ckanapi and ckanext-scheming

CKANCon 2015

Ian Ward

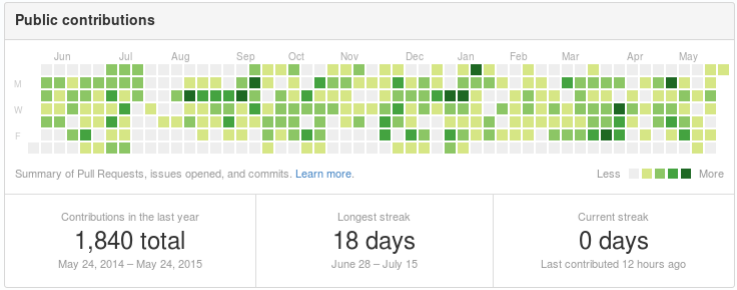
open.canada.ca

www.datacats.com

DataCats

Python + Open Source developer

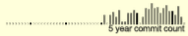
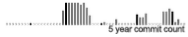
github.com/wardi



DataCats

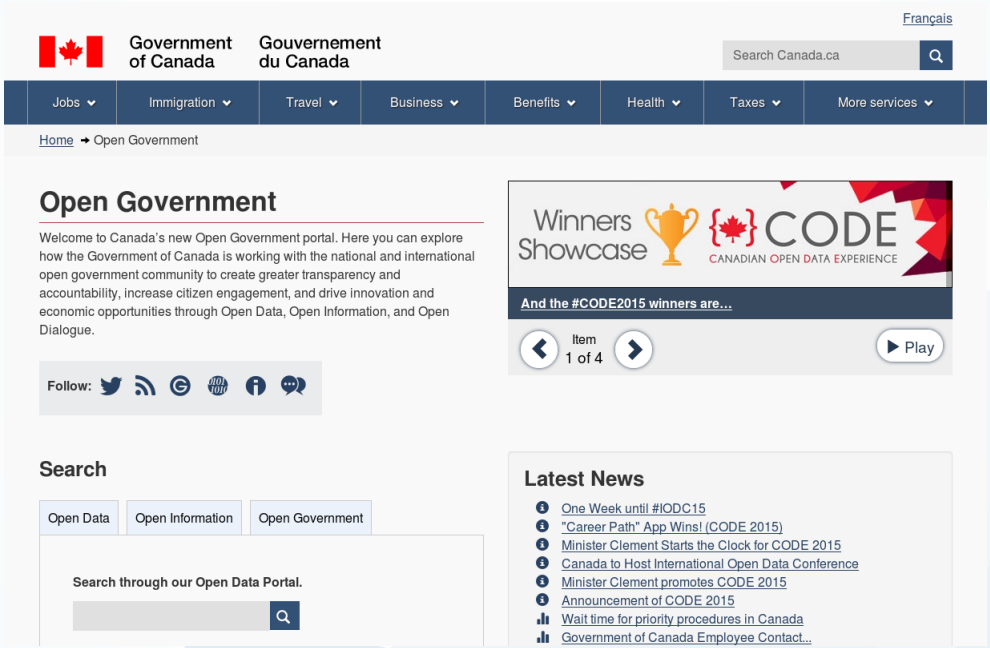
CKAN core

Top Contributors

Name	Kudos	12 Month Commits	All Time Commits	5 Year Trend	Primary Language	First Commit	Last Commit
amercader	7	321	915		Python	over 3 years ago	5 days ago
Ian Ward	7	188	284		Python	over 2 years ago	about 1 month ago
Vítor Baptista	7	154	460		Python	over 2 years ago	8 months ago
David Read	7	144	1026		Python	over 4 years ago	12 days ago
Brook Elgie	6	101	101		Python	8 months ago	5 days ago
Sean Hammond	8	93	1749		Python	over 3 years ago	3 months ago
joetsoi	7	53	163		Python	about 2 years ago	10 days ago
Ross Jones	7	48	574		Python	over 3 years ago	3 months ago
Nigel Babu	7	18	235		Python	over 2 years ago	9 months ago
nigelb	6	16	42		Python	about 1 year ago	4 months ago

