

TWSMP 對應 EML 詮釋資料項目檢核驗證與擴充評估報告書

一、 國土資訊系統的詮釋資料標準 TWSMP 之簡介

國土資訊系統 (National Geographic Information Systems, NGIS) 為我國以地理資訊系統技術促進國家整體發展之跨單位機制，歷經多年發展，已累積大量具有空間特性之地理資料。基於資料分散儲存之特性，須由基礎建設 (infrastructure) 之觀點，建立國家級地理資料流通環境，以避免資料重複建置、加速地理資料流通、落實地理資料之多目標應用及促進業務單位應用之結合。達成上述目標之關鍵在於有效掌握國土資訊系統中之各類空間資源 (georesource)，並創造使用者查詢、取得、交換與應用空間資源的便利環境。國家級詮釋資料標準約制空間資源之描述內容與架構，有利於全國地理資料流通機制之發展與提升整合應用之成效。因應近年地理資訊系統技術之「開放性」及「互操作性」趨勢，詮釋資料標準之制定必須考量國際地理資訊系統標準與網際網路技術之發展現況，以善用標準技術之優勢，減低異質性資料流通之障礙。

詮釋資料是「資料的資料」 (data about data)，亦即針對現有資料之不同特性建立描述性資料，以協助現有資料之流通、解讀與應用。由國家級地理資料流通機制發展之觀點，詮釋資料可被廣泛應用於資料發現、資料適用性評估、資料存取、資料使用、資料傳輸轉換、資料管理等不同用途，主要包括以下之描述內容：

What：是什麼資料？ 資料有什麼內容？

When：資料代表之時間狀態？更新頻率或時間？

Who：誰建置資料？誰供應資料？誰提供資料之說明？

Where：哪裡可以取得資料？

How：如何取得資料？如何應用資料？

基於詮釋資料為全球地理資訊系統產業之共同需求，國際標準組織（International Organization for Standardization）制定編號為 ISO 19115 之標準文件，正式名稱為地理資訊－詮釋資料（Geographic Information－Metadata）。ISO 19115 標準為詮釋資料之基本架構（framework），可供各應用領域或國家制定詮釋資料標準之參考。基礎於 ISO 19115 標準而制定之各詮釋資料標準將具有共同之架構，方便搜尋及流通機制之發展。國土資訊系統地理資料詮釋資料標準（以下簡稱本標準）之設計基礎於 ISO 19115 標準，可視為 ISO 19115 標準之特定領域子標準(profile)，英文名稱為 TWSMP(TaiWan Spatial Metadata Profile)，適用於我國各類空間資源之描述。遵循本標準而建立之詮釋資料有利於我國各地理資料及地理資訊服務之流通交換。

詮釋資料標準之制定須依實際需求設定詮釋資料之角色，以設計合適之詮釋資料項目，並進一步規定各項目之名稱、資料型別、值域、內容及編碼方式，最後制訂為標準文件。本標準針對國土資訊系統地理資料流通之需求而設計，透過建置完成之詮釋資料，資料需求者可搜尋符合條件之現有資料、解讀資料之各類特性、評估資料之內容是否合乎需求、了解資料應用之限制及獲得如何處理資料之資訊。

二、 EML (Ecological Metadata Language) 詮釋資料標準簡介

在以往生態學研究中，研究人員各自收集自己研究相關的資料加以分析，進行較小規模或尺度的研究。但是在現今許多重大的生態與環境議題上，例如全球暖化問題、生物多樣性消失、外來種的入侵或自然資源的耗竭等研究上，逐漸朝向以大規模的時間與空間尺度來探討複雜的生態系統，其所要分析的資料也非傳統單一範圍的資料，研究人員需要收集更多相關的資料，並且開始需要功能更強大的資料庫來達成資料共享、整合及資料探勘（data-mining）（陸聲山、林朝欽，2005；Cushing et al，2003）

且在生態學的研究中，生物間及生物與環境間的關係非常的複雜，生態現象與過程具有時間尺度的動態性及空間尺度上的異質性。有鑑於此，1980年美國國家科學基金會（The National Science Foundation, NSF）開始補助學術界，進行長期生態研究（Long-Term Ecological Research）。目的是希望整合在同一個生態系中的研究成果，創造長期與短期研究共同結合的基礎。

長期的資料蒐集將累積數量龐大，分散且性質差異甚大的許多資料。LTER 為了有效的管理資料，從一開始便強調保存長期資料的重要性。在計畫初始便包括了資料管理的概念，並且將此列為每個研究站研究計畫的一部份。隨著網際網路的蓬勃發展，相關技術提供生態研究團隊傳播與交換訊息的新方式，其應用於生態學研究上的部分逐漸走向如何整合各領域的研究成果，以提供更有用的知識。

