國立政治大學資訊科學系

Department of Computer Science

National Chengchi University

碩士論文

Master’s Thesis

雲端委外語意式資料保護

Protection of Outsourcing Semantic Data in the Cloud

研究生：鄭國平

指導教授：胡毓忠

中華民國一百零二年三月

March 2013

雲端委外語意式資料保護

Protection of Outsourcing Semantic Data in the Cloud

研 究 生：鄭國平　　Student：Kuo-Ping Cheng

指導教授：胡毓忠　　Advisor：Yuh-Jong Hu

國立政治大學
資訊科學系
碩士論文

A Thesis
submitted to Department of Computer Science
National Chengchi University
in partial fulfillment of the Requirements
for the degree of
Master
in
Computer Science

中華民國一百零二年三月

March 2013

雲端委外語意式資料保護

摘要

企業基於業務需求會蒐集大量的個人資料。近來，企業基於成本考量往往會將資料委外儲存到雲端資料庫服務中，並自行建置資料查詢介面給使用者。但將資料委外到雲端資料庫服務中，雲端資料庫服務提供者便可能侵犯資料擁有者隱私。此外，資料查詢介面也必須根據不同使用情境提供不同揭露程度的資料給使用者，像是基於診療目的的醫生應該使用完整的資料，而醫療研究人員便只能使用匿名處理的資料。如何根據使用情境提供不同揭露程度的資料來確保資料使用上的隱私保護是一個重要的議題。

本研究將探討結構化單一資料源的資料委外和委外資料使用時的隱私保護，藉由在資料委外前以加密結合分割的方式處理資料，以及運用 OWL 本體論和 LP 規則語言設計三種規範：Access Control Policy (ACP)、Data Handling Policy (DHP)和 Data Releasing Policy (DRP)，並且以規則語言來落實規範。透過這三種規範的合作和分工來提供揭露程度不同的資料確保資料委外和使用上的隱私保護。

Protection of Outsourcing Semantic Data in the Cloud

Abstract

Enterprise collects a great amount of personal identifiable information (PII) for business service. Recently, these PII data are outsourced for Database-as-a-Service (DaaS) in the cloud to reduce the enterprise's data administration cost. However, this might provide an opportunity for cloud service providers (CSPs) to infringe data owners' privacy. In addition, a data controller of enterprise should provide an interface for data users with different level of data usage based on its access context. For example, a doctor can use a patient's complete PII when (s)he performs the medication practice. On the other hand, a researcher for medical analysis can only use de-identifiable PII. Therefore, the research challenge is to ensure how privacy protection criteria are satisfied for various data disclosure when using associated data usage context.

In this study, we address the research problem for outsourcing data protection from a single data source in the cloud. We apply encryption and fragmentation techniques for these outsourcing data to avoid privacy violation risk from CSPs. Furthermore, we use OWL-based ontologies to represent there types of data protection policies, i.e., Access Control Policy (ACP), Data Handling Policy (DHP), and Data Releasing Policy (DRP). In addition, we use Logic-Program (LP) rules to enforce these policies. Finally, through integration and collaboration of these policies, we ensure the satisfaction of privacy protection criteria and achieve different level of outsourcing data disclosure in the cloud.

致謝

首先感謝我的指導教授胡毓忠博士，因為教授才能讓我進入語意網的領域，除此之外，教授在課餘時間也常常藉由反覆的討論來讓我學到不少研究思維，並且在我研究生的兩年中指點我方向，學生滿懷感謝。

另外實驗室的同學們更是我互相討論的好伙伴，不論是一起度過一段時光的學長竣展、協達和迪嶸，還是同學雅玲，在實做上或是研究方向都幫了我不少忙。專班的世麒學長、郁婷學姊以及博班的穩男學長，都是陪我在研究的路上，並且提供了我許多寶貴的意見，這些點點滴滴我會牢記心頭的。

最後，謹以此文獻給我摯愛的雙親。

目錄

圖目錄

表目錄

第1章

導論

1.1 研究動機

隨著網路傳輸速率和電腦處理效率的提升，由組織所收集來的資料藉由網路來完成分享的動作已經是一種趨勢。而因為業務或管理上的需求，組織在蒐集大量的個人資料後會將其建置成資料庫如醫院的病患資料庫或企業的營運、薪資資料庫等，資料庫龐大的資料量使得組織必須花額外的費用在資訊人員的聘請和資料庫安全架構的建置等。通常，組織並沒有足夠的成本去應付這龐大的開銷，因此資料委外成為了最好的選擇。

既有的資料委外是指將資料庫委外儲存到以雲端技術建置資料庫服務當中，是儲存空間的租用，稱為雲端資料庫[1]，如 Amazon 推出的 S3（Simple Storage Service）或 Google Cloud SQL 等。它的好處在於” Pay-as-you-go”的概念，租用雲端資料庫的費用是隨著使用量的增減而隨時調整，如以 Gigabyte 收費，這可以讓組織減少租用資料庫的費用，對於雲端資料庫服務提供者而言也能藉此動態調整提供雲端資料庫服務的成本[2][3]。一般而言，只有組織有權限進行委外資料的存取，但由於雲端資料庫服務提供者直接管理整個雲端資料庫服務的運作，所以雲端資料庫服務提供者有能力基於牟利或好奇心而私自窺探、使用委外資料[4]。因此，資料委外時如何保護資料擁有者的隱私是一項重要的議題。

為了使用委外資料，組織會建置資料查詢介面提供給使用者，但使用者必須經由個人資料保護法的概念驗證後才能使用資料，同時並不是每一個使用者都能使用自己所需的資料，否則可能會導致使用者侵犯資料擁有者的隱私。使用者可分為兩類：一般使用者與統計分析使用者。舉例而言，一般使用者如主治醫生知

道越多病患資料越有利於對病患做出適合的診斷，同時病患也同意主治醫生查詢自己所有的資料，因此以診斷目的下主治醫生可以查詢病患任意的資料是較為洽當；而統計分析使用者如醫療研究人員也需要病患資料做醫療研究用途，在此情況下病患或許同意自己的資料作為醫療研究用途，但並不希望醫療研究人員可從揭露的資料裡推論出資料擁有者的身分進而侵犯隱私，因此會先將資料 De-identify(不能識別)後才揭露[5]。De-identify 是指讓資料具有匿名性的處理，也就是將個人可辨識資料 (PII)做處理，如遮罩 ID 或 SSN 等欄位，但 PII 除了上述單一識別特性的欄位外，某些欄位單獨檢視時並無法 Re-identify(再度識別)資料擁有者的身分，但一起檢視時便可 Re-identify 資料擁有者的身分，如 Gender、DoB 和 ZIP 這個欄位組合[6]。因此，如何讓這兩類型的使用者，如醫生或者醫療研究人員，分別使用不同揭露程度的資料，來保護資料擁有者的隱私是一項重要的議題。

本研究將以委外資料到雲端資料庫為背景，探討如何在單一結構化資料源的資料委外時進行隱私保護，並以本體論與規則語言設計三種規範：取控管規範(Access Control Policy,ACP)、資料處理規範(Data Handling Policy,DHP)和資料釋放規範(Data Releasing Policy, DRP)來提供 Subject-based Query (SBQ)和 Pattern-based Query (PBQ)這兩種查詢模式[7]，來完成委外資料使用的隱私保護。

1.2 研究目的

本研究的主要目地有二個，第一是當組織決定將資料庫委外儲存到雲端資料庫時，如何處理資料以確保隱私不被侵害。第二，為了保護委外資料使用上的隱私如何提供 SBQ 和 PBQ 給使用者查詢資料，並且 PBQ 須避免讓使用者 Re-identify 資料擁有者的身分的情況發生。本研究會透過設計 ACP、DHP 和 DRP

這三種規範來落實上述隱私保護。主要的研究方向如下：

- 資料委外前如何處理資料才能保護委外資料的隱私。
- 如何將委外資料的整合方式與本研究的規範結合，使得委外資料能夠在符合個人資料保護法的概念下被使用者使用。
- 如何在本體論中定義 PII，提供 SBQ 和 PBQ，並確保使用 PBQ 後的資料揭露並不會讓使用者辨別資料擁有者身分進而侵犯隱私。
- 如何挑選 SDC 處理可能引起 Re-identify 的資料，確保使用者不能利用該資料辨別資料擁有者身分。

1.3 各章節概述

本文第二章是研究架構的背景說明；第三章則是對於相關研究說明；第四章會針對研究架構作完整的描述，並且針對 ACP、DHP 和 DRP 以及落實資料揭露的步驟來詳述；第五章則是本研究的方法驗證；第六章則為總結本研究。

第2章

研究背景

2.1 資料揭露的隱私風險

隨著電腦運算能力增強和網路傳輸效率的提升，透過網路來做資料的分享與散佈已經是一種習慣。舉例來說，政府機關會定期公布一些歷史性的資料如證券交易以及企業的營運資料或客戶資料揭露給組織內部、外部的使用者使用等。隨著資料分享，使用者侵犯資料擁有者隱私的風險也提高。我國於民國 101 年實施的新版個人資料保護法第八條：「公務機關或非公務機關依第十五條或第十九條規定向當事人蒐集個人資料時，應明確告知當事人下列事項：一、公務機關或非公務機關名稱。二、蒐集之目的。三、個人資料之類別。四、個人資料利用之期間、地區、對象及方式。…」。由上述可知，在蒐集資料時除了盡到明確告知的義務外，為了保護資料擁有者的隱私，也必須以其概念如使用者的身分、目的、地區和方式等來驗證使用者是否可以使用資料。根據被揭露的資料的種類，使用者侵犯隱私的方式也不同。被揭露的資料可分成兩種：Microdata 和 Macrodata[5]。Microdata 是指揭露原始欄位的資料，像是使用者查詢 Name、Sex 和 DoB 欄位的資料而使用者得到的資料也是這三個欄位的資料。Macrodata 則是指揭露經過統計運算過後的數據，像是使用者查詢 Sex 為男性的總人數而使用者得到的是加總的結果。

Microdata 的使用者可分為兩類：一般使用者與統計分析使用者，如醫生或醫療研究人員。前者像是主治醫生為了做出正確的診斷在得到病患的同意下可查詢病患的任何資料；但後者像是醫療研究人員雖然病患同意將部分資料用於醫療研究用途，但卻不希望醫療研究人員從部分資料裡推論出病患的身分進而侵犯隱

私，因此會將揭露資料中的 PII 做處理，像是去除單一識別欄位如 ID 或 SSN 後，使其具有匿名性，再揭露給使用者。但 PII 不只有單一識別欄位，舉例而言，Gender、DoB 和 ZIP 這個欄位組合雖然個別欄位無法識別個人身分，但同時檢視時卻能 Re-identify 資料擁有者的身分[6]。因此如何提供不同揭露程度的資料給使用者是 Microdata 的隱私保護議題。而對於揭露 Macrodata 而言，使用者看到的是資料經過統計運算後的數據，一般而言較難以再度識別資料擁有者的身分，但使用者還是可以透過多次查詢來再度識別資料擁有者的身分。舉例來說，使用者可先查詢 Sex 為男性的人數，接著再查詢 Sex 為男性並且 ZIP 為 252 的人數，假如兩次查詢結果相差為 1，那就代表使用者已經再度識別資料擁有者的身分；而假如兩次查詢結果數據相同，使用者也可藉此得知 ZIP 為 252 的條件無法再度識別資料擁有者的身分而更換查詢條件。

本研究只會探討 Microdata 揭露的情況下的隱私保護，也就是如何提供不同揭露程度的資料給使用者的議題。本研究將會設計 ACP、DHP 和 DRP 這三種不同的規範以個人資料保護法的概念驗證使用者並提供 SBQ 和 PBQ 來解決上述議題，確保資料使用上的隱私保護。本研究並不探討哪些欄位組合可以辨別個人身分，而是集中在已知這些組合的情況下如何確保匿名性。

2.2 雲端環境的安全保護

雲端運算的概念就是將多個伺服器串聯並且藉由一個控制中心統整、分派資源給所需的工作。這種概念的好處在於，當不需要資源的時候，控制中心就會將資源收回，而需要的時候，就會酌量給予，可避免掉不必要的資源浪費[2][3]，運用雲端運算建構服務會是未來主流之一。一般而言，雲端服務提供商會以額外的防護機制如硬體防火牆、入侵偵測等，打造一個安全的雲端環境提供給使用者。但是這樣子的防護機制只針對來自雲端外的攻擊，並沒有考量到來自雲端內部的

攻擊。舉例來說，惡意使用者向雲端服務提供商租用雲端服務後，惡意使用者便有雲端服務提供商提供的介面可以登入雲端服務，因此便繞過雲端服務的防護機制。

為了建立安全的雲端架構，EU OpenTC 提出 TVD 技術[8]，主要包含兩個部分：Trust Computing 和 Isolation。Trust Computing 是藉由 Trust Platform Module (TPM)晶片將具有信賴度的雲端服務藉由特殊 Function 運算成一筆運算碼儲存，並且每次啟動服務時會立即再運算一遍並且和已經儲存的運算碼比較。如果相同代表雲端服務仍然可以被信賴，但若不相同則代表雲端服務可能已遭更改變成不可信賴。而 Isolation 則是使用虛擬網路 (VNET)技術來分隔每個雲端服務所使用的虛擬網路資源，使用相同虛擬網路資源的雲端服務所形成的集合稱為 TVD。使用相同虛擬網路資源的雲端服務群即屬於同一個 TVD，不同的 TVD 使用不同的虛擬網路資源。

運用 TVD 技術建置雲端環境是保護雲端和資料的必要方式[9][10]，本研究中的資料委外也是將資料庫委外到以 TVD 技術保護的雲端資料庫中，以確保資料的安全。

2.3 資料委外的隱私保護

傳統的資料委外是聚焦在資料庫的委外，資料庫服務提供者會提供資料庫服務租用者適合的 API 進行資料查詢、新增、修改和刪除等動作。而運用雲端運算建置資料庫服務稱為雲端資料庫服務[1]，和傳統的資料庫服務有些差異在於服務的提供是藉由雲端虛擬化技術，因此提供量、效能等會具有彈性，也因此就像家庭用電一般，是採用” Pay-as-you-go”的概念[2][3]。而無論是傳統的資料委外和現在的雲端資料庫委外，資料庫服務租用者都是將資料庫整個委外到資料庫服務，而資料庫服務提供者則有責任確保資料庫不被任意第三方使用。資料庫服務