DataCats

open.canada.ca



DataCats

ckanext-fluent

github.com/open-data/ckanext-fluent

EN Books: Franklin

FR Books: Benjamin

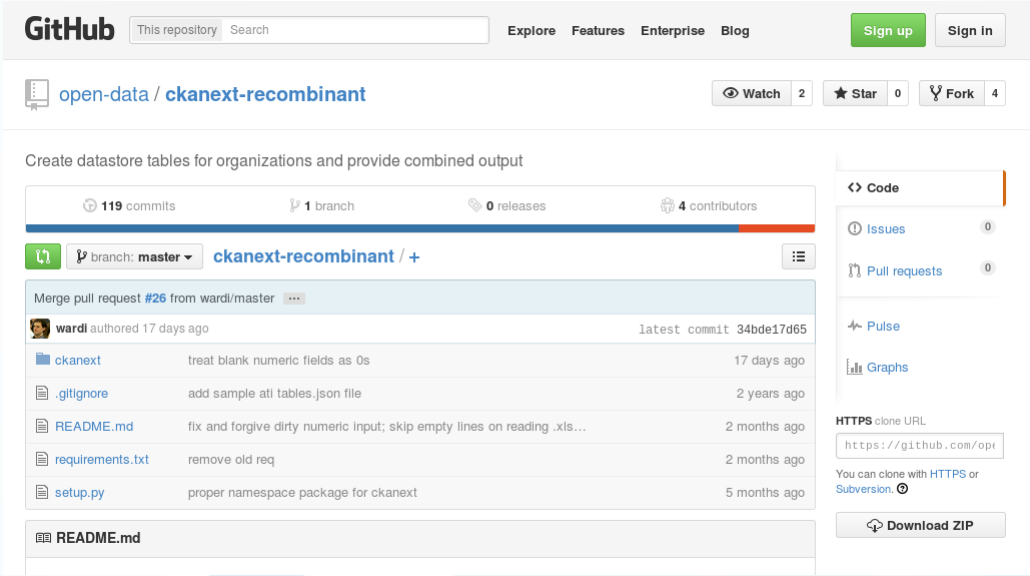
Field	Value
Books	en Franklin fr Benjamin

```
"books": {"en": "Franklin", "fr": "Benjamin"}
```

DataCats

ckanext-recombinant

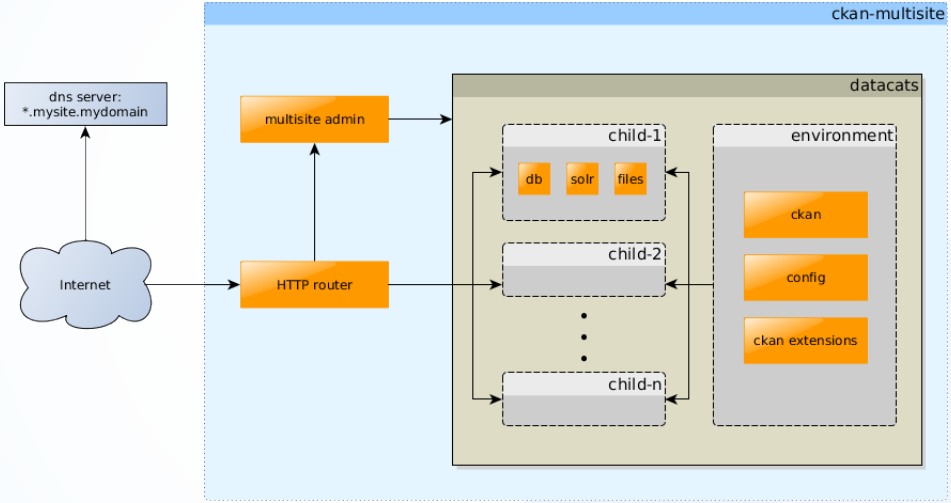
github.com/open-data/ckanext-recombinant



