

# 澳洲分散式系統技術中心與都柏林核心集

吳政叡 (Cheng-Juei Wu)

輔仁大學圖書資訊系專任副教授

Associate Prof.

Department of Library & Information Science

Fu-Jen University

E-mail: lins1022@fujens.fju.edu.tw

## 中文摘要

澳洲的「分散式系統技術中心」(DSTC)推動了一系列的研究計畫，重點在研究如何利用「都柏林核心集」來改進一般搜尋引擎的效能，如何來協助使用者在WWW和網際網路上搜尋資料，如何來結合都柏林核心集和其他的檢索技術(如URN)、網路協定(如HTTP和Z39.50)等，可以說是將都柏林核心集在檢索上的可能應用，發揮的淋漓盡致。本文介紹了其中三個跟都柏林核心集密切相關的計畫—TURNIP、HotOIL、MetaWeb。

---

---

### DSTC and the Development of the Dublin Core

#### Abstract

The Distributed Systems Technology Center in Australia is currently developing a series of research projects related to the Dublin Core. The focus of the DSTC research is to use the Dublin Core to improve the efficiency of current search engines on the World Wide Web. In this paper, I introduce three research projects of the DSTC which are closed related to the Dublin Core: the TURNIP project, the HotOIL project, and the MetaWeb project.

---

---

關鍵字：元資料，分散式系統技術中心，都柏林核心集，Metadata，DSTC，Dublin Core，HotOIL，MetaWeb。

## 一、前言

都柏林核心集 (Dublin Core) 創始於1995年3月由國際圖書館電腦中心 (Online Computer Library Center, 簡稱OCLC) 和 National Center for Supercomputing Applications (NCSA) 所聯合贊助的研討會, 是五十二位來自圖書館、電腦、網路方面的學者和專家共同研討下的產物。目的是希望建立一套描述網路上電子文件特色的方法, 來協助資訊檢索。研討會的中心問題是--如何用一個簡單的元資料記錄來描述種類繁多的電子物件? [註 1] 主要的目標是發展一個簡單有彈性, 且非圖書館專業人員也可輕易了解和使用的資料描述格式, 來描述網路上的電子文件。

都柏林核心集最近一次的研討會為第五次研討會, 於1997年10月6-8日在芬蘭的赫爾辛基舉行, 由於在本文撰寫期間, 第五次研討會的正式報告尚未出版, 因此先根據澳洲國家圖書館的一位與會者--Bemal Rajapatirana的報告, 先行介紹第五次研討會的情況與成果 [註2], 待第五次研討會的正式報告出爐後, 作者會另撰專文來加以介紹。

根據Bemal Rajapatirana的報告, 與會者達成了如下的幾項共識:

- (一) 加快標準化的腳步—由於都柏林核心集的15個基本項目架構, 自第四次研討會以來已普遍獲得認同, 同時都柏林核心集也得到世界各國很多研究者的肯定, 並且嘗試建造系統, 此時若無一定的標準來遵循, 將使系統的建造者無所適從和系統的更改頻繁。因此基於都柏林核心集已趨成熟的共識, 決定推派代表撰寫RFC的草案, 呈交給 IETF進行標準化的過程。
- (二) 區分簡單和複雜兩種都柏林核心集格式—簡言之, 所謂簡單 (simple) 和複雜 (complex) 格式的區分, 一般而言主要是以有無使用任何修飾詞作為標準來劃分的。由於都柏林核心集的15個基本項目已有共識, 因此簡單都柏林核心集的標準化過程將會較早開始。
- (三) 語法上採用HTML和RDF格式為主—HTML的格式目前是使用4.0版本, 寫法請參見作者的另一篇文章 [註 3]。
- (四) 成立工作小組—針對一些尚未有定論的議題, 組成工作小組進行研討, 主要有
  - (1) 內容或格式尚未有定論的基本項目, 如Date、Relation、Rights Management等項目。
  - (2) 修飾詞。

(3) 特殊性議題，如都柏林核心集和Z39.50間的互換。

(五) 次項目（或類別修飾詞）的制定原則

(1) 與基本項目一致，都是可省略的選擇項。

(2) 次項目須能進一步協助詮釋項目的內容。

(3) 祇展開一層，免得結構過於複雜。

(4) 數目盡可能精簡，有可能需要類別修飾詞的基本項目，將限於Title、Creator、Contributor、Publisher、Date、Relation、Coverage等。

