

## E-ARK SIP 提交信息封包介紹

E-ARK Submission Information Package(提交信息封包，簡稱 SIP)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。

### 一、 SIP 結構與數據模型

SIP數據模型描述提交檔案的封包，該封包由陳述訊息<sup>1</sup>（提交的數據和詮釋資料）和詮釋資料組成，如圖1所示。

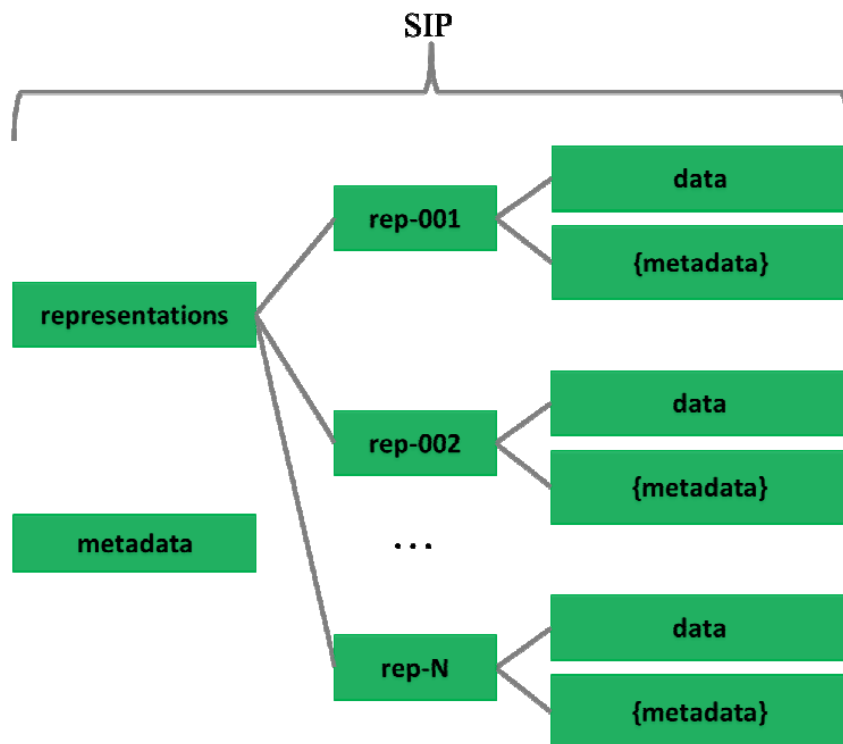


圖1 SIP 資料模型

註<sup>1</sup>: Representations 陳述訊息為儲存文件及結構性詮釋資料的數位物件提供智能實體的描述  
<http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>

由於 SIP 可由多個陳述訊息組成，如圖 1 的 Rep-001 和 Rep-002。如果在資訊封包階段儲存所有詮釋資料，則不需要在陳述訊息層使用詮釋資料。在這種情況下，陳述訊息內的詮釋資料目錄是可選擇的，同樣如下：

- 文檔文件夾 - 用於包含說明內容或其用途的文件（例如使用手冊）。
- 模式文件夾-詮釋資料目錄中 XML 文件的模式直接加入到封包。

信息封包文件夾必須包含一個名為“METS.xml”的強制性核心詮釋資料文件(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)，其中包含核心信息，用於識別和描述封包本身及其它元件的結構。因此最小的 SIP 結構如下（圖 2）。陳述訊息的名稱可以自由選擇。

