

To Educate or to Enforce

The Case for Underhanded Methods to Improve Research Data

Joshua Ramon Enslin¹ 

¹Freies Deutsches Hochstift / Frankfurter Goethe Museum

*Correspondence: Joshua Ramon Enslin, jenslin@freies-deutsches-hochstift.de

Abstract

museum-digital, founded 2009, has over the years developed into one of the central hubs for managing and publishing museum collection data in Germany. Over a thousand museums — of all sizes and different disciplinary backgrounds, as well as with different national backgrounds — use the platform to publish their collection data, roughly half of which also use it to manage their collections.

To make this diverse set of data interoperable, various techniques are used on an organizational, interface, and entirely technical level. First, all participating museums work with one set of controlled vocabularies, which are redacted and enriched by a centralized editing team. Second, users are encouraged to select controlled entries from drop-down lists rather than entering new ones. Finally, a range of reconciliation and cleanup routines run in the background when users enter a new term or one is to be imported —from the automatic correction of terms based on known aliases on a general ("Frankfurt a.M." becoming "Frankfurt am Main") or museum level ("Frankfurt" becoming "Frankfurt am Main" for a museum in Hessen) to reconciliation based on external norm data references.

Similarly, centralizing development and hosting allows for a simplification of workflows, e.g. when the data is to be packaged for long term archival or presentation in overarching portals such as the German Digital Library.

This talk will detail these and similar methods to show how the generation of high-quality research data can be simplified, encouraged and enforced without explicitly educating users about the underlying concepts. Thus, colleagues —many of whom are volunteers —who may have no academic background at all can be easily embedded into the generation of research data while a sufficiently high level of quality is maintained. Even those with a graduate or post-graduate level of education profit from a more usable interface and shared resources. Finally, by providing best-practice solutions (e.g. the use of norm data) implicitly, users can become accustomed to them slowly, increasing the likelihood of their acceptance.

The success of this approach can be seen in the discrepancies between the stated use of norm data according to the Institute for Museum Research's recent survey on digitization [1] and the size of museum-digital's user base. On the other hand, such an approach is necessarily limited. Collection data is relatively uniform by itself and can,

with some abstractions, be modeled and structured in a uniform fashion. It thus lends itself to generalized solutions, that may not be possible in other domains of data. Published collection data is also commonly focused on presenting the outcome of research rather than the corresponding research process. The talk will close with a discussion of such limitations of less explicit approaches to increasing research data quality.

Resources

- museum-digital. <https://www.museum-digital.org> “museum-digital”
- musdb <https://de.about.museum-digital.org/software/musdb/> “museum-digital:musdb”
- Poster: Normdaten bei museum-digital. <https://doi.org/10.5281/zenodo.15304745>. “Poster: Normdaten bei museum-digital”
- Bereinigung bei Importen zu museum-digital. <https://doi.org/10.5281/zenodo.15305169>. “Presentation: Bereinigung bei Importen zu museum-digital.”
- Open Source Libraries of museum-digital. <https://gitea.armuli.eu/museum-digital>. “Open Source Libraries of museum-digital”

Author contributions

Competing interests

The authors declare the following competing interests: The author is the primary developer of museum-digital.

Funding

None.

Acknowledgements

References

References

- [1] I. für Museumsforschung, *Zahlen und Materialien aus dem Institut für Museumsforschung*, Bd. 77, de, P. Rahemipour and K. Grotz, Eds. Arthistoricum.net, 2023. DOI: [10.11588/IFMZM.2023.1](https://doi.org/10.11588/IFMZM.2023.1).