Research Reproducibility with Confidential Data: Certifying the Uncertifiable

Authors: Kamel Gadouche (Director, CASD), Christophe Hurlin (University of Orleans), Christophe Pérignon (HEC Paris), and Roxane Silberman (scientific advisor, CASD)

A growing fraction of research is nowadays conducted using confidential data, such as highly-granular government data which require specific accreditation and controlled secure access. Analyzing such rich datasets allows researchers to conduct extremely innovative research programs and to address research questions that they could not address by only relying on public data; hence significantly pushing the frontiers of knowledge and having a positive impact on Society.

However, researchers using confidential data are inexorably challenged when it turns to research reproducibility. Indeed, how can they show that that their research is reproducible when their peers, referees, editors, etc cannot have access to these unique data? How can they signal the reproducibility nature of their research?

In this paper, we present a joint initiative conducted in France between the CASD (French Secure Data Access Center) and cascad (Certification Agency for Scientific Code and Data). They jointly propose to CASD users to attest that the numerical results in a given scientific article (tables and figures) can be reproduced from the code and confidential data used by the researcher. This certification consists in a rigorous evaluation process jointly conducted by a referee specialized in the software used by the researcher and by an expert from the particular scientific field (an Editor). At the end of the process, a certification rating is delivered to the researcher, with RRR being the highest potential rating. The researcher can transmit the reproducibility certificate along with the manuscript when submitting a paper to an academic journal. Hence, such certification process enriches the « peer review » process of research.