

# DCC Spring Training: The role of good RDM in accelerating scientific progress

Presenters: Joanne Yeomans & Kristina Hettne, Leiden University Libraries

Facilitator: Margriet Miedema, LCRDM

Guest: Barbara Magagna, independent semantic expert, knowledge facilitator and expert on FAIR data management

THIS GOOGLE DOCUMENT WILL BE CLOSED FOR COMMENTS AT THE END OF JUNE AND THE PADLET WILL BE SAVED AND CLOSED.


9th June 2022, 9:30 - 12:30







## Contents of this document:




- [Agenda](#)
- [Code of conduct](#)
- [Who's who in the session](#) \*
- [What do you wish to learn from this session?](#) \*
- [Instructions and notes for activities](#)
- [Feedback and additional tips or resources](#)

\* (when you have time)

## Agenda:

9:30	 <b>Welcome and introductions (Margriet)</b>
	<b>Part 1: Reproducibility</b>
9:35	Introduction to the session (Joanne) / Explanation breakout room exercises (Joanne) Set up breakout rooms (10 rooms of 3-4 people) (Margriet)
9:45	Activity 1 in breakout room (All)

	<ul style="list-style-type: none"> <li>• Breakout rooms will contain 3 or 4 people.</li> <li>• Everyone opens the padlet and everyone can type: <a href="https://edu.nl/838xr">https://edu.nl/838xr</a></li> <li>• One person can share their screen if it helps.</li> </ul> <p> Round of introductions within breakout room: (1 minute each).</p> <p> Brainstorm “How to make research results reusable?” (15 minutes)</p> <p>Answer the following question in the same padlet:</p> <ol style="list-style-type: none"> <li>1. How can researchers make their results, data and code reusable?</li> </ol>
10:00	 Summary (5 min) (Joanne)
10:05	<p>Introduction to Reprohack and how it works (Kristina).</p> <ul style="list-style-type: none"> <li>• Presentation</li> </ul>
10:35	 <b>Break 15 minutes</b>
10:50	<b>Part 2: FAIR (meta)data in practice</b>
10:50	FAIR (meta)data basics (Kristina)
11:20	<p>Developing the I-ADOPT standard (Barbara Magagna)</p> <ul style="list-style-type: none"> <li>• Presentation</li> <li>• Questions</li> </ul>
11:55	 <b>Break 5 minutes</b>
12:00	<p>Activity 2 in breakout room (All)</p> <ul style="list-style-type: none"> <li>• set up breakout rooms (10 rooms of 3-4 people) (Margriet)</li> <li>• <a href="#">Find your group</a> in the GoogleDoc</li> <li>• One person can share their screen if you want.</li> </ul> <p> Brainstorm: Improving the metadata of a dataset (All)</p>

12:10	 <b>Summary</b> (10 min) (Kristina)
12:20	 <b>Wrap up:</b> Feedback ( <a href="#">one up, one down</a> ) & Additional Resources (Kristina & Joanne) <ul style="list-style-type: none"> <li>• What's next? Look out for training opportunities (DCC, GO FAIR, funders). In Leiden we are developing training as well.</li> <li>• As a basic: help researchers with their metadata</li> </ul>
12:30	 The end

## Code of Conduct

- Be inclusive and constructive
- Be patient with the technology/each other/yourself
- Let everyone have a chance to speak
- Be respectful of different viewpoints and experiences
- Gracefully accept constructive criticism
- Show courtesy and respect towards other community members
- Give, as well as take

What do you wish to learn from this session?

- 

## Improving the metadata of a dataset for reusability

These links relate to a research project about the chemical analysis of the ultramarine pigment in historical paints.

Step 1: Go to the publication:


<https://www.science.org/doi/10.1126/sciadv.aay8782>

Read the abstract of the paper to get an idea about the topic.

