

資源描述架構在都柏林核心集的應用介紹

吳政叡(輔仁大學圖書資訊系專任副教授)

摘要

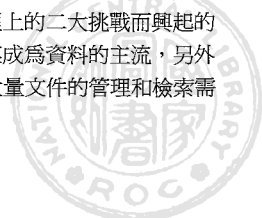
元資料是將來用來描述(網路)資源的格式,由於資源的種類複雜,多種元資料共存共榮乃為未來必然的趨勢,因此需要有一種適當的工具,來同時攜帶多種元資料來往於網路上,而「資源描述架構」(RDF)即為此種工具之一。本文主要是介紹「資源描述架構」及其在「都柏林核心集」的可能應用,資源描述結構是由W3C主導和結合多個元資料團體(如都柏林核心集等)所發展而成的一個架構,可用來攜帶多種不同的元資料來往於網際網路和 WWW 上。本文介紹了 RDF 的核心資料模式,以及「具體化」(Reification)和「聲明」(Assertion)等的機制,最後是一個使用 XML 語法和 RDF 核心資料模式,來攜帶都柏林核心集記錄的實例。

關鍵詞:元資料(Metadata) 資源描述架構(RDF) 都柏林核心集(Dublin Core)

一、前言

資源描述結構(Resource Description Framework, 簡稱RDF)是一個用來攜帶多種不同的元資料來往於網路上的工具(註1)。元資料(Metadata)最常見的英文定義是“data about data”,可直譯為描述資料的資料,主要是描述資料屬性的資訊,用來支持如指示儲存位置、資源尋找、文件紀錄、評價、過濾等的功能。以圖書館的角度來看,就

其本義和功能而言,元資料可說是電子式目錄,因為編製目錄的目的,即在描述收藏資料的內容或特色,進而達成協助資料檢索的目的(註2)。因此元資料是用來揭示各類型電子文件或檔案的內容和其他特性,其典型的作業環境是電腦網路作業環境。換言之,元資料是因應現代資料處理上的二大挑戰而興起的:一是電子檔案成為資料的主流,另外一個是網路上大量文件的管理和檢索需求。



至於元資料的種類，下面是一些常見的清單。首先，國際圖書館協會聯盟 (International Federation of Library Association and Institutions，簡稱 IFLA) 在描述元資料資源的首頁中 (註 3)，列舉了以下的元資料種類：Dublin Core、EAD (Encoded Archival Description)、FGDC's Content Standard for Digital Geospatial Metadata、DIF (Directory Interchange Format)、GILS (Government Information Locator Service)、IAFA/whois⁺⁺ templates、MARC、PICS (Platform for Internet Content Selection)、RDM (Resource Description Messages)、SOIF (Summary Object Interchange Format)、SHOE (Simple HTML Ontology Extensions)、TEI、URC (Uniform Resource Characteristics)、X3L8 Proposed ANSI standard for data representation。

其次是在『Judy And Magda's List of Metadata Initiatives』的網頁中，按類別提出一些經常被廣泛使用或具有潛力的元資料如下(註 4)：

- (一)通用描述型—MARC、Dublin Core、Edinburgh Engineering Virtual Library (EEVL)、Semantic Header for Internet Documents、GILS、URC、X3L8 Proposed ANSI standard for data representation、IAFA

Templates、NetFirst、Header for HTML documents、SOIF、MCF (Meta content Format)、PICS。

- (二)文字檔描述型—TEI、BibTex、Gruber Ontology for Bibliographic Data、RFC 1807。
- (三)數據資料類—ICPSR Data Documentation Initiative、SDSM (Standard for Survey Design and Statistical Methodology Metadata)。
- (四)音樂類—SMDL (Standard Music Description Language)。
- (五)圖像與物件類—CDWA (Categories for the Description of Works of Art)、CIMI (Consortium for the Computer Interchange of Museum Information)、VRA Core Categories、MESL Data Dictionary。
- (六)地理資料類—FGDC's Content Standards for Digital Geospatial Metadata。
- (七)檔案保存類—EAD、Z39.50 Profile for Access to Digital Collections、Fattahi Prototype Catalogue of Super Records。

由以上的列表和清單可知，因為網

路資源的種類複雜，用途殊異，因此多種元資料共存共榮實為不可避免的趨勢，因此需要有一種適當的工具，來同時攜帶多種元資料來往於網路上，而「資源描述架構」即為此種工具之一。