DataCats

CKAN Multisite

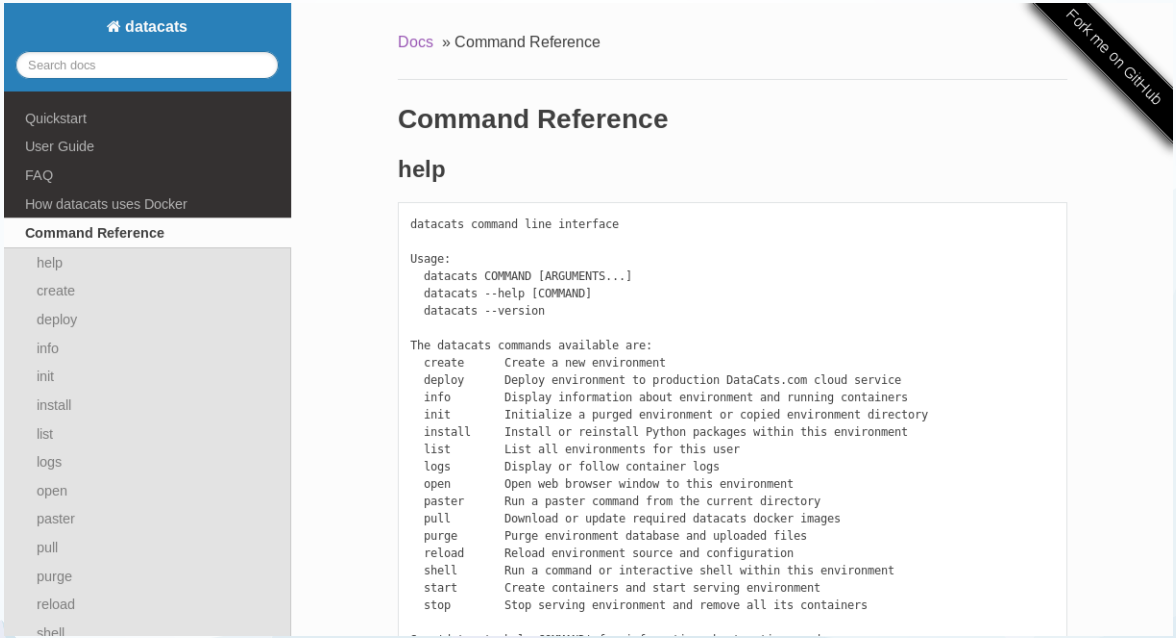
github.com/boxkite/ckan-multisite



DataCats

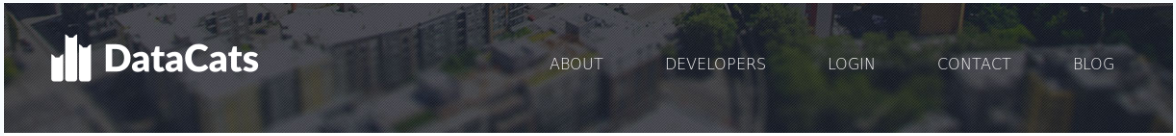
datacats

github.com/datacats/datacats



DataCats

www.datacats.com



ABOUT

DataCats provides Open Data software for cities. Our software helps governments increase transparency and accountability, better engage with citizens, and make meaningful decisions based on the data they collect.

BENEFITS

Create APIs around

Enable civic technology

DataCats

ckanapi

github.com/ckan/ckanapi

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

build passing tested under Python 2.6, 2.7, 3.2, 3.3 and pypy

ckanapi CLI

The ckanapi command line interface lets you access local and remote CKAN instances for bulk operations and simple API actions.

Actions

Simple actions with string parameters may be called directly. The response is pretty-printed to STDOUT. e.g.:

```
$ ckanapi action group_list -r http://demo.ckan.org
[
  "data-expolorer",
  "example-group",
  "geo-examples",
  ...
]
```

DataCats

ckanext-scheming

github.com/open-data/ckanext-scheming

This extension provides a way to configure and share CKAN schemas using a JSON schema description. Custom template snippets for editing and display are also supported.

Requirements

This plugin is compatible with CKAN 2.3 or later.

Configuration

Set the schemas you want to use with configuration options:

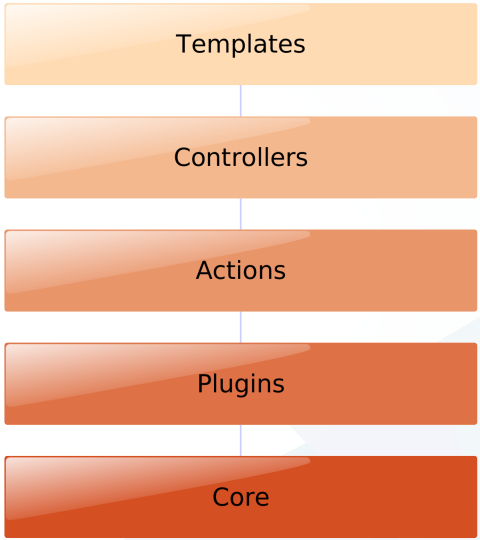
```
ckan.plugins = scheming_datasets

#   module-path:file to schemas being used
scheming.dataset_schemas = ckanext.spatialx:spatialx_schema.json
                           ckanext.spatialx:spatialxy_schema.json
#   will try to load "spatialx_schema.json" and "spatialxy_schema.json"
#   as dataset schemas
#
```

