

內部稽核範例

—非法旅宿查核作業

壹、前言

在大數據時代，透過網路此一巨大資料庫蒐集資料，進而分析作為決策參據已是當今趨勢，其中網路爬蟲(Web crawler)係運用網路自動化抓取(Web scraping)技術，抓取所需網頁之數筆資料，包括文字內容、圖片等，抓取者可藉由取得之資料做後續的資料分析及加值應用。本範例係協助機關內部稽核人員，運用網路爬蟲技術自網站抓取旅宿業者資料，結合行政院主計總處開發之「通用稽核模組」^{註1}執行資料整理及比對，有效勾稽疑似非法旅宿業者作為稽查作業之參考，提升整體稽核效率。

貳、緣由

臺灣為促進觀光事業，著力於發展在地多元文化特色，包括歷史故事、傳統技藝、種族風俗、農場體驗及生態旅遊等，旅宿業者亦藉著旅遊群聚效應，於觀光勝地周邊拓展住宿商機吸引各地觀光客。為確保旅客住宿安全與消費者權益，以及保障合法旅宿業者之經營權，近年來各地方政府致力於稽查取締非法旅館、民宿及日租套房等旅宿業，而現行稽查方式多為被動接獲民眾檢舉或由稽核人員定期至訂房網站逐筆核對是否屬交通部觀光局臺灣旅宿網之合法旅宿業者，此種稽查方式不僅耗力費時，亦無法適時遏止非法營業，爰運用前開自動抓取等技術工具，針對刊登於各訂房網站之國內旅宿是否合法登記營業，進行大範圍且有效率之清查作業。

參、稽核步驟

為有效清查國內非法旅宿，經運用 Chrome 瀏覽器下載免費 Web Scraper 應用程式，藉由圖像化方式擷取訂房網站相關資料，並將擷取之資料連同交通部觀光局臺灣旅宿網公開之合法旅宿資料，透過「通用稽核模組」整理、比對，以獲取適切之稽核證據，茲將作業步驟說明如下：

註1：「通用稽核模組」操作方法請詳行政院主計總處官網／主要業務／政府內控與內稽／內部控制監督作業範例。

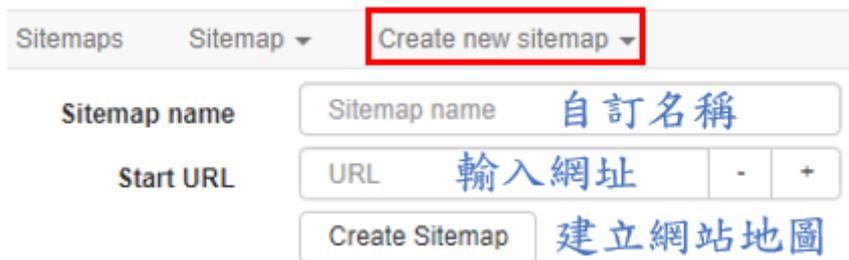
一、瞭解資料：瞭解訂房網站架構、抓取目標（旅宿名稱、地址）、臺灣旅宿網之合法旅宿資料，以及相關網站資料使用限制。

二、抓取資料 (Scrape)

(一) 首先在欲抓取的訂房網站資料頁面按 F12 開啟「開發人員工具」，找到 Web Scraper 應用程式，並將工具置於頁面下方，以利操作。

(二) 建立網站地圖：命名並輸入欲抓取網頁資料之網址後建立網站地圖（圖 1），建立完成後會出現新增新選擇器（Add new selector）選項。

圖 1、建立網站地圖



The screenshot shows the 'Sitemaps' section of the Web Scraper tool. At the top, there is a dropdown menu labeled 'Sitemaps' with a sub-menu 'Sitemap' and a button 'Create new sitemap' which is highlighted with a red box. Below this, there are two input fields: 'Sitemap name' with the text '自訂名稱' and 'Start URL' with the text '輸入網址'. At the bottom, there is a 'Create Sitemap' button with the text '建立網站地圖'.