租用者必須去信賴資料庫服務提供者不會窺探、利用資料，但這是難以保證的事情，畢竟資料庫服務提供者有這個能力去窺探、利用資料牟利而資料庫服務租用者也未必會知道這個事情的發生，所以僅能相信資料庫服務提供者會提供正常的服務給資料庫服務租用者，而對於資料隱私的保護還必須由資料庫服務租用者自己落實[4]。因此，美國加州參議院於 2003 年制定法律，規定「資料在委外前必須經過加密、分割的方式處理後，才能被委外」[11]。

本研究將探討在將資料委外儲存到雲端資料庫服務的情況下，如何處理才能確保資料擁有者的隱私。並且進一步探討委外資料使用時的隱私保護。

2.4 本體論與規則語言

本體論(Ontology)最早是從哲學而來的名詞，現在則被應用到語意網技術中。我們可以利用本體論來架構一個領域知識(Domain Knowledge)，利用它可以實現半自動甚至全自動的操作，像是資料整合、資料搜尋與檢索或代理人的溝通。Ontology 中包含下列三項要素：概念(Class)、屬性(Property)和實體(Instance)。概念代表本體論中對某類實體的集合或概念；屬性代表本體論中實體與實體或概念與概念之間的關係。而實體代表本體論中的個別真實例子。

OWL(Web ontology Language)是W3C推薦的本體論語言，是由DAML及OIL所結合演變而來。OWL是第一個本體論語言結合敘述邏輯的元素，以網路標準語言XML、RDF的方式呈現。而隨著表達能力等級不同，OWL又分為三種，分別是OWL DL、OWL DL和OWL Full。OWL所擁有的描述邏輯能力是以類別為基礎和關聯性的推論為主，舉例來說一台電腦如果有運算速度很快的CPU便是屬於快速電腦的OWL可以寫成”Computer∩∃hasCPU.∃hasSpeed.HighSpeed⊆FastComputer”。

規則(Rule)則是可以在特定的條件下推論出新結論，通常規則的形式是由標

頭(head)和主體(body)所構成。標題是一個原子公式(atomic formula)，代表可能符合的事實而主體是一串原子公式(atomic formula)，當主體的原子公式都滿足就代表標題的原子公式所描述的事實成立。SWRL(Semantic Web Rule Language)是一套語意規則語言，是由Web Ontology Language(OWL)子語言OWL-DL與OWL Lite以及Unary/Binary Datalog RuleML為基礎的規則描述語言[12]，而SQWRL為SWRL的加強[13]。本體論結合規則可以強化原有本體論不足的推論能力。

本研究假設情境為醫院在進行資料委外和資料使用時的隱私保護，並且醫院為具有龐大結構的醫院，如有不同分院和所屬的病患資料庫等，因此本研究採用語意網技術來設計三種規範ACP、DHP和DRP。本研究運用本體論可以描述概念的方式，描述資料型態和資料屬性，判斷那些資料的揭露會讓使用者判斷出資料擁有者的身分，進而保護隱私。本研究用OWL DL和SWRL以及SQWRL設計三個規範ACP、DHP和DRP，提供SBQ和PBQ以確保在資料使用上能保護資料擁有者的隱私。

2.5 資料整合

資料整合是在不同資料來源下透過單一的介面來查詢各個資料來源中的資料。而關聯式資料庫的整合主要著重於 Schema 的整合，也就是以不同資料庫的 Local Schema 中的定義先產生各自的 View，代表該資料庫可供整合的部分；另外產生一個 Global Schema(Mediators)作為使用者查詢時的介面，當查詢進行時，藉由將 Global Schema 的查詢式轉換成 Local Schema 的查詢式來對每一個資料庫做查詢，最後再將結果回傳給 Mediators[14]。

由於 View 的產生有時會連帶產生一些限制，而這些限制又必須使用一個方式搜集到上層在合併時一併加以考慮，因此 Diego Calvanese 等人提出運用本體論作為 Local Schema 的表達[15][16]。好處是本體論作為 Global Schema 是以概

念(Concept)的方式進行整合，並且有較強的描述能力去描述諸多限制，也因此原先對應的方式也從 SQL Query to SQL Query 轉成了 SWRL to SQL Query，由於表達能力等級的落差，必須經由 Query-Rewriting 的技術才能讓轉換對應完成。

Query-Rewriting 是藉由本體論 Schema 和關聯式資料庫之間的對應關係來完成。對應關係有三種：Local-As-View (LAV)[17]、Global-As-View (GAV)和 Global-Local-As-View (GLAV)。LAV 是指以關聯式資料庫的 View 對應到上層本體論 Schema 的 Query；GAV 則是以上層本體論 Schema 的 View 對應到關聯式資料庫的 Query；而 GLAV 則是上層本體論 Schema 的 Query/View 和關聯式資料庫的 View/Query 互相對應。

第3章

相關研究

3.1 隱私導向的存取控管系統

為了落實資料使用時的隱私保護，研究[18]運用本體論和規則語言來設計三種不同的規範：Access Control Policy、Data Handling Policy 和 Releasing Policy[19]，來落實保護隱私的資料揭露。Access Control Policy 是由企業所設定，負責進行使用者存取服務/資料時的驗證。而 Data Handling Policy 則是當資料擁有者的委外資料要被企業或第三方查詢、使用時，負責驗證使用者的使用情境是否符合資料擁有者設定，如果符合則可揭露，反之。假如有第三方要求將資料庫整個搬回時，Data Handling Policy 也必須和資料庫一併流出，並且在第三方的資料使用也須經由 Data Handling Policy 的驗證。Releasing Policy 負責驗證使用者的 PII 資料流出的使用情境是否符合使用者設定，如果符合則可揭露，反之。

表 1、Access Control Policy、Data Handling Policy 和 Releasing Policy 的設定範例

規範名稱	設定方	允許所需的條件
Access Control Policy	ACME 公司	連絡電話和信用卡號
Data Handling Policy	Alice	限定於匿名化交易時可提供
Releasing Policy	Alice	必須是有上市上櫃的公司

舉例來說，有一個使用者叫做 Alice， 她想要去存取 ACME 公司的網站服務時，就必須經過這三種規範的驗證，而 Alice 和 ACME 公司設定的三種規範內容請參考上表。一開始 Alice 會先經過 ACME 網站的 Access Control Policy 驗證，驗證過程中 ACME 網站會要求 Alice 的連絡電話以及信用卡號這兩個 PII 資料，接著 Alice 的 Releasing Policy 便會去驗證是否 ACME 網站是否是使用者設定可

揭露的範圍，因此 ACME 網站必須設法去證明自己為上市上櫃公司，假如證明成功，Releasing Policy 則會將聯絡電話和信用卡號揭露給 ACME 網站並通過 Access Control Policy 的驗證；假如反之，則不會揭露資料並且使用者也無法通過 ACME 網站的 Access Control Policy 的驗證。通過 Access Control Policy 後，也代表 Alice 的聯絡電話和信用卡號交給 ACME 網站，在傳送資料的同時 Alice 也會將自己 PII 資料的隱私偏好一併交給 ACME 網站。當 ACME 公司或第三方想要使用 Alice 的聯絡電話和信用卡號時，必須參考 Alice 所設定 PII 資料的隱私偏好，由上表可知，Alice 的聯絡電話和信用卡號只能用於 ACME 公司或第三方的匿名化交易中。

雖然透過這三種規範能夠控管資料揭露的使用情境並滿足資料擁有者的預想，但是並沒有探討資料委外時的隱私保護以及考量提供不能 Re-identify 資料給使用者的情況，因此距離本研究的研究目的仍然有一段差距。

3.2 委外資料的查詢

資料委外前，依法必須將委外資料以加密或分割的方式來落實隱私保護。此舉雖然可以有效防止資料庫服務提供者私自利用資料，但也需要額外的轉換機制來轉換查詢式，來完成資料查詢。根據不同的處理方式如加密或分割，轉換機制的方式也有所不同[4]。假如委外資料是以加密的方式處理，委外資料會以密文的形式儲存在資料庫中，而使用者送出的查詢式是屬於明文的形式，因此轉換機制需要將查詢式轉換為密文的形式後才能完成資料查詢。常見的有兩種方式：額外增加 Index 欄位或將查詢條件轉為密文。額外增加 Index 欄位會在資料委外時，額外增加 Index 欄位去對應到未加密前的資料。因此轉換機制會將查詢式的查詢條件轉換為 Index 值，透過 Index 值找到正確的資料。而將查詢條件轉為密文則是直接將查詢式的查詢條件也以相同的加密函數轉換為密文的形式，因此查詢時

藉由密文的比對去找到正確的資料。

而委外資料以分割的方式處理時，資料庫會切割為數個子資料庫，因此轉換機制也須將查詢式轉換為對應到子資料庫的子查詢式，最後再整合子查詢式的資料。整合的方式是以資料委外時額外設定的 Index 欄位來判斷哪些資料屬於同一個資料擁有者。舉例來說，使用者想要查詢出生於 1970/01/01 前並且沒有吃藥的病患的病患姓名和疾病，使用者會送出查詢式 Q，而查詢式 Q 會經由轉換機制轉換為對應的子查詢式 Q1 和 Q2，最後在比對 IndexKey 欄位來整合不同子資料庫中的資料，查詢式和子查詢式之間的轉換如下圖所示：

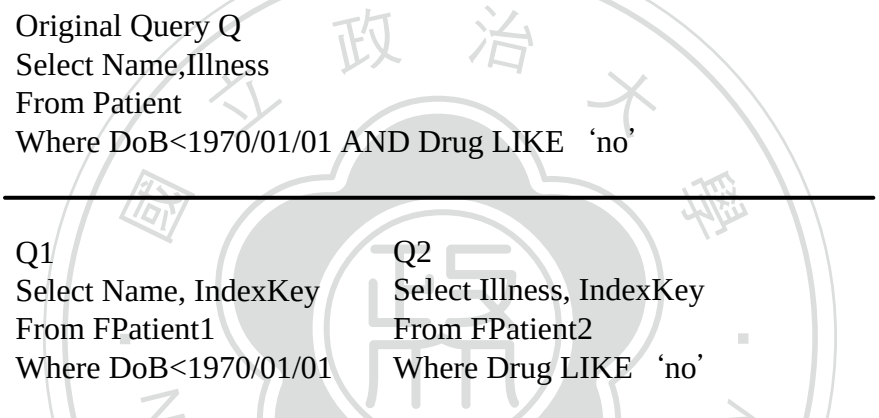

圖 一、查詢式與子查詢式轉換概念圖

而隨著查詢式的查詢條件不同，子查詢式也是有可能先執行其中一條，接著根據第一條子查詢式查詢結果來設定第二條子查詢式的查詢條件，最後在整合子查詢式的資料。

本研究的研究議題之一為落實委外資料使用時的隱私保護，委外資料需要經過整合後才能使用，但本研究和既有研究的整合方式不同的是，本研究只整合符合資料擁有者的隱私偏好的委外資料後再提供給使用者查詢，而上述研究是整合所有符合查詢條件的委外資料揭露給使用者。因此既有研究的委外資料使用並沒有考量到資料擁有者的隱私偏好的概念，也沒有提供資料的匿名性來保護隱私。

3.3 企業隱私授權語言

EPAL(Enterprise Privacy Authorization Language)[20]是由 IBM 公司全球八大研究中心之一的蘇黎士研發中心公司致力發展的一種規範語言，它提供了企業蒐集大量客戶個人資料時，隱私規範設定的標準化規格。EPAL 主要構成的六大要素為：

1. 資料使用者(Data users)：存取資料的個體。
2. 行為(Actions)：存取資料的動作，像是讀取或寫入資料。
3. 資料種類(Data Categories)：定義企業擁有的資料種類。
4. 目的(Purposes)：表示資料使用於特定目的。像是行銷或訂貨。
5. 條件(Conditions)：表述一些在存取資料前必須滿足一些條件。
6. 義務(Obligations)：允許存取下企業必須採取額外的步驟。

透過 EPAL，企業在使用資料時便能識別哪些資料是客戶同意使用以及不同意使用的欄位。但 EPAL 只提供預設的詞彙來進行隱私偏好的設定，因此不能滿足不同使用情境下所需要的要素，像是資料擁有者如果考量存取地點跟存取時間，則必須新增要素 Location 和 Time 的需求。此外，EPAL 是能讓資料擁有者設定自己的隱私規範，也就是資料擁有者的隱私偏好，但並不能用以描述某些會引起 Re-identify 的資料。

第4章

研究架構

4.1 研究架構設計

本研究的研究目的有兩者：實現單一結構化資料源的資料委外時和委外資料使用時的隱私保護。因此，本研究的研究架構會建立在資料委外模式上，在講解本研究架構前，先行介紹資料從蒐集到利用流程中的每個角色。

一開始，企業會依照各自的需求會向資料擁有者蒐集大量的個人資料並將其建成資料庫，在此企業為資料控制者。之後向雲端資料庫服務提供者租用以 TVD 技術建置的雲端資料庫服務並且將資料庫委外儲存。最後，企業會利用雲端資料庫服務提供者所提供的資料存取介面，建置自己的資料查詢介面給使用者使用，使用者可能是個人、其他組織或者是組織內部的員工。上述概念如下圖所述：

