

E-ARK DIP 傳遞信息封包介紹

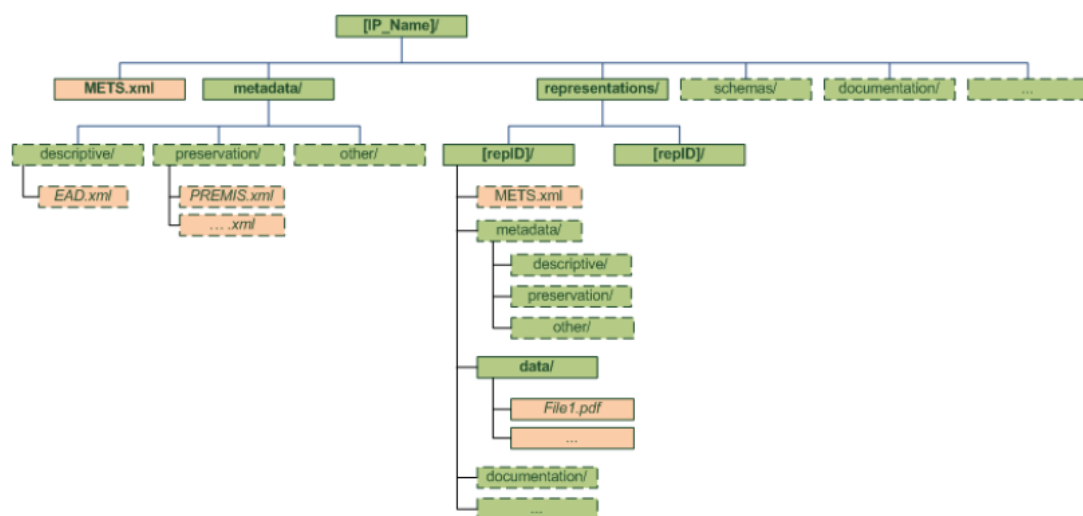
E-ARK Dissemination Information Package (傳遞信息封包，簡稱 DIP) 是源自於歐盟的E-ARK (European Archival Records and Knowledge Preservation) 計畫，該計畫為歐盟委員會的ICT方案贊助。E-ARK旨在解決電子檔案在長期保存及應用存取作業上發生無法讀取問題，藉由開放檔案資訊系統(Open Archive Information System, OAIS)模型下，發展解決電子檔案讀取問題的方案，以適用於歐盟各國家檔案產生者、各種規模檔案管理系統，以及各種型態的電子檔案能有依循的方向。

E-ARK專案根據OAIS模型架構，主要為預擷取(Pre-ingest)、擷取、資料管理與檔案儲存(Ingest, Data Management and Archival Storage)及存取(Access)3個階段。此外，數位資訊以信息封包的形式存在；而E-ARK專案中信息封包分為3種：（一）由資訊生產者傳送給OAIS系統的「提交信息封包」(Submission Information Package，簡稱SIP)；（二）儲存於典藏系統中的「存檔信息封包」(Archival Information Package，簡稱 AIP)；（三）當資訊消費者發出取用要求，從系統提供給使用者的「傳遞信息封包」(Dissemination Information Package，簡稱DIP)。

一. E-ARK信息封包結構介紹

(一)E-ARK信息封包結構

E-ARK將封包分成陳述訊息，詮釋資料和其他元件等獨立資料夾部分。



信息封包結構如下：

信息封包必須包含在一個實體資料夾內（即“Information Package folder”）。換句話說：在上層結構，信息封包必須由唯一的一個資料夾組成；信息封包資料夾應以信息封包的ID或名稱命名；

1. METS.xml

信息封包資料夾必須包含名為“METS.xml”的詮釋資料，其中包含關於封包及其元件的辨識和結構的訊息。

2. metadata

信息包資料夾必須包含一個“metadata”的資料夾，該資料夾必須包含與整個封包相關的詮釋資料。

(1) preservation

如保存性詮釋資料可用，則應將其存放在“preservation”資料夾。

(2) descriptive

如描述性詮釋資料可用，則應將其存放在“descriptive”資料夾。

(3) other

如有其他詮釋資料可用，則可包含在單獨的子資料夾中，例如另外名為“other”的資料夾。

3. representation

信息封包資料夾必須包含一個名為“representation”的資料夾；
“representation”資料夾必須包括每個單獨陳述信息的子資料夾(例如，
陳述訊息資料夾的名稱/或創建日期)，以唯一的辨識字串命名陳述訊息；

(1) data

陳述訊息資料夾必須包括一個名為“data”的子資料夾，其中包含構成陳述訊息的所有數據。

(2) METS.xml

陳述訊息文件夾可以包括“METS.xml”的詮釋資料文件，其中包含有關陳述訊息的辨識和結構的資訊。

(3) metadata

陳述訊息文件夾可包含“metadata”的子資料夾，其包含相關特定陳述訊息的所有詮釋資料。

4. schemas

信息封包資料夾和陳述訊息資料夾可以使用其他子資料夾進行擴展：我們建議將XML模式的詮釋資料納入在封包中。這些模式應該被放入信息封包文件夾中的“schemas”子資料夾。

5. documentation

建議將包含整個封包的其他文件或特定陳述訊息放入封包中。這些文件應該放在信息封包資料夾或陳述訊息資料夾的子資料夾“documentation”中。

E-ARK DIP的定義如下：

1. 在資料庫環境中發送(或準備發送)給使用者的資訊封包。
2. 支援E-ARK工具即透過特定的資料庫軟體呈現。

一、實體架構及詮釋資料規範

(一)DIP 數據模型及資料夾結構

E-ARK DIP的結構須符合E-ARK通用規範中的概述準則。基本E-ARK信息封包結構如下圖所示。它封裝在ZIP或TAR文件中，最頂層的資料夾包含DIP的唯一名稱。DIP由“METS.xml”文件組成，並提供其內容的清單（包裝信息）。任何其他詮釋資料文件放在“metadata/”資料夾中（例如PREMIS，EAD）。“representations”資料夾包含陳述訊息內容。其他必要文件都可以放在這些資料夾中，並且可以選擇放入“schemas /”資料夾和“documentation /”資料夾。DIP結構如下圖。