澳洲的分散式系統技術中心（Distributed Systems Technology Center，簡稱DSTC）所推動的一系列計畫，是目前為止在都柏林核心集的應用和實作系統上，最著名和成功的系列研究計畫之一，這由都柏林核心集的第四次研討會在澳洲的國家圖書館舉行可以窺知。分散式系統技術中心（DSTC）是在1992年由十個機構參與設立，主要宗旨是研究建設一個全球分散式資訊系統所需的技術，並取得澳洲政府和其他參與機構七年（1992-1999）總共6千萬美元的研究經費贊助，因此正致力於推動一些與資訊管理和檢索相關的計畫。[註 4] 而DSTC選用都柏林核心集作為資源描述的工具，因此有以下與都柏林核心集相關的主要計畫：[註 5]

(一) 將HTML文件中內嵌的都柏林核心集作成索引，並提供檢索的介面。

(二) 使用都柏林核心集作為URN服務（TURNIP計畫）中的資源描述工具。

(三) 使用都柏林核心集來呈現WWW伺服器（使用HTTP通訊協定）和圖書館書目資料（使用Z39.50通訊協定）檢索服務的結果（如HotOIL計畫）。

(四) 在X.500服務中，使用都柏林核心集和政府資訊服務（GILS）。

(五) 使用資源描述結構（Resource Description Framework，簡稱 RDF，詳細介紹請參見第六章）來裝載都柏林核心集。

(六) 擴充PURL來支持都柏林核心集。

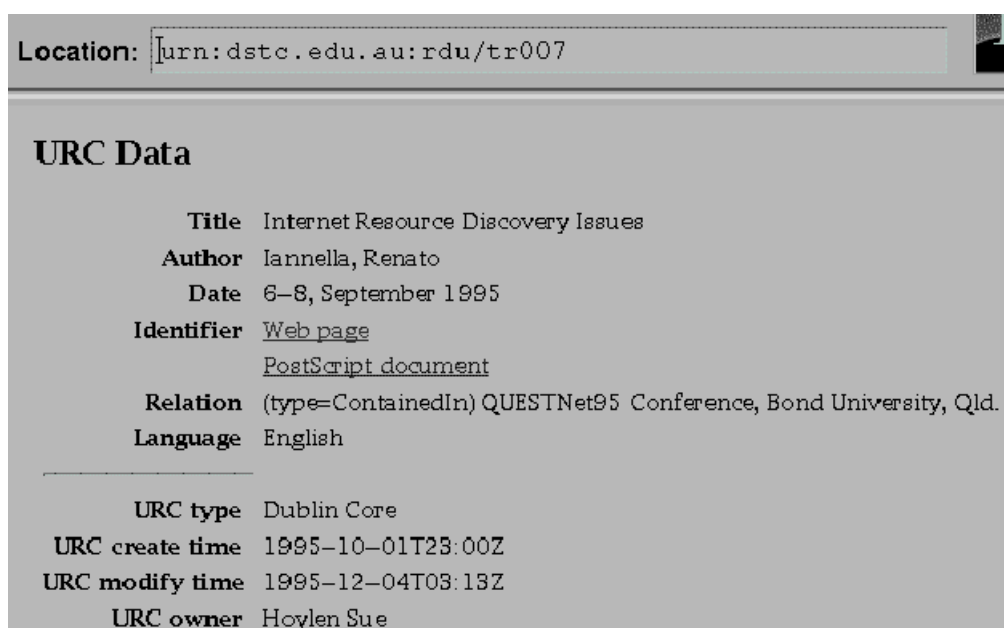
以下將介紹其中三個跟都柏林核心集最密切相關的計劃—TURNIP、HotOIL、MetaWeb。

## 二、TURNIP

DSTC的全球資源識別名稱互通研究計畫（The URN Interoperability Project，簡稱TURNIP），使用URN作為資源的識別名稱和檢索點（如圖1中的urn:dstc.edu.au:rdu/tr007，有關urn字串的解釋，請參見『三個元資料格式的比較分析』一文 [註 6]），TURNIP系統有以下的特色：[註 7]

