Katy Granath

LIS 545

Mayernik

03.14.23

**Additional Dataset Information**

*Recommended Data Citation*

Note about citation: including the time stamp is recommended because AO3, like YouTube, Twitter, or TikTok, is a highly dynamic online repository with tens of thousands of daily updates and uploads.

Archive of Our Own “AO3” (2021-03-21 21:35:00 UTC). [Selective data dump for fan](https://archiveofourown.org/admin_posts/18804)

[statisticians](https://archiveofourown.org/admin_posts/18804). Originally published at

<https://archiveofourown.org/admin_posts/18804>.

*Considerations for Long-Term Preservation.*

For long-term preservation, these files should be stored in a repository that does not have file size limitations, or the files could be divided into smaller ones. Updates or new versions of the dataset should be made and stored together.

That being said, while columns and rows of information in CSV format are relatively straightforward to preserve, their substance is not. The long-term preservation of the *meaning* of this data relies on the preservation of metadata and context about the data. Like many other types of literature, fanfiction is a notoriously contextual type of media. Common types of meaning in literary works which require context to translate include metaphors, allusion, and vernacular specific to a time period. To be sure, all of these and more are present in fanfiction, but additionally, crucially, fanfiction arises directly from an original work, and inhabits the milieu of internet fandom and fanfiction community. Preserving this context is necessary to preserve the meaning of this dataset. Not doing so would be like preserving a copy of iconic Amanda Bynes film *She’s The Man* without also preserving the context that it is a retelling of Shakespeare’s *Twelfth Night*.

One way of encouraging/ensuring the preservation of this context might be to request that the repository also solicit and accept interpretations of the dataset from people within the community. For example, you can find data analysis write-ups about the data published on Ao3 itself (Eg. <https://archiveofourown.org/works/36682951/chapters/91505053#workskin>). In the case of highly contextual information like this, preserving the interpretations and questions that the dataset generated within its own community would be exceedingly valuable.

*Copyright*

The copyright license appropriate for this dataset is a [CC0 1.0 Universal](https://creativecommons.org/publicdomain/zero/1.0/) license. This aligns with the values and practices of AO3, which support a free and open Internet. AO3 provided this data to the public to be copied, modified, distributed, and utilized even for commercial purposes without needing to request permission.

*Human Subject Considerations*

AO3 accounts do not require identity verification and are not intentionally linked to a user’s offline personal information. This dataset includes creator usernames, but no other metadata about the creator. As such, this data does not include protected personal identifiable data about creators.