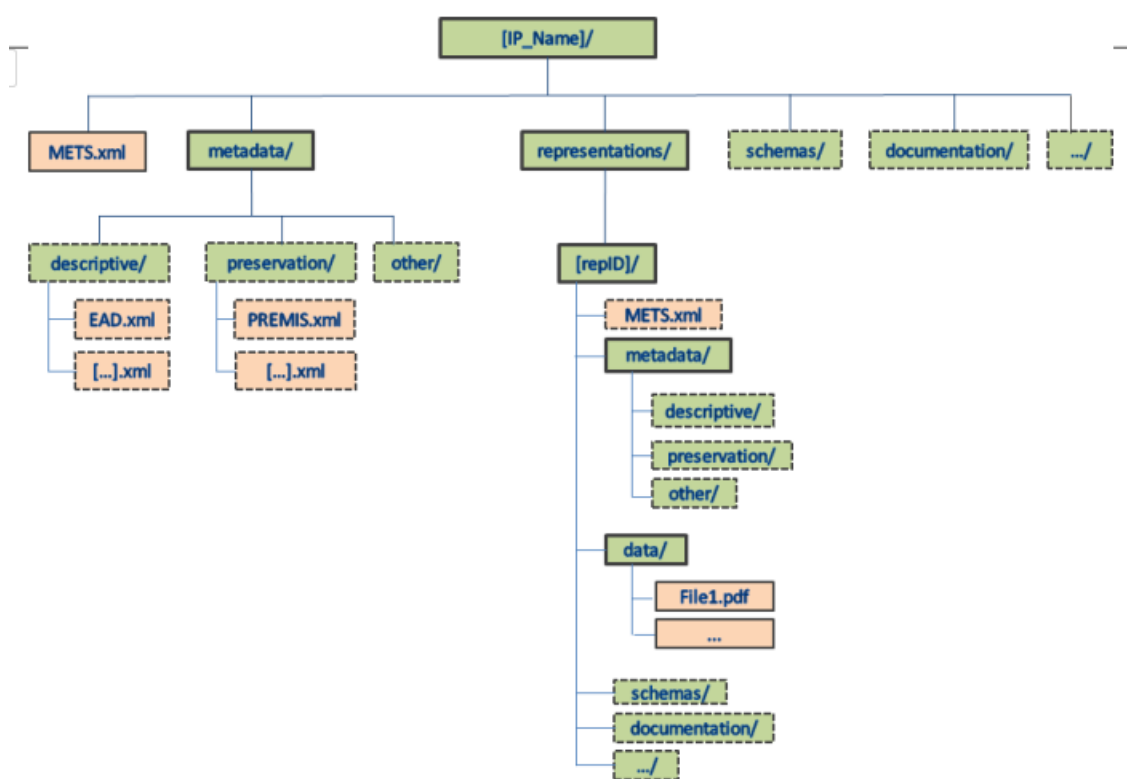


圖1 E-ARK DIP訊息封包結構

(二)DIP 詮釋資料

DIP詮釋資料涵蓋三個詮釋資料類別：結構（METS）；保存（PREMIS¹）；和描述性（EAD²）。

註¹ Preservation Metadata: Implementation Strategies 簡稱 PREMIS，數位保存詮釋資料，為基於 OAIS 架構用於資料長期保存的詮釋資料格式。

註² Encoded Archival Description 簡稱 EAD，檔案描述編碼格式，為專門用於描述檔案的格式。

METS文件位於根目錄資料夾中，其綱要文件（mets.xsd）位於/schemas資料夾中。大部分存取軟體假設EAD存在root / metadata/descriptive資料夾中且PREMIS存在root / metadata / preserve資料夾中。

因此，假設相關的綱要文件(ead3.xsd; premis.xsd)存在於/ schemas文件夾。

1.METS

METS（詮釋資料編碼和傳輸標準）是編碼描述，使用XML表達的管理性和結構性詮釋資料件(相對於我國推行的線上簽核電子封裝檔 xml，包含該文件檔案的詮釋資料及對應的 DTD 描述相關的欄位要求)。以下是METS文件的概述。

Elements				Values	Comments
mets	metsHdr				
		agent			software or archivist creating the DIP
	dmdSec				
		mdRef		EAD	Information about the EAD file
	amdSec				
		mdRef		PREMIS	Information about the PREMIS file
	fileSec				
		fileGrp		Common Specification root	
			fileGrp	metadata	
			fileGrp	representations	normally only one repr. in the DIP
			fileGrp	schemas	
			fileGrp	documentation	
	structMap				
		div		metadata	
		div		representations	mets pointer to mets file for the repr.
		div		schemas	
		div		documentation	

圖2 METS資料夾架構

METS配置文件中有5個主要部分：

- <metsHdr> - METS標頭（關於IP的創建者，聯繫人等的詮釋資料）。
- <dmdSec> - 描述性詮釋資料。
- <amdSec> - 管理性詮釋資料（有關如何創建和存儲文件的信息及知識產權等）。
- <fileSec> - 文件單元，列出所有文件包含的內容（也可能包含有關文件的詮釋資料）。
- <structMap> - 結構圖，描述了數位物件和所有封包（即物件+詮釋資料）的層級結構。

2.PREMIS

PREMIS為適用於長期保存及技術可用性的標準。

從存取的角度來看，PREMIS符合紀錄陳述訊息資訊的相關要求。在不同訊息的時候Access軟體如何表示，例如DIP包含那些陳述訊息格式。因此E-ARK Access軟體假設根據以下規範可使用PREMIS詮釋資料。

I.關於陳述訊息及Access軟體的詮釋資料

在PREMIS中，陳述訊息包括結構性詮釋資料，需要完整描寫智能實體。

此外，當在E-ARK中執行PREMIS時必須選擇是否PREMIS存在陳述訊息階段或是根目錄階段。

在PREMIS陳述訊息以單元”objectIdentifier”來表示。

為了確保有效呈現，PREMIS需要三個訊息:呈現陳述訊息辨識、確保軟體辨識及建立兩者之間的關係。

如下說明：

1. 描述DIP中所包含的陳述訊息格式。
2. 描述使用哪些Access軟體呈現特定的DIP陳述訊息格式。
3. 描述Access軟體及DIP陳述訊息之間的關係。

i. 說明1-紀錄DIP陳述訊息格式

為了描述特定的DIP表示格式，使用語義元件

“importantProperties”。 以下範例：

注意物件類型為陳述訊息及objectIdentifierType值為

“filepath”。

```
<object xsi:type="representation">
  <objectIdentifier>
    <objectIdentifierType>filepath</objectIdentifierType>
    <objectIdentifierValue>xlink:href="representations\AVID.SA.18006.rep0"</objectIdentifierValue>
  </objectIdentifier>
  <significantProperties>
    <significantPropertiesType>DIP representation format</significantPropertiesType>
    <significantPropertiesValue>SIARD2</significantPropertiesValue>
  </significantProperties>
  <!-- PREMIS file continues but further elements left out in this example-->
```