- (一) 資料範圍涵蓋全球。
- (二) 識別名稱全球唯一性--沒有兩個資源有相同的識別名稱。
- (三) 持久性 -- 祇要負責命名的機構存在，此機構所命名的識別名稱即會存在。
- (四) 可擴充性 -- 識別名稱可輕易隨資料量的增加來調整。
- (五) 支持現存的系統，如URL。
- (六) 延伸性佳 --可輕易加入新功能。
- (七) 獨立性 -- 命名的機構彼此獨立，祇須負責維持所管轄資訊的識別名稱具唯一性即可。

由於URN是所謂の間接指引，須使用URC來對映到資源的實施儲存位置，因此須要透過URN解譯器（URN Resolver）自資料庫中找出相對映的URC（用來存放資源的相關資訊，包括檔案的實際位置--URL），而TURNIP 系統的URC則是採用都柏林核心集來作為描述資源的主要工具（如圖1中的URC type）[註 8]，整個系統的架構如圖2所示。[註 9]



Location: urn:dstc.edu.au:rdu/tr007

### URC Data

<b>Title</b>	Internet Resource Discovery Issues
<b>Author</b>	Iannella, Renato
<b>Date</b>	6-8, September 1995
<b>Identifier</b>	<a href="#">Web page</a> <a href="#">PostScript document</a>
<b>Relation</b>	(type=ContainedIn) QUESTNet95 Conference, Bond University, Qld.
<b>Language</b>	English

---

<b>URC type</b>	Dublin Core
<b>URC create time</b>	1995-10-01T23:00Z
<b>URC modify time</b>	1995-12-04T03:13Z
<b>URC owner</b>	Hoyleen Sue

圖1. DSTC的URN和URC對映的結果畫面

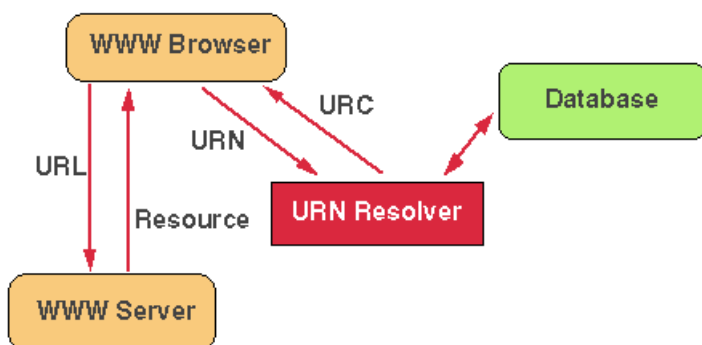


圖2. DSTC的URN和URC系統架構圖

### 三、HotOIL

HotOIL提供一個整合的搜尋介面給使用者，當一個查詢送入HotOIL系統如圖 3所示時，系統會將使用者的查詢作必要的格式處理後，傳遞到所有使用者選擇的資料庫（可以多個），同時將各個資料庫或搜尋引擎傳回的結果綜合整理後顯示給使用者。[註 10] 如圖 4是作者查詢“metadata”後，HotOIL將所有的結果綜合整理後顯示的畫面，使用者可以很方便的來進一步縮小或擴大查詢的範圍，圖 5是作者選擇“metadata model”主題後，HotOIL呈現的查詢結果，此畫面與一般搜尋引擎提供的畫面類似。

**HotOIL**  
**New Search**

Enter Keywords:

Select search engine(s)

<input checked="" type="checkbox"/> AltaVista 10 results	<input checked="" type="checkbox"/> LookSmart 10 results
<input checked="" type="checkbox"/> Lycos 10 results	<input checked="" type="checkbox"/> MEDLINE(R)--Z39.50 10 results
<input checked="" type="checkbox"/> Melvyl-System(R)--Z39.50 10 results	<input checked="" type="checkbox"/> Unilinc--Z39.50 10 results

Search results format:

Cluster and Browse (Hyper-index browser)

List matching results directly

圖3. HotOIL檢索介面

**HotOIL**

**Current Focus**

You can choose to see the [results of the search on metadata](#).

