

NCHC Data Marketplace Introduction

<https://scidm.nchc.org.tw>

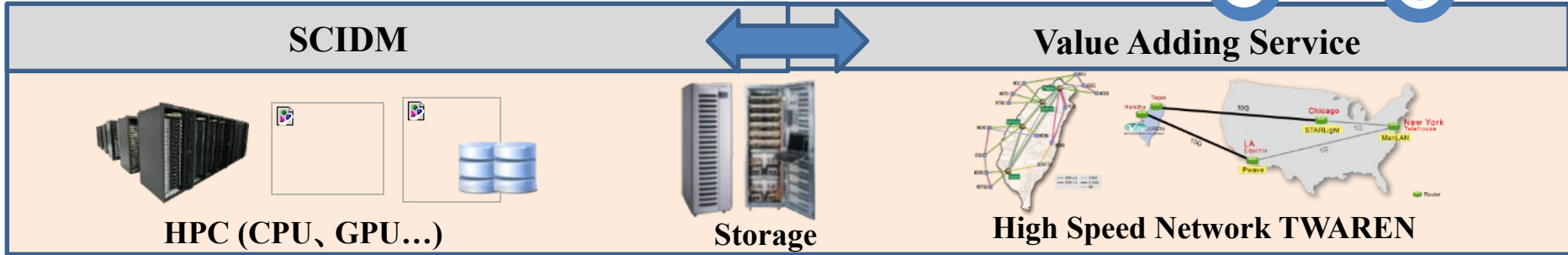
Data Service Team, NCHC



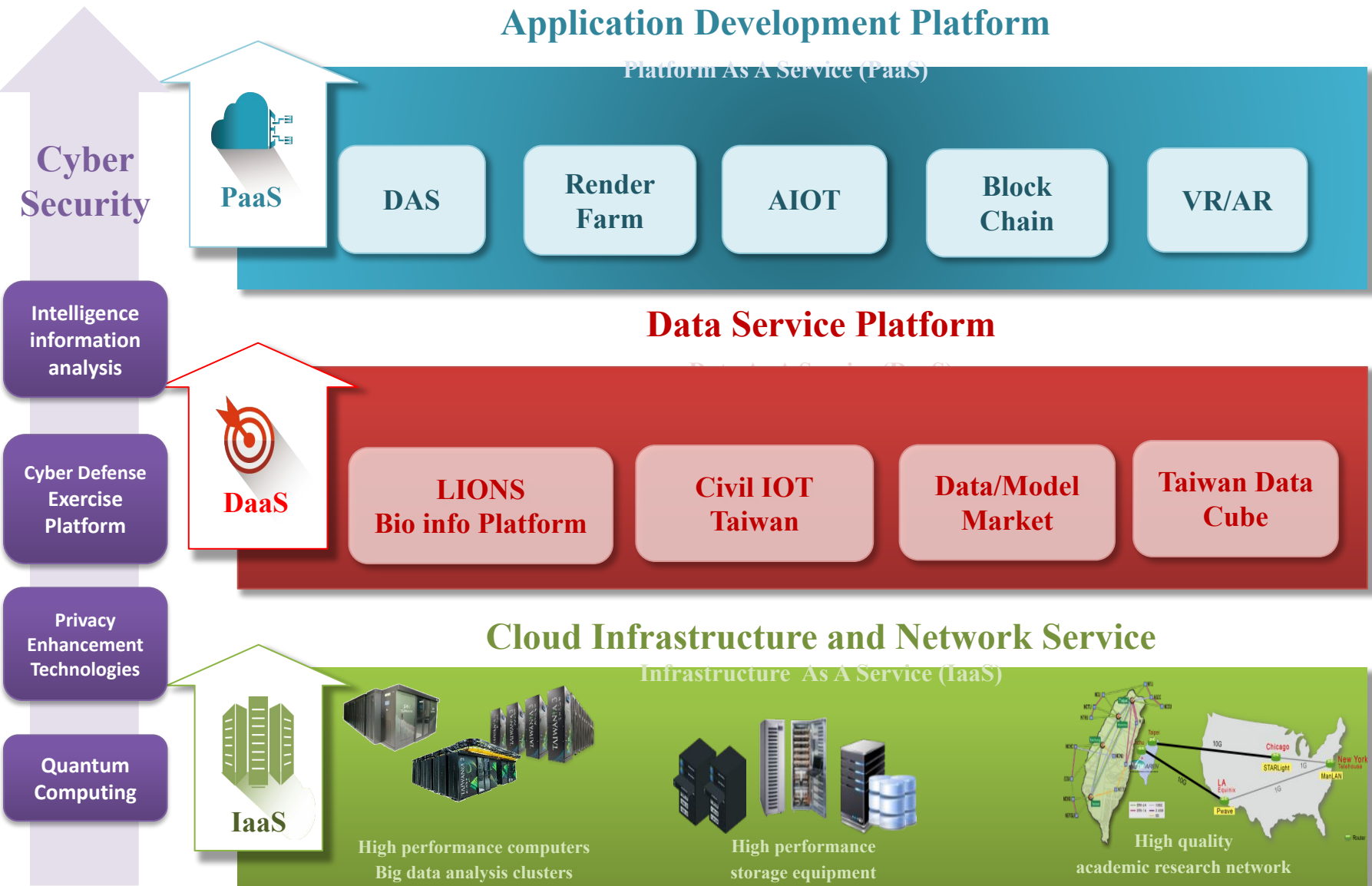
Outline

- Version of Data Service
- Introduction
- Current Status
- Feature Work