(三) 依訂房網站網頁呈現方式，分就其設定及選擇欲抓取之目標資料(旅宿名稱、地址)，簡要說明如下：

1. 所有旅宿及目標資料皆呈現於同一頁面（第 1 種型態）

首先點選新增新選擇器（Add new selector），為抓取各旅宿資料，第一層類型（Type）選擇元素（Element），再選取（Select）各旅宿資料區塊，選取後點選完成選取（Done selecting），並勾選多筆數（Multiple）後點選儲存選擇器（Save selector），次為抓取旅宿名稱及地址，接續點選前開已建立之第一層選擇器進入第二層，第二層須分別建立抓取旅宿名稱及地址 2 個選擇器，類型選擇文字（Text），再選取（Select）旅宿名稱（或地址）（圖 2）。設定抓取流程後，可先至選擇器圖解（Selector graph）檢視抓取邏輯是否正確，再點選抓取（Scrape）執行指令（圖 3）。

圖 2、抓取流程示例（第 1 種型態）

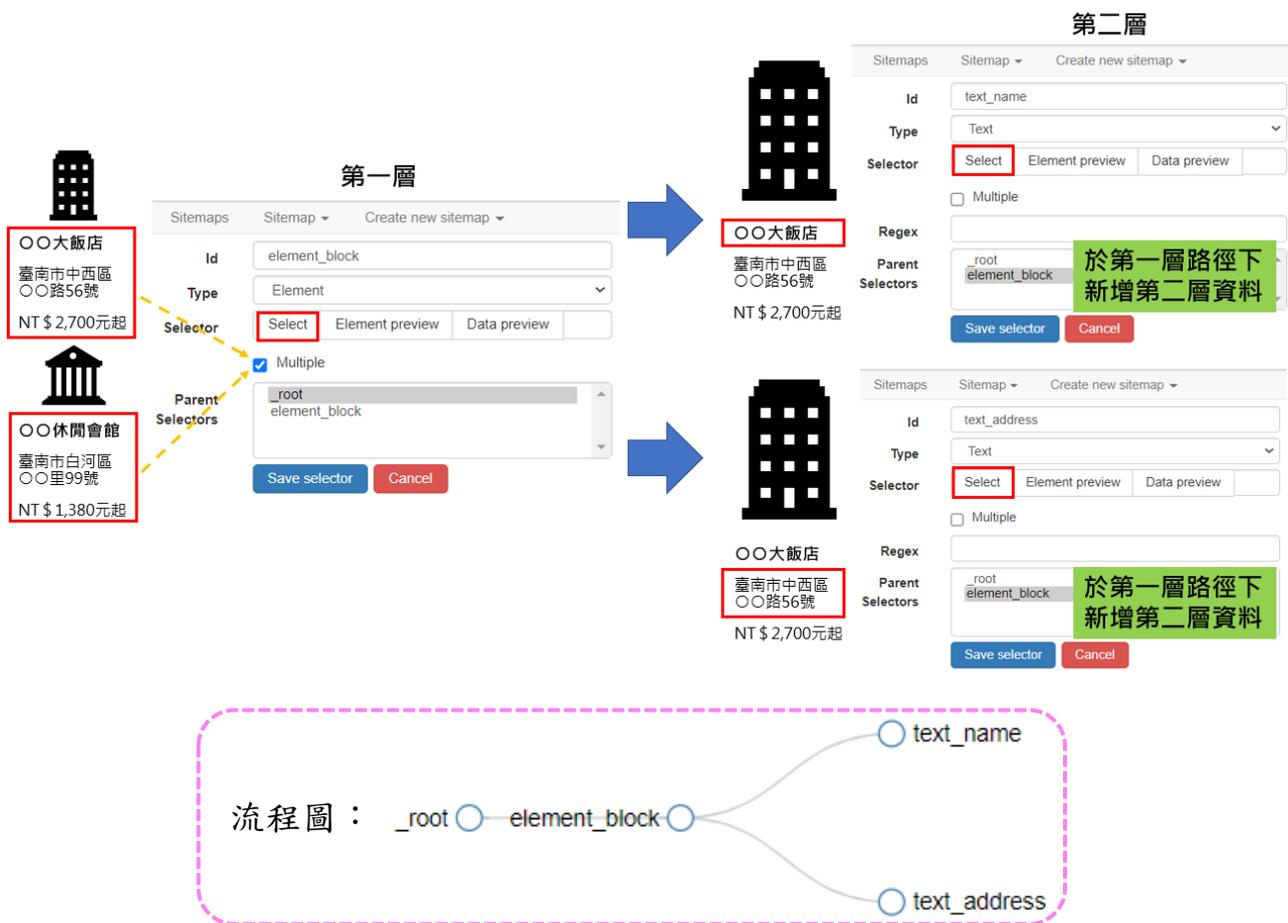
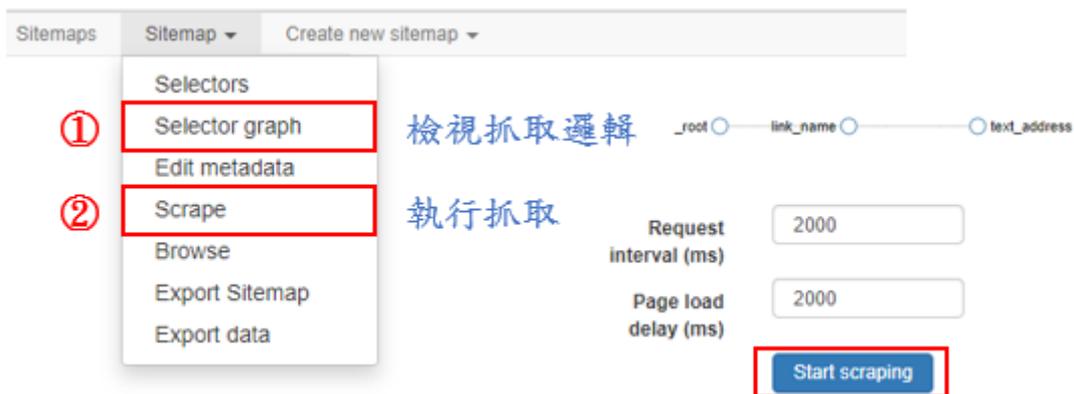