**Refinements**

Or choose a refinement from the following:

<a href="#">nsgic metadata</a>	<a href="#">metadata videoconference</a>	<a href="#">fgdc metadata</a>
<a href="#">250k metadata</a>	<a href="#">benefits of metadata</a>	<a href="#">metadata model</a>
<a href="#">metadata sources</a>	<a href="#">roads metadata</a>	<a href="#">metadata crosswalk</a>
<a href="#">metadata for rnr2lahar</a>	<a href="#">ausliq metadata</a>	<a href="#">metadata page</a>
<a href="#">metadata in database</a>	<a href="#">entity-attribute-value metadata</a>	<a href="#">semantic metadata</a>
<a href="#">syp and metadata</a>	<a href="#">metadata in geosciences</a>	

**Enlargements**

Or choose an enlargement from the following: *Sorry, no enlargements could be generated.*

---

**New Search**

圖4. HotOIL的查詢結果畫面


**HotOIL**

**Current Focus**

The current focus is **`metadata model`**. There are no refinements, so the results are displayed below. Either select a resource from the result list, or modify the search by choosing an *enlargement* from below.

**Results**

Total of 10 matches retrieved.

 [Metadata Information Day Lund, Sweden October 11, 1996](#)  
**URL:**<http://www.oclc.org:5046/~weibel/lund/part2/meta-lund.html>  
**Subject:** Metadata Information Day Lund, Sweden October 11, 1996. Part 2: The Metadata Workshop Series. Stuart L. Weibel. Senior Research Scientist. OCLC Office of..


 [metadata](#)  
**URL:**<http://ecsinfo.hitc.com/sec2/graham.html>  
**Subject:** EOSDIS Core System Metadata Design. Graham Bland Earth Observation Sciences, Ltd 301/925-0433, grahamb@eos.hitc.com. Pervading the ECS is the Earth...

圖5. HotOIL呈現的最後查詢結果畫面

若是使用者偏好一般搜尋引擎的工作方式，可使用”List matching results directly”選項，如圖 6所示，則HotOIL會跳過如圖 4的步驟，直接呈現查詢的結果，如圖 7所示。

圖6. HotOIL中使用 List matching results directly 選項畫面

圖7. 使用 List matching results directly 選項的查詢結果畫面

在系統的設計上，HotOIL使用URN來標示資源，URC則負責告訴系統，資源的實際儲存位置、如何來查詢、如何來詮釋查詢的結果等的資訊 [註 11]，這些資訊正是元資料的功能，沒有元資料的協助，系統將無法同時來面對眾多不同的資料庫設計或搜尋引擎，也無法提供如圖 4的整理功能，而如同前面提及的，DSTC的URC是採用都柏林核心集來作為描述資源的工具。

在網路檢索的設計上，HotOIL同時支持HTTP 和Z39.50兩種通訊和檢索協定，將來還能查詢支援SQL的資料庫和使用DAP 或LDAP 通訊協定的X.500資料庫，如此一來幾乎將網路上所有的系統一網打盡。

## 四、MetaWeb

MetaWeb 是元資料工具和服務計畫（Metadata Tools and Services Project）的簡稱，計畫的期間是從1997 年7月到1998年6月，此計畫的主要宗旨為發展元資料欄位、使用者工具、索引服務等來提昇元資料在網際網路上的應用，其合作夥伴包括澳洲國家圖書館。[註 12] 計畫預計達成如下的重要目標：[註 13]

- (一) 建立一個澳洲的元資料網站。
- (二) 建立一個測試用的元資料資料庫。
- (三) 發展一套工具來協助元資料的創造和管理。
- (四) 擴充都柏林核心集來發展一套適合澳洲在資源發現（Resource Discovery）上的標準。
- (五) 推廣數位簽名在網路上電子文件的使用。