Step 2: Take a look at the README file for the code to get an idea of the contents and to read a summary of the datasets:

<https://github.com/alessaan/rhapsody-in-blue>

Step 3: Go to the data: <https://doi.esrf.fr/10.15151/ESRF-DC-186933507> You will need to 'log in as Anonymous' in order to reach the dataset. You should see the following page:



DOI > 10.15151/ESRF-DC-186933507

Data collection

DatasetOpen access

**SULFUR K-EDGE MICRO- AND FULL-FIELD XANES IDENTIFIES MARKER FOR PREPARATION METHOD OF ULTRAMARINE PIGMENT FROM LAPIS LAZULI IN HISTORICAL PAINTS**

Alessa GAMBARDILLA ; Marine COTTE ; Wout DE NOLF ; Kokkie SCHNETZ ; Rob ERDMANN ; Roel VAN ELSAS ; Victor GONZALEZ ; Arie WALLERT ; Myriam EVENO ; Katrien KEUNE.

DOI

DOI 10.15151/ESRF-DC-186933507

Licence (for files)

Creative Commons Attribution 4.0

**Abstract**

Ultramarine blue pigment from lapis lazuli is one of the most valued natural artist's pigments used throughout history for its brilliant hue. Historical recipes describe various pigment preparation methods, however, little is understood about why such treatments were performed or how to distinguish them a posteriori on historical paintings. One such historically relevant treatment is heating of the lapis lazuli rock prior to extracting lazurite. X-ray absorption near-edge structure spectroscopy (XANES) at the sulfur K-edge in microbeam and full-field modes (analyzed with non-negative matrix factorization, NMF) is employed to monitor the changes in the sulfur species within lazurite following treatments. Sulfur signatures in lazurite show dependence on the heat-treatment of lapis lazuli rock from which it is derived. In particular, peaks at 2469.0 and 2471.2 eV, attributed to contributions from the trisulfur radical (S<sub>3</sub><sup>•-</sup>), which is responsible for the blue color of lazurite, increase in relative intensity with heat of treatment ≥ 600 °C, paralleled by an intensified blue hue. The peak at 2472.5 eV (presumed as S<sub>8</sub>), on the other hand, decreases in relative intensity. The sulfur signature following heat-treatment is not only unique but also retained following accelerated aging of each pigment mixed as a paint with linseed oil. Further, XANES spectra using an attenuated focused microbeam were gathered on lazurite particles from five historical paint samples. In each, profiles matching that of pigment derived from heat-treated rock were identified, providing a marker for artists' pigments that had been extracted from heat-treated lapis lazuli.

Proposals	Beamlines	Publication year
HG-139	ID21	2019
HG-94	ID21	
HG-62	ID21	

Experimental report

One report has been found for this proposal.

90755\_A.pdf

Experimental data

The data can be accessed by clicking on the link below

Access data

Reference

Below is the recommended format for citing this work in a research publication.

Gambardella A., Cotte M., De nolf W., Schnetz K., Erdmann R., Van elsas R., Gonzalez V., Wallert A., Eveno M., Keune K. (2019). Sulfur K-edge Micro- and Full-field XANES Identifies Marker for Preparation Method of Ultramarine Pigment from Lapis Lazuli in Historical Paints. European Synchrotron Radiation Facility (ESRF). doi:10.15151/ESRF-DC-186933507

Step 5: Click on the big blue "Access data" button. You should see the dataset list as below.

Dataset List 23

Search

		Date	Sample	Dataset	Definition	Files	Size	Download
	<input type="checkbox"/>	15:44 26 Jul 2019	ultramarine	dataset_43	XANES	40	1.2 MB	<a href="#">Download</a>
	<input type="checkbox"/>	15:44 26 Jul 2019	ultramarine	dataset_42	XANES	3	648.1 KB	<a href="#">Download</a>
	<input type="checkbox"/>	15:44 26 Jul 2019	ultramarine	dataset_41	XANES	1	8.0 MB	<a href="#">Download</a>
	<input type="checkbox"/>	15:43 26 Jul 2019	ultramarine	dataset_40	XANES	332	276.1 MB	<a href="#">Download</a>
	<input type="checkbox"/>	15:38 26 Jul 2019	ultramarine	dataset_RAD	XANES	2	52.4 KB	<a href="#">Download</a>

Explore the data by clicking on the magnifying glass on the left hand side for the first dataset (dataset\_43).

There are three tabs: Summary, Files and Metadata List. See screenshot below.

