

# 9th International Digital Curation Conference Workshop on Data Citation Principles

Feb 27 2014

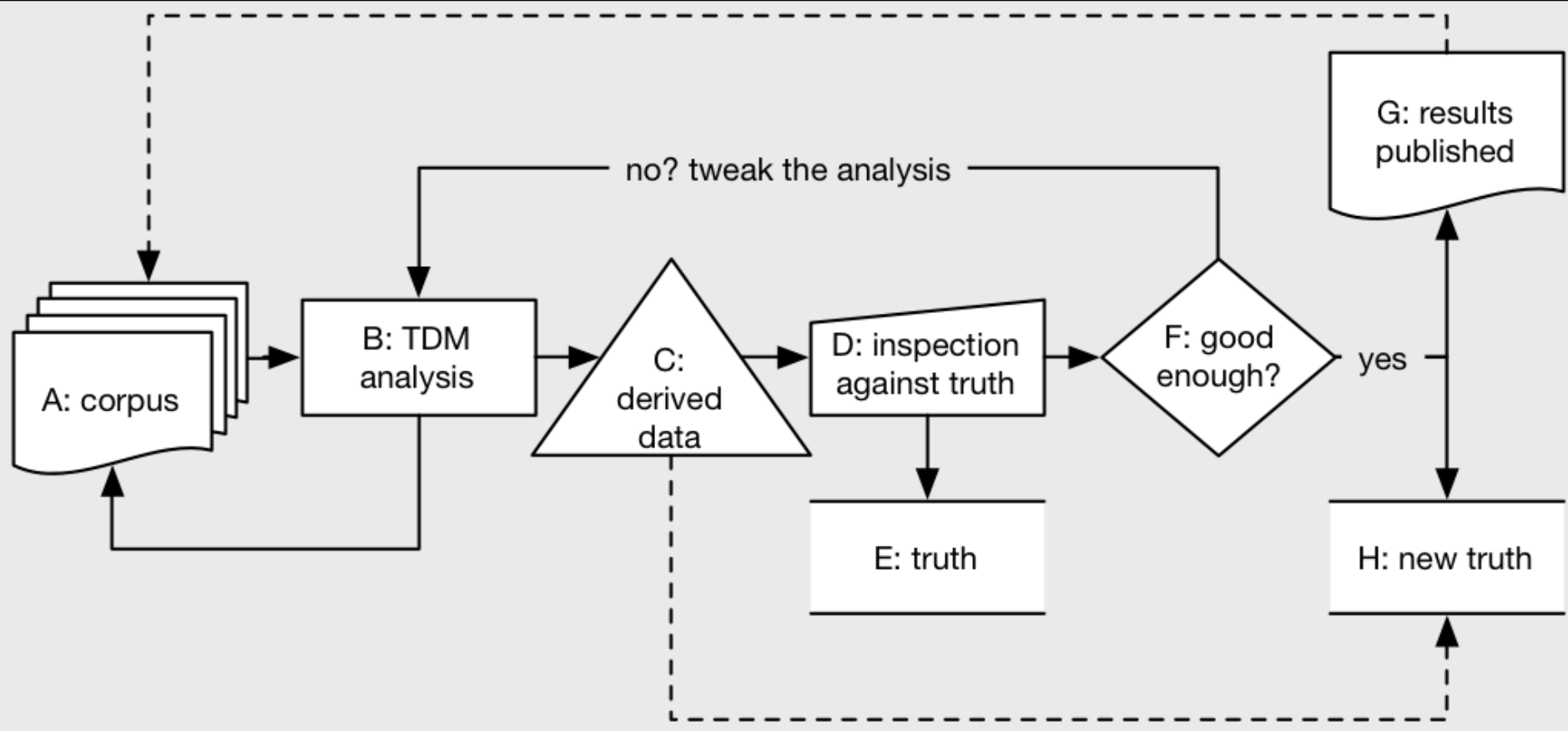
## Considerations for Text and Data Mining (TDM), Data Citation and Licensing