Data Service Overview



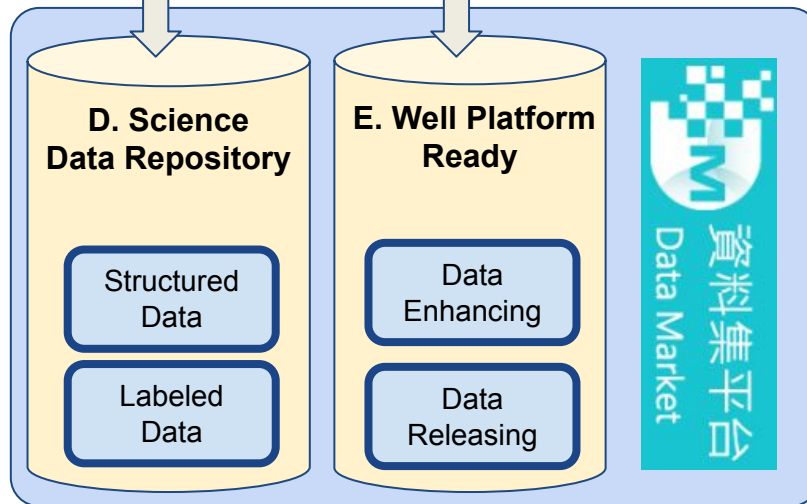
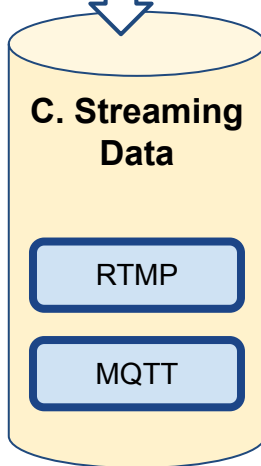
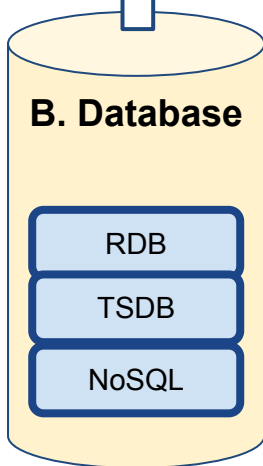
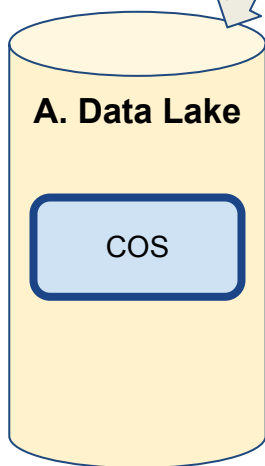
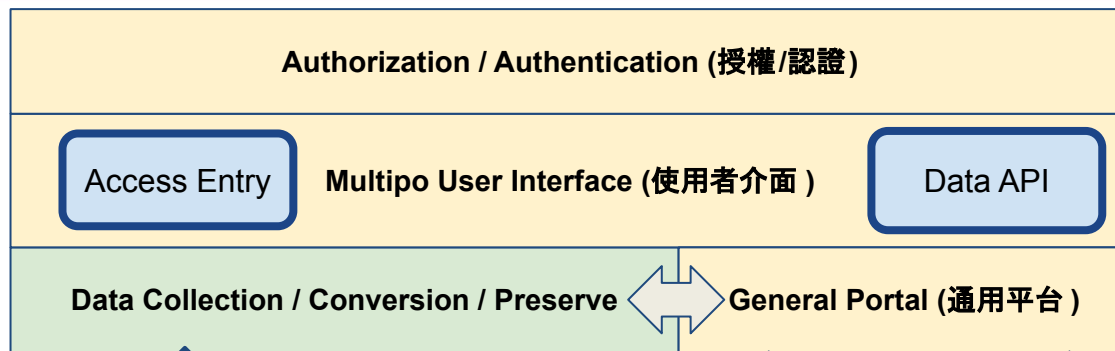
NCHC Core Service Structure