圖3 DIP陳述訊息內容

ii. 說明2-紀錄Access軟體

在PREMIS 3.0中對環境的描述為物件本身。單元

environmentFunction為描述不同階段的環境物件。單元

environmentDesignation通過使用人類語言辨識訊息達成在數

位存取庫之外被理解。範例：

```

<object xsi:type="intellectualEntity">
  <objectIdentifier>
    <objectIdentifierType>local</objectIdentifierType>
    <objectIdentifierValue>DBVTK</objectIdentifierValue>
  </objectIdentifier>
  <environmentFunction>
    <environmentFunctionType>software</environmentFunctionType>
    <environmentFunctionLevel>1</environmentFunctionLevel>
  </environmentFunction>
  <environmentFunction>
    <environmentFunctionType>software application</environmentFunctionType>
    <environmentFunctionLevel>2</environmentFunctionLevel>
  </environmentFunction>
  <environmentDesignation>
    <environmentName>Database Visualization Toolkit</environmentName>
    <environmentVersion>2.4.1</environmentVersion>
    <environmentDesignationNote>Lightweight web viewer for relational databases, specially
if preserved in SIARD 2, that uses SOLR as a backend, and allows browsing, search, and export.
Documentation at github.com/eark-project/software/DBVTK</environmentDesignationNote>
  </environmentDesignation>
</object>

```

圖4 紀錄Access軟體

iii. 說明3:紀錄陳述訊息及Access軟體之間的關係

為了建立呈現DIP陳述訊息格式及Access軟體之間的連結，需要使用單元relationship。關係元素可綁定非環境物件及環境物件，且可將環境物件與其他環境物件綁在一起。以下範例顯示DIP表示格式如何與Access軟體相關：“。範例如下：

```

<object xsi:type="representation">
  <objectIdentifier>
    <objectIdentifierType>filepath</objectIdentifierType>
    <objectIdentifierValue>xlink:href="representations\AVID.SA.18006.rep0"</objectIdentifierValue>
  </objectIdentifier>
  <significantProperties>
    <significantPropertiesType>DIP representation format</significantPropertiesType>
    <significantPropertiesValue>SIARD2</significantPropertiesValue>
  </significantProperties>
  <!-- The following is the relation between the software and the DIP representation -->
  <relationship>
    <relationshipType>dependency</relationshipType>
    <relationshipSubType>requires</relationshipSubType>
    <relatedObjectIdentifier>
      <relatedObjectIdentifierType>local</relatedObjectIdentifierType>
      <relatedObjectIdentifierValue>DBVTK</relatedObjectIdentifierValue>
    </relatedObjectIdentifier>
    <relatedEnvironmentPurpose>render</relatedEnvironmentPurpose>
  </relationship>
</object>

```

圖5 紀錄Access軟體

由上圖可看出關係的性質使用<relationshipType>如dependency，個人關係使用<relationshipSubType>如requires。為了辨識呈現陳述訊息的Access軟體，則使用<relatedObjectIdentifier>及<relatedEnvironmentPurpose>提供提示。

