CKAN 同好會

議題討論綱要

2018 研究資料管理工作坊
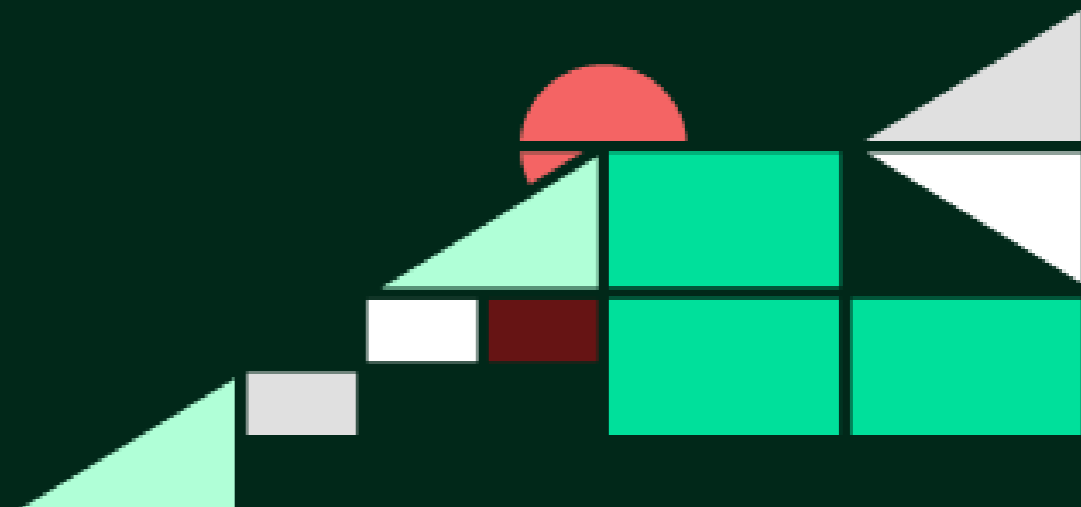
李承鑫 (Sol)

中研院 資創中心

December 13, 2018

今天實際報告內容（CKAN / depositar 簡介）

https://docs.depositar.io/zh_TW/6.3.3/user-guide.html

About Me

- Data analysis, GIS programming, information system design, and database management
- Working on depositar
- Slides Archives

  https://www.slideshare.net/u10313335/presentations

About Me (Cont'd)

- Contributions to CKAN (part)
  - https://github.com/ckan/ckan/commits?author=u10313335
  - https://github.com/ckan/ckanext-geoview/commits?author=u10313335
  - ckanext-geoview: wmts_view and shp_view
  - ckanext-wikidatakeyword: Wikidata-powered keyword field for CKAN