為了整合各領域，原始資料在描述資訊上的不足將成為整合的瓶頸。要整合資料，並非僅是將所有資料庫的資料納入一個資料庫系統。目前國際上的組織與計畫中，資料庫交換與整合的重點趨向於發展詮釋資料的標準；先以詮釋資料建立交換的機制，而後再進行資源整合。

EML 是由美國長期生態研究社群（US LTER）所發展的一種詮釋資料標準。這是一種根據生態學科而發展出的詮釋資料資料規格，

也是一種依循 XML 標準建立的標記語言。其目的就是對生態研究的資料進行詳盡且結構化的描述，進而讓不同的應用程式，有規則可尋地找到資料。利用這種方式，生態資料可以被詳細的規格化且可重複被使用。若從語意網的概念來看，這些資料日後便可以被各種程式所利用，達成自動化（automation）、整合（integration）與重用（reuse）資料的目標。2004 年，EML 2.0.1 發表後，開始推廣到各個生態研究社群，目前 EML 版本已發展至 EML 2.1.0。

EML Schema 為一模組設計，可重複使用，且具有自我描述的能力；資料本身被內嵌在 EML 文件化的檔案裏。EML 模組就像一個包裝容器，以一定結構的方式描述這些不同類型的研究資料。而模組化的一個特性就是具有延展性，如果 EML 在某一個特殊領域上沒有合適或相關的模組可利用，便可經由產生一個新的模組來描述該領域，擴大 EML 的資料集。

EML 的另一項特點在於同時處理 raw data 及詮釋資料，透過 eml-datable、eml-entity 等模組，對資料內容進行描述，讓使用者瞭解其資料內涵、資料儲存記載方式以及資料使用運算的方法與限制等。

三、 EML（版本 2.1.0）與 TWSMP（版本 2009.06）詮釋資料標準兩者之架構。

表 1 國土資訊系統詮釋資料標準與 EML 之版本

	國土資訊系統的詮釋資料標準	EML 的詮釋資料
版本	2009.06	2.1.0
推行單位	內政部資訊中心	Knowledge Network for Biocomplexity (KNB)

<p>摘要說明</p>	<p>可以分為下列 10 項應用綱要：</p> <ol style="list-style-type: none"> 1. 詮釋資料資訊 2. 識別資訊 3. 限制資訊 4. 資料品質資訊 5. 資料歷程資訊 6. 空間展示資訊 7. 供應資訊 8. 範圍資訊 9. 維護資訊 10. 引用資訊 11. 參考系統資訊 12. 展示目錄資訊 13. 詮釋資料擴充資訊 	<p>可以分為下列 23 項 Module：</p> <ol style="list-style-type: none"> 1. eml 2. eml-access 3. eml-attribute 4. eml-constraint 5. eml-coverage 6. eml-dataset 7. eml-dataTable 8. eml-entity 9. eml-literature 10. eml-methods 11. eml-party 12. eml-physical 13. eml-project 14. eml-protocol 15. eml-resource 16. eml-software 17. eml-spatialRaster 18. eml-spatialReference 19. eml-spatialVector 20. eml-storedProcedure 21. eml-text 22. eml-unitTypeDefinitions
-------------	---	---

四、TWSMP 詮釋資料必填項目之比對與分析

於第四與第五節中分析 TWSMP 與 EML 之架構，將其各自的必填項目列出，再做交叉比對。

分析 TWSMP 的架構，其將詮釋資料項目分為三級：「必要項目」（Mandatory, M）、「條件項目」（Conditional, C）及「選擇項目」（Optional, O）等三類情形，TWSMP 將詮釋資料項目分為若干「資訊」，見圖 1。

