Industry Innovation

Data Integrity & Ethical Considerations

Data is power. Its far-reaching influence can have a significant impact on improving lives and communities. It is our new pulse on the world. Every day we are confronted with numbers and statistics. However, we must be mindful of the integrity that data can potentially have or lack and the ramifications therein.

It is essential to leverage the correct tools for understanding how to frame and design the problem (or data) that you are presented with. Then there is a second step of verifying and ensuring that the data you have is accurate. Lastly, the ethical considerations that surround our data environments should always be kept in mind. Since data has such a significant impact on how we perceive the world, ethics is an encompassing umbrella around the whole analytical process.

Throughout history, major events have occurred to compromise data integrity and expose security flaws. In today's world, we have a necessary blind trust with data requests. How many forms have you filled out that ask you if you agree to the Terms and Services, but the only choice is…'yes'? Additionally, it is all too common for impatient users to skip reading Terms and Services altogether. Whether you are the one working with the data or the one who benefits from the data as the consumer, the responsibility to uphold data integrity is critical in today's connected environment.

Data

There are those that input data and those that process data. Listed below are some of the main ways that data can be compromised and how one can minimize risk:

- ***Human error*** occurs when the user or operator inadvertently changes the

value or accessibility of data. One method to combat risk of human error is to restrict the user interface so only relevant and controlled alterations can be made.

- **Unintended alterations** occur due to human error or incorrect permission settings. One method to minimize this risk is to have consistent error checking routines/queries and validation procedures in place.
- **Viruses, malware, and hacking** is always a potential source of data compromise. The best way to combat these is prevention, such as, having proper cybersecurity protocols and correct user permissions set at every level.
- **Data transfer – device to device glitches** and **hardware crashes** are some of the most frustrating occurrences of potentially losing valuable data. To minimize risk, having adequate backup and data recovery protocols in place is vital. Always ensure you have redundancy by having data duplication protocols in place.

When looking at ethical considerations surrounding data, data stewardship is an important responsibility. Make sure data management personnel—including analysts, database managers, and researchers—are knowledgeable of the guidelines, regulations, applications, and procedures regarding the ethical use of data. Ensure compliance with confidential data from all angles. Below are a few examples of how direct identifiers, indirect identifiers, and linked data need to be considered in ensuring protection of confidential data.

- **Direct Identifiers** = name, address, social security number
- **Indirect Identifiers** = combination of location, health information, ethnicity, etc. that could lead to exposure of identifying an individual
- **Linked Data** = combining multiple tables of data, where one of the tables houses data that could potentially identify an individual inadvertently.

Analytical Procedures

Analyzing data in a powerful and accurate manner is key. Listed below are data considerations that can be encountered in the analytical process and how one can minimize risk.

- **Data sources:** before one can start on drilling down insights, the actual data source should be checked. Adequate data is needed to run meaningful analytic procedures.
- **Analytical and statistical methods** need to be reviewed and decided on before analyzing data. Best way to way to approach this is by following/establishing ethical evaluation method protocols for your company.
- **Eliminate bias** by focusing on your sample design procedures. Also, keep in mind the data you are including and/or excluding. In some circumstances, you may need to deal with missing and/or incomplete data and the best way to handle these is by documenting your methods and following company guidelines.

- *Outliers* should be considered when looking at a data set. Maintain complete documentation of all related decisions when including or excluding outliers.

The American Statistical Association's Ethical Guidelines for Statistical Practice are intended to help statistics practitioners, researchers and general data personnel make decisions ethically. Always have methodology documentation when analyzing data. Ensure that your process is reproducible to maintain reliability and always avoid potential disclosure issues with confidential data.

Data Reporting

A data report is very influential and requires the due diligence of unbiased reporting. Listed below are data reporting considerations and how one can minimize risk:

- *Statistically significant findings* should be reviewed before publishing/reporting. Critically evaluate the reasons for inclusion or exclusion of data and avoid cherry-picking results.
- *Graphics/visualizations* are very important in reports and need to be chosen critically. Avoid misleading or confusing graphics and visualizations.
- *Variability/error* should be analyzed before publishing to ensure finding are statistically significant. Critically evaluate methods of statistical testing and reporting findings.
- *Limitations of data* are very important to include. Discussions of estimate variability, error and limitations should be recorded and included in your methodology section of a report.

All possible sources of data bias should be addressed and documented. Documents, contracts, presentations, and all communications should be routinely inspected for possible disclosure of confidential information.
Each of us handle significant data on a regular basis. As you encounter data on a project, keep in mind the acquisitions, ownership, and management of the data. Ensure that you put an appropriate emphasis on user permissions and access. Ensure you are using a secure network or portal for data transmissions that contain confidential information. An organizational culture of compliance with policies and procedures is key when implementing company guidelines for data.

*Please contact **innovation@necanet.org** if you have any questions.*

Suite 1200, Washington, D.C. 20004, United States, 202-991-6300