Outline

- Extensions

- UI/UX

- CMS Integration

- Interoperability

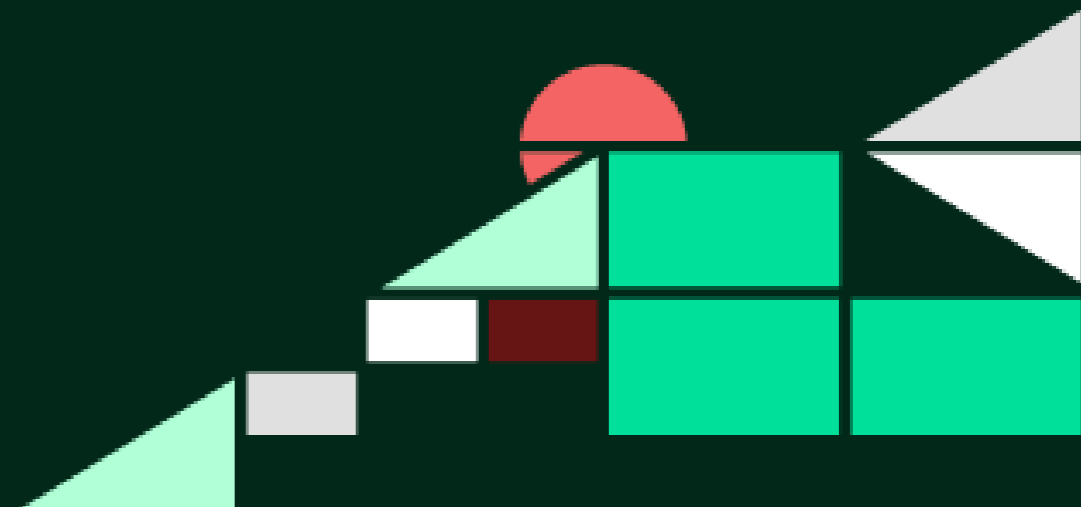- Deployment & Maintenance

- Promotion

- General Discussion
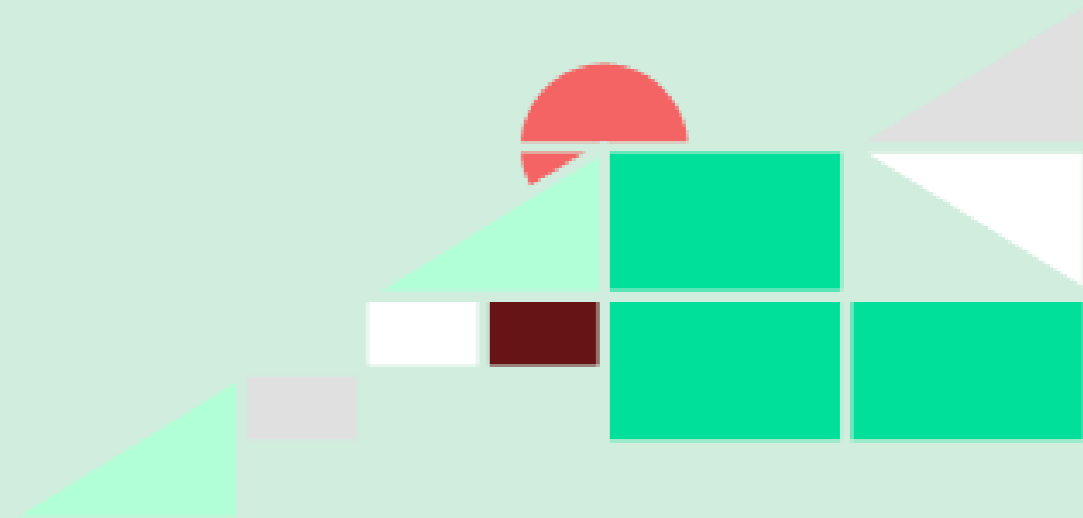
CKAN Core

(We use v2.7.x for now.)

https://github.com/ckan/ckan/

depositar customized extensions

https://github.com/depositar/

Extensions

Extensions and Plugins

Each extension contains one or more plugins.

E.g. ckanext-geoview includes geojson_view,

wmts_view, geo_view…, etc.

Find extensions & plugins

- https://extensions.ckan.org/
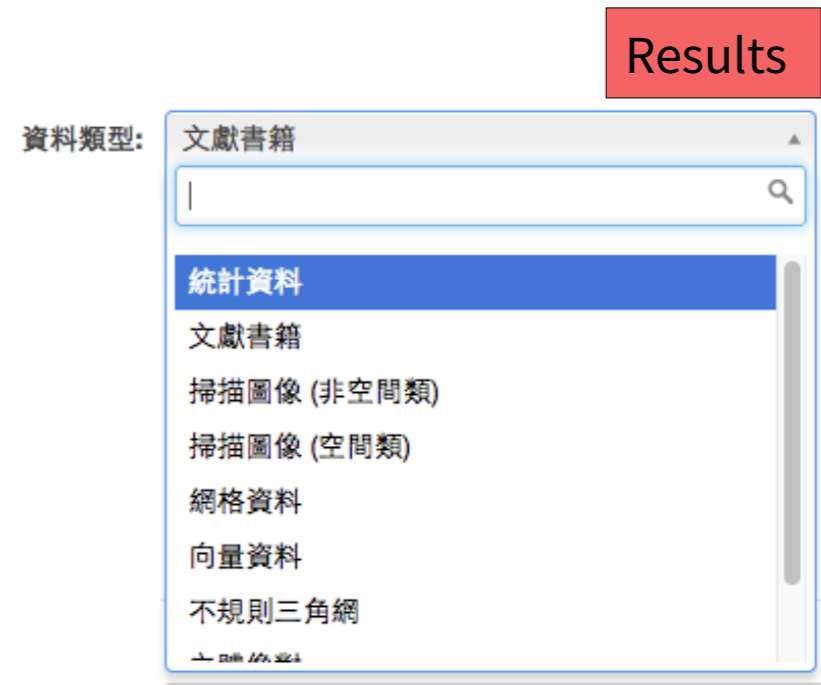
- SITE_URL/api/action/status_show

