

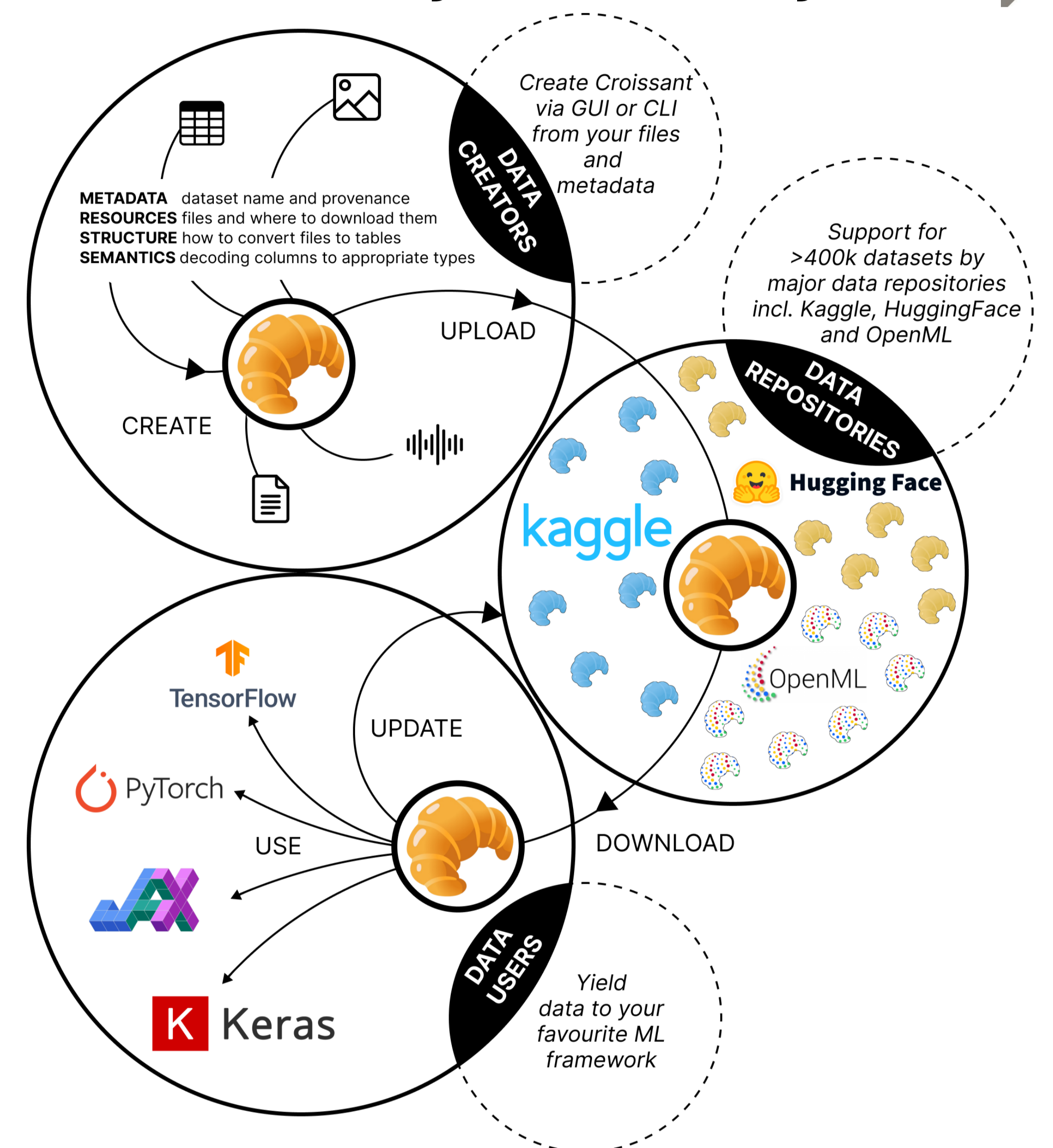
## Towards a Standardised Data Documentation Format for Cultural Heritage Collections

Nitisha Jain, Albert Meroño Peñuela, Elena Simperl

### Croissant

- **Data** is a critical resource for different domains, yet working with data remains a **key friction point**.
- **Croissant** is a **metadata format** for datasets that simplifies how data is used by ML tools and frameworks.
- **Croissant** makes datasets more **discoverable**, **portable** and **interoperable**, enabling **dataset management** and **responsible AI (RAI)**.
- Support from **Hugging Face**, **Google Dataset Search**, **Kaggle**, **Open ML**, **TFDS**, **Meta**, **NASA**.
- **Extensions** to digital heritage, health, geospatial domains.

### Croissant lifecycle and ecosystem



### MuseIT Objectives

- To build a **Multimodal Knowledge Graph** to capture multisensory representations of cultural heritage (CH) assets.
- To enable **digital rights management** for digital heritage collections, build **data trust** and **provenance** solutions.

### Croissant for MuseIT

- Attributes for documenting **user demographics**, **provenance** and **data lineage**, data use and **limitations**.
- **Open-source** python library, user-friendly **visual editor**.

### Multimodal cultural artefacts

