(一) 強制車險賠案賠款資料欄位

項 目 名 稱		備 註
5	賠案受理日期	
5.1	受理年(西元)	
5.2	受理月	
5.3	受理日	
7	賠付／追償次數	一賠案的總處理次數
9	結案日期	
13	被保險車輛資料	
13.2	原始發照年(西元)	
13.3	製造年份(西元)	
13.4	廠牌車型代號	
13.5	車輛種類代號	
13.6	排氣量	
13.7	引擎/車身號碼	
13.8	牌照號碼	
15	本期違規肇事理賠等級	
16	總保險費	
16.1	保費＋／－記號	
16.2	總保險費	
17	調整後純保險費	
17.1	保費＋／－記號	
17.2	調整後純保險費	
18	業務費用	
18.1	費用＋／－記號	
18.2	業務費用	
19	費率實施年月	
19.1	費率實施年(西元)	
19.2	費率實施月	
20	本車被保險人資料欄	
20.1	身分證統一編號	
20.2	出生年份(西元)	
20.3	性別代號	
20.4	婚姻別代號	

21	本車肇事駕駛人資料欄	
21.1	身分證統一編號	
21.2	出生年份(西元)	
21.3	性別代號	
21.4	婚姻別代號	
22	本車肇事駕駛人與被保險人關係	
23	出險原因代號	
24	本車肇事責任代號	
25	肇事責任百分比資料	
25.1	本車肇事責任百分比	
25.2	對方車肇事責任百分比	
25.3	其他肇事責任百分比	
27	出險日期	
27.1	出險年(西元)	
27.2	出險月	
27.3	出險日	
28	出險地區代號	
29	給付類別代號	
30	每人保險金額	
30.1	保額＋／－記號	
30.2	保險金額(元)	
31	理賠人數	
33	未含健保理賠金額	
33.1	金額＋／－記號	
33.2	未含健保理賠金額(元)	
34	理賠費用	
34.1	費用＋／－記號	
34.2	理賠費用(元)	
35	給付健保金額	
35.1	金額＋／－記號	
35.2	給付健保金額(元)	
38	賠付/追償日期	
38.1	賠付/追償年(西元)	
38.2	賠付/追償月	
38.3	賠付/追償日	

(二) 強制車險受害人賠款資料欄位

項 目 名 稱		備 註
5	賠案受理日期	
5.1	受理年(西元)	
5.2	受理月	
5.3	受理日	
7	賠付／追償次數	一賠案的總處理次數
12	受害人資料欄	
12.1	身分證統一編號	
12.2	出生年份(西元)	
12.3	性別代號	
12.4	出事當時乘坐狀況	
12.5	傷亡情形	
13	出險日期	
13.1	出險年(西元)	
13.2	出險月	
13.3	出險日	
14	出險地區代號	
16	出險原因代號	
17	給付類別代號	
18	每人保險金額	
18.1	保額＋／－記號	
18.2	保險金額(元)	
20	未含健保理賠金額	
20.1	金額＋／－記號	
20.2	未含健保理賠金額(元)	
21	理賠費用	
21.1	費用＋／－記號	
21.2	理賠費用(元)	
22	給付健保金額	
22.1	金額＋／－記號	
22.2	給付健保金額(元)	
24	賠付/追償日期	
24.1	賠付/追償年(西元)	
24.2	賠付/追償月	
24.3	賠付/追償日	

(三) 任意險賠案賠款資料欄位

項 目 名 稱	備 註
3 . 賠案受理日期	
5 . 賠付／追償次數	分險種的總次數
1 0 . 原始發照年月	
1 1 . 製造年份	
1 2 . 廠牌車型代號	
1 3 . 車輛種類代號	
1 4 . 排氣量	
1 5 . 引擎/車身號碼	
1 6 . 牌照號碼	
1 8 . 駕駛人資料欄	
(1) 駕駛人駕照號碼	
(2) 駕駛人出生年份	
(3) 駕駛人性別	
(4) 駕駛人婚姻別	
1 9 . 出險原因代號	分險種
2 0 . 被保險人資料欄	
(1) 被保險人身分證統一編號	
(2) 被保險人出生年份	
(3) 被保險人性別	
(4) 被保險人婚姻別	
2 2 . 出險日期	
2 3 . 出險地區代號	
2 4 . 承保地區代號	
2 7 . 保險種類代號	
2 8 . 給付類別代號	分險種
3 0 . 被保險人賠款紀錄	分險種
(1) 賠款紀錄係數＋／－號	
(2) 賠款紀錄係數	
3 1 . 保險金額	分險種
(1) 保額＋／－記號	
(2) 保險金額	
3 2 . 自負額	分險種
(1) 自負額單位類別代號	
(2) 自負額數量	
3 3 . 實際理賠人數	
3 5 . 理賠自負額	分險種
3 6 . 實際理賠金額	分險種
(1) 金額＋／－記號	
(2) 實際理賠金額	
3 7 . 理賠費用	分險種

項 目 名 稱	備 註
(1) 費用 + / - 記號	
(2) 理賠費用	
4 1 . 賠付日期	分險種

附錄 3、人車監理違規資料欄位說明

欄 位 名 稱	型 態	寬 度	欄 位 說 明
XACT_NO	char	10	流水號
PLT_NO	char	9	車號*
VIL_DT	char	7	違規日
VIL_TIME	char	5	違規時間
VIL_ADDR	char	6	違規地址
REG_SEQ	char	11	車籍序號
SPRVSN_NO	char	7	原始監理單位
TKT_NO	char	10	單號*
ID_NO	char	11	駕駛人證號
BIRTHDAY	char	7	駕駛人生日
RULE_1	char	9	條款一
RULE_2	char	9	條款二
RULE_3	char	9	條款三
RULE_4	char	9	條款四
ACCUSE_NO	char	5	告發單位
ACCUSE_TP	char	2	舉發類型
CLOSE_NO	char	2	結案代號
OP_DATE	char	7	出廠年月
BRAND_NAME	char	16	廠牌
MODEL_NO	char	5	標準車型
CC	char	6	排氣量
BODY_NO	char	21	車身號碼
ENGINE_NO	char	21	引擎號碼
MARK	char	2	記點
AREA_CODE	char	1	區域屬別
UNI_CODE	char	20	違規唯一碼

附錄 4、人車監理違規資料之專家訪談紀錄

1. 進行違規唯一碼的整併時，同樣一個違規唯一碼(單號+車號)，理論上有些資料應該要一樣，例如告發單位、舉發類型、條款以及違規日等資料，實際上為何會發生不一樣的狀況和問題呢？雖然此類資料筆數不多。

鍵入的問題，可能可以看它的違規日，如果早期的資料，如民國 90 年前，發生資料重複鍵入的狀況特別多，之後雖類似的情況減少了許多，而且某些代號，舉發單位可能有他們自定的格式，目前舉發單位的代號，通常是由告發單號來判斷，所以其實這樣的狀況可以忽略。實際上，從前面兩碼就可以看出是哪個單位(經由全國警察局開會決定)，但有可能是因為警察系統方面仍未轉換過來，有的還停留在以前純數字的表示方式。雖然希望可以做到全國唯一，但由於不同的警局裡可能包括 40-50 個單位，所以不能確定不會重複到。所以就算是不一致的告發單號，只要看到前面兩個數字是一樣的，其實就可以拿這個當作告發單號。

2. 例如有兩筆需整併的資料，想要取結案的那一筆資料作為代表，但可能這筆資料沒有另外一筆的齊全，應該取哪一筆？

扣掉先結，即違規案件未寫到系統，民眾先來結案；民眾只要先攔下來告發，就能馬上繳費，但如果是用電腦繳費，就要等警局先處理完(約 7 天)才能繳。如果先繳，則系統寫入的紀錄就會較少。比如在台北違規，但在高雄繳費，則會出現兩個結案，在高雄的資料為先結，又叫先收，在台北的資料為他收。如遇此狀況，優先以結案為主，並參照較完整(所填欄位較多者)的資料，但若出現欄位的資料都不一致時，建議都不要採用。取 data 中欄位資料一致的，其實也不必在意舉發單位的不同。

3. 如何選出一筆較完整的資料？

重要的資料是否都有留下，如違規日、車號、生日、條款、舉發類型、

車型、排氣量...等等。又比如說，車子被違規照相，但車主澄清駕駛者是他人，則需寫切結書，然後辦移轉，可能過程中會補條款，造成條款不一致，就比較麻煩。建議：條款很重要，但只要檢查前 2 碼就好，因為法律是條、項、款，只要檢查到條就好。

4. 資料如何紀錄？如發現不同區的資料有所出入時該如何抉擇？

同一個案件在不同區的檔案紀錄，由於認知上的不同或處理程序上的先後，可能會有所差別，又例如一案件，有一筆資料顯示其列管，另一筆資料顯示其免罰，則以免罰為主，因為可能是舉發有問題，已澄清過後的結果。而在同一區，銷案資料會覆蓋過去，並且紀錄異動時間，以及員工代號，可以有所依據，所以理論上應不會出現兩筆相同的資料。而列管的資料會一直保留，直到結案才會清除，但繳款資料不會清除。

5. 資料庫的形式是？

目前有七個資料庫，各自完全獨立，各管各的，可能由於疏忽，已結案的案件會留下未結案的紀錄，沒有作更改，所以大概幾個月就會到外區統整移出的結案；而有些案件在各區也可能較難分的一致，例如現在的罰單有些採分期付款制，但就有可能會記錄成未結案或結案兩種形式，這是較麻煩的地方。

附錄 5、期中審查意見處理情形

交通部運輸研究所合作研究計畫(具委託性質)

☒期中 ☐期末報告審查意見處理情形表

計畫名稱：道路交通事故相關資料整合系統雛形建置研究(II)—資料與系統擴充

(2/2)