目前MetaWeb系統使用的測試版搜尋引擎名為「HotMeta」，HotMeta是一個元資料搜尋引擎，可以將網路上電子文件的元資料，提取出來作成索引後，來加以查詢。系統包含兩個元件，一個是類似一般搜尋引擎自動抓取程式的「收集者」（Gatherer），「收集者」負責將HTML文件中的元資料抽出，以SOIF（Summary Object Interchange Format）[註 14] 格式送到另外一個元件 -- 「經紀人」（Broker）存放，「經紀人」負責管理存放的資料（含其索引），並提供如圖 8的檢索介面。圖 9是使用HotMeta查詢 metadata後的結果顯示畫面，值得一提的，在目前的實驗階段，HotMeta的資料庫將祇涵蓋澳洲網站的資料，所提取的元資料也祇限定在都柏林核心集，如圖 10中的 ”DC.title” 所示。



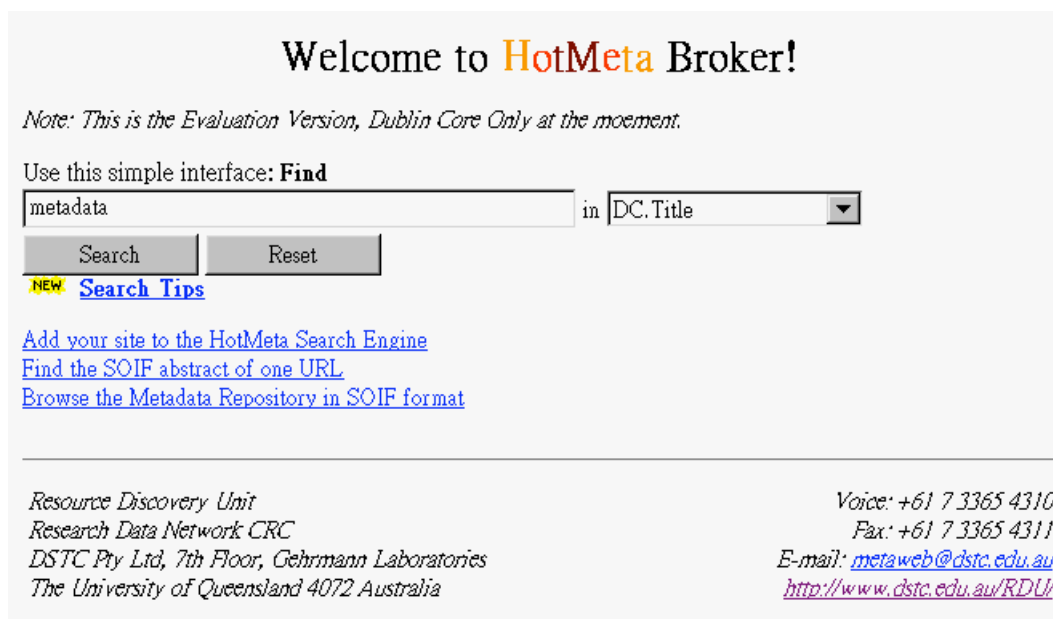


圖8. HotMeta元資料搜尋引擎的檢索介面

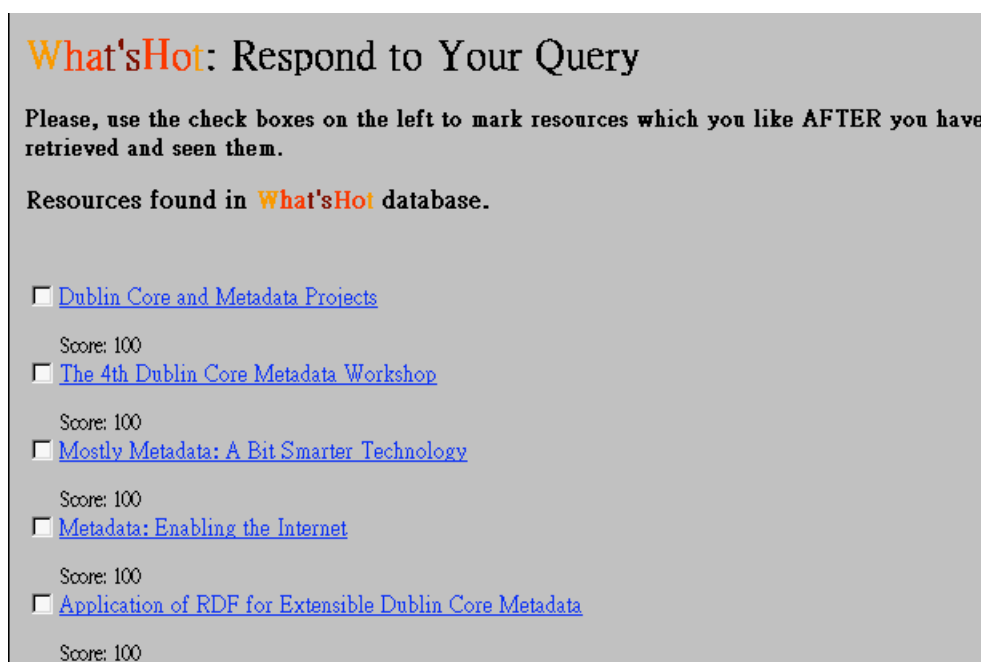


圖9. 使用HotMeta查詢 metadata後的結果顯示畫面



圖10. HotMeta元資料搜尋引擎中使用DC.title選項