資料服務藍圖 (Updated in 2022, Q4)

(NCHC Data Service Blueprint)

- 僅需 Url + token 即可存取資料
- 透過 token 與安全機制, 確認使用者是否合法存取資料
- 針對不同協定適用的資料導入介面



For any format file-based data

ex : nchc s3

Data be present in the DB

rdb ex : PostgreSQL
NoSQL : cassandra
TSDB : Prometheus

For streamig/IoT data

MQTT: 即時訊息
RTMP: 影音串流

For labeled, well structured data (file-based)

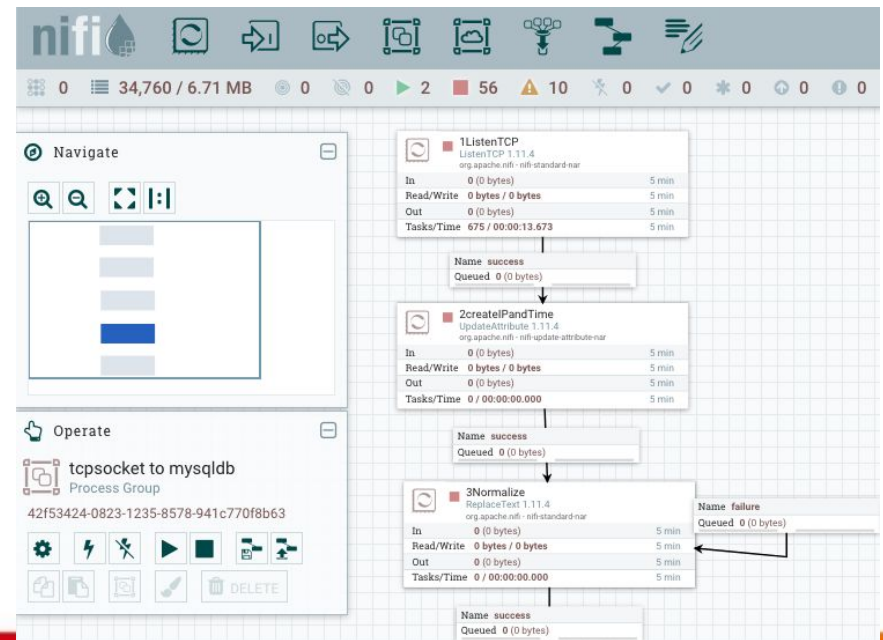
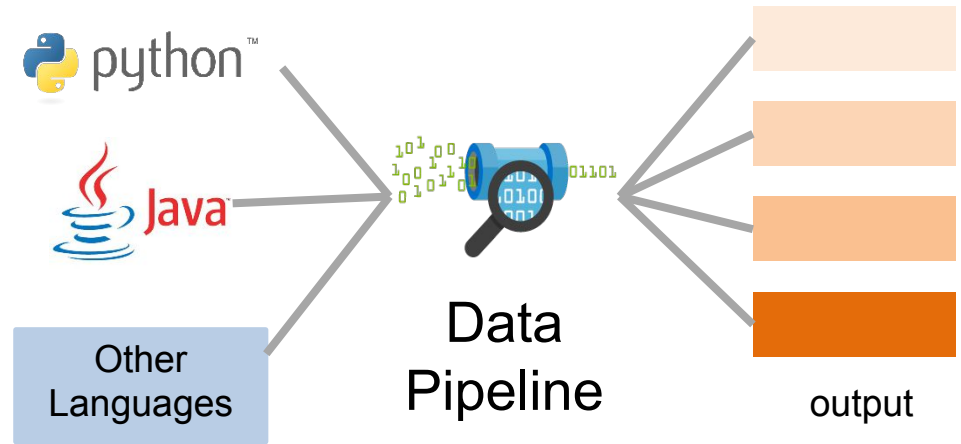
ex : 結構化資料與標記資料

