Open Science mandates exist across Europe indicating that science must be made accessible beyond the borders of academia affecting publishers and scientists alike. For example, in the UK, the UKRI requires that all funded science outputs are made accessible either via OA publishing or in institutional repositories. Similarly, in Spain, the Ministerio de Ciencia, Innovación y Universidades promotes open science for research publications and adherence to FAIR principles for research data.

However, simply sharing science as publicly available does not make it accessible. Scientists are constrained to write in the languages accepted by major publication venues and the languages of reviewers. Science communication written in a mother tongue is accessible only to a home audience, limiting international impact. A recent analysis of a large repository of full text scientific articles revealed over 30% to be from non-English sources, with European languages such as Spanish, Portuguese, German and French featuring in the top 10 publication languages. Automated machine translation is a proposed solution to the democratisation of cross-lingual science communication. Faithful machine generated translations have the potential to aid readers to interact with science outside of their known languages as well as to allow scientists to access publication venues beyond their usual range. Science communication requires accessibility to build impact. If the output of translation is complex technical language we only overcome one accessibility barrier, but many more may remain.

Translation is not enough. We need simple, clear and accessible cross-lingual science communication. The goal of the TINE project is to enable both producers and lay readers of science communication to access clear scientific information in their own language, promoting multilingualism. To enable this, TINE will leverage machine translation tools along with publicly accessible corpora to create new science-specific machine translation foundation models. These models will work on language pairs and will be adapted for specific scientific domains. Text will be adapted to plain language across multiple languages, focussing on interventions that apply cross-lingually. Terminological alignment between multilingual scientific vocabularies will be categorised for complexity to allow appropriate simplification of terminologies. Finally, we will engage citizen panels throughout the lifecycle of the project to determine guidelines, perform evaluation and disseminate findings.