

Meeting minutes: [☰ Croissant Task Force Minutes](#)

Slava's slides: [🟡 Croissant ML standard in Dataverse and beyond](#)

- What is Dataverse?
 - 116 installations in 36 countries
 - Get a DOI for your data, many other [features](#), [open source](#)
- Demo
 - Cars dataset landing page:
<https://dev3.dataverse.org/dataset.xhtml?persistentId=doi:10.5072/FK2/DZRHUP>
 - Cars dataset exported as Croissant:
<https://dev3.dataverse.org/api/datasets/export?exporter=croissant&persistentId=doi%3A10.5072/FK2/DZRHUP>
- Croissant is the new Schema.org JSON-LD
 - Different enough from the old format for Dataset Search that we'll treat Croissant as a new format (and avoid duplicated fields this time)
- Questions
 - [citeAs](#): for a paper, sure, but where do I put the DOI of a dataset? In `identifier` like we did in the old format?
 - "@id": "https://doi.org/10.7910/DVN/TJCLKP",
 - "identifier": "<https://doi.org/10.7910/DVN/TJCLKP>",
 - citeAs: bibtex, ok to put dataset citation here
 - Where does the file hierarchy go? e.g. `code/compute.py`
 - `@id` of a file? filename or database id? What about the file path?
 - Welcome to open an issue about what to do with file path
 - @id has to be unique
 - Can/should [summary statistics](#) go into Croissant?
 - Dropping author (creator is enough)
 - Dropping provider (publisher is enough)
 - What about [multiple formats](#) for a file? Original (e.g. Stata) vs. archival (tab delimited). Do you want the original? For encodingFormat etc.?
 - What if the Croissant file is huge, due to many files (distribution) and columns (recordSet). Still put it in the <head> of the dataset landing page?
 - When adding Dataverse installations to Croissant Online Health, first query each of the 116 installations for Croissant support?
<https://github.com/mlcommons/croissant/issues/530>
 - Where should we get answers to additional questions? This meeting? The mailing list? Discord? GitHub issues?
 - GitHub issues are best
- Beefs
 - md5 is supported (good!) but not mentioned in spec
 - 1.0 as a dataset version should be valid
 - <https://github.com/mlcommons/croissant/issues/609>