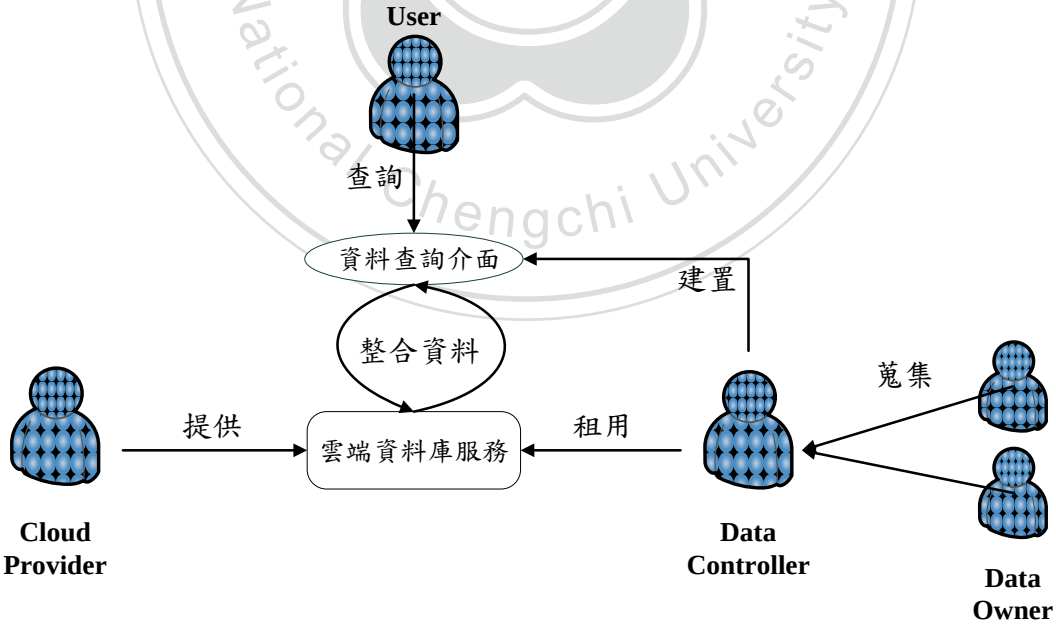

圖 二、資料從蒐集到利用的概念圖

由上面的概念圖可知，可能侵犯隱私的時機有二：1. 委外儲存到雲端資料

庫服務時可能會被雲端資料庫服務提供者私自窺視、利用。2. 不同使用者查詢資料時，必須揭露不同程度的資料。如果讓使用者查詢過多的資料便可能讓使用者有機會侵犯隱私。因此，本研究的研究架構也分成兩部分：資料委外和資料使用的隱私保護。資料委外的隱私保護：為了讓雲端資料庫服務提供者不能使用資料，資料委外前必須使用加密或分割的方式處理資料。在處理資料的同時，也必須留下資料還原的線索以便使用資料。

資料使用的隱私保護：為了限制使用者不能查詢過多的資料，本研究參考文獻[7]中 SBQ 和 PBQ 的概念。SBQ 是指使用者查詢的任何資料都會揭露，如醫生查詢病患資料則必須全數揭露以利正確的診斷;而 PBQ 則是指使用者只能查詢部分資料，並且在查詢可辨別個人身份的資料時必須特別以 Statistic Disclosure Control (SDC)處理後才能揭露資料[5]。本研究將上述的概念以規範的方式描述，並以本體論和規則語言設計三種不同的規範：存取控管規範 (Access Control Policy,ACP)、資料處理規範(Data Handling Policy,DHP)和資料釋放規範(Data Releasing Policy, DRP)來提供 SBQ 和 PBQ。ACP 是由企業所設定，會根據使用者的使用情境判別使用者是否有權限查詢資料以及授權使用者能使用的查詢模式是屬於 SBQ 或 PBQ，如果是 SBQ 則使用者可以得到所需的資料；而如果是 PBQ 使用者只能得到部分資料且可辨別個人身分的資料必須以 SDC 的方式處理後才能揭露。DHP 是透過描述資料擁有者對於被蒐集的資料使用上的限制，也就是隱私偏好，讓資料的使用可以滿足於資料擁有者的預想。它負責從委外的資料庫中整合出符合使用者的使用情境，也就是滿足該隱私偏好的資料來給使用者查詢。DRP 則是負責資料的揭露，根據使用者被授權的查詢模式屬於 SBQ 或 PBQ 來決定如何揭露資料。假如是 SBQ 則將使用者查詢的資料全部揭露；而當 PBQ 時則必須將可辨別個人身分的資料以 SDC 的方式處理後，才將使用者所查詢的資料揭露。

此外，本研究之所以設計三種規範 ACP、DHP 和 DRP 是因為 DHP 單純是

資料擁有者憑著自身的喜好以及經驗設定，也就是說，資料擁有者並不一定知道揭露甚麼程度的資料能讓使用者 Re-identify 資料擁有者的身分，因此設定的隱私偏好並不一定能達到隱私保護的功用。舉例來說，資料擁有者讓醫療研究人員查詢 Name、Gender、ZIP、DoB 和 Disease 欄位，雖然 Name 並不能辨別資料擁有者的身分，但 Gender、ZIP 和 DoB 三個欄位組合後便可辨別資料擁有者的身分進而侵害隱私。DHP 並不能確保隱私不被使用者侵犯，因此額外設定 DRP 來判斷哪些資料可辨別資料擁有者的身分以便保護隱私。

ACP、DHP 和 DRP 彼此有著分工流程的關係，當使用者進入資料查詢介面時會先經過 ACP 的驗證並授權 SBQ 或 PBQ。接著 ACP 會啟動 DHP，DHP 將使用者能夠使用，也就是滿足的隱私偏好所屬資料從雲端資料庫中整合出來給使用者查詢。最後 DRP 會被啟動，根據 ACP 授權使用者的 SBQ 或 PBQ，判斷是否直接揭露資料或者是以 SDC 處理資料後才揭露。本研究架構圖如下所示：

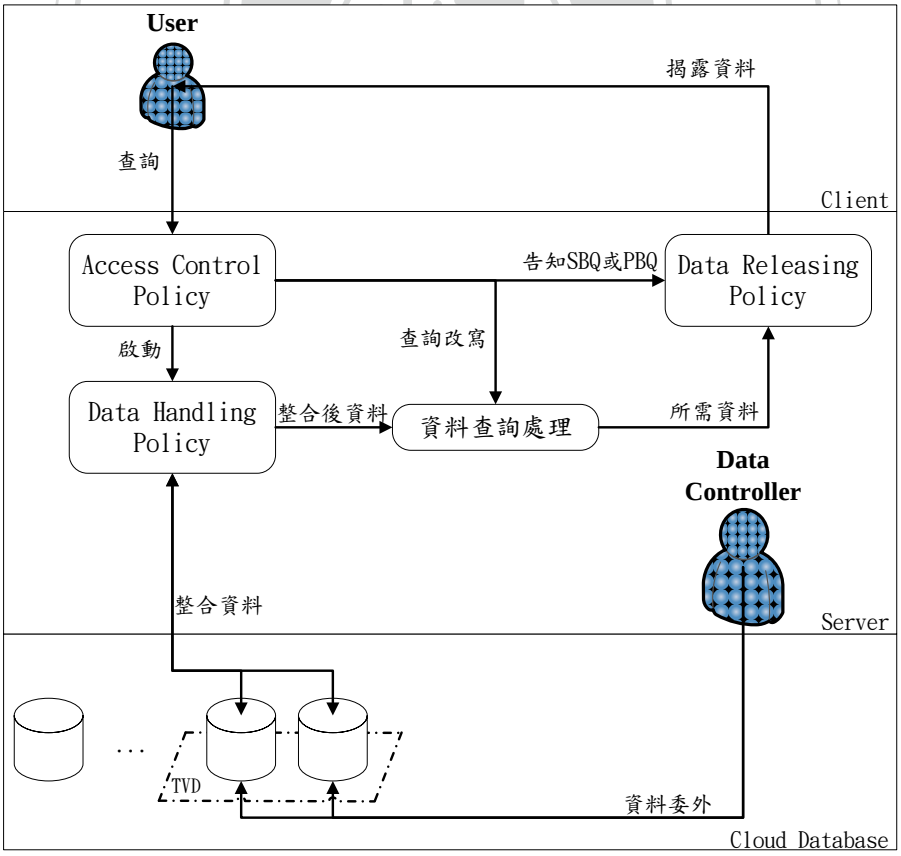

圖 三、研究架構流程圖

本研究的研究情境是以 a 醫院的病患資料庫委外和使用為例。a 醫院會蒐集大量的病患資料並建成病患資料庫但並沒有足夠的成本購買伺服器硬體、聘請資訊人員，因此向雲端資料庫服務提供者承租雲端資料庫服務。在委外儲存病患資料庫前，先以加密結合分割的方法處理資料，達成資料委外的隱私保護。接著利用雲端資料庫服務提供者提供的資料存取介面建立資料查詢介面讓使用者查詢資料。a 醫院同時也設計自己的 ACP、DHP 和 DRP，確保委外資料使用上的隱私保護。

本研究採用 OWL DL 和 SWRL 以及 SQWRL 來描述跟落實 ACP、DHP 和 DRP。在下面的 4.2 節會介紹企業在進行資料委外前如何運用加密結合分割的方式處理資料；4.3 節分別描述 ACP、DHP 和 DRP 這三種規範設計的細節。

4.2 資料委外

4.2.1 委外資料的保護方式

委外資料的隱私保護的概念有分成兩個面向：保護資料本身和保護資料的關聯性。保護資料本身是指讓資料本身處於不可識別的情況下而保護資料的關聯性是指讓單一資料本身處於可識別但資料跟資料之間的關聯性處於不可識別的情況。依據這兩種面向，資料委外保護的方式可分為加密、分割、自行保管資料與加密結合分割這四種[4][21][22][23]。

- 加密：運用加密函式將資料加密後變成密文再委外到外部資料庫。由於服務提供者缺乏加密函式和加密、解密金鑰，因此能保護資料的隱私。但是資料處於加密的情況下，資料的使用還必須經過資料解密的流程。為了正確查詢資料，通常會將查詢條件轉換為密文或者是增加額外的欄位代表密文的特徵值，否則便須將整個密文資料庫傳給服務租用者後才行解密，但這相當沒有

效率以及花費高額的成本

- 分割：將資料庫欄位依據能夠辨識個人身分的資料 (PII)進行分割後成為數個子資料庫再委外到外部資料庫，PII 有兩種：
  1. 單一欄位違反隱私：單一欄位即可指出特定人士，像是身份證字號。此類欄位要進行分割通常是欄位內分割，像是身份證字號分成前 5 碼後 4 碼。
  2. 多個欄位違反隱私：多個欄位在一起檢視時即可違反隱私，像是出生年月日、性別和郵遞區號。此類欄位的分割通常是將欄位間分開，並且還須加上抖亂每筆資料的次序以確保關聯性的不會輕易被串起。

  由於資料之間的關聯性被隱藏，服務提供者便無法使用資料。為了正常使用資料，資料跟資料之間必須留下還原、合併的線索，通常會在每個子資料庫新增欄位，用欄位之間的關聯性來還原資料。
- 自行保存敏感資料：將資料庫中的敏感欄位保留在服務租用者的資料庫中，而委外剩餘欄位的資料庫。保留在服務租用者和委外到服務提供者的資料庫之間也需要設定額外的欄位提供還原、合併的線索。自行保存敏感資料的好處在於服務提供者絕對沒辦法使用資料違反隱私，但通常資料庫都是相當大量，因此自行保存敏感資料只能在服務租用者能自行建置資料庫保存部分資料的情況下。
- 加密結合分割：將資料庫的欄位依據服務租用者的設定同時落實加密跟分割的保護後，才委外到外部資料庫。使用加密結合分割的好處是在於，因為儲存成本的問題，分割的動作不適合用於單一欄位資料而加密的動作又需要經由還原相當耗費時間計算，因此將單一欄位違反隱私的資料以加密保護而多個欄位違反隱私則用分割是相當普遍。

本研究將採用加密結合分割的方式來保護被委外儲存到雲端資料庫服務的委外資料。詳細步驟如下節所述。

4.2.2 資料委外保護

本研究假設 a 醫院的病患資料庫具有 ID、Name、Gender、DOB、ZIP、Disease、Doctor 和 Cholesterol 等 8 個欄位，資料內容如下表所示：

表 2、a 醫院病患資料庫欄位與資料內容

ID	Name	Gender	DOB	ZIP	Disease	Doctor	Cholesterol
A139181122	Tsai Ya Ting	Male	790109	112	CAHD	Huang Fu Yuan	203
A239726212	Yang Jing Yi	Female	781222	105	CAHD	Ma Huei Ming	231
A163190073	Tsai Ya Ping	Male	790110	111	Allergy	Wang Ho	148
A118814368	Chang Yu Yun	Male	780422	111	CAHD	Ma Huei Ming	210
A239726214	Li Guo Yuan	Female	750422	103	Allergy	Wang Ho	180
A118814369	Tsai Jian Hung	Male	770102	111	CAHD	Huang Fu Yuan	231
A239726213	Pai Yu Yun	Female	780423	110	CAHD	Huang Fu Yuan	199
A118814370	Yang Yu Yun	Male	770712	110	CAHD	Ma Huei Ming	232
A298663018	Yang Jian Hung	Female	781022	106	Allergy	Wang Ho	178
A192142787	Yang Guo Yuan	Male	801012	100	Allergy	Wang Ho	236

為了以加密、分割或結合加密和分割的方式處理資料，企業必須先行判斷欄位是屬於單一欄位違反隱私、多個欄位違反隱私或不屬於上述兩類的欄位。判斷的依據必須參考現實生活中可能發生的情況或者既有研究的內容。本研究參考文獻[10]中分類方式，在 a 醫院的病患資料庫中，ID 屬於單一欄位違反隱私、Gender、ZIP 和 DoB 為多個欄位違反隱私，而 Name、Disease、Doctor 和 Cholesterol 並不屬於上述兩類。本研究採用以結合加密和分割的方式處理委外資料，原因在於如果只以加密或分割處理資料，前者花費太多處理時間而後者處理屬於單一欄位違反隱私的欄位又太過沒效率。因此，本研究採用加密結合分割的方式，加密處理屬於單一欄位違反隱私的欄位而分割處理多個欄位違反隱私的欄位。委外資料處理的步驟如下所述：

1. 將 ID 用特定的加密演算法加密，加密演算法的挑選可以是對稱加密演算法或非對稱加密演算法。
2. 將資料庫分成兩個子資料庫，屬於多個欄位違反隱私的欄位不可完全在同一個子資料庫中，並且屬於多個欄位違反隱私的欄位也不可在兩個子資料庫中重複出現。最後，額外在各個子資料庫增加欄位 IndexKey 以及 IndexKeyTable，