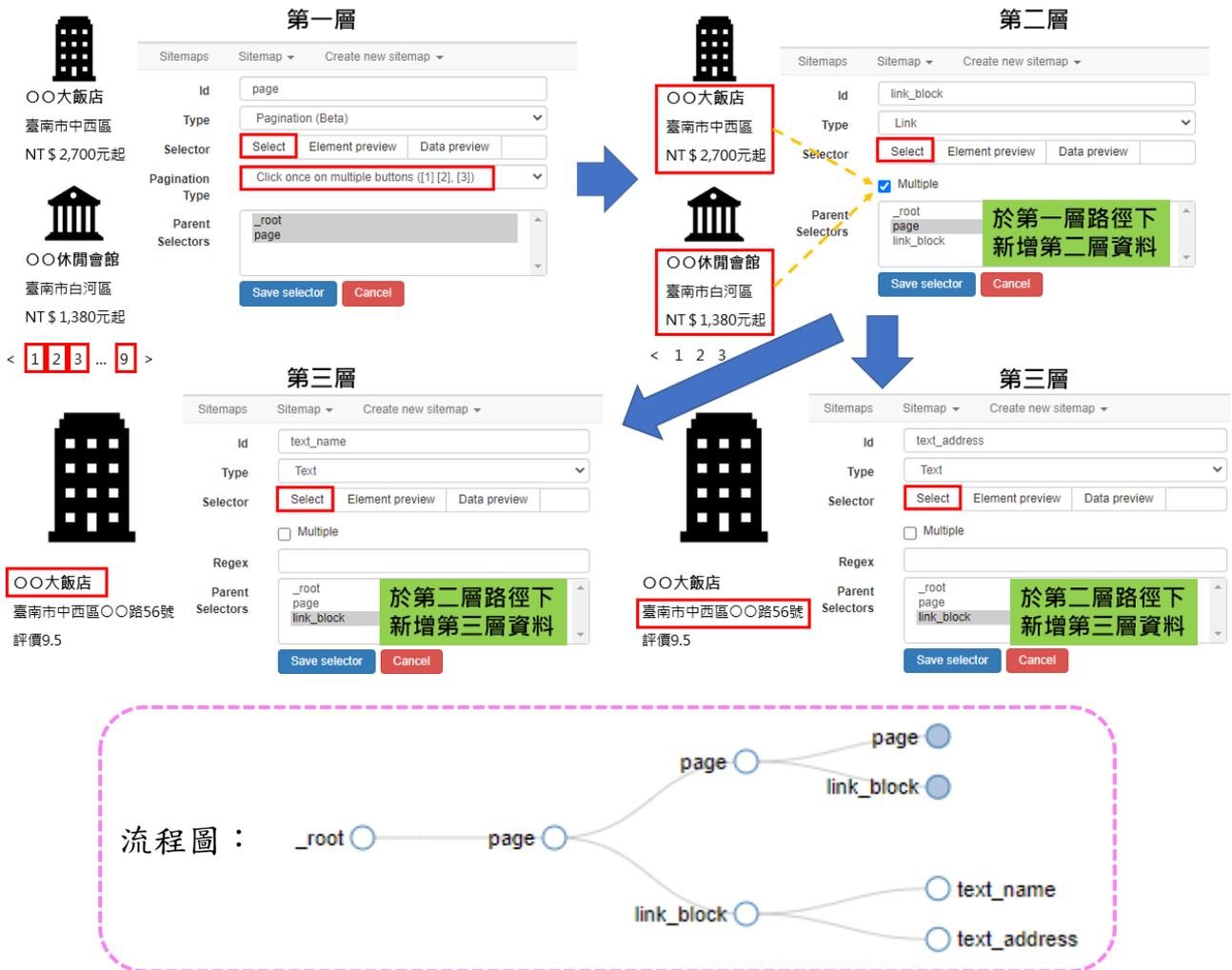
圖 3、檢視邏輯並執行抓取



2. 旅宿資料呈現於不同分頁，目標資料透過連結旅宿資料取得（第 2 種型態）
 首先為抓取各分頁之旅宿資料，第一層類型選擇分頁（Pagination），分頁類型選擇在多個按鈕點擊一次（Click once on multiple buttons）

註²，再選取(Select)各個頁碼按鈕，選取後點選完成選取(Done selecting)，次為抓取各住宿資料，接續點選前開已建立之第一層選擇器進入第二層，第二層類型選擇連結(Link)，再選取(Select)各住宿資料區塊(或住宿名稱)，並勾選多筆數(Multiple)，最後為抓取住宿名稱及地址，點選任一住宿資料區塊(或住宿名稱)進入下一頁面，並點選前開已建立之第二層選擇器進入第三層，第三層類型選擇文字(Text)，再分別建立選擇器選取(Select)住宿名稱及地址(圖4);如果住宿名稱及地址無須透過連結(Link)住宿資料取得，則可參照前述第1種型態設定抓取流程。檢視抓取邏輯及執行指令同圖3。

圖4、抓取流程示例(第2種型態)