DataCats

I ♥ CKAN

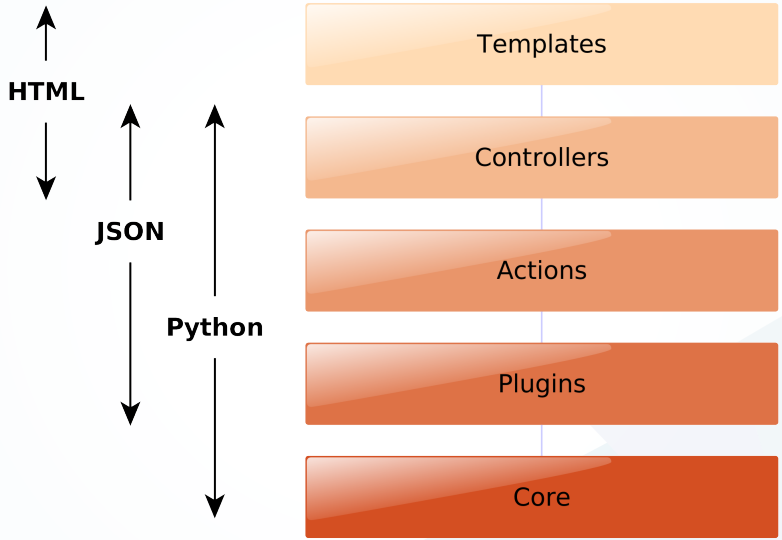
DataCats

CKAN layers



DataCats

Extending CKAN



DataCats

ckanapi

github.com/ckan/ckanapi

1. Python library for calling ckan actions
2. Command-line client for calling ckan actions

Python library

Pass strings, lists, dicts, file objects like normal method calls

```
import ckanapi

ckan = ckanapi.RemoteCKAN('http://exampleckan', apikey=...)

with open('mydata.csv') as csv:
    ckan.action.package_create(name='best-dataset', title='Best evar')
    ckan.action.resource_create(package_id='best-dataset', upload=csv)
```

DataCats

Safe

- Raises exceptions on errors, no manual error checking
- Clean context and data_dict for every call

Universal

- Python 2.6, 2.7, 3.2+ (source-compatible)
- Generic, no per-action custom code

Consistent interface

- LocalCKAN for core or extensions
- RemoteCKAN for remote API calls
- TestAppCKAN for tests

ckanapi Python library

```
pip install ckanapi
```

```
import ckanapi
```

github.com/ckan/ckanapi

CKAN Action API Reference

DataCats

Command-line client

- Single and bulk actions
- Easy to pipeline
- Local and remote

DataCats

Single actions

Pass strings as simple parameters

```
ckanapi action package_create name=best-dataset title="Best evar"
ckanapi action package_list
```

DataCats

Bulk actions

```
ckanapi dump datasets
ckanapi dump organizations
ckanapi dump groups
```

DataCats

Bulk actions

```
ckanapi load datasets
ckanapi load organizations
ckanapi load groups
```

DataCats

Bulk actions

```
ckanapi load datasets -p 3
ckanapi load organizations -p 3
ckanapi load groups -p 3
```

DataCats

Pipelining

```
ckanapi dump datasets | ssh otherbox ckanapi load datasets -p 3
```

Local and remote

```
ckanapi dump datasets -r http://sourceckan | ckanapi load datasets -p 3
```

DataCats

Bulk data format

- JSON Lines
- Consistent order with sorted keys

Track metadata in git

```
ckanapi dump datasets > datasets.jsonl
git diff datasets.jsonl --stat

datasets.jsonl | 52 ++++++++++++++++++++++++++++++++++++----------------
1 file changed, 36 insertions(+), 16 deletions(-)
```

DataCats

Distributed load

```
split -n l/3 datasets.jsonl part
ckanapi load datasets -r http://web1 -a ... < partaa &
ckanapi load datasets -r http://web2 -a ... < partab &
ckanapi load datasets -r http://web3 -a ... < partac &
```

DataCats

Summaries

```
head -5 datasets.jsonl | jq .title

"Exports of Dairy Products by Country of Destination Dairy Year 2001/02"
"Historical Border Wait Times"
"Number of Head Slaughtered in Federally Inspected Plants 2009 Hogs"
"Canadian Dairy Exports Month Calendar Year 2001 September"
"Federal Corporations"
```

