

**CONSISTENCY OF OUTPUT
FROM REMOTE ACCESS
SERVERS, MICRODATA, AND
TABULAR DATA RELEASE:
THE NEED FOR COMMON
PERFORMANCE METRICS**

**Muralidhar,
Sarathy,
Trottini**

**DATA GATHERED
BY GOVERNMENT
AGENCIES ARE AN
IMPORTANT
SOURCE OF
INFORMATION TO
THE GENERAL
PUBLIC.**



TYPES OF DATA ACCESS

- Microdata
 - Tabular Data
 - Remote Analysis Systems
- It is likely that all three types of access will be employed ... at least in the short run

CONSISTENCY OF OUTPUT

- Necessary in order to
 - Ensure that output from one type of access does not compromise data in another access type of access
 - Convince the public that the statistical disclosure methods employed are effective both in providing meaningful data for analysis purposes and in preventing disclosure of confidential information.

IMPORTANT ISSUES TO CONSIDER

- Access method, disclosure risk, and information loss
 - Development of global measures applicable across all methods of access
 - Difficult problem as evidenced by the extensive literature on the tabular data/microdata comparisons

IMPORTANT ISSUES TO CONSIDER

■ Data Characteristics

- The underlying characteristics of the data may dictate our selection of the access provided and the adoption of the disclosure risk and information loss measures
- One size fits all is unlikely to succeed

IMPORTANT ISSUES TO CONSIDER

- **Input versus Output masking**
 - This age old question still remains unanswered
 - Input masking for microdata
 - Output masking for tabular data
 - End result – inconsistencies
 - Both are feasible for remote analysis but which one is preferable?

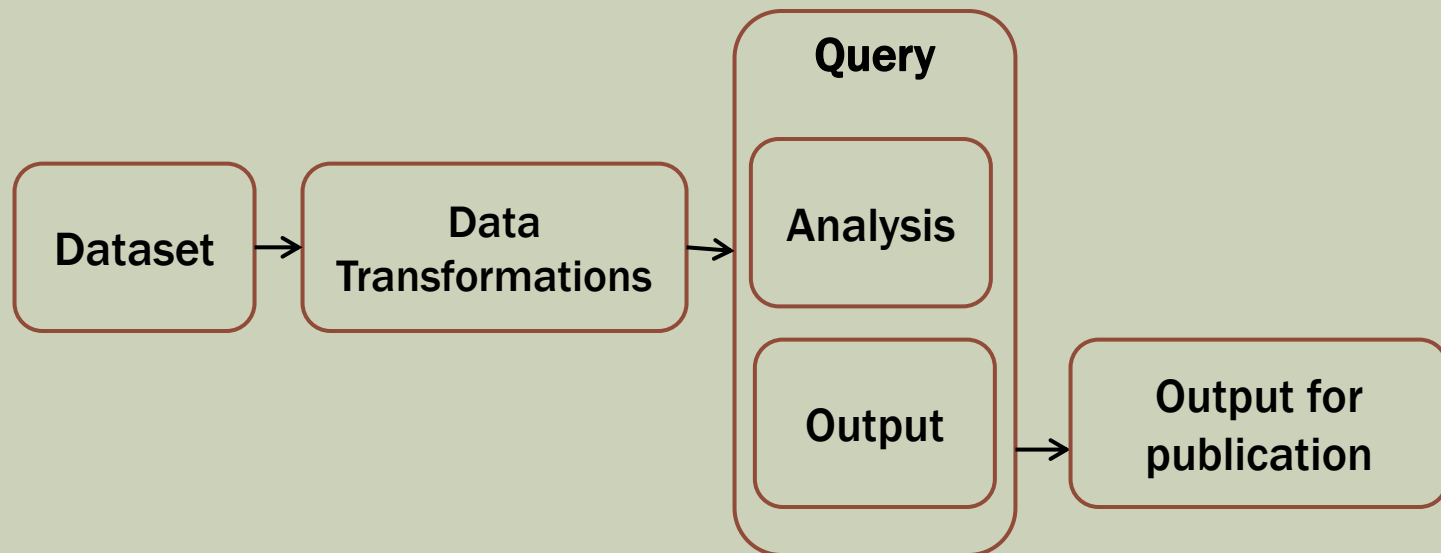
ADDITIONAL CRITERION?

- **Flexibility**

- Let us consider the combination that offers the most flexibility in terms of both access and masking

FLEXIBLE ACCESS METHOD

- Remote Analysis Systems
 - A single system that responds to all queries
 - tabular queries, statistical queries, perhaps even microdata access (if necessary)



FLEXIBLE MASKING MECHANISM

- **Input or source masking**
 - Mask original data and use the masked data to respond to all queries
 - Response consistency
 - Ability to assess information loss and disclosure risk characteristics prior to release
 - Unsuitable for data that changes frequently

POLICY QUESTION

- Data administrator's have to decide on the appropriate mode of access and data masking mechanism to ensure that they provide the public with the information that is accurate, safe, and easy to access
- Policy issues (including legal requirements) are likely to play a greater role in this decision than just technical issues

THE DESIRED END RESULT

- An integrated system with a single data source offers the best solution to protect the confidentiality of the sensitive data while simultaneously providing the users with analytical results that are meaningful and consistent.

THANK YOU

- **Comments, suggestions, questions?**