Is there a case for accepting purely machine translated material to be deposited in repositories?

This document summarizes discussions about Machine Translation undertaken by several members of the COAR Task Force on Supporting Multilingualism and non-English Content in Repositories and complements the blog post that was published on May 8, 2023

Pros	Cons
Purely MTed material can be useful for "gisting", i.e. to give an idea of what is being conveyed in an unfamiliar language. On the other hand, having translated abstract and multilingual keywords would already go a long way for such purposes. Users can then upload the document in an MT Tool would they wish to go further. Can the translated abstract be done by an MT tool?	MT is not universally reliable. There are large differences in the output quality capacity between language directions.
"Good enough" MT can help readers decide whether there is an interest in further professional human translation or post-editing of the content. But do readers need the MTed content to be uploaded into IR to do that?	When MT produces output in a very fluent way, there is no absolute guarantee that it is accurate. This means that it can sound very convincing but not always accurately reflect the original content, without the reader always being aware of this. On the other hand, there is no absolute guarantee that human translation is accurate either, but at least humans have the capacity to express doubt and feel self-doubt, to perceive uncertainty about their own translation, ask for help and engage in communication.
Archived MTed material may stimulate the interest of specialised readers in improving this technology.	There is a problem of accountability and liability for authors and content producers of original material, the translated version of which they do not control. And are repositories the best places/tools to build that corpus of MTed content?
Archived MTed material can encourage readers to adapt the vocabulary and phrases of their specialist field into the target languages, rather than borrowing duplicate words and expressions from English into their everyday language. This can stimulate creative terminology in the target languages. On the other hand, multilingual keywords can already go a long way to ensure discoverability and use of "proper"/adapted terminology.	MT of scientific articles cannot be taken at face value, as perceptions of the technology used to deliver scientific content may vary between different scientific communities. Some may feel that MT is not up to the challenge of communicating complex content in contrast to simpler content for the general public. Also, any MTed content is by definition short lived as the modules behind MT learn continuously (and therefore change output). On the other hand, the temporality of the translation output is also a reason to store it at one point in time.

Archived pure MTed material can encourage a dynamic to correct such translations, which can then be used as new material to train MT algorithms.	ChatGPT's reputation for being able to "trick" scientists into producing undetectable fake abstracts may also extend to MT producing undetectable, convincing but inaccurate translated versions.
Archived pure MTed material may be a starting point for new paths of research in translation studies, both as research and teaching corpora.	Current research in MT is seeking technological improvements and deep learning material to improve this technology in the specific application of scientific content translation. This shows that more work is needed to achieve a satisfactory level of reliability in this context.
On licensing and liability: open licensing enables MT and we have to learn to label it properly, clearly identifying what it was generated by AI and what level of review the content receives. Clear labeling that the content was generated by a MT tool will diminish legal issues.	Are MT algorithms legal entities that can be held liable for the content they produce? Can a person who uploads MTed content be accountable in case legal issues concerning the translation arise? And legal issues can arise if people start uploading MTed-only content for which they don't have the rights. For those reasons MTed content shouldn't be part of scholarly records as such, unless it's a collaborative research community effort.
The use of MT is quick, efficient, and potentially of a high impact in increasing access across languages: MT "technologies can play a key role to facilitate inter-language communication: MT in fact can deliver output cheaply, timely and quickly, and can tackle large volumes of data to be translated."[1]	Human translation is expensive, slow and difficult to scale to large quantities of content. "However, the quality of the MTed output remains lower than the quality of the output delivered by professional language professionals, especially for translations from/to low-resource languages and text loaded with cultural significance."[2]

[1] Borgonovi, F., J. Hervé and H. Seitz (2023), "Not lost in translation: The implications of machine translation technologies for language professionals and for broader society", OECD Social, Employment and Migration Working Papers, No. 291, OECD Publishing, Paris, https://doi.org/10.1787/e1d1d170-en

[2] Ibid