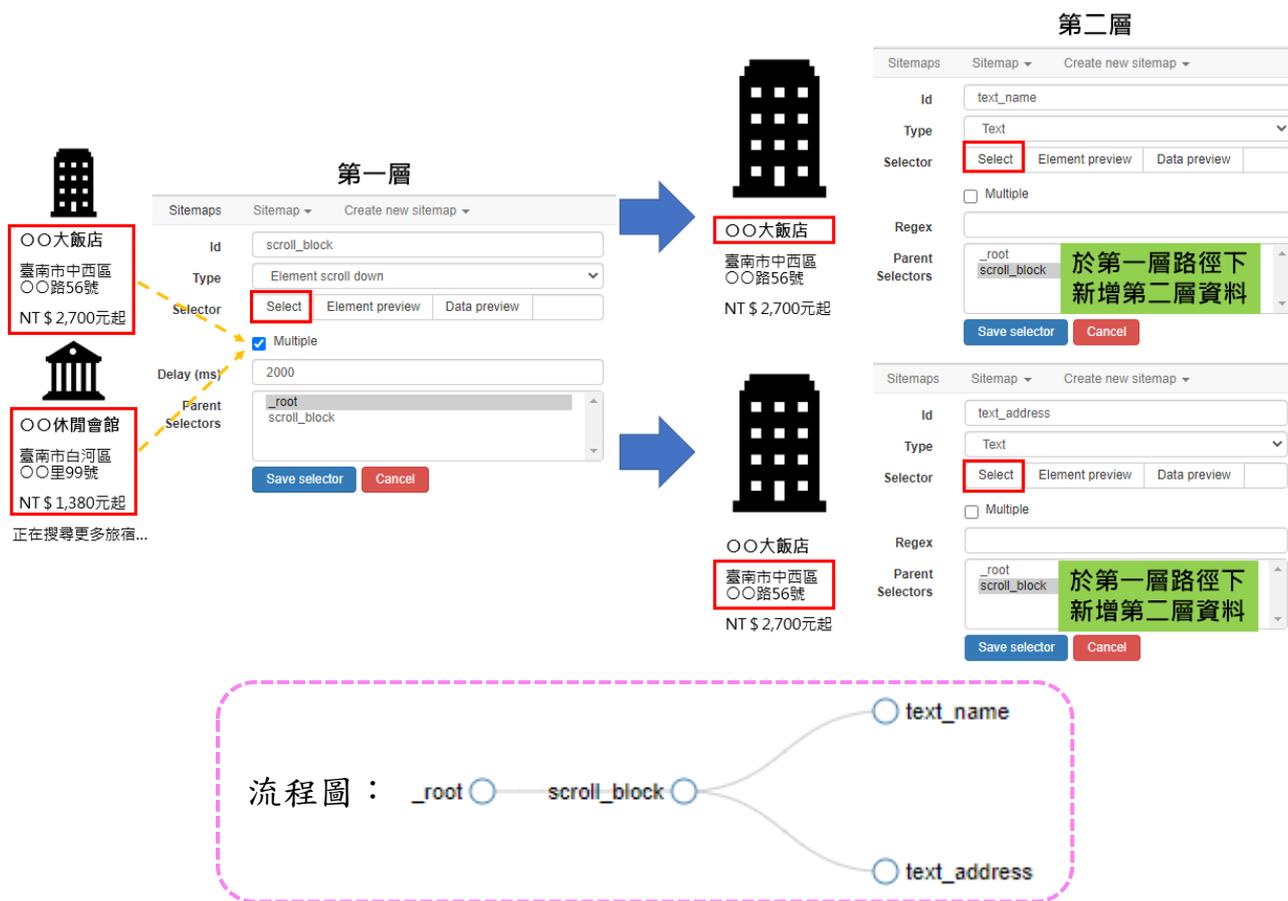


註2：如屬須按下一頁或載入更多按鈕，類型選擇在下一頁/載入更多按鈕點擊多次 (Click multiple times on next/more button)，再選取下一頁或載入更多按鈕。

3. 所有旅宿及目標資料雖呈現於同一頁面，但須向下滾動以瀏覽更多項目（第3種型態）

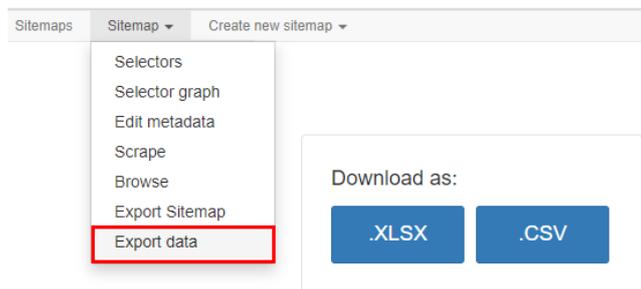
首先為隨著向下滾動頁面以抓取更多項目，第一層類型選擇元素向下滾動（Element scroll down），再選取（Select）各旅宿資料區塊，選取後點選完成選取（Done selecting），並勾選多筆數（Multiple），次為抓取旅宿名稱及地址，接續點選前開已建立之第一層選擇器進入第二層，第二層類型選擇文字（Text），再分別建立選擇器選取（Select）旅宿名稱及地址（圖5）；資料取得如須透過連結（Link）抓取，則可參照前述第2種型態設定抓取流程。檢視抓取邏輯及執行指令同圖3。

