Contribution ID: **47** Contribution code: **2-34** Type: **Poster**

FAIR DO Application Case for Composing Machine Learning Training Data

The application case for implementing and using the FAIR Digital Object (FAIR DO) concept aims to simplify usage of label information for composing Machine Learning (ML) training data.
Image data sets curated by different domain experts usually have non-identical label terms. This prevents images with similar labels from being easily assigned to the same category. Therefore, using the images collectively for application as training data in ML comes with the cost of laborious relabeling. To automate this process, machine-actionable decisions for label information must be enabled. For this purpose the FAIR DO concept is used. A FAIR DO is a representation of scientific data and requires at least a globally unique Persistent Identifier (PID), relevant metadata, and a type.
We show the requirements for specifying and using FAIR DOs when applied to ML data. Based on an application case with Scanning Electron Microscopy (SEM) image data, a Proof-of-Principle approach shows the potential of the concept for usage in ML related data management.
This work has been supported by the research program 'Engineering Digital Futures'of the Helmholtz Association of German Research Centers and the Helmholtz Metadata Collaboration (HMC) Platform.

Please assign your poster to one of the following keywords.

Processes/Policies

Please assign yourself (presenting author) to one of the stakeholders.

Scientist/ Data Producer

Please specify "other" (stakeholder)

In addition please add keywords.

FAIR, Metadata, image data, label,

Primary author: BLUMENROEHR, Nicolas (Karlsruhe Institute of Technology, Steinbuch Centre for Computing)

Co-authors: PFEIL, Andreas (KIT); STOTZKA, Rainer; JEJKAL, Thomas

Presenter: BLUMENROEHR, Nicolas (Karlsruhe Institute of Technology, Steinbuch Centre for Computing)

Session Classification: Postersession II

Track Classification: Postersession