The Status of Data Marketplace — on its way to the National Data Service

Thomas Tsai, Ceasar Sun, Steven Shiau, Yen-Wei Huang, Weicheng Huang*

{thomas, ceasar, steven, ellery, whuang}@nchc.narl.org.tw

NATIONAL CENTER FOR HIGH-PERFORMANCE COMPUTING, NARL, TAIWAN

Abstract

Based on an existing open source project, CKAN[1], a platform, the SCIDM, of the data marketplace is deployed with minimum efforts. To further enhance the user experience, several features such as API and synchronization function, the Ckansync, were developed.

Currently, various data sets are available on the platform, including open data as well as those with limited access control.

The SCIDM is currently under the integrating process with computing resources to provide an one-stop service for data related research to the data research scientists and experts from enterprise alike. In the near term, FUSE[2] will be implemented for Ckansync and provide mount feature to the users thus to provide working scenario that is familiar to the users.

Data

Data Access
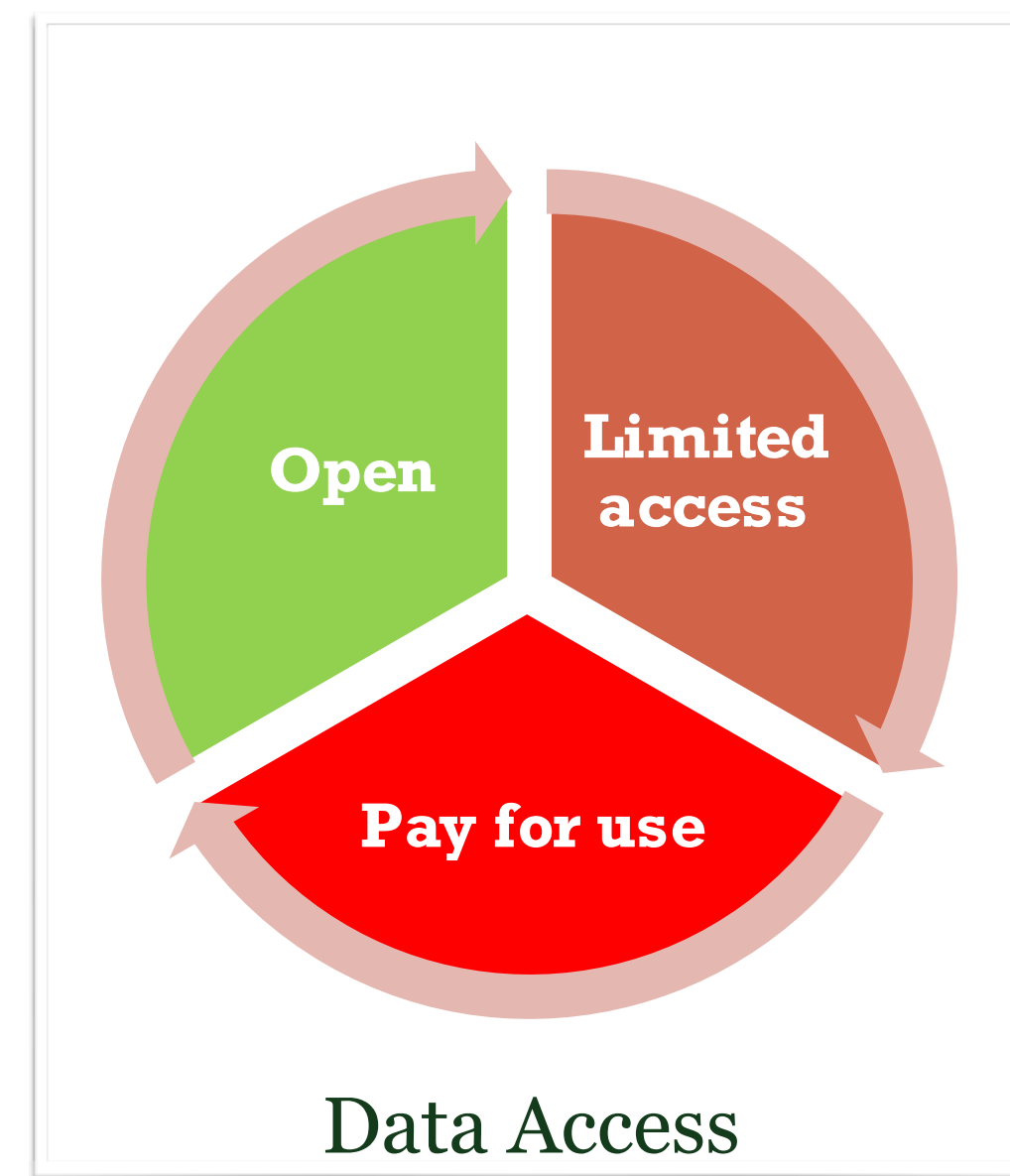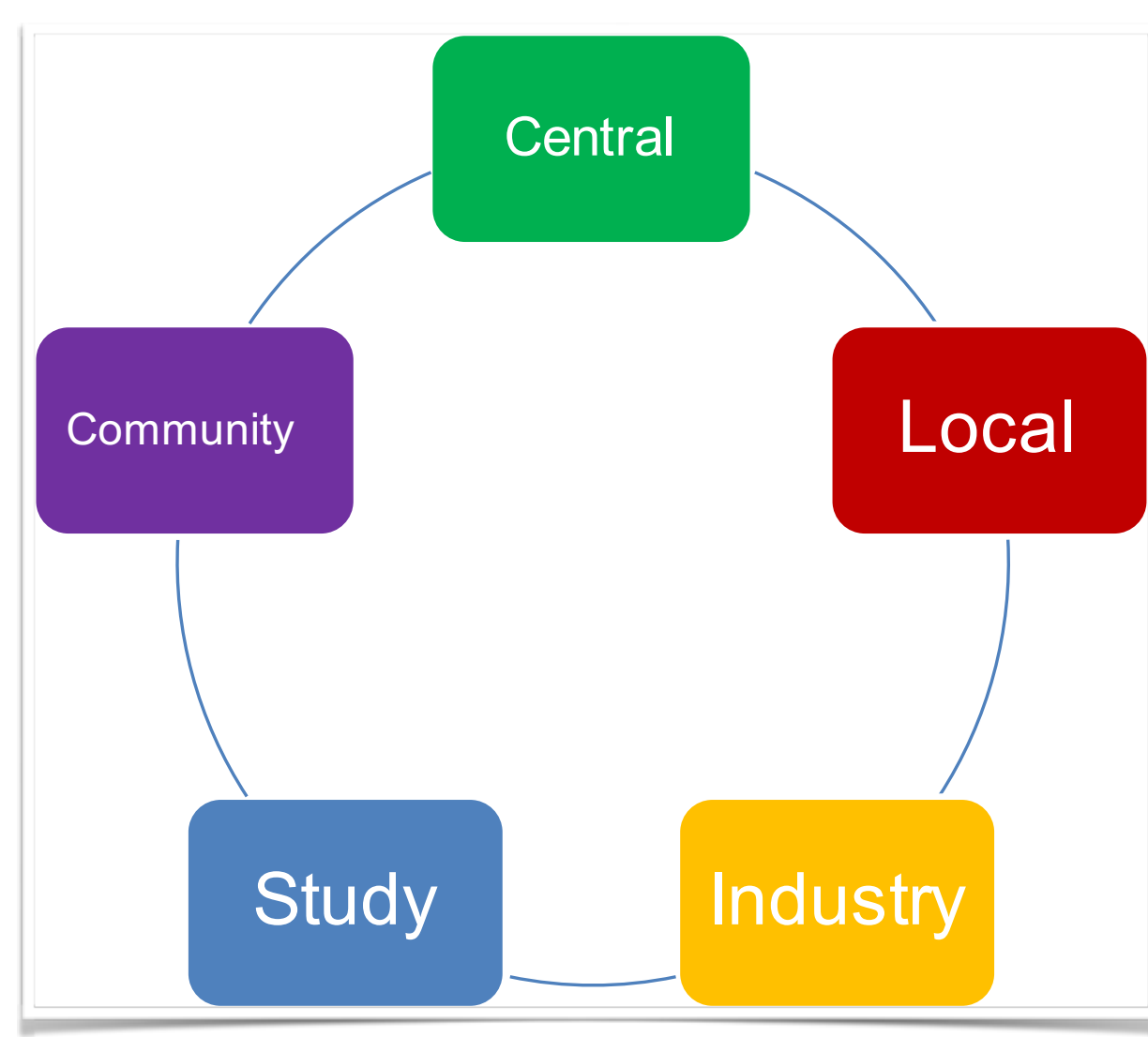
Datasets' Feature:

