# FAIR for whom?
# Commentary on Hofmann et al. (2021)

STEFAN MÜNNICH[1]
*University of Basel, Switzerland*

ABSTRACT: In this commentary on Hofmann et al. (2021), the notion of ethnomusicology and some of its underlying biases are questioned and reflected in the light of applying FAIR data principles to musicological research data from outside a Western canon and its musical practices.

THE fact that the study of "all the musics of the world in their various social and cultural contexts" (Hofmann et al., 2021) still requires a separate label in 2021 should at least give pause for thought. About 130 years after the first steps of the discipline and its forerunners, there is still the necessity to speak of *ethno*musicology instead just of musicology when it comes to music outside of European or American traditional lines. A strategy of separation can be observed behind this labeling practice until today, that exposes a musicological self-understanding as "Kunstwissenschaft" (study of music as art; Janz, 2013), which all too often focuses on a solely music-historiographical canon dominated by white male composers (Kijas, 2020). Such an imbalance is also reflected in the musical research data available today, be it edition projects, analytical examples, sound recordings, or training data for MIR procedures. The equal integration of research data from outside a Western canon could lead to an increased inclusivity and diversity of musico-cultural practices, methods, and approaches.

To get a little closer to this goal, Hofmann et al. present a distributed framework along the so-called FAIR principles, which were established in 2016 by Wilkinson et al. and form an integral part of various Open Science initiatives and projects such as OpenAIRE[2], FAIRsFAIR[3] or the Research Data Alliance RDA[4] (Bangert et al., 2019). The idea of the authors is to identify individual, infrastructural, organizational, but also political and institutional needs. Based on their findings, they propose necessary developments and efforts to ensure and promote the findability, accessibility, interoperability, and reusability (FAIR) of research data of any musics. The very thoroughly designed and well-balanced framework has been developed in close exchange and contact with the research community and has the potential to provide an unprecedented amount of historical and "real-life" data to a wider audience, if permitted and desired. It addresses issues of distributed data repositories, standardized metadata descriptions, the interconnection of musicological data via Linked Data vocabularies as well as the (semi-)automatic support of MIR tools. Thus, for the first time, research data from the study of the various musics of the world could serve on a vast scale as ground truth for the development of corresponding digital approaches. To leverage these rich resources, as Laurent Pugin pointed out in 2015, "will result in a transformation of the way we access music[s of the world] and the extent to which we make use of it." (Pugin, 2015)

One question, however, should not be lost sight of in the digital domain and should be an explicit part of an open and critical discourse. It is that of: FAIR for whom? Who has the access, the possibilities, resources, and capacities to access and maintain the relevant content? And who actually controls this access, or the infrastructure in the end, even in the realm of Open Science and FAIR data? As it turns out, human biases, most of which are anchored in language, concepts, or behavior, can be projected through algorithms or models into the digital realm (Caliskan et al., 2017). Furthermore, initiatives like *Whose Knowledge?*[5] show that the majority of the knowledge available on the Web today is not only covering less than 7% of the world's nearly 7000 languages and dialects (Anderson, 2004), but that it is put online mainly by white men from Europe or North America, although nearly half of the Internet users are female and about 60-75% of

the Internet users come from countries of the Global South (UN-ITU, 2019; IWS, 2020). The issue of appropriate copyright licenses, as discussed by Hofmann et al. in their paper, adds another perspective to the same problem: Since it is a primarily Western concept that focuses on individual authors and does not allow authors to be groups or even non-humans (animals, natural spirits), there are no existing copyright licenses that could be applied to the type of data in question here. Personal rights of individuals or groups of individuals or groups recorded on audio or video recordings may have been violated. The question of how to deal with cultural testimonies that may have been produced illegally or at least under dubious circumstances needs to be discussed and clarified in the research community regarding their digital accessibility. Consideration should be given to whether it should be possible to implement a "right to be forgotten" (Whitman, 2004), i.e., the complete deletion of the corresponding data, within the proposed framework.

Care must be taken that analogue biases are not transferred to and reinforced through digital space, but that a truly equal, inclusive, and participatory research infrastructure can be established in which the musics, music practices, and cultural contexts of the world can meet each other on an unbiased and equal footing.

## ACKNOWLEDGMENTS

## NOTES

[1] Correspondence can be addressed to: Stefan Münnich, University of Basel, Department of Musicology, Petersgraben 27/29, CH-4051 Basel, E-mail: stefan.muennich@unibas.ch.

[2] https://www.openaire.eu/

[3] https://www.fairsfair.eu/

[4] https://www.rd-alliance.org/

[5] https://whoseknowledge.org/

## REFERENCES

Anderson, S. R. (2004). How Many Languages Are There in the World? FAQ available from the Linguistic Society of America, Washington, DC. Retrieved from https://www.linguisticsociety.org/content/how-many-languages-are-there-world

Bangert, D., Hermans, E., van Horik, R., de Jong, M., Koers, H., & Mokrane, M. (2019). Recommendations for Services in a FAIR data ecosystem. *Zenodo.* https://doi.org/10.5281/zenodo.3585742

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics Derived Automatically from Language Corpora Contain Human-Like Biases. *Science, 356*(6334), 183-186. https://doi.org/10.1126/science.aal4230

Hofmann, A., Miksa, T., Knees, P., Bakos, A., Sağlam, H., Ahmedaja, A., Yimwadsana, B., Chan, C. & Rauber, A. (2021). Enabling FAIR use of Ethnomusicology Data - through Distributed Repositories, Linked Data and Music Information Retrieval. *Empirical Musicology Review, 16*(1), 47-64. https://doi.org/10.18061/emr.v16i1.7632

Internet World Stats (IWS). (2020). Internet Usage Statistics. The Internet Big Picture. World Internet Users and 2020 Population Stats. Retrieved from https://www.internetworldstats.com/stats.htm

Janz, T. (2013). Musikwissenschaft als Kunstwissenschaft? In M. Calella & N. Urbanek (eds.), *Historische Musikwissenschaft* (pp. 56-81). Stuttgart: J.B.Metzler. https://doi.org/10.1007/978-3-476-05348-0_4

Kijas, A. (2020). What does the data tell us?: Representation, Canon, and Music Encoding. Keynote text delivered at the Music Encoding Conference, University of Maryland, May 24, 2018. Retrieved from https://medium.com/@kijas/https-medium-com-kijas-what-does-the-data-tell-us-926ba830702f

Pugin, L. (2015). The challenge of data in digital musicology. *Frontiers in Digital Humanities 2*(4), https://doi.org/10.3389/fdigh.2015.00004

United Nations International Telecommunication Union (UN-ITU). (2019). ICT Facts and Figures 2019. Retrieved from https://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2019.pdf

Whitman, J. Q. (2004). The Two Western Cultures of Privacy: Dignity Versus Liberty. *Yale Law Journal 113*(6), 1151-1221. https://doi.org/10.2307/4135723

Wilkinson, M., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Bonino da Silva Santos, L. O., Bourne, P., Bouwman, J., J. Brookes, A., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C., Finkers, R., & Mons, B. (2016). The fair guiding principles for scientific data management and stewardship. *Scientific Data 3*(160018). https://doi.org/10.1038/sdata.2016.18