Datasets and other research artifacts are a major topic in the scientific community in the recent years. Many ongoing projects focus on improving the standardization, publication and citation of these artifacts. Currently, the dblp team is involved in three of them: NFDI4DataScience, NFDIxCS, and Unknown Data. As part of these projects, we are happy to announce that datasets and artifacts have now been added as true "first-class citizens" to dblp, just like any other research contribution.

To this end, we updated our internal tools to better support datasets as a type of publication. We started to index some dataset repositories, such as Zenodo and IEEE DataPort, with many more to come. A first batch about 3,700 data publications is already part of the dblp dataset. The metadata of these data publications is treated with the same care as classical printed publications, including the disambiguation of authors and the attribution with PIDs such as ORCIDs and DOIs whenever available.

`<data>` records and UI

From a technical viewpoint, we added a new record type to the dblp data model. The new `<data>` record represents any form of "data-like" research artifact, be it a CSV data file, a piece of software, a trained model, or even a whole virtual execution environment.

Since the `<data>` field has been part of the dblp.dtd for several years now and no new fields have been added at this point with the recent changes, the latest version of the dblp.xml file should still be valid with existing software as usual. However, we plan to release a new and expanded DTD in the near future that will allow for additional content in `<data>` records.

BibTeX for dataset publications

We also updated our website to better present these new dataset records. This includes an update to one of our most used features: The BibTeX export. Since there are no de facto standards for datasets in BibTeX (yet?) we started with a first, prototypical implementation.

```
[+]
[-]  BibTeX record data/10/CabreraOFM21

> Home > data/10/CabreraOFM21

download as .bib file

@misc{DBLP:data/10/CabreraOFM21,
  author       = {Oscar Cabrera and
                  Marc Oriol and
                  Xavier Franch and
                  Jordi Marco},
  title        = {A Context-Aware Monitoring Architecture for Supporting System Adaptation
                  and Reconfiguration - Software code and ontology},
  publisher    = {Zenodo},
  year         = {2021},
  month        = jan,
  howpublished = {\url{https://doi.org/10.5281/zenodo.4482144}},
  note         = {Accessed on YYYY-MM-DD.},
  url          = {https://doi.org/10.5281/zenodo.4482144},
  doi          = {10.5281/zenodo.4482144},
  timestamp    = {Thu, 30 Mar 2023 11:15:16 +0200},
  biburl       = {https://dblp.org/rec/data/10/CabreraOFM21.bib},
  bibsource    = {dblp computer science bibliography, https://dblp.org}
}
```

To collect your feedback on how you would expect dataset bibTeX to be handled in dblp, we have set up a brief survey about our current BibTeX export of data publications. We very much appreciate any feedback and experiences you are willing to share with us, so that we can improve and update that feature in the near future. Thank you very much for your kind support!

Re-labeling of existing records

As part of the rework of datasets as a type of publications, we also reorganized the type labelling of some of our existing records. Previously, a smaller number of dedicated

journals, such as the Journal of Open Source Software (JOSS), Elsevier's SoftwareX, MDPI's Data, or the Dagstuhl Artifacts Series (DARTS), had been labelled as "Data and Artifacts". This was always a little bit of a misnomer, as while the content represented by those articles were closely related to datasets, they were still articles by nature, and not the datasets themselves.

Hence, we changed the type labeling for those items to "Journal Article" in order to more properly reflect the true nature of the items behind the records, and to better distinguish them from the datasets themselves. From now on, only true datasets will be labeled as "Data and Artifacts".

Work in progress

All these new features are still under development and you can expect further updates and changes in the future. The amount of indexed dataset publication repositories will also increase. If you have any feedback, we are always very happy to listen! You can reach us via dblp@dagstuhl.de.