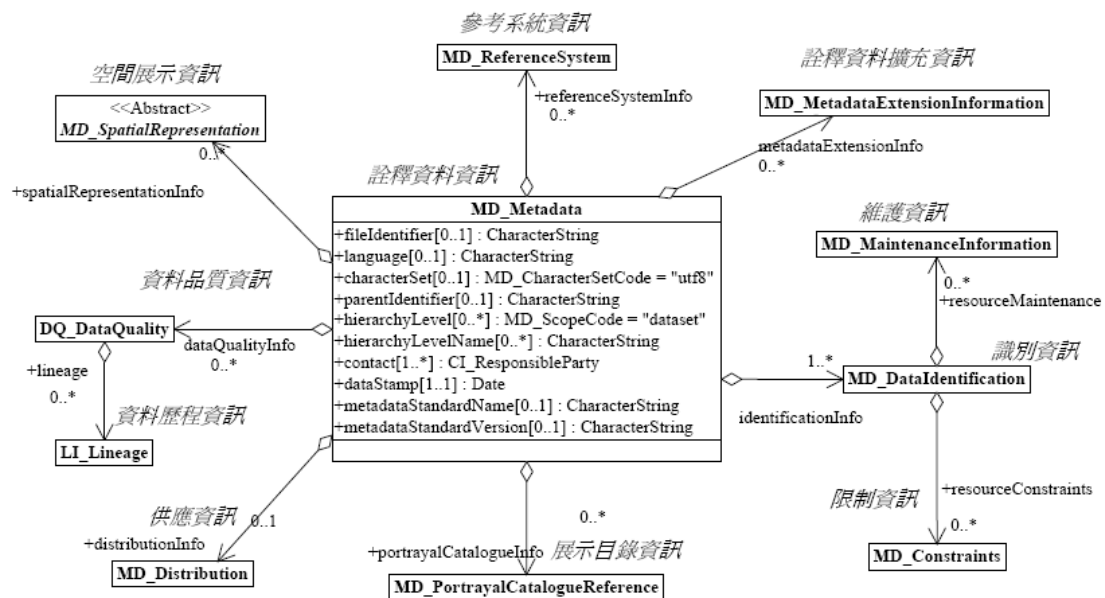


圖 1 TWSMP 整體架構

事實上 TWSMP 是以圖中間的「詮釋資料資訊」為主，去串連其他各項資訊，由圖 1 可知其中只有「識別資訊」與「詮釋資料資訊」之間是發生一至多次的關係(1..*)，其他資訊則是零至多次(0..*)，因此真正必填項目應在「識別資訊」與「詮釋資料資訊」中。

再參考詮釋資料標準中之資料典部分，可知真正為必填的項目有：

1. 「詮釋資料資訊」的必填項目、條件填答項目以及其下「聯絡資訊」的子項目共 9 項。
2. 「識別資訊」中的必填與條件填答項目與其下的子項目共 11 項。

共計 20 項必須填答之欄位，其中有 4 項 EML 無相對應的欄位，列於表 2。

表 3 列出所有其餘 16 項比對成功的項目，以供參考核對。

表 2 EML 無相對應的 TWSMP 項目

項次	類別	屬性名稱或關係	定義	選填條件	資料型別
5	MD_Metadata 詮釋資料資訊	hierarchyLevel 描述層級	詮釋資料描述對象之階層分級	C	MD_ScopeCode <<CodeList>>
6		hierarchyLevelName 描述層級名稱	詮釋資料所描述對象之層級名稱	C	CharacterString
27	MD_DataIdentification 資料識別資訊	topicCategory 主題分類	地理資料描述現象之主題分類	C	twsmpMD_Topic CategoryCode <<Enumeration>>
187	EX_Extent 範圍資訊	verticalElement 垂直元素	地理資料之對應垂直範圍	C	Association

表 3 TWSMP 與 EML 的對應項目

TWSMP					EML			
項次	類別	屬性名稱或關係	定義	選填條件	資料型別	module	類別	資料型別
2	MD_Metadata 詮釋資料資訊	language 語言	詮釋資料使用之語言	C	CharacterString	resource	language	NonEmptyStringType
3		characterSet 字元集	詮釋資料之字元編碼標準	C	MD_CharacterSetCode <<CodeList>>	physical	characterEncoding	NonEmptyStringType
7		contact 聯絡資訊	詮釋資料之權責單位或個人	M	CI_ResponsibleParty	resource	metadataProvider	ResponsibleParty
8		DateStamp 時間	詮釋資料之建置時間	M	Date	resource	pubDate	yearDate
170	CI_ResponsibleParty	individualName	聯絡人的姓名及頭	C	CharacterString	party	individualName	person

	權責單位資訊	個人姓名	銜資訊				e	
171		organisationName 單位名稱	權責單位之名稱	C	CharacterString		organizationName	NonEmptyStringType
172		positionName 職稱	權責單位之聯絡人 職稱或角色	C	CharacterString		positionName	NonEmptyStringType
174		role 角色	權責單位之角色	M	CI_RoleCode<<CodeList>>	resource	role	RoleType
19	MD_DataIdentification 資料識別資訊	abstract 摘要	摘要說明地理資料 之重要特性	M	CharacterString	resource	abstract	TextType
25		language 語言	地理資料記錄時所 採用之語言	M	CharacterString	resource	language	NonEmptyStringType
26		characterSet 字元集	地理資料之字元編 碼標準	C	MD_CharacterSetCode <<CodeList>>	physical	characterEncoding	NonEmptyStringType

167	CI_Citation 引用資訊	title 名稱	引用來源之名稱	M	CharacterString	resource	title	NonEmptyStringT ype
180	CI_Date 日期資訊	date 引用時間	引用資料和服務之 參考時間	M	Date	resource	pubDate	yearDate
184	EX_Extent 範圍資訊	description 範圍描述	地理資料之空間、垂 直及時間範圍	C	CharacterString	Coverage	geographicDesc ription	NonEmptyStringT ype
185		geographicElement 地理元素	地理資料之對應空 間範圍	C	Association		geographicCov erage	geographicCovera ge
186		temporalElement 時間元素	地理資料之對應時 間範圍	C	Association		temporalCover age	temporalCoverage

