

Approaching open science at SENTINEL

8th October 2020



Context & guiding principles

SENTINEL

- ▶ brings together a very diverse set of modelling communities
- ▶ researchers are used to very different workflows

Goals:

- ▶ facilitate reuse of research output,
- ▶ build on each others work,
- ▶ be inclusive: accomodate *diverse* workflows

Our approach

the format 1/2

- ▶ rich metadata to capture research context, license, etc
- ▶ tabular data, with certain minimum schema requirements, e.g.

<u>time</u>	<u>primary_col</u>
...	...
...	...

Our approach

the format 1/2

- ▶ rich metadata to capture research context, license, etc
- ▶ tabular data, with certain minimum schema requirements, e.g.

time	primary_col	cola	colb
...
...

Our approach

the format 2/2

- ▶ Frictionless datapackages
- ▶ an open specification, mostly for tabular datasets.
- ▶ datapackage = metadata + datasets

The screenshot shows the 'DATA PACKAGE CREATOR' interface. On the left, there are buttons for 'Upload', 'Validate', and 'Download'. Below these are sections for 'Metadata' (Name, Title, Profile, Description, Home Page, Version, Author) and 'My Data Package' (Tabular Data Package). The main area is titled 'Resources' and contains a table with columns 'Name', 'View', 'Path', and 'Level'. There are two resource cards visible: 'facts' and 'colors'. Below them are 'books' and 'energy_eff' cards. On the right, a 'Preview' pane shows the JSON metadata for a resource, including fields like 'name', 'path', 'profile', and 'table'.

<https://create.frictionlessdata.io/>

Adoption strategy

Encourage user buy-in:

- ▶ by providing tools to work with datapackages, easing transition,
- ▶ with features like data validation (wip) & visualisation (planned) provide additional utility