

Query: How is AI used in bioimage analysis?

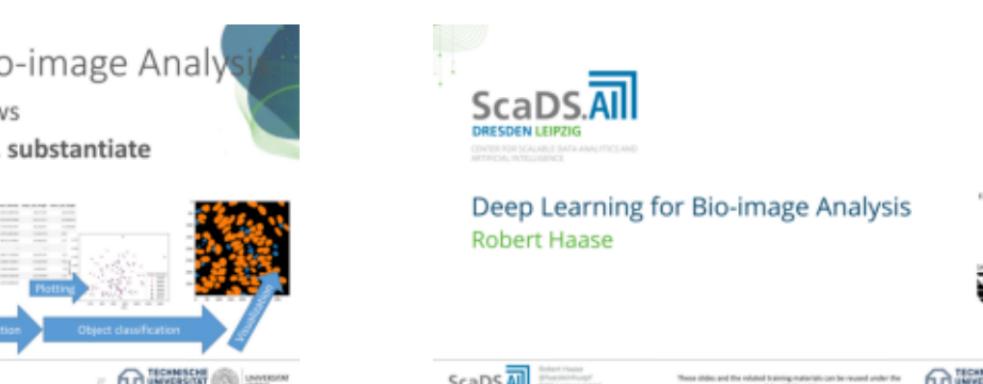
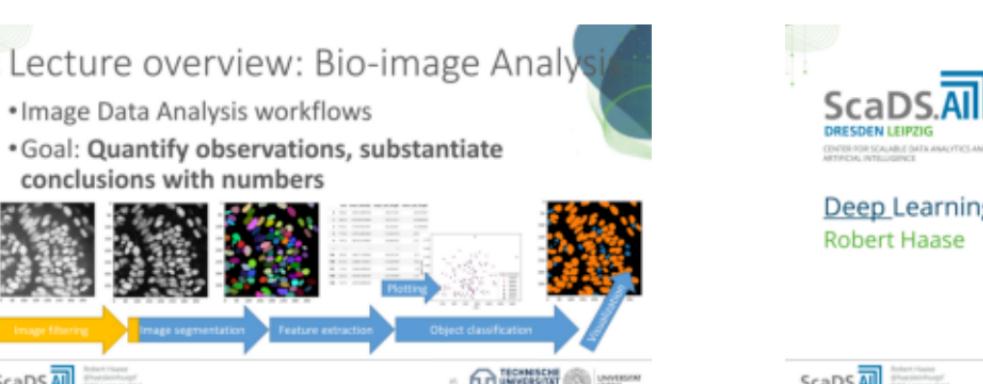
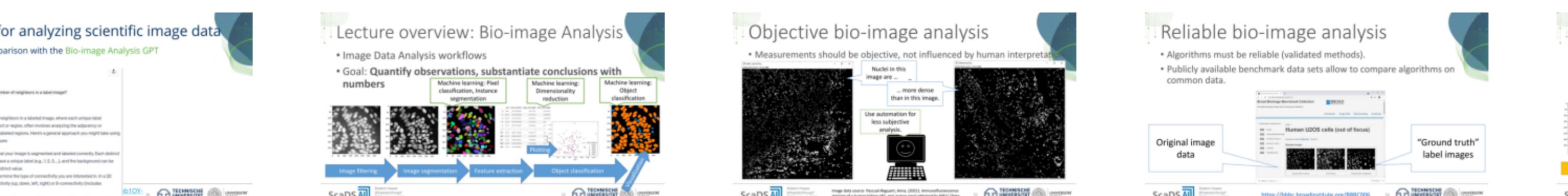
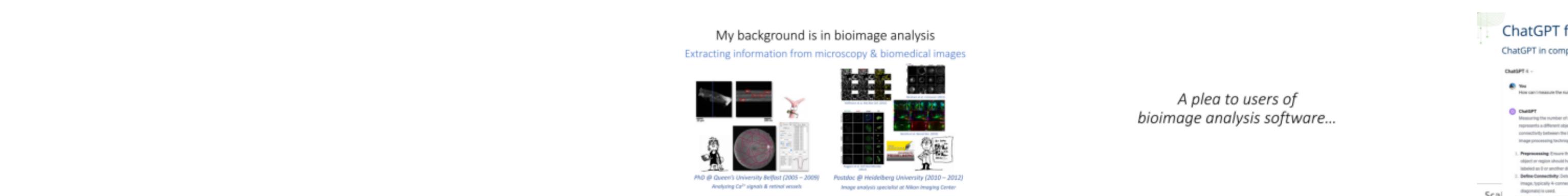
Approach A

The collage consists of 12 distinct screenshots arranged in a grid:

- Other points to think about:** A slide with five yellow circles containing question marks and corresponding text: Vision, Standards, Initiatives, Resources, and AI.
- Artificial intelligence in daily programmer's life:** A Jupyter notebook titled "How it started" showing code and output.
- Quick survey:** A screenshot of a survey interface with a QR code.
- ChatGPT for analyzing scientific image data:** A Jupyter notebook titled "How it started" showing ChatGPT processing an image.
- BiA-Bob: Vision models:** A Jupyter notebook titled "How it started" showing a neural network diagram.
- Training (TA5):** A Jupyter notebook titled "Jupyter Book + Search Index" showing a search interface.
- Supervised and Unsupervised Machine Learning for Bio-image Analysis:** A Jupyter notebook titled "How it's going" showing a GitHub copilot interface.
- ScaDS.AI:** A screenshot of the ScaDS.AI website featuring the logo and navigation menu.
- Supervised and Unsupervised Machine Learning for Bio-image Analysis:** A Jupyter notebook titled "How to use this resource" showing a GitHub repository link.
- ScaDS.AI:** A screenshot of the ScaDS.AI website featuring the logo and navigation menu.
- Supervised and Unsupervised Machine Learning for Bio-image Analysis:** A Jupyter notebook titled "How to use this resource" showing a GitHub repository link.
- ScaDS.AI:** A screenshot of the ScaDS.AI website featuring the logo and navigation menu.

Query: How is AI used in bioimage analysis?

Approach B



Lecture overview: Bio-image Analysis

Query: What are the benefits of using machine learning in microscopy?