利用 IndexKeyTable 紀錄 IndexKey 值之間的關係代表

兩個子資料庫的資料之間的關聯性以便日後使用資料時可以還原資料。

委外資料處理的步驟流程和 IndexKeyTable 如下所示：

ID	Name	Gender	DOB	ZIP	Disease	Doctor	Cholesterol
A139181122	Tsai Ya Ting	Male	790109	112	CAHD	Huang Fu Yuan	203
A239726212	Yang Jing Yi	Female	781222	105	CAHD	Ma Huei Ming	231
A163190073	Tsai Ya Ping	Male	790110	111	Allergy	Wang Ho	148
A118814368	Chang Yu Yun	Male	780422	111	CAHD	Ma Huei Ming	210
A239726214	Li Guo Yuan	Female	750422	103	Allergy	Wang Ho	180
A118814369	Tsai Jian Hung	Male	770102	111	CAHD	Huang Fu Yuan	231
A239726213	Pai Yu Yun	Female	780423	110	CAHD	Huang Fu Yuan	199
A118814370	Yang Yu Yun	Male	770712	110	CAHD	Ma Huei Ming	232
A298663018	Yang Jian Hung	Female	781022	106	Allergy	Wang Ho	178
A192142787	Yang Guo Yuan	Male	801012	100	Allergy	Wang Ho	236

Encryption

ID	Name	Gender	DOB	ZIP	Disease	Doctor	Cholesterol
＊＊＊＊＊	Tsai Ya Ting	Male	790109	112	CAHD	Huang Fu Yuan	203
＊＊＊＊＊	Yang Jing Yi	Female	781222	105	CAHD	Ma Huei Ming	231
＊＊＊＊＊	Tsai Ya Ping	Male	790110	111	Allergy	Wang Ho	148
＊＊＊＊＊	Chang Yu Yun	Male	780422	111	CAHD	Ma Huei Ming	210
＊＊＊＊＊	Li Guo Yuan	Female	750422	103	Allergy	Wang Ho	180
＊＊＊＊＊	Tsai Jian Hung	Male	770102	111	CAHD	Huang Fu Yuan	231
＊＊＊＊＊	Pai Yu Yun	Female	780423	110	CAHD	Huang Fu Yuan	199
＊＊＊＊＊	Yang Yu Yun	Male	770712	110	CAHD	Ma Huei Ming	232
＊＊＊＊＊	Yang Jian Hung	Female	781022	106	Allergy	Wang Ho	178
＊＊＊＊＊	Yang Guo Yuan	Male	801012	100	Allergy	Wang Ho	236

Fragment

ID	Name	Gender	DOB	IndexKey
＊＊＊＊＊	Tsai Ya Ting	Male	790109	Α
＊＊＊＊＊	Yang Jing Yi	Female	781222	Β
＊＊＊＊＊	Tsai Ya Ping	Male	790110	Γ
＊＊＊＊＊	Chang Yu Yun	Male	780422	Δ
＊＊＊＊＊	Li Guo Yuan	Female	750422	Ε
＊＊＊＊＊	Tsai Jian Hung	Male	770102	Ζ
＊＊＊＊＊	Pai Yu Yun	Female	780423	Η
＊＊＊＊＊	Yang Yu Yun	Male	770712	Θ
＊＊＊＊＊	Yang Jian Hung	Female	781022	Λ
＊＊＊＊＊	Yang Guo Yuan	Male	801012	Φ

ZIP	Disease	Doctor	Cholesterol	IndexKey
111	Allergy	Wang Ho	148	γ
103	Allergy	Wang Ho	180	ε
110	CAHD	Ma Huei Ming	232	θ
112	CAHD	Huang Fu Yuan	203	α
105	CAHD	Ma Huei Ming	231	β
100	Allergy	Wang Ho	236	φ
106	Allergy	Wang Ho	178	λ
111	CAHD	Ma Huei Ming	210	δ
111	CAHD	Huang Fu Yuan	231	ζ
110	CAHD	Huang Fu Yuan	199	η

圖 四、委外資料處理的步驟流程

表 3、IndexKeyTable 內容，代表不同子資料表之間 IndexKey 的對應關係

table1	table2
Α	α
Β	β
Γ	γ
Δ	δ
Ε	ε
Ζ	ζ
Η	η
Θ	θ
Λ	λ
Φ	φ

資料委外前將委外資料以加密結合分割的方式處理後，能夠保護資料擁有者的隱私。但是資料委外後，資料已經不是以完整的資料存在，因此直接使用委外資料可能會有語意上的缺陷。本研究會以規範的方式從委外資料庫整合出完整資料後才提供給使用者使用，而非直接使用委外資料庫中的資料，因此不會有語意上的缺陷。

4.2.3 資料委外模式

根據不同資料委外的模式，落實資料委外保護技術的細節也有所不同，資料委外模式總共可以分為兩種[10]：

1. NON-COMMUNICATING SERVERS：企業將子資料庫分別委外儲存到不同的雲端資料庫服務提供者所提供的雲端資料庫服務並且雲端資料庫服務提供者之間並不互相知道該企業有將資料委外儲存到對方的雲端資料庫服務中。
2. COMMUNICATING SERVERS：企業將子資料庫委外到同一個雲端資料庫服務提供者所提供的雲端資料庫服務或者是不同雲端資料庫服務提供者所提供的雲端資料庫服務並且雲端資料庫服務提供者彼此互相知道該企業將

資料委外儲存到對方的雲端資料庫服務中。

對於以加密處理資料而言，因為資料是以密文的形式儲存在雲端資料庫，並且加密演算法和金鑰都是握在資料控制者手中，因此不同的資料委外模式並不會有影響。而以分割處理資料而言，處理的細節就必須有所不同。這是因為資料被分割成數個子資料庫，因此需要增加額外的欄位 IndexKey 來描述子資料庫資料之間的關聯性。如果在 NON-COMMUNICATING SERVERS 下，IndexKey 欄位值可以簡單的方式設定，像是 IndexKey 值相同，即代表同一筆資料。而在 COMMUNICATING SERVERS，由於雲端資料庫服務提供者彼此互相知道，所以不能排除雲端資料庫服務提供者們可能會藉由交換資料庫的方式來得到完整的資料庫，因此不能僅僅以 IndexKey 值相同就代表同一筆資料，而是以亂數的方式設定 IndexKey 值並且再額外以 IndexKeyTable 紀錄 IndexKey 值之間的關係，以便整合出同一筆資料。

本研究是假設 a 醫院的病患資料庫委外到同一個雲端資料庫服務提供者所提供的雲端資料庫服務，屬於 COMMUNICATING SERVERS，因此以 IndexKey 的方式記錄各個子資料庫資料之間的關聯性，以便整合出同一筆資料。不同的是，

本研究並沒有實際使用 IndexKeyTable，而是將 IndexKeyTable 的關聯性紀錄和資料擁有者的隱私偏好結合紀錄於規範中，以便達成只整合符合隱私偏好的委外資料的目標。

4.3 規範的設計和資料查詢處理

在資料委外隱私保護完成後，接著要針對的是委外資料使用時的隱私保護。本研究設計 ACP、DHP 和 DRP 三種規範，並且透過三種規範的合作來完成委外資料使用時的隱私保護。

4.3.1 研究情境說明

本研究假設使用者的使用情境的各項元素會自動對應到 ACP 和 DHP 中，下面將以 a 醫院病患資料庫的一般使用者和統計分析使用者，也就是 SBQ 和 PBQ 分開說明。

假如有一名醫生在診斷目的下，想要查詢 ID 為 A239726214 的資料擁有者膽固醇指數，經過醫院的 ACP 驗證後會授權 SBQ 給該醫生。接著將 DHP 裡的使用情境與醫生的使用情境做比對，整合相同使用情境的委外資料。整合出完整資料庫後，系統會將醫生的查詢條件，經由查詢改寫後下到整合出的完整資料庫中挑選 ID 為 A239726214 的資料擁有者膽固醇指數。由於該醫生可使用 SBQ，因此 DRP 會將該筆資料直接揭露給醫生。

假如有一名醫療研究人員，想要研究在地區、性別、年紀和膽固醇指數與疾病的關聯性。經過醫院的 ACP 驗證後會授權 PBQ 給該醫療研究人員，接著系統讓醫療研究人員勾選所需的欄位 ZIP、Gender、DoB、Cholesterol 和 Disease。SBQ 和 PBQ 的 DHP 處理步驟相同，會整合相同使用情境的委外資料並且藉由查詢改寫將使用者所需的欄位挑選出來。由於該醫療研究人員是使用 PBQ，因此 DRP 必須會先判斷該使用者所使用的資料中是否會侵害隱私，而在本研究中 ZIP、Gender 和 DoB 會構成 Quasi-identifiers 以及使用者查詢敏感性欄位 Disease，因此必須要求該醫療研究人員從 ZIP、Gender 和 DoB 中挑選一欄位以 SDC 處理，醫療研究人員可根據自身的需求來取捨哪個欄位以 SDC 處理，而醫療研究人員挑選完欄位與 SDC 後，系統呼叫 SDC 處理結束才會揭露資料。

4.3.2 Access Control Policy

ACP 負責的是驗證使用者是否可以進入資料使用介面以及授權使用者能夠使

用的查詢模式是 SBQ 或 PBQ。像是醫生基於診療目的要查詢病患的資料便可使用 SBQ 而醫療研究人員基於研究目的要使用病患的資料便是使用 PBQ。ACP 授權 SBQ 或 PBQ 是根據使用者的使用情境來決定。使用情境是由使用情境要素構成，像是

DataUser、Action 和 Purpose 等，每一種使用情境都有各自的使用情境構成要素以及授權的查詢模式如 SBQ 或 PBQ，而本研究參考 P3P 和 APPEL 中的要素以及現實生活中存取控管時需要參考的因素設計使用情境構成要素。主要的使用情境構成要素如下表所示：

表 4、ACP 的使用情境、使用情境構成要素以及對應到的查詢模式

Data User	Action	Purpose	Location	Time	QueryType
Doctor	Read	Diagonosis	Hospital	Work	SBQ
Researcher	Read	Analysis	Research Center	Work	PBQ
Researcher	Read	Analysis	Research Center	Work	PBQ

由上表可知，當一名在 a 醫院上班中的醫生在診療病人的目的下是可以使用 SBQ 來查詢病患所有的資料而一名在研究中心上班中的醫療研究人員只能使用 PBQ 來查詢病患部分的資料。

當有一個使用者想要進入資料查詢介面時，會將使用者的使用情境與 ACP 中描述的使用情境互相比對，假如使用者的使用情境符合 ACP 中所描述的使用情境，便會將 ACP 中符合的使用情境授權使用的查詢模式授權給使用者。假如不符合任何 ACP 中所描述的任何使用情境，則拒絕使用者進入資料查詢介面。

本研究運用 OWL DL 去塑模上述使用情境以及使用情境要素的概念，整個架構如下圖所示：

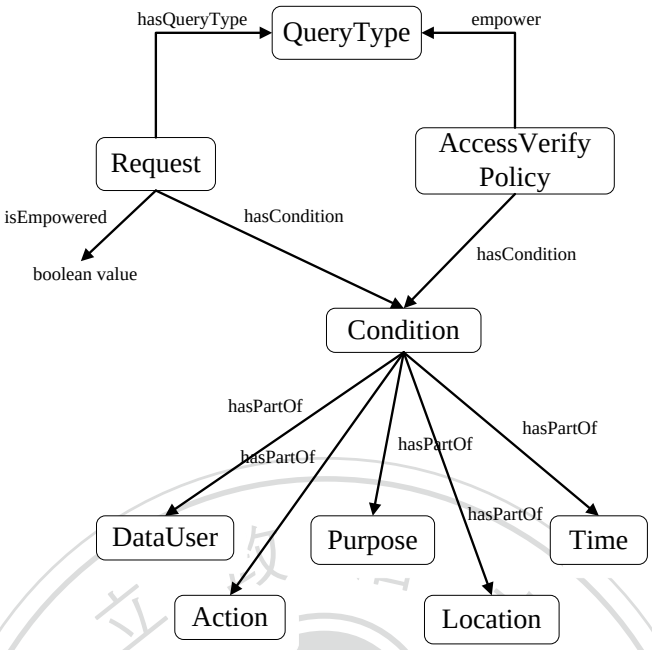

圖 五、ACP 的本體論設計

使用情境以及各種使用情境要素皆是類別(Class)的概念，而類別的實體(Instance)則代表一種使用情境或是一個使用情境要素，像是 Researcher 和 Doctor 為 DataUser 的實體而 Diagonosis 和 Statistics 為 Purpose 的實體。在上圖中，類別跟類別之間的箭頭實線代表類別之間的關係，如使用情境和各個使用情境要素之間有 hasPartOf 代表使用情境中包含各個使用情境要素的關係或者是 AcceessVerifyPolicy 或 Request 和 Condition 之間有 hasCondition 關係，代表使用者和 ACP 都有各自的使用情境。在此本研究運用 hasPartOf 來代表使用情境包含的各個資料使用情境要素，是一種結構的概念。而在語意網技術中除了以 hasPartOf 描述結構外，還有一種稱為 Functional Symbol 也能用來描述結構[24]。但不同的是，Functional Symbol 代表的是一個函數的概念，而 hasPartOf 代表類別的概念。

而箭頭實線如果不是位於類別之間則是代表類別對應到一種資料型態，像是 isEmpowered 將 Request 關連到 BooleanValue，BooleanValue 為 1 代表 Request 被驗證通過而 0 反之。Request 和 AccessVerifyPolicy 都有各自的 Condition，前者

為使用者在送出 Request 時的資料使用情境，而後者為資料控制者設定 ACP 的各種資料使用情境。

本研究假設當使用者進入資料查詢系統時，系統會以各種方式蒐集使用者的各種使用情境要素，像是 DataUser 使用帳號密碼驗證身分以及 Location 使用 IP 驗證，並自動對應到 ACP 中各種使用情境要素。當使用者和使用情境、使用情境要素之間的對應皆完成後才會開始 ACP 的驗證和授權 SBQ 和 PBQ 的動作。

