Alternatives to "Cold Rooms" for Restricted-use Data

Introduction

- More and more studies are producing data with disclosive and sensitive variables.
- As the risk of re-identification and harm increases, so does the security level needed to ensure protection to participants.
- Modifying data for public use may reduce the analytic value tremendously.
- Challenge is to provide appropriate data protections while making data as accessible as possible.

Options for Providing Access to Restricted-Use Data

- Physical enclaves or "cold rooms" are often used to make data having increased risk of re-identification and potential for harm available.
- Virtual analysis systems and batch systems with synthetic data provide the same level of protections as cold rooms and are more accessible to researchers.



COLD ROOMS & THEIR ALTERNATIVES

Traditional Cold Rooms

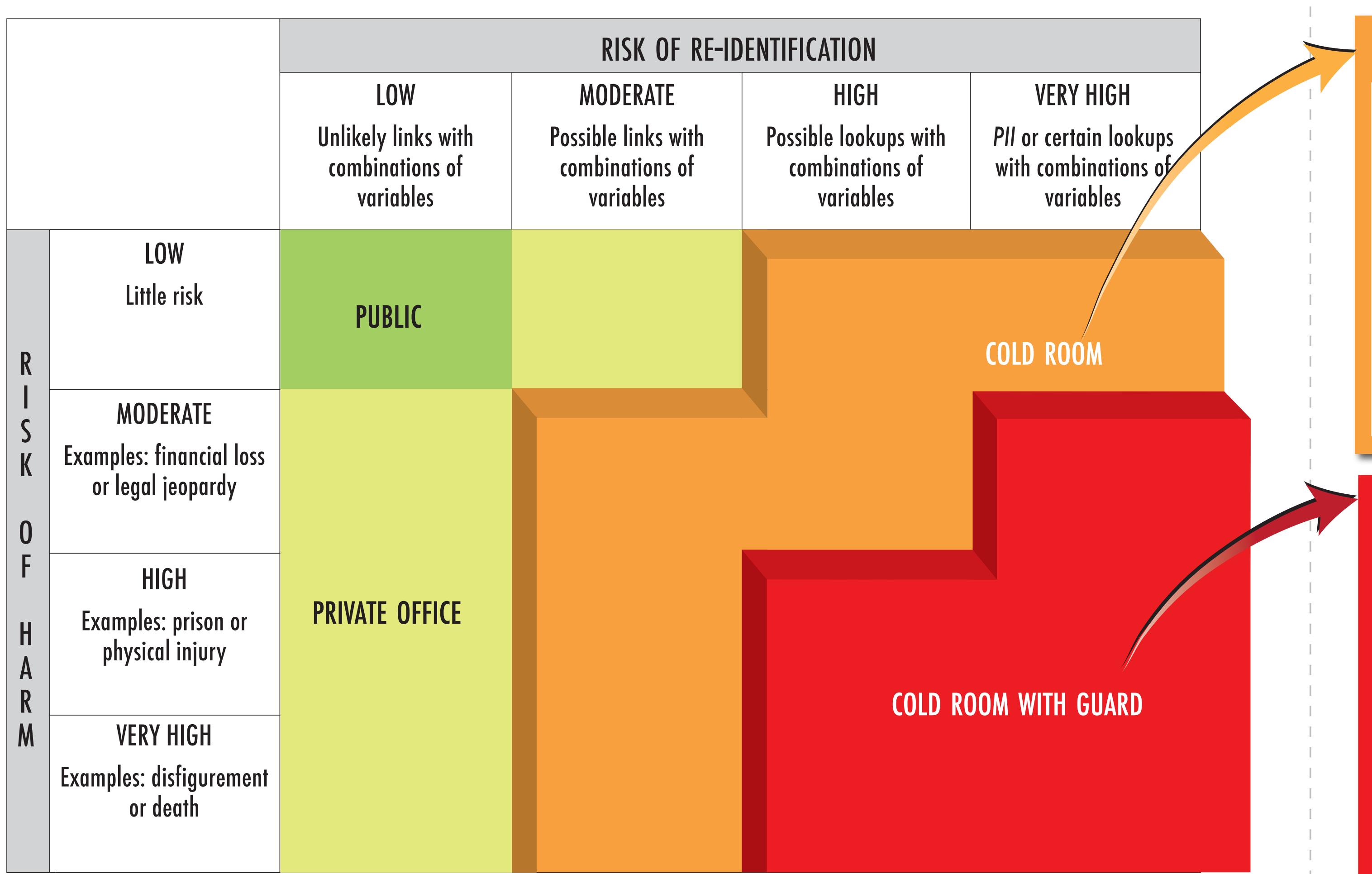
- Types: Unguarded and Guarded.
- Computers not connected to public network or printers.
- USB ports for connecting flash drives are disabled and inaccessible.
- Researchers must submit requests to remove output.
- Output released to researcher only after passing disclosure review.

Benefits

- + Data cannot be copied.
- + In guarded room, data manipulation is monitored and notes may not be removed from the room.

Disadvantages

- Inconvenient; often requires travel.
- Personnel to maintain and guard rooms makes them costly to operate.
- Space is often a premium resource at universities and other research organizations.



Virtual Analysis System

• Encrypted connection to online server for data analysis while blocking connections to the Internet.

- Alternatives to Cold Rooms -

- Files can't be copied from the server.
- Researchers submit requests to remove output.

Benefits

+ Researchers connect to the server from a specified, private location (e.g., office).

- Like Cold Rooms, note taking isn't prevented.

Disadvantages

- + Facilitates collaboration; each member of project team can connect to server from an approved location.
- + Server may offer more processing power than a desktop computer.

Batch System & Synthetic Data

- Submitted data analysis programs run on a server with the restricted-use data.
- No access to restricted-use data.
- Output is only provided to researcher after passing disclosure review.

Benefit

- + No travel required to submit programs.
- + Less costly than a guarded cold room.
- + Prevents unauthorized note taking.

Disadvantages

- Researchers must wait to receive output.
- Programming mistakes may take days to detect.
 - To address this, synthetic data can be used to debug programs. Fully tested programs can then be submitted to run on the restricted-use data.