Data stored in well functional platforms

資料加值 ex : DOI
資料交易 ex : 點數兌換

Data Service Framework

- Automatically create data workflow
 - Data extraction
 - Data processing
 - Information analytics
 - Data visualization
- Process and migrate data between different computing and storage service
- Users can drag and drop the components to perform the data stream on the web UI
- Command interface allows users to quickly create cloud services
- Users can execute functions written in different programming languages through the cloud platform



Outline

- Version of Data Service
- **Introduction**
- Current Status
- Feature Work

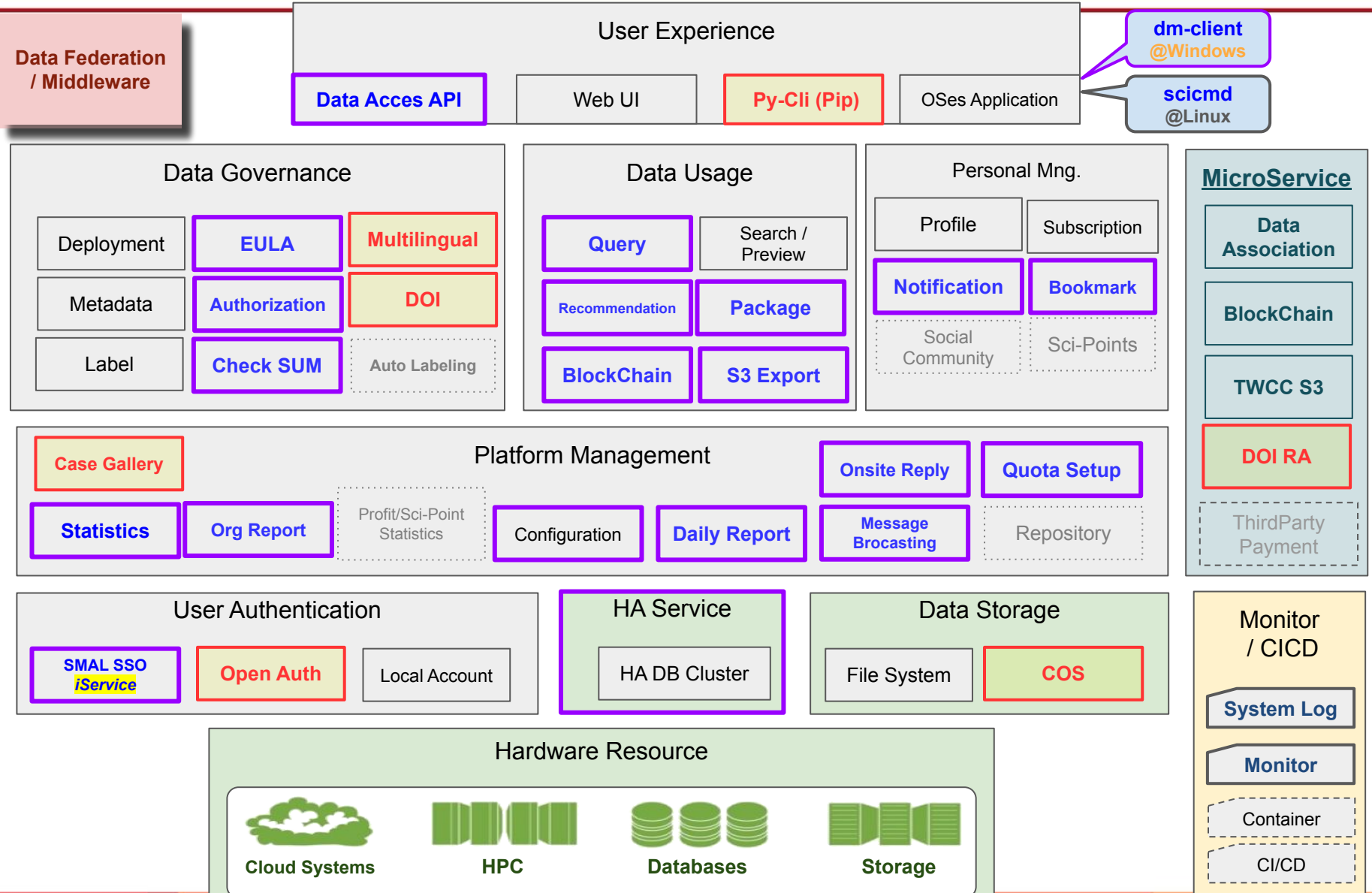
NCHC Data Marketplace: Sci-DM

<https://scidm.nchc.org.tw>

- Platform Goal :
 - Provide Data SHARING, Data Co-Management and Data Provision
 - Provide the connection platform between PEOPLE and DATA
- Function :
 - Provides access/management of data via Web UI/API
 - Provide interpretive data (metadata) management
 - Provide usage statistics query function
 - Provide data/organization/group/authority management functions
 - Provide information to apply for review/valuation process
 - Provide simple viewing function of structured data
- Target User : normal user / scientist
 - Data provider
 - Data user
- Basic System :
 - CKAN (Current : v2.7.6)

Data Market Architecture

(*Data PID/Multilingual/Py-Cli* function , **release at 2023Q3~Q4 , expected**)



NCHC Data Marketplace (Sci-DM)

I. 資料庫簡介 (英文)

In order to increase the capacity of national data services power, the NCHC builds a data service platform, named NCHC Data Mart , which integrates with NCHC's computing and storage resources.

The NCHC Data Mart platform : Sci-DM ,it is a data service platform established by the National Center for High-performance Computing (NCHC), aimed at facilitating data sharing, collaborative management, and service provisioning. The platform supports diverse data categories and file formats, offering various user interfaces for service access. Data release types include fully open, user agreement required, and user authorization required. Users can easily access and manage data, with support for interpretive data management and usage statistics queries. Overall, the platform's goal is to connect people with data, providing a secure and controllable framework for scientists, general users, and data providers.