由於無法一直使用同個Access軟體呈現DIP陳述訊息格式，因此必須在PREMIS中的幾個軟體建立關係及序列。

3.EAD

描述性詮釋資料用於描述存檔的相關內容，並支持發掘及理解各訊息封包。E-ARK項目允許封包內含有描述性詮釋資料，如下圖所示

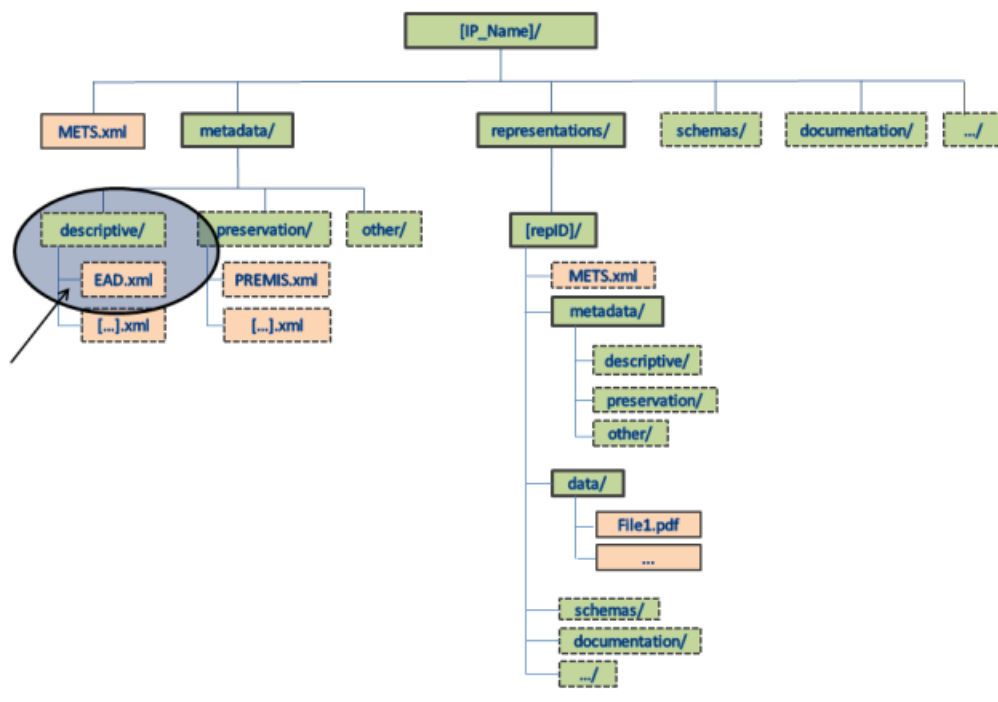


圖6 E-ARK封包描述性詮釋資料

EAD文件包含三個主要內容：

- (1)檔案描述:包含主要的檔案描述(包含組合及類別的詮釋資料)。

(2) 內容:包含鏈結<dao>元素及 @base屬性類別的文件及資料夾。

(3) 附加詮釋資料:不適合EAD標準元素的特定資訊可以儲存於EAD元素內的<odd> 元素或localtype屬性。

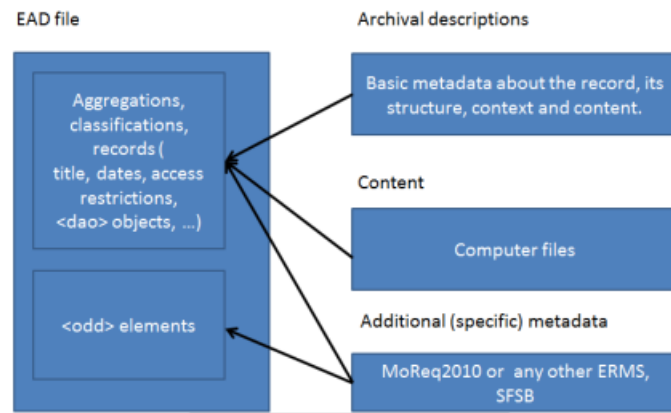


圖7 EAD文件

I. 工具

E-ARK項目構建的工具在默認情況下只能對應EAD3，因為EAD3是E-ARK選擇的描述詮釋資料標準。但這些工具也可以處理其他類的描述性詮釋資料。

E-ARK中使用EAD的工具如下：

- ERMS Export Module (EEM)
- EAD Editor
- RODA
- ESS tools
- E-ARK Web including Apache Hadoop, Lily and Solr.
- Access Software
 - Search GUI
 - Order Management Tool

○ DIP Viewer

II. 分層檔案描述

元件標籤(<c>)是描述特定物件附屬部分的元素。將使用無編號的元件標籤(<c>)來描述層級檔案描述(描述單位)。

Access軟體開發用於可以解決檔案集合內所有層級階段的描述，由此提供使用者有效描述於詮釋資料文件內的所有智慧內容階段資訊。

下面提供如何描述IP顯示器內呈現層級階段的描述：

EAD使用聚合值作為元素<archdesc> 及<c>的階層屬性來說明描述所屬的聚合級別。

```
<archdesc level="fonds">
  ...
  <dsc>
    <c level="series">
      ...
      <c level="file">
        Records and computer files
      </c>
    </c>
  </dsc>
</archdesc>
```

圖8 使用“level”屬性指定聚合級別的EAD範例

聚合級別的確切名稱取決於數據生產者及歸檔間的協議。EAD為此目的定義一組值(class, collection, file, fonds, item, otherlevel, recordgrp, series, subfonds, subgrp, subseries)，但如果使用@otherlevel屬性定義則允許使用其它值。

```

<archdesc level="collection">
  <did>
    <abstract>...</abstract>
  </did>
  <dsc>
    <c level="series">
      <did>
        <abstract></abstract>
      </did>
      <c otherlevel="case"> <!--A new aggregation Level-->
        <did>
          <abstract>Records and computer files from ....</abstract>
        </did>
      </c>
    </c>
  </dsc>
</archdesc>

```

圖9 使用<otherlevel>的EAD範例

為了能夠從EAD文件中引用實際智能實體，使用兩種機制：數位檔案物件<dao>標籤及@base屬性。

@base使用於引用文件夾，<dao>使用於文件。

這些機制的目標為：

將查找援助中的結果連結到訊息封包內特定的描述性單元。

允許E-ARK Access軟體使用者介面內的數據及EAD詮釋資料適當可視化。

數位檔案物件<dao>標籤是用於連結產生數位檔案或被描述的數位陳述訊息的子元素。<did>為封包元素，其包括辨識檔案的必要訊息。

III. Representations 陳述訊息

無論陳述訊息何時創建，都必須創建新的EAD文件且包含新的陳述訊息的文件項目。其替代了其中具有日期的舊文件。訊息封包仍然包含一個有效的EAD文件(在‘metadata/descriptive/’資料夾)並且引用訊息封包裡的所有陳述訊息。

IV. 存取限制

<accessrestrict>標籤是“有關影響物件描述可用性的信息元素條件”。關於最終用戶的訪問權限信息必須在EAD中可用，而不是在PREMIS中，而<accessrestrict>用於此目的。原因如下：

- i. 應該可以在一個地方找到訪問權限信息，即描述性詮釋資料，默認情況下於存取軟體（查找幫助和不同的查看者）中顯示的詮釋資料。
- ii. EAD支持在封包訊息內複雜層級的潛在描述，因此如果必要，將訪問限制降到單個文件級別。
- iii. 描述性詮釋資料通常在採集和查找後可以隨時添加，因此可以立即填充此類信息。

4. 陳述訊息資訊

陳述訊息資訊為了解及呈現數據及詮釋資料的資訊。

陳述訊息可以細分為三類：

結構訊息:描述應用於位元流的格式及數據結構概念，從而產生更有意義的值，如符號或像素。

語意訊息:結構訊息上需要此點。如果數位物件是將結構訊息解釋為字元符號序列，則語意訊息應包含正在表達何種語言的細節。

其他陳述訊息資訊:包含相關軟體、硬體、儲存媒體、加密或壓縮演算法等訊息。

在E-ARK項目中以不同的方式涵蓋陳述訊息，例如使用作為容器呈現的文件資料夾，可以放置在預攝取階段收集所有信息，以幫助檔管人員和消費者更了解存儲在封包訊息中的數據。該文件滿足從生產

者到歸檔間轉移檔案材料的要求。

透過遵循此策略可以建立一個遵守當前規範的陳述訊息資訊網路，因此我們實際上可以利用現在項目正在開發的保存措施工具來保護陳述訊息資訊。此外此種標準方式交換陳述訊息資訊將增加機構、企業、保存專家間的協作同時降低外部風險。

二、資料庫規範: SIARDDK、SIARD1 及 SIARD2

(一)SIARD 版本、SIARD 及格式結構

使用的 SIARD³版本可以是 SIARD 1.0，SIARDDK⁴或 SIARD 2.0，由 E-ARK 推薦使用版本於關係資料庫。SIARD 格式由 SIARD 存檔文件和 SIARD 存檔文件外的相關文件（不適用於 SIARD 1.0）。這些相關文件引用 SIARD 存檔文件中的 SIARD 表，並且列於信息封包的 METS 文件中。SIARD 歸檔文件和 SIARD 歸檔文件外的相關文件在下文中稱為 SIARD 數據庫。

