

國外電子檔案最新技術或制度-KOST-Val 驗證工具簡介

KOST-Val 驗證工具是由 PREFORMA(PREservation FORMAts for culture information/e-archives)計畫及歐盟委員會的 FP7-ICT 方案共同贊助。KOST-Val 是一個開放源碼的格式驗證工具(可驗證 TIFF、SIP、PDF/A、SIARD、JPEG 及 JP2 等格式)，可幫助人員驗證檔案，並提供分析檔案格式功能。

KOST-Val 透過圖形化介面操作使用，進行本機 TIFF、SIP、PDF/A、SIARD、JPEG 及 JP2 等檔案格式檢查，且支援單一檔案與批次處理作業檢查，支援相關格式說明如下：

一、支援格式

(一) TIFF:標記圖像格式 (Tagged Image File Format)

TIFF (tagged image file format) 是一種跨平台及廣泛使用的靜態圖形檔案格式，其副檔名為「.tif」或「.tiff」。最初由 Aldus 公司與微軟公司一起為 PostScript 列印開發的一種標準影像格式，適用於不同的解析度、不同的色彩模式和不同的壓縮方式，幾乎所有影像處理軟體都有支援，所以大多數使用者都會以 TIFF 格式(又稱標籤資訊格式)儲存圖檔。TIFF 有以下的特色：

1. 使用標籤 (tag)：TIFF 使用標籤來標示圖形檔案中相關參數，如解析度、色彩深度、定義圖像資料排列方式，以及圖像壓縮方式等，俾利影像處理軟體了解如何處理檔案中的數據。

2. 彈性佳：TIFF 之標籤功能標記圖形檔案相關資訊，TIFF 可以將多張掃描的影像集結在單一檔案中；也可以將其他的圖像格式（如 JPEG）包含在 TIFF 檔內。
3. 適合數位典藏：與 JPEG 不同的是 TIFF 在儲存資料時採非失真性壓縮，不會損失部分圖形資訊，因此 TIFF 為大部分數位典藏機構所採用的圖像儲存檔案格式。

(二) SIP: 提交信息封包(Submission Information Package)

開放檔案資訊系統(Open Archive Information System, OAIS)模型下負責接收生產者的提交信息封包，該封包由陳述訊息¹（提交的數據和詮釋資料）和詮釋資料組成。

