

Data Risk Assessment Matrix

NCAR Example

Slides by Matt Mayernik, 7/16/2019

Motivating questions

- How to clarify what data are at risk, and what risk factors they face?
- How to make data risks more transparent, internally and/or externally?
- How to prioritize mitigation efforts for large/diverse data collections?

Findings from prior exercises

- No one way of categorizing risks
- May need to categorize risks in different ways depending on the situation

Data Risk Assessment Matrix

Risk Factors (more not shown, 21 total)

<i>Categorization Method</i>	Legal status	Missing files	Lack of metadata	File format obsolescence	Data mislabeling	Loss of funding
<i>1. Length of recovery time</i>						
<i>2. Impact on user</i>						
<i>3. Who is responsible for addressing the problem</i>						
<i>4. Cause of problem (e.g. internal vs ext)</i>						
<i>5. Degree of control</i>						
<i>6. Proactive vs reactive response</i>						
<i>7. Severity of risk</i>						
<i>8. Likelihood of occurrence</i>						
<i>9. Nature of mitigation</i>						
<i>10. Resources required for mitigation</i>						

NCAR Example - Analog data within NCAR Library



Use of Matrix - Step 1

Identify which risk factors were most important for these materials

Most immediate risks

Other important risks

Lack of use	Lack of documentation & metadata	Data mislabeling	Legal status for ownership and use	Media deterioration	Lack of planning	Poor data governance
--------------------	---	-------------------------	---	----------------------------	-------------------------	-----------------------------

Use of Matrix - Step 2

Identify which categorization method is most applicable

1. Length of recovery time

3. Who is responsible for addressing the problem

9. Nature of mitigation

10. Resources required for mitigation

Use of Matrix - Step 3

Try to fill in some cells. More guidance for this step is needed.

For the “Length of recovery time” - I decided to use numbers to indicate how big of a problem this was. In other words, “3” means that this will be hard to mitigate (could take a long time)

<i>CATEGORIZATION METHOD</i>	Lack of use	Lack of documentation & metadata
1. Length of recovery time		
<i>O Clima de Portugal</i>	3	2
<i>USSR WDC Remote Stations</i>	3	3
<i>Taiwan, Aerological Observations</i>	3	2
<i>Taiwan, Annual report of weather bureau</i>	3	2
<i>Alaska High Latitude Geophysical data</i>	3	1

Use of Matrix - Step 3 (part 2)

For “Resources required for mitigation”, numbers didn’t make sense. A text description seemed more appropriate.

CATEGORIZATION METHOD	Lack of documentation & metadata
10. Resources required for mitigation	
<i>O Clima de Portugal</i>	Would need to create new metadata for library catalog, then transform to ISO for inclusion in NCAR DASH Search
<i>USSR WDC Remote Stations</i>	Same, but with added challenge of needing to look at microfilm files (no current working reader in Library)
<i>Taiwan, Aerological Observations</i>	Same, but with added challenge of possibly needing to translate language
<i>Taiwan, Annual report of weather bureau</i>	Same, but with added challenge of possibly needing to translate language
<i>Alaska High Latitude Geophysical data</i>	Would need to create new metadata for library catalog, then transform to ISO for inclusion in NCAR DASH Search

Preliminary takeaways

- Matrix was very useful as "something to think with"
- It jump starts the process for doing the risk assessment
 - You don't need to spend a time thinking about what the risk factors are, or how to categorize them, because they are laid out in the matrix.
- It is not a maturity model that you use to say "we meet some range of acceptable risk"
- It is also not a rubric, though could perhaps be applied/customized to create a rubric (e.g. through creating a numeric scoring process for categories where that is appropriate)