執行單位：國立中央大學

參與審查人員 及其所提之意見	合作研究單位 處理情形	本所計畫承辦單位 審查意見
(一) 台北榮民總醫院醫務企管部李偉強副主任		
1. 健保局因財務問題，而車禍有代位求償的規定，故應對此資料頗感興趣；我們的健保資料中有些資料是美國所沒有的，故此系統所得的資料相當有價值。	敬悉。	悉。
2. 個人基於長久以來利用健保門診資料進行研究的經驗，除了本案目前利用 ICD 產生的 AIS 評估指標外，還有其他指標也是可以使用多重 ICD 編碼來產生，此方面若有需要，可提供協助。	謝謝。	悉。
3. FCI 的評估，主要應是美國的健康保險制度中，會依據被保險人的資源耗用狀況而將被保險人的危險度分級，故需評估創傷的後續使用資源情形。	敬悉。	悉。
4. 本案所建立的資料，應可進行二項課題的研究：台灣醫院創傷中心的資源分佈狀況，此對衛生政策的評估應有所幫助；以 FCI 分析長期醫療資源的使用狀況。	所提課題研究將會列入未來建議的應用課題。	同意。
5. 在網路上提供醫療資訊，為目前的趨勢。本案網路查詢是否有設定對象？是否需申請密碼？還是完全開放？	網路查詢是放在運研所的運輸安全專業網站中的技術專區，進入此專區者需先填寫資料，申請成為會員，故會有一些會員審查及管理的機制。	悉。
6. 在網路上可供查詢的資訊，若太過詳細，因本案的資料欄位很多，如：有年、月、日等，查詢者有可能可以自己透過許多變數的設定來進行串連，而車禍後續的處理又會涉及各種權利責任問題，此應謹慎。	謝謝提醒。本案提供網路查詢的資料，並不含個人隱私資料，且出生年月日也不會提供，此外，諸如年齡等的屬人的資料欄位均會再進行群組分類。	請依以往計畫中，在處理健保資料時的原則處理。
7. 台灣 ICD 碼的登錄情形，以糖	遵照辦理。	請加入建議事項中。

尿病為例，準確率為 75%；臨床醫護人員只知就醫者受到創傷，但並不知其是否為車禍所致，建議本案可提出於門診資料中加一欄標註車禍的建議，以訓練第一線醫護人員在建置門診資料時，即可鍵入。		
8. 建議資料可抽出一樣本(如：500份)進行 ICD 編碼正確性的驗證工作；因本案有警察資料，已可知其創傷來自車禍，故可再進行就醫者的醫院資料檢視，來瞭解 ICD 碼正確程度。	據了解，本案衛生署及健保局當初提供個人資料時即強調，不可依據個人資料進行追蹤研究。未來若運研所要執行此種驗證工作，需要本團隊參與，自當樂意參與研究。另外，有關驗證工作，本案 92 年計畫中，已針對連結前後的資料型態，進行過比較分析，結果顯示差異不大。	同意。
9. 目前因就醫資料需寫入晶片中，故需每天上傳醫院的處置資料，且榮總是由醫師直接鍵入 ICD；一般認為，大醫院的 ICD 建置品質應會比診所好。	敬悉。	悉。
10. 台灣的健保資料最多可鍵入 3 組 ICD，而美國可鍵入 8-9 組，因此，我們的資料在此部份應用上有其限制，此應在報告中說明。惟健保資料才建立 10 年，屬相當年輕的資料，大家都在學習，且在國內是很好的一個資料庫，相當具研究與應用價值。	遵照辦理，將會於報告中補充說明台灣資料特性所可能產生的限制。	此項 ICD 應用上的限制，請補充於未來報告中的結論。
(二) 交通部管理資訊中心紀美妃組長		
1. 應於報告中說明資料淨化的作法。	有關資料淨化部份，已於 93 年計畫中建立 SOP；期末報告中將補充相關重點說明。	悉。
2. 目前設計的網頁資料傳送方式，採實體隔離，為目前資通安全推動方向，是屬非常安全的設計。	敬悉。	悉。
3. 網路查詢的作業方式如何考量及設計，應於報告中說明。	遵照辦理。期末報告中將會補充說明查詢用資料庫的產生與傳送設計，包括軟硬體部份。	同意。
(三) 國泰產險汽車保險部孫騰敏經理		
1. 依照規定，新車領牌時需先投保強制險，此時因尚未有車牌號	謝謝提供實務經驗，但本案所取得的資料是出險理賠資料，	相關說明請補充於期末報告中。

碼，故投保資料中無法登記，雖依規定保險公司需在事後回補資料記錄，但可能會有遺忘回補現象，若本案資料來自投保資料，則可能是因未回補車牌資料而使得事故跟強制險的連結率不到 20%。	並非投保資料，且目前所呈現的連結率是以駕駛者身份證號串連的結果。不過委員所提此種現象有可能發生在新車尚未領牌，以臨時牌在路上行駛便發生車禍，而警察登錄的是臨時牌照，但強制險可能尚未有牌照號碼，故當以車牌號碼連結時，此種案例便會連不到。	
2. 如果警察以備案方式處理車禍，則受害者也可到保險公司申請理賠，保險公司也會受理；當然也有可能是乘客受傷的問題。	謝謝提供實務經驗；將會補充說明於期末報告中。	請研究團對再加強對保險理賠實務面的瞭解，以釐清資料目前產生的疑義。
3. 健保局為了財務問題，今年曾把警政署的車禍受傷者的就醫資料，提供給保險公司，讓保險公司查是否有理賠，但因仍有一些疑義需釐清，故目前仍未溝通如何處理；因保險公司提供理賠的資料給健保局，當然也會有個資法需考慮。	敬悉。	悉。
(四) 金融監督管理委員會保險局陳定輝專員		
1. 強制險與健保門診資料的相符程度應較高。	敬悉。	悉。
2. 理賠記錄中，依照規定需記錄所有駕駛者、乘客的身份證字號；依據保發中心統計，每年強制險理賠約 10-12 萬件，而每年約有 6 萬件健保局成功完成代位求償作業；金管局與健保局雖在此過程中雙方會互相核對，但仍會有些黑數，未來需更精確反應。	敬悉，不過目前本團隊所使用的出險理賠資料中，只有被保險人及肇事駕駛者二個身份證號；若有其他乘客的身份證號則會對連結率的提高，有明顯的幫助。	本所將會澄清是否有其他乘客之身份證號可供使用，若能取得，且取得後之資料格式可能不同時，請研究團隊能配合於明年維護合約結束前，仍能協助解決強制險與事故資料連結率低的問題。
3. 從今年 6 月 22 日開始，已有會議決議強制險的理賠，需回歸法規中規定的互相理賠機制，若此規定爾後正式實施後，保險公司便須知道所理賠的所有對方受害者的資料，屆時，資料的準確度與完整性便會提高。	敬悉。	意見同上。
4. 新車掛牌與車輛異動時的強制險資料回補問題，目前保險局正與數據分公司在商談如改善回	敬悉。	悉。

補率。		
5. 報告中提到，平台有二種管制層級，一為提供給其他政府單位使用，此可視為公益用途，另一則為供進階研究使用，此即市場化問題；本案是否會提出後續資料市場化的相關作法？	由國外網頁查詢系統的搜尋過程可知，收取資料查詢費用是一種可行且常見的市場化作法，本團隊將會在期末報告中補充此項建議。	同意。
6. 本案應提出如何訓練第一線人員正確建置資料的建議；因好的資料是一切分析與應用的基礎。	由於本案是在技術可處理範圍內，提高整合性資料的可使用性，有關資料建置的品質改善問題，雖不在計畫範圍內，但將會於期末報告中說明資料特性所可能產生的限制。	同意。
(五) 中華電信數據通信分公司資訊處丁思瑩工程師		
1. 本案中有 6 個不同來源的資料庫，未來隨著時間經過、法令變遷等，如何進行資料的更新機制？	若資料欄位無增減，只是定義或值域改變，依據國外文獻顯示，可考慮用版本別方式區別資料記錄，再透過說明文件提供必要的補充說明。若欄位改變，則須要作一些轉換處理或資料庫結構的調整。	請於期末報告中補充說明此部份的建議。
2. 本案的系統，未來要新增資料庫時，其限制為何？	目前的設計並不會有太大限制。	請於期末報告中補充說明此部份的建議。
3. 本案之資料來源多，資料敏感度高，未來資料的提供對象是否有限制？	有關網路查詢部份，是放在運研所的運輸安全專業網站中的技術專區，進入此專區者需先填寫資料，申請成為會員，故會有一些會員審查及管理的機制。另外依據 93 年經驗，比較詳細的資料是透過專案向運研所申請。	同意。
(六) 林大煜教授		
1. 第 2.2.1 節的文獻回顧中，提出 FCI 指標，此如何應用於本案 (p5-2)？	目前是規劃利用此指標來作為長期結果的一種定義參考，不過此需俟正式取得相關對照表或軟體後，透過初步的分析及探討，才能有較具體的建議。	同意。由於目前正透過 APEC 官方管道取得此資料中，有可能取得時已近期末，請針對計畫執行可能需要的調整與因應，預為考量。
2. 網頁資料查詢系統 (介紹國內、國外各機關許多統計網頁資料)，本案例中如何達到由使用者自行設定查詢變數之線上統計	本案的網頁查詢系統，將會參考 Utah CODES 的設計，讓查詢者可自行在所提供的變數中，挑選變數進行統計。	同意。