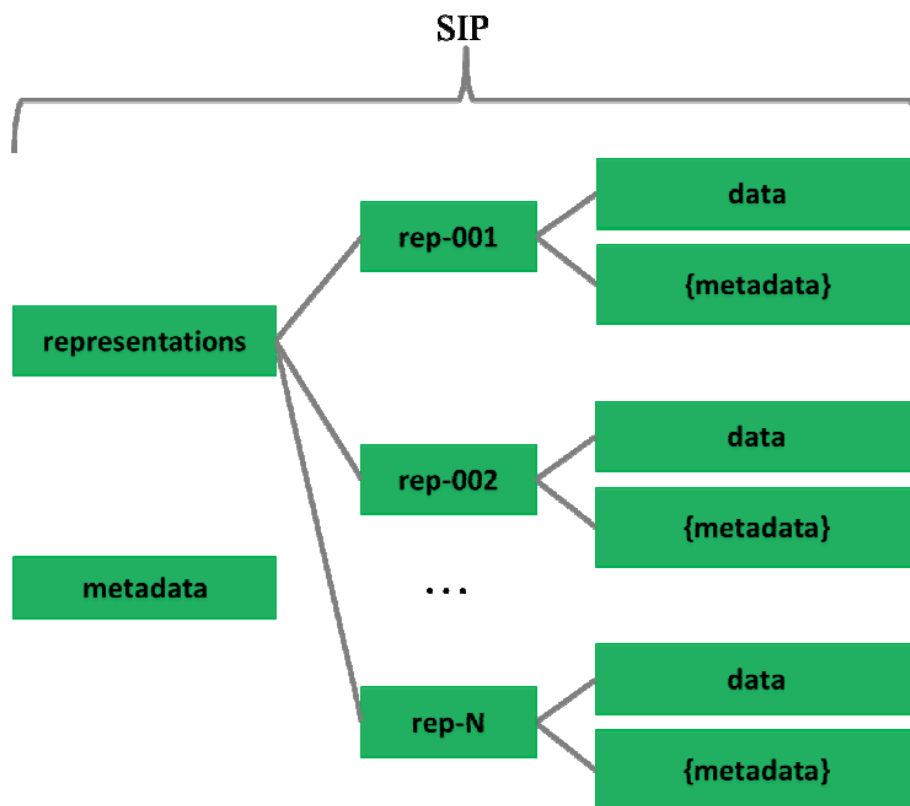


圖1 SIP 資料模型

¹ 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>

(三) PDF/A : 可攜式文件格式 (Portable Document Format/Archive)

PDF 於 1990 年代早期開發，用於分享包括文字及內建影像的文件，並且可以跨平台操作。規範是由國際標準組織 (ISO) 所制定，PDF/A 適用於電子文件的長期保存格式。

(四) SIARD: 典藏關聯式資料庫格式 (Software Independent Archiving of Relational Databases)

SIARD 為瑞士聯邦檔案館開發用來儲存關聯式資料庫裡的資料，SIARD 不受於作業系統影響，可於 Windows、Linux、Mac OS X. 等作業系統執行。

(五) JPEG: 聯合圖像專家小組格式 (Joint Photographic Experts Group)

JPEG 為針對影像使用的失真壓縮方法，其不適合於線條繪圖 (drawing) 和其他文字或圖示 (iconic) 的圖形，因為這些類型的圖形用此種壓縮方法得到的結果不好，而在色調及顏色平滑變化的相片或是寫實繪畫 (painting) 上可以達到它最佳的效果，該格式目前為最普遍的被用來儲存和傳輸相片的格式。

(六) JP2: 聯合圖像專家小組 2000 格式 (Joint Photographic Experts Group 2000)

JP2 是一種圖檔格式在壓縮時具有良好的兼容能力、精

確度及分辨率等優點，JG 2 被認為是未來取代 JPEG 的下一代圖像壓縮標準。

一、驗證方式

KOST-Val 針對支援的格式與其檔案相關附件，透過 KOST-Val 制定驗證政策執行檢查，檢查結果產生 XML 報告。

KOST-Val 符合以下六種格式驗證：

(一)TIFF 驗證：使用 JHOVE 驗證 TIFF 格式結構及內容，並以 ExifTool 驗證壓縮、色彩空間及多頁面等 TIFF 格式特性(ISO 12639:2003)。

(二)SIARD 驗證：驗證 SIARD (eCH-0165 v1) 格式結構和內容。

(三)PDF/A 驗證：透過使用 3-Heights 修復工具來驗證 PDF 格式結構和內容。可將不同的錯誤訊息歸納成幾個主要類別、如字型、圖像和詮釋資料。(ISO 19005-1:2005)

(四)JP2 驗證：使用 Jpylyzer 驗證 JP2 格式 (ISO 15444) 結構和內容。

(五)JPEG 驗證：使用 Bad Peggy 驗證 JPEG 文件 (ISO 10918-1) 格式結構和內容。

(六)SIP 驗證：使用 SIP 規範要求驗證，驗證範圍包括目錄結構、架構驗證及校驗。

二、軟體介紹

KOST-Val V1.8.3 軟體操作如下：

- (一) 選擇要驗證的檔案，選擇批次驗證(Choose Folder)或單一檔案驗證(Select File)，點選「Validate」執行驗證(如圖 39)。