本研究運用 SWRL 規則來比對使用者和 ACP 的資料使用情境是否相同，假如

相同則授權該使用情境能使用的查詢模式給使用者，SWRL 規則如下所示：

Request(?r) ∧hasCondition(?r, ?c) ∧Condition(?c) ∧Condition(?ac) ∧ hasCondition(?avp, ?ac) ∧AccessVerifyPolicy(?avp) ∧sameAs(?ac, ?c) ∧ empower(?avp,?qt) ∧QueryType(?qt) → isEmpowered(?r, 1) ∧hasQueryType(?r,?qt)

在 SWRL 的規則中提供 sameAs 的 Atom 來進行限制 Instance 的相同。舉例來說，假如有 Fred 和 Freddy 這兩個不同的實體，則 sameAs(Fred, Freddy)這個 Atom 便會不會成立。假如使用者的資料使用情境符合 ACP 中所規範的使用情境，則 Request 和 AccessVerifyPolicy 其中之一藉由 hasCondition 關係對應到的 Condition 的實體必會相同，因此運用 sameAs 的 Atom 來表達實體相同的限制，其概念如下圖所示：

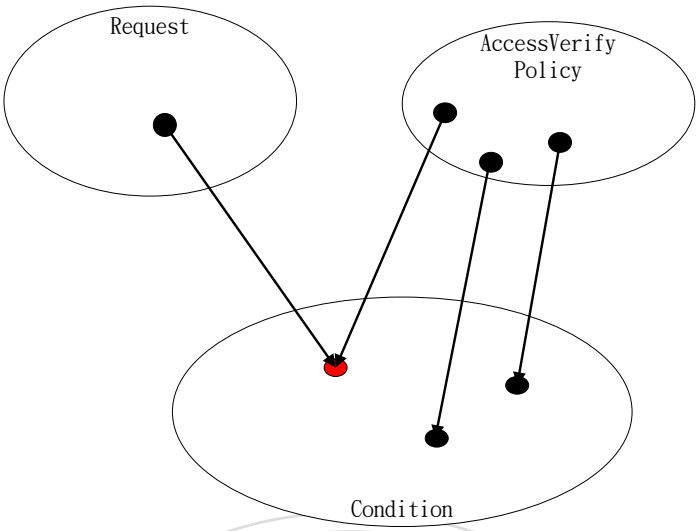

圖 六、透過 sameAs 限制實體相同的概念

在上述的 SWRL 規則中，假如使用者的資料使用情境符合 ACP 中所規範的使用情境，則會將 Request 的 isEmpowered 關係所對應到的 BooleanValue 設定為 1，並授權可以使用的查詢模式?qt，最後啟動 DHP 去整合委外資料。而如果 BooleanValue 設定為 0，代表使用者不能使用資料，則會拒絕使用者查詢。

4.3.3 Data Handling Policy

DHP 負責讓資料的使用可以滿足資料擁有者的隱私偏好。首先將使用者的使用

情境與 DHP 中資料擁有者設定的隱私偏好中的使用情境相比，假如兩者相同才會從委外資料庫中整合相對應的資料；假如兩者並不相同則相對應的資料並不會被整合。透過上述的概念讓委外資料的使用滿足於資料擁有者的隱私偏好。

企業在蒐集資料擁有者資料時，必須將設定隱私偏好的模板一併交給資料擁有者，而資料擁有者將個人資料交給企業時，會一併將設定的隱私偏好交給企業，而企業在使用資料前必須參考資料擁有者的隱私偏好，在符合的情況下才能使用資料。本研究參考 P3P 裡的要素來建置隱私偏好中的使用情境要素，共有四項：

DataUser、Action、Purpose 和 Location。每一個資料擁有者可以替自己的資料設定一個以上的使用情境，以因應不同的使用情境揭露不同程度的資料，舉例來說當醫生基於診療目的下可揭露 ID、Name、Gender、ZIP 和 Disease 這五個欄位，而當醫療研究人員基於研究目的下只能揭露 Gender、ZIP 和 Disease 這三個欄位。

ACP 和 DHP 都各有資料使用情境的要素，並且兩者大部分皆為相同。但其意義有所不同。DHP 是資料擁有者憑著自身的喜好以及經驗設定，也就是讓資料使用能夠滿足資料擁有者的預想；而 ACP 則是企業也就是資料控制者所設定，用來驗證使用者的身分和授權使用者可以使用的查詢模式。

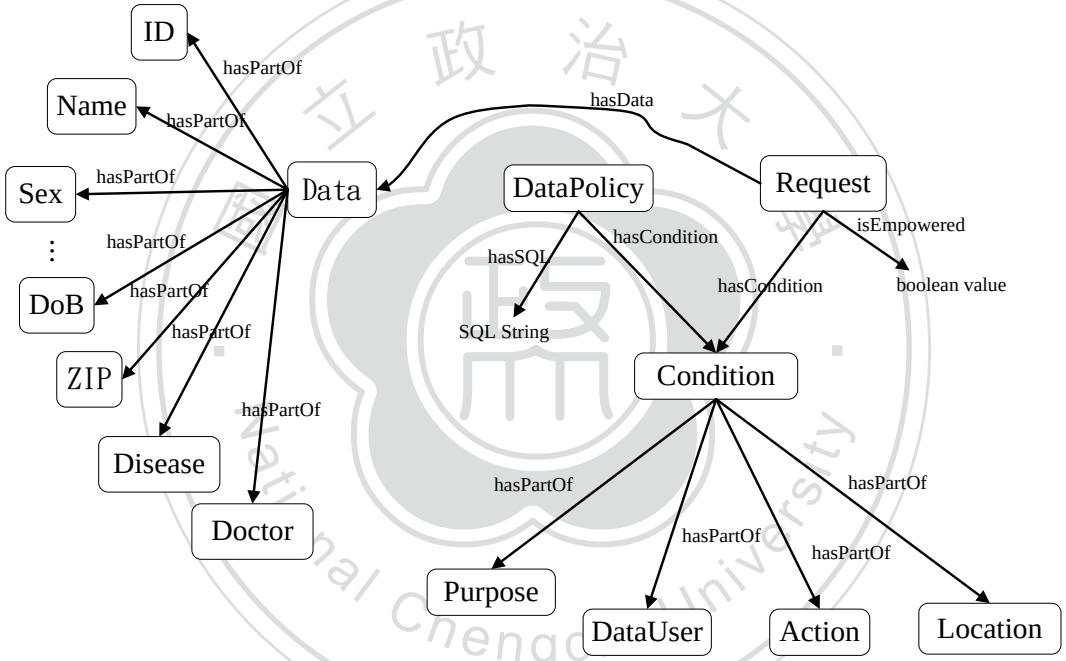

圖 七、DHP 的本體論設計

一個完整的 DHP 的本體論設計如上圖所示，在 DHP 中也用 Condition 和 hasPartOf 代表資料使用情境和使用情境要素。此外，DHP 使用 hasSQL 這個關係將 DataPolicy 對應到 SQL String，SQL String 是兩個 SQL 查詢式構成，並且利用其來整合委外資料。SQL 查詢式是用來查詢任一子資料庫中的查詢式，如下所示：

```
SELECT `ID` , `Name` , `Gender` , `DOB` FROM `p1` WHERE `IndexKey` =1
```

資料委外後，會將資料庫分成兩個以上的子資料庫，因此 SQL String 會是兩個 SQL 查詢式以上的結合，而在本研究中只分成兩個子資料庫，如下所示：

SELECT `ID` , `Name` , `Gender` , `DOB` FROM `p1` WHERE `IndexKey` =1

and

SELECT `ZIP` , `Disease` , `Doctor` , `Cholesterol` FROM `p2` WHERE `IndexKey`=101

既有的資料委外中，對於分割處理的資料必須利用 IndexKey 值的對應來整合資料。在本研究的情境中租用的雲端資料庫服務屬於 COMMUNICATING SERVERS，因此也需要額外的 IndexKeyTable 去紀錄 IndexKey 值的對應關係，而本研究則將 IndexKey 值的對應記錄在 SQL String 當中，如同上述的 SQL String 裡的兩個 SQL 查詢式中都指定 IndexKey 值，並且因為 COMMUNICATING SERVERS 所以 IndexKey 值並不相同。此外，一個資料擁有者可能會對一筆資料設定一個以上的隱私偏好，代表不同使用情境可以揭露的不同程度的資料，因此在 SQL 查詢式中根據不同的使用情境而調整 SQL 查詢式中的查詢欄位。

DHP 的 SWRL 規則如下所示，它會比對使用者的使用情境與 DataPolicy 的使用情境，並且將比對結果相同 DataPolicy 所屬的 SQL String 全數列出。系統再根據列出的 SQL String 從委外的子資料庫中整合出完整資料庫的資料。在此完整資料庫的資料並非全數都是使用者所需，必須在查詢改寫後經由使用者的查詢條件挑選後才是使用者所需的資料。

Request(?r) ∧ hasCondition(?r, ?c) ∧ Condition(?c) ∧ DataPolicy(?dp) ∧hasCondition(?dp, ?dc) ∧ Condition(?dc) ∧ sameAs(?c, ?dc) ∧ hasSQL(?dp, ?s) → sqwrl:select(?s)

此外，不論使用者被授權使用 SBQ 或是 PBQ，DHP 只會根據使用者的使用情境

挑選該被整合的委外資料，SBQ 或 PBQ 的資料揭露限制會在 DRP 中處理。

在既有資料委外的研究[4]，資料整合的處理不能於資料庫服務的伺服器中進行，以免資料庫服務提供者藉機窺視資料。如圖三所示，本研究中資料是儲存在雲端資料庫服務中，而資料查詢介面是由資料控制者如企業所建置。由於企業會被允許觀視資料，因此整合資料的處理會在資料查詢介面中進行，也就是資料並不會在委外資料庫中被整合，而是直到資料到達資料查詢介面的 Server 後才會進行，經由資料查詢和 DRP 的推論後才將資料送往使用者所在的 Client 進行揭露。

4.3.4 資料查詢處理

本研究在提供查詢服務的介面上是採用本體論 Schema，屬於語意式的查詢。運用本體論 Schema 管理資料庫會具有降低管理複雜度、為了特殊用途而改變資料庫內容與資料整合三種優勢[27]，而本研究則是為了透過本體論 Schema 定義資料屬性和資料類型，進而判斷哪些資料可以辨別資料擁有者的身分以及可處理該資料的 SDC。但是本體論和資料庫之間的語意層級並不相同，因此要讓本體論 Schema 的查詢可以下到資料庫必須要經過適當的轉換。在本研究中，本體論 Schema 並非是直接對應到委外的子資料庫，而是對應到完整的資料庫，這是因為 DHP 已經負責將子資料庫的資料整合為完整的資料庫資料，因此在本研究中只需處理本體論 Schema 和資料庫 Schema 之間的對應，而不用在考量資料委外還原的部分，如下圖所示：

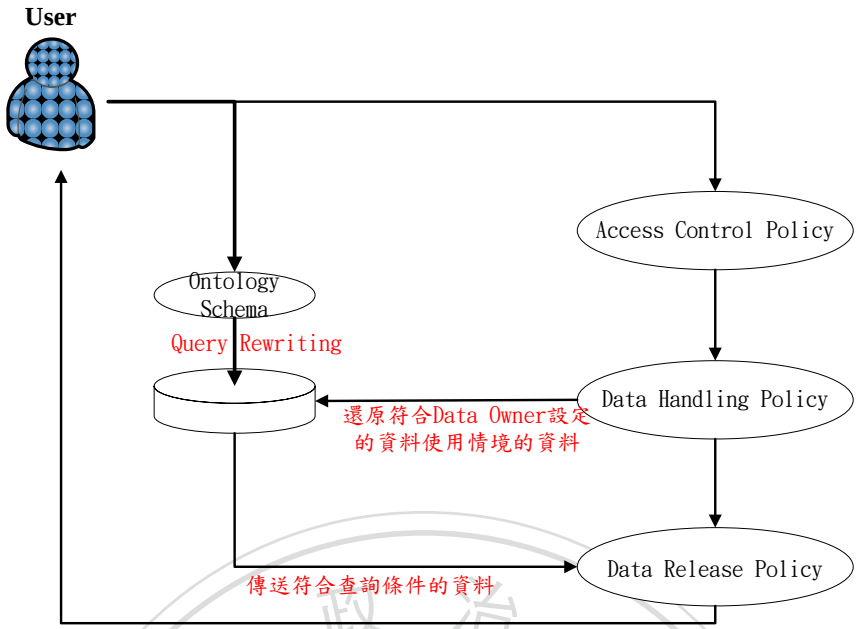

圖 八、資料整合與查詢流程

根據 Diego 的研究[15]，是以 Description Logic 作為整合的語言，並將資料本身放在下層的資料庫中，為了讓本體論 Schema 和資料庫 Schema 之間能相互對應，必須經過兩個步驟：

1. 對應(Mapping)：該研究是以 OWL2 作為上層整合的語言，底下則是使用資料庫作為資料載體，並使用 GLAV 作為對應方式也就是 GAV 與 LAV 的一般化情形。
2. 查詢改寫(Query Rewriting)： 由於 OWL2 的表達能力對於本體論中會出現一些隱含性的事實(Inferred)，一般來說隱含性的事實雖然可以使用推論引擎如 Pellet 推論出來，但是在查詢改寫的當下必須要即時的判斷是否有隱含性的事實，該事實必須要一併的被找出來，目前則是尚未支援，因此必須要採取的是手動性的查詢改寫。

由於該研究是採用 GLAV 的對應方式，因此在查詢改寫必須分成 GAV 和 LAV 的部分進行改寫。GAV 的部分只需將該 View 直接進行展開(Unfolding)的動作即可，而 LAV 的部分則再需使用其他的演算法進行處理，如 Minicon Algorithm、Bucket Algorithm。而在本研究架構中只處理單一本體論 Schema 和單一完整資料

