

HPC 專案空間使用手冊

目錄

第一章、 HPC 專案空間說明	3
一、 使用目的	3
二、 JUPYTERLAB 連線使用說明	5
三、 磁碟空間使用說明	6
四、 TERMINAL 基本操作指令說明	7
五、 使用私有套件儲存庫安裝 PYTHON PACKAGES	8
六、 容器開關 API 使用說明	9
(一) 開啟專案容器(POD)環境	9
(二) 關閉專案容器(POD)環境	10
第二章、 資料資源儲存庫	11
一、 使用目的	11
二、 使用說明	11
第三章、 團隊進場資料檢視相關作業說明	16
第四章、 附錄：影像序列的醫療費用申報資料檔欄位內容	17

第一章、HPC 專案空間說明

一、使用目的

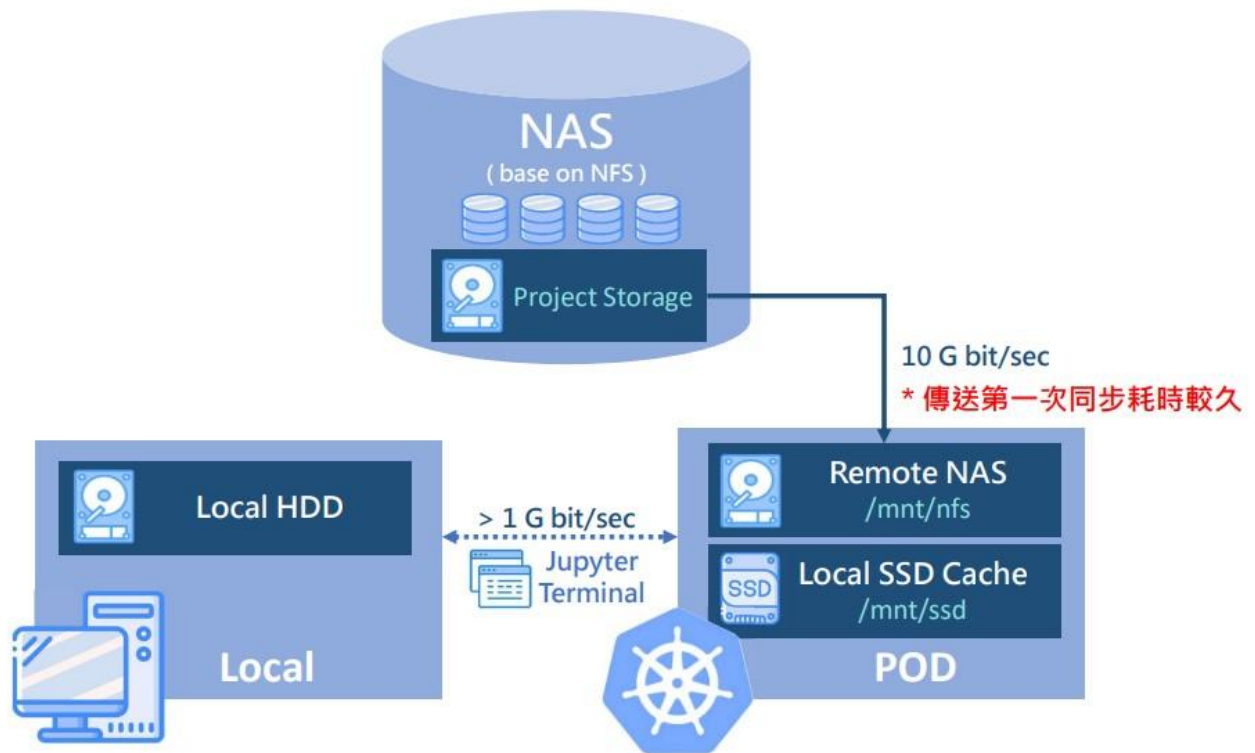
本署分派予團隊專屬使用之 HPC 專案空間為「AI 服務作業區-PC 工作站、POD 容器組與資料資源儲存庫」等三部分組成，如[圖 1]所示。「PC 工作站」為團隊實際操作模型程式之 Windows 10 Professional 作業主機，署內會提供團隊專屬登錄帳號與密碼；「POD 容器組」為團隊實際操作模型程式運作服務作業之操作區，提供新一代 Integrated Development Environment (IDE) 整合式網頁操作 Jupyterlab 環境，俾利於進行模型開發與執行等應用服務；而「資料資源儲存庫」為存放團隊作業所需資料與模型程式之靜態儲存庫，提供團隊對於相關操作資源之操作管理應用。

團隊申請的研究資料與處理後的資料與模型程式，皆會被儲存於資料資源儲存庫(以下稱 NAS)專屬的 volume 中，當團隊登入 PC 工作站後，即可藉由 POD 容器組內之網頁式 Jupyter Notebook 整合 Jupyterlab 環境服務操作介面進行資料與模型程式之存取應用，亦可將欲資料檢視或標註之影像資料下載至 PC 工作站本機端硬碟空間，或將每日完成處理之資料或模型程式上傳回 volume 中存檔備份。於 Jupyterlan 環境開啟 Terminal 之 Linux 環境中，亦可藉由/mnt/nfs 同步存取操作資料資源儲存庫；同時，亦因採用快取模式掛載，頻繁取用之檔案將可快速存取，而非每次經由網路 I/O 存取。

POD 容器組內提供「兩組」磁碟空間予以運用：

1. 高速暫存空間：「/mnt/ssd」，高速但容器環境關閉後，資料會清除。
2. 網路永久空間：「/mnt/nfs」，透過快取存取網路資料夾，資料將保存至專案結案離開。