Internal Plugins

- stats

- multilingual_dataset/group/tag/resource

- datastore

- datapusher

- text_view, recline_view, recline_map_view, image_view, webpage_view…

- example_*

- … and more

https://github.com/ckan/ckan/blob/ckan-2.7.4/setup.py#L96

ckanext-scheming

```
{

 "field_name": "data_type",

 "label": {"en": "Data Type", "zh_TW": " 資料類型 "},

 "preset": "select",

 "form_attrs": {"data-module": "autocomplete"},

 "choices": [

  {"value": "statistics", "label":{"en": "Statistics", "zh_TW": " 統計資料 "}},

  ...

 ]

}
```

資料類型:　文獻書籍

統計資料
文獻書籍
掃描圖像 (非空間類)
掃描圖像 (空間類)
網格資料
向量資料
不規則三角網

```
{

 "preset_name": "select",

 "values": {

 "form_snippet": "select.html",

 "display_snippet": "select.html",

 "validators": "scheming_required scheming_choices"

 }

}
```
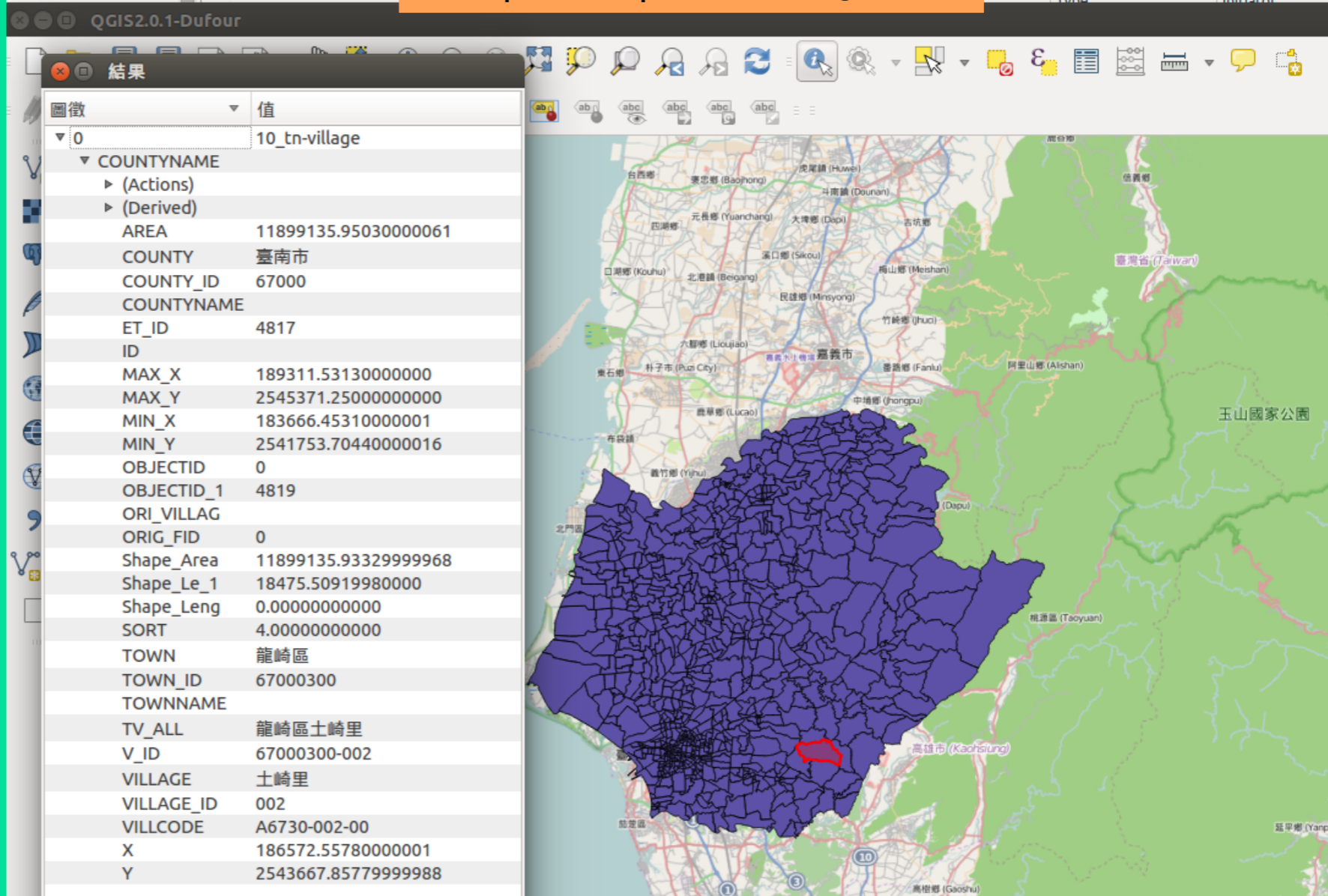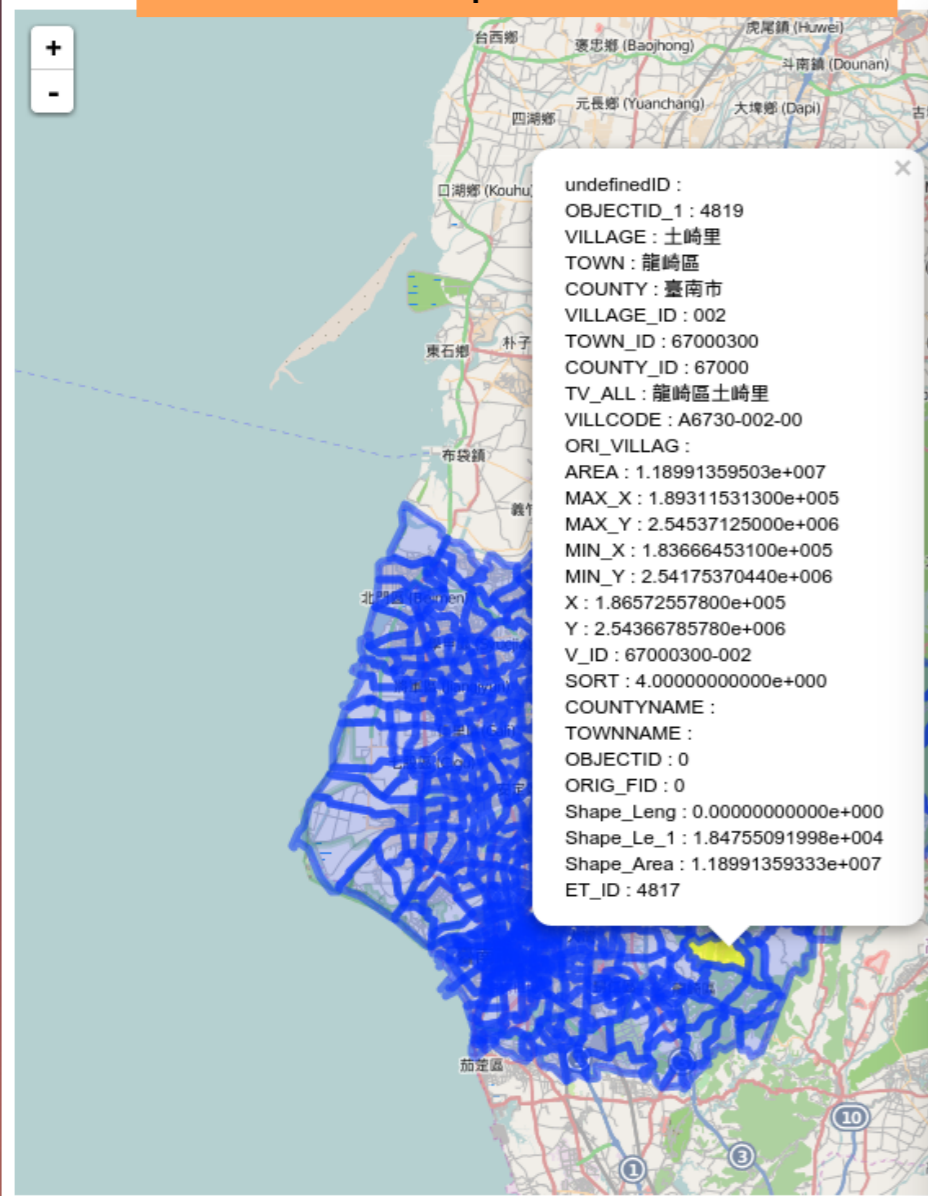
11