HotMeta 元資料搜尋引擎也提供其他兩種查詢的方法，其中一種是直接使用 URL 來查看一個特定資源的元資料，其畫面如圖 11 所示。

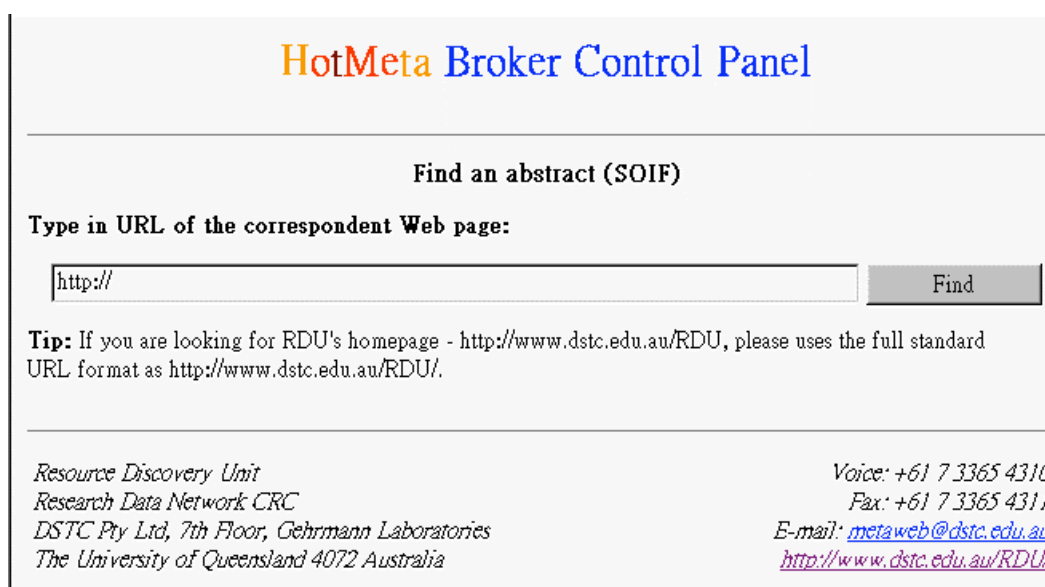


圖11. HotMeta中直接使用URL來查看一個特定資源的元資料畫面

## 五、結語

以「都柏林核心集」在國外的發展現況來看，1997年10月的第五次研討會已有專門的議程，來針對都柏林核心集的實作系統進行展示和討論，這是以前四次研討會所沒有的，顯示都柏林核心集已漸趨成熟和受到肯定。除了已開發系統的介紹外，也有一些正在籌建中的都柏林核心集相關系統的宣佈，以下是它們的

簡介：[註15]

- (一) 丹麥政府決定自西元1997年起將所有政府的出版物上網，系統的主要規格之一，是採用都柏林核心集來描述文件和協助查詢。
- (二) 荷蘭國家圖書館將發展一種新的全球資訊網服務，系統的主要做法是要在所有已蒐集的網頁中，加入都柏林核心集的資料，新的網頁將要求提供者先自行加入都柏林核心集的資料後再送呈，將來荷蘭國家圖書館的搜尋引擎會利用這些元資料來協助檢索。
- (三) 英國的UKOLN正在推行一個名為BIBLINK的計劃，在出版社和國家書目中心間建立一條網路通訊管道，來直接交換書籍紀錄和資訊，這套系統是使用都柏林核心集作為其基本的格式。