五、 EML 詮釋資料必要項目之比對與分析

EML 詮釋資料標準之架構是由多個模組所組成，模組下可能又有一個模組，一層一層的包覆下去，最上層的 EML 模組就如同容器(container)一般，但必須依據次一層的模組(sub-module)進一步地加以說明與描述，但自身必須提供基本的鑑別(identification)資訊。而 EML 其下大大小小的模組共計有以下 23 個：

1. eml
2. eml-access
3. eml-attribute
4. eml-constraint
5. eml-coverage
6. eml-dataset
7. eml-dataTable
8. eml-entity
9. eml-literature
10. eml-methods
11. eml-party
12. eml-physical
13. eml-project
14. eml-protocol
15. eml-resource
16. eml-software
17. eml-spatialRaster
18. eml-spatialReference

19. eml-spatialVector
20. eml-storedProcedure
21. eml-text
22. eml-unitTypeDefinitions
23. eml-view

從 EML 詮釋資料標準之架構來看，EML 的上層模組，即 EML 資源模組(eml-resource module)，包含所有資源的基本資訊。主要有四個基本的資源類型(resource type)可供利用，包括資料集資源(dataset resource)、文獻資源(literature resource)、協議資源(protocol resource)和軟體資源(software resource)等四個類型。各資源類型的描述雖有其特殊性，但之間仍具有一些共通的欄位，例如標題、建置者等。而其它同一層級之模組，如 access 模組為描述文件或資料之權限資訊，而 additionalMetadata 模組是當 EML 在某一個特殊領域上沒有合適或相關的模組可利用，經由產生一個新的模組來描述該領域，擴大 EML 的資料描述之額外的詮釋資料模組。而 EML 第二階層下又有許多層級，其它模組有可能互相散布於其中，一同組成整個 EML 詮釋資料架構。

國際長期生態研究組織(ILTER)在建置 EML 詮釋資料時，除必填項目 title、creator/contact 及 abstract 等資訊外，亦建議選填其它項目資料，以求取最大公約數資料項目，加強生態資訊的完整性。現將其 EML 詮釋資料必填及建議選填項目與 TWSMP 詮釋資料對應整理，如表 4。

由表 4 可知，EML 詮釋資料之必填項目皆與 TWSMP 詮釋資料相對應；而 EML 建議選填項目部份，除 dataTable 資料表外，其餘項目資訊亦皆與 TWSMP 詮釋資料相對應，而 dataTable 資料表實為記錄資料屬性之定義與描述，這部份之闡述可由各領域的資料標準另訂之。

故總結上述之情形，EML 詮釋資料必填與建議選填項目之部份，對照 TWSMP 詮釋資料時，皆有其相對應的項目，因此評估 TWSMP 詮釋資料不需針對 EML 詮釋資料項目進行額外擴充。

表 4 EML 與 TWSMP 之對應項目

EML				TWSMP			
模組	屬性	定義	選填條件	項次	類別	屬性	選填條件
Resource	title	資料名稱	M	167	CI_Citation 引用資訊	title 名稱	M
	creator	建置者	M	7	MD_Metadata 詮釋資料資訊	contact 聯絡資訊	M
	abstract	資料摘要	M	19	MD_DataIdentification 資料識別資訊	abstract 摘要	M
	keywordSet	關鍵字	O	52	MD_Keywords 關鍵字資訊	keyword 關鍵字	O

	distribution	資料來源	O	270	MD_Distribution 供應資訊		O
Dataset	contact	資料來源之單位或個人的聯絡資訊	M	7	MD_Metadata 詮釋資料資訊	contact 聯絡資訊	M
Coverage	geographicDescription	資料對應之地理空間描述	O	184	EX_Extent 範圍資訊	description 範圍描述	C
	geographicCoverage	資料對應之地理範圍	O	185	EX_Extent 範圍資訊	geographicElement 地理元素	C
	temporalCoverage	資料對應之時間範圍	O	186	EX_Extent 範圍資訊	temporalElement 時間元素	C
Physical		實體檔案資訊描述	O	284	MD_Format 格式資訊		O
Methods		資料歷程	O	86	LI_ProcessStep 處理步驟資訊		O

DataTable		資料屬性描述與定義	O			
SpatialVector		空間向量資料描述	O	176	MD_VectorSpatialRepresentation 向量空間展示資訊	O
SpatialRaster		空間網格資料描述	O	157	MD_GridSpatialRepresentation 網格空間展示資訊	O