庫 Schema 之間的對應，而資料庫數量也不會有變動性，因此本研究只採用 GAV 對應來完成查詢改寫。對應的方式如左：“Name(?name)→Select Name From Patient_db”。在完成查詢改寫後，便能依據使用者的查詢條件去挑選使用者所需的資料，最後將資料經由 DRP 處理後揭露。

4.3.5 Data Releasing Policy

DRP 負責處理資料的揭露，避免向使用者揭露過多的資料。假如使用者被 ACP 授權 SBQ 則 DRP 會將使用者查詢的資料全數揭露；但假如使用者被 ACP 授權使用 PBQ，則 DRP 會判斷使用者所查詢的資料是否可能侵害資料擁有者的隱私，如果不會則直接揭露資料，但如果會則讓使用者從可辨別資料擁有者身分的資料中挑選欄位，使剩下的欄位不能辨別資料擁有者身分，並將使用者挑選的欄位以使用者使用者挑選的 SDC 方式處理後才全數揭露資料。因此，DRP 的重點在於判斷哪些資料可以辨別資料擁有者的身分以及使用者如何挑選 SDC 的方式處理資料。

文獻[5]中列出四種資料屬性(Data Attribute)：Identifiers、Quasi-identifiers、Confidential 和 Non-confidential。Identifiers 是指能辨別資料擁有者身分的單一欄位，如 ID 或 SSN。Quasi-identifiers 是指能辨別資料擁有者身分的欄位組合，如 Gender、ZIP 和 DoB 等。Quasi-identifiers 並不百分之百都能辨別資料擁有者的身分，只能提升辨別資料擁有者的身分的機率。Confidential 是指敏感性資料，如 Disease 或 Cholesterol。Non-confidential 則是並不屬於上述三類的資料。在本研究中，將 Identifiers 和 Quasi-identifiers 這兩類的資料列為能夠辨別資料擁有者身分的資料。本研究的 DRP 將描述資料與其資料屬性的關係，藉此判斷使用者查詢哪些資料時會辨別資料擁有者身分或侵犯隱私。

SDC 則可分為兩種：Masking 和 Synthetic[5][25][26]。Masking 會將資料做

修改或隱藏的轉換，而 Synthetic 則會將資料轉換成具有統計特性的資料。Masking 又可分為 Non-perturbative 和 Perturbative。Non-perturbative 並不會修改資料內容但會藉由隱藏資料的方式來保護隱私，如 Local Suppression 或 Top-Coding；而 Perturbative 則會以資料修改的方式來保護隱私，如 Lossy Compression。SDC 分類圖如下：

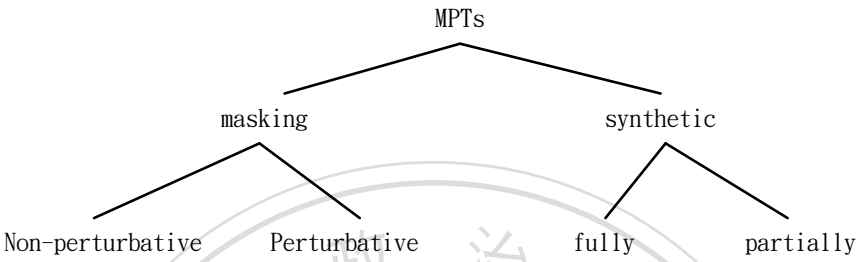

圖 九、Statistics Disclosure Control 的分類

每個 SDC 的處理方式都不相同，如 Local Suppression 能將資料直接遮罩、Generalization 是藉由將資料分類的方式，將同一類的資料以相同替代資料表示；而 Top-Coding 是將超出某個特定範圍的資料以＞n 來表示，因此資料至少必須要是可以進行排序。由上述可知，不同 SDC 能夠處理的資料會依據其資料型態(Data Type)不同而不同。資料型態共有兩種：Continuous 和 Categorical[5]。Continuous 是指可以數學運算處理的資料，如加減乘除等，像是 Cholesterol 和 Cost 等，而 Categorical 則反之，像是 Name 和 Disease 等。舉例而言，Top-Coding 只能處理 Continuous 的資料，而 Local Suppression 則可處理 Continuous 和 Categorical 這兩類的資料。雖然 SDC 方法眾多，但每一種 SDC 都只能用於單一資料型態或者兩者皆可，為了簡化，本研究採用的 SDC 只集中於 Non-perturbative，並且 DRP 會描述資料型態跟 SDC 之間的關係讓使用者挑選可以使用的 SDC 去處理可辨別資料擁有者身分的資料後才揭露。

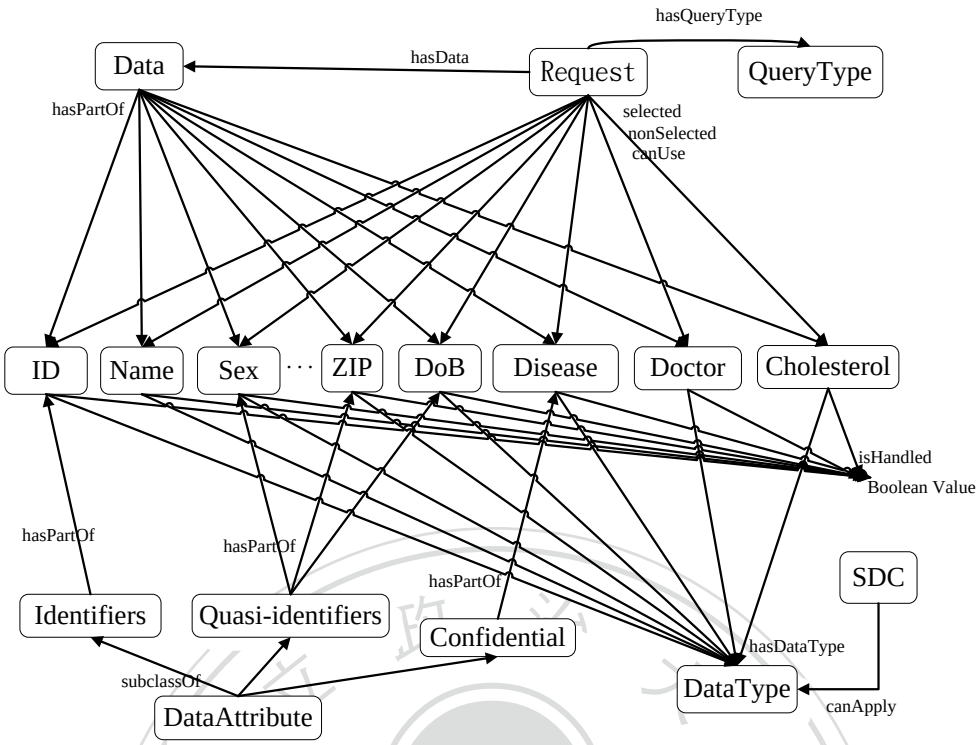

圖 十、DRP 的本體論設計

完整的 DRP 設計如上所示，DataAttribute 和 DataType 分別代表資料屬性和資料型態的類別，但兩者意義不同。DataAttribute 是一種結構的概念，像是 Quasi-identifiers 中每一個實體都是一個構成 Quasi-identifiers 的情況，因此用 hasPartOf 關連到資料，而 DataType 則有兩個實體：Continuous 和 Categorical，每一種資料都以 hasDataType 關聯到自己擁有的資料型態。SDC 和 DataType 之間的關係則以 canApply 表示。

而 SWRL 推論的規則流程分為三階段，各階段有不同的規則負責推論。第一階段：判斷使用者為 SBQ 或 PBQ，假如為 SBQ 則直接將資料揭露給使用者，而 PBQ 則判斷是否侵害隱私，並進入下一階段。其規則有兩條，分別判斷 SBQ 和 PBQ 的情況，如下所示：

Request(?r) ∧ hasData(?r, ?d) ∧ hasPartOf(?d, ?pod) ∧ hasQueryType(?r, Sub) → canUse(?r, ?pod)

和

Request(?r) ∧ hasData(?r, ?d) ∧ hasPartOf(?d, ?pod) ∧ hasQueryType(?r, Pat) ∧ sqwrl:makeSet(?rs, ?pod) ∧ sqwrl:groupBy(?rs, ?r) ∧ Quasi-identifiers (?qui) ∧ hasPartOf(?qui, ?qpod) ∧ sqwrl:makeSet(?qs, ?qpod) ∧ sqwrl:groupBy(?qs, ?qui) ∧ sqwrl:contains(?rs, ?qs) ∧ Confidential (?c) ∧ hasPartOf(?c, ?dc) → sqwrl:selectDistinct(?qui, ?qpod)

如上所述，第一階段第一條規則負責推論 SBQ 的情況下使用者可以使用完整的資料；而第二條規則則去推論 PBQ 的使用者所查詢的資料是否會侵害隱私。第二條規則會先用 sqwrl:contains 這個 SQWRL 提供的 Atom 來判斷使用者所查詢的資料是否包含構成 Quasi-identifiers 的資料，是一種結構包含的判斷，只有當構成 Quasi-identifiers 的資料全都被包含時，才能代表使用者所查詢的資料可以辨別資料擁有者的身分。單純辨別資料擁有者的身分並不一定侵犯隱私，像是只檢視 ID 或 Gender、ZIP 和 DoB 這個欄位組合，雖然可以辨別資料擁有者的身分，但缺乏 Confidential 類型的資料，並不能侵犯隱私。因此第二條規則除了判斷是否能辨別資料擁有者的身分外，還考量使用者是否查詢 Confidential 類型的資料。第二條規則最後會列出滿足的 Quasi-identifiers 以及其組成資料。假如推論出有滿足 Quasi-identifiers 以及其組成資料，則系統列出滿足的 Quasi-identifiers 以及其組成資料讓使用者從挑選任一欄位進行 SDC 的處理，使用者挑選後系統會設定 selected 關係，代表使用者和所挑選的欄位。

第二階段：根據第一階段的判斷，假如 PBQ 的使用者沒有侵犯隱私，則直接揭露資料，但如果侵犯隱私則讓使用者挑選要用何種 SDC 去處理使用者所挑選的欄位，而最後揭露的資料便不可辨別資料擁有者身分或侵犯隱私。規則如下所示：

Request(?r) ∧ hasData(?r, ?d) ∧ hasPartOf(?d, ?pod) ∧ hasQueryType(?r, Pat) →

canUse(?r, ?pod)

和

Request(?r) ∧ hasData(?r, ?d) ∧ hasPartOf(?d, ?b) ∧ selected(?r, ?b) ∧ hasDataType(?b, ?tp) ∧ DataType(?tp) ∧ SDC(?sdc) ∧ canApply(?sdc, ?tp) → sqwrl:select(?b, ?sdc)

第二階段第一條規則代表如果 PBQ 的使用者並未侵犯隱私，則直接揭露資料予使用者。第二階段第二條規則根據資料型態和 SDC 的對應關係，列出可處理使用者所挑選欄位的 SDC，並且根據使用者的挑選呼叫 SDC 處理資料。

第三階段：判斷是否 SDC 已經處理完畢，如果是則將所有資料一併揭露，意即使用者所挑選侵害隱私的欄位以 SDC 處理後揭露和使用者沒挑選的欄位直接揭露。規則如下：

Request(?r) ∧ hasData(?r, ?d) ∧ hasPartOf(?d, ?b) ∧ selected(?r, ?b) ∧ hasPartOf(?d, ?a) ∧ isHandled(?b, 1) ∧ notSelected(?r, ?a)
→ canUse(?r, ?b) ∧ canUse(?r, ?a)

上述規則運用 isHandled 來判斷使用者挑選的欄位是否處理完畢。

以使用者查詢為例來講解 DRP 的處理流程，當有一個被授權使用 PBQ 的使用者想要查詢 Gender、ZIP、DoB、Disease 和 Doctor 欄位時， Gender、ZIP 和 DoB 屬於一個 Quasi-identifiers，再加上使用者同時查詢敏感性欄位 Disease，因此以 DRP 規則推論得知使用者侵害隱私，故要求使用者從 Gender、ZIP 和 DoB 三個欄位中取一個欄位以 SDC 處理。使用者可根據自身的需求取捨哪個欄位，假如挑選 ZIP，則 DRP 會再次查詢可以處理使用者挑選欄位的 SDC 如 Local Suppression 和 Generalization 等，並供使用者挑選。當使用者挑選 SDC 後，DRP 便呼叫該 SDC 處理資料，並於資料處理後揭露資料於使用者。如下圖所示：

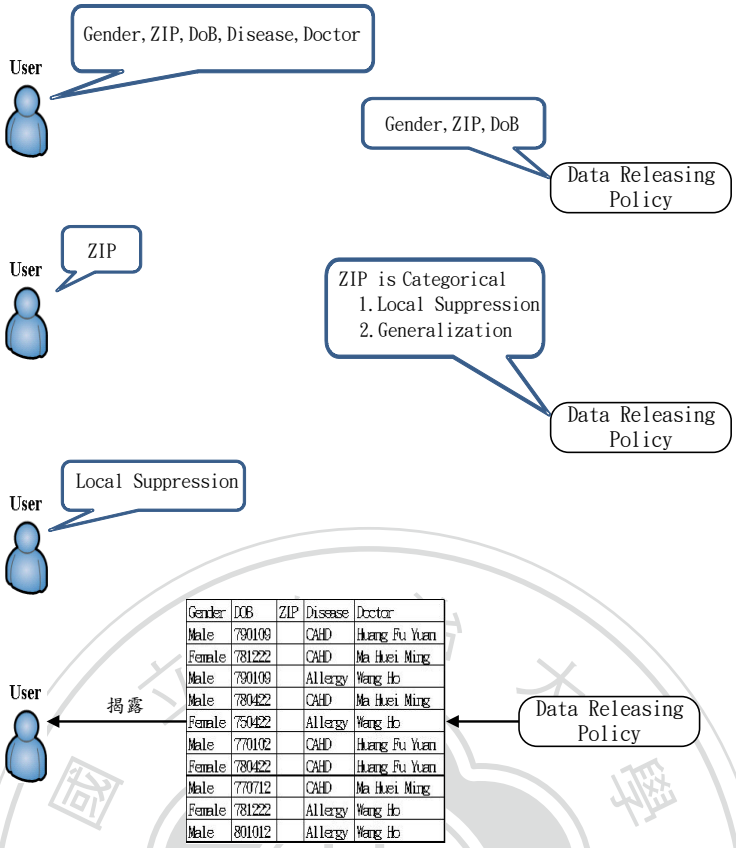

圖 十一、DRP 處理範例流程

此外，本研究將要比較資料委外的隱私保護與資料使用的隱私保護概念，兩者的詞彙對應關係如下表：

表 5、資料委外與資料使用詞彙對應

資料委外	資料使用
單一欄位違反隱私	Identifiers
多個欄位違反隱私	Quasi-identifiers
敏感性欄位	Confidential attribute