功能？可達到何種地步？		
3. 雛形系統與軟體平台建置		
(1) 「事故關聯資料分析與評估系統」軟體之改善部份，p.3-4 及 p.3-7 列出許多的改善方法，如：軟體系統功能的改善、資料編碼系統的改善等，是否已實質於各機關資料的輸入格式中加以改善。	此部份是屬以改善完成部份，目前正由運研所承辦人員測試與檢查中；後續將會配合運研所意見，進行必要的修改與調整。	同意。
(2) 人車監理（人車基本資料）、違規（人違規及車違規）資料庫，是屬較封閉的資料庫，目前在資料整理、淨化、整併過程分析的很清楚，但仍有許多需要協調、討論、處理之事宜，是否均已切實作到？	有關需要討論與處理的問題，已於 97.7.25 期中審查會後，就教於中華電信數據分公司之丁工程師，並與運研所承辦人員討論處理方式，將會於期末報告前完成處理。	同意。
4. 系統連結結果部份，事故資料串連強制險資料後只有 20%連結成功率，此比率太低，目前經改善後仍提升有限，後續之繼續探討（如 p.4-14），預期可提升多少百分比？	若以目前所取得的資料初步來看，能改善的幅度可能不大，但若可取得陳委員所提到的乘客身份證號資料，則應可較大的連結率改善空間。惟仍將繼續完成相關探討，以釐清目前系統中資料所能達到的連結程度。	請就本計畫目前已取的資料進行分析，以釐清連結的極限；至於陳委員所提資料的處理，請詳本所針對陳委員的第 2. 點審查意見。
5. 二篇論文是否已有所準備？	已在準備中，謝謝委員提醒。	悉。
（七）財團法人保險事業發展中心		
1. 車主與駕駛者不同的投保案例，當發生交通事故時，本案是如何連結？	本案是利用肇事駕駛者的身份證號進行連結。	悉。
2. 事故資料與強制險資料連結時，會出現二個落差，一為孫經理所提的新車領牌問題，一為車主與駕駛者不同的投保案例問題。	(1) 本案所取得的資料是出險理賠資料，並非投保資料，且目前所呈現的連結率是以駕駛者身份證號串連的結果。不過委員所提第一種落差有可能發生在新車尚未領牌，以臨時牌在路上行駛便發生車禍，而警察登錄的是臨時牌照，但強制險可能尚未有牌照號碼，故當以車牌號碼連結時，此種案例便會連不到。(2) 本案是利用肇事駕駛者的身份證號進行連結。	(1) 請進行保險理賠實務面的瞭解時，確認此項認知。(2) 悉。

3. 會後將提供本中心的資訊人員聯絡方式，以供澄清出險資料中是否有除了車主、肇事駕駛者以外的其他人的身份證號可供使用，以協助本案改善事故與強制險資料連結率低的問題。	謝謝，本團隊將會配合運研所進行相關作業。	本所將會進行相關接洽工作。
(八) 交通部道路交通安全督導委員會		
1. 請於第一章中，清楚說明本案的研究目的。	遵照辦理。	悉。
2. 第二章的文獻回顧請加強表達方式，以協助閱讀者參閱文獻。	遵照辦理。	悉。
3. 請增加有關各個資料庫的建檔方式，以及資料結構、格式、定義等基本描述。	遵照辦理。	悉。
4. 未來如何處理資料中的個人隱私問題，應於報告中清楚說明。	遵照辦理。	悉。
(九) 交通部運輸研究所陳一昌組長		
1. 簡報時所提的違規唯一碼重複的用詞，應表達清楚，以免造成誤解。	遵照辦理。	悉。
2. 本案資料中，身份證號重複的比例高嗎？	93 年計畫中的檢查結果顯示，以死因資料（其他資料不容易分清楚是重號，還是同一個人的不同事件）來看，比例很低。	悉。
3. 在資料淨化過程中，為何會刪除身份證號此項本系統的重要資料。	93 年所訂之針對身份證號的淨化，是處理空白、不符合身份證號編碼原則的資料，此種資料會產生許多無效的連結作業及結果，故需在連結前先處理。	悉。
4. 如何利用本系統的資料進行後續應用，請提出具體建議。	遵照辦理。	請彙整更新 92-93 二年所提出的相關建議內容，加入未來報告的建議中。
5. 與會委員對於隱私問題的疑義，應於報告中再說明清楚，以釐清疑義。	遵照辦理。	悉。
6. 於健保門診資料中建議增加標註車禍的代碼，此是否可行，請研究團隊考量。	遵照辦理。	請加入建議事項中。
(十) 運安組意見		
1. 網頁查詢系統是今年的重點之一，且其查詢結果的正確性與相關說明資料亦相當重要。請先以較小時間範圍的資料為測試基	遵照辦理。	悉。

礎，加快網頁查詢的設計與測試進度，基於計畫時程的考量，請儘量於8月底前完成初步測試。		
2. (1)系統中違規資料部份的問題，處理進展較為緩慢；(2)目前系統測試結果，在第一層及第二層的正确性均佳，但第三層的某些輸出流程仍須調整。有關此二部份的問題，基於計畫時程的考量，請儘量於9月中旬前解決，並與本所完成測試。	遵照辦理。	悉。
3. 在明年維護期限結束前，仍請合作單位能配合本所今年度資料更新時間，協助完成資料的更新與整理。	遵照辦理。	悉。
4. 雖然目前系統的資料架構已定，但期中審查會中獲知強制險資料中可能有乘客的身份證號資料，故未來若本所能順利取得此項資料，且會有助於提升事故與強制險資料的連結率時，在明年維護期限結束前，仍請中央大學能協助改善本系統此部份連結率低的問題。	遵照辦理。但建請考量若修改幅度較大時，可能需投入較一般維護作業為多的資源。	悉；惟請研究團隊針對此部份能配合的程度，預作考量並與本所研商，俾利 貴我雙方互為因應。
5. 利用本案部份成果發表的論文，請於期末報告書中，列示相關發表資訊；並請於所發表文章中刊登致謝說明。	遵照辦理。	悉。
(十一) 主席結論：		
1. 本期中審查會原則上認為，研究團隊對本案投入心力多，且研究有具體成果。	謝謝。	
2. 委員的疑慮、所提問題部份，請再依據錄音進行整理，重新列表對照，於工作會議中逐項討論，何項可達到、何項無法達到。	遵照辦理。	悉。
3. 委員所提的具體建設性建議，請研究團隊參考。	遵照辦理。	悉。
4. 本案之期中報告為可接受，請研究團隊與本所辦理後續行政事宜。	遵照辦理。	悉。

附錄 6、期末審查意見處理情形

交通部運輸研究所合作研究計畫(具委託性質)

☐期中 ☒期末報告審查意見處理情形表

計畫名稱：道路交通事故相關資料整合系統雛形建置研究(II)－資料與系統擴充(2/2)

執行單位：國立中央大學

參與審查人員 及其所提之意見	合作研究單位 處理情形	本所計畫承辦單位 審查意見
一、成功大學醫學院呂宗學副教授		
1. 報告書前後配合度請一致，譬如：1.2 節的研究範圍與對象與 1.3 節研究內容與工作項目，以及與各章節內容之配合度問題，在寫作技巧上應改進。	1. 遵照辦理。將會依委員意見修改報告。	悉。
2. 網上查詢之主要使用者如果是運安組人員，文獻回顧就應該針對專業人士，而非大眾讀者。應補充說明如何由國外供一般使用者使用之網頁查詢文獻回顧中，綜整其中之優缺點，設計國內供專業人士使用的網頁查詢。	2. 遵照辦理。有關如何由國外網頁資料，擷取其優缺點來設計本計畫之網頁部份，將補充於定稿報告中。此外，本網頁供專業人士用，未來有關 on line help 的設計也以專業人士為目標。	同意。
3. 建議應積極推動研究者使用資料庫。	3. 此項將會列入定稿報告書的建議事項中。	同意。
4. 建議應鼓勵推廣民眾可獲取之資料查詢。	4. 此項將會列入定稿報告書的建議事項中。	同意。
二、國泰產險孫騰敏經理		
1. 強制險連結率低的現象，應該是保險公司的問題；雖為已決賠案，因保險公司實務面問題，可能車號仍未回補。	1. 敬悉。	悉。
2. 建議可考慮再比對健康保險局向保險公司申請代位求償的資料。	2. 此項將列入定稿報告書的建議事項中。此外，今年曾協助健保局進行代位求償的探討，其最後結果後續可進一步洽詢，看看是否有相關資訊予以比較。	相關回覆處理請列入定稿報告的建議事項中。
三、中華電信數據分公司丁思萱工程師		
1. 新車領牌時，因依監理規定新車領牌前需先投保，故投保時(在保險公司的資料中)無車號，但監理單位有車號，可申請新領牌的車牌資料。	1. 敬悉。若此項資料勾稽無須額外付費，本團隊願意進行可行性測試。	悉。本所將會繼續處理相關資料取得事宜。

2. 目前詐騙集團追蹤串連個人資料的技術相當好，本研究的網路查詢系統，應注意有關使用者的管理問題。	2. 敬悉。	請將委員的意見列入定稿報告的結論中。
3. 今年監理資料在車種分類上，新增大型重型機車（需配合 C.C. 數資料共同判別）及租賃車，此請納入考慮。	3. 敬悉。此部份將會納入處理。	請將此項資訊補充於定稿報告書的相關章節中。
4. 所建立的系統將如何測試，包括：程式正確性、bug 以及資料驗證等。	4. 所設計之資料擷取流程，即可驗證資料正確性，測試部份則配合運研所進行。	同意。
5. 建議後續應考慮系統維運問題。	5. 此將會列入定稿報告的建議事項中。	同意。
四、金融監督管理委員會保險局陳定輝專員		
1. 本研究確具前瞻性及實用性。	1. 敬悉。	悉。
2. 新車領牌應不是強制險連結率低的主因，因本研究所使用的是強制險理賠資料，而民眾申請理賠時一定會填車牌，新車約有 4 成未回補車牌是事實；主要應該是管理上的問題，但今年 6 月 22 日時，已立法公布強制險應回歸互相理賠的機制，因此，未來資料應該會越來越好。	2. 敬悉。此將會列入定稿報告書中，有關強制險連結結果的討論內容。	同意。
3. 關於資料正確性之加強，請持續研商。	3. 敬悉。	悉。本所將會繼續處理相關資料取得事宜。
4. 一定要納入市場性之考慮，俾社會大眾得以利用。	4. 建議可透過辦理專案主題（如：代位求償等）的研討會，邀請各相關單位參加，進行促銷，但此需透過政府機制，營建繼續推動的環境。此將會納入定稿報告書的建議事項中。	同意。
5. 本研究請在財源無虞前提下，繼續持續推動。為了系統持續運作及市場化，應找到財源來支援計畫進行所需的經費，可考慮的財源為中央相關主管機關（如：交通部及金管會需資料進行強制險從人因素探討）、縣市政府（因各地方政府均需資料改善其肇事地點、設置 EMS 站、執法取締人力配置等）、公會及保險公	5. 敬悉。委員相關意見將會納入定稿報告書的建議事項中。	同意。