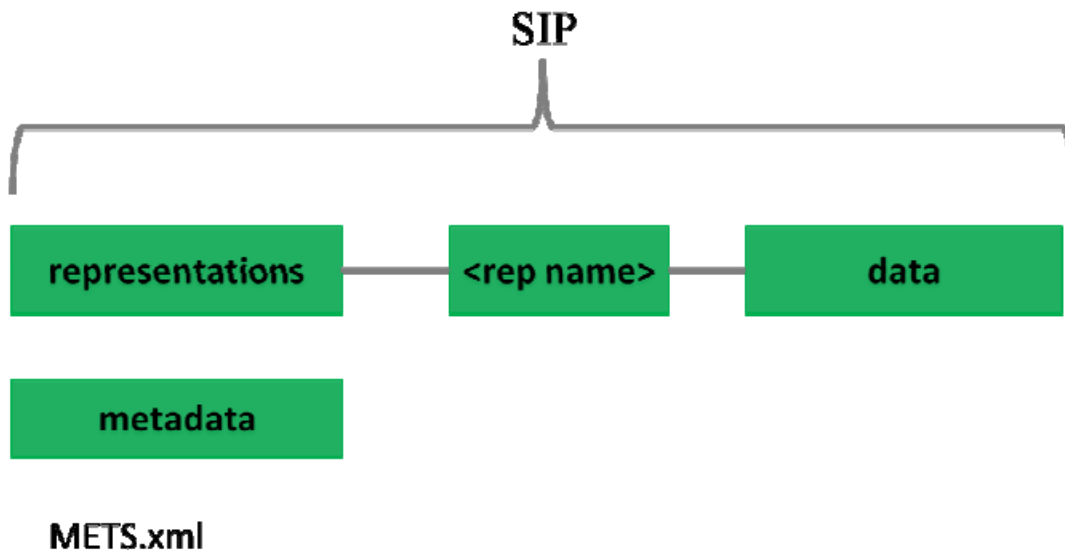


圖2 SIP 最小結構

## 二、SIP 詮釋資料

SIP 詮釋資料基於 METS<sup>2</sup> 標準，並作為配置文件呈現。METS 旨在詳細描述檔案內容，提供檔案作者和程序員創建符合以下條件的 METS 檔案格式所

<sup>2</sup>: Metadata Encoding and Transmission Standard 詮釋資料編碼和傳輸標準，為 XML 的檔案格式標準

需的準則。

METS配置文件中有5個主要部分：

- <metsHdr> - 標題（關於訊息封包的創建者，聯繫人等的詮釋資料）。
- <dmdSec> - 描述性詮釋資料（引用EAD，EAC-CPF）。
- <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> - 結構地圖，描述了數位物件和所有封包（即物件+詮釋資料）的層級結構，將物件分層呈現。

所有這些部分按照以下結構以表格形式呈現METS元素的SIP要求：

- 元素 - 元素或屬性在附帶模式中使用的元素名稱。
- 定義 - 定義元素的功能。 包含元素的說明及範例值。

### (一) 根目錄

METS文檔根目錄包含一些可選擇的屬性，命名空間 (xmlns :) 和引用METS紀錄的綱目範例位置 (xsi :) 如表1所示。

表1 根目錄

元素	定義	解釋
Identity/ Content ID 身分/ 內容ID	封包識別	唯一辨識SIP和提交的數位物件之代碼。使用UUID或GUID來創建唯一識別碼。 範例: "UUID:550e8400-e29b-41d4-a716-446655440004"
Description/P ackage name 描述/封包 名稱	封包描述	簡短文字描述封包 範例: "SIP for delivery of personnel information"
Content type /General	封包類型	必須使用TYPE屬性來識別封包類型，例如ERMS，RDBMS，數位建立計劃。

content type 內容類型/ 一般內容類 型		範例: ”ERMS”
Profile 輪廓	配置文件名稱	描述使用的METS文件。該名稱應包含在文件路徑中的版本號。 ” http://eak-project.com/resources/METS/v02/METS.xml”
Content type Specification 內容類型 規範	用於內容類型的規範	依此規範添加屬性。它描述使用內容的信息類型規範。屬性值固定於以下詞彙表中： 1. SMURFERMS 2. SMURFSFSB 3. SIARD1 4. SIARD2 5. SIARDDK 6. GeoVectorGML 7. GeoRasterGeotiff 隨著額外的內容信息類型規範的開發，詞彙是可擴展的。 Example: “SMURFERMS”