- Open : a variety of open data
- Centralized : data are next to the computing facility
- Uniqueness : research data, government data

New Datasets to come:

- medical images, disaster mitigation data, media data, satellite images

Service online

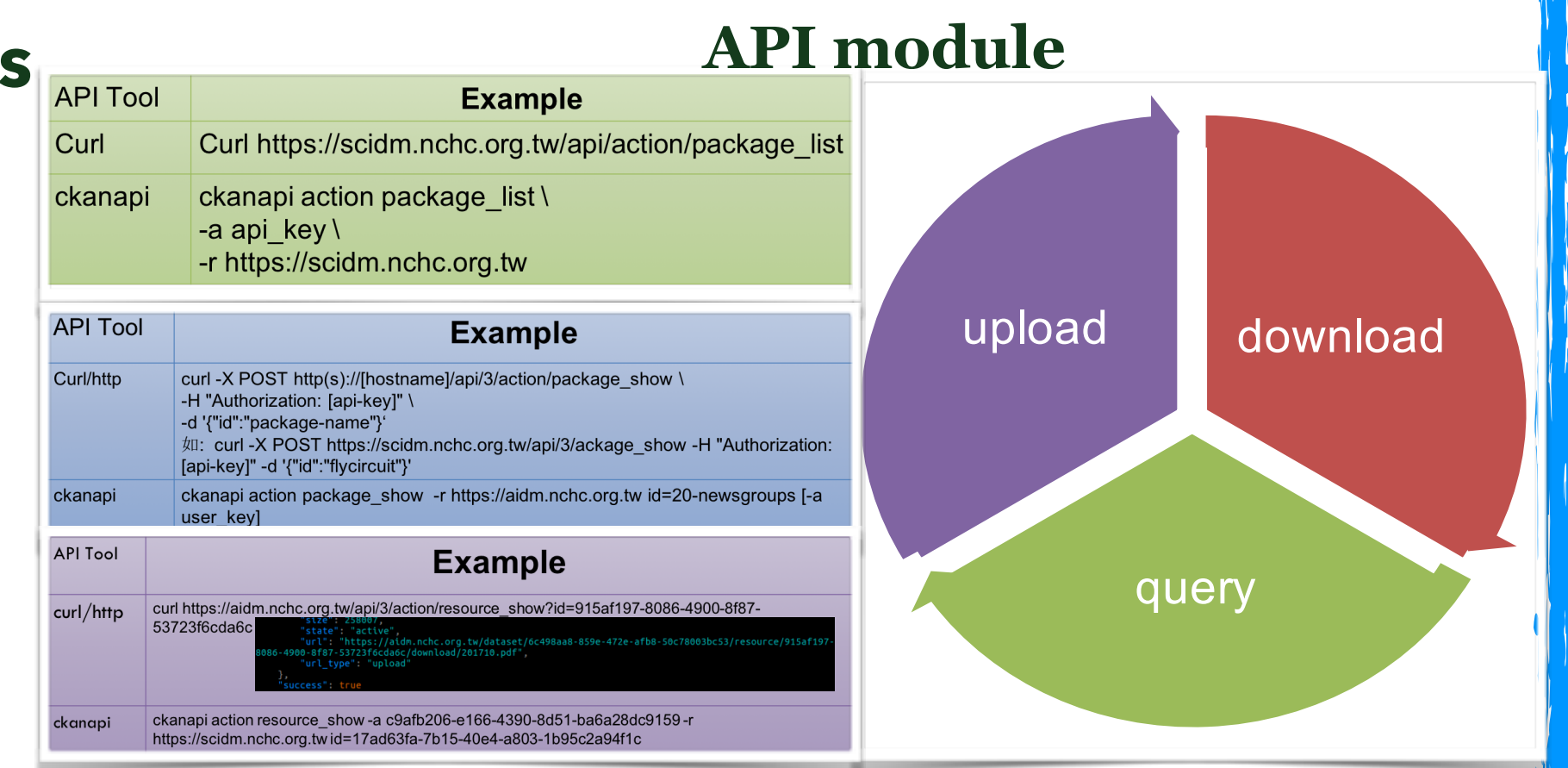The Data Service Platform is official online.