https://github.com/ckan/ckanext-scheming

ckanext-geoview: shp_view

https://github.com/ckan/ckanext-geoview/pull/54

CKAN Shapefile Preview

Open Shapefile with QGIS

https://data.depositar.io/dataset/proj4-29/resource/c2bbe543-dc9d-41c2-82d5-d16a93b87835
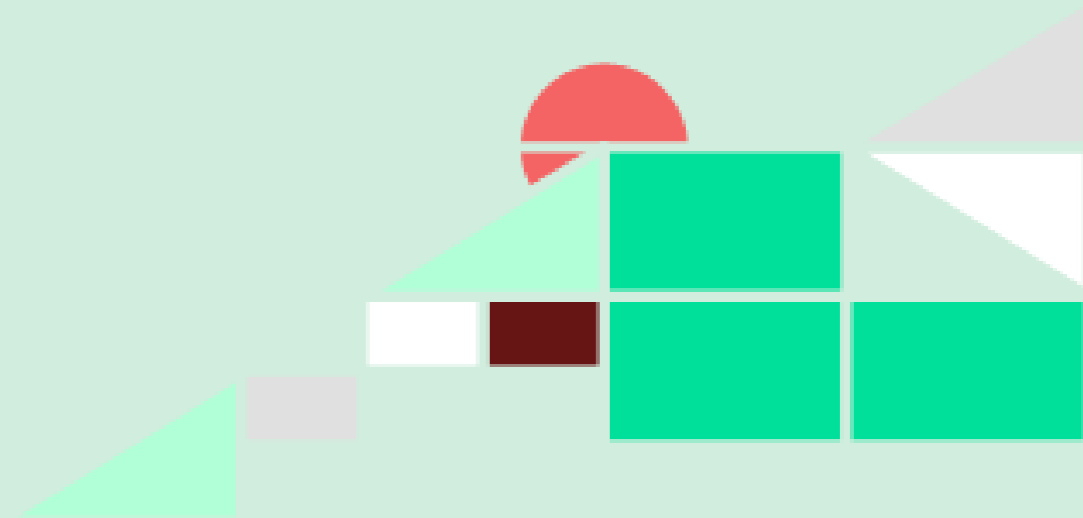
討論

- 是否有其他覺得實用的套件，或覺得應該要有但沒有的套件

- 格式預覽是擴充套件的大宗，但是否仍有不足?

- 是否有自行開發套件的經驗? 若想擴充功能，會透過自行開發套件完成嗎?

UI/UX

Visuals

- Jinja 2 Template

- CSS (Less.js)

- CKAN JavaScript Module

i18n

Translation Target	Related Modules
UI Strings	Babel & i18n folders (GNU gettext)
Metadata	ckanext-scheming extension
Search facets Dataset / Group /Organization title	Internal multilingual_* plugin
Keyword search	Solr schema + tokenizer (e.g. mmseg4j)
License list	A json file

Note: This is not an exhaustive list.

i18n (cont'd)

- Column name length limit for datastore upload
  - https://github.com/ckan/ckan/issues/2804

  - PostgeSQL's 63 characters limit

  - Python's len() method: Not for UTF-8

- zh_Hant_TW and zh_Hans_CN (CKAN 2.8.x)
  - https://github.com/ckan/ckan/issues/4413
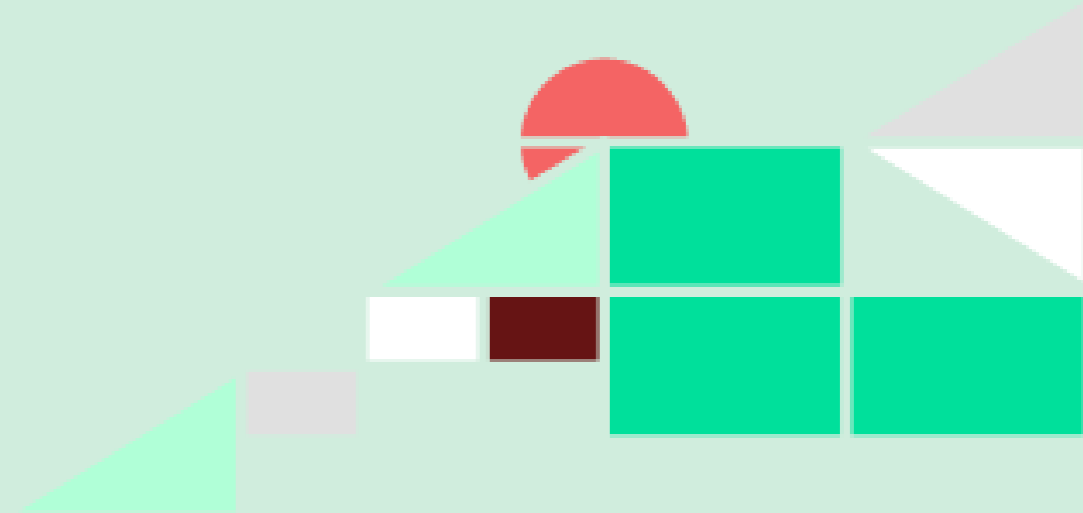
Registration