(二)詮釋資料

1. PREMIS

除了 SIARD 文件中的標準 E-ARK DIP 和 SIARD 元數據外，還有以下內容應使用 PREMIS 語義單位存儲信息：

- (1) DIP SIARD 的詮釋數據是否從 AIP 中的原始資料庫或修改版本產生。如果是從修改成的，那麼與修改的詮釋資料相關。

³ Software Independent Archiving of Relational Databases，簡稱 SIARD，典藏關連式資料庫軟體。

⁴ SIARDDK 是自 2010 年以來丹麥所使用的格式。

- A. 如果 SIARD 封包是從原始資料庫生成的，則由
<creationApplicationExtension> 或 <formatNote> 顯示。
- B. 如果其為原始資料庫的修改版本生成的，則透過
<creationApplicationExtension> 或 <formatNote> 來
顯示。在這種情況下，詮釋數據是必要的，因為
此物件是從原始衍生，所以陳述訊息之間的關係
也必須如下給出：
- relationshipType： “派生”
 - relationshipSubType： “去規範化”

2. EAD

為了在 EAD 中描述資料庫，建議將屬性級別的值設
為 “otherlevel”，並將其他級別的屬性值設置為
“database”，最高層級的資料庫描述處於如下示例：

```
<archdesc level="otherlevel" otherlevel="database" lang="eng">
  <did>
    <unittitle>Northwind database</unittitle>
    <dao daotype="borndigital" href="/representations/AVID.SA.18006_rep0/data/northwind.siard"></dao>
    <unitdate>2000-2008</unitdate>
    <abstract lang="eng">Northwind Traders Access database is a sample database from Microsoft Office
suite. The Northwind database contains the sales data for a fictitious company called Northwind
Traders, which imports and exports specialty foods from around the world. The database contains 13
tables and also includes pictures of foods and employees. </abstract>
    <!-- EAD file continues but further elements left out in this example-->
  </did>
</archdesc>
```

如果有相關表格，屬性級別的值應設為 “otherlevel”，並且該
值在其他級別的屬性應設置為 “table”：

```
<c level="otherlevel" otherlevel="table">
  <did>
    <unittitle>Orders</unittitle>
    <abstract>Table0. This table contains information about the orders that Northwind Traders
had</abstract>
  </did>
</c>
<c level="otherlevel" otherlevel="table">
  <did>
    <unittitle>Products</unittitle>
    <abstract>Table1. This table contains information about the products that Northwind Traders
had</abstract>
  </did>
</c>
```

(三)存取腳本

第一步是搜索使用者要求的信息封包資料。使用者可能發現一個或多個相關的 AIP。則需要考慮以下問題：

1. 敏感數據、版權或其他法律規定的存取限制。
2. 使用者需要哪些信息封包的部分？檔案管理員須根據用戶的具體要求作出決定，以確保使否將信息封包整個資料內容傳遞給用戶，或者只足以傳遞信息封包的一部分）。
3. 必須創建多少個 DIP 才能滿足使用者的要求？
4. 資料內容、相關詮釋資料以及 DIP 的必要文件如何呈現給使用者。必須選擇適當的工具提供資料給使用者。

基於上述因素，最簡便的方式是將 AIP 的資料內容傳送給使用者，資料庫可以使用者友好介面呈現。為此資料庫視覺化工具可以滿足此需求。

在多數情況下，產生 DIP 是複雜的。

1. 在找到相關 AIP 後，檔管人員須決定哪部分的 AIP 生成 DIP 的存取限制及使用者要求。包含如下

(1) 選擇相關資料庫

- (2) 選擇表的相關紀錄
 - (3) 選擇表的相關欄位
 - (4) 如果需要，更改/移除相關敏感性資料
 - (5) 選擇適當的工具修改，如 E-ARK DB Viewer，在多數情況下存儲在 AIP 中的 SIARD 資料庫必須從 AIP 中提取，並使用 E-ARK DBPTK⁵導入 RDBMS⁶。之後新的 SIARD 資料庫可透過 DBPTK 從 RDBMS 產出。
2. 在某些情況下，必須修改資料模型已達成可理解或允許複雜得分析。將 SIARD 數據庫導入到 RDBMS 中之後，可使用 SQL 指令修正。修改後的資料庫如下呈現：
- (1) 產生新 SIARD 資料庫的視覺化資料庫標準工具。
 - (2) 透過特殊工具直接連結到 RDBMS(使用 Quest Software 的 SQL Navigator 或丹麥國家檔案館開發的 Sofia 等工具)。
3. 某些情況下需要開發新的呈現工具（例如 Oracle APEX 申請特殊報告）。
4. 在具體的情況下，RDB 可以轉換為資料倉庫（DW），因此可以基於它創建 OLAP 資料集，例如 Oracle Business Intelligence 等工具可能提供分析原始 RDB 資料。

三、資料倉儲: OLAP cube

(一)數據倉儲

⁵ Database Preservation Tool Kit 簡稱 DBPTK，資料庫保存工具包。

⁶ Relational Database Management System 簡稱 RDBMS，關聯資料庫管理系統。

創建和使用 OLAP 物件的方式需要一個數據倉儲，數據倉儲有不同類型的數據來源：關係數據庫（RDB）、電子表格、統計文件格式（SPSS，SAS 等）、XML 格式的數據等。這些數據將被提取並導入到暫存區域，並且從那裡將數據轉換並加載到特定類型的第三正常格式 RDB（關係資料庫）和“維度模型”。

1. OLAP Cube 線上分析處理立方體

線上分析處理 (On-Line Analytical Processing，簡稱 OLAP) 立方體為多元資料集，通常來自資料市場。OLAP Cube 的定義及創建無統一交換格式。因此，OLAP 立方體定義，OLAP 立方體的供應商特定的輸出格式不可在不同的 OLAP 應用程序之間互換。

2. 數據倉儲儲存為 RDB 於 DIP SIARD 格式中

E-ARK 制定數據倉庫階段之後，所有完成的數據內容將以資料集/資料庫呈現，即去標準化的數據被導出為 SIARD 格式，所有具體信息將存儲在 DIP 中的文件資料夾。

(二) 實體結構

含有關係數據庫 SIARD 格式的 DIP 資料夾結構完全符合 E-ARK 信息封包的文件夾結構。

在訊息封包含有多個數據庫的陳述訊息的情況下，需使用 PREMIS 詮釋資料來表示陳述訊息間的差異。

不同的陳述訊息之間的關係可用語義單位”relationship”說明。

該語義單元 “relationshipType” 和 “relationshipSubType” 作為語義元件。有關使用 PREMIS 的建議，兩個元件值應來自受控詞彙。