圖5、抓取流程示例（第3種型態）



(四) 匯出資料 (Export data): 可選擇以 XLSX 或 CSV 檔案格式匯出資料 (圖 6)。

圖 6、匯出資料



三、整理資料

運用通用稽核模組之「資料整理」功能，將所擷取訂房網站資料及臺灣旅宿網取得之旅宿名稱及地址（市縣別、鄉鎮市區、里、鄰、郵政區號、中英文地址等）進行資料齊一化處理（如表），以利後續資料比對。

表、齊一化處理後資料

旅宿名稱	市縣別	鄉鎮市區	地址
○○大飯店	臺南市	中西區	○○路 1 號
○○休閒會館	臺南市	白河區	○○街 2 號
○○商旅	臺南市	永康區	○○一路 3 號
○○民宿	臺南市	安平區	○○路 4 巷 5 弄 6 號

四、資料比對

運用通用稽核模組之「跨表比對」功能，將整理後之訂房網站及臺灣旅宿網資料進行二階段比對，步驟如下：

- (一) 比對旅宿名稱：將訂房網站資料列主檔，臺灣旅宿網資料列次檔，以旅宿名稱為關聯欄位，點選「有對應成功」，匯出確認兩者資料是否為同一旅宿（圖 7），再點選「未對應成功」，匯出並保留此檔案（圖 8）。