請團隊依自身應用系統或軟體需求，斟酌使用與調整。



[圖1] HPC 專案空間架構圖

二、 Jupyterlab 連線使用說明

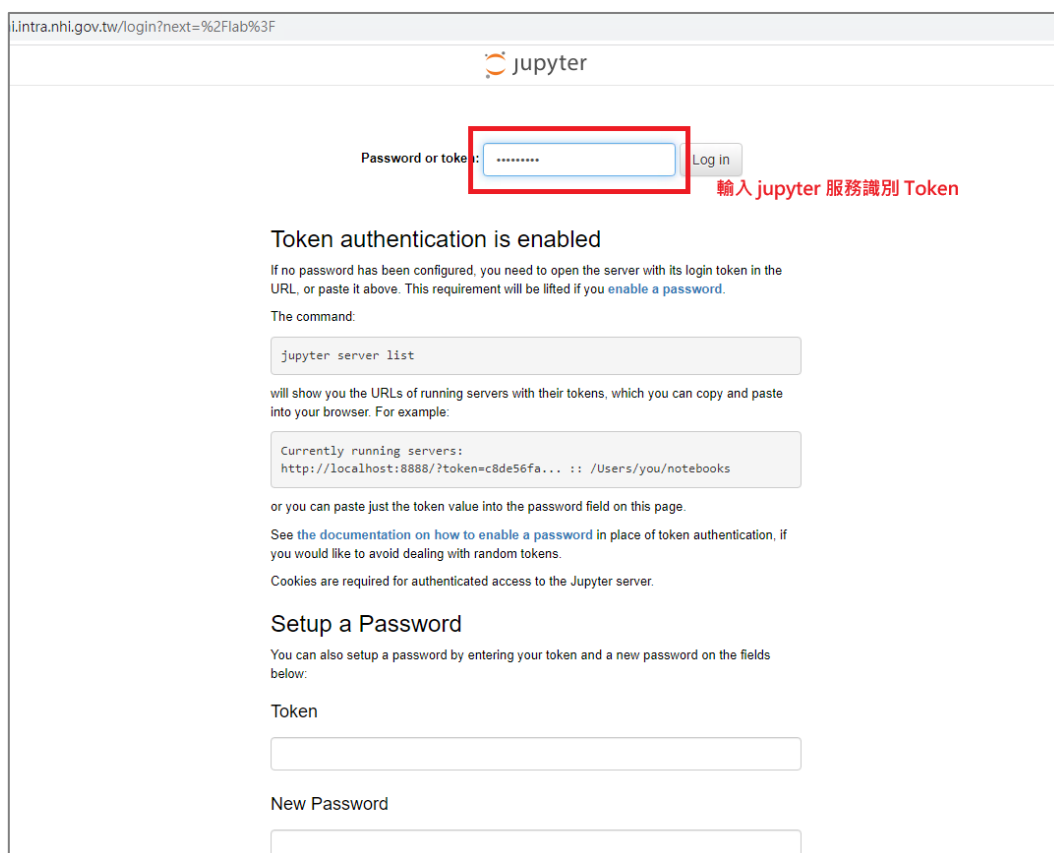
1. 使用者連線資訊

以下為團隊於專屬 HPC 空間所需使用到的連線資訊，實際資訊請與署內負責人員索取。

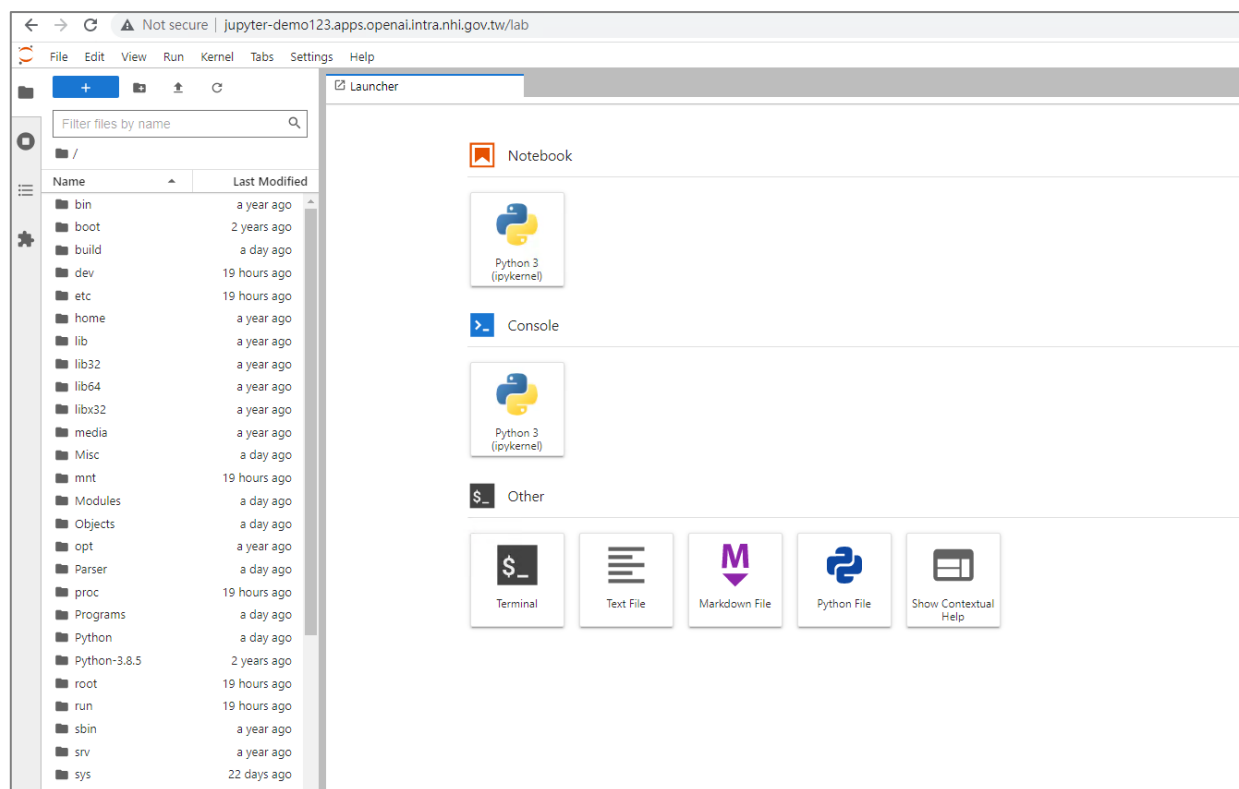
連線參數	參數名稱
Jupyter Notebook Location	Jupyter連線網址
Jupyter Notebook Token	Jupyter網頁預設初始密碼
TensorBoard Location	Tensorboard連線網址

2. 連線至 Jupyter Notebook

開啟 Firefox 瀏覽器，在網址搜尋處輸入署內負責人員提供之 Jupyter Notebook Location，並在“Password or token”空格處輸入署內提供之 Jupyter Notebook Token，如[圖 2]。連線完成畫面如[圖 3]。