## (二) Header標題

標題<metsHdr>描述提交信息封包的創建者，聯繫人等的詮釋資料，如圖3所示。

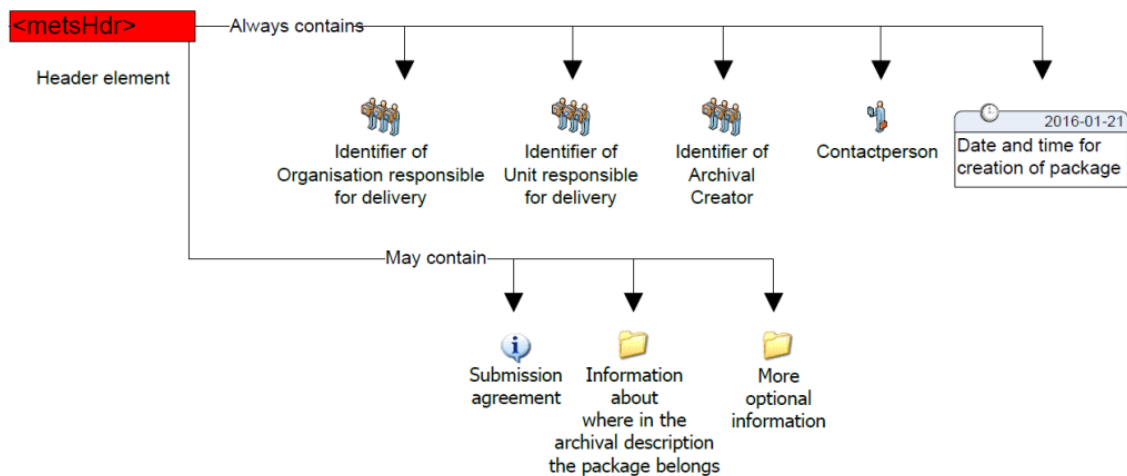


圖3 標題

表2 訊息封包詮釋資料

元素	定義	解釋
Id	METS標題部分的唯一ID	METS標題部分的唯一ID 範例: "ac0e8400-e29b-41d4-a716-446655440004"
Date and time 日期時間	封包創建時間	根據XML標準描述封包創建日期和時間。當封包及包封資料夾創建會被此時戳紀錄狀態。 範例: "2012-04-26T12:45:00+01:00"
Status 狀態	封包狀態	顯示封包狀態的方式，使其了解如何處理此封包 範例: "TEST"
OAIS type of package OAIS類型的封包	封包類型	資訊封包的類型。可能的值是SIP, AIP, DIP, AIU, AIC。
DocID	METS 文件ID	METS的唯一辨識碼。 範例: "SIP20150127.xml"
Submission agreement 提交協議	參考使用的提交協議	完全參考提交封包的提交協議。 範例: RA 13-2011/5329; 2012-04-12
Previous submission agreement 先前的提交協議	先前提交協議的信息屬於被記錄的情況。	如果信息可用，先前提交協議的信息屬於被記錄的。 範例: "FM 12-2387/12726, 2007-09-19"
Archival reference code 歸檔參考碼	描述檔案的參考代碼	可以給出一個參考代碼，指示封包檔放置的層級位置。 範例: SE/RA/123456/24/P
Previous reference code 以前參考代碼	描述檔案中較早使用的參考代碼	如果SIP源於其他機構的維護參考代碼結構，這個元素可用於記錄這些參考代碼，因此支持來源封包檔尚未提交時。 範例: "SE/FM/123/123.1/123.1.3"
Archival creator 檔案創建者	檔案創建者的名稱	數據的原始創建者（組織）的名稱。 範例: "The Swedish health agency"

Archival creator identification code 檔案創作識別碼	檔案創建者的識別碼	檔案作者的獨特識別碼。代碼使用字首根據 vcTypeOfIdentificationCode 範例: "VAT:SE201345098701"
Submitting organisation name 提交組織名稱	提交封包歸檔的組織名稱	將封包提交到存檔的組織名稱。用來說明和延伸信息。 範例: "The agency, Personnel"
Delivering organisation identification code 傳遞組織識別碼	傳遞組織的特殊識別碼	傳遞組織的特殊識別碼。 代碼使用字首根據vcTypeOfIdentificationCode. 範例: "VAT:SE2098109810-AF87"
Contact person Name 聯絡人名稱	提交聯絡人名稱	在提交時記錄時提交的聯繫人可能是有用的。 範例: "Sven Svensson"
Contact person contact information 聯絡人聯繫資訊	聯絡人的連絡資訊	聯繫人的電話號碼和電子郵件。 範例: "08-12 34 56" "sven.svensson@fm.se"
Software 軟體	使用於創建封包的軟體	metsHdr必須至少包含一個描述軟件的代理程序，該代理程序用於創建軟體封包。
Preservation organisation name 保存組織名稱	保存封包檔的組織名稱	負責保存提交封包檔的組織名稱。 範例: "National Archives of Hungary"
Preservation organisation identification code 保存組織識別碼	保存組織識別碼	負責保存的組織的特殊識別碼，代碼使用字首根據 vcTypeOfIdentificationCode 範例: "ORG:2010340987"