資源描述結構(Resource Description Framework, 簡稱RDF)是由全球資訊網協會(W3C)主導和結合多個元資料團體(如都柏林核心集等)所發展而成的一個架構，可用來攜帶多種不同的元資料來往於網際網路和WWW上。因為W3C先前曾致力發展一個元資料—PICS(Platform for Internet Content Selection)(註5)，因此RDF受到PICS很深的影響，在語法上則是遵循另一個W3C致力推廣的架構—XML(Extensible Markup Language)(註6)，由於目前XML已受到業界廣泛的支持，如瀏覽器的兩大霸主 Netscape (註7)和Internet Explorer(註8)都已經各自製作使用XML格式的元資料規格，並且也已呈送W3C審核，因此XML與RDF的發展可說是備受矚目。

二、RDF的核心資料模式與聲明的機制

以下根據W3C的RDF工作小組的草案(註9)，來對RDF模型做更進一步的介紹，基本上RDF是一個與任何特定(電腦)語法無關的抽象的資料(表達)模式，用來呈現一個特性與其值。而所謂的

「特性」(Property)，可能是：

資源的屬性：如題名、著者等，都柏林核心集的題名欄位即可歸屬於這類。

資源間的關係：如都柏林核心集的關連(Relation)和來源(Source)兩欄位即屬於這類範疇。

RDF的另外一個特點是語法獨立性，因此兩段看起來差異很大的RDF陳述，事實上可能是描述相同的一件事，這是因為RDF是一個抽象的資料模式。由於這個抽象的特點，各種不同的元資料(如都柏林核心集)均可利用此種抽象的資料模式，來表達它們的內容。

RDF的核心資料模式(RDF Core)定義如下：

(一)N：一個點(Node)的集合(Set)，此處的「集合」是一個數學的名詞和概念，在此的意義和用法正如在數學中一般。而「點」可以是一個資源(如網頁)或是物件(Object)，甚至可以是一個「特性」(Property)。《作者註：「特性」的意義請參見前面的描述。》

(二)P(特性型態)：是一個N的子集合(Subset)。

(三)T：一個含有三個元素的「有序對」(Tuple)，其形式為(P, N, V)，即有序對中的第一個元素來自前面的集合P，第二個元素來自前面的集合N，第

三個元素V可以是來自集合N，
或者是一個單純的值（如字串
「吳政勸」）。

例子：吳政勸是網頁 <http://mes.lins.fju.edu.tw>的著者，可用RDF的有序對表示如下：

{著者, {<http://mes.lins.fju.edu.tw>}, [吳政勸]}

上述的有序對中，著者是一個「特性」，[吳政勸]是此特性的值，網頁 <http://mes.lins.fju.edu.tw>是一個點(Node)。

從另外一個角度，可把RDF核心資料模式的三個元素有序對(P, N, V)，以數學中的圖學表示如下：

N — P —> V

即將N和V視為點，P是從N到V弧線的標示，因此上述的例子又可表示為 [<http://mes.lins.fju.edu.tw>] — 著者 —> [吳政勸]

此外又可透過所謂的「具體化」(Reification)將「特性」(Property)變成一個新的點(假設為X)，從而產生三個新的有序對如下：

(一) {PropName, X, P}。

(二) {PropObj, X, N}。

(三) {PropValue, X, V}。

以上的例子來說，若將「特性」著者具體化為新的點X後，將產生如下的三個新有序對：

(一) {PropName, X, 著者}。

(二) {PropObj, X, [<http://mes.lins.fju.edu.tw>]}

。

(三) {PropValue, X, [吳政勸]}。

若將描述同一個資源的眾多特性的有序對集結起來，即成為RDF的「聲明」(Assertion)，例如描述網頁 <http://mes.lins.fju.edu.tw> 的兩個有序對：

(一) {著者, [<http://mes.lins.fju.edu.tw>], [吳政勸]}

(二) {題名, [<http://mes.lins.fju.edu.tw>], [吳政勸的首頁]}

組合起來即構成RDF的「聲明」。

三、一個都柏林核心集記錄的RDF實例

都柏林核心集(Dublin Core)為備受矚目的元資料之一，是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，經過五十二位來自圖書館、電腦和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10)？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，由於在寫作本書時，第五次研討會的正式報告尚未出版，祇好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11)，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12)。