[圖 2] Jupyter Notebook 連線畫面



[圖3] Jupyter Notebook 連線成功

三、磁碟空間使用說明

1. 磁碟空間類別詳細說明

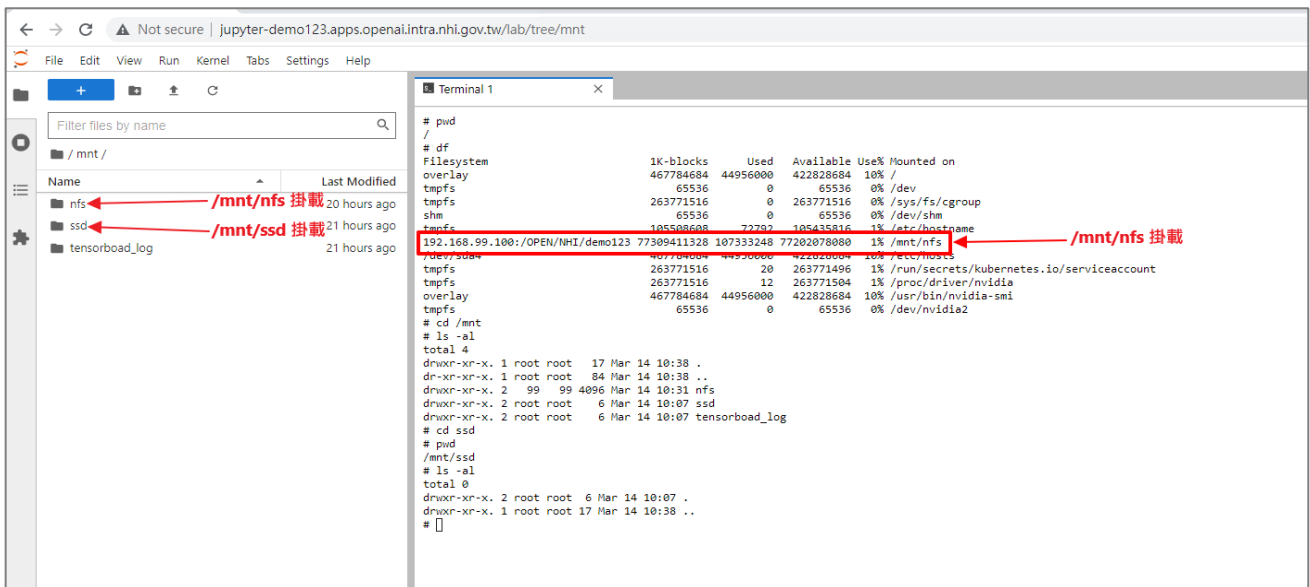
本作業環境提供團隊於 POD 容器組內計兩組磁碟空間予以運用，分別為「高速暫存空間」與「網路永久空間」。

- (1) 高速暫存空間：主要為高速固態硬碟(Solid-State Disk, SSD)，I/O 速度快，將受限於 POD 容器運作與否生命週期相關，為容器底層作業系統空間掛載，並於 POD 容器啟動時自動掛載於「/mnt/ssd」，空間大小 1.8TB。基於模式運作效能考量，建議將訓練或執行運作之模型與資料上傳至此空間，惟容器環境關閉後，空間內資料與模型將自動清除。
- (2) 網路永久空間：主要為資料與模型之靜態儲存與管理應用，其不受 POD 容器運作與否生命週期影響，為外部磁碟空間掛載(即資料資源儲存庫 NAS)，資料資源將保存至專案結案時離

開，並於 POD 容器啟動時自動掛載於「/mnt/nfs」，空間大小 2TB。專案所需初始化資料皆會置放於此磁碟空間，建議訓練處理後的資料與模型，請由「/mnt/ssd」定期複製回此磁碟空間，避免 POD 不小心關閉而致資料遭系統刪除而遺失。

2. 掛載作業說明與檢視確認

「/mnt/ssd」與「/mnt/nfs」磁碟空間不須團隊自行掛載啟動，空間將於 POD 容器啟動時自動掛載，如[圖 4]。可由 Jupyter Notebook 操作環境中使用 df 進行查詢(僅能顯示外部磁碟掛載空間「/mnt/nfs」，因「/mnt/ssd」為系統底層掛載而無法顯示)，或由左側視覺化 UI 介面進行檢視確認，若發現掛載點不存在時，請立即通知署內負責人員協處。

