

Case description from the project “FAIRify Humanities” of the National Data management Forum 2019.

2020-02-06

Title: Jens Bille's Ballad Book

Name and affiliation: Dorthe Duncker and Hanne Ruus, UCPH

Research area: Denmark's mediaeval ballads were committed to paper in the Renaissance. Hundreds of ballads are found in handwritten ballad books from 1550 to 1700. The full corpus consists of 9 ballad books handed down from Renaissance ballad collectors (1553-1591). See more at <https://duds.nordisk.ku.dk/english>

Data set(s): Jens Bille's Ballad Book belongs to a corpus of the oldest Danish ballad tradition. Jens Bille's Ballad Book (1555-1559) contains 7 epic ballads, 22 lyrical ballads, 31 ballads of chivalry, 11 historical ballads, 6 ballads of the supernatural, 3 jocular ballads, 4 heroic ballads, and 4 hymns. The texts have partly been re-keyed from scholarly editions, partly transcribed from photographs of the manuscript. All texts are in XML format with tokenization, stanza and verse segmentation, orthographical neutralization, lemmatization, POS-tagging, and homograph discrimination. Individual text words are numbered consecutively.

What efforts was carried out to make your data more FAIR?

An XML version of the texts with annotation was already created but the dataset was FAIRified by zipping the XML files and adding metadata in English. As the ballad book was produced in 1555-1559, no publishing rights make the FAIRification process difficult. The dataset was released in the CLARIN.DK repository, where metadata is both harvestable and searchable.

The dataset now has a PID: <http://hdl.handle.net/20.500.12115/41>. It can be searched in repository.clarin.dk and the harvestable metadata can be found in vlo.clarin.dk¹.

The use of XML formatted data enables interoperability and makes the data more reusable. This version of the ballad book was also given a version 1.1 and the citation will be “Ruus, Hanne and Duncker, Dorthe, 2020, *DUDS Jens Bille's Ballad Book*, v. 1.1, CLARIN-DK-UCPH Centre Repository, <http://hdl.handle.net/20.500.12115/41>.”

The dataset is released for academic use and licensed under [CLARIN-ACA-NC](https://creativecommons.org/licenses/by-nc/4.0/).

What was the biggest challenges to make your data more FAIR?

It was easy to FAIRify, as the data was already in an open format, and prepared for publication and sharing.

1

https://vlo.clarin.eu/record/http_58__47__47_hdl.handle.net_47_20.500.12115_47_41_64_format_61_cmdi?1&q=duds+ballad&index=0&count=1