Puneet Kishor, Science and Data Policy

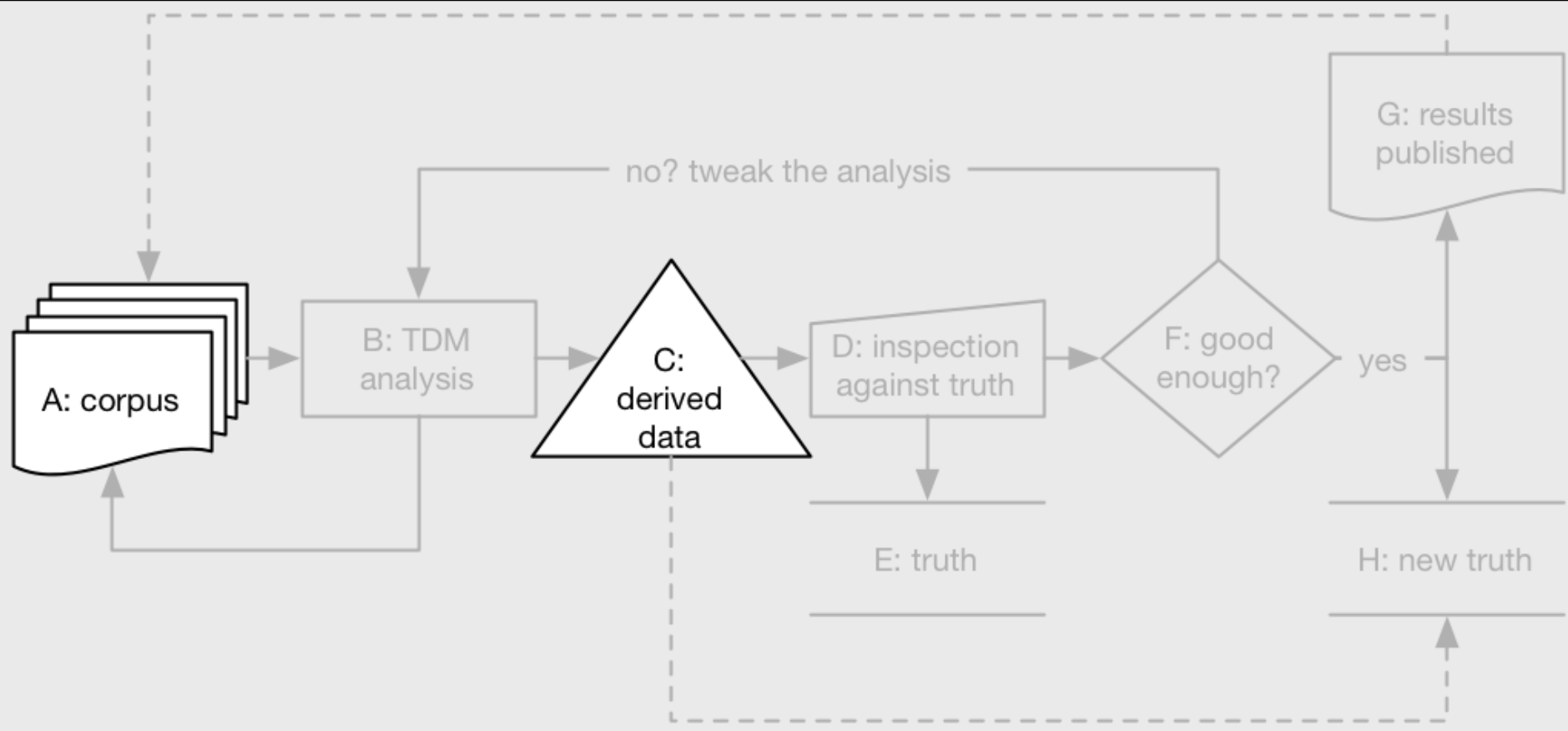
Creative Commons

Placed in the public domain under the [CC0 Public Domain Dedication](#)