DataCats

Reports

```
jq 'select(.organization.name!="nrcan-rncan")' -c datasets.jsonl | wc -l

8613
```

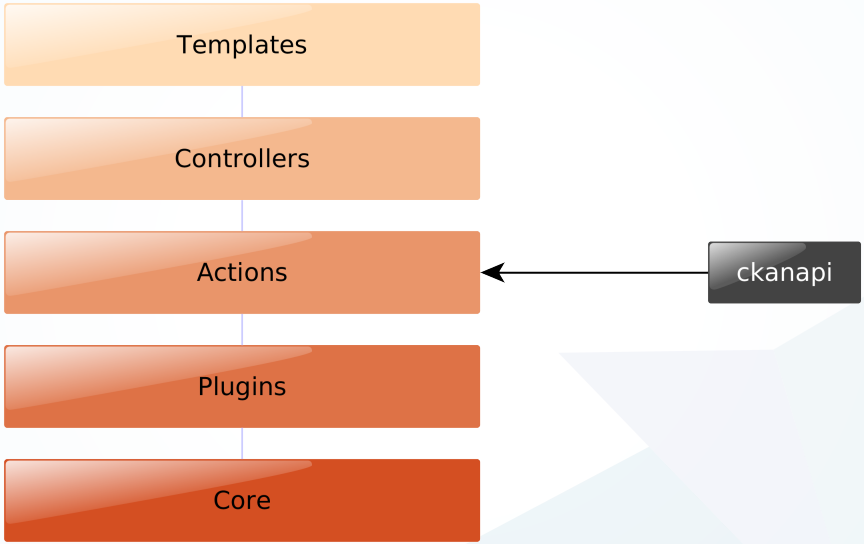
DataCats

Command-line client

```
pip install ckanapi
ckanapi -h
```

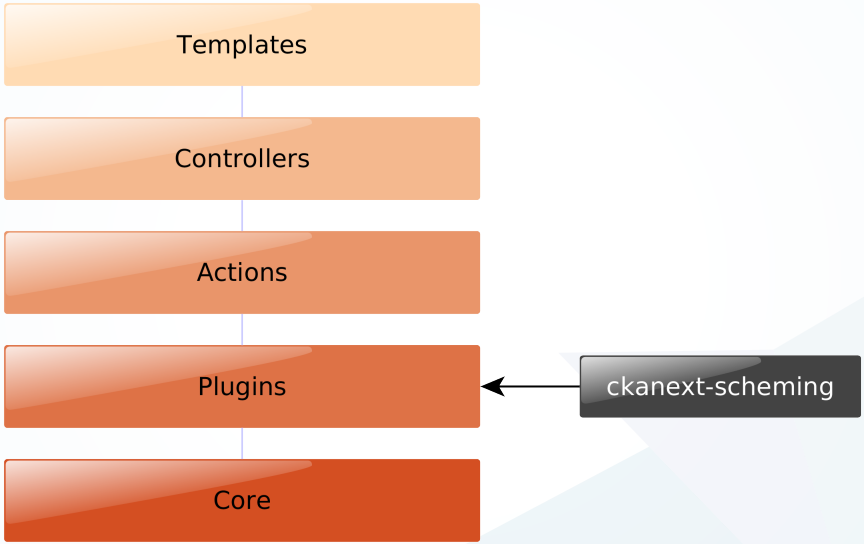
DataCats

CKAN layers



DataCats

CKAN layers



DataCats

IDatasetForm is awesome

- Define dataset types
- Custom templates for edit + show
- New fields + validation rules for datasets + resources

DataCats

IDatasetForm is hard

- Python plugin code
- Nested Jinja2 templates
- navl schema + validators

ckanext-scheming is easier

- Code optional
- Templates included
- Combined JSON schema
- Custom validators with IValidators

DataCats

ckanext-scheming is sharable

- scheming_dataset_schema_list
- scheming_dataset_schema_show

DataCats

Example schema

```
{
  "dataset_type": "camel-photos",
  "dataset_fields": [
    {
      "field_name": "title",
      "label": "Title",
      "form_placeholder": "eg. Larry, Peter, Susan",
      "form_snippet": "large_text.html",
      "form_attrs": { "data-module": "slug-preview-target" },
      "validators": "if_empty_same_as(name) unicode"
    },
    {
      "field_name": "name",
      "label": "URL",
      "form_placeholder": "eg. camel-no-5",
      "form_snippet": "slug.html",
      "validators": "not_empty unicode name_validator package_name_validator"
    },
```