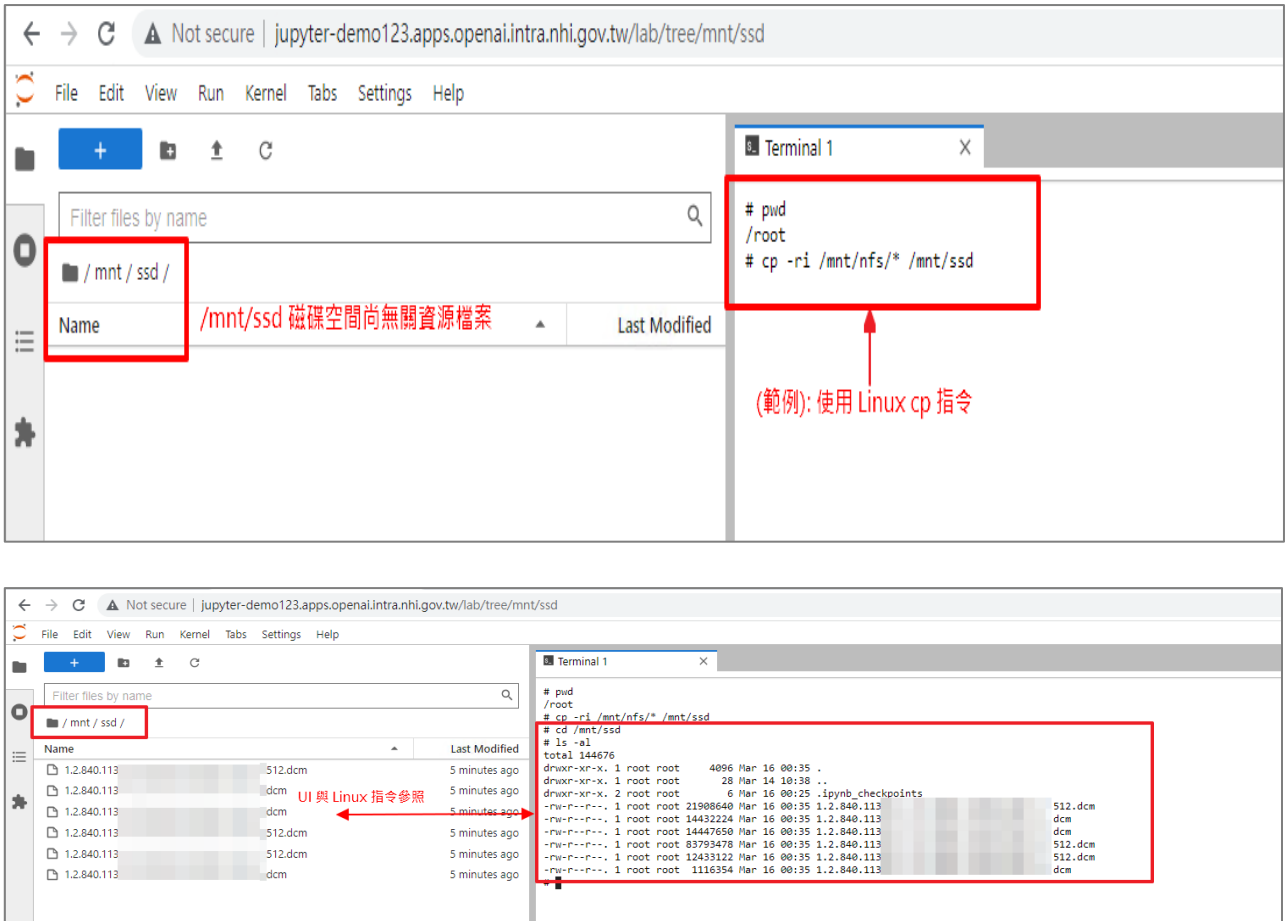


[圖4] 磁碟空間掛載狀態檢視

四、 Terminal 基本操作指令說明

1. 查詢專案資料目錄與取得研究資料

於 POD 環境下進行資料操作，可直接使用 Linux 指令(如: cp 或 mv 等)於目錄「/mnt/nfs」與「/mnt/nfs」間存取轉換，或由 Jupyter Notebook 網頁式操作環境左側頁框進行視覺化「上(下)載」等基本檔案操作(如[圖 5])。



[圖5] 查詢專案資料目錄與進一步取得資訊

2. 關閉專案空間及解除 GPU 資源綁定

團隊結束署內作業環境後，需關閉團隊的專案空間，並解除 GPU 綁定狀態。請注意，每次關閉專案空間就會將掛載的 SSD 清空(「/mnt/ssd」)，請務必確認操作後之檔案或資料是否已同步完成更新至 NAS (「/mnt/nfs」)中後再行關閉。並請務必參酌署內規定，針對申請於「白日上午時間或夜間延長使用專案空間」時，請配合署內負責人員指示關閉專案空間，以免影響收費等權益。

五、 使用私有套件儲存庫安裝 Python Packages

現行 POD 執行環境內所提供之 Python 套件及版本，若不符合團隊應用

需求，請聯繫署內負責人員提出欲新增之套件項目及版本，後續將新增之套件推送至 Private Index Server 提供團隊使用(POD 執行環境之 pip 已預設指向 Private Index Server)，以下提供套件安裝操作範例：

1. 安裝套件

```
pip install [package_name]
```

2. 安裝指定版本套件

```
pip install [package_name]==[package_version]
```

六、 容器開關 API 使用說明

(一) 開啟專案容器(POD)環境

功能說明	開啟專案容器環境
函式名稱	https://h301gw-openai2.intra.nhi.gov.tw/project/environment
Protocol	JSON over HTTP POST
Headers	Content-Type=application/json nhitoken=BASE64ENCODE(專案代號)
傳入參數說明	專案代號 - "ProjectName": <專案代號>"
傳出參數說明	回傳 HTTP response 200 OK 及包含以下內容的陣列 專案代號 - "ProjectName": "<project_name>" 專案階段 - "ProjectStage": "Customize" 專案狀態 - "ProjectStatus": "Start" 時間戳記 - "timestamp": "<current_time>"
處理邏輯	調用 oc 指令設定 OKD 中該專案 deployment 的 replicas=1 並重新建立 PVC

操作範例	<pre># 以專案代號 E202501010001 為例 curl \ -X POST \ -H 'Content-Type: application/json' \ -H "nhitoken: \$(echo -n 'E202501010001' tr 'A-Z' 'a-z' base64)" \ -d "{\"ProjectName\": \"\$(echo -n 'E202501010001' tr 'A-Z' 'a-z')\"}" \ https://h30lgw-openai2.intra.nhi.gov.tw/project/environment</pre>
------	---

(二)關閉專案容器(POD)環境

功能說明	關閉專案容器環境
函式名稱	https://h30lgw-openai2.intra.nhi.gov.tw/project/environment
Protocol	JSON over HTTP DELETE
Headers	Content-Type=application/json nhitoken=BASE64ENCODE(專案代號)
傳入參數說明	專案代號 - "ProjectName": <專案代號>
傳出參數說明	<p>回傳 HTTP response 200 OK 及包含以下內容的陣列</p> <p>專案代號 - "ProjectName": "<project_name>" 專案階段 - "ProjectStage": "Customize" 專案狀態 - "ProjectStatus": "Stop" 時間戳記 - "timestamp": "<current_time>"</p>
處理邏輯	調用 oc 指令設定 OKD 中該專案 deployment 的 replicas=0 並刪除 PVC。
操作範例	<pre># 以專案代號 E202501010001 為例 curl \ -X DELETE \ -H 'Content-Type: application/json' \ -H "nhitoken: \$(echo -n 'E202501010001' tr 'A-Z' 'a-z' base64)" \ -d "{\"ProjectName\": \"\$(echo -n 'E202501010001' tr 'A-Z' 'a-z')\"}" \ https://h30lgw-openai2.intra.nhi.gov.tw/project/environment</pre>

第二章、資料資源儲存庫

一、使用目的

署內人員匯出團隊申請所需「去識別化影像及其相關醫療費用申報資料檔」，將會放置於 NAS 中，提供團隊需要進行研究使用時，可由 POD 容器組中所提供之整合式 Jupyterlab 環境操作下載儲存於 PC 工作站本機端，進行部分影像檔案挑選，少量而快速檢視研究資料內容。