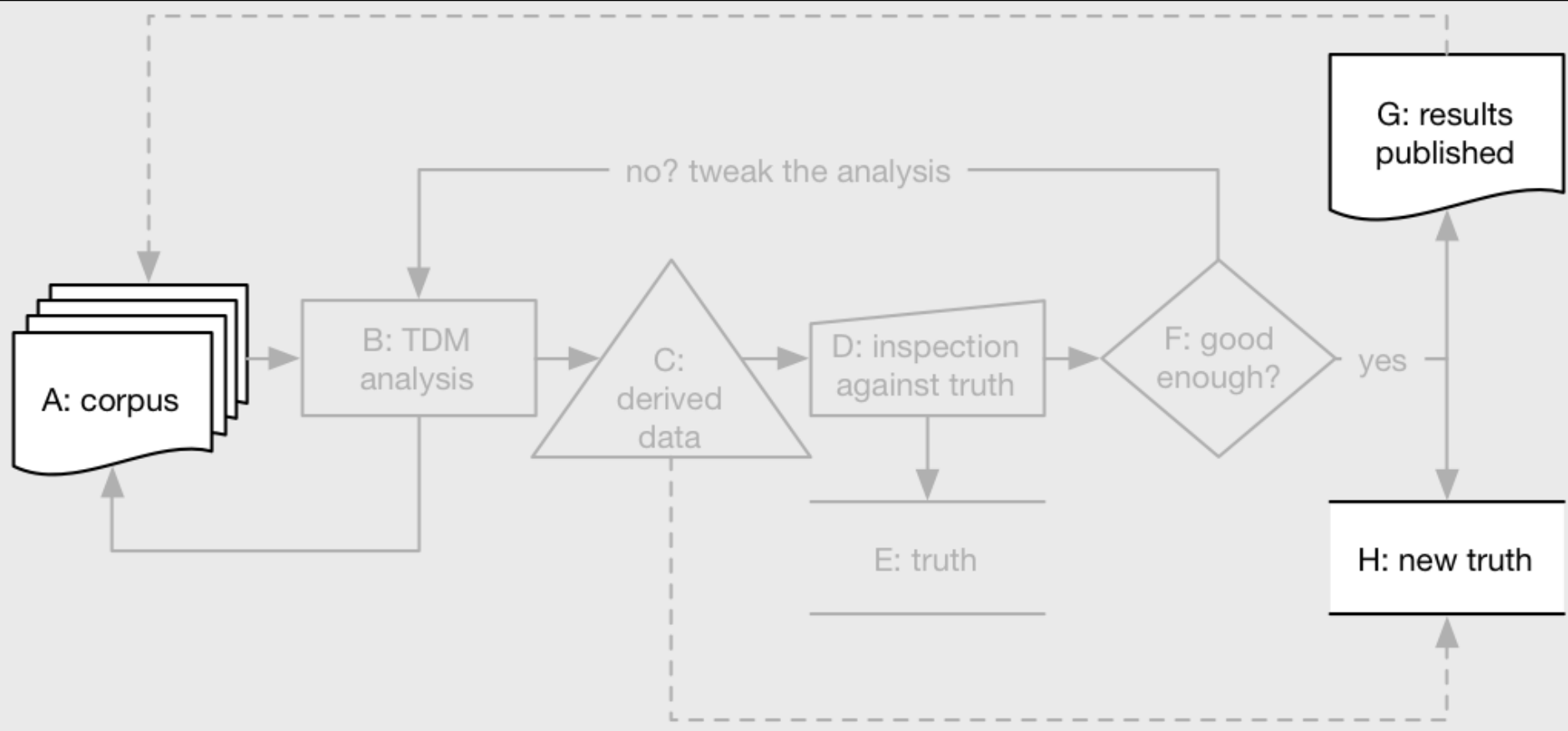


# Typical Text and Data Mining (TDM)



Data should be citable. But what is data in TDM? Is it the corpus being mined, or the product derived from the corpus?