由上表可知，資料委外與資料使用的概念其實相通。資料委外著重於讓雲端資料庫服務提供者無法辨識資料擁有者的身分，因此針對可辨別資料擁有者的身分的資料進行處理；而資料使用時則著重於讓使用者無法侵犯資料擁有者的隱私，因此也是對於可辨別資料擁有者的身分的資料進行處理，使其不能辨別資料擁有者的身分，便無法侵犯隱私。

第5章

模式驗證與系統展示

5.1 規範推論的驗證

本研究的研究方法是運用 OWL DL 和 SWRL 以及 SQWRL 來塑模 ACP、DHP 和 DRP 三種規範提供 SBQ 和 PBQ，進而達成委外資料的整合以及資料使用的隱私保護。本研究為了驗證規範推論的正確性，採用的工具為 Protégé。Protégé 是 Stanford 大學所開發的本體論編輯軟體。目前有數千個使用者社群。Protégé OWL Plugin 是 Protégé 提供發展擴充技術，像是大量的 plugin 來支援以編輯 OWL 本體論，並且可客製化圖形化工具集，可以不同的格式載入及儲存 OWL 的檔案。而 SWRL 規則的推論需要使用 Jess 推理引擎來推理 SWRL 規則出結果。Protégé 用來執行 SWRL 規則所提供的 plugin 為 SWRL Tab，它會將 SWRL 規則轉換為 Jess 推理引擎可處理的格式並且將推論結果轉回 OWL，完成推論。由上述可知，Protégé 是支持 OWL DL 和 SWRL 以及 SQWRL 語意塑模和推論上的驗證，因此本研究採用其作為規範建置和推論的工具。

本研究使用 Protégé 3.4.8 工具，根據前一章所建構的三種規範的本體論來設定類別與屬性。在建構好 ACP、DHP 和 DRP 後，本研究接著利用 Protégé plugin 的 SWRL Tab 的功能來完成建置 SWRL 和 SQWRL 規則，如先前第四章所提之 SWRL 和 SQWRL 規則。

ACP、DHP 和 DRP 規則推論結果畫面分別如下：

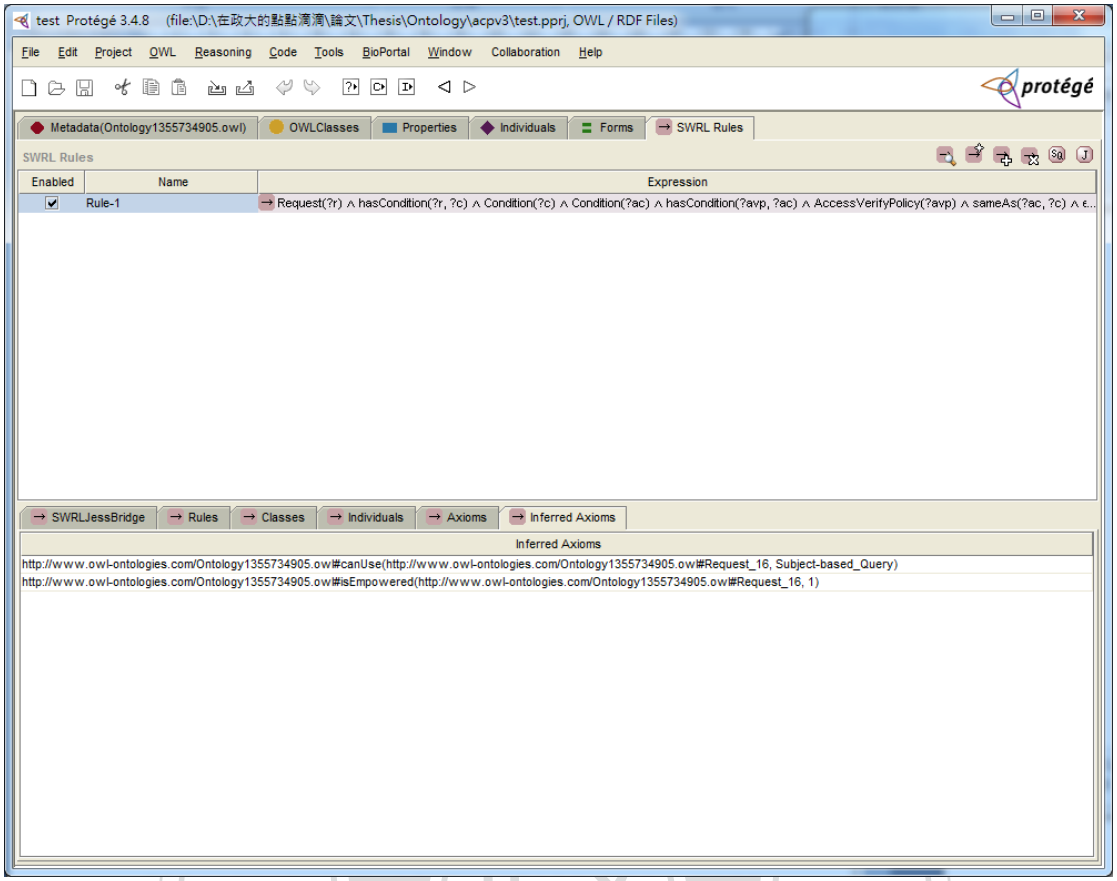

圖 十二、Protégé SWRL Tab 推論 ACP 的規則畫面

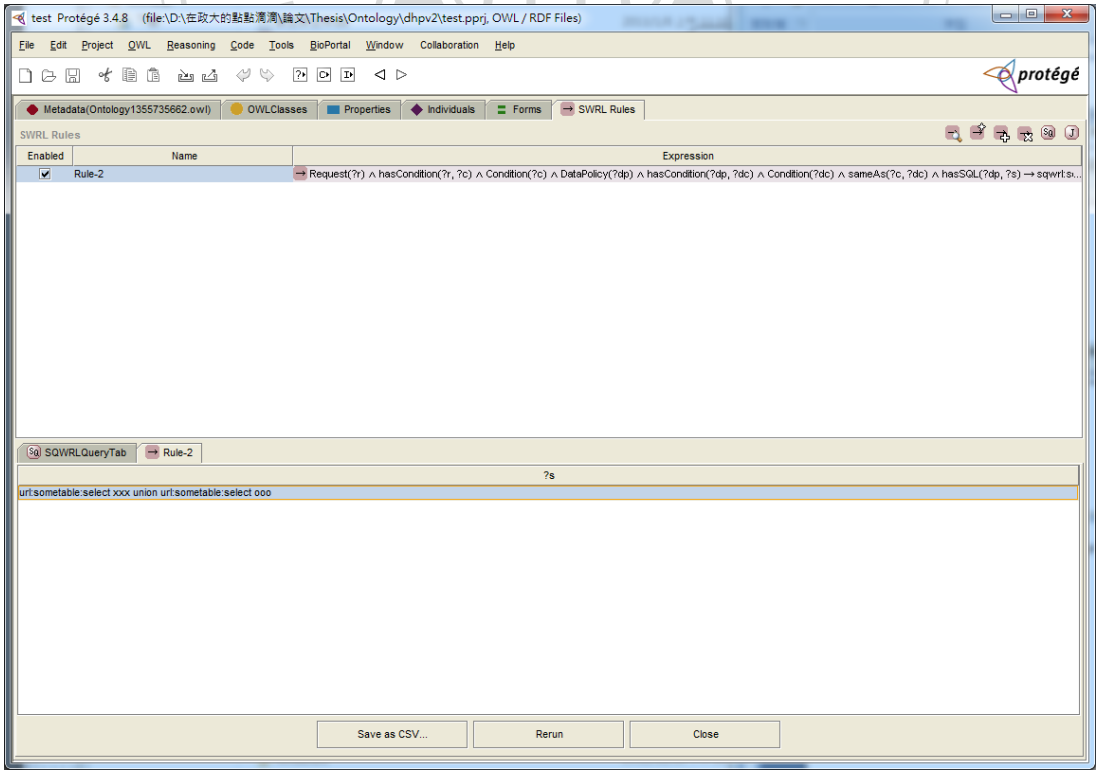

圖 十三、Protégé SWRL Tab 推論 DHP 的規則畫面

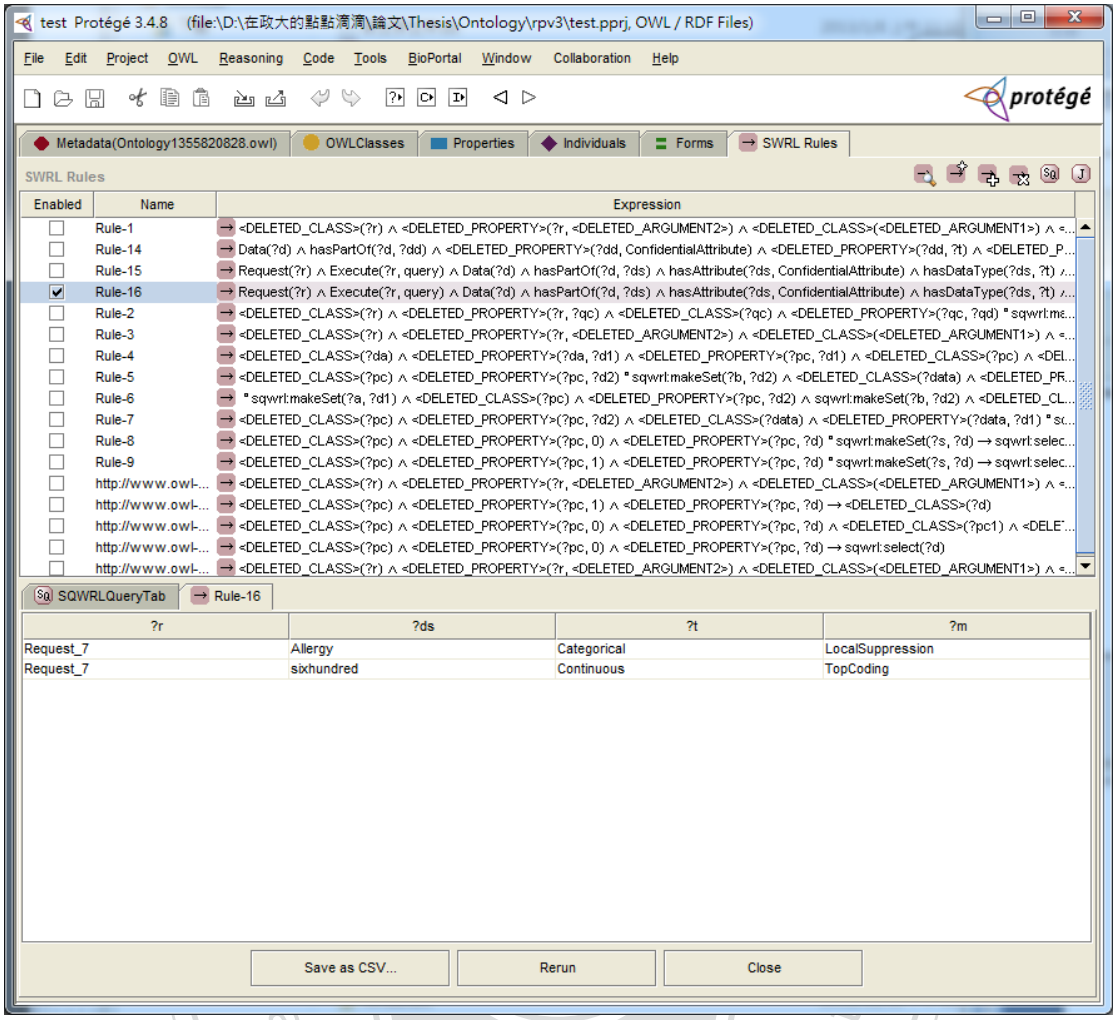

圖 十四、Protégé SWRL Tab 推論 DRP 的規則畫面

5.2 系統展示

本研究的研究架構分為兩個步驟：第一在資料委外的時候必經先將資料以加密結合分割的方式處理，並且記錄 IndexKey 的關係。第二是在 Protégé 設定好 ACP、DHP 和 DRP 三種規範和相對應的 SWRL 或 SQWRL 規則，並且由人進行手動推論。由於本研究所設計的系統需實現動態推論，因此採用 Protégé 所提供的