二、使用說明

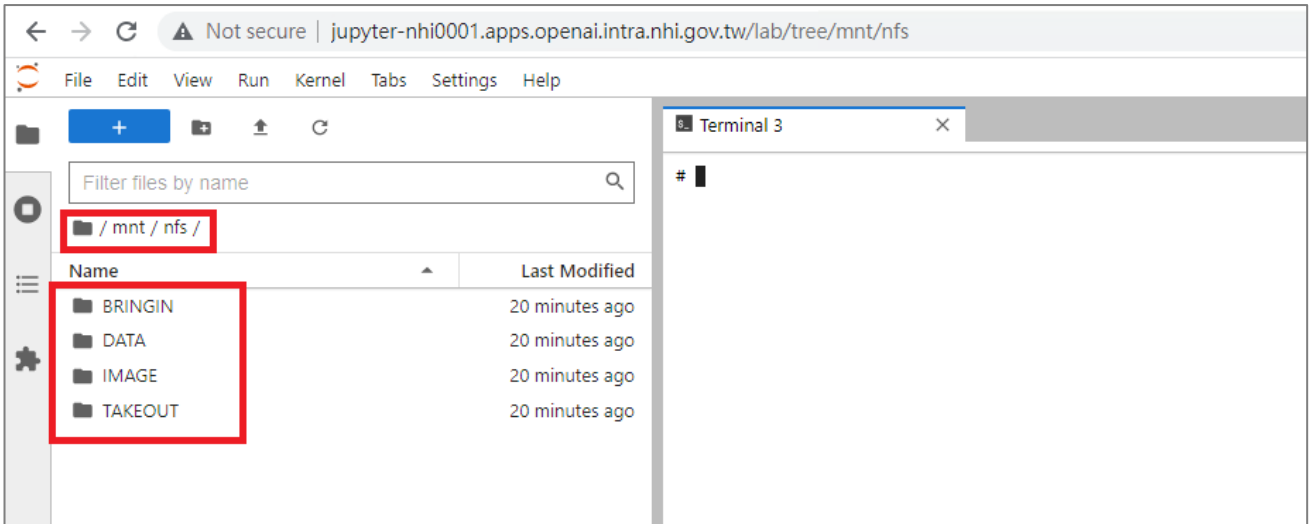
1. 檔案存放注意事項

請注意於 NAS 存放檔案時，建議務必使用英文命名相關檔案及目錄，符號僅可使用“_”(底線)及“-”(連接線)，不可使用中文或其他特殊符號，避免產生檔案無法傳輸的情形，且請建立好完整管理目錄，避免目錄下過多檔案。

2. 檔案預設存放位置

NAS 包含「影像、影像相關醫療費用申報資料檔與團隊需要攜出入之資料」。檔案置放位置如[圖 6]所示，儲存位址名稱及內容說明如下：

- DATA/：影像序列的醫療費用申報資料檔（檔案副檔格式包含 sqlite 與 csv）、影像序列的相關歷史申報資料 csv 檔。
- IMAGE/：完成去識別化後的影像檔案。
- BRINGIN/：團隊攜入資料放置處；經署內審核通過後協助攜入。
- TAKEOUT/：團隊攜出資料放置處，需另行填寫申請表單，經署內審核通過後協助攜出。團隊須於此目錄下，以辦理攜出日期作為資料夾名稱，將該日欲攜出的檔案全數置於此資料夾下。