(三) 詮釋資料

1. METS

協議是內容信息類型的每個作者提供 xml 文件範例的鏈接。

2. PREMIS

除了標準的 E-ARK DIP 和 SIARD 2.0 詮釋資料外，還應儲存以下信息：

- (1) 詮釋資料實際陳述訊息為 SIARD 封包（具有多種類型的文件），或者是於原始軟體環境中整個原始數據倉儲的模擬。該信息應由 PREMIS 語義單位提供 <objectCharacteristics> 或 <format>。如果 DIP 中存在多個陳述訊息，則必須使用語義單元 <relationship> 來指定它們之間的關係。
- (2) 詮釋資料於實際的 SIARD 數據庫是從修改後的模式生成還是從原始數據庫模式生成。如果是從修改後生成的，那麼與該修改的詮釋資料相關。如果 SIARD 封包是從原始數據庫生成的，則由 <creatingApplicationExtension> 或 <formatNote> 顯示。在這種情況下，詮釋資料是可選的。
- (3) 關於 DIP 中的 OLAP 物件的信息可以由 PREMIS 語義單元 <environmentDesignationNote> 顯示。但是，用

戶必須具備所有必要信息根據存檔的 RDBMS 來重建特定的 OLAP 對象，這樣信息可能包括文件，特定的定義和導出文件等。因此，關於這些文件的作用的所有必要信息也應該在語義單元 <environmentDesignationNote> 中提供，或者在一個 “starting point document” 中提供，這將在 <environmentDesignationNote> 中被引用，但被放置在 “Documentation” 文件夾。

(四)存取腳本

可以通過 E-ARK DBVTK⁷將數據倉庫（無 OLAP 物件）呈現 SIARD 數據庫。也可以使用 DBPTK 加載到 RDBMS 中，並由供應商特定的應用程式（如 Oracle BI，Oracle BI 數據可視化）或內部開發的應用程式存取。

存取及呈現 OLAP 物件相當複雜。首先，SIARD 資料庫包含資料，必須加載到 RDBMS 中。之後，可能會發生以下情況：

1. 特定供應商特殊 OLAP 定義文件存儲於 DIP（例如，Oracle 維度物件定義可以導出/保存在 XML 模式或 EIF 格式中）。

(1) 定義文件必須導入到特定的 RDBMS 中。在 Oracle 的情況下可以做到：

A. 透過使用 PLSQL 封包，而定義文件（name）為封包參數。

B. 透過使用 Oracle Analytic Workspace Manager。

⁷ Database Visualization ToolKit 簡稱 DBPTK，資料庫可視化工具包。

- (2) 需使用供應商特定工具對 OLAP 資料集執行可能/必需的分析流程。在 Oracle 的情況下，此具體工具是 Oracle BI。

2. DIP 僅包含產生(定義)OLAP 物件(在文檔資料夾中)所需的描述和文件。這些描述可以是供應商特定的，但也可以是通用的，不需要指定的軟件。

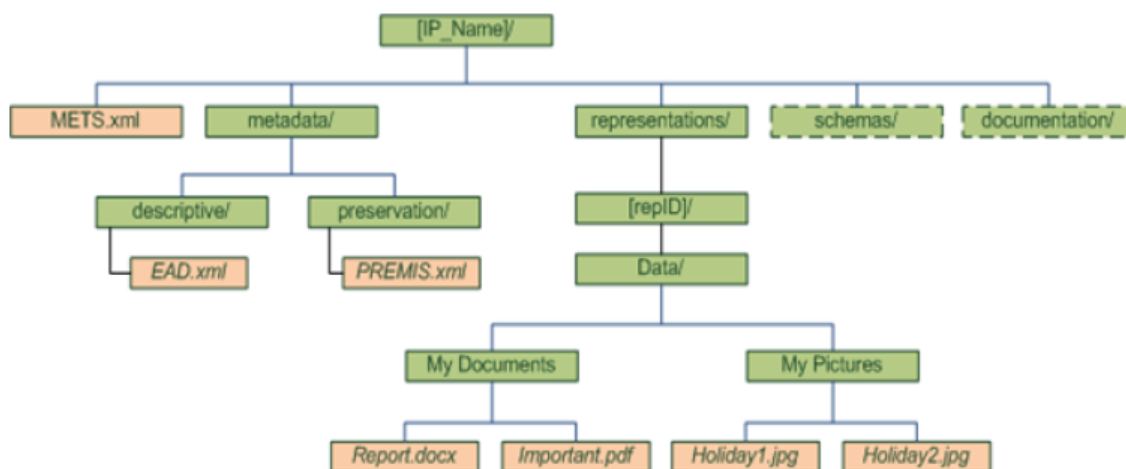
當資料倉儲及分析工具為 AIP 的一部分情況下，資料可以在原始環境傳送給使用者，資料倉儲可以在原始環境中提供給使用者。

- (1) 模擬環境需恢復到原始的硬體及作業系統。
- (2) 需確認存於資料倉儲中的資料取用限制。如果需要，檔案管理員必須控管資料倉儲中的資料進行。當用戶不需要整個資料倉儲時，則可以刪除。
- (3) 使用者可以取用內容並使用原始工具分析資料倉儲。

如果只是一個存於 AIP 中的資料倉儲的模擬版本，在這種情況下這些傳遞給使用者的部分資料倉儲，可以在原始環境或外部創建。在另種情況下，資料倉儲的資料需從原始資料倉儲導出，資料倉儲需要 OLAP 的物件並於新系統創建。

四、簡易文件系統基礎檔案

(一)實體結構



簡易文件系統基礎檔案如上圖，強制性要求結構如下：

1. DIP 只能包含一個陳述訊息(即在“representation”資料夾內，具有 id 或 mane 的資料夾)。
2. DIP 的所有詮釋資料需再 root “metadata” 資料夾中(即“representation”資料夾中不允許使用詮釋資料)。
3. DIP “metadata” 資料夾的 “descriptive” 資料夾中包含一個 EAD.xml 文件。
4. DIP “metadata” 資料夾的 “preservation” 資料夾中包含一個 PREMIS 文件。

(二) 詮釋資料

1. METS

SMURF⁸ SFSB⁹ 軟件包必須在 DIP 的根階層中包含一個 METS.xml 文件。這代表陳述訊息資料夾中不允許有更多的 METS 文件。

⁸ Semantically Marked-Up Records Format simple file-system based records 簡稱 SMURF，語意標記記錄格式。