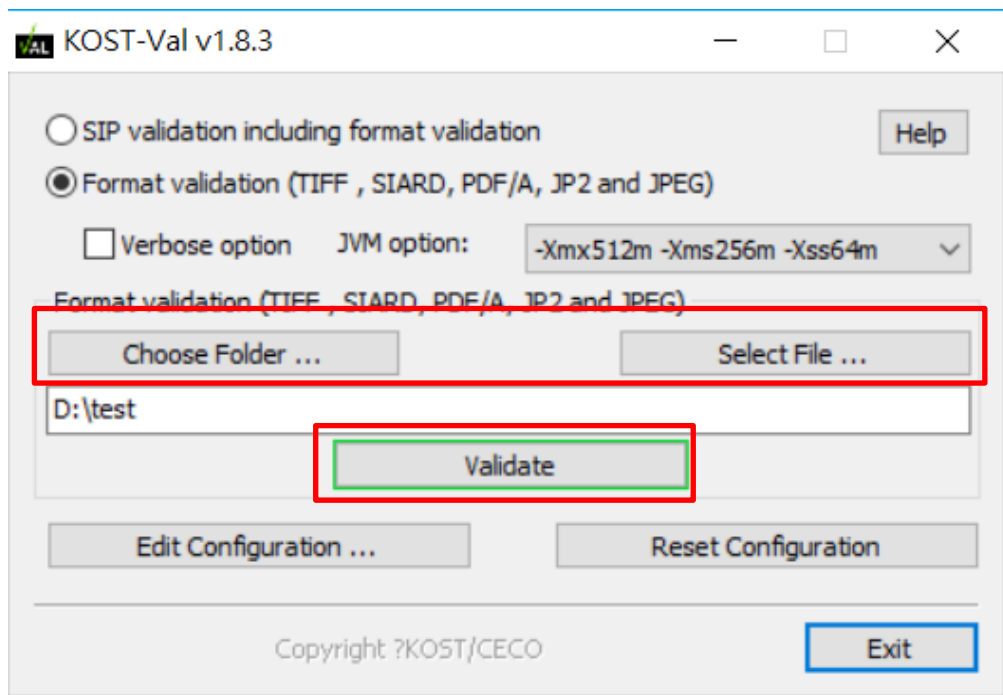
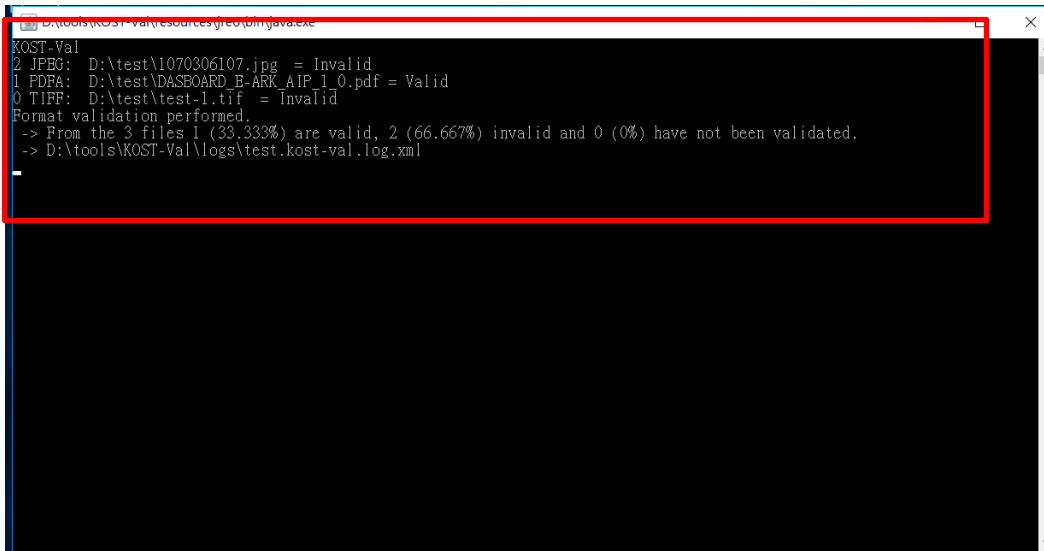