- Email Verification
  - depositar/ckanext-data-depositario@da25b14

- SSO
  - OAuth 2
    - https://github.com/conwetlab/ckanext-oauth2
    - ORCID
  - LDAP
    - https://github.com/NaturalHistoryMuseum/ckanext-ldap

討論

- 是否有介面客製化的經驗？

- 若要進行介面客製化，會傾向
  - 直接修改 CKAN CSS 或 JavaScript
  - 結合 CMS (Drupal / WordPress)
  - 其他
- 使用 CKAN 上覺得不順手的地方？

3 / 6

CMS Integration
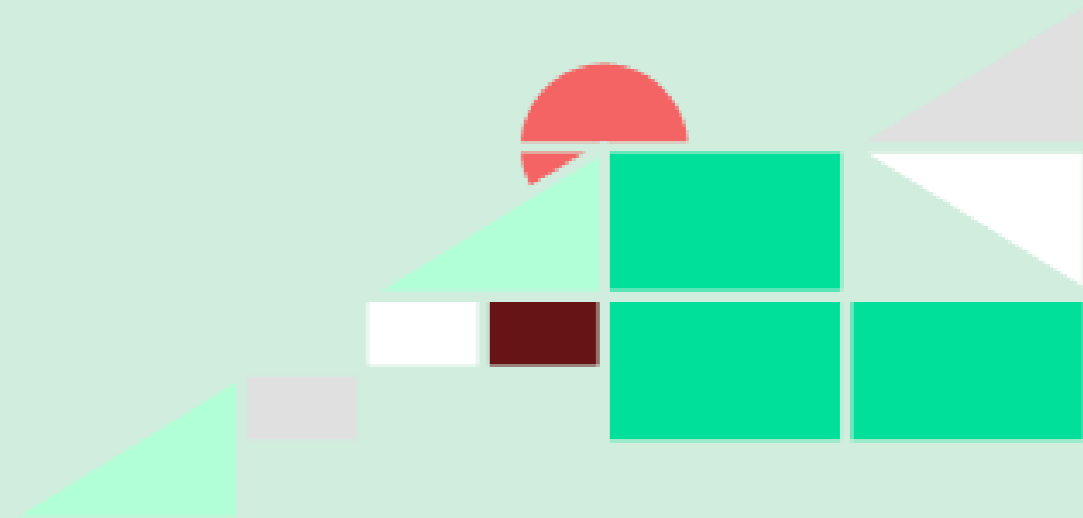
CMS Integration

- Pros & Cons

- WordPress
  - opendata.swiss
    - Their slides
  - data.gov

- Drupal
  - data.gov.uk (old version)
  - G Space Information Center (Japan)
    - blog post: CKAN vs. CKAN + Drupal vs. **DKAN**

CMS Integration

- CMS 整合的挑戰

Interoperability

ckanapi

- A command line interface and Python module for accessing the CKAN Action API

- https://github.com/ckan/ckanapi

- Bulk dumping and loading (JSON lines text files), deleting

- Useful when doing metadata migration (bulk updating)

- vs. CKAN's Data API

- Possible WebUI?

ckanext-harvest

- Provide a common harvesting framework for CKAN extensions and adds a CLI and a WUI to CKAN to manage harvesting sources and jobs.

- https://github.com/ckan/ckanext-harvest

ckanext-harvest (Cont'd)

```python
from ckan import plugins as p

class SRDAHarvester(SingletonPlugin):

    p.implements(IHarvester)

    def info(self):
        (...)

    def gather_stage(self, harvest_job):
    (…)

    def fetch_stage(self, harvest_object):
    (…)

    def import_stage(self, harvest_object):
    (…)
```