References

[1]. https://github.com/KSP-CKAN/CKAN

[2]. https://github.com/libfuse/libfuse

API Gateway

CKAN API commands

- datasets' lookup
- data resource list
- file download via wget/http/curl

API module

Application of API

air quality data
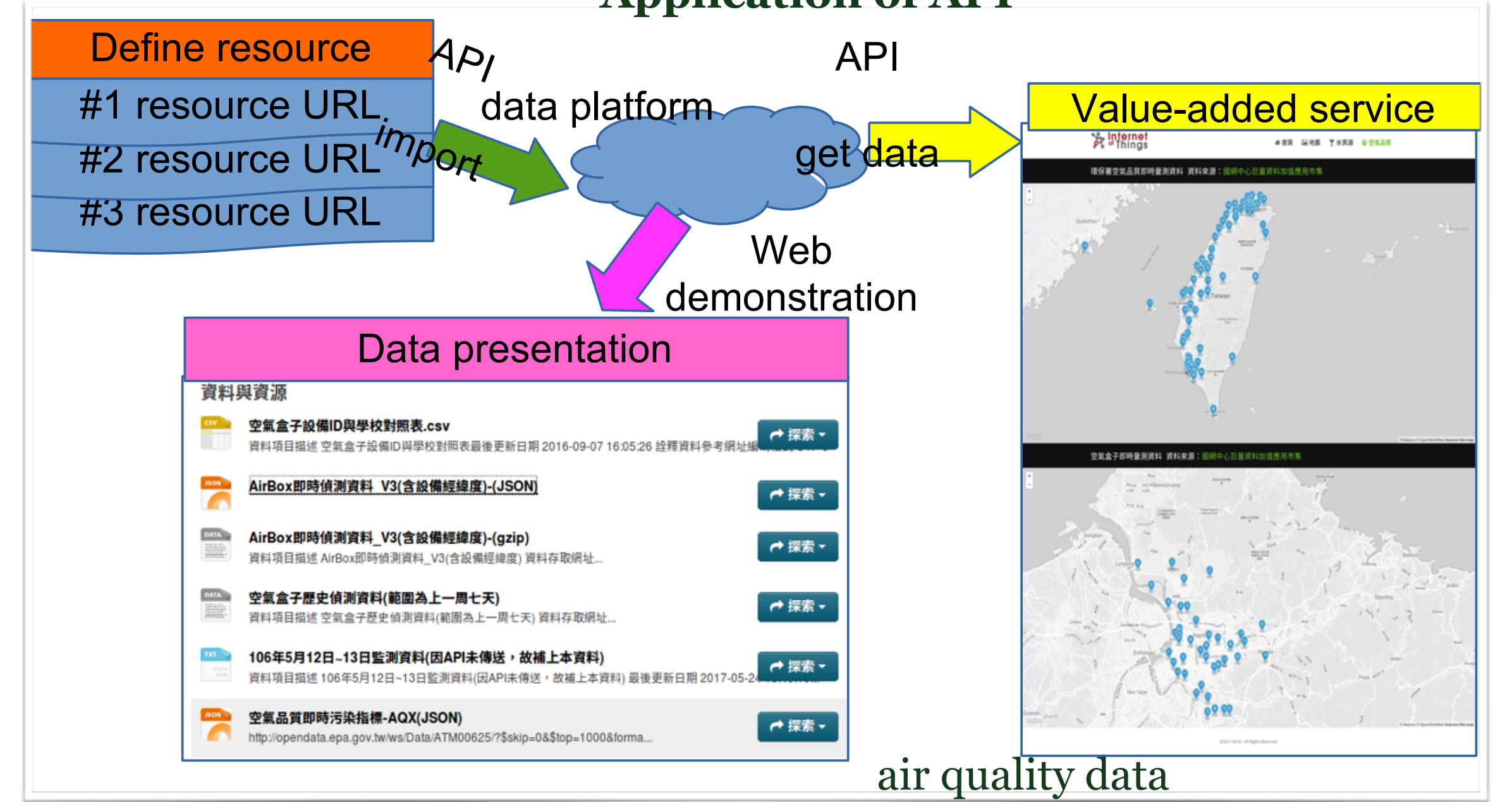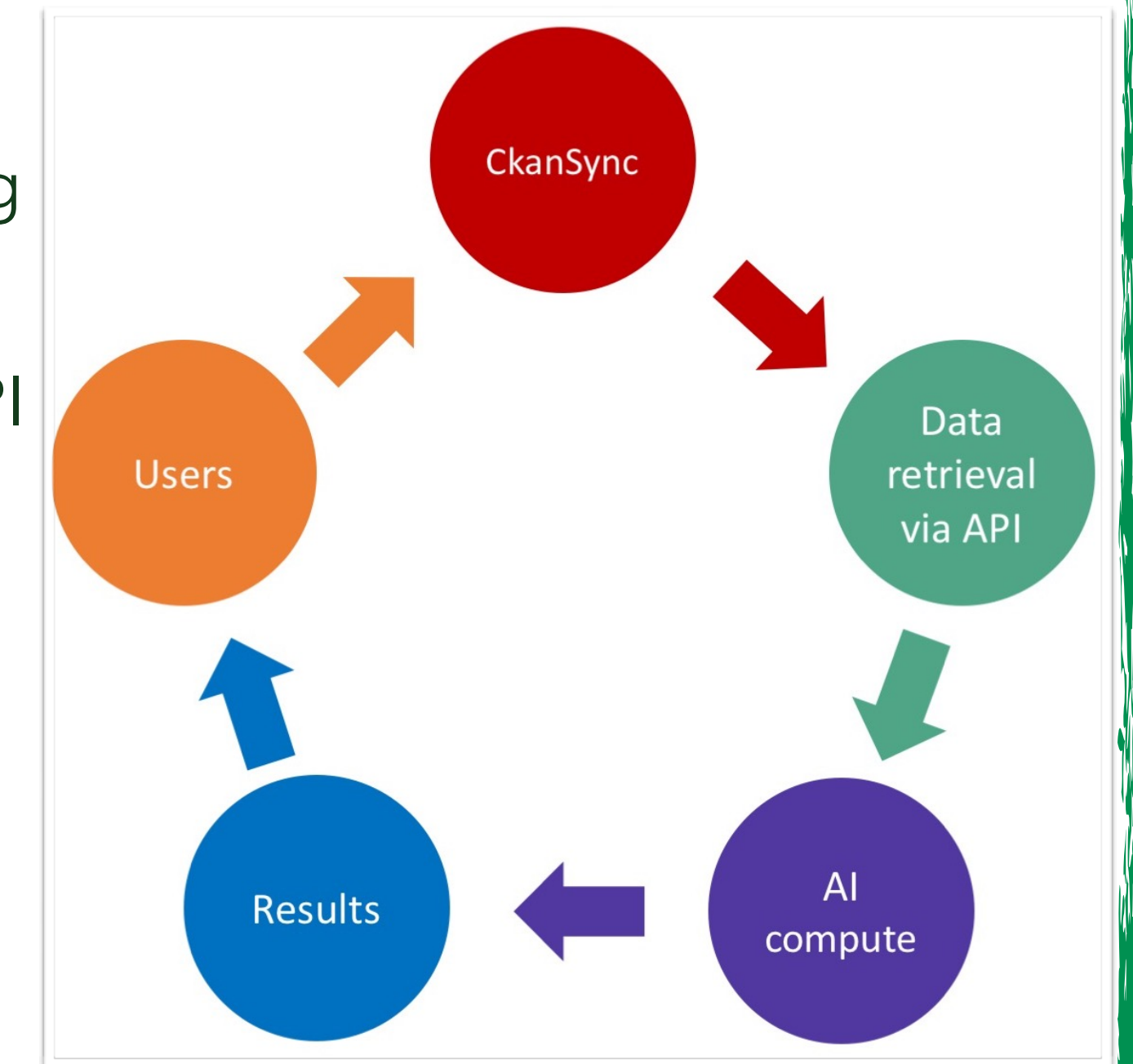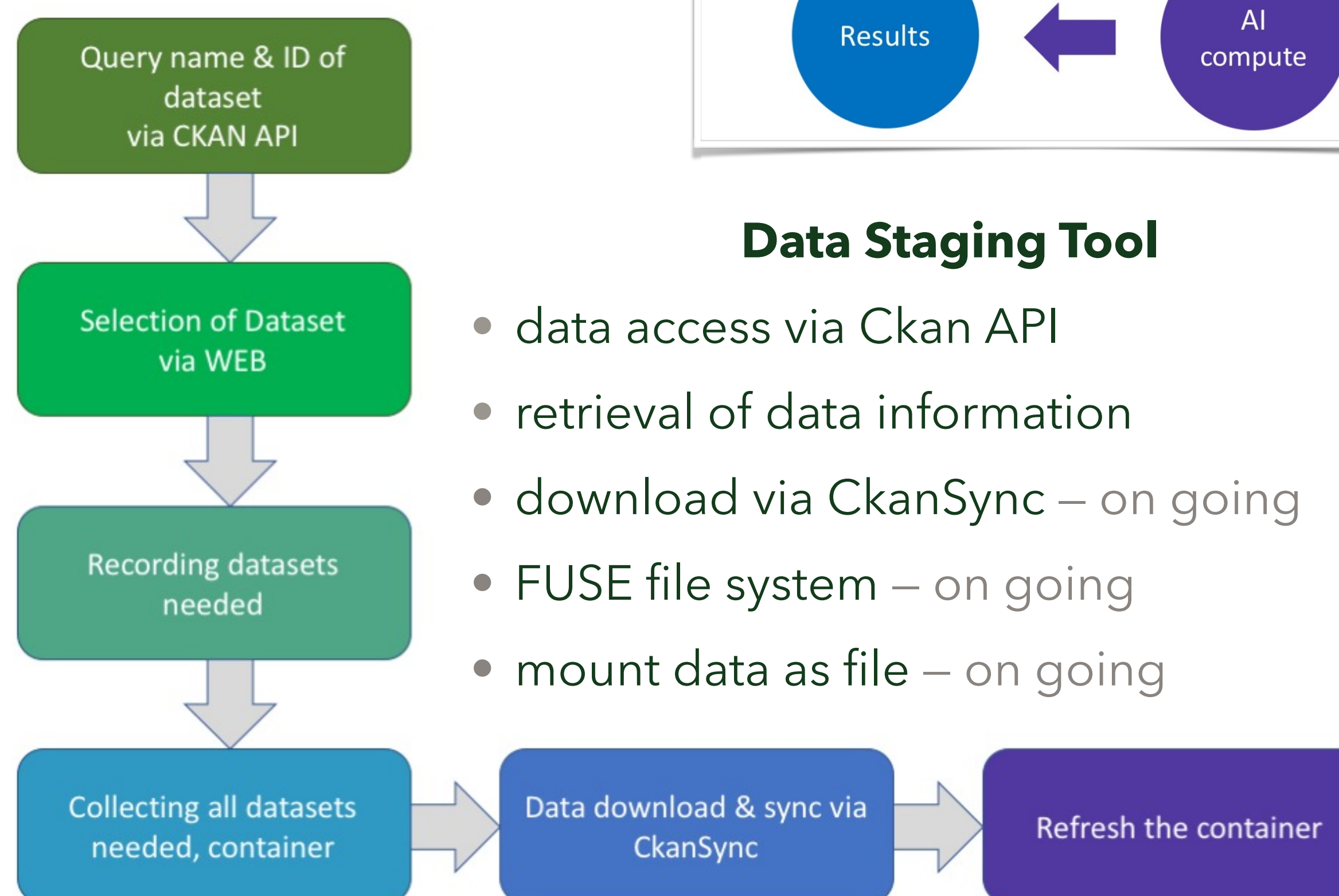
CkanSync