(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。

1997年10月公布的資料著錄項目列表如下(註13)：

(一)主題和關鍵詞 (Subject)：作品所屬的學術領域，控制語彙用scheme註明出處如LCSH，亦可包含分類號如杜威十進分類號(Dewey Decimal Number)。例子：Subject =都柏林核心集。

(二)題名(Title)：作品名稱。

例子：Title = 都柏林核心集與元資料實驗系統。

(三)著者 (Creator)：作品的創作者或組織。

例子：Creator =吳政叡。

(四)簡述 (Description)：文件的摘要或影像資源的內容敘述。

(五)出版者(Publisher)：負責發行作品的組織。

(六)其他參與者(Contributors)：除了著者外，對作品創作有貢獻的其他相關人士或組織。〔註：如書中插圖的製作者。〕

(七)出版日期(Date)：作品公開發表的日期，建議使用如下格式—YYYY-MM-DD和參考下列網址：http://www.w3.org/TR/NOTE_datetime。在此網頁中共規範有六種格式，都是根據國際標準日期暨時間格式—ISO(國際標準組織) 8601 制定而成，是ISO8601的子集合(subset)，現在列舉和解說如下以供參考(註14)：

例子：1997-09-07(西元1997年9月7日)。

(八)資源類型(Type)：作品的類型或所屬的抽象範疇，例如網頁、小說、詩、技術報告、字典等，建議參考下列網址：<http://sunsite.berkeley.edu/Metadata/types.html>。

例子：Type =Text.Dictionary

例子：Type = 文字.技術報告

(九)資料格式(Format)：告知檢索者在使用此作品時，所須的電腦軟體和硬體設備，例如text/html (MIME 格式)、ASCII、Postscript (一種印表機通用格式)、可執行程式、JPEG(一種通用圖像格式)。亦可擴展至非電子文件，例如 book(書本)、叢書、期刊。

例子：Format = text/html。

(十)資源識別代號 (Identifier)：字串或號碼可用來唯一標示此作品，例如URN、URL、ISSN、ISBN等。

(十一)關連(Relation)：與其他作品(不同內容範疇)的關連，或所屬的系列和檔案庫。

例子：Relation=<http://mes.lins.fju.edu.tw/>。

(十二)來源(Source)：作品從何處衍生而來(同內容範疇)，例如莎士比亞的某個電子書出自那個紙本。

(十三)語言(Language)：作品所使用的語言，建議遵循RFC 1766的規定，請參考下列網址：<http://ds.internic.net/rfc/rfc1766.txt>，RFC 1766 是使用ISO 639的二個字母的語言代碼(註15)。

例子：Language=en(註16)

(十四)涵蓋時空(Coverage)：作品所涵蓋的時期和地理區域。

(十五)版權規範(Rights)：作品版權聲明和使用規範。

以下是使用XML語法和RDF核心資料模式來攜帶一個都柏林核心集記錄的實例：

```
<xml:: namespace href="http://
purl.oclc.org/metadata/dublin_core
_elements" as="DC">
```

```
<xml:: namespace href="http://
www.w3.org/schemas/rdf-schema" as="
RDF">
```

```
<RDF:serialization>
```

```
<RDF:assertions href="http : /
/mes.lins.fju.edu.tw/mes/">
```

```
<DC:creator>吳政觀</DC:creator
```

```
>
<DC:title> 元資料實驗系統 </DC
:title>
```

```
<DC:subject> 都柏林核心集 </DC
:subject>
```

```
<DC:subject>元資料</DC:subject
```

```
>
<DC:description>有鑑於元資料對
資料著錄和檢索的重要性，作者建立了一個相關的實驗系統—元資料實驗系統(Metadata Experimental System，簡稱MES，網址：http://140.136.85.194/mes 或 http://mes.lins.fju.edu.tw/
```

mes)，作者建立MES 目的，除了是讓讀者透過這個系統，對元資料及其未來的可能運作方式，有更具體的認知外；也希望利用此一實驗系統，來測試和驗證元資料的功能和效用，例如都柏林核心集這種簡易的資料描述格式，是否如制定者們所預期的，足以滿足大部分網路文件著錄和檢索的需求。MES 是一開放性的實驗系統，歡迎任何人上站著錄自己的網頁或文件，以供他人查詢和檢索。

```
</DC:description>
```

```
<DC:date>
```

```
<IS08601>1997-009</IS08601>
```

```
</DC:date>
```

```
<DC:type>homepage</DC:type>
```

```
<DC:format> text / html </DC:
format>
```

```
<DC:identifier><url>http://140.136.85.194/mes</url></DC:identi-
fier>
```

```
<DC:rights>所有版權屬於吳政觀<
/DC:rights>
```

```
</RDF:assertions>
```

```
</RDF:serialization>
```

