



Active Data Management Plans – a vision of what we can achieve

Tomasz Miksa

Vienna University of Technology, Austria

Perspective change

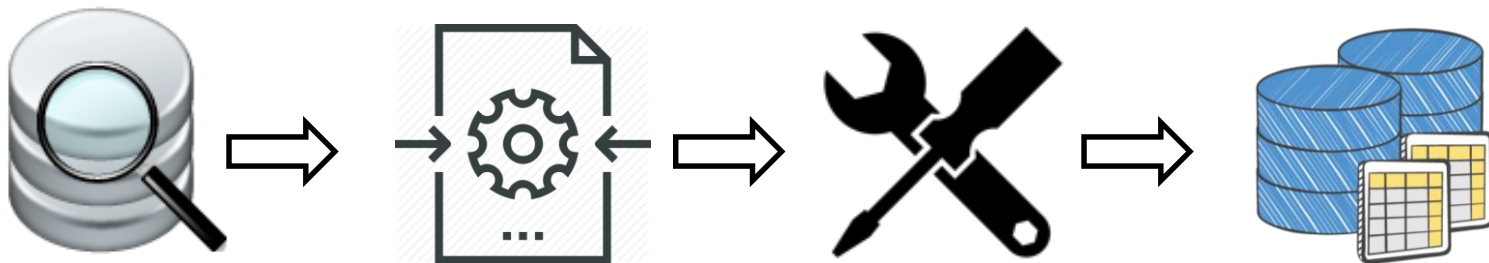
- Change perspective to identify the gap between needs of researchers and reality
- Become a scientist reusing others data
- Focus on your researcher needs
- Forget (for a while) about DMPs

- Do we need DMPs at all?



What does your life look like?

- During your research you want to
 1. find related projects and data
 2. re-execute them and validate them
 3. reuse, modify, and build on top of them
 4. publish results and get credit



What do you use?

1. find related projects and data

- search engines
- scholarly communications

2. re-execute and validate

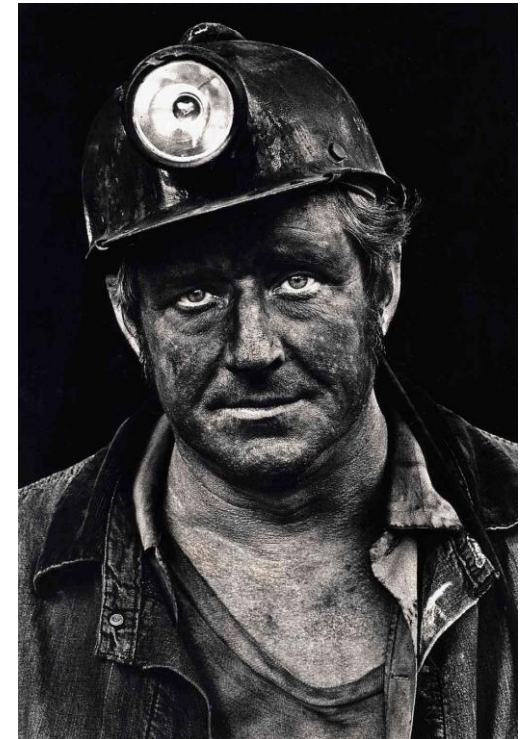
- provenance
- validation data
- system configuration
- deployment scripts or containers

3. reuse, modify, and build on top

- documentation, metadata

4. publish and get credit

- repositories, persistent identifiers, etc.



What do you use?

1. find related projects and data

- search engines
- scholarly communications

*What about experiments
that used the same input data?
Can we go beyond text search?*

2. re-execute and validate

- provenance
- validation data
- system configuration
- deployment scripts or containers

3. reuse, modify, and build on top

- documentation, metadata

4. publish and get credit

- repositories, persistent identifiers, etc.

What do you use?

1. find related projects and data

- search engines
- scholarly communications

*What about experiments that used the same input data?
Can we go beyond text search?*

2. re-execute and validate

- provenance
- validation data
- system configuration
- deployment scripts or containers

*How many of these really gets published?
How much time and money I need to spend to make it run?*

3. reuse, modify, and build on top

- documentation, metadata

4. publish and get credit

- repositories, persistent identifiers, etc.

What do you use?

1. find related projects and data

- search engines
- scholarly communications

*What about experiments that used the same input data?
Can we go beyond text search?*

2. re-execute and validate

- provenance
- validation data
- system configuration
- deployment scripts or containers

*How many of these really gets published?
How much time and money I need to spend to make it run?*

*Do I know the standard?
What's the coverage?
Will I be able to maintain it?*

3. reuse, modify, and build on top

- documentation, metadata

4. publish and get credit

- repositories, persistent identifiers, etc.

What do you use?