圖2 KOST-Val 選擇驗證檔案

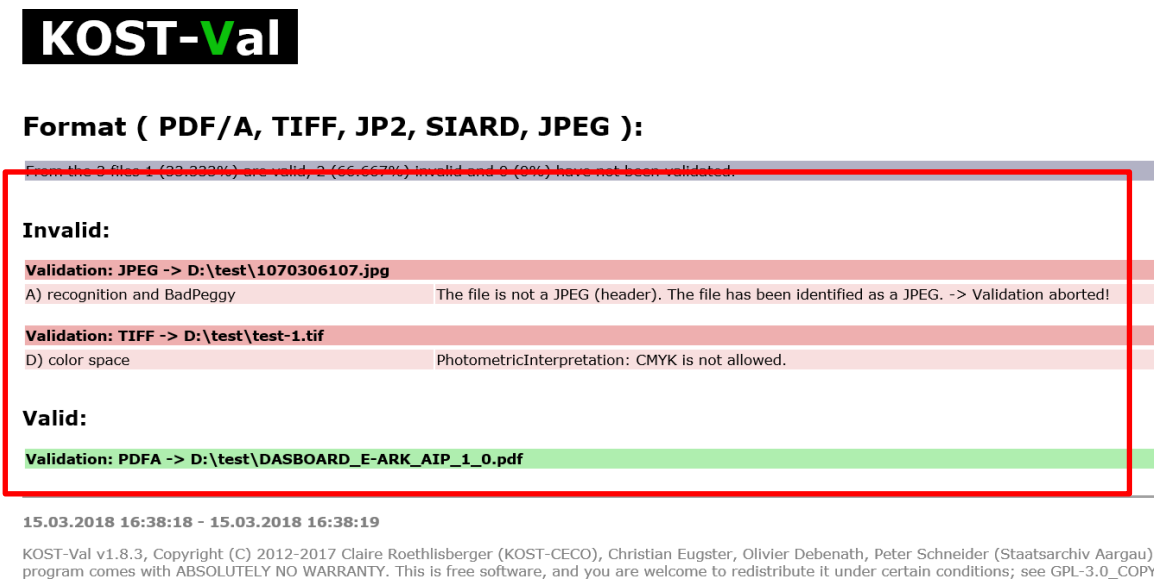
(二) 執行驗證畫面 (如圖 40)。



```
D:\tools\KOST-Val\resources\jre\bin\java.exe
KOST-Val
2 JPEG: D:\test\1070306107.jpg = Invalid
1 PDF/A: D:\test\DASHBOARD_E-ARK_AIP_1_0.pdf = Valid
0 TIFF: D:\test\test-1.tif = Invalid
Format validation performed.
-> From the 3 files 1 (33.333%) are valid, 2 (66.667%) invalid and 0 (0%) have not been validated.
-> D:\tools\KOST-Val\logs\test.kost-val.log.xml
```

圖3 KOST-Val 執行驗證畫面

(三) 完成驗證後產出報告(圖 41)



KOST-Val

Format (PDF/A, TIFF, JP2, SIARD, JPEG):

From the 3 files 1 (33.333%) are valid, 2 (66.667%) invalid and 0 (0%) have not been validated.

Invalid:

Validation: JPEG -> D:\test\1070306107.jpg
A) recognition and BadPeggy The file is not a JPEG (header). The file has been identified as a JPEG. -> Validation aborted!

Validation: TIFF -> D:\test\test-1.tif
D) color space PhotometricInterpretation: CMYK is not allowed.

Valid:

Validation: PDF/A -> D:\test\DASHBOARD_E-ARK_AIP_1_0.pdf

15.03.2018 16:38:18 - 15.03.2018 16:38:19

KOST-Val v1.8.3, Copyright (C) 2012-2017 Claire Roethlisberger (KOST-CECO), Christian Eugster, Olivier Debenath, Peter Schneider (Staatsarchiv Aargau) program comes with ABSOLUTELY NO WARRANTY. This is free software, and you are welcome to redistribute it under certain conditions; see GPL-3.0_COPY

圖4 KOST-Val 報告產出

由於 KOST-Val 可驗證 TIFF、SIP、PDF/A、SIARD、JPEG 及 JP2 等格式，透過驗證檔案結構，協助各機關(構) 電子檔案管理之發展及達到電子檔案長期保存的目標。

三、參考資料來源

- (一) <http://www.preforma-project.eu/kost-val.html>
- (二) <https://github.com/KOST-CECO/KOST-Val/releases>
- (三) <http://www.eark-project.com/>
- (四) https://help.adobe.com/zh_TW/acrobat/pro/using/WSd7ad1123a4c5dffe-5183c8012b97790144-8000.html
- (五) <https://zh.wikipedia.org/wiki/JPEG>
- (六) https://zh.wikipedia.org/wiki/JPEG_2000