### (三)Descriptive metadata描述性詮釋資料

描述詮釋資料<dmdSec>引用的檔案描述詮釋資料（如EAD，EAC-CPF等），如圖4所示。

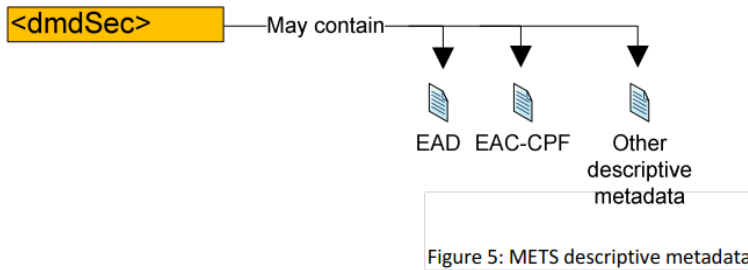


圖4 敘述詮釋資料

檔案信息可以包含在METS封包中。通常這些信息以EAD<sup>3</sup>和EAC-CPF<sup>4</sup>格式提交。

根據上述提到的METS執行指南，使用<dmdSec>，將EAD和EAC-CPF包含在METS配置文件中。配置文件中列出<dmdSec>裡所有元素和屬性的完整規則，參考和嵌入時使用的特定元素如下所示。

表3 EAD 詮釋資料

元素	定義	解釋
EAD metadata file EAD詮釋資料文件	METS文件中引用EAC-CPF格式的詮釋資料文件	在METS文檔中引用EAD格式的詮釋資料。另外還有唯一的識別碼（@ID），本節中的詮釋資料的創建日期（@CREATED）和必須記錄xlink:href中使用的定位器類型（@LOCTYPE）。
EAC-CPF metadata file EAC-CPF詮釋資料文件	METS文件中引用EAC-CPF格式的詮釋資料文件	在METS文檔中引用EAC-CPF格式的詮釋資料。另外還有唯一的識別碼（@ID），本節中的詮釋資料的創建日期（@CREATED）和必須記錄xlink:href中使用的定位器類型（@LOCTYPE）。

註<sup>3</sup>: Encoded Archival Description 檔案描述編碼格式，為專門用於描述檔案的格式。

註<sup>4</sup>: Encoded Archival Context for Corporate Bodies, Persons, and Families 是使用 XML 描述檔案相關內容的標準。

#### (四) Administrative metadata 管理詮釋資料

管理詮釋資料<amdSec>引用技術和保存詮釋資料，如圖5所示。

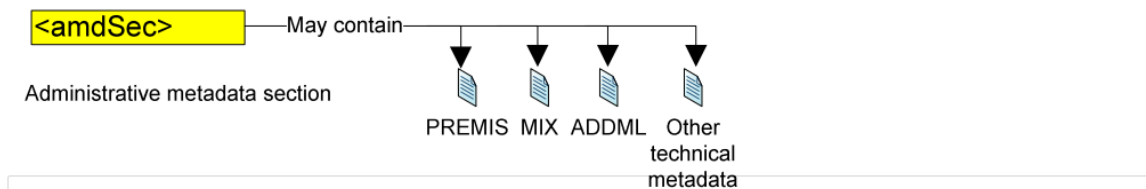


圖5 管理詮釋資料

METS封包中可以包含保存詮釋資料。建議使用PREMIS<sup>5</sup>保存詮釋資料。

表4 PREMIS詮釋資料

元素	定義	解釋
PREMIS metadata file PREMIS 詮釋資料文件	引用METS文的PREMIS格式詮釋資料檔	當PREMIS格式的詮釋資料檔引用METS文件和嵌入時需在提交協議中註明。

註<sup>5</sup>: PREservation Metadata: Implementation Strategies 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。



## (五) Files 文件

文件元素<fileSec>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），如圖6所示。

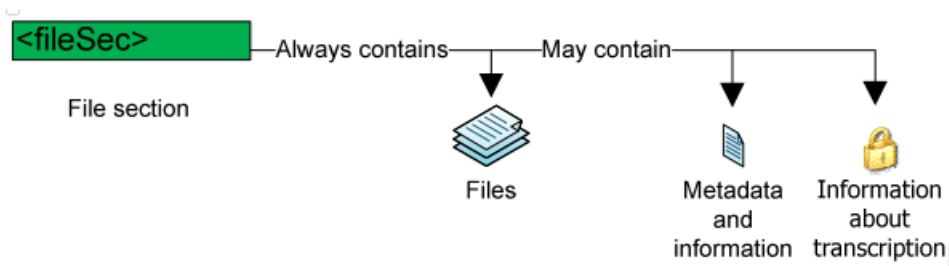


圖6 文件

所有在提交封包中找到的文件都應被引用一次在METS文檔中描述提交資料。元素和屬性皆以相同的方式的提交。當描述METS中的內容和文檔文件時，它們將被放置在一個或多個fileGrp的fileSec元素中。fileGrp元素可用於以不同的方式將文件分類在一起。在這個配置文件中，不會將文件分在不同的類別中，並且只使用一個必要的fileGrp。