Python Plugin

Get the list of datasets

Fetch datasets

Import datasets

*SRDA=Survey Research Data Archive
（中研院人社中心調研中心學術調查研究資料庫）

Source type:
- ● CKAN
- ○ SRDA
- ○ CSW Server
- ○ Web Accessible Folder (WAF)
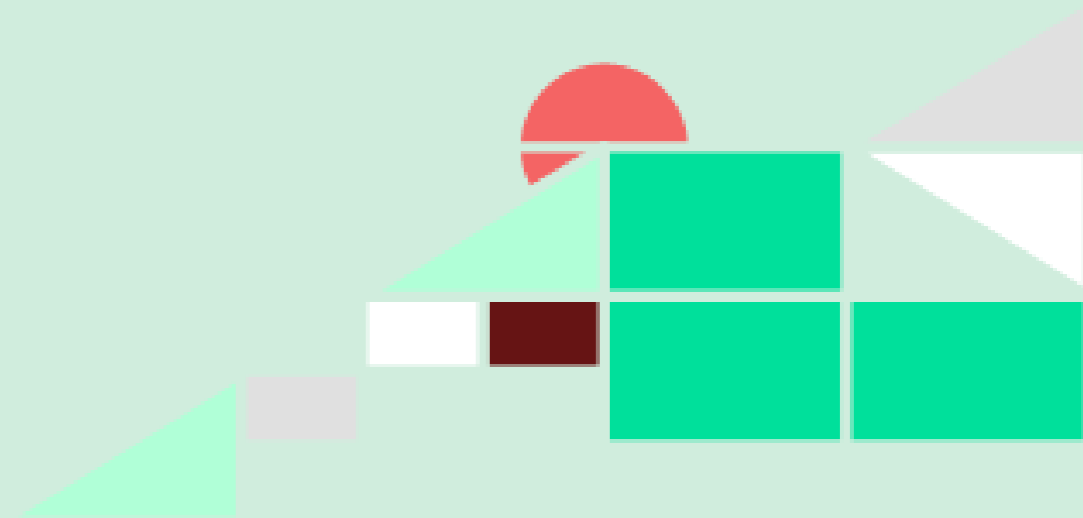- ○ Single spatial metadata document

Results

Source: http://goo.gl/LS79s4

ckanext-dcat

- Allow CKAN to expose and consume metadata from other catalogs using RDF documents serialized using DCAT.

- https://github.com/ckan/ckanext-dcat

- Google Dataset Search compatible

ckanext-doi

- DOI (Digital Object Identifier) integration

- https://github.com/NaturalHistoryMuseum/ckanext-doi

Deployment & Maintenance

Depositar's Current Setup

- A KVM image in a rack server

- 8 cores CPU, 16 GB RAM, 900 GB disk

- Ubuntu 14.04.5 LTS

- AP and DB run on the same machine
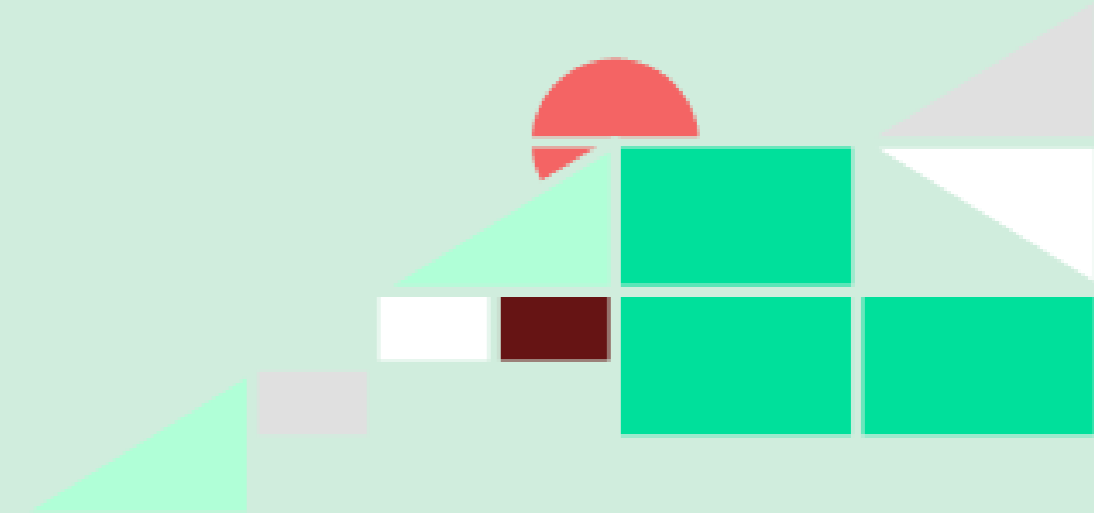
ckanext-security