Approach A

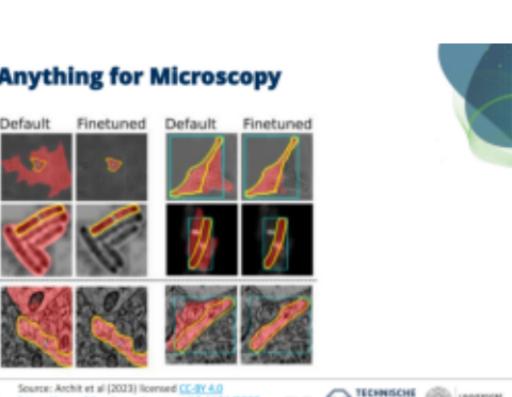
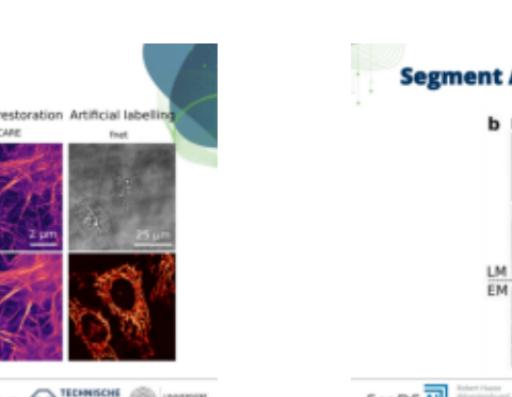
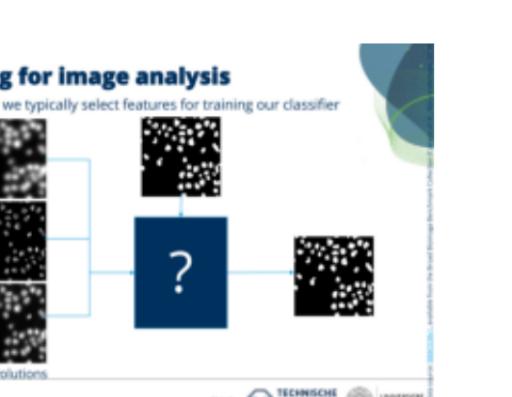
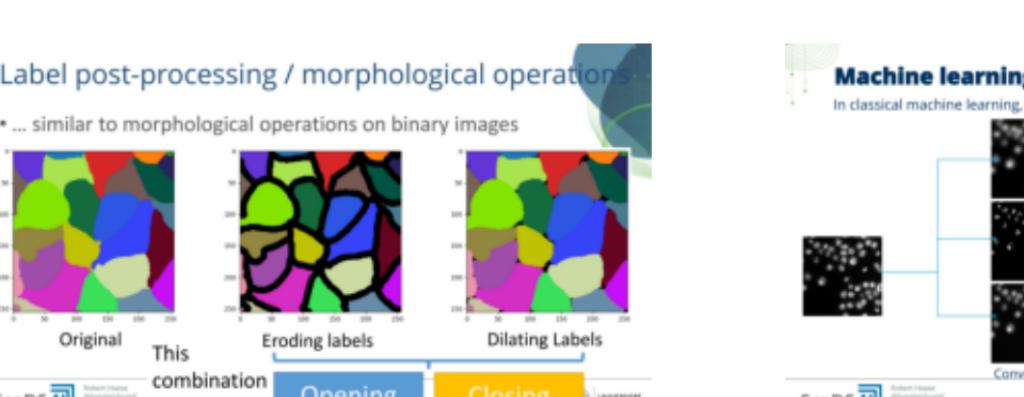
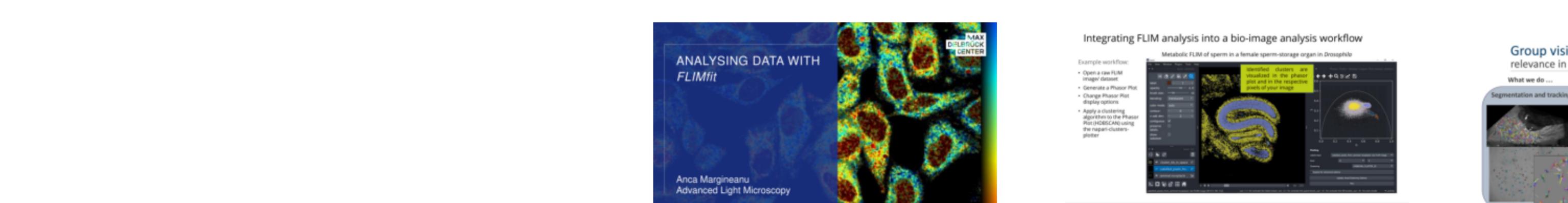
The collage consists of ten screenshots arranged in two rows of five. The top row includes:

- "Where to share?" showing bioRxiv manuscripts and F1000 reviews.
- "Deep Learning for Microscopy" showing segmentation, denoising, and image restoration tools.
- "Machine learning" showing a circular diagram of Artificial Intelligence, Machine learning, and Deep Learning, along with a timeline of events.
- "Licensing: Permissive versus restrictive" comparing CC-BY 4.0 and other licenses.
- "Preliminary schedule" showing a timeline from April 2nd to July 2nd, 2024.

The bottom row includes:

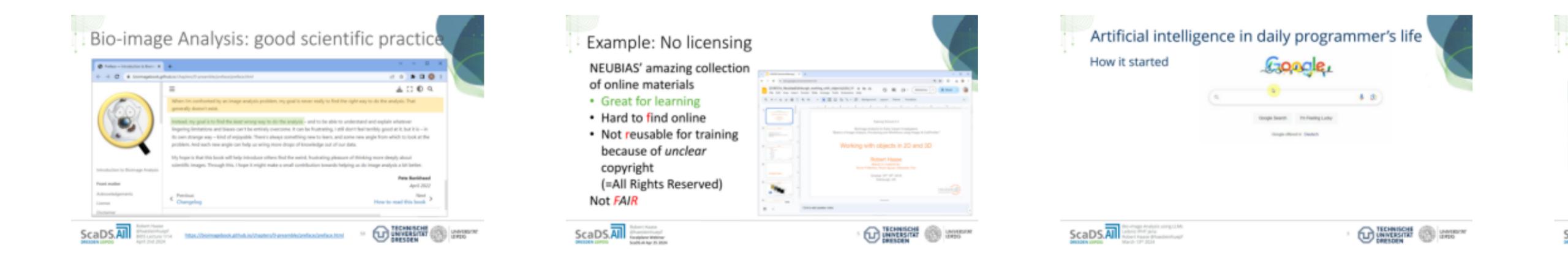
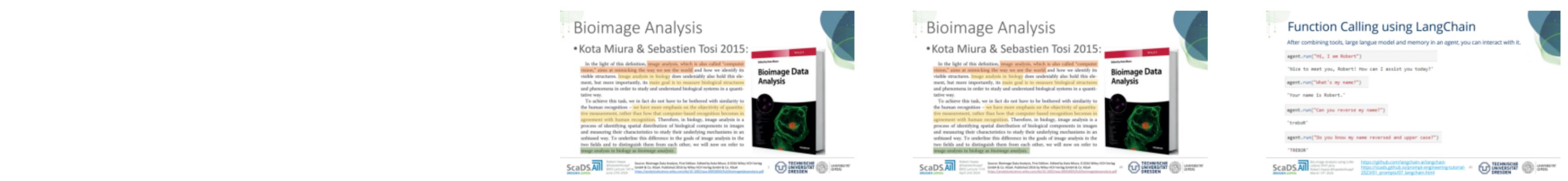
- "Machine learning" showing a timeline from April 9th to June 25th, 2024.
- "Licensing: Permissive versus restrictive" showing a timeline from April 15th to June 25th, 2024.
- "Machine learning" showing a timeline from April 15th to June 25th, 2024.
- "Where to share?" showing bioRxiv manuscripts and F1000 reviews.
- "Group vision: from images to insight and clinical relevance in collaboration with life scientists" showing a timeline from April 15th to June 25th, 2024.

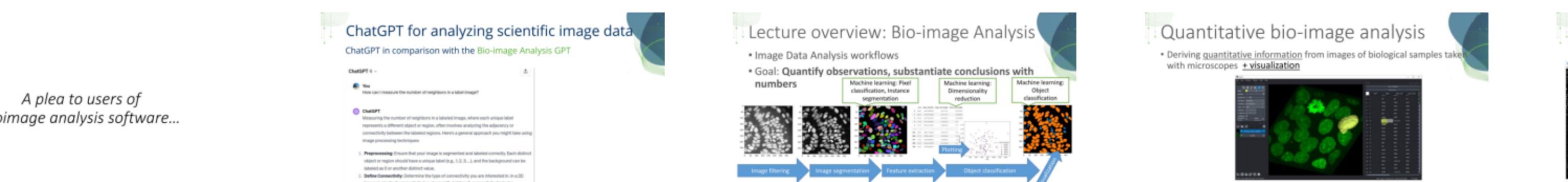
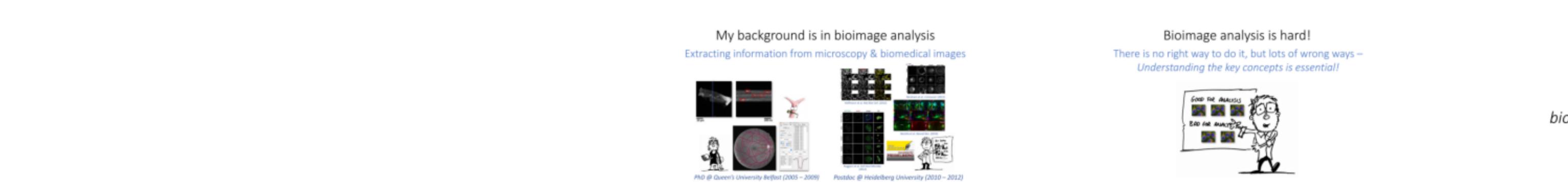
Query: What are the benefits of using machine learning in microscopy?
Approach B



Query: What is bioimage analysis?

Approach A





Query: What tools are commonly used for image analysis in bioimaging?

Approach A

(S)lide 01 OMERO training material page 7 of 14

Subchapter 07.3 introduces the concept of ontologies. Experience shows that many beginners in OMERO are not yet familiar with ontologies. OMERO has no strict rules about how to use ontologies. Here we provide a suggestion in the context of bioimaging. We mainly relate to the Recommended Metadata for Biological Images (REMBI) [33]. For the annotation of ontology-derived terms in Key-Value Pairs, we suggest combining REMBI items in the Keys with ontology-derived natural language terms as values when adequate based on previously proposed standards (e.g.: [SA](#), [IRB](#), [video link](#): <https://zenodo.3604042/SA.pdf>)

(S)lide 02 Using OMERO together with Fiji (video link: <https://zenodo.3604042/FyO2FkG0D9>)

Fiji(Fiji) [18, 20] is the most widely used image analysis software for researchers using bioimaging [7, 21]. Many other, open as well as proprietary, image analysis software can be used in combination with OMERO. However, here we focus on the connection of Fiji and OMERO to showcase the immediate benefit of using OMERO as a data organization when working with the data ('not data'). The Fiji plugin for OMERO is introduced, and we point to community-developed FijiImageJ macro extensions for OMERO [22].

Chapter 08: More features, functions, and ways to use OMERO (no video)

Having focused on absolute beginners for the use of OMERO as the primary target group, here we provide links and information for the many additional features that OMERO offers to advanced users. Where available, we point to online tutorials and training material and/or to original publications. Chapter 9 is not presented as a video tutorial.

How to use this material

The intended use of our material is to download the slideshows, change the layout as required. Notes for presenters may give hints on which slides should likely be adjusted. Yellow markings highlight where slide content requires manual changes. The content including the figures and text (but excluding logos) is – unless marked otherwise – published with a CC-BY 4.0 license as an open educational resource. We intend to provide FAIR training material by versioning in open formats with permissive license accompanying metadata, and versioning [23]. Suggestions for slide improvements may be communicated to the authors directly for adoption in updated versions of the material. Contributions may be addressed to bioimaging@zenodo.org. Furthermore, if you use the material at your own institution, we would greatly appreciate being notified to get in touch for feedback over time.

OMERO-interoperable analysis tools

OMERO and image analysis software or platforms (examples)

- Fiji (Chapter 8) <https://github.com/quewha/fijopath-extension-omero>
- QuPath <https://github.com/quewha/fijopath-extension-omero>
- Galaxy Imaging https://usegalaxy.eu/github/reports/2020/11/23/OMERO_galaxy
- Napari <https://www.napari-hub.org/plugin/napari-omero>
- BioImage IT <https://doi.org/10.1371/journal.pone.1081174>
- Jupyter Notebook <https://omero-guides.readthedocs.io/en/latest/cellprofiler/cellprofiler.html>
- & Cell Profiler: <https://omero-guides.readthedocs.io/en/latest/cellprofiler/cellprofiler.html>

See more: https://omero-guides.readthedocs.io/en/latest/external_tools.html

ADD LOGO SMALL

Quick survey

Question

Who has tried ChatGPT, Dall-E or other large language model based tools before?

Who is using them on a daily basis?

Who knows how they work under the hood (roughly)?

Number of raised hands

<https://doi.org/10.5281/zenodo.10811747>

NFDI4BIOIMAGE Training Portfolio

LangChain is used to combine tools.

It uses chatGPT under the hood.

Building applications with OMERO through compatibility +

Function Calling using LangChain

Q: Which tools do you use in your daily work life?

A Training Portfolio is being established in the consortium

FAIR data life cycle as applied to bioimaging

DMPs for microscopy data sets

How to publish bioimaging data sets

Bioimage data repositories

OMERO as a research data management platform

Microscopy file formats: OMExaml as a cloud-ready file format for large N-dimensional arrays

Licensing of material

Platform for bioimage management

Electronic lab notebook

Data repository

Measurements / feature extraction

Statistics / tabular data wrangling

Advanced workflows / big data

Unit-test pass-rate (n=10)

combine_columns_of_tables	create_map	measure_intensity_over_line	measure_properties_of_regions	count_number_of_touching_neighbors	stack_image_processing
1.0	0.8	1.0	0.9	0.7	0.1
1.0	1.0	1.0	0.9	1.0	0.3
1.0	1.0	1.0	1.0	1.0	0.3
1.0	1.0	1.0	1.0	1.0	0.3
1.0	1.0	1.0	1.0	1.0	0.3
1.0	1.0	1.0	1.0	1.0	0.3
1.0	1.0	1.0	1.0	1.0	0.3
1.0	1.0	1.0	1.0	1.0	0.3
1.0	1.0	1.0	1.0	1.0	0.3
1.0	1.0	1.0	1.0	1.0	0.3

How to use this material

After combining tools, large language model and memory in an agent, you can interact with it.

Allowing others to do your experiment again.

“nice to meet you, Robert! How can I assist you today?”

“What’s my name?”

“Your name is Robert.”

“Can you reverse my name?”

“reverse”

“agent.run(“Do you know my name reversed and uppercase?”)

“TREBON”

Benchmarking LLMs for Bio-image Analysis

RDM CAMPUS SURVEY 2023

How to use this material

Unit-test pass-rate (n=10)

combine_columns_of_tables	create_map	measure_intensity_over_line	measure_properties_of_regions	count_number_of_touching_neighbors	stack_image_processing
1.0	0.8	1.0	0.9	0.7	0.1
1.0	1.0	1.0	0.9	1.0	0.3
1.0	1.0	1.0	1.0	1.0	0.3
1.0	1.0	1.0	1.0	1.0	0.3
1.0	1.0	1.0	1.0	1.0	0.3
1.0	1.0	1.0	1.0	1.0	0.3
1.0	1.0	1.0	1.0	1.0	0.3
1.0	1.0	1.0	1.0	1.0	0.3
1.0	1.0	1.0	1.0	1.0	0.3
1.0	1.0	1.0	1.0	1.0	0.3

Reproducible bio-image analysis

A fabulous ecosystem of open source bioimage analysis software makes this possible...

How to use this material

The intended use of our material is to download the slideshows, change the layout as required. Notes for presenters may give hints on which slides should likely be adjusted. Yellow markings highlight where slide content requires manual changes. The content including the figures and text (but excluding logos) is – unless marked otherwise – published with a CC-BY 4.0 license as an open educational resource. We intend to provide FAIR training material by versioning in open formats with permissive license accompanying metadata, and versioning [23]. Suggestions for slide improvements may be communicated to the authors directly for adoption in updated versions of the material. Contributions may be addressed to bioimaging@zenodo.org. Furthermore, if you use the material at your own institution, we would greatly appreciate being notified to get in touch for feedback over time.

Function Calling using LangChain

After combining tools, large language model and memory in an agent, you can interact with it.

“nice to meet you, Robert! How can I assist you today?”

“What’s my name?”

“Your name is Robert.”

“Can you reverse my name?”

“reverse”

“agent.run(“Do you know my name reversed and uppercase?”)

“TREBON”

Data From Survey on Research Data Management on CMCI/PoL Campus, October 2023

ScaDS ALL Bioimage analysis using LangChain (bioimage-analysis-with-langchain) <https://zenodo.3604042/bioimage-analysis-with-langchain>

Scalable Agent-based Workflow for Bioimage Analysis (scalable-agent-based-workflow-for-bioimage-analysis) <https://zenodo.3604042/scalable-agent-based-workflow-for-bioimage-analysis>

Open Access Bioimage Analysis Using LangChain (open-access-bioimage-analysis-using-langchain) <https://zenodo.3604042/open-access-bioimage-analysis-using-langchain>

Open Access Bioimage Analysis Using

Query: What tools are commonly used for image analysis in bioimaging?
Approach B

The collage consists of ten screenshots arranged in two rows. The top row shows:

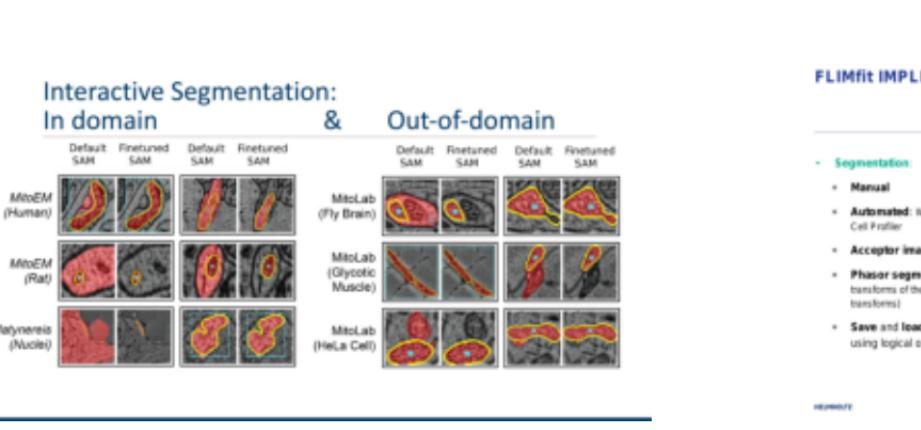
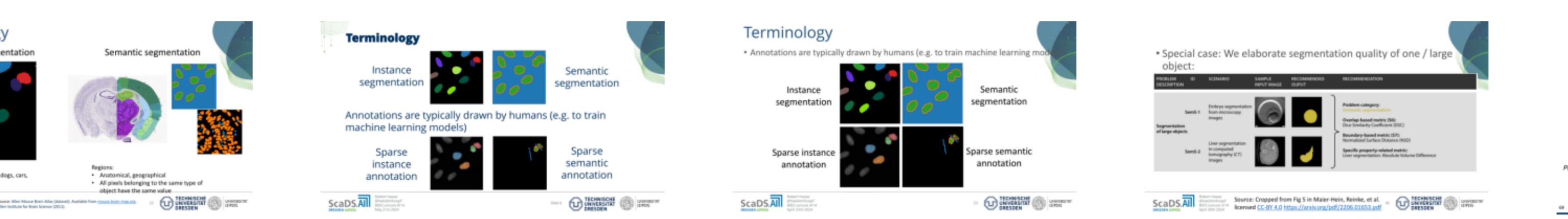
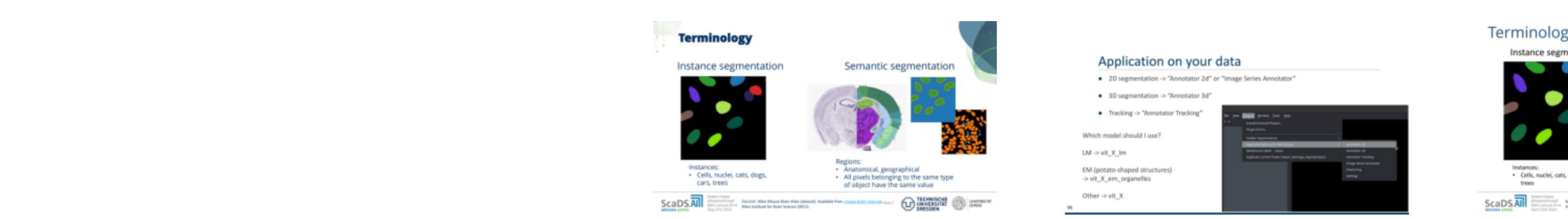
- ChatGPT for analyzing scientific image data**: A screenshot of a ChatGPT interface asking about measuring neighbors in a labeled image.
- Lecture overview: Bio-image Analysis**: A flowchart showing a pipeline from image filtering to object classification, with steps for image segmentation and feature extraction.
- Exercise: Image manipulation**: A screenshot of a software interface with multiple windows showing microscopy images and processing tools.
- High-throughput imaging**: A flowchart showing a pipeline from image filtering to object classification, with steps for image segmentation and feature extraction.
- Bioimage Analysis**: A screenshot of a software interface showing a 3D volume and various analysis tools.

The bottom row shows:

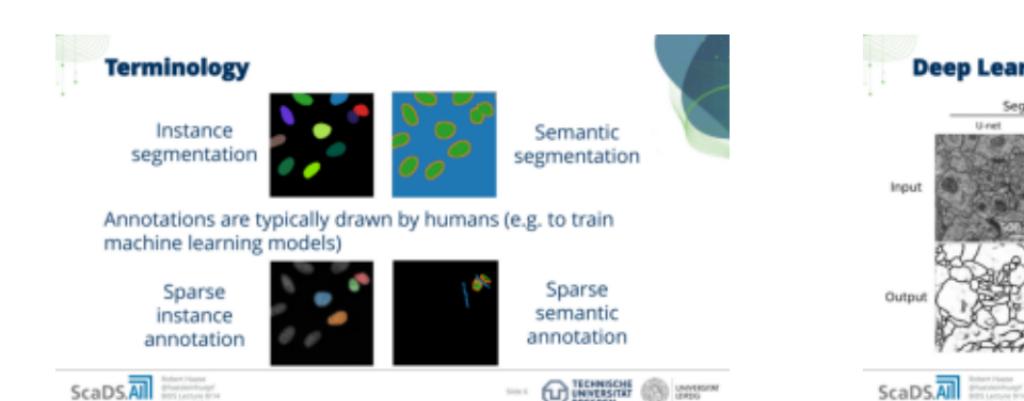
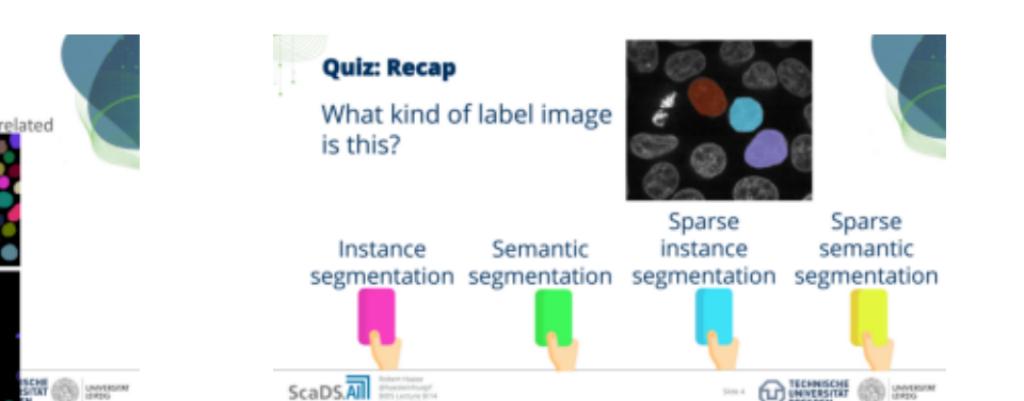
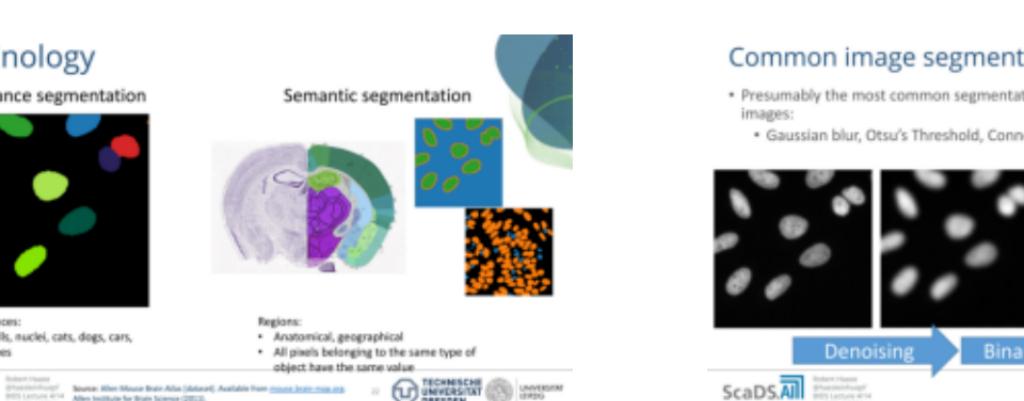
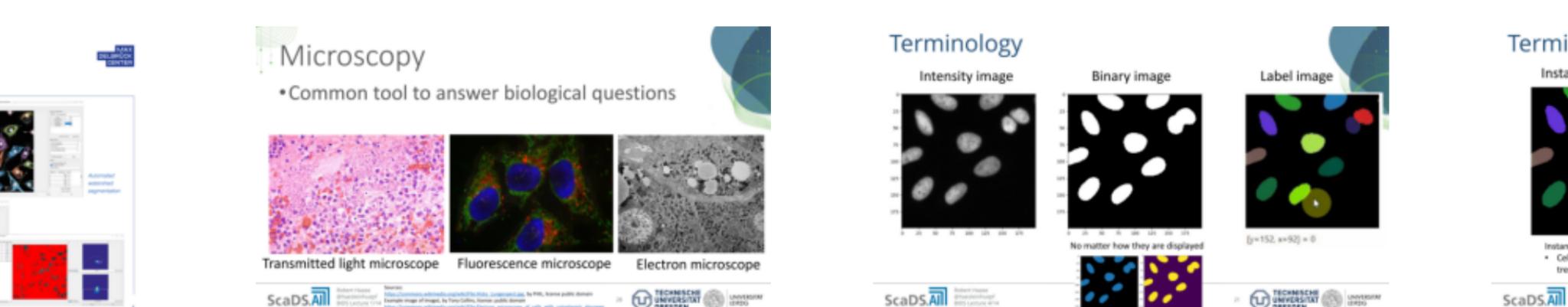
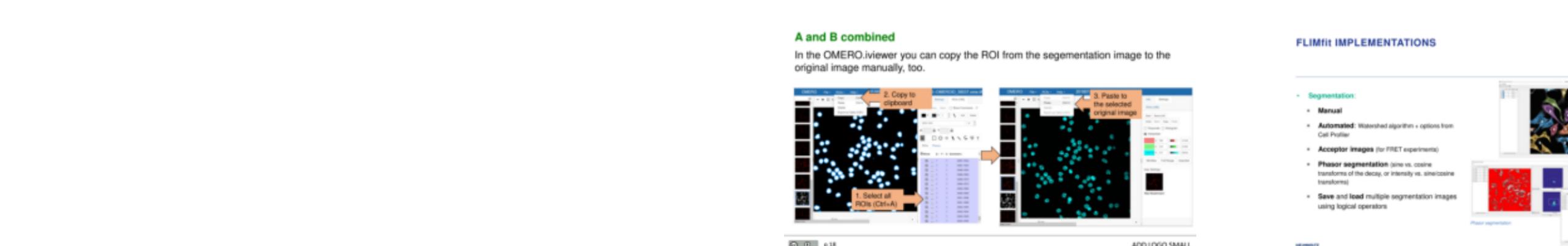
- Extracting information from microscopy & biomedical images**: A screenshot of a software interface showing microscopy images and analysis tools.
- ChatGPT in comparison with the Bio-image Analysis GPT**: A screenshot of a ChatGPT interface comparing its capabilities with Bio-image Analysis GPT.
- Exercises: Prompting image analysis tasks**: A screenshot of a software interface showing a pipeline from image filtering to object classification.
- Lecture overview: Bio-image Analysis**: Another flowchart showing a pipeline from image filtering to object classification, with steps for image segmentation and feature extraction.
- Deep Learning for Bio-image Analysis**: A screenshot of a software interface showing a 3D volume and various analysis tools.

Query: What are different segmentation methods used for microscopy data analysis?

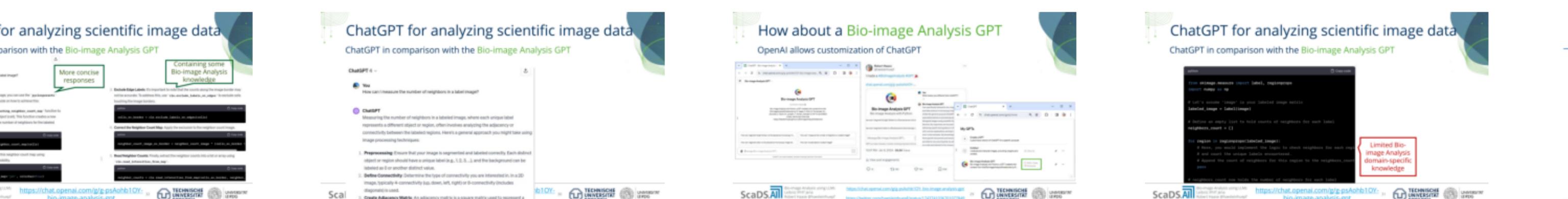
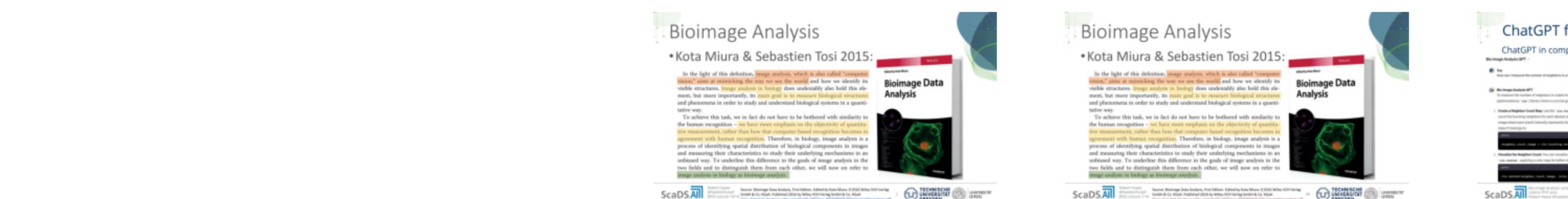
Approach A



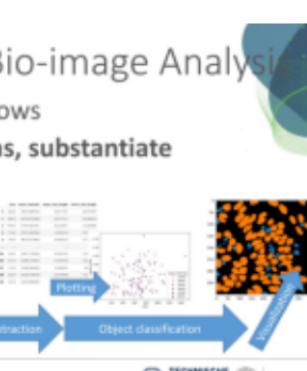
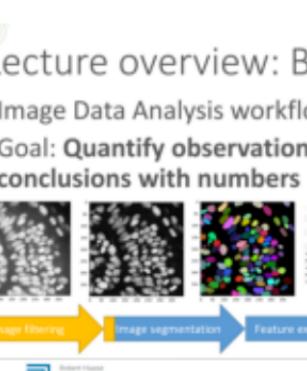
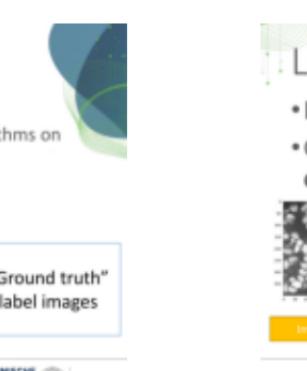
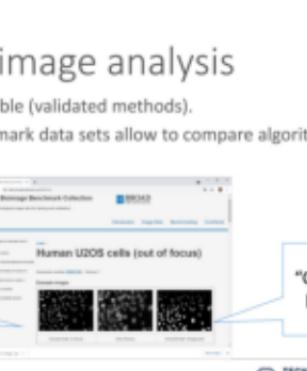
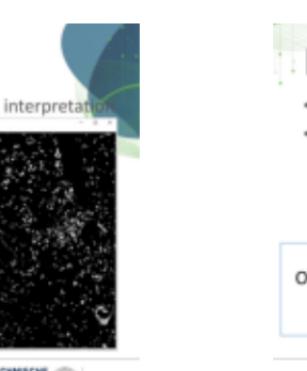
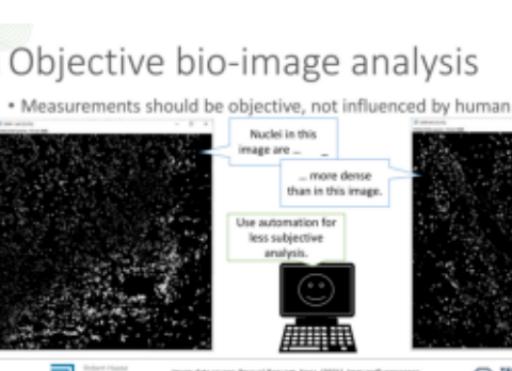
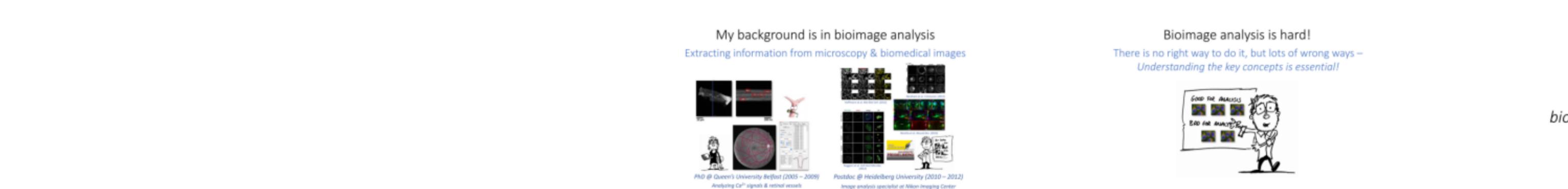
Query: What are different segmentation methods used for microscopy data analysis?
Approach B



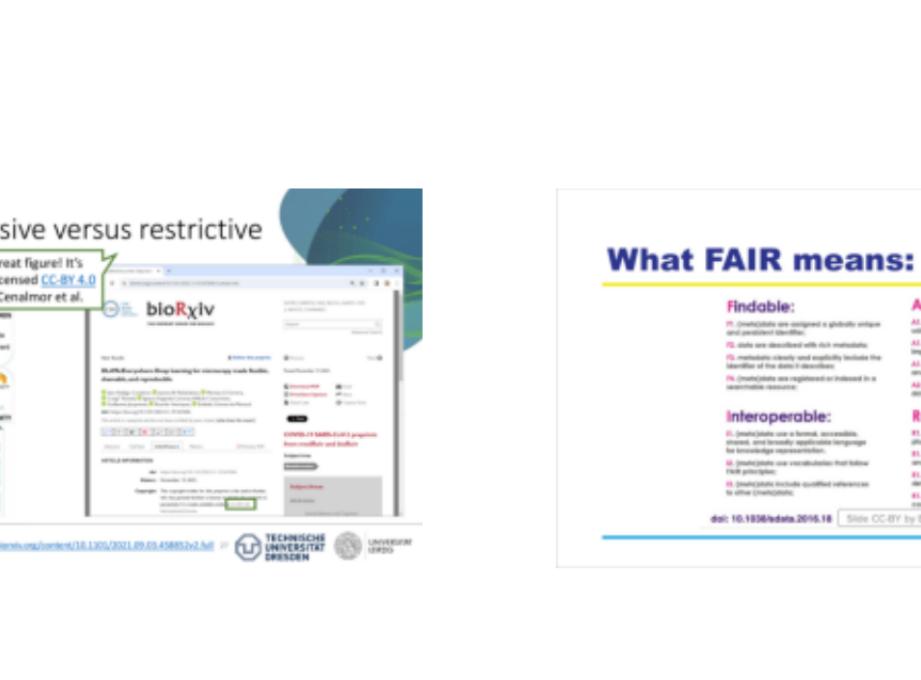
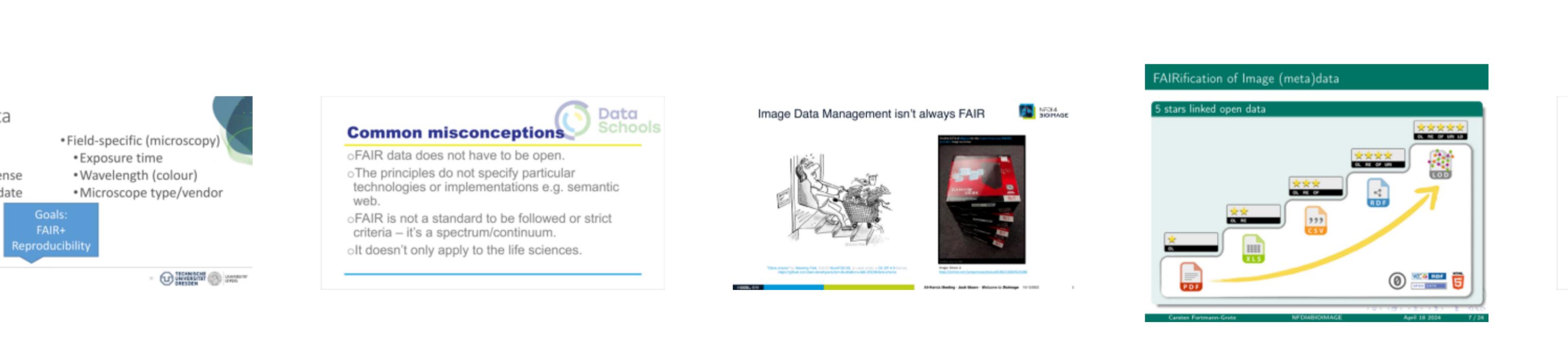
Query: What's the difference between bioinformatics and bioimage analysis?
Approach A



Query: What's the difference between bioinformatics and bioimage analysis?
Approach B

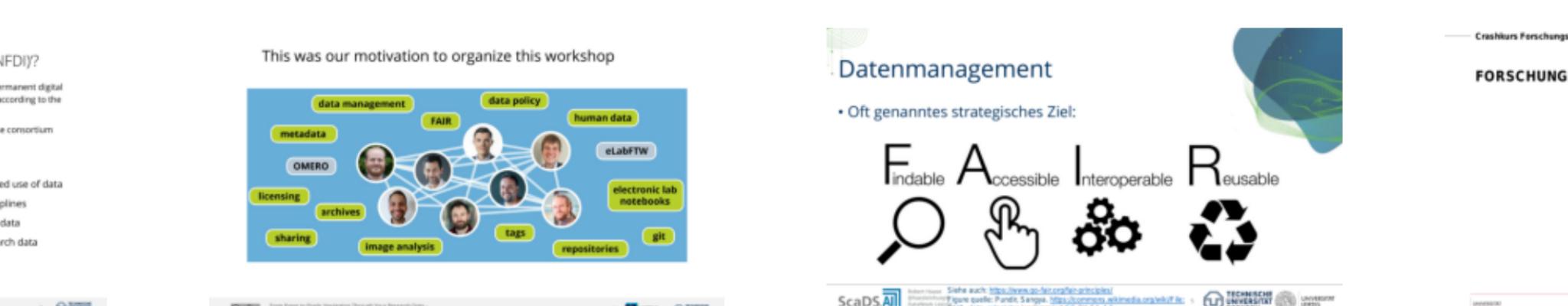


Query: How can I make my microscopy data FAIR? Approach A



Query: What are the FAIR principles in data management?

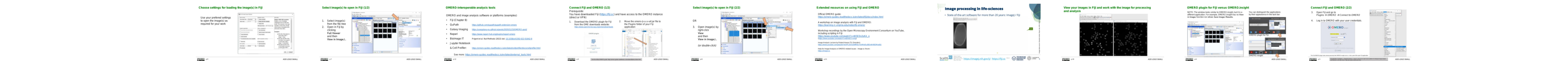
Approach A



Query: What are the FAIR principles in data management?
Approach B

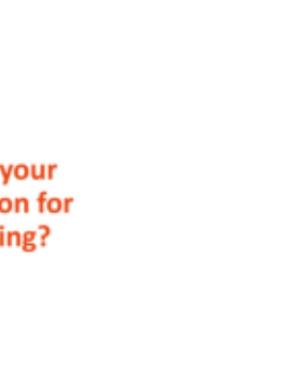
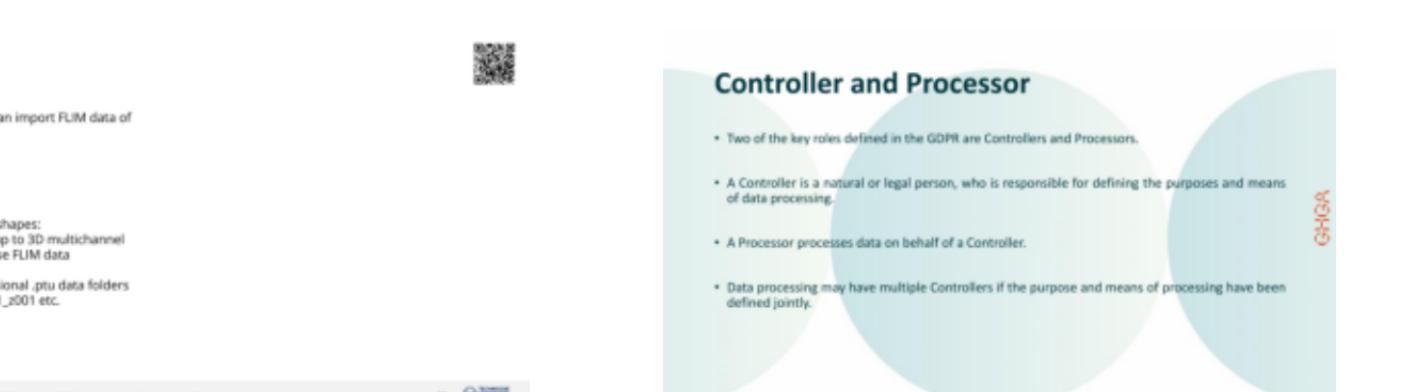
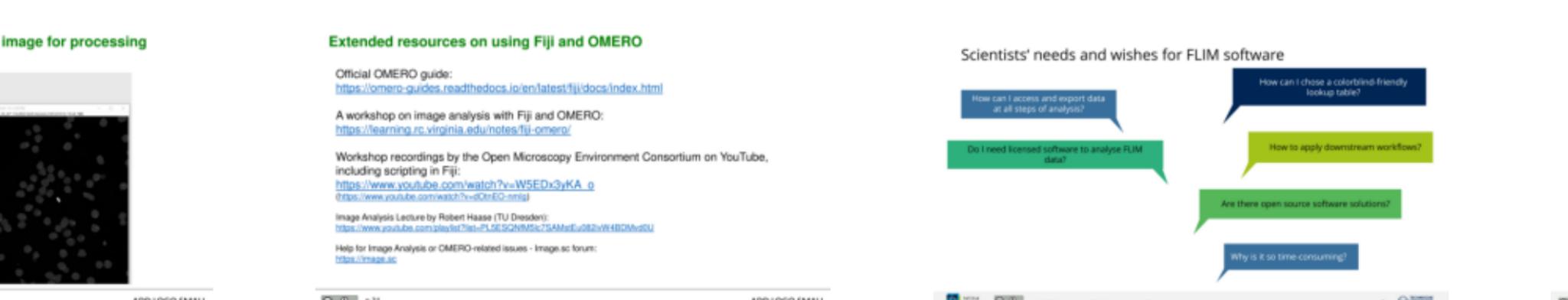
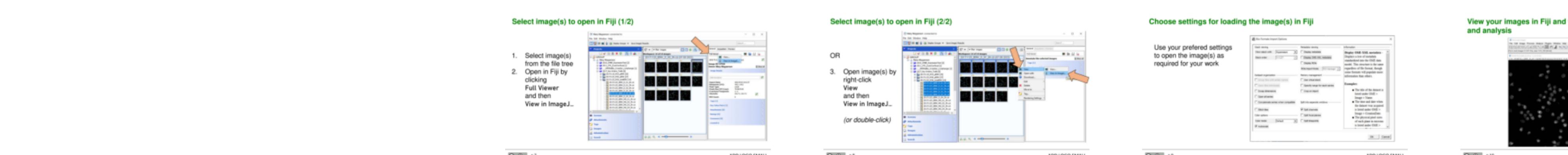
The collage consists of six distinct panels, each featuring a different theme or perspective on the FAIR principles:

- Data Schools:** A slide titled "The FAIR Principles" with a circular diagram showing "Managed data" at the top, "FAIR data" in the center, and "Open data" at the bottom, with arrows indicating relationships between them.
- Increasing that which is FAIR & data Schools:** A slide titled "The FAIR Principles open" with a circular diagram showing "the wild" on the left, "Managed data" at the top, "FAIR data" in the center, and "Open data" at the bottom, with arrows indicating relationships between them.
- Research Data Management:** A slide titled "Research Data Management" with a section on "Interoperable" principles (Logistical necessity, Ensures Integrity, reproducibility and transferability) and a section on "Reusable" principles (I1. (Meta)data are richly described with a plurality of accurate and relevant attributes, I2. (Meta)data are released with a clear and accessible data usage license, I3. (Meta)data use vocabularies that follow FAIR principles, I4. (Meta)data are associated with detailed provenance, I5. (Meta)data include qualified references to other (meta)data).
- The FAIR-principles:** A slide titled "The FAIR-principles" with a section on "Interoperable" principles (I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation, I2. (Meta)data use vocabularies that follow FAIR principles, I3. (Meta)data include qualified references to other (meta)data) and a section on "Reusable" principles (R1. (Meta)data are richly described with a plurality of accurate and relevant attributes, R2. (Meta)data are released with a clear and accessible data usage license, R3. (Meta)data use vocabularies that follow FAIR principles, R4. (Meta)data are associated with detailed provenance, R5. (Meta)data meet domain-relevant community standards).
- Survey:** A slide titled "Survey" with the question "Think about the FAIR principles for data sharing, which one is wrong?" followed by four colored boxes labeled "Findable" (blue), "Accessible" (yellow), "Interoperable" (green), and "Reproducible" (orange).
- Survey:** A slide titled "Survey" with the question "Think about the FAIR principles for data sharing, which one is wrong?" followed by four colored boxes labeled "Findable" (blue), "Accessible" (yellow), "Interoperable" (green), and "Reproducible" (orange).



Query: What is Fiji and how is it used?

Approach B



Query: What is Napari and when should I use it?
Approach A



ScaDS@
DRESDEN
LEIPZIG

Technische Universität
Dresden

Universität Leipzig

Scalable Data Science
Lectures 2024

May 21st 2024

Slide 28

TECHNISCHE
UNIVERSITÄT
DRESDEN

UNIVERSITÄT
LEIPZIG

Scalable Data Science
Lectures 2024

May 21st 2024

Slide 29

TECHNISCHE
UNIVERSITÄT
DRESDEN

UNIVERSITÄT
LEIPZIG

Scalable Data Science
Lectures 2024

May 21st 2024

Slide 30

TECHNISCHE
UNIVERSITÄT
DRESDEN

UNIVERSITÄT
LEIPZIG

Scalable Data Science
Lectures 2024

May 21st 2024

Slide 31

TECHNISCHE
UNIVERSITÄT
DRESDEN

UNIVERSITÄT
LEIPZIG

Scalable Data Science
Lectures 2024

May 21st 2024

Slide 32

TECHNISCHE
UNIVERSITÄT
DRESDEN

UNIVERSITÄT
LEIPZIG

Scalable Data Science
Lectures 2024

May 21st 2024

Slide 33

TECHNISCHE
UNIVERSITÄT
DRESDEN

UNIVERSITÄT
LEIPZIG

Scalable Data Science
Lectures 2024

May 21st 2024

Slide 34

TECHNISCHE
UNIVERSITÄT
DRESDEN

UNIVERSITÄT
LEIPZIG

Scalable Data Science
Lectures 2024

May 21st 2024

Slide 35

TECHNISCHE
UNIVERSITÄT
DRESDEN

UNIVERSITÄT
LEIPZIG

Scalable Data Science
Lectures 2024

May 21st 2024

Slide 36

TECHNISCHE
UNIVERSITÄT
DRESDEN

UNIVERSITÄT
LEIPZIG

Scalable Data Science
Lectures 2024

May 21st 2024

Slide 37

TECHNISCHE
UNIVERSITÄT
DRESDEN

UNIVERSITÄT
LEIPZIG

Scalable Data Science
Lectures 2024

May 21st 2024

Slide 38

TECHNISCHE
UNIVERSITÄT
DRESDEN

UNIVERSITÄT
LEIPZIG

Scalable Data Science
Lectures 2024

May 21st 2024

Slide 39

TECHNISCHE
UNIVERSITÄT
DRESDEN

UNIVERSITÄT
LEIPZIG

Scalable Data Science
Lectures 2024

May 21st 2024

Slide 40

TECHNISCHE
UNIVERSITÄT
DRESDEN

UNIVERSITÄT
LEIPZIG

Scalable Data Science
Lectures 2024

May 21st 2024

Slide 41

TECHNISCHE
UNIVERSITÄT
DRESDEN

UNIVERSITÄT
LEIPZIG

Scalable Data Science
Lectures 2024

May 21st 2024

Slide 42

TECHNISCHE
UNIVERSITÄT
DRESDEN

UNIVERSITÄT
LEIPZIG

Scalable Data Science
Lectures 2024

May 21st 2024

Slide 43

TECHNISCHE
UNIVERSITÄT
DRESDEN

UNIVERSITÄT
LEIPZIG

Scalable Data Science
Lectures 2024

May 21st 2024

Slide 44

TECHNISCHE
UNIVERSITÄT
DRESDEN

UNIVERSITÄT
LEIPZIG

Scalable Data Science
Lectures 2024

May 21st 2024

Slide 45

TECHNISCHE
UNIVERSITÄT
DRESDEN

UNIVERSITÄT
LEIPZIG

Scalable Data Science
Lectures 2024

May 21st 2024

Slide 46

TECHNISCHE
UNIVERSITÄT
DRESDEN

UNIVERSITÄT
LEIPZIG

Scalable Data Science
Lectures 2024

May 21st 2024

Slide 47

TECHNISCHE
UNIVERSITÄT
DRESDEN

UNIVERSITÄT
LEIPZIG

Scalable Data Science
Lectures 2024

May 21st 2024

Slide 48

TECHNISCHE
UNIVERSITÄT
DRESDEN

UNIVERSITÄT
LEIPZIG

Scalable Data Science
Lectures 2024

May 21st 2024

Slide 49

TECHNISCHE
UNIVERSITÄT
DRESDEN

UNIVERSITÄT
LEIPZIG

Scalable Data Science
Lectures 2024

May 21st 2024

Slide 50

TECHNISCHE
UNIVERSITÄT
DRESDEN

UNIVERSITÄT
LEIPZIG