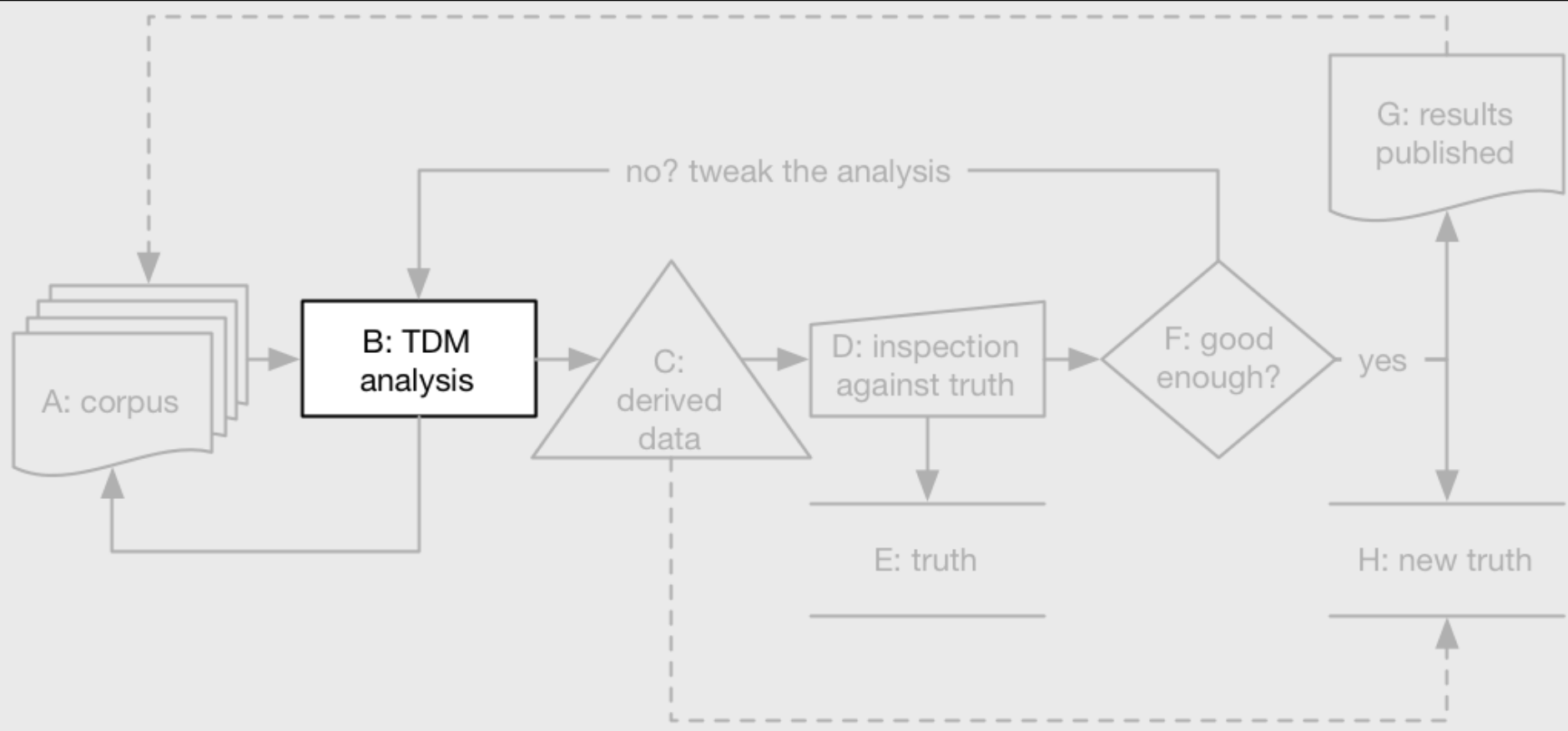




Scholarly credit and legally required attribution should be provided. If the corpus is made of half a million articles, would all of them have to be cited?