protégé-owl api，來實現動態編輯 OWL DL 本體論和 SWRL 或 SQWRL 規則推

論。

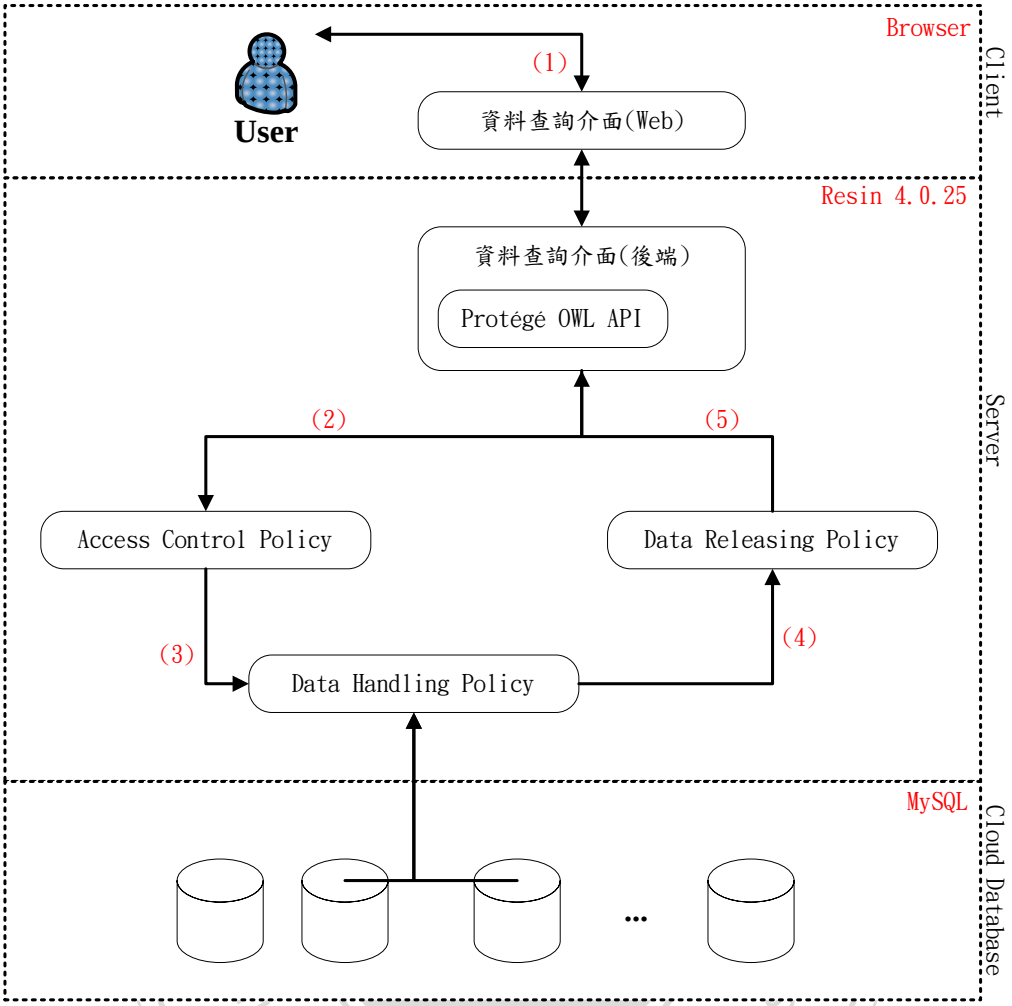

圖 十五、系統實作架構圖

本研究的系統實作架構圖如上所示。protégé-owl api 是以 Java 語言撰寫的函式庫，為了系統整合方便本研究採用 JSP 語言撰寫資料查詢介面和後端，以調用 protégé-owl api 執行動態編輯 ACP、DHP 和 DRP 本體論以及落實 SWRL 或 SQWRL 規則推論。整個系統執行的流程便如同第四章所述：一開始，ACP 會去驗證進入資料查詢介面的使用者，在驗證通過後便授權該使用者可使用的查詢模式，接著啟動 DHP 整合使用者可使用的資料和資料查詢處理步驟，並且將使用者所需的資料傳遞給 DRP，DRP 在最後執行資料揭露的動作。當一個 PBQ 使用者所查詢的資料會侵害隱私時，系統執行圖如下所示：

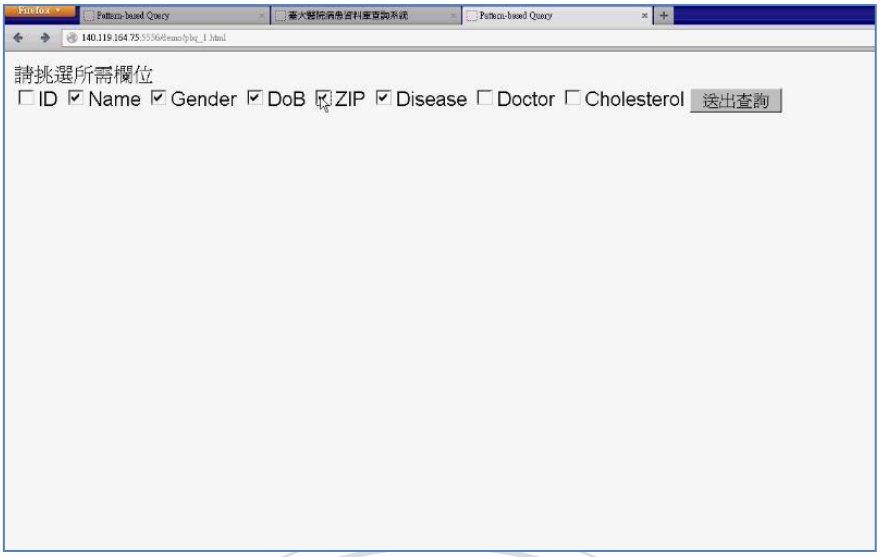

圖 十六、PBQ 使用者勾選所需資料頁面

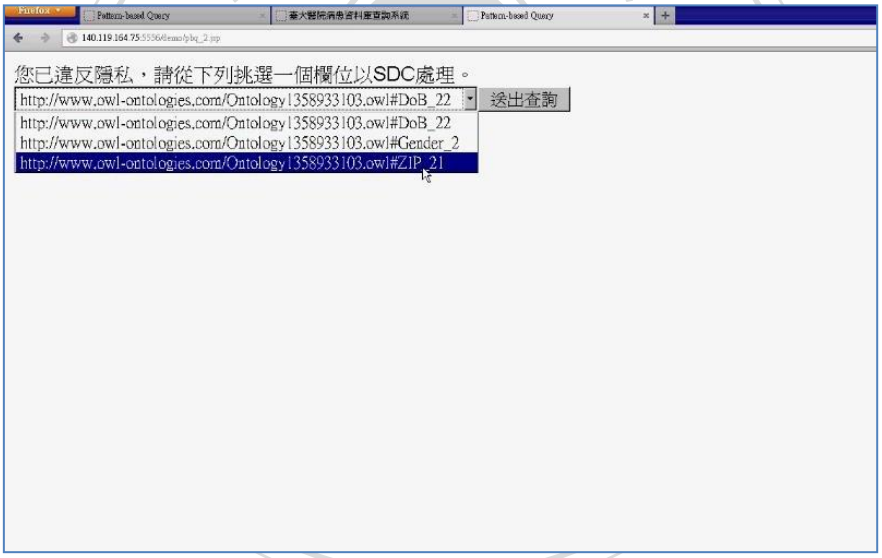

圖 十七、侵害隱私的 PBQ 使用者需從 Quasi-identifiers 的組成元素中挑選一個欄位

圖 十八、侵害隱私的 PBQ 使用者需挑選 SDC 去處理上一步驟所挑選的欄位

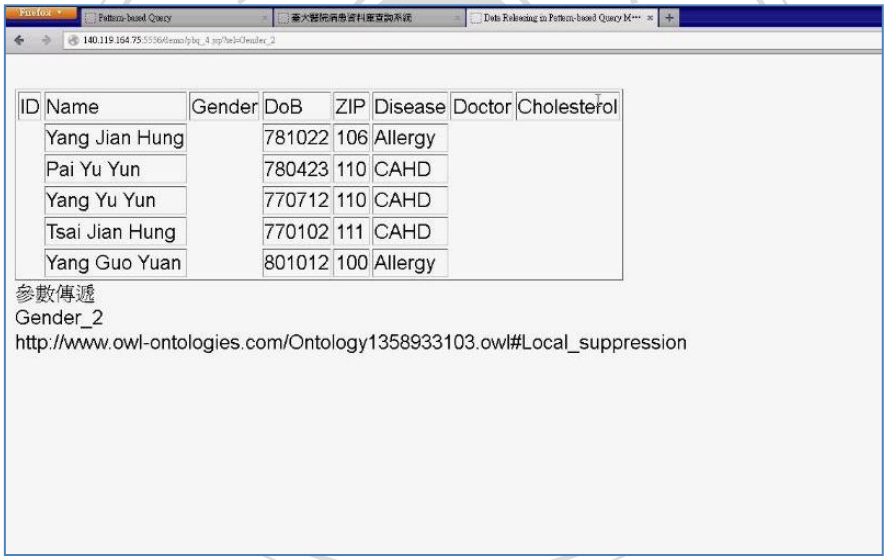

圖 十九、最後揭露給 PBQ 使用者的資料

第6章

結論與未來展望

本研究在資料委外前先行以加密結合分割的方式處理資料以確保資料委外的隱私保護，並以本體論和規則語言設計 ACP、DHP 和 DRP 這三個規範。ACP 負責驗證使用者以及授權使用者能使用的查詢模式，DHP 則將符合資料擁有者的隱私偏好使用情境的委外資料從雲端資料庫服務中整合出來，最後 DRP 負責最後進行對使用者的資料揭露。透過這三個規範合作、分工來提供 SBQ 和 PBQ 確保資料使用上的隱私保護。

而本研究中只有考量到結構化單一資料源的 Microdata 揭露一次時的隱私保護，並沒有考量使用者多次查詢不同欄位資料，接著再藉由比對資料確認資料擁有者身分來侵害隱私的情況。同時對於結構化多資料源的 Microdata 揭露以及 Macrodata 揭露時該落實的保護也是本研究沒有探討的部分，前者像是透過多資料源的不同欄位資料的比對，可以辨別一個人的身分[6]進而侵犯隱私，而後者像是多次查詢 Macrodata 等。如何運用本體論和規則語言來塑模 Audit Log 針對上述情況進行資料使用上的保護，還需要進一步的探討。此外，本研究是探討在關聯式資料庫的委外隱私保護，但資料的儲存方式正在轉往語意格式儲存，如歐洲等國正逐步推動以 RDF (Resource Description Framework)格式儲存資料，如何確保委外的 RDF 格式資料的隱私也是未來需要探討的部分。

參考文獻

[1] H. Hakan, "Providing Database as a Service," 2002, pp. 0029-0029.

[2] M. Armbrust, A. Fox, et al., "Above the Clouds: A Berkeley View of Cloud Computing," EECS Department, University of California, Berkeley UCB/EECS-2009-28, February 10 2009.

[3] H. Takabi, J. B. D. Joshi, et al., "Security and Privacy Challenges in Cloud Computing Environments," IEEE Security and Privacy, vol. 8, pp. 24-31, 2010.

[4] P. Samarati and S. D. C. d. Vimercati, "Data protection in outsourcing scenarios: issues and directions," Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security, Beijing, China, 2010.

[5] V. Ciriani, S. Capitani di Vimercati, et al., "Microdata Protection," in Secure Data Management in Decentralized Systems. vol. 33, 2007, pp. 291-321.

[6] L. Sweeney, "k-anonymity: a model for protecting privacy," Int. J. Uncertain. Fuzziness Knowl.-Based Syst., vol. 10, pp. 557-570, 2002.

[7] R. Popp and J. Poindexter, "Countering Terrorism through Information and Privacy Protection Technologies," IEEE Security and Privacy, vol. 4, pp. 18-27, 2006.

[8] OpenTC. Available: http://www.opentc.net/

[9] S. Cabuk, C. I. Dalton, et al., "Towards automated security policy enforcement in multi-tenant virtual data centers," Journal of Computer Security, vol. 18, pp. 89-121, 2010.

[10] S. Berger, R. C, et al., "Security for the cloud infrastructure: trusted virtual data center implementation," IBM J. Res. Dev., vol. 53, pp. 560-571, 2009.

[11] California Senate Bill SB 1386, 2002.

[12] SWRL: A Semantic Web Rule Language Combining OWL and RuleML. Available: http://www.w3.org/Submission/SWRL/

[13] Connor, M. O. and A. Das (2009). "SQWRL: a Query Language for OWL." Proceedings of the 6th International Workshop on OWL: Experiences and Directions (OWLED 2009).

[14] D. Calvanese and G. D. Giacomo, "Data integration: a logic-based perspective," AI Mag., vol. 26, pp. 59-70, 2005.

[15] D. Calvanese, G. Giacomo, et al., "Data Integration through DL-Lite A Ontologies," in Semantics in Data and Knowledge Bases, 2008, pp. 26-47.

[16] D. Calvanese, G. Giacomo, et al., "Using OWL in Data Integration," in Semantic Web Information Management, 2010, pp. 397-424.

[17] A. Y. Levy, A. Rajaraman, et al., "Querying Heterogeneous Information Sources Using Source Descriptions," Proceedings of the 22th International Conference on Very Large Data Bases, 1996.

[18] C. A. Ardagna, M. Cremonini, et al., "A privacy-aware access control system," J. Comput. Secur., vol. 16, pp. 369-397, 2008.

[19] C. A. Ardagna, J. Camenisch, et al., "Exploiting cryptography for privacy-enhanced access control: A result of the PRIME Project," J. Comput. Secur., vol. 18, pp. 123-160, 2010.

[20] The Enterprise Privacy Authorization Language(EPAL). Available: http://www.w3.org/2003/p3p-ws/pp/ibm3.html

[21] S. De Capitani di Vimercati and S. Foresti, "Privacy of Outsourced Data Privacy and Identity Management for Life." vol. 320, 2010, pp. 174-187.

[22] V. Ciriani, S. De Capitani di Vimercati, et al., "Keep a Few: Outsourcing Data While Maintaining Confidentiality Computer Security – ESORICS 2009." vol.

5789, 2009, pp. 440-455.

[23] V. Ciriani, S. D. C. D. Vimercati, et al., "Combining fragmentation and encryption to protect privacy in data storage," ACM Trans. Inf. Syst. Secur., vol. 13, pp. 1-33, 2010.

[24] S. Ceri, G. Gottlob, et al., "What You Always Wanted to Know About Datalog (And Never Dared to Ask)," IEEE Trans. on Knowl. and Data Eng., vol. 1, pp. 146-166, 1989.

[25] N. R. Adam and J. C. Worthmann, "Security-control methods for statistical databases: a comparative study," ACM Comput. Surv., vol. 21, pp. 515-556, 1989.

[26] J. Mateo-Sanz, A. Martínez-Ballesté, et al., "Fast Generation of Accurate Synthetic Microdata," in Privacy in Statistical Databases. vol. 3050, 2004, pp. 298-306.

[27] M. Lenzerini, "Ontology-based data management," Proceedings of the 20th ACM international conference on Information and knowledge management, Glasgow, Scotland, UK, 2011.