司（許多保險商品需損失資料）、汽車公司（可評估車輛受損部位等）、健保局（每季會代位求償，但此可透過一比例參數的研究統一納入保費中）。		
五、交通部運輸研究所運安組張開國副組長		
1. 本計畫以本所而言，是一項基底的研究計畫，其主要是在支援其他研究計畫，跟先進安全車計畫是帶頭計畫的性質不同。	1. 敬悉。	悉。
2. 在資料處理的功能選項，各項功能是否有執行上的順序要求，可加上編號等指引，讓使用介面更友善並確保資料處理結果與預期一致。	2. 遵照辦理。將會補充操作順序及提供操作手冊。	請配合修改程式。
3. 資料淨化時，被剔除的資料可否設計成儲存到檔案中，此可提供給對此方面資料的探討有興趣的研究者進行研究。	3. 遵照辦理。將會於後續維護工作中，調整此部份的設計，將會按照投影片第 19 頁的流程，逐步輸出被剔除的資料。	目前的系統已可輸出各種被剔除的資料，相關系統細節修飾同意可於維護期間因應使用需求逐步進行。
4. 車險連結率如何計算？目前強制險理賠資料量相當大，有許多並未向警察通報，若以事故資料庫為分母，是否可提高連結率。	4. 以事故資料為分母計算連結率所得比例仍不高（約 20%）。	悉。
六、財團法人保險事業發展中心韓令元小姐		
1. 身份證字號可能會填護照或居留證號碼。	1. 敬悉。惟理賠資料中，身份證字號錯誤或空白的比例並不高。	悉。
2. 保發中心有二個關於理賠的資料庫，一個是統計用，一個則是關貿資料。目前提供給本研究的是統計用的理賠資料，此資料並不對身份證字號及車牌號碼進行嚴格檢查。	2. 建議可由研究角度，再取得關貿資料庫的資料，測試一下連結率；惟此需先由運研所出面取得資料。	同意；本所將會繼續處理相關資料取得事宜。
七、交通部道安委員會劉昭正技正		
1. 資料正確性的檢查，除了筆數外，是否亦應針對實質內容部份。	1. 有關實質內容部份，是在淨化程序中進行。	同意。
2. 警察登錄之事故資料刪除比例明顯較其他資料庫為高，原因為何，可否補充說明。	2. 此部份資料剔除比例是逐年降低中，顯示原始資料的改善。本研究最重要的是原始資料源頭需檢核，本研究只能得知其錯誤率，至於原	同意。

	因，則僅能就所知儘量補充於定稿報告中。	
八、運安組書面意見		
1. 有關資料格式更新所造成的前後年度資料庫版本不同的情形，請說明使用者如何知道其所使用之資料格式版本。	1. 此係透過新增一資料定義版本的變數方式，予以辨別。將會加強說明於定稿報告中。	請配合修改程式。
2. 請於 1.1 節中，將本計畫 92-94 年三年間，發展系統的策略及各階段性成果與後續修改等歷程，作一說明，以讓閱讀者瞭解系統發展過程中的重要考量。	2. 遵照辦理。	悉。
3. 請於 4.3.4 節「事故關聯資料分析與評估系統」，補充有關違規資料部份資料連結流程圖說。	3. 遵照辦理。	悉。
4. 關於 5.1-5.2 節，請補充說明本研究依據本系統資料特性所進行之機率連結測試的設計架構，及依此架構彙整相關連結測試結果。	4. 遵照辦理。	悉。
5. 請修改初稿中關於各來源資料庫在錯誤檢核、連結處理等步驟的結果推論表達方式；至於其他有關初稿內容的遺漏字、錯字、表達正確性、一致性等，請另見所提供的書面資料。	5. 遵照辦理。	悉。
6. 請於系統驗收時，依規定提供相關系統設計的說明文件。	6. 遵照辦理。	悉。
九、主席結論		
1. 經過與會學者專家的討論，本計畫之期末報告初稿可予以接受。	1. 謝謝。	
2. 有關系統使用介面的親和性部份，請於期末報告定稿中改善。	2. 遵照辦理。	悉。
3. 與會學者專家的建議，請納入期末報告定稿的結論或建議中。	3. 遵照辦理。	悉。
4. 與會學者專家的意見請整理成清單，逐條回覆。	4. 遵照辦理。	悉。
5. 請研究團隊於 1 個月內提交期末報告定稿，後續辦理事項請與本組承辦同仁聯繫處理。	5. 遵照辦理。	悉。

附錄 7、工作會議紀錄

主題：94 年 2 月工作會議 MEMO	
計劃編號及名稱：MOTC-IOT-94SDB004 道路交通事故相關資料整合系統雛形建置研究(II)—資料與系統擴充(2/2)	
時間：94.2.18 pm 3:30~5:30	地點：運研所 7 樓會議室
主持人：陳一昌	
合作單位（中央大學）出席人員： 董基良、陳苑蕙、黃維信、許峻嘉、宋文旭	
運研所（運安組）出席人員： 賴靜慧	
主要結論： 1. 為使本（運安組）組同仁能瞭解及應用本計畫所整合的 yes-chain 資料，並能將本組同仁需求回饋本計劃，請中央大學提供一個下載資料的 friendly interface，俾利對本組同仁進行說明，同時請備有一份 user manual。有關此部份，中央大學將以 3 月底為努力目標。 2. 若本計劃後續將定期有資料匯入，則請中央大學於本計畫報告書中，提出編列年度例行維護經費的建議。 3. 中央大學將會朝機率連結方向努力，以提高連結的紀錄筆數。 4. 運研所將會按照以往取得各相關原始資料的作業方式，先行聯繫以獲得各相關資料，供中央大學進行處理。	

主題：94 年 3 月工作會議 MEMO	
計劃編號及名稱：MOTC-IOT-94SDB004 道路交通事故相關資料整合系統雛形建置研究(II)—資料與系統擴充(2/2)	
時間：94.3.28 pm 2:00~3:30	地點：運研所 7 樓會客室
主持人：陳一昌	
合作單位（中央大學）出席人員： 董基良、黃維信	
運研所（運安組）出席人員： 賴靜慧	
主要結論： 1. 期中及期末審查會議時，可邀請衛生署、健保局、警政署等資料提供單位，以及國健局等資料可能應用單位參加，聽取各方意見及擴大資料影響層面。 2. 合作單位在資料探勘的部份已投入不少研究人力，請考慮於下次工作會議中，就此項技術及其應用提出專案報告，屆時可請本組其他同仁一起參與工作會議，瞭解此資料處理技術。 3. 有關投稿論文期刊部份，建議本計劃可朝 SCI、SSCI 類期刊論文的 submit 為努力方向。	

主題：94 年 4 月工作會議 MEMO	
計劃編號及名稱：MOTC-IOT-94SDB004 道路交通事故相關資料整合系統雛形建置研究(II)—資料與系統擴充(2/2)	
時間：94.5.2 am 9:45~12:00	地點：運研所 7 樓會議室
主持人：陳一昌	
合作單位（中央大學）出席人員： 蔡興國、黃維信、許峻嘉	
運研所（運安組）出席人員： 賴靜慧、洪憲忠、吳熙仁、葉祖宏、張仲杰	
專案報告：資料探勘的簡介 結論： 1. 有關警大的交通學報邀稿（本計劃 93 年投稿於運輸安全與執法研討會的論文--「肇事駕駛者交通事故與違規記錄關聯性之探討」）乙事，請合作單位撥冗增修原研討會論文內容，協助轉投至交通學報上。 2. 94.5 月請合作單位開始進行有關網頁設計部份的工作。 3. 請合作單位繼續進行有關機率連結的工作。 4. 94.5 月請盡可能完成有關違規資料的釐清工作。 5. 94.5 月本所將會開始進行新年度資料的取得。	

940527 工作會議備忘錄

會議時間: 94.5.27 PM 2:20~6:10

會議地點: 運研所七樓會議室

參與人員: 運研所 張開國副組長 賴靜慧

中央大學 黃維信

會議流程:

一、研究單位工作進度說明:

工作項目	執行情行
1. 健康相關生活品質量表收集	1. 收集 6 種與健康相關生活品質之量表 2. 量表需採受測者自填或是由調查員進行訪談，需長期的調查，不適合本研究使用。 3. 收集由 NHTSA 委託 Johns Hopkins 所進行的 FCI 之應用研究。 4. 與 FCI 原始研究單位，以取得更進一步之研究資訊。
2. 機率連結文獻研讀	1. ORLS 歸納的資料錯誤類型。 2. 連結前的資料處理方式。 3. 利用語音進行資料分類處理方式。
3. 國內外線上查詢系統資料收集	1. 國內網路線上統計資料下載，多以年報、季報等固定格式提供下載，較缺少可線上即時互動之下載界面。 2. 國外除制式表格下載外，亦有提供自訂表格之下載服務，大部份為免費，但部份資料下載仍需收費。 3. 佛羅里達州的 FMCSA 系統，將下載資料先進行分類，再由使用者自行下載統計表格。 4. 威斯康辛州的 CODES 系統所提供的 Cube Analyzer 除固定格式之下載外，亦提供自訂分析欄位之下載。
4. 類神經網路測試	1. 使用類神經網路裡的倒傳遞類神經網路做為資料分析處理之方式。 2. 使用 SPSS Clementine 做為資料分析工具。