下面的RDF文法是摘錄自W3C 的RDF工作小組 1997年10月 2日公開的草案(註17)，此文法是以電腦界通用的BNF (Backus-Naur Form) (註18)形式呈現，同時由於工作小組的草案是會隨時增修的，請自行連上 W3C 的網站 (<http://www.w3.org/Metadata/RDF/Group/WD->

rdf-syntax)查看最新的發展。

- (一)RDF ::= ' <RDF:serialization>
' node* </RDF:serialization>
' >'
(二)node ::= resource | assertions
| aggregate
(三)resource ::= ' <RDF:resource
' idAttr? ' > property* </RDF:
:resource>
(四)assertions ::= ' < RDF :
assertions' idRefAttr* ' >
property* </RDF: assertions
' >
(五)aggregate ::= sequence | bag |
alternatives
(六)sequence ::= ' <RDF:seq' idAttr
' ? ' > aggnode* ' </RDF:seq>
(七)bag ::= ' <RDF:bag' idAttr?
' > aggnode* ' </RDF:bag>
(八)alternatives ::= ' < RDF:
alternatives' idAttr? ' >
aggnode* </RDF: alternatives
' >
(九)aggnode ::= node | ' <RDF:li'
hrefAttr ' />
(十)idRefAttr ::= hrefAttr | idAttr
(十一)hrefAttr ::= ' href = "
' resource URI "'
(十二)idAttr ::= ' id=" IDsymbol
' ,"
(十三)resourceURI ::= (see RFC
1738)

- (十四)IDsymbol ::= (any legal
XML name symbol)
(十五)property ::= ' < prop Name
idAttr? ' > ' propValue' </
' propName ' > | ' < prop
Name idRefAttr ' />
(十六)propName ::= name | name
Prefix ' : ' name
(十七)propValue ::= node | string
(十八)name ::= (any legal XML name
symbol)
(十九)namePrefix ::= (any legal
XML namespace prefix)
(二十)string ::= (any XML text
)

四、結語

元資料的興起和WWW 與搜尋引擎的盛行頗有關連，WWW 盛行後，為因應檢索網頁內容的需要而有搜尋引擎的產生，搜尋引擎運作的方式，基本上是屬於全文檢索，主要是透過自動抓取程式在網際網路上抓取網頁，然後以自動拆字(或詞)作索引的方式來建立其資料庫，做為檢索的基礎，這種操作方式的特點是高運作效率和一網打盡，因此有高回收率與低精確率的特性，這個低精確率的缺點，隨著WWW 網頁數量的急遽膨脹，成為無法忍受的致命傷。

很明顯的，我們需要更多的資訊，來從回覆的款目當中，挑選我們真正需

要的資料，而這些資訊必須由資料提供者來提供，因此如何制定一套資料描述格式，來有效率的描述收藏的資料，成爲一個重要的課題，這正是元資料日漸受到重視的原因。這種對資料須加以適當描述的體會，正是圖書館製作目錄的動機，於是這個古老的經驗又得到再一次的肯定。

都柏林核心集(Dublin Core)是一個簡單有彈性，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。這種簡單有彈性和適合各種專業人員的特性，正是它在國外越來越受到歡迎的主要因素，也是作者特別青睞都柏林核心集的原因，這是因爲作者同時具有圖書館學和電腦的背景，了解到在現階段，一種適合各專業人士的簡易元資料的必要性；一方面傳統的機讀編目格式過於繁瑣，也繼承太多的傳統包袱，同時傳統圖書館的著錄方式並不適合非圖書館專業的人。另一方面以作者對目前人工智慧、類神經元網路、模糊邏輯等相關學科的了解，知道創造一個具有現今一般圖書館員智慧的自動化系統，在現階段仍是一個遙不可及的夢想，因爲至今我們連模仿一個三歲小孩說和聽故事的智力都有困難，更別說是模仿一個成年的專業人士。所以綜合來說，在現階段資料的描述仍需以人工作業爲主，同時以今日網際網路上資料膨脹的速度來看，光靠圖書館員來處理是不夠的，由（眾多專業的）文件或資料的創造者本身