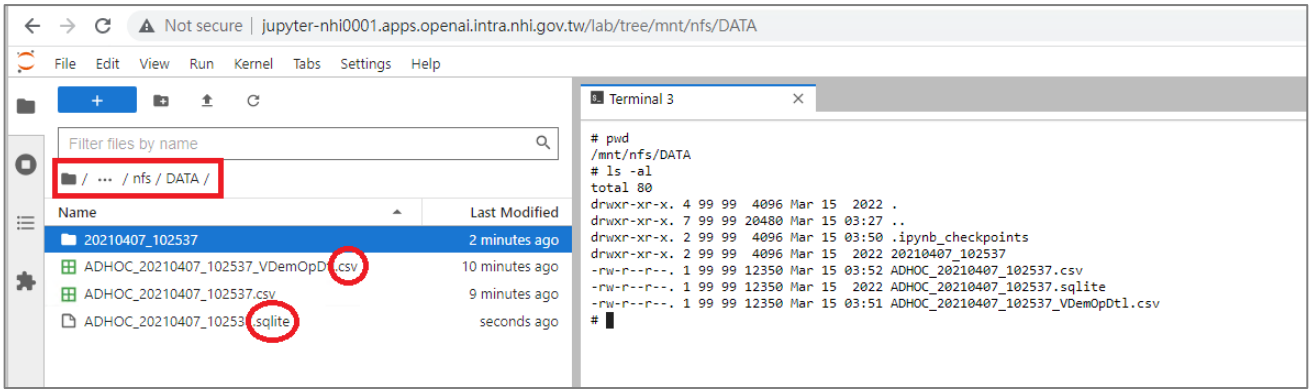


[圖6] NAS (/mnt/nfs) 空間-檔案預設存放位置目錄結構

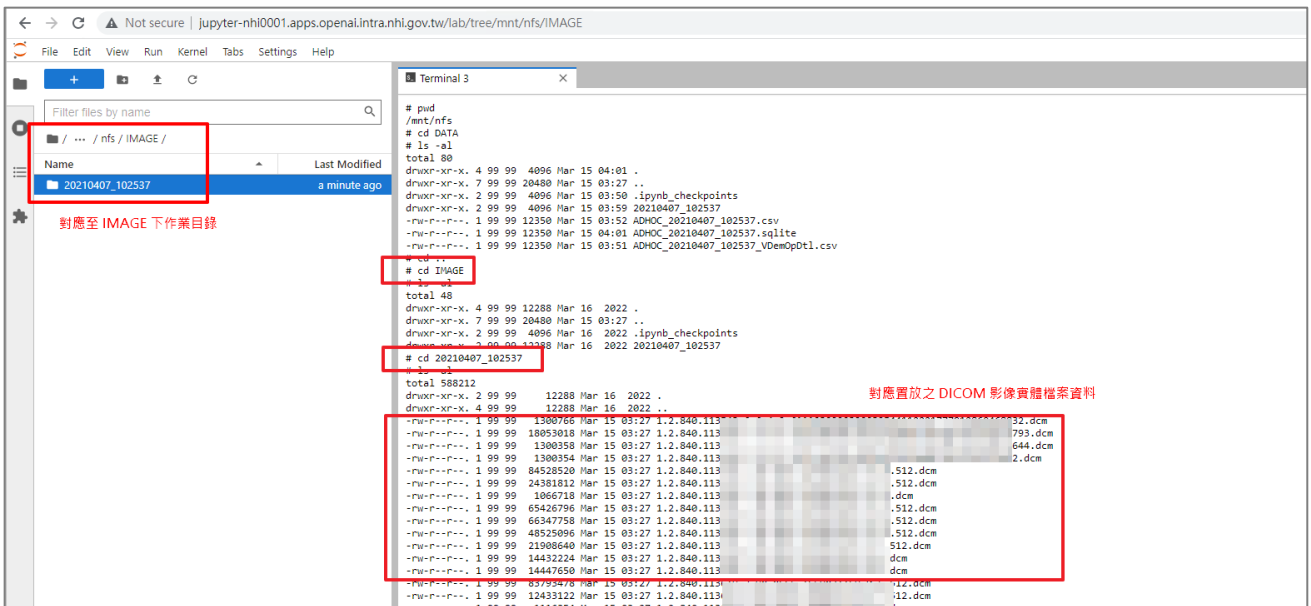
團隊研究申請所需釋出的研究資料會以「一組 8 碼日期加 6 碼流水號」的代碼作為該次釋出的識別依據，分別標示於影像序列的醫療費用申報資料檔案名稱結尾處，以及去識別化影像資料夾名稱，用以對應同一筆釋出資料。如[圖 7]中，在”DATA/”資料夾下以”20210407_102537”結尾的檔案，為對應[圖 8]中”IMAGE/”資料夾下同一筆”20210407_102537”影像資料的相關文字內容。

”DATA/”資料夾下，檔名包含同一識別代碼的檔案共有三個項目，如[圖 7]，分別為：

- “ADHOC_[識別代碼].csv”：影像序列的醫療費用申報資料檔。詳細欄位內容請見附錄。
- “ADHOC_[識別代碼]_VDemoOpDtl_[年份].csv”：影像序列的相關歷史申報資料檔，結尾為團隊所申請的歷史資料年份。
- “ADHOC_[識別代碼].sqlite”：影像序列的醫療費用申報資料檔，內容與 “ADHOC_[識別代碼].csv” 相同。