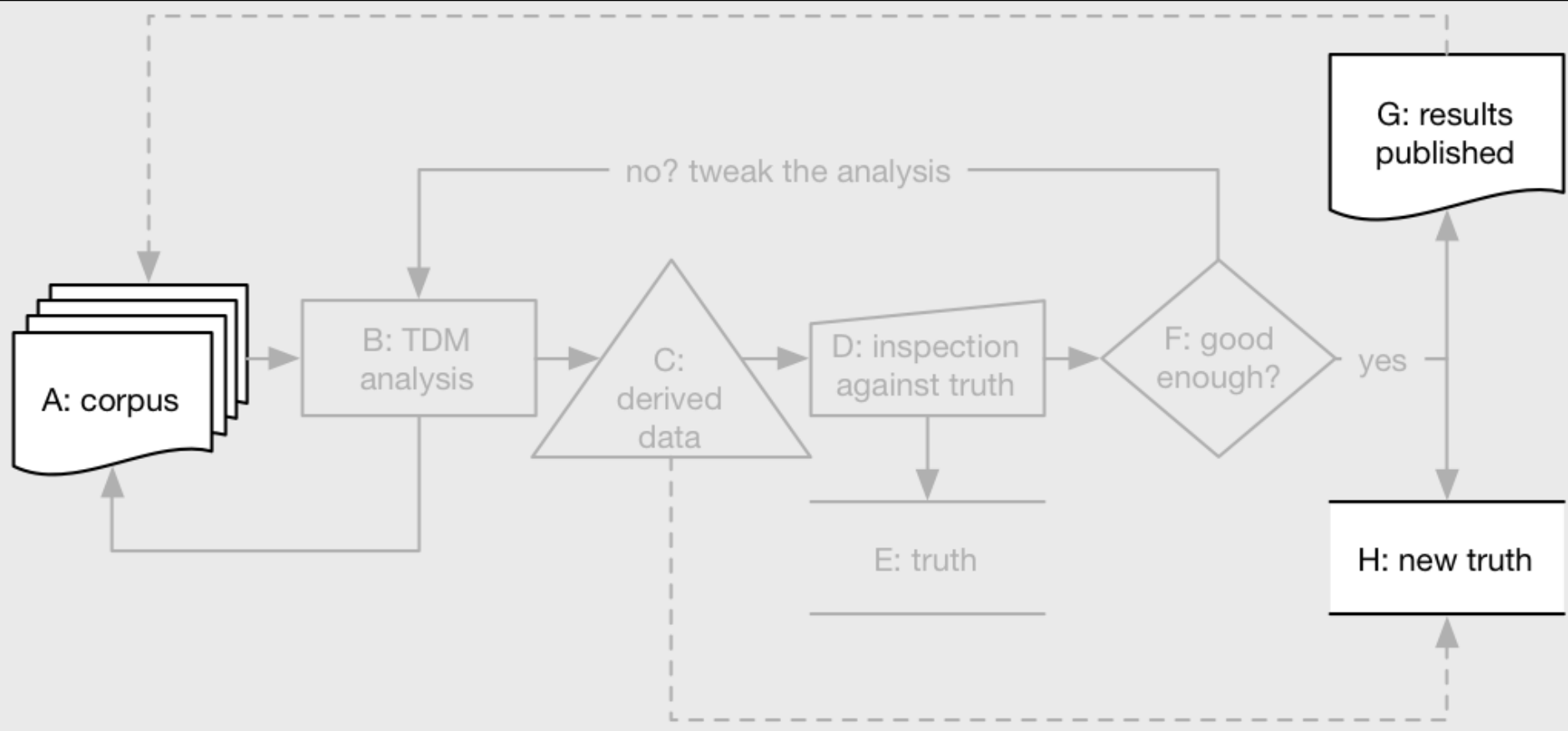
The **Credit and Attribution** Principle



Copyright allows TDM as fair use (in the US). Note: a CC License does not apply to uses (such as TDM) that qualify as Exceptions and Limitations, and the user does not need to comply with terms and conditions of the license if the TDM doesn't implicate copyright

