Management Science Data Provenance Policy January 2019

Current Policy

We certify that the authors of the manuscript have the right to use the data and publish the research results contained in the manuscript. This includes data that may be collected from public sources such as open web-sites. When data are collected from public sources, they should not violate the fair use policy of the data owner, e.g., the web-site. In the cases where corporate or other legal permissions are required to use the data or the results derived from the data, these must be obtained before submitting the manuscript for review. A violation of the data provenance may result in withdrawal of the manuscript from Management Science (even after acceptance, retroactively). To understand fair use policy, authors are encouraged to review Measuring Fair Use: The Four Factors **here**.

The Intent of Policy and Interpretations

The core intent of the policy is for authors to not use data obtained by means that materially harm individual, business, public sector, or societal interests. Typically, in the United States, IRB reviews protect the interests of individuals and society. However, it is not clear, for example, whether business interests are kept in mind when reviewing IRB applications. It is the intent of data provenance policy to protect "material" business interests of an entity.

The biggest challenge in applying this policy remains in the area of data obtained from publicly available online sites and portals. Data scraping seems to be a core tool for researchers to discover, explain and theorize emerging phenomenon. Most of the commercial web sites explicitly ban collection of data through scraping. However, a lack of case law and findings make it difficult to interpret whether sites that display this scrape data publicly remain rightful owners of data to enforce collection of data unless they can demonstrate a material harm that may come to its users or its operations (e.g., due to aggressive scrapping strategies that may affect operational performance of a business owners' servers). Therefore, operationally, we suggest the following interpretation and implementation of the data provenance policy:

1) We would allow review or publication of the papers that use scraped data. However, such papers may be retroactively withdrawn if an entity complains to INFORMS and demonstrates that data collection inflicted significant material harm to the business interests of that entity (e.g., operational performance of servers were affected by the aggressive scraping strategies).
    a. The interpretation does not apply to the results of the studies, i.e., if the results discover and demonstrate findings that may be interpreted as inflicting material harm that in itself cannot be a reason to retract the paper.
2) We would not allow review or publication of the papers that use the data that was "originally" obtained in illegal manner (e.g., stolen or hacked) regardless of the availability of the data in public forum. The only way to use such data for research purposes is to obtain explicit permission from a company to use this publicly available data – in many instances a company may grant this permission since the data is already available and they may not see any further harm.
    a. Examples of such data may be wikileaks data or data stolen from companies such as Ashley Madison.

b. The rationale for the exclusion is to prevent unethical acts such as stealing data and putting them on public forums. In addition, it clearly violates the norm of not using data that materially harmed business or public interests.
c. In case private data of a company is used, authors must provide evidence that they have the right to use these data for research purposes.