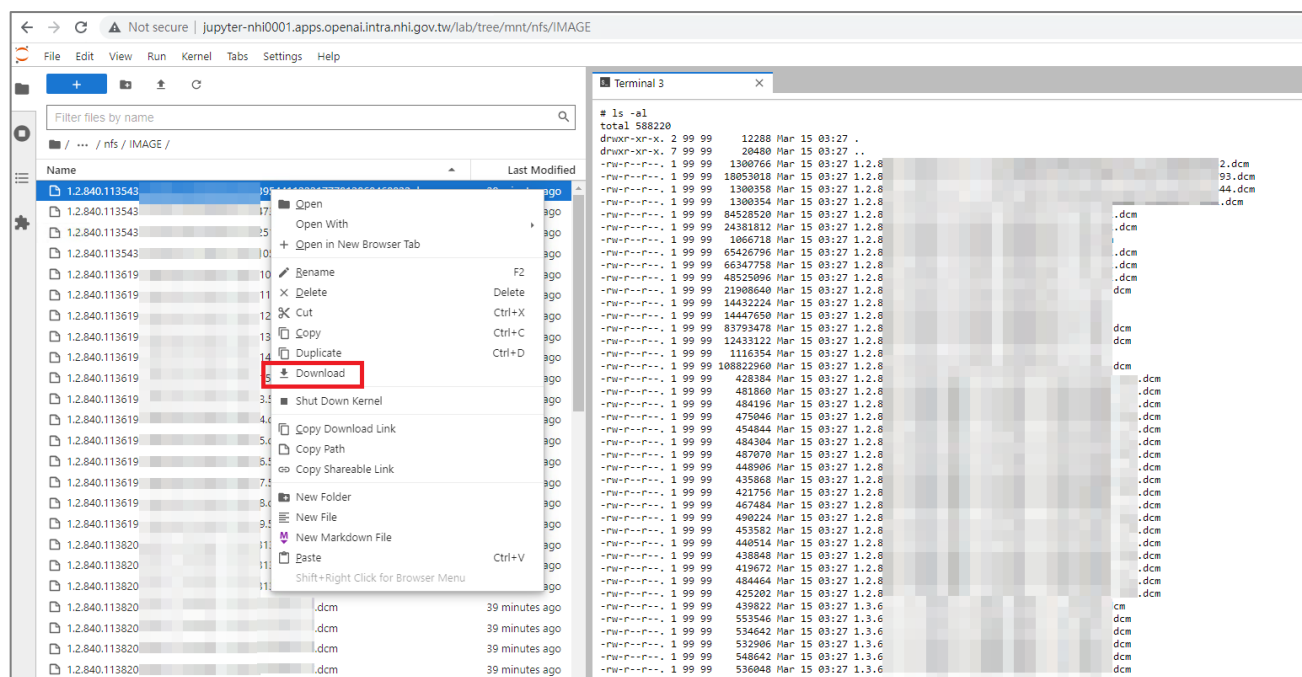
[圖7] 影像序列的醫療費用申報資料



[圖8] 對應置放已去識別化實體影像檔案

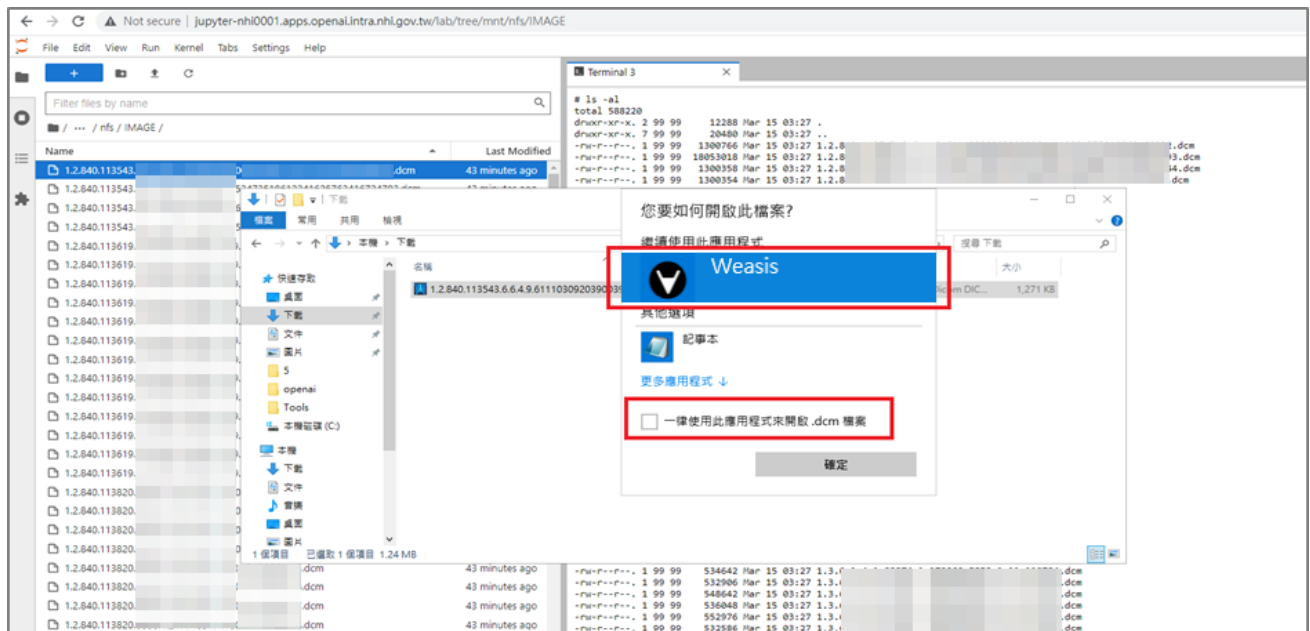
3. 查看影像與下載檢視

在專案對應掛載的雲端網路磁碟空間目錄(/mnt/nfs)下，選取所需影像檔案，選擇 Download 服務功能操作，即可進行影像下載至 PC 工作站作業本機端，如[圖 9]。因應 PC 工作站本機端硬碟空間有限且為網頁式下載影像資料速度較慢，建議用於少量檢視影像使用。

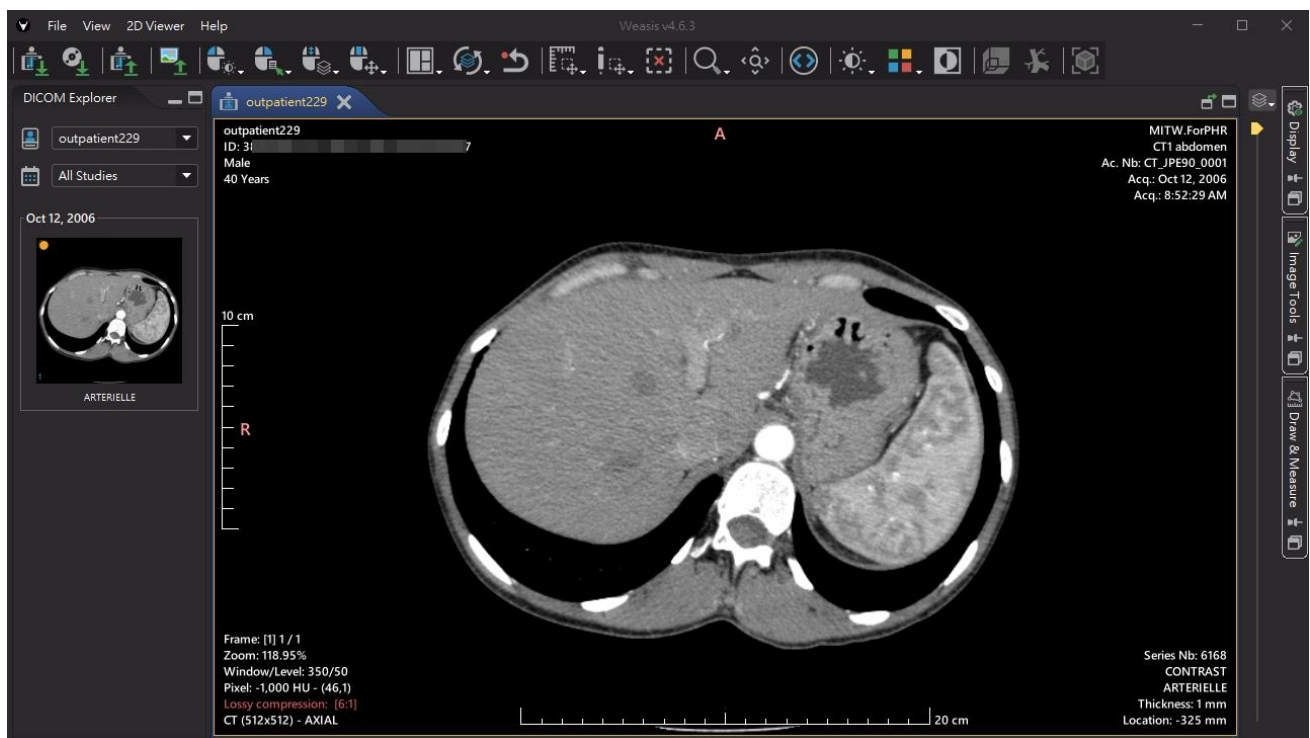


[圖9] 使用 Jupyter Notebook 網頁式操作環境下載影像至本機端

將所需檔案儲存於自選的本機資料夾下後，到檔案儲存位置並選擇以 Weasis 開啟，如[圖 10]，即可檢視影像及進行簡易編輯，如[圖 11]。
(此處瀏覽方式是藉由 POD 容器組以安全性連線將遠端資料下載至本機端進行檢視，團隊無法直接用 Weasis 選取「/mnt/nfs」中目的影像檔案透過網路 I/O 遠端開啟。



[圖10] 選取影像瀏覽器開啟 DICOM 影像檔案



[圖11] Weasis 畫面(範例: CT)

第三章、 團隊進場資料檢視相關作業說明

團對於資料檢視期間，可申請進入署內先行檢視爾後研究作業所需影像資料狀況，進行資料挑選或不必要資料剔除等作業，俾利於研究作業順遂。以下針對資料檢視作業規則說明。

- 影像資料檢視方法：

[1] 團隊自行於 /mnt/nfs/IMAGE 目錄下針對不需要的影像資料直接刪除。但提醒您，一旦刪除後將無法再重新還原。

[2] 團隊可利用 /mnt/nfs/DATA 目錄下的 csv 檔案中影像索引資訊，並正向表列，將欲保留的影像資料填具於新的 csv 檔案中，檔案名稱與資料欄位格式規定說明如下：