CKanSync-API

- data integration w/computing machines
- data transfer tool via Ckan API
- integration of web & computing power

Query name & ID of dataset via CKAN API → Selection of Dataset via WEB → Recording datasets needed → Collecting all datasets needed, container → Data download & sync via CkanSync → Refresh the container
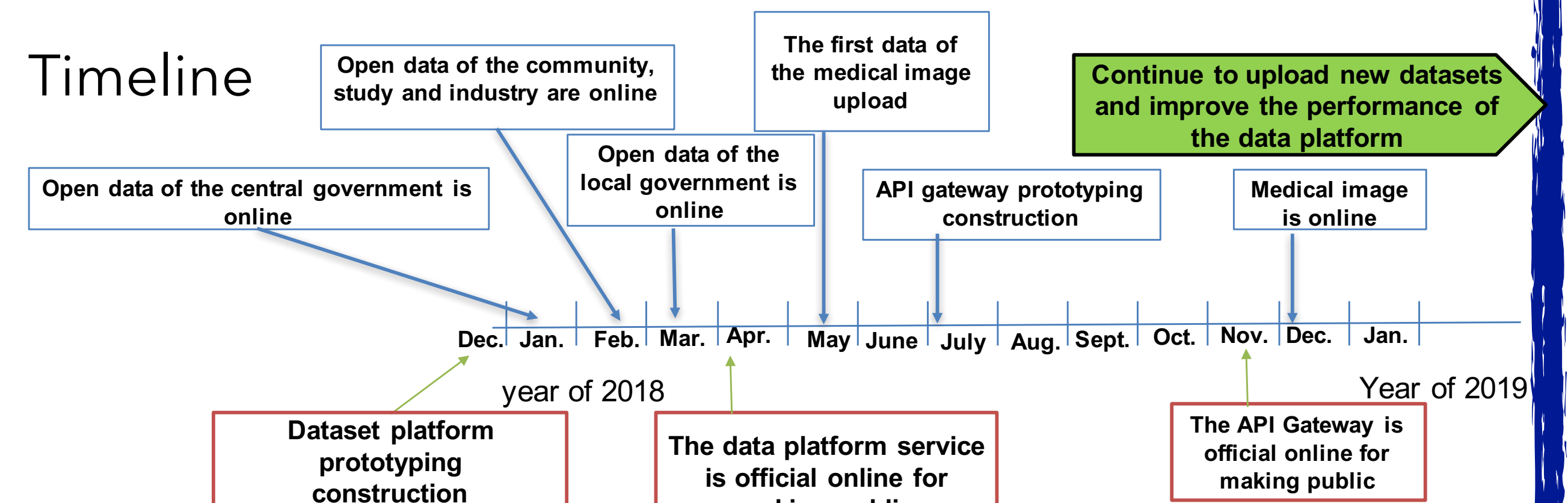
Data Staging Tool

- data access via Ckan API
- retrieval of data information
- download via CkanSync – on going
- FUSE file system – on going
- mount data as file – on going

Future Work

- Timeline

- Near term

   * Integration with Computing facility
   * FUSE as file system
   * CkanMount to mount data