NCHC Data Marketplace (Sci-DM)

1. 資料庫簡介 (英文)

The NCHC Data Marketplace (Sci-DM) is to provides the data service for co-sharing, co-management and co-construction with a security controllable platform.

2. Data Category:

- Transportation
- Public Info
- Government administration
- Audio/Voice
- Information Security
- Environment
- Academic
- Medical

3. 建置目的及日期(英文)

In order to increase the capacity of national data services power, the NCHC builds a data service platform, named Sci-DM , started since Q4 of 2017
Contact : Chenkai Sun (Ceasar)
<datamarket@narlabs.org.tw>

4. Release Type:

- **Fully open**
- **User Aggrement Required (only for registered user)**
- **User Authorization Required (only for registered user)**

5. How to use Sci-DM :

- Data user:
 - List、show、preview dataset and data
 - Search data via keyword
 - Download the physical resource (with the right)
- Data provider:
 - Create dataset and publish the resources via web UI
 - Management the data of metadata 、resources and access right.

Platform Manual :

<https://scidm.nchc.org.tw/pages/pg002>

6. 資料庫是否有資安防護措施 (英文)

- SSL/HTTPS in network transmission
- Firewall for host and storage resource

7. 智財權歸屬 (英文)

Copyright : Each data provider

8. 資料應用舉例說明 (英文)

- Geo-info added for Taiwan mask open data
- The application of satellite image data in science research

9. 資料庫/平台網頁:

<https://scidm.nchc.org.tw/>



NCHC Data Market: Sci-DM

See : <https://scidm.nchc.org.tw>

Goal : Provide Data SHARING, MNG. and BUILD platform

- **Feature :**

- Base on open source , full compatible with **CKAN 2.7**
- Provides multiple user interface : **Web/API/Oses client**
- Create lots of **data / mng. functions**
- Provide **4 control types of data releasing**
- Integration with **NCHC core tech/service (HA-DaaS/storage/BC/...)**

- **Target Clients: Provider / User**

- **Current Status:**

- Over **31k datasets , 20 external orgs** (GOVs, EDUs, COMs)
- Over **1000 data request form** in last 1 year , over **1.5M** download files per year.