1. find related projects and data

- search engines
- scholarly communications

*What about experiments that used the same input data?
Can we go beyond text search?*

2. re-execute and validate

- provenance
- validation data
- system configuration
- deployment scripts or containers

*How many of these really gets published?
How much time and money I need to spend to make it run?*

*Do I know the standard?
What's the coverage?
Will I be able to maintain it?*

3. reuse, modify, and build on top

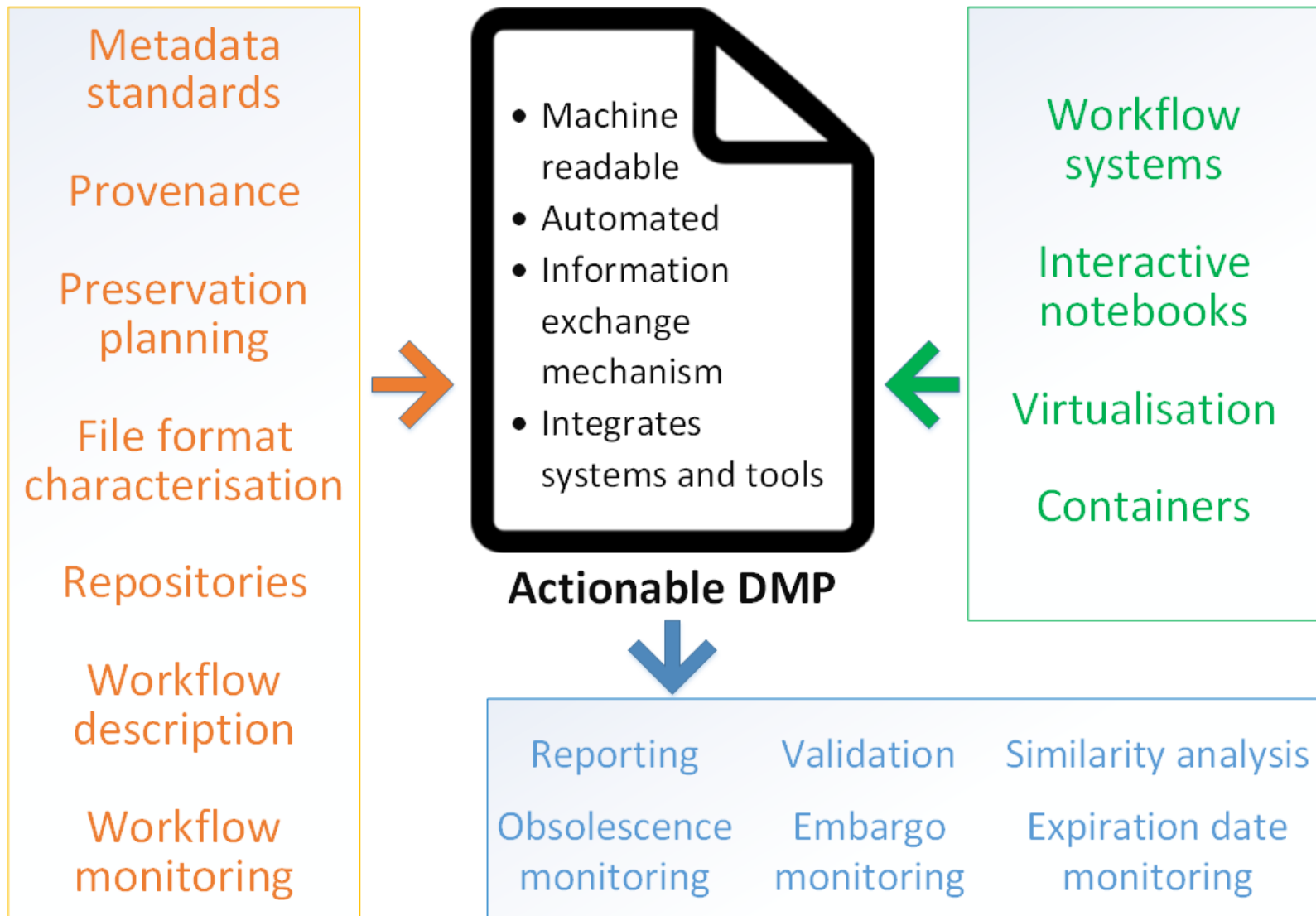
- documentation, metadata

4. publish and get credit

- repositories, persistent identifiers, etc.

*Which one to choose?
How will others find my data?*

What if we had all this information in one place?



What would your life look like?

1. find related projects and data

- publically available landing page of a project
- similarity search

2. re-execute and validate

- data is always available
- authenticity confirmed
- software requirements captured

3. reuse, modify, and build on top

- better coverage of information and higher trust

4. publish results and get credit

- automatically generated DMP of my project
- easier to find and to reference by others



Open questions

- How to exchange the information?
 - system integration (APIs)
 - open format/protocol/standard
- How to model information?
 - how to integrate and reuse existing standards?
 - which models, and dictionaries we need to develop?
- Do we need an open format/protocol/standard for DMPs?
 - can we have an open standard to which all conformant tools can write?
 - or do we focus on developing particular tools and integrating those?

Conclusion

- It is more than just work on DMPs!
- DMPs are bringing these developments together
- We have many solutions in place – we need to iteratively integrate
- We need use cases that are gradually more complex
- This is not utopia!

