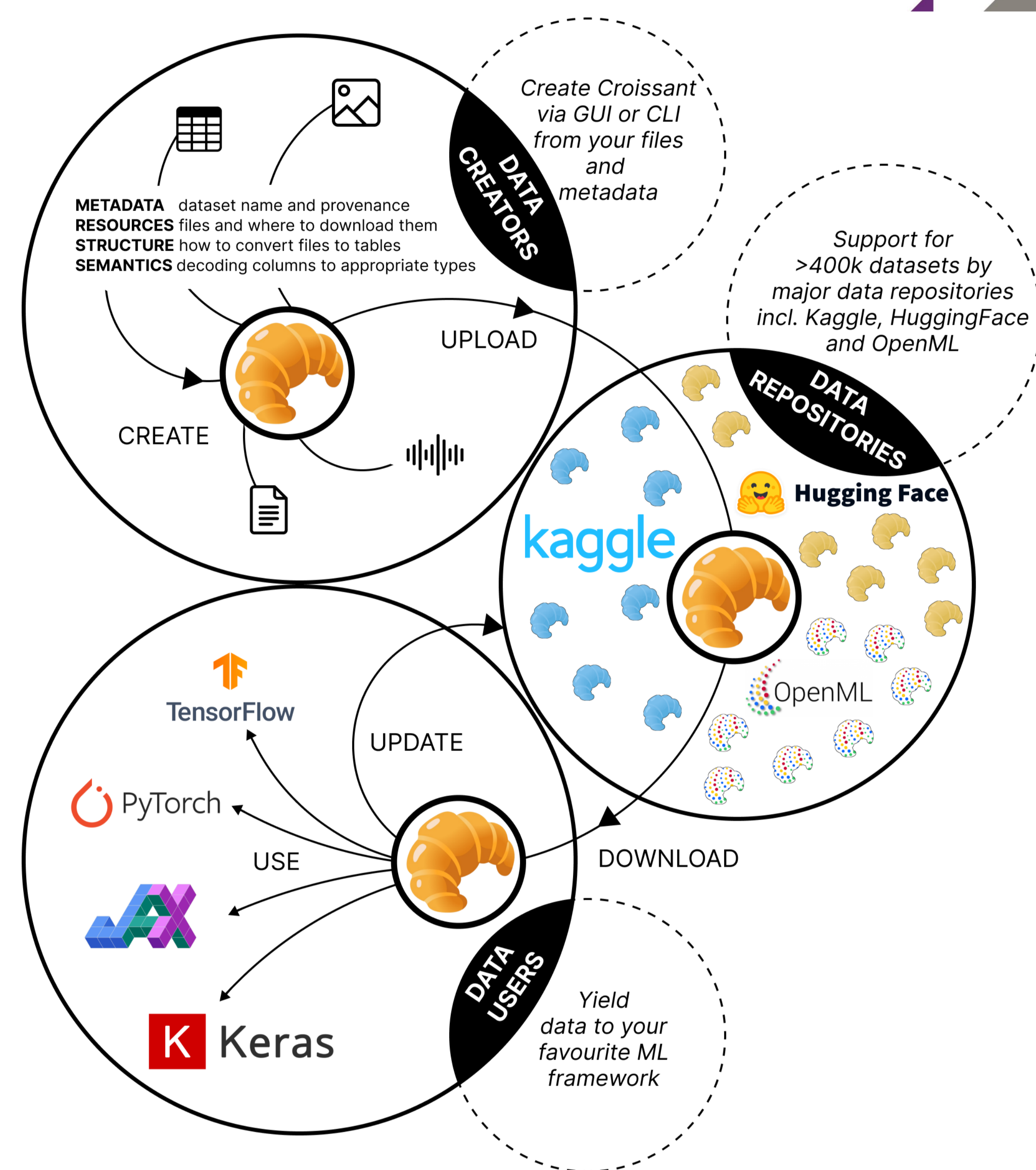


## Croissant: Fostering Responsible AI with a Standardized Data Documentation Format

Nitisha Jain, Mubashara Akhtar, Elena Simperl

### Background

- Data is a critical resource for Machine Learning (ML) 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**.
- Support from Hugging Face, Google Dataset Search, Kaggle, Open ML, TFDS..



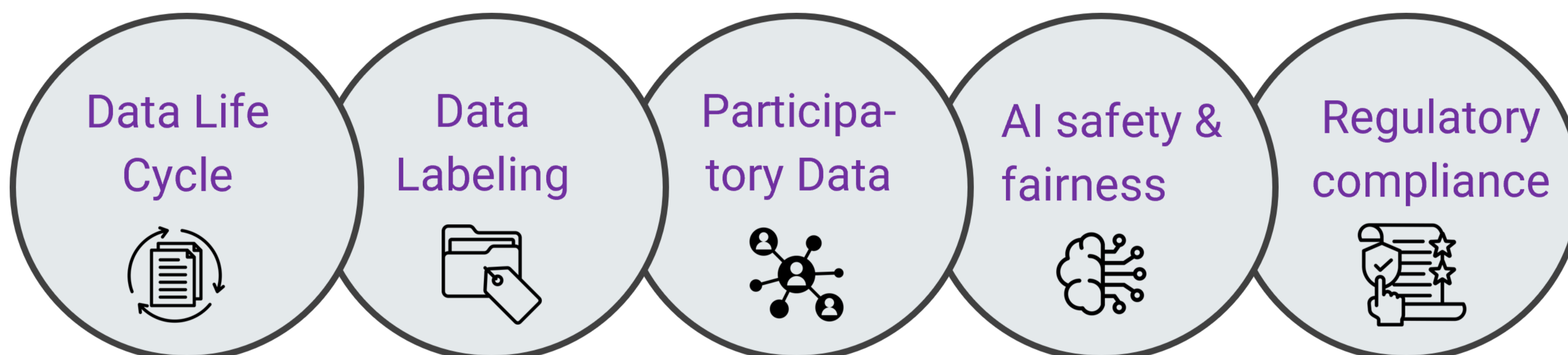
Croissant lifecycle and ecosystem

### Croissant is for

- Creators and maintainers of ML datasets
- ML researchers and practitioners
- RAI researchers and practitioners
- Policy makers

### Croissant offers

- Open-source Python library
- Open-source visual editor
- Extensions for RAI, geospatial datasets, life sciences, digital humanities



Use Cases for Croissant-RAI Vocabulary