Contact: datamarket@narlabs.org.tw

Data Category:

- Environment / Transportation
- Public Info
- Government administration
- Text/Audio/Voice/Corpus
- AI ready
- Academic
- Medical /Agriculture

Release Type:

- Fully open
- User Agreement Required (only for registered user)
- User Authorization Required (only for registered user)
- Chargeable (Sci-points)

Sci-DM Quick View



組織 / NCHC-資料服務專區

資料集 2 | 動態 30

最新專題

ImageNet ILSVRC2012

更新頻率: 不定期, 瀏覽次數: 3993, 下載次數: 239

Large Scale Visual Recognition Challenge 2012 (ILSVRC2012), 因資料授權因素, 本資料集同仁進行研究使用。若非本中心人員有使用需求, 請參考 ImageNet 官方原始授權 (<http://www.image-net.org/download-taq>) ImageNet is an...

MNIST

更新頻率: 不定期, 瀏覽次數: 2526, 下載次數: 222

THE MNIST DATABASE of handwritten digits mnist 手寫辨識資料 <http://yann.lecun.com/exdb/mnist/> in csv 格式, 出自於kaggle <https://www.kaggle.com/odrdonale/mnist-in-csv/data> Yann...

MNIST 示範展示資料

更新頻率: 不定期, 瀏覽次數: 8403, 下載次數: 231

該資料為 ai-market 專家展示示範使用

CIFAR-100

更新頻率: 不定期, 瀏覽次數: 1564, 下載次數: 57

The CIFAR-100 dataset This dataset is just like the CIFAR-10, except it has 100 classes of images each. There are 500 training images and 100 testing images per...

Data Dictionary

Column	Type	Label
Time	numeric	
溫度	numeric	
溼度	numeric	
紫外線	numeric	
風速	numeric	

其他資訊

欄位	值
Data last updated	十月 30, 2021
Metadata last updated	十月 30, 2021
建立	十月 30, 2021

彰化縣路口號誌經緯度資料.GeoJSON.csv

網址: <https://adm.nchc.org.tw/dataset/6458a8a8-859e-472e-ab5-50c70030c353/resource/Seed4d2-6eee-4961-ab11-8903ca1c700/download...>

新增篩選條件

Grid 圖表 地圖 3249 records 1 - 100 Search data... 篩選條件

地圖顯示了彰化縣路口的號誌經緯度資料，以藍色圓點標記在地图上。

欄位: Latitude / Longitude fields, GeoJSON field, Geometry field (GeoJSON), 經緯度, Update, Auto zoom to features, Cluster markers

應用欄: 智慧製造(AI), 人工智慧(AI), 智慧聯網(AIoT), 物聯網(IoT), 工業自動化, 自動駕駛, 飛行模擬, 教育學習, 科學研究, 模擬訓練, 資訊安全, 民生醫療, 基礎建設, 軍事國防

Meta Data /Dictionary / Preview/ Apply Form

從我 <datamarket@narlabs.org.tw> ★

主旨 <國網中心資料集平台>非公開資料集申請通知:[teeth0112]>>[六大毒蛇標示]

給 牙牙 ★

【國網中心資料集平台】 非公開資料集申請

組織管理員您好:
我是 牙牙 (帳號: teeth0112@gmail.com)
已於西元2021年11月12日申請六大毒蛇標示資料資料集
<https://scidm.nchc.org.tw/dataset/f61b5bd5-fcb-495b-a9cf>
申請說明:

Backend Management

• 帳號: 已審核 待審核 尚簽署 已刷單 全部

• 時間區間: 西元年/月/日 - 西元年/月/日

• 使用者: 查詢

匯出歷史CSV

審核	編號	資料集名稱	使用者名稱	單位	申請日期	地圖	應用領域	簽署日期	狀態
申請	非公開	Elley Huan	g	2020/06/12	生活	快	通		

- 01.學術用-需同意授權之使用條款
- 02.需審核之申請說明與使用條款
- 03.需付費之說明與使用條款

API/ Windows client

Services

<https://scidm.nchc.org.tw/api/> 國網中心資料集平台

```

get CKAN's Action API is a powerful RPC-style API that exposes all of CKAN's core features to API clients.

GET /action/tag_list 取得tag列表 - Get a list of tags
GET /action/status_show 取得網站狀態 - Get the site status
GET /action/package_list 取得所有套件 (或資料集) - Get a list of all packages (datasets)
GET /action/package_search 尋找符合查詢條件的套件 (或資料集) - Find packages (datasets) matching query terms
GET /action/package_show 取得指定套件 (或資料集) 之詳細資料 - Get metadata about one specific package (dataset)
GET /action/package_activity_list 取得指定套件 (或資料集) 之活動串流 - Get the activity stream of a package (dataset)
GET /action/package_activity_list_html 取得指定套件 (或資料集) 的活動串流及HTML格式 - Get the activity stream of a package (dataset) HTML format
    
```