- 檔案名稱命名規則：專案代碼_[識別代碼].csv

(如: 2021010001_20210407_102537.csv)

- 檔案置放位置： /mnt/nfs/TAKEOUT 目錄下

- 檔案內容格式：

由左至右，分別為 STUDY_INSTANCE_UID,

SERIES_INSTANCE_UID 與 DICOM File Name，團隊可以依據實際應用需求，保留到單張或檢查或序列目錄下之影像資料作業，如下圖範例說明。

- 團隊請特別注意，若選用第 2 種方式進行選取保留影像資料處理時，請務必配合本署作業要求規範，以免影響後續使用權利。

	A	B	C	
1	STUDY_INSTANCE_UID	SERIES_INSTANCE_UID	DICOM File Name	
2	1.2.68.0.1.6	71	1.3.12.2.11	70 1.3.12.2.1 1.dcm
3	1.2.68.0.1.6	71	1.3.12.2.11	70 1.3.12.2.1 8.dcm
4	1.2.68.0.1.6	71	1.3.12.2.11	79 1.3.12.2.1 0.dcm
5	1.2.68.0.1.6	71	1.3.12.2.11	73 1.3.12.2.1 4.dcm
6	1.2.68.0.1.6	71	1.3.12.2.11	73 1.3.12.2.1 5.dcm
7	1.2.68.0.1.6	71	1.3.12.2.11	73 1.3.12.2.1 6.dcm
8	1.2.68.0.1.6	1	1.3.12.2.11	7
9	1.2.68.0.1.6	1	1.3.12.2.11	7
10	1.2.68.0.1.6	1	1.3.12.2.11	7
11	1.2.124.113	37		
12	1.2.124.113	35		
13	1.2.124.113	31		

保留單張影像(IMAGE)資料

保留整個序列(SERIES)目錄下所有的影像資料

保留整個檢查(STUDY)目錄下所有的影像資料

第四章、 附錄：影像序列的醫療費用申報資料檔欄位內容

欄位名稱	欄位中文名稱	欄位定義	備註
HOSP_ID	醫事機構代碼編號	串接影像 8key	hashID
HOSP_DATA_TYPE	醫事類別	串接影像 8key	11 門診西醫診所 12 門診西醫醫院 13 門診牙醫 14 門診中醫 15 門診洗腎 19 門診其他醫事機構 20 住診代辦業務 21 住診西醫診所 22 住診西醫醫院 29 住診其他醫事機構 30 特約藥局 40 物理治療所 50 特約檢驗所
FEE_YM	費用年月	串接影像 8key	
APPL_TYPE	申報類別	串接影像 8key	1- 送核 2- 補報 3-
APPL_DATE	申請/申報日期	串接影像 8key	
CASE_TYPE	案件分類	串接影像 8key	0 統扣案件核減 11 牙醫一般案件 12 牙醫急診 13 牙醫門診手術 14 牙醫師至無牙醫師鄉服務獎勵措施 15 牙周統合照護 16 牙醫特殊專案醫療服務項目 19 牙醫其他專案 A3 預防保健 B6 職災案件 B7 代辦門診戒菸治療試辦計畫 1 一般案件 一般案件 2 論病計酬論病例計酬案件 3 特定案件特定案件 4 試辦計畫支付制度試辦計畫 5Tw-DRGs 6 安寧療護 7 愛滋病確診服藥滿 2 年後案件依本署 106.01.12 健保醫字第 1060032557 號書

欄位名稱	欄位中文名稱	欄位定義	備註
			函辦理 A1 代辦勞保局一般案件 A2 代辦勞保局論病例計酬案件 A3 代辦勞保局特定案件 A4 代辦勞保局支付制度試辦計畫 AZ 職業傷病住院膳食費 B1 代辦行政院衛生署精神病嚴重病人送醫及強制 C1 代辦疾病管制局 AIDS 案件 C2 結核病痰抹片陽性個案住院≤14 天內案件 (102/12/18 疾管愛核字第 1020003965 號函取消) C3 結核病 MDR 及抗結核藥物副作用住院 ≤6 C4 無健保結核病患之醫療費用 DZ 低收入戶住院膳食費
SEQ_NO	流水號	串接影像 8key	
ORDER_SEQ_NO	醫令序號	判斷案件是否為同一人	
ORDER_CODE	醫令代碼		
HOSP_CNT_TYPE	特約類別	分析醫學中心與區域醫院等不同層級個案量。	1 醫學中心 2 區域醫院 3 地區醫院 4 基層醫療單位 5 特約藥局 6 居家照護 7 社區精神復健 8 助產所 9 醫事檢驗機構 A 物理治療所 B 放射所 C 職能治療所 D 呼吸照護所
STUDY_DATE	檢驗日期	檢驗日期	
PATIENT_ID	病患 ID		加密 ID
PATIENT_AGE	年齡	個案年齡	
PATIENT_SEX	性別	個案性別	

欄位名稱	欄位中文名稱	欄位定義	備註
MODALITY	影像種類	影像種類	
CURE_PATH	診療部位	檢查的部位	H --診療部位 頭部 A --診療部位 頸部 F--診療部位 頭頸部 U --診療部位 胸部 N --診療部位 頸椎 J--診療部位 胸椎 K --診療部位 腰椎 G --診療部位 胸部及上腹部 C --診療部位 上腹部 I --診療部位 上腹部及骨盆腔 P--診療部位 骨盆腔 E--診療部位 肢部 Q --診療部位 臂神經血管叢 M --診療部位 骨骼肌肉系統 V --診療部位 心血管系統 S--診療部位 MRS(磁振頻譜) T--診療部位 MRA(磁振血管攝影) O --診療部位 其他
DEATH_CODE	死亡註記		Y—死亡
DEATH_M	死亡年月	死亡日期	
ICD1	主診斷代碼	疾病代碼	
ICD2	次診斷代碼(一)	疾病代碼	
ICD3	次診斷代碼(二)	疾病代碼	
ICD4	次診斷代碼(三)	疾病代碼	
ICD5	次診斷代碼(四)	疾病代碼	
StudyInstanceId	Study 識別碼	Study 識別碼	
SeriesInstanceId	Series 識別碼	Series 識別碼	