由以上對都柏林核心集現況的簡介中，可知都柏林核心集是已經被普遍接納的一種元資料，因此國外很多的研究計畫也將都柏林核心集納入，本文介紹了其中的一個由「分散式系統技術中心」(DSTC)推動的系列研究計畫。澳洲的分散式系統技術中心(DSTC)是在1992年由十個機構參與設立，主要宗旨是研究建設一個全球分散式資訊系統所需的技術，並取得澳洲政府和其他參與機構七年(1992-1999)總共6千萬美元的研究經費贊助，因此正致力於推動一些與資訊管理和檢索相關的計畫，而分散式系統技術中心(DSTC)所推動的這一系列計畫，是目前為止在都柏林核心集的應用和實作系統上，最著名和成功的系列研究計畫之一。本文介紹了其中三個跟都柏林核心集最密切相關的計畫—TURNIP、HotOIL、MetaWeb。

DSTC所推動的這一系列關於全球資訊存取(Global Information Access)的研究計畫，都是採用都柏林核心集做為描述資源的格式，因此是以都柏林核心集來協助使用者在WWW和網際網路上搜尋資料，並以都柏林核心集來和其他的檢索技術(如URN)、網路協定(如HTTP和Z39.50)等相結合，可以說是將都柏林核心集的檢索效用發揮的淋漓盡致。

DSTC在使用者檢索介面上的設計(如HotOIL研究計畫)非常值得借鏡，不但可以自動替使用者將查詢送至各個搜尋引擎和資料庫，還會將傳回的結果先行整理，大大的減輕了使用者過濾的負擔，這是作者所見過最好的介面設計之一。另外一個DSTC檢索系統的優良設計，是直接介面的設計上，應用了都柏林核心集的欄位來縮小查詢的範圍(如MetaWeb研究計畫，參考圖10)，這將有助於檢索效能的提昇。

## 註釋

- 註 1：吳政叡，「三個元資料格式的比較分析」，中國圖書館學會會報 57 期（民 85 年 12 月），頁39。
- 註 2：B. Rajapatirana, "The 5th Dublin Core Metadata Workshop: a report and observations," 2 Dec. 1997, <<http://www.nla.gov.au/nla/staffpaper/helsinki.html>>.
- 註 3：吳政叡，「元資料實驗系統和都柏林核心集的發展趨勢」，國立中央圖書館臺灣分館館刊 4 卷 2 期（民 86 年 12 月），頁18。
- 註 4：R. Iannella, "Resource Discovery Unit," <<http://www.dstc.edu.au/RDU/>>, (20 Nov. 1997).
- 註 5："Dublin Core and Metadata Projects," <<http://www.dstc.edu.au/RDU/DCMD/>>, (20 Nov. 1997).
- 註 6：同註1，頁41。
- 註 7：R. Iannella and H. Sue, "Basic URN Service (BURNS)," <<http://www.dstc.edu.au/RDU/TURNIP/burns.html>>, p. 3.
- 註 8："Dublin CoreDown Under," <<http://www.dstc.edu.au/RDU/pres/warwick/>>, p. 3.
- 註 9：同註7，頁2。
- 註10：N. Ward, "HOTOIL," <<http://www.dstc.edu.au/BDU/APAP/HotOIL/HotOIL.html>>, (26 Jan. 1998), p. 1.
- 註11：同註10，頁1。
- 註12：D. Campbell, "The MetaWeb Project," 22 January 1998, <<http://www.dstc.edu.au/RDU/MetaWeb/>>, p. 1.
- 註13：D. Campbell, "The MetaWeb Project: Project activities and timeline," 25 November 1997, <<http://www.dstc.edu.au/RDU/MetaWeb/timeline.html>>, p. 1.
- 註 14：D. Wessels, "The Summary Object Interchange Format (SOIF)," 31 Jan. 1996, <<http://harvest.transarc.com/afs/transarc.com/public/trg/Harvest/user-manual/node151.html#SECTION000120000000000000000000>>.
- 註 15：同註2，頁3-4。