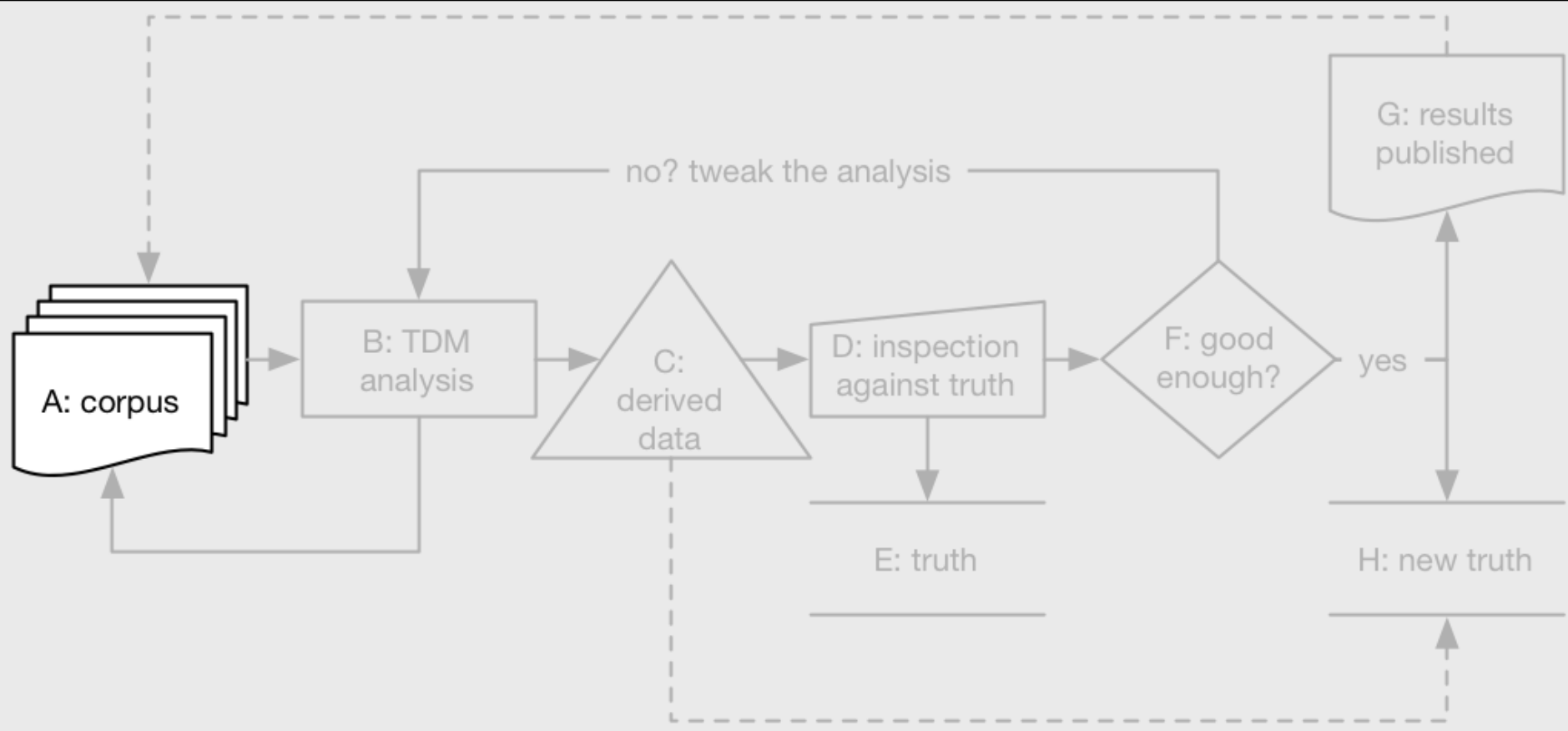


TDM as Fair Use



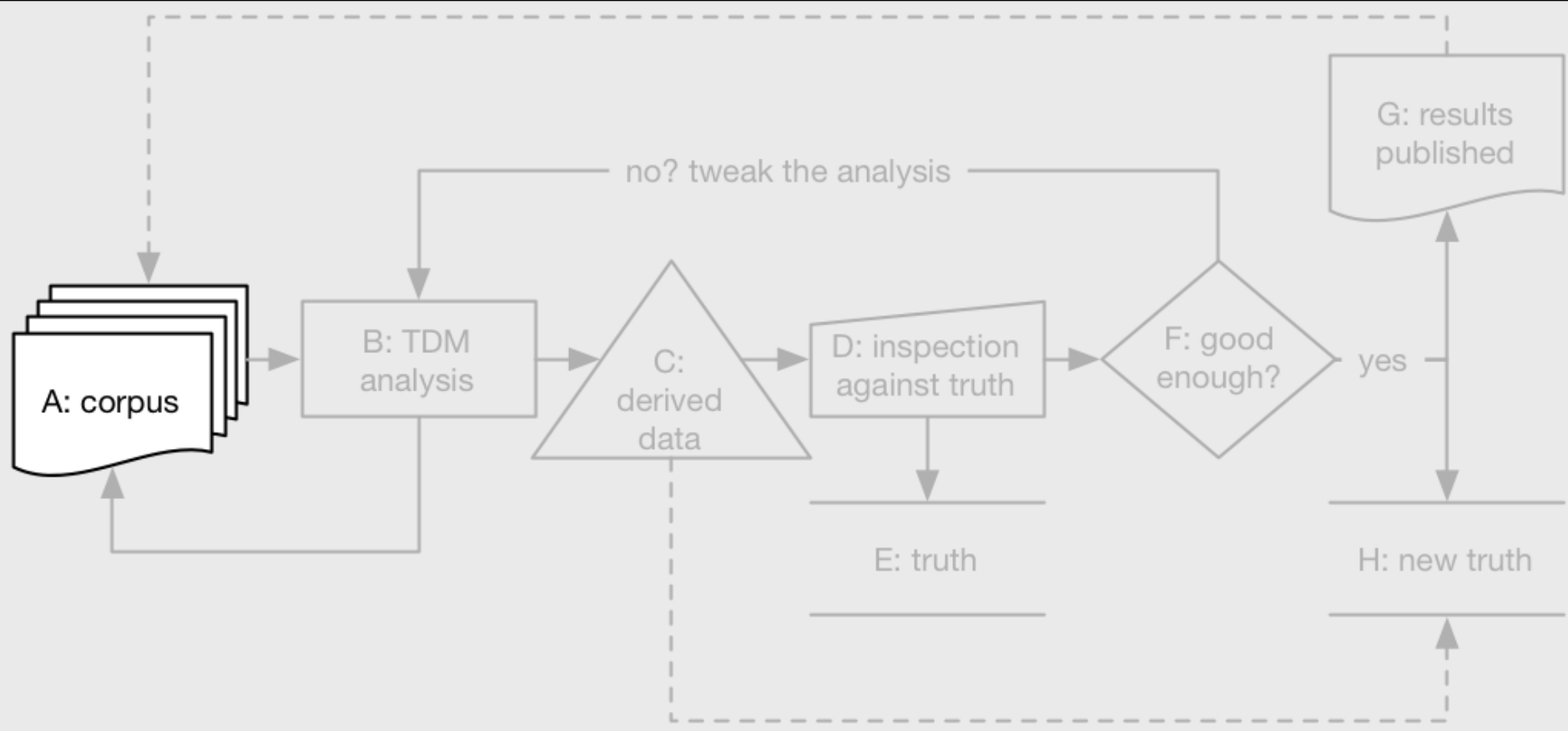
Licenses/contracts on the corpus can affect the license under which the final results and data are published. What would happen if the corpus were made of articles under different kinds of licenses?





Should scholars be able to get back to the corpus?





Should the corpus be open access and be available online for downstream users?



The **Access** Principle