來自行加以描述，已是必然的趨勢，這正是類似都柏林核心集這種元資料受重視的原因。

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，也說明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：

- (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 正在推行一個名爲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爲其基本的格

式。

(四)在商業的應用上，一個稱為 STARTS的協定正在發展中，它可以辨識網頁中的元資料，來協助使用者過濾和排比查詢結果，STARTS已決定包含都柏林核心集。

綜觀以上的發展，顯示都柏林核心集已漸成熟和廣受肯定，以系統的實作而言，歐洲和澳洲可說是居於領先的地位，歐洲較注重都柏林核心集在圖書館相關服務上的應用，澳洲的DSTC則較偏重都柏林核心集在WWW 相關服務上的應用。

由於類似都柏林核心集這類的元資料正逐漸獲得肯定和使用，因此相關的攜帶工具也成為研究者注目的焦點。這是因為元資料的種類複雜且用途殊異，將來多種元資料共存共榮的局面已成為共識，因此一種可同時攜帶多種元資料來往於網際網路和WWW 上的架構，成為不可或缺的工具。基於此種認知，W3C 乃主導和結合多個元資料團體發展出「資源描述架構」(RDF)。雖然在第二次都柏林核心集的研討會中，也提出一個類似的多個元資料的攜帶工具—「瓦立克架構」(註19)，但是由於W3C 在網際網路和WWW 界的影響力甚鉅，作者預期RDF 終將獲得採用而取代「瓦立克架構」，成為攜帶都柏林核心集的主要工具，因此撰寫本文來介紹資源描述架構在都柏林核心集的可能應用方式。

註 釋：

- 註 1 E. Miller and B. Schloss, "Resource Description Framework (RDF) Model and Syntax," 2 Oct. 1997, <<http://www.w3.org/TR/WD-rdf-syntax-971002>>, p. 2.
- 註 2 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（民國86年12月），頁12。
- 註 3 IFLA, "Digital Libraries: Metadata Resources," 24 March 1997, <<http://www.nlc.bnc.ca/ifla/II/metadata.htm>>。
- 註 4 J. Ahronheim, "Judy and Magda's List of Metadata Initiatives," 2 Nov. 1997, <<http://www-personal.umich.edu/~jaheim/alcts/bibaccess.htm>>.
- 註 5 P. DesAutels, "Platform for Internet Content Selection," 18 July 1997, <<http://www.w3.org/PICS/index.htm>>.
- 註 6 D. Connolly, "Extensible Markup Language (XML)," 1 Dec. 1997, <<http://www.w3.org/XML/index.htm>>.
- 註 7 A. Hopmann, "Web Collections



- using XML," 9 March 1997, <<http://www.w3.org/TR/NOTE-XMLsubmit.html>>.
- 註 8 R. V. Guha and T. Bray, "Meta Content Framework Using XML," 24 June 1997, <<http://www.w3.org/TR/NOTE-MCF-XML/index.htm>>.
- 註 9 同註1, 頁2-6。
- 註10 吳政勸, 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11 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12 同註2, 頁18。
- 註13 S. Weibel and E. Miller, "Dublin Core Metadata Element Set: Reference Description," 2 Oct. 1997, <http://purl.oclc.org/metadata/dublin_core_elements>.
- 註14 M. Wolfand C. Wicksteed, "Date and Time Formats," 15 Sept. 1997, <<http://www.w3.org/TR/NOTE-datetime>>。
- 註15 H. T. Alvestrand, "Tags for the Identification of language," March 1995, <<http://ds.internic.net/rfc/rfc1766.txt>>, p. 2.
- 註16 "Guide to Creating Core Descriptive Metadata," 13 April 1996, <<http://www.ckm.ucsf.edu/people/jak/meta/mguide3.html>>, p. 7.
- 註17 同註1, 頁6-7。
- 註18 A. V. Aho, R. Sethi, and J. D. Ullman, *Compilers: Principles, Techniques, and Tools*, (Addison-Wesley: Reading, Massachusetts) 1988, p. 159.
- 註19 同註2, 頁14-15。