表5 文件詮釋資料

元素	定義	解釋
Identification of the file 鑑定文件	識別文件物件	引用structMap元素的METS文件特殊辨識代碼。建議依據UUID或GUID或本機識別碼使用字首“ID”。ID遵循xml的規則，屬性XML:ID。 範例: "ID550e8400-e29b-41d4-a716-4466554400bg"
File location 文件路徑	文件名稱	文件的路徑及名稱 文件名稱必須在前面加上file://。該屬性LOCTYPE是描述如何查找文件並使用METS中值列表的必需條件。 範例: "file://personnelexport.xml"
Date and time 日期時間	文件的時戳	時戳可在文件上看到並用於驗證文件。多數情況下，為最後修改日期。使用xml類型的datetime規則描述。 範例: "2012-04-20T13:30:00+01:00"
MIME type 類別	最簡單描述文件類型的方式	最簡單描述文件類型的方式 範例: "text/xml"

File format name 文件格式名稱	文件格式名稱	如果在提交協議中尚未同意使用PREMIS時，需要給出更詳細的名稱。 範例： "PDF/A"
File format version 文件格式版本	文件格式版本	提交協議中尚未商定PREMISE的文件格式和使用版本。 範例： "1.0"
Format registry name 格式註冊名稱	格式註冊時的識別名稱	在提交協議中尚未同意使用PREMIS時的格式註冊識別的名稱。 範例： "PRONOM"
Format key 格式金鑰	註冊表的格式的金鑰	在提交協議中尚未達成一致時使用PREMIS時註冊的文件格式的金鑰。 範例： "fmt/101"
File size 文件大小	文件大小以bytes為單位	文件大小以bytes為單位 範例： "8765324"
Function 功能	標示文件的功能	有時是有助於記錄文件功能的訊息。 範例： "Submission file"
Checksum type 校驗類型	使用演算法創建及校驗	使用演算法創建及校驗，在METS預先定義值，應於提交協議中說明使用的算法。 範例： "SHA-256"
Checksum value 校驗值	計算校驗文件	檢查文件的總和。 範例： "574b69cf71ceb5534c8a2547f5547d"
Transformation algorithm 轉換演算法	轉換演算法	用於任何文件轉換（加密/解密）的演算法。此演算法使用於提交協議。屬性"TRANSFORMATIONTYPE"根據預設值設定轉換的類型。 範例： "DES"
Transformation key 轉換金鑰	文件轉換金鑰	轉換文件的轉換金鑰。屬性"TRANSFORMATIONTYPE"根據預先定義的值來表示轉換類型。

### (六) Structure結構

<structMap>描述了數位物件的層級結構，將物件分層呈現，如圖7所示，

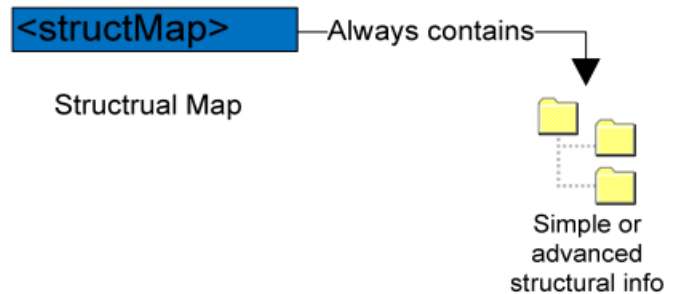


圖7 結構

### 三、 結論

隨著網際網路發展，政府、公司及金融機構也逐漸地將業務電子化，如何有效歸檔及取用則愈趨重要。

為了有效地使用詮釋資料來支持歸檔功能，本文單獨定義SIP METS為各種詮釋資料功能的容器，詮釋資料包含標題<metsHdr>描述提交信息封包的創建者，聯繫人等的詮釋資料，描述詮釋資料<dmdSec>用於描述詮釋資料標準，管理詮釋資料<amdSec>描述所引用相關保存技術及其他功能與標準。文件<fileSec>列出文件包含的內容。結構<structMap>描述了數位物件的層級結構，將檔案詮釋資料分為上述五個部分仔細描述。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。

#### 四、 資料來源

(一)<http://www.eark-project.com/>

(二)<http://www.dashboard.eu/specifications/sip>

(三)<http://eark.xorxor.hu/home>