	3.使用發生時間、天候、年齡等做為輸入變數，探討何變數對於平均醫療費用高低有較顯著之影響。
5.人車監理違規資料整理	1.中華電信數據分公司所提供之資料，部份欄位內容與原始定義不同，會造成字串拆解之問題。 2.解決字串拆解及資料匯入資料庫之問題。 3.澄清資料整併前資料變數使用原則。

二、討論項目與結論：

1. 對於所提的健康相關生活品質量表，須再確認於本研究中之應用範圍。
2. 建議：可參考 WHO Burden 研究所提之 DALYS，是否能應用在本研究中。
3. 對於本研究所欲進行的機率連結，目前所研讀之資料已初步能了解國外機率連結的大致內容，唯國外研究所常使用的姓名、生日、地址等，並不完全適合目前所使用的資料庫。

建議：

- (1) 國外研究論文之研讀可暫告一段落，應回歸服務建議書所提及欲使用機率連結解決之問題。
- (2) 為改善事故資料庫與強制險資料庫連結成功率過低之問題，可依下列三個步驟進行。
 - (i) 先針對資料淨化過程中所刪除之資料進行回補處理，可參考使用前九碼進行處理。
 - (ii) 將資料淨化後未能與事故資料連結的強制險資料，先與健保資料進行連結。並以出險日做劃分，以確定強制險與健保資料連結成功率。
 - (iii) 整理 PO-VI 流程圖中各步驟之資料處理筆數，以澄清不同連結順序所可能資料連結情形。
4. 已初步了解國內外線上查詢系統所提供之功能，應可規劃本研究所欲採用的線上查詢系統雛型。

建議：

- (1) 可參考威斯康辛州 CODES 系統所使用的 Cube Analyzer，但需先定

義本研究線上查詢所欲提供變數的 Haddon Matrix。

(2) 須再與甲方討論欲提供線上查詢的資料內容。

5. 資料探勘已嘗試多種分析模式。

建議：須先澄清資料探勘在長期照護及機率連結上，所能提供之研究功用。

6. 人車監理資料庫資料內容除與中華數據分公司人員討論外，亦可針對資料內容做更進一步之分析。

建議：

(1) 先整理利用違規唯一碼做區別時，同一筆違規資料筆數之分佈，以做為資料整併之優先順序，估計一筆與兩筆的資料應會佔多數。

(2) 當以兩筆資料做整併處理時，先做各欄位中有資料相異的筆數，先判斷那些欄位較容易出現資料相異之情形，是否與研究欲使用之欄位有關。

(3) 將資料做縱向二分法處理，以做為後續資料處理之優先順序判斷。

940905 工作會議備忘錄

會議時間: 94.9.5 PM 2:30~3:30

會議地點: 運研所七樓會議室

參與人員: 運研所 陳一昌組長、賴靜慧

中央大學 鄭銘章教授、黃維信

會議流程:

一、研究單位工作進度說明:(另見簡報內容)

二、討論結論:

1. 今年後續工作重心:

(1) linkage system 的 level 3 輸出結果正確性確認。

(2) 違規資料的整併處理及勾稽的前置作業。有關勾稽作業，將俟強制險及任意險資料取得後，便會送請中華電信勾稽，依據以往經驗，因需處理之資料龐大，且今年需將資料分成新年度 ID、舊年度 ID、新年度車牌、舊年度車牌 4 個檔，故請中央大學預為處理手邊現有資料。

(3) 完成機率連結的研究。

(4) 完成網頁查詢系統。目前的查詢已具雛形，有關界面及資料呈現方式等細節，請於近期內再與運安組承辦人討論。

2. 運安組將會再確認中央大學所提設備費用的支用狀況。

3. 明年維護期間，中央大學將會配合事項:

(1) 強制險與任意險資料格式改變，需因應調整的相關資料處理程式與結構。

(2) 新年度違規資料取得後，與就年度資料的分離及更新工作。

(3) 因應 FCI 取得狀況的後續發展。

940927 工作會議備忘錄

會議時間: 94.9.27 AM 9:30~10:30

會議地點: 運研所七樓會議室

參與人員: 運研所 陳一昌組長、賴靜慧

中央大學 鄭銘章教授、黃維信

會議流程:

一、研究單位工作進度說明:(另見簡報內容)

二、討論結論:

1. 運安組承辦人將會進行下列 9 月份成果的檢核與測試:

(1) 整合系統輸出結果的正確性。

(2) 網頁查詢結果的正確性及查詢效率。查詢效率測試包括網路離線、連線狀態。

2. 有關設備採購內容及規格部份,本所將會於 94.9.30 前與中央大學確認。

3. 有關網頁查詢將會改善查詢進行時的狀況訊息提供。

4. 有關整合系統將會改善:(1)層名的命名;(2)說明各種資料起迄日期。

5. 中華電信勾稽作業所需的新年度 ID、舊年度 ID、新年度車牌、舊年度車牌 4 個檔的資料來源均已到齊,請中央大學著手準備此 4 個勾稽檔,俟本所秘書室與中華電信簽訂採購合約後,便可提供中華電信進行勾稽作業。

6. 92-93 強制險及任意險資料已到,將交付中央大學進行處理。(因格式改變,依 8 月工作會議決議,可於明年完成匯入系統的工作)

7. 92-93 年強制險資料中有受害人資料,為應證期中審查委員認為,此資料可提高與警察登錄的事故資料間的連結率,請中央大學抽一樣本,針對此部份議題進行探討。

附錄 8、簡報資料

道路交通事故相關資料整合系統雛形 建置研究(II)－資料與系統擴充(2/2)

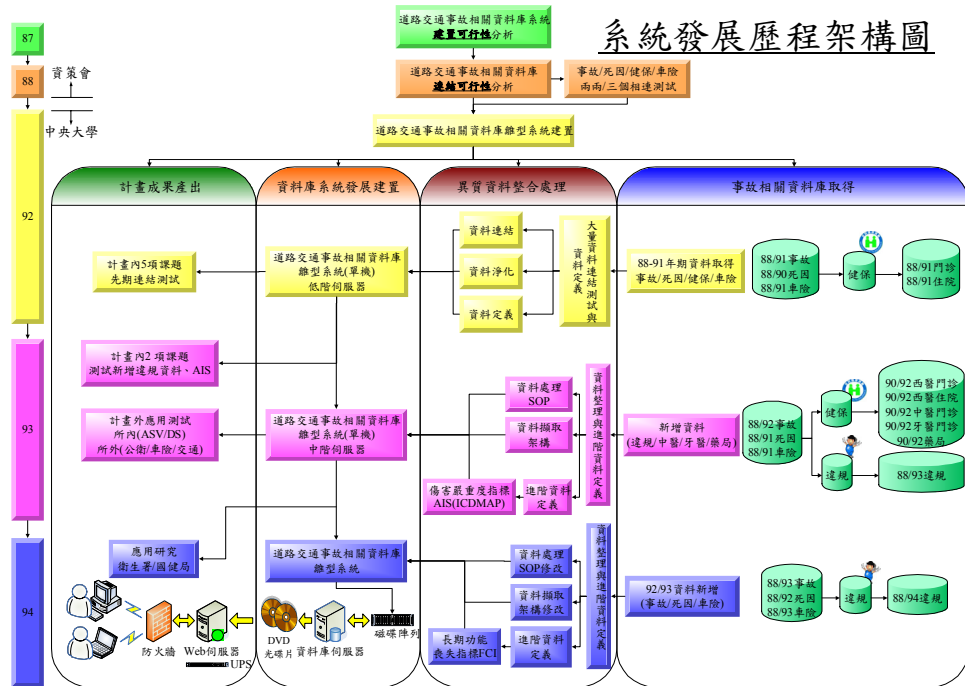
交通部運輸研究所

1

計畫背景

- ☐ 國內與道路交通事故相關之資料庫，由所屬單位維護，資訊系統不盡相同，且在資料取得上會有時間框的差異問題。
 - 道路交通事故資料庫(警政)
 - 健保資料庫(衛生)
 - 死因資料庫(衛生)
 - 汽車保險資料庫(財政)
 - 人車監理違規資料庫(交通)
 - ☐ 缺乏有效整合，無法作深入及跨領域的探討
-

2



87-88年系統發展中斷之原因

□ 軟體效能之影響

- SQL Server 7.0、Borland C++ Builder、NT 4.0、Windows 95

□ 硬體效能之影響

- 伺服器：Pentium II 500、30G硬碟、256MB記憶體
- 用戶端：Pentium 233、30G硬碟、128MB記憶體

□ 應用研究之影響

- 未能有實際具體之應用研究方向

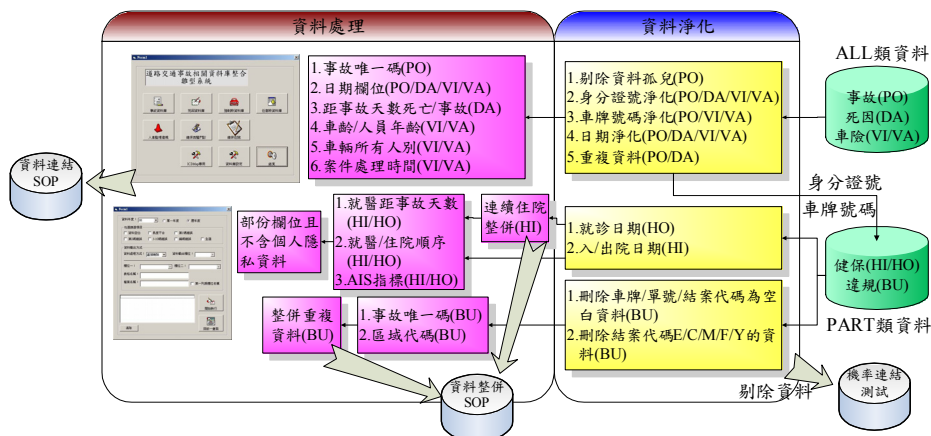
事故相關資料庫取得-資料庫內容

	主管機關	資料匯集單位	資料庫名稱	資料年份	取得方式	資料格式	檔案大小 (mb)	資料筆數	工作 年份		
ALL類	警政單位	內政部警政署交通組	道路交通事故資料	88-93	原始資料	Access mdb 檔	477	1,104,527	92-94		
	財政單位	財團法人保險事業發展中心	強制險已決賠資料	88-93	原始資料	純文字檔	164	727,583			
			強制險受害人資料	92-93			41	416,660			
			任意險已決賠資料	88-93			620	3,260,003			
PART類	衛生單位	行政院衛生署	死因資料	88-92	原始資料	純文字檔	48	632,953	92-93		
		中央健康保險局	健康保險資料	西醫門診	88-92	進駐健保局數據中心作業後取得。	Access mdb 檔	10,720		5,756,831	
				中醫門診	88-92			1,120		7,497,976	
				牙醫門診	88-92			1,155		5,906,312	
				住院門診	88-92			318		5,014,067	
				藥局	90-92			955		1,288,847	
				交通單位	中華電信數據分公司			人車監理違規資料		88-94(視各工作區而定)	提供資料由數據分公司代為連結。
		總計						18,762		47,084,096	

