Good Practice Advice for Managing Multilingual and non-English Language Content in Repositories

Iryna Kuchma and COAR Task Force on Supporting Multilingualism and non-English Content in Repositories

COAR Annual Conference - Anticipating the Repository of the Future, Gothenburg, Sweden, June 6, 2024
Multilingualism presents a particular challenge for the discovery of research outputs. Although researchers and other information seekers may only be able to read in one or two languages, they want to know about all the relevant research in their area, regardless of the language in which it is published. Yet, discovery systems such as Google Scholar and other scholarly indexes tend to provide access only to the content available in the language of the user. In addition, the language of a scholarly resource is often not labelled appropriately, meaning a large portion of non-English resources are excluded from search results. Furthermore, many scholarly communications infrastructures are sub-optimal in their support for a variety of languages since little attention was paid to this issue during their design process.

On October 30, the COAR Task Force on Supporting Multilingualism and non-English Content in Repositories published 15 recommended practices for repositories to support multilingual and non-English content in repositories. The recommendations identify good practices for repository managers and repository software developers, and focus on the topics of metadata, multilingual keywords, user interfaces, formats, and licences that will improve the visibility, discovery and reuse of repository content in a variety of languages. Read the report.

The report is also available in Spanish and Catalan

Summary of Recommendations

COAR Task Force on Supporting Multilingualism and non-English Content in Repositories

Dealing with multilingualism and non-English content in open repositories: Challenges and perspectives

Dony, Christophe; Kuchma, Iryna; Ševkušić, Milica

Abstract

Several organizations and initiatives have recently called for more support of multilingualism in research to promote epistemic plurality and raise awareness of the adverse effects of an anglocentric research ecosystem. But this support for and practice of multilingualism and linguistic diversity cannot happen in a digital or technological vacuum. Open repositories can play an important role in ensuring that research infrastructures have the ability to implement and promote multilingualism at scale in an Open Science environment. This implementation, however, is complex and does not come without its own theoretical and technical challenges. One of these challenges is to recognize that the implementation of multilingualism in open repositories can hardly be dissociated from wider concerns of discoverability, research assessment practices, and the anglocentric nature of digital infrastructures and metadata standards or protocols. Drawing on the COAR (Coalition of Open Access Repositories) recommendations report produced by the COAR Task Force on Supporting Multilingualism and non-English Content in Repositories, this article presents and critically examines how and why three particular recommendations of this document are particularly well suited to support a decolonial trajectory for the management of multilingualism in open repositories. More specifically, this article discusses the decolonial aspects and praxis underlying guidelines such as declaring the language(s) of the resource and of its metadata, writing personal name(s) using the writing system used in the deposited document while providing a persistent identifier to disambiguate author/s identification and, overlapping with the latter, enabling UTF-8 support so as to promote use of the original alphabet / the writing system whenever possible, without negating the possibility to transiterate metadata by means of recognized standards (e.g. ISO). In so doing, we argue that these recommendations enable a multifaceted technology and politics of recovery that promotes a form of linguistic revitalization and strengthens linguistic diversity.

Hi @VictorVenema and @irynakuchma, I wanted to follow up on this thread since we’ve recently drafted changes related to this topic!

The DataCite Metadata Working Group has shared a new Request for Comments with several proposed schema changes. This proposal includes making the Language property repeatable and recommended, as well as documentation updates to clarify supported language codes.

Details on how to provide feedback are included in the draft proposal document and summarized in this blog post: https://doi.org/10.5438/cvnf-vs86

We welcome your input on the Request for Comments through May 6, 2024. Should you have any questions, please let me know.
Next steps?

We did not discuss the role of translations in helping to evaluate content – arXiv’s question

Anything else we should be focusing on?
Contributing Task Force Members

Iryna Kuchma, EIFL (chair), Ukraine
Jagadish Aryal, Social Science Baha, Nepal
Andreas Czerniak, Bielefeld University – Library, Germany
Christophe Dony, ULiège Library, Belgium
Joe Cera, Berkeley Law Library, University of California, USA
Sebastiano Giorgi-Scalari, Open University of Catalonia, Spain
Gussun Gunes, Marmara University Information and Records Management Department, Türkiye
Gultekin Gurdal, Izmir Institute of Technology İYTE, Türkiye
Johanna Havemann, AfricArXiv, Germany
Libio Huaroto Pajuelo, Universidad Peruana de Ciencias Aplicadas, Peru
Alan Ku (Gu Liping), National Science Library, Chinese Academy of Sciences, China
Pierre Lasou, Bibliothèque de l’Université Laval, Canada
Norma Aída Manzanera Silva, Centro de Investigaciones sobre América del Norte, Universidad Nacional Autónoma de México, Mexico
Lautaro Matas, LA Referencia, Spain/Latin America
Ayako Mikami, Hokkaido University, Japan
Tomoki Nagase, National Institute of Informatics, Japan
Tomasz Neugebauer, Concordia University, Canada
Jean-Francois Nomine, INIST, France
Milica Sevkusic, ITS SASA/EIFL, Serbia
Kathleen Shearer, COAR, Canada/International
Freddy Sumba, CEDIA, Ecuador