圖 7、比對旅店名稱流程示例 (1)

關聯欄位	主檔	次檔
資料欄位1	G	C
資料欄位2	J	D
資料欄位3	K	F

旅館名稱	市縣	鄉鎮市區	地址	市縣	鄉鎮市區	地址
○○門建	台南	中西區	○○路77	臺南市	中西區	○○路77號
○○大飯	台南	中西區	○○路6號	臺南市	中西區	○○路6號
○○趣	台南	中西區	○○街32	臺南市	中西區	○○街322
○○旅棧	台南	中西區	○○路二	臺南市	中西區	○○里○○
○○大飯	台南	中西區	○○路12	臺南市	中西區	○○路126
○○慢慢	台南	中西區	○○街10	臺南市	中西區	○○街106

圖 8、比對旅店名稱流程示例 (2)

關聯欄位	主檔	次檔
資料欄位1	G	B
資料欄位2	J	
資料欄位3	K	

旅館名稱	市縣	鄉鎮市區	地址
○○起」	台南	中西區	○○街1號
○○美墅	台南	中西區	○○路二段
○○Station	台南	中西區	○○街188
○○夢想	台南	中西區	○○路三段
○○Design	台南	北區	○○路120
○○Door	台南	東區	○○一街3

(二) 比對旅店地址：將前一步驟之「未對應成功」資料列主檔，臺灣旅店網資料維持次檔，以旅店地址為關聯欄位，點選「有對應成功」，匯出確認兩者資料是否為同一旅店(圖 9)，再點選「未對應成功」，匯出疑似非法旅店資料(圖 10)。

圖 9、比對旅店地址流程示例 (1)

主檔	主檔	次檔	次檔
檔案名稱	旅店名稱未對應.xlsx	臺灣旅店網.ods	
工作表名稱	未對應項	工作表1	
關聯欄位	D	F	
資料欄位1	全	B	
資料欄位2		C	
資料欄位3		D	
資料欄位4			
資料欄位5			

A	B	C	D	E	F	G
地址	旅店名稱	市縣	鄉鎮市區	旅店名稱	市縣	鄉鎮市區
○○街200	○○ Song	台南	中西區	○○輕鬆	臺南市	中西區
○○街70	○○ Old F	台南	中西區	○○IN	臺南市	中西區
○○路二	○○ A 聖	台南	中西區	○○大飯	臺南市	中西區
○○街7	○○ Way	台南	中西區	○○ Way	臺南市	中西區
○○大道	○○花園	台南	安南區	○○花園	臺南市	安南區
○○路38	○○ Scho	台南	北區	○○家屋	臺南市	北區

圖 10、比對旅店地址流程示例 (2)

主檔	主檔	次檔	次檔
檔案名稱	旅店名稱未對應.xlsx	臺灣旅店網.ods	
工作表名稱	未對應項	工作表1	
關聯欄位	D	F	
資料欄位1	全		
資料欄位2			
資料欄位3			
資料欄位4			
資料欄位5			

A	B	C	D
地址	旅店名稱	市縣	鄉鎮市區
○○街1號	○○起」	台南	中西區
○○路二	○○美墅	台南	中西區
○○街18	○○ Static	台南	中西區
○○路三	○○夢想	台南	中西區
○○路12	○○ Desig	台南	北區
○○一街	○○ Door	台南	東區

肆、稽核發現與結論

經運用 Web Scraper 及「通用稽核模組」輔助稽核結果發現，國內有多家旅店業者未列於臺灣旅店網之合法登記名單，疑似為非法經營之旅店業者，依據發展觀光條例第 55 條之 1 規定，未依該條例規定領取營業執照或登記證而經營觀光旅館業務、旅行業務、觀光遊樂業務、旅館業務或民宿者，以廣告物、出版品、廣播、電視、電子訊號、電腦網路或其他媒體等，散布、播送或刊登營業之訊息者，可處以罰鍰。

建議主管機關可就上述疑似非法旅宿業者加強查核取締，並輔導非法旅宿業轉型為合法旅宿或伺機退場，進而確保消費者住宿安全與權益及保障合法旅宿業者之經營權，強化國內旅宿業管理效能。