資料取得之困難：
 1.各單位對資料之保密
 2.資料取得之聯繫過程
 3.資料連結之處理(派員進駐)
 4.逐年經費之編列

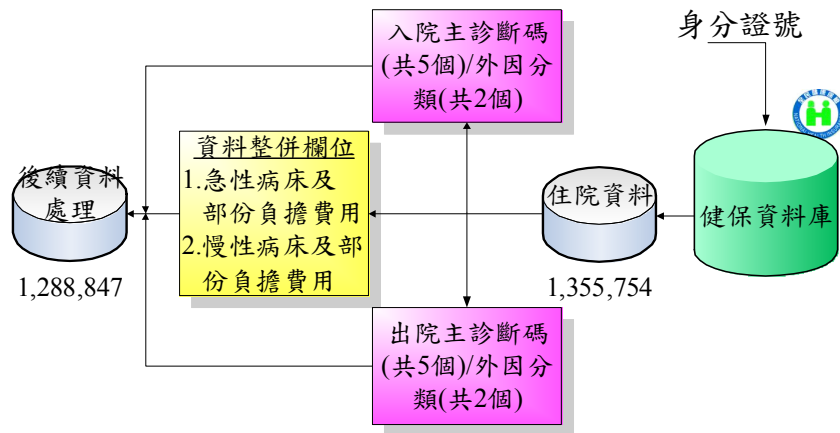
7

異質資料整合處理-資料處理SOP



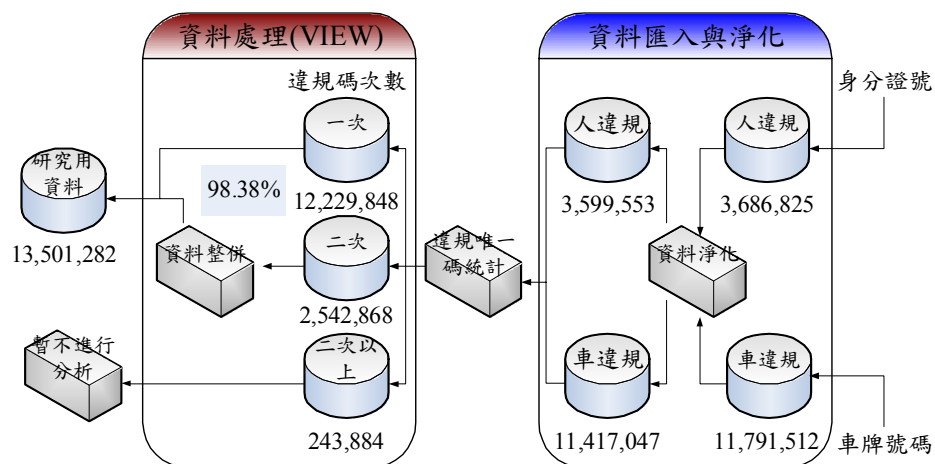
8

異質資料整合處理-住院資料整併SOP



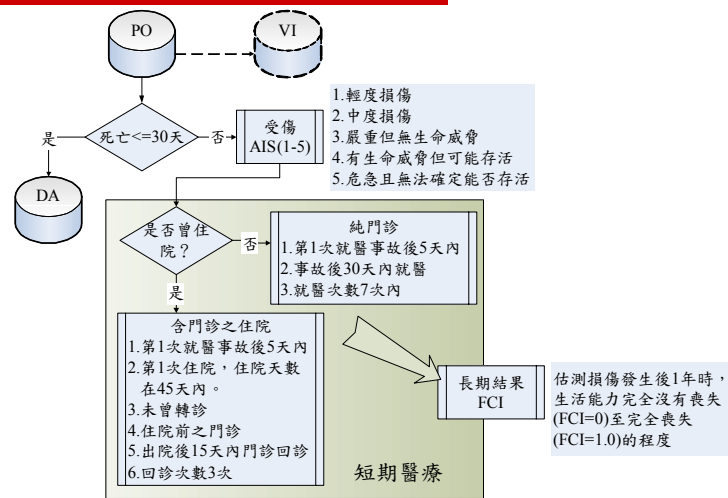
9

異質資料整合處理-人車監理違規資料處理SOP



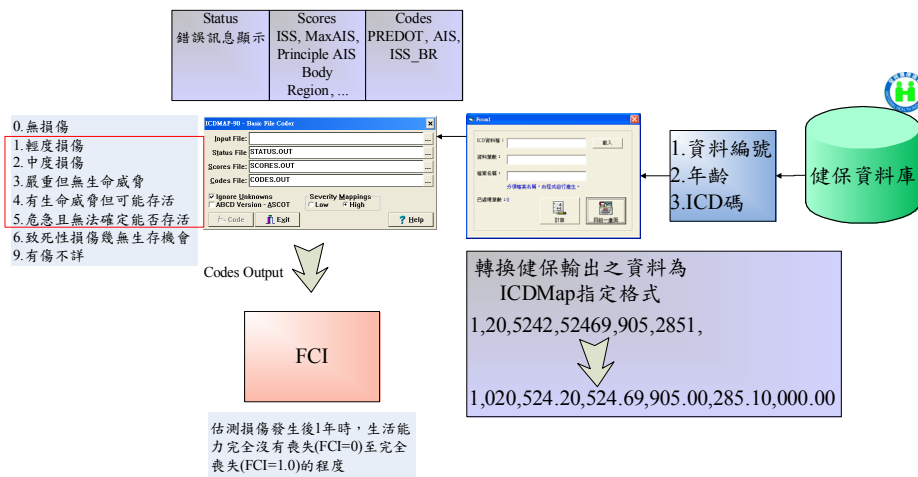
10

異質資料整合處理-資料連結定義SOP



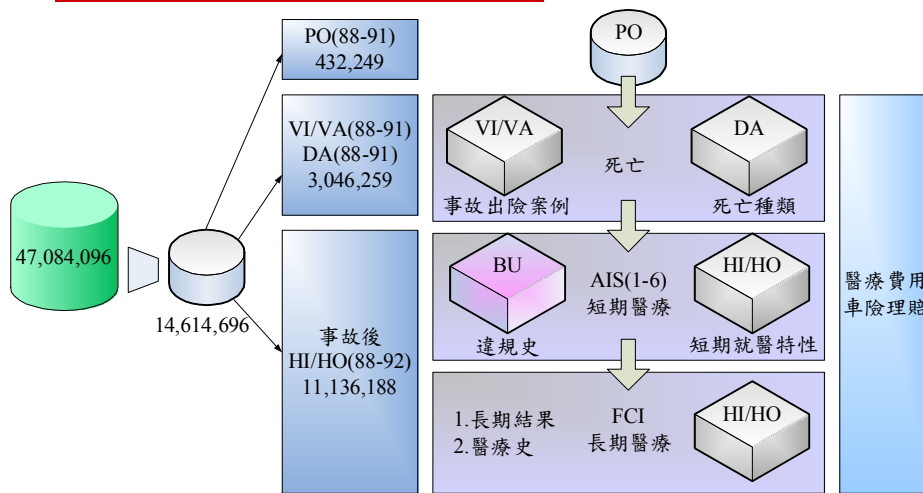
11

AIS與FCI資料之取得



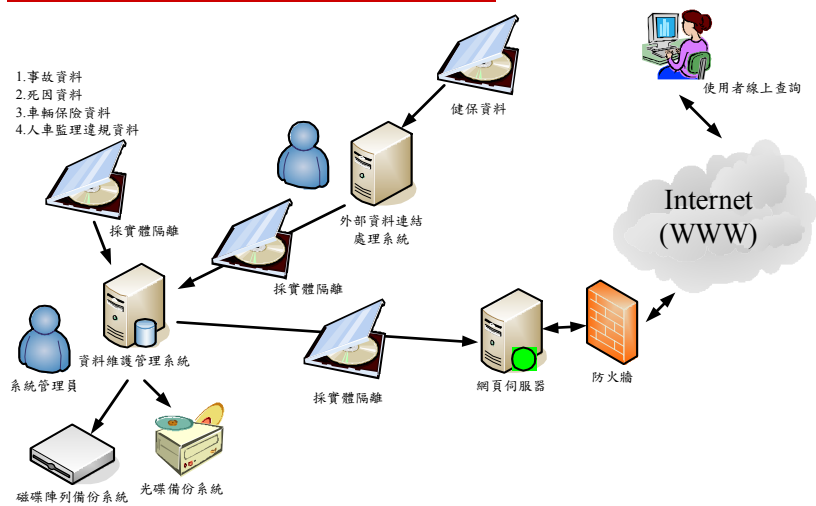
12

異質資料整合處理-資料連結應用關聯圖



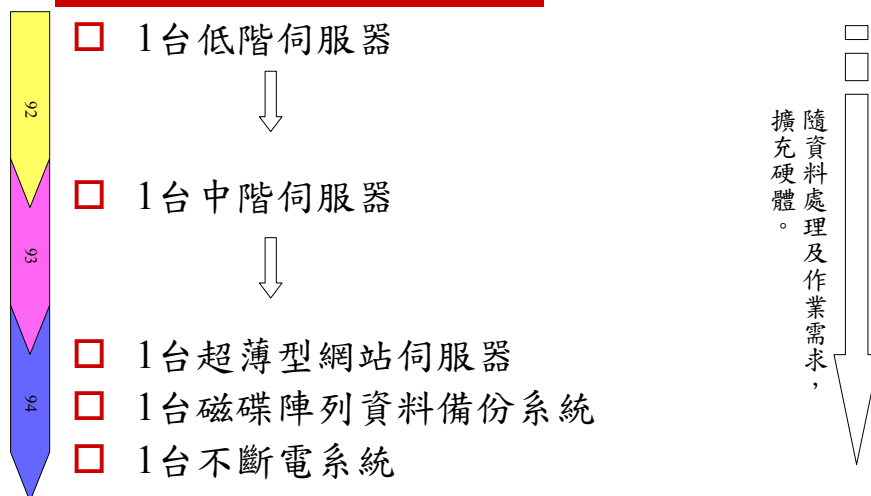
13

資料庫系統發展建置-資訊系統架構圖



14

資料庫系統發展建置-硬體發展歷程



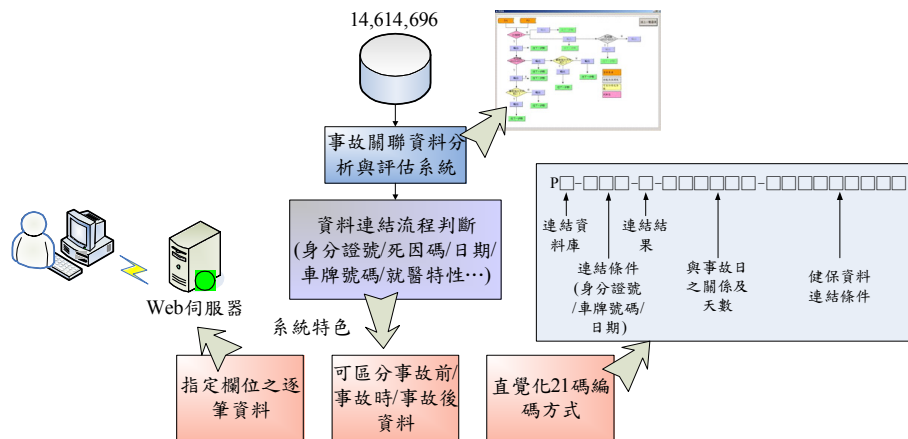
15

資料庫系統發展建置-軟體發展歷程



16

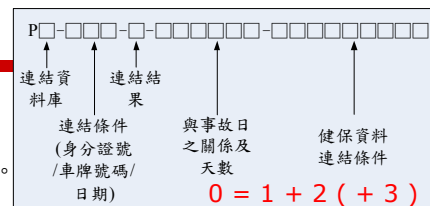
事故關聯資料分析與評估系統



17

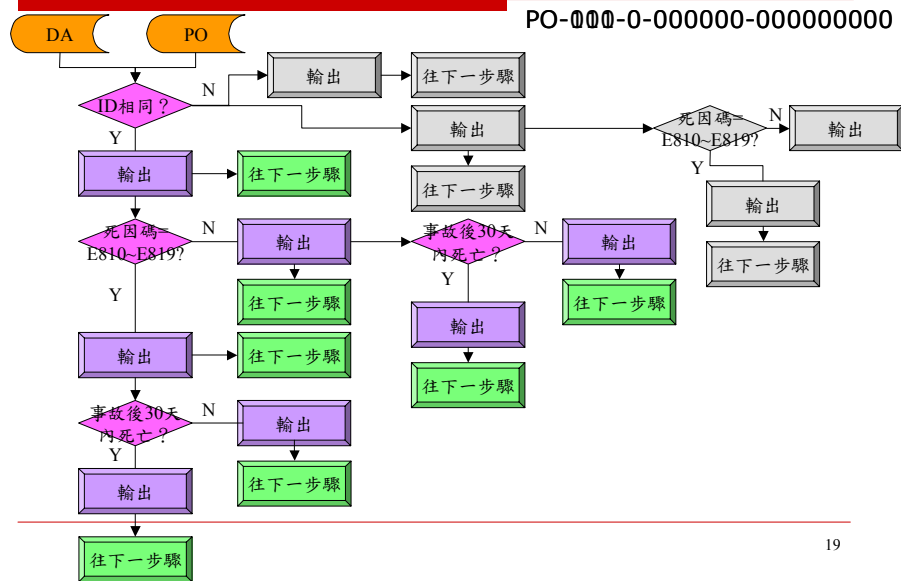
21碼編碼規則

- 連結條件代碼定義(3-5碼)
 - 0：不須考慮；1：相同；2：不同
- 連結成功為0，失敗用資料庫代碼。
- 與事故日之關係(第7碼)
 - 0：不考慮；1：事故前；2：事故當天；3：事故後
- 與事故相距天數(8-10碼)，000：當天；999：一年以上。
- 就醫史或違規史之連結(11-12碼)
- 健保就醫情形(13-21碼)
 - 有無轉院(第13碼)
 - 長期住院(第14碼)，0：不須考慮；1：長期住院；2：非長期住院
 - 住院天數(15-17碼)，000表不考慮是否為長期住院
 - 第1次就醫記錄(第18碼)：0：不考慮；1：5天內；2：非5天內
 - 就醫科別(第19碼)，0：不須考慮；1：10科內；2：10科外
 - 就醫時間框類別(20-21碼)



18