DataCats

Example schema

```
{
  "dataset_type": "camel-photos",
  "dataset_fields": [
    {
      "field_name": "title",
      "label": "Title",
      "form_placeholder": "eg. Larry, Peter, Susan",
      "preset": "title"


    },
    {
      "field_name": "name",
      "label": "URL",
      "form_placeholder": "eg. camel-no-5",
      "preset": "dataset_slug"

    },
```

DataCats

Example schema

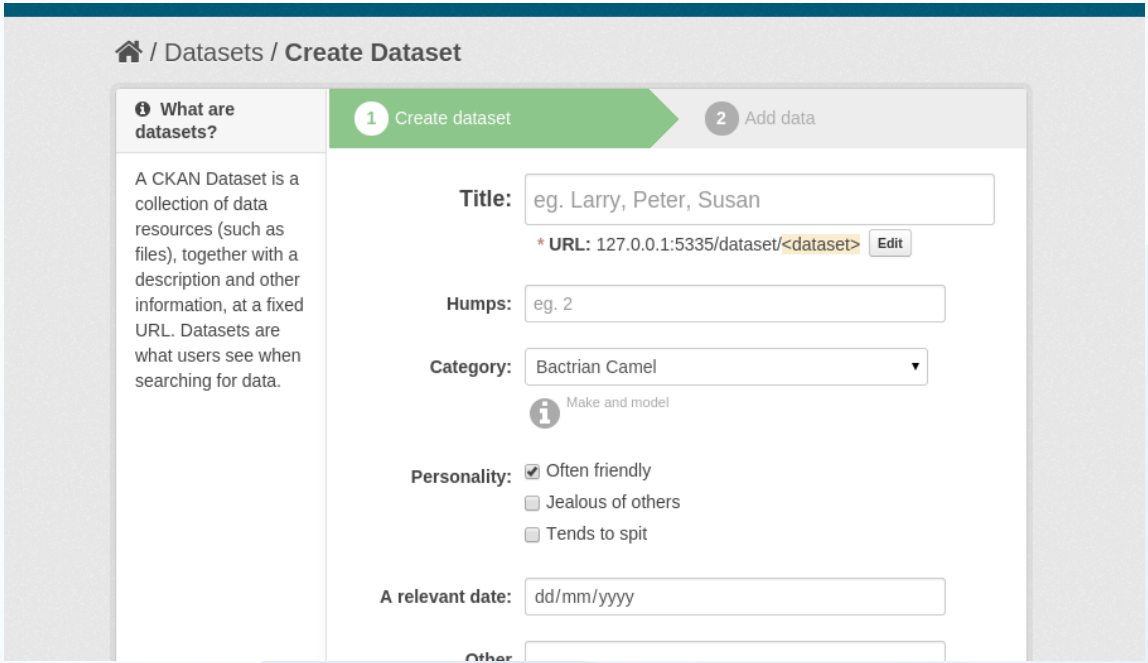
```
{
  "field_name": "category",
  "label": "Category",
  "help_text": "Make and model",
  "help_inline": true,
  "preset": "select",
  "choices": [
    { "value": "bactrian", "label": "Bactrian Camel" },
    { "value": "hybrid", "label": "Hybrid Camel" },
    { "value": "f2hybrid", "label": "F2 Hybrid Camel" },
    { "value": "snowwhite", "label": "Snow-white Dromedary" },
    { "value": "black", "label": "Black Camel" }
  ]
},
```

DataCats

Example schema

```
{
  "field_name": "personality",
  "label": "Personality",
  "preset": "multiple_checkbox",
  "choices": [
    { "value": "friendly", "label": "Often friendly" },
    { "value": "jealous", "label": "Jealous of others" },
    { "value": "spits", "label": "Tends to spit" }
  ]
},
{
  "field_name": "a_relevant_date",
  "label": "A relevant date",
  "preset": "date"
},
```

Example form



DataCats

Future

- Groups and organizations
- Search facets
- Schema registry?

ckanapi and ckanext-scheming

- github.com/ckan/ckanapi
- github.com/open-data/ckanext-scheming

Questions?

ian@datacats.com

DataCats