資料集搜尋工具

方式一: 輸入設定檔

下載設定檔:

下載資料集名稱:

方式二: 下拉選擇

系統: 高中文 NOST-AI (for CIMs) | 資料集: [搜尋] 顯示申請審核 (demo-app)012 |

選擇下載目錄: C:\Users\cesan\Desktop\Download-SCDM

下載Log:

1 | NTThomas@Vimpep(6) | 國網中心資料集平台 | Copyright 2021

Poster for Sci-DM

SciDM 資料集平台

資料集平台為國網中心承接前瞻數位建設計畫下所打造的資料集平台服務，目的是讓資料科學能更易用與完整。除了持續擴充式開放與非開放資料集的資料使用與下載服務、也提供使用者上架資料集，歡迎分享使用。

非營利的
搜尋、發布、分享
之特色資料集平台
SCIENCE
DATA
MARKET

開放性

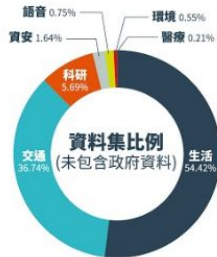
涵蓋多類之開放與特色資料，其開放資料不需任何帳號即可使用

集中性

資料整合計算與儲存資源於國網中心提升資料使用效率

唯一性

唯一提供科技部學研資料、科專計畫資料成果



特色資料集

通用智慧醫療、智慧語音、環境防災、智慧製造發展所需資料集

容易性

平台操作簡單直覺註冊會員只需三分鐘

免費上傳資料集

成為會員後管理員審核通過，即可上傳



資料市集團隊
datamarket@narlabs.org.tw
https://scidm.nchc.org.tw



NAR Labs 財團法人國家實驗研究院
國家高速網路與計算中心
National Center for High-performance Computing

SciDM 資料集平台

特色
資料集



想做AI智慧醫療 醫療影像資料集

- 國際TCIA癌症醫學影像庫 (免費開放)
- 美國衛生研究院NIH醫療影像庫 (免費開放)
- 科技部AI醫療影像資料庫 (申請授權使用)
- 衛福部疾管署-登革熱、毒蛇標示影像 (免費開放)



練習深度學習 訓練AI資料集

- 國網中心 AIDM 資料集 (免費開放)
- UCI 資料集 (免費開放)



認識周遭的生活 環境資料集

- FLOSS 開放社群-空氣盒子即時量測資料 (免費開放)
- NOAA氣象資料 (免費開放)
- 科學園區空氣汙水用電數據 (免費開放)
- 氣象局 pm 2.5 資料集 (免費開放)



發展人工語音不可少 語音資料集

- 科技大播台語賽AI語音數據集 (申請授權)
- 中研院語料庫：廣義知網中文詞知識庫、中文新聞語料庫、漢語平衡語料庫、中文詞庫、中研院中文句結構樹資料庫3.1版、中文分詞語料庫、具音段標記之中文對話語音資料庫、口語翻譯語料庫暨工具平台庫、亞洲口音英語跨國語音資料庫 (主機學研帳號免費使用)



其他特色資料集

- 風傳媒：傳媒資料集 (專案合作)
- 太空中心：衛星影像圖集 (申請授權)
- 洞見資料集：專家整理的政府相關資料

最有特色的資料集 可能就是
你的資料集

- 免費上傳、分享、使用
- 提高資料集的能見度
- 國家品質機房維護



資料市集團隊
datamarket@narlabs.org.tw
https://scidm.nchc.org.tw



NAR Labs 財團法人國家實驗研究院
國家高速網路與計算中心
National Center for High-performance Computing



Service Contact

- Platform:
 - <https://scidm.nchc.org.tw/>
- Manual :
 - <https://scidm.nchc.org.tw/en/pages/pg002>
- Email
 - datamarket@narlabs.org.tw