	Date	Sample	Dataset	Definition	Files	Size	Download
	<input type="checkbox"/>	15:44 26 Jul 2019	ultramarine	dataset_43	XANES	40	1.2 MB <a href="#">Download</a>

Summary

Files 40

Metadata List

Name

dataset\_43

Definition

XANES

Start

3:44:34 PM

Sample

ultramarine

Description

/data/id21/inhouse/ultramarine/HG139/ultramarine/dataset\_43

[Download](#)

Write down your ideas for improving the metadata about dataset 43. What do you miss in order to understand what the data is about?

For example:

1. Is it clear what you are looking at and how to navigate the files?
2. Do you recognize these file types and know how to open them?
3. Can you say something about the use of standard vocabularies?
4. Does the ESRF archive system enable you to give clearer metadata? Do you have thoughts about how the archive itself could be better designed?

Step 4: Find a csv file and try downloading it.

1. Are the column headers clear?
2. Does the filename give you any information that could be useful?
3. Does the code documentation in github help you to understand what you're looking at? For example, dataset 43 is used to create figures in the paper, see this codebook documentation:

[https://github.com/alessaan/rhapsody-in-blue/blob/master/17\\_FigurePaint\\_Treatment\\_2018.ipynb](https://github.com/alessaan/rhapsody-in-blue/blob/master/17_FigurePaint_Treatment_2018.ipynb). Is the information in the codebook enough for you to understand the csv files?

Please make some notes of your thoughts below:

Group 1

1.

Group 2

1.

Group 3

1.

Group 4

1.

Group 5

1.

Group 6

1.

Group 7

1.

Group 8

1.

Group 9

1.

Group 10

1.

## One up, one down feedback

Please tell us what you liked and what can be improved with this session:

### One up

- Good presentations, interesting topics
- Reprohack was new to me
- Interesting questions from the audience
- Great to share experiences with other people with different backgrounds
- Nice to have a quite practical perspective of research supporters, this is very valuable and applicable to my own work.
- Great topics and presentations. Also enough room for discussion
- Very good and clear presentations (I'm definitely going to use the Reproducible/Robust/ etc. diagram)
- All of the items mentioned above..... → I second this.
- Like the initiative

### One down

- I really enjoyed the small group discussions and the different views/outcomes, maybe induce a couple more during the training.
- The practical implication on how to make data FAIR (especially the I) is still difficult to grasp, maybe for future use examples of variables that a lot of researchers collect (like age, gender, bmi) → agreed (I second this)

- Give some more info about what a datamanager can do with the reprohack or I-ADOPT, in practice. Both were new for me. How can I use it to help the researchers
- My breakout group only had two members so our discussion was a bit limited (but very “gezellig”!)
- Giving only one up is impossible!
- Would like to have more hands-on tips

*Please feel free to share with everyone additional tips or resources that you are aware of, or comments.*

#### Pre-registration information (shared in 2021)

- VU Data Conversations: “Personal Journeys into Preregistration”  
[https://zenodo.org/record/4455208#.YMs0\\_GgzaUk](https://zenodo.org/record/4455208#.YMs0_GgzaUk)
- Pre Registration <https://osf.io/>
- [jupyter notebooks](#) are heavily used in astronomy
- The Turing Way: A Handbook for Reproducible Data Science  
<https://zenodo.org/record/3233986#.YMs3Z2gzaUk>

#### A selection of data archiving solutions (shared in the 2021 training):

- YODA: <https://www.uu.nl/en/research/yoda>
- CLARIN: <https://www.clarin.eu/faq/what-clarin>
- Genbank: <https://www.ncbi.nlm.nih.gov/genbank/>

#### Resources useful for training (shared in the 2021 training):

- Data sharing and management snafu <https://www.youtube.com/watch?v=N2zK3sAtr-4>
- Slides and a guide for running a reproducibility workshop:  
<https://www.repro4everyone.org/resources>
- Reproducibility self-assessment tool: <https://www.protocols.io/reproducibility>

Please add your own suggestions above or below..

Other interesting resources:

- 

## NOTE:

The materials used to deliver this training session are available to use under [Creative Commons Attributions 4.0](#). Leiden University Libraries, Centre for Digital Scholarship