- A CKAN extension to hold various security improvements for CKAN

- https://github.com/data-govt-nz/ckanext-security

討論

- Improve performance
  - Remote database
  - Load balancing

- Where's the limit?
  - Performance tips for large imports

  - CKAN tested on 2 million datasets

- Containerization for easier deployment
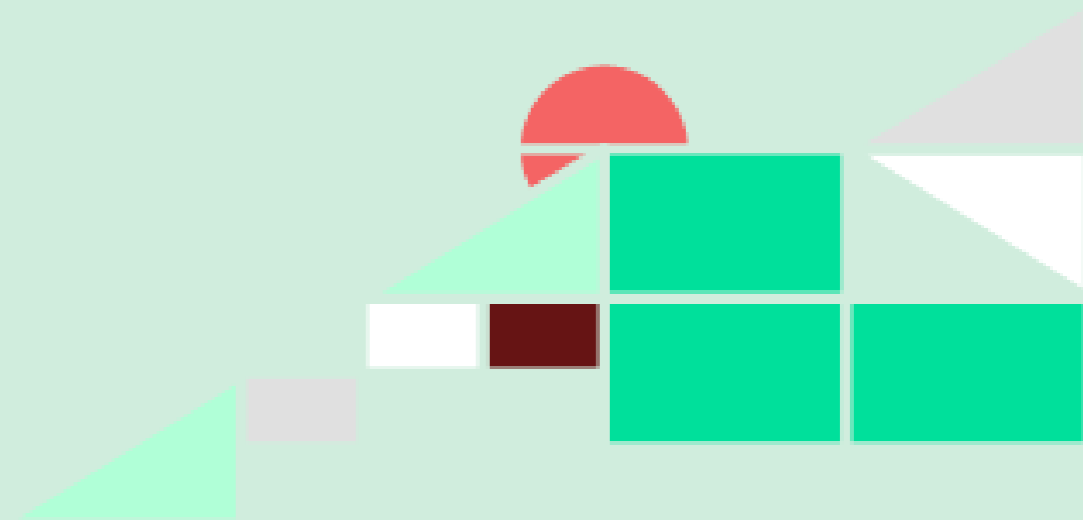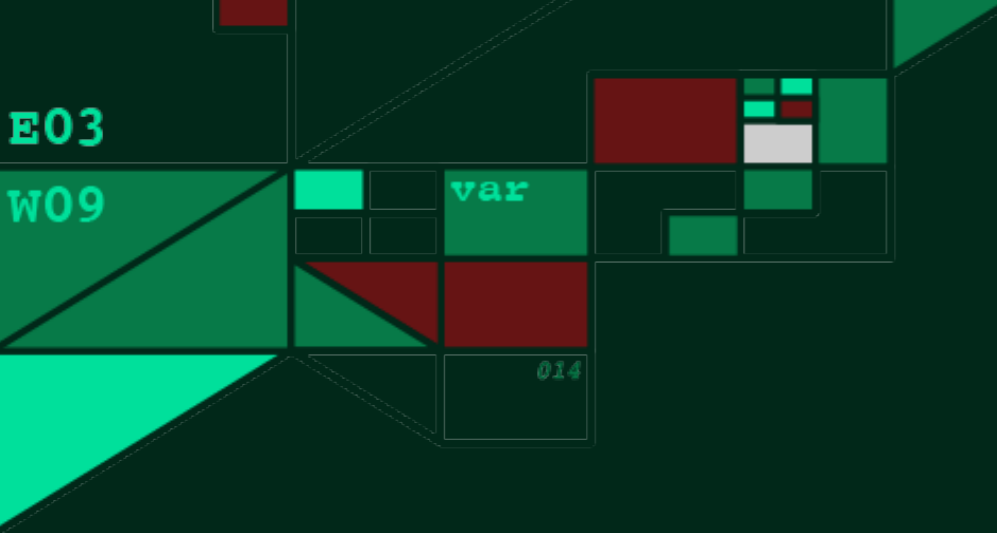 - https://github.com/datacats/datacats (docker)

Promotion

Keep in Touch

- ckan-dev mailing list (both general and technical)

- https://lists.okfn.org/mailman/listinfo/ckan-dev

- Contribute upstream!

- Discussion group in Taiwan?

Documentation L10n

- CKAN2 安裝使用教學

  https://ckan-docs-tw.readthedocs.io

- Depositar's documentation https://docs.depositar.io

- Helper wanted

General Discussion

d_ depositar

Thank you!

Please contact us at
data.contact@depositar.io

Or visit us at
https://data.depositar.io