事故關聯資料分析與評估系統流程圖



19

關於機率連結

- ❑ 缺乏唯一識別碼欄位
 - SSN、姓名、性別、生日、郵遞區號、職業……等
- ❑ 從Newcombe提出記錄連結的觀念，到Fellegi及Sunter提出數學模式
 - CODES、CIREN、GAMLIS、GRLS、WADLS、VAES、ORLS等
 - Identity Systems、LinkageWiz、ChoiceMaker Inc.、GRLS、LinkSolv、LinkPro等。
 - 不適用雙位元語系、價格昂貴
- ❑ 應用範圍
 - 人口普查、農業調查、犯罪偵查、醫藥預防保健、疾病統計、商業活動調查等

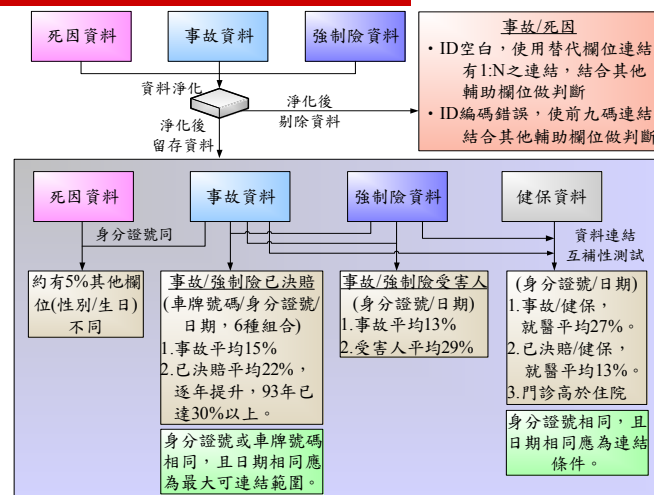
20

資料連結概念之運用

- 國外使用的連結欄位與國內之情形不甚相同，本研究以ID為主，結合機率連結概念。
 - 如ID九碼同，以其他欄位為輔(性別、生日、死因碼)
- 部份發展經驗可做為資料連結分析參考
- 不同資料庫間連結互補性測試
 - 事故、強制險、健保
- 資料探勘之應用測試

21

資料連結結果及分析



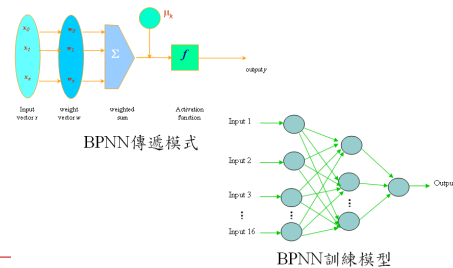
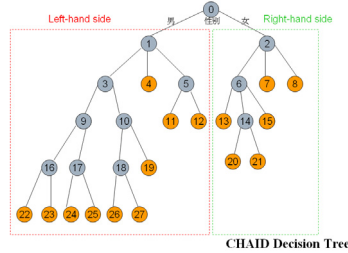
22

資料探勘之應用測試

- 利用分類、關連性、序列分析、群集分析、機器自我學習及其它統計方法，從大量的資料中萃取出隱含的、不為人所知、可信而有效的知識，以做為決策輔助或預測的參考。
- 國內外已有蠻多研究應用DM做為分析方法，但本研究著重於可行性之探討。

■ CHAID、CART 及倒傳遞類神經網路

- 機車交通事故之醫療費用預測模型研究



23

已發表之研究成果(含計畫要求)

- 2003年健康與管理學術研討會
 - 事故後住院、門診就醫時間框之探討
- 2004年道路安全與交通事故傷害防制研討會
 - 高齡者道路交通事故受傷嚴重性與醫療成本分析
 - 利用全民健保資料分析機動車交通事故傷害之傷害嚴重度—初步探討
 - 道路交通事故整合資料庫與其應用之介紹
- 93年道路交通安全與執法國際研討會
 - 駕駛者住院受傷嚴重性與住院成本之探討
 - 肇事駕駛者交通事故與違規記錄關連性之探討
- 2005國際醫學資訊研討會
 - 道路交通事故相關資料整合系統研究
- 94年道路交通安全與執法國際研討會
 - 機車交通事故之醫療費用預測模型研究

24

協助完成之研究成果_2004年(續)