⁹ Simple file-system based records 簡稱 SFSB，簡單文件系統檔案，包含基礎文件系統的文件夾或文件。

與 E-ARK 通用規格相比，這代表：

- (1) METS 文件的<fileSec>元素必須列出陳述訊息資料夾的所有元件(在適當的<fileGrp>和<file>子元素中)。
- (2) METS 文件的<structMap>元素必須列出陳述訊息資料夾的所有元件，而不使用<mptr>子元素。

METS 文件需使用 SMURF SFSB 內容信息類型規範。透過將<mets CONTENTTYPESPECIFICATION> 屬性值設為“SMURFSFSB”。

```
<mets xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xlink="http://www.w3.org/1999/xlink" xmlns="http://www.loc.gov/METS/"
PROFILE="http://www.eark-project.com/METS/IP.xml" TYPE="ERMS"
CONTENTTYPESPECIFICATION="SMURFSFSB" OBJID="5d378f86-28a1-41d8-a2b9-264b10fbd511"
LABEL="METS file describing a DIP with simple computer file content."
xsi:schemaLocation="http://www.loc.gov/METS/ schemas/IP.xsd http://www.w3.org/1999/xlink
schemas/xlink.xsd">
```

2. PREMIS

SMURF SFSB DIP 必須包含 PREMIS 詮釋資料文件，其中包含適用於記錄保存事件的訊息以及與呈現給終端使用者查看和使用檔案的環境相關訊息（如 DIP viewer）。

建議將所有 PREMIS 詮釋資料格式化為單一詮釋資料文件，也會嵌入所有技術詮釋資料。遵循此建議將增加封包的可讀性。

3. EAD

應用 SMURF SFSB 配置文件的必要先決條件是使用 EAD 創建簡單的歸檔描述。在最常見的情況下，此 EAD 描述將包括

幾個聚合及簡單描述內容。有幾個 EAD 元素被認為是強制性的，以允許開發的通用訪問工具，強制性元素如下：

- (1) 每個聚合的標題（element did / unittitle）。
- (2) 檔案或聚合的創建日期。

完整範例如下：

```
<?xml version="1.0" encoding="utf-8"?>
<ead xmlns="http://ead3.archivists.org/schema/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xsi:schemaLocation="http://ead3.archivists.org/schema/ ../../schemas/ead3.xsd">
  <control>
    <recordid>f323992d-bfa0-4586-bc1b-090d06ec6a72</recordid>
    <filedesc>
      <titlestmt>
        <titleproper>Personal letters of Mr Private Person</titleproper>
      </titlestmt>
    </filedesc>
    <maintenanceagency>
      <agencyname>National Archives of Utopia</agencyname>
    </maintenanceagency>
    <maintenancehistory>
      <maintenanceevent>
        <eventtype value="created"></eventtype>
        <eventdatetime>2016-07-08T08:50:52.120000</eventdatetime>
        <agenttype value="human">text</agenttype>
        <agent>Utopian Recordsmanager</agent>
      </maintenanceevent>
    </maintenancehistory>
  </control>
  <archdesc level="series">
    <did>
      <unitid label="current">FNS.11-2</unitid>
      <abstract>Personal letters of Mr Private Person</abstract>
      <origination label="Creator">
        <name>
          <part>Mr Private Person</part>
        </name>
      </origination>
    </did>
  </archdesc>
</ead>
```

```

        </name>
    </origination>
</did>
<dsc>
    <c level="item">
        <did>
            <unitid locality="current">FNS.112-2-1</unitid>
            <unittitle>Report 01</unittitle>
            <unitdate datechar="created">20.01.2008</unitdate>
            <abstract>Report No. 3</abstract>
            <physdescstructured coverage="whole"
            physdescstructuredtype="spaceoccupied">
                <quantity>0.0138</quantity>
                <unittype>MB</unittype>
            </physdescstructured>
            <dao id="c90fc089-1986-4473-8a9f-55d6a73ee9dc"
            linktitle="repA136.doc"
            href="file:../../representations/rep1/data/My
            Documents/important issue.doc" />
        </did>
    </scopecontent>
    <p>A document on an extremely important issue</p>
</scopecontent>
</c>
</dsc>
</archdesc>
</ead>

```

(三)存取脚本

在這種情況下，資料作為一組文件和/或資料夾結構在歸檔內進行管理，例如從硬碟歸檔。在這種情況下，產生較少的描述性詮釋資料。然而，我們假設有一個基本的歸檔階層可用，而依次使用 EAD 的核心元素進行描述。

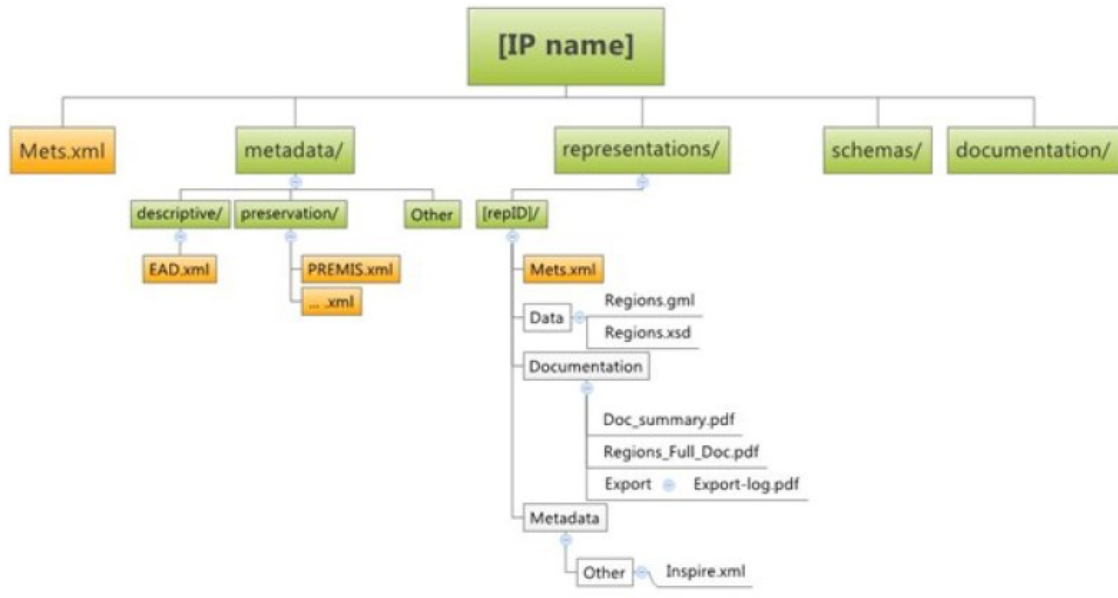
關於存取使用案例，使用者能夠執行全文搜索或瀏覽基本層次結構。一旦用戶找到合適的文件（通過全文搜索）或聚合（通過分層瀏覽或目錄中的搜索）。

