



Practical Examples for Data Masking and Data Sharing

Seema Sayala, Ph.D

July, 2023

Approved for Public Release, NGA-U-2023-01653

Data Confidentiality: Definitions for This Presentation

OPM uses the terms data masking and data suppression, however, definitions are specific to this presentation:

Data Masking

Hiding/removing data in order to prevent or limit the disclosure of individual respondents.

Some agencies may also mask data to prevent use of unreliable estimates.

- Data suppression can be one form of data masking; it can mean not showing data in order to prevent disclosure
- The exact masking procedure will depend on the type of data you have

Of note: data suppression is also known as secondary cell suppression...
....one method of data masking is global recoding ...Local suppression can also be a form of data masking... Global recoding is also known as simply collapsing categories...

- Make sure to understand exactly what data masking procedures were used on the data you receive, or clearly describe what procedures were used on the data you are reporting out on!



Things to Consider with Your Data Before Reporting

- What is the business question being asked?
- What are my organization's policies for protecting PII and sensitive data?
- What data am I using, and therefore what masking thresholds should I apply when reporting out?
- Who is receiving the data report?

Data Scenario 1:

- The Finance Director has heard (anecdotally) that females in their Florida office feel that there aren't enough female supervisors, and this is affecting their perception of career growth opportunities there.
- What information can you give her to help her understand whether this is true, but also ensure that individual identities or sensitive information is not revealed in the data report?

Finance Division Data: Females in Supervisory Positions

	Original Data				
Supervisory Status?	Males %	Females %	Males n	Females n	Supervisory Status n
Non-Supervisor	50%	50%	25	25	50
Team Lead	50%	50%	8	8	16
Supervisor	76%	24%	19	6	25
Manager	90%	10%	9	1	10
	Males %	Females %	Males n	Females n	
Gender Overall Finance (FL)	60%	40%	61	40	



Note: Data is representational and does not contain actual data from any organization/NGA

Females in Supervisory Positions, Suppressed Data

Data Masking Rule: Hide cells that have fewer than 4

Supervisory Status?	Suppressed Data		Original Data		
	Males %	Females %	Males n	Females n	Supervisory Status n
Non-Supervisor	50%	50%	25	25	50
Team Lead	50%	50%	8	8	16
Supervisor	76%	24%	19	6	25
Manager	--	--	9	1	10
	Males %	Females %	Males n	Females n	
Gender Overall Finance (FL)	60%	40%	61	40	



Note: Data is representational and does not contain actual data from any organization/NGA

Females in Supervisory Positions, Data Report

Supervisory Status?	Males %	Females %
Non-Supervisor	50%	50%
Team Lead	50%	50%
Supervisor	76%	24%
Manager	--	--

	Males %	Females %
Gender Overall Finance (FL)	60%	40%

Note: Counts that are too small to report (< 4) are indicated with dashes

Report Conclusions:

- Females have higher representation than expected in the Team Lead position (50% vs 40%), but are underrepresented at the supervisory level (24% vs 40%).
- Additionally, females are underrepresented at the manager level (data too small to report).
- Essentially, females don't appear to be progressing much past the team lead position



Note: Data is representational and does not contain actual data from any organization/NGA

Data Scenario 2:

- The Finance Director also hears that their office in the Northeast does not have an inclusive environment, and it's rumored that minority employees are looking for new jobs.
- What information can you give her to help her understand what's going on, but also ensure that no individual identities or sensitive information is revealed in the report?

Data Scenario 2a:

- To help the Finance Director to better understand the context of the results you will give her, you first provide the overall minority composition of the Northeast office

Finance Division Data: Overall Minority Composition, Northeast Office

Data Masking Rule: Mask cells that have fewer than 4

Minority Status	Original Data		Suppressed Data
	%	N	%
White	46%	90	46%
Black or African American	14%	28	14%
Hispanic	13%	25	13%
Asian	13%	25	--
Native American/Alaskan Native	0.5%	1	--
Native Hawaiian or Other Pacific Islander	0.5%	1	--
Two or More Races	13%	25	13%
Total	100%	195	

Hide this too, since only 1% is left, and counts could be back-calculated if total N is discovered.



Note: Data is representational and does not contain actual data from any organization/NGA

Overall Minority Composition, Northeast Office, Data Report

Minority Status	%
White	46%
Black or African American	14%
Hispanic	13%
Asian	--
Native American/Alaskan Native	--
Native Hawaiian or Other Pacific Islander	--
Two or More Races	13%
Total	

Note: Counts that are too small to report (< 4) are indicated with dashes

Data Summary:

- Whites are nearly in the majority in the northeast division (46%), with the next highest race/ethnicity category being Black or AA at 14% representation.



Note: Data is representational and does not contain actual data from any organization/NGA

Leave Intentions from the Employee Engagement Survey

Survey Question: I really want to be at Agency B a year from now

Response rate 80%, not all respondents provided their race/ethnicity data

Data Masking Rule: Mask cells that have fewer than 10

Minority Status	Original Data		
	% Agree	Agree n	Total respondents
White	80%	40	50
Black or African American	10%	2	20
Hispanic	20%	4	20
Asian	80%	12	15
Native American/Alaskan Native	0%	0	0
Native Hawaiian or Other Pacific Islander	0%	0	0
Two or More Races	20%	4	20



Note: Data is representational and does not contain actual data from any organization/NGA

Leave Intentions from the Employee Engagement Survey, Data Report

Survey Question: I really want to be at Agency B a year from now
Response rate 80%, not all respondents provided their race/ethnicity data

Collapsed Category	Minority Status	% Agree
	White	80%
	Black or African American, Hispanic, Two or More Races	17%
	Asian	80%
	Native American/Alaskan Native	0%
	Native Hawaiian or Other Pacific Islander	0%

Report Conclusions:

- Not all minority groups want to leave; the employee climate seems to be impacting leave intentions for the employees that are Black, Hispanic, and Two or More Races.



Note: Data is representational and does not contain actual data from any organization/NGA

Closing Thoughts

- There is no one-size-fits-all for how to mask data.
- It is a balance between maintaining data privacy and making sure the data is useful.
- Establishing formalized policies for data masking in your organization can help justify to decision-makers/customers why they can or cannot see certain data.



Know the World, Show the Way...from Seabed to Space

Approved for Public Release, NGA-U-2023-01653