- 公衛子題—交通事故之受傷嚴重度、醫療費用及住院天數間之關係
 - 結合交通警察的登錄資料，可以大幅提高事故傷害流行病學研究的廣度。
- 警察登記傷處與ISS部位分類一致性約為53.2%；與AIS部位之一致性為47.8%

NEW傷處 * RECODISS Crosstabulation

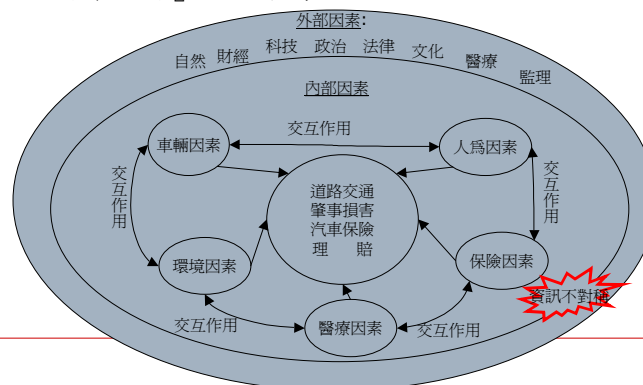
% of Total

		RECODISS					Total
		頭/頸/臉部	胸部	腹部	四肢	其他	
NEW傷處	頭/頸	20.9%	.5%	.5%	3.7%	4.5%	30.1%
	胸部	1.0%	1.1%	.3%	1.0%	.6%	3.8%
	腹部	.1%	.1%	.5%	.2%	.3%	1.1%
	手腕部/腿部	3.1%	.3%	.3%	26.1%	3.8%	33.5%
	其他	11.6%	1.2%	1.5%	12.5%	4.6%	31.4%
Total		36.6%	3.1%	3.0%	43.5%	13.7%	100.0%

27

協助完成之研究成果_2004年(續)

- 保險子題—強制汽車責任保險理賠與肇事風險因素分析-討論資訊不對稱的影響
 - 理賠『死亡給付』與『殘廢給付』因素屬性
 - 『醫療給付』的因素屬性



28

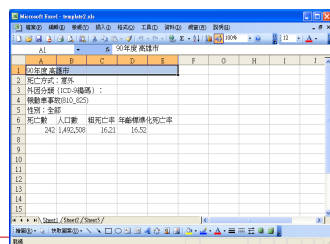
協助完成之研究成果_2004年(續)

- 交通子題—道路交通事故整合資料之研究—車種別肇因分析
 - 使用貝氏定理，檢驗雙車事故中之肇事原因，以判定不同運具之肇事原因差異。
- 高齡者交通事故之短期醫療支出研究
 - 依據TRL低報率修正係數修正後，初步估測高齡者因車禍產生的短期醫療總費用為6~10億。
 - **TRL建議亞洲地區：**
 - 死亡人數低報率25%~50% $\rightarrow 0.2\sim0.4\text{億} \times 1.25 = 0.3\sim0.5\text{億}$
 - 傷者數：死者數約20~30：1 $\rightarrow 0.3\sim0.5\text{億} \times 20 = \mathbf{6\sim10\text{億}}$

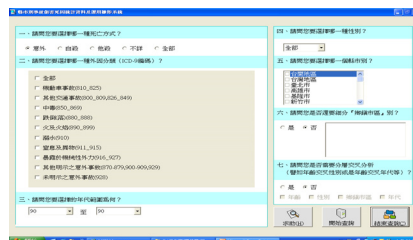
29

協助進行之研究計畫_2005年

- 國民健康局
 - 建立縣市別事故傷害死因統計資料及運用
 - 自行車事故傷害探討與防治策略
- 衛生署
 - 緊急醫療救護系統醫療不良事件之研究及病患安全體系建立



	A	B	C	D	E	F	G	H	I	J
1	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數
2	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數
3	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數
4	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數
5	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數
6	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數
7	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數
8	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數
9	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數
10	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數
11	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數
12	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數
13	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數
14	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數
15	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數	死亡人數



Microsoft Word - 2005年研究計畫

一、請問您想查詢哪一種死亡方式？

二、請問您想查詢哪一種死亡原因？

三、請問您想查詢哪一種死亡地點？

四、請問您想查詢哪一種性別？

五、請問您想查詢哪一種年齡？

六、請問您想查詢哪一種職業？

七、請問您想查詢哪一種教育程度？

八、請問您想查詢哪一種婚姻狀況？

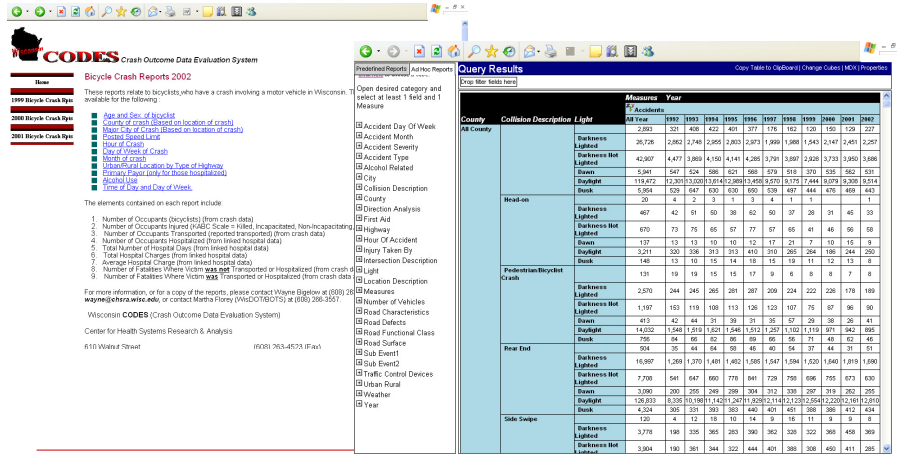
九、請問您想查詢哪一種宗教信仰？

十、請問您想查詢哪一種其他因素？

30

國內外線上查詢系統

Wisconsin CODES-Cube Analyzer



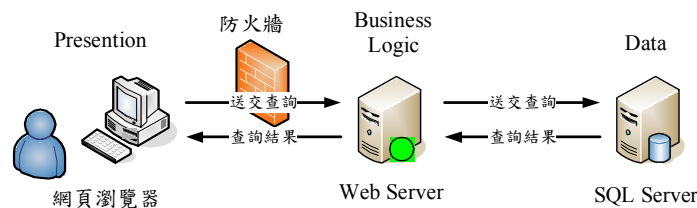
31

線上查詢系統建置規劃

採用三層式的架構做開發，並使用標準的 HTML，可提供跨平台的查詢使用。

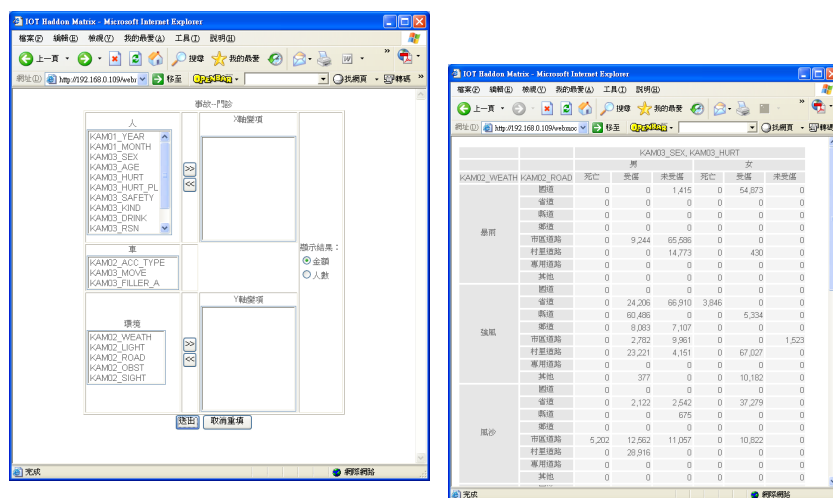
資料查詢類別

- 事故資料與死因資料；事故資料與門診資料；事故資料與住院資料。



32

線上查詢系統成果



33

結論

□ 系統的規劃與建置

- 已有初步之完整性，應可滿足未來短期內之需求。

□ 道路交通事故相關資料維護整合系統

- 已建立完整之資料維護步驟，並完成新舊資料表之整合。
- 具有擴充性，未來可依使用需求再做功能性擴充。

34

結論(續)

- 事故關聯資料分析與評估系統
 - 建立標準化之資料連結流程
 - 提供立即檢核功能，可確保連結資料筆數之正確性。
 - 應可滿足大部份之使用需求
- 線上查詢系統
 - 提供「事故與死因」、「事故與門診」及「事故與住院」三大類的資料查詢
 - 具備類似Excel樞紐分析之功能
 - 使用實際記錄說明取代資料代碼，增加可讀性

35

結論(續)

- 資料連結與機率連結
 - 已評估過國外機率連結之軟體，並不完全適用於本研究之資料，且目前使用確定性連結之資料筆數已相當多。
- 長期照護
 - FCI指標應是一個可行之探討方向，但因資料取得較晚，且內容複雜，尚須時間了解，故不在本研究中應用。

36

未來建議

□ 研究用資料取得

- 事故及死因資料可在既現有的模式下，持續資料的更新即可。
- 車險與事故可連結資料約有13萬，應可先著重於已連結資料之進一步分析探討。
- 健保資料為相當寶貴之資料，建議未來應繼續取得，以增加本系統之資料連結效益。
- 人車監理違規資料可做為觀察事故前後之駕駛習慣，未來可再作深入探討。

37

未來建議(續)

□ 資料系統維護

- 當進行大量之資料處理時，仍須借助SQL指令之操作，未來應可就資料庫的維護及效能調校再進行更深入的探究。
- 當有新格式資料匯入時，會需變更資料處理及欄位之設定，未來應可朝參數化及一致化的欄位名稱進行改善。
- 資料系統的改善為長期性之工作，編列經費以進行長期性之發展。

□ 連結資料應用

- 特定主題及常態性之研究
- 與其他單位之研究者之合作

38