五、Geodata 地理資訊

(一)實體結構

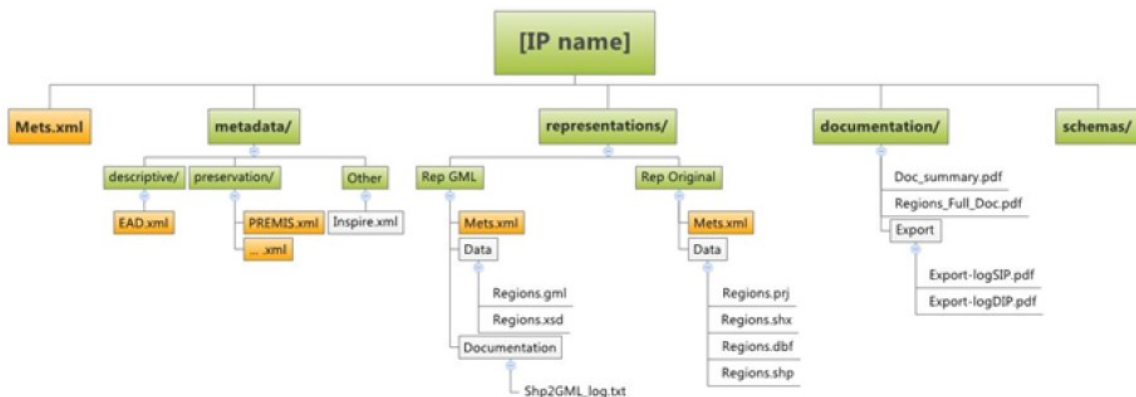
1. DIP 陳述訊息包含地理數據和文件

地理數據是空間中物件的圖形表示及其描述的組合屬性。地理資訊 - GML 的長期保存格式存放在陳述訊息中。地理數據的附加文件存放在陳述訊息資料夾。



2. DIP 於頂層包含多個向量表示和文檔

在這種情況下，DIP 封包含一種 GML 格式的陳述訊息形式，並以 ESRI Shapefile 格式包含相同數據的原始陳述訊息。使用 GML 文件的陳述訊息還包含特定陳述訊息的文件，例如從 shapefile 到 GML 的導出記錄檔。所有其他的文件放在頂層文件資料夾。



3. 陳述訊息目錄 - data 資料夾

陳述訊息目錄包含至少一個地理數據的陳述訊息，以及以 SFSB 描述的 SIP SMURF 格式的正確呈現信息所需的所有附加信息。

(二) 詮釋資料

1. METS

封包含有地理資訊則 DIP 根目錄中須有一個 METS.xml 文件。也就是說陳述訊息中不能有更多的 METS.xml 文件。

- (1) METS 文件的<fileSec>元素須列出陳述訊息資料夾的所有元件（在適當的<fileGrp>和<file>子元素中）；
- (2) METS 文件的<structMap>元素須列出陳述訊息資料夾的所有元件，而不使用<mptr>子元素；

METS 文件必須指示使用 GEODATA 內容信息類型規範。透過設置<mets CONTENTTYPESPECIFICATION>屬性值為“GEODATA”來表示。

```
<mets xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xlink="http://www.w3.org/1999/xlink" xmlns="http://www.loc.gov/METS/"
PROFILE="http://www.eark-project.com/METS/IP.xml" TYPE="ERMS"
CONTENTTYPESPECIFICATION="GEODATA" OBJID="5d378f86-28a1-41d8-a2b9-264b10fbd511" LABEL="METS
file describing a DIP with simple computer file content."
xsi:schemaLocation="http://www.loc.gov/METS/ schemas/IP.xsd http://www.w3.org/1999/xlink
schemas/xlink.xsd">
```

2. PREMIS

地理數據 DIP 必須包括一個 PREMIS 詮釋資料文件，其中包含記錄保存事件以及最終用戶查看和使用地理數據記錄所需相關環境的所有資訊。

3. EAD

當地理數據訊息封包的內容以影響位置精度的方式更改時，EAD 詮釋資料描述必須反映新的精確值。必須更新元素“Lineage”，並將元素“Spatial resolution”更新為新值。

在下面的情況下，行政單位的空間數據集被推廣。精度從 1：2880 變為 1：50000。

```
<odd>
  <head>Lineage</head>
  <p>This dataset was created by manually digitizing the original cadastral maps in various
    scales (1:1000, 1:1440 and 1:2880). The whole dataset was created by merging
individual
    units, that were digitised by different authors, since different municipalities had
    different service providers. The base datasets were created in the period from 1984 to
    1994 and were subsequently added to create the final dataset. Boundaries were
    generalised for a faster display.</p>
</odd>
<odd>
  <head>SpatialResolution</head>
  <p>50000</p>
</odd>
```

(三)存取腳本

1. 地理資訊搜尋

終端使用者可以在歸檔中的資料搜尋地理資訊。除了標準的 E-ARK 搜尋工具（全文搜索，目錄搜索）地理搜尋是唯一可用地理數據指定搜尋。地理搜尋需要具有空間索引的獨立工具，可作為執行地理搜尋功能的基礎。使用者可選擇想要的區域，或可以使用過濾的方式來限制搜尋結果。地理網絡服務允許終端使用者設置不同的過濾器來優化搜索結果，

所有搜尋都會導致終端使用者訂購一個或多個資訊封包，然後取用 DIP 地理數據。由於地理數據需要特殊的地理數據庫查看器，因此出現了特定 ACCESS 的情況。

2. 取用地理資訊

檔案館可以提供兩種不同方式取用地理資訊：非結構化文件；並透過 Web 服務呈現地理資訊。取用地理資訊的方式取決於使用者處理原始地理資料集的能力及其對地理資料工具知識（QGIS 等）。

六、結論

為了有效地使用詮釋資料來支持歸檔功能，本文介紹DIP為各種詮釋資料功能的容器。由於各國機關歸檔欄位不盡相同，檔案機構可參考本篇內容提到的欄位定義制定提交檔案欄位規範，使得檔管人員可有效歸檔避免檔案流失，以達成檔案機構想要有效管理檔案。

七、資料來源

(一)<http://www.dashboard.eu/specifications/dip>

(二)<http://www.eark-project.com/>

(三)<http://www.dashboard.eu/content-types/smurf>