Responsible AI vocabulary for documentation of Cultural Heritage datasets

As AI advances at rapid speed there is increased recognition among researchers, practitioners and policy makers that we need to explore, understand, manage, and assess its economic, social, and environmental impacts [1]. Dataset documentation is one of the main instruments to operationalise responsible AI (RAI) that makes existing documentation solutions easier to publish, share, discover, and reuse.

Croissant [2,3] is a metadata format that introduces a standardised vocabulary for dataset attributes, the resources they contain, and their structure in order to streamline their usage and sharing within the community. Croissant aims to address significant dataset management challenges and foster responsible AI practices. The RAI vocabulary [4] being developed as an extension to the Croissant aims to address various important RAI use cases such as data life cycle, data labelling, AI safety and fairness evaluation, traceability, regulatory compliance, and inclusion.

In the cultural heritage domain, we are building a multimodal knowledge graph of artefacts for enabling wider accessibility to people with different sensory limitations as part of the MuseIT project [5]. In this effort, dataset documentation aligned with the above RAI aspects are vital. Comprehensive documentation, including origins, contributors, biases, and limitations, fosters trust among stakeholders and ensures accurate representation of diverse perspectives. Transparent documentation also enhances credibility and facilitates understanding of data lineage, context, and reliability. Prioritising inclusion and diversity ensures equitable representation of cultural heritage, while clear provenance information safeguards against misinformation. Embracing these practices strengthens dataset integrity, accessibility, and impact, fostering deeper understanding and accessibility of cultural heritage.

1.https://link.springer.com/book/10.1007/978-3-030-30371-6
2. https://mlcommons.org/2024/03/croissant_metadata_announce/
3. https://blog.research.google/2024/03/croissant-metadata-format-for-ml-ready.html
4. https://mlcommons.github.io/croissant/docs/croissant-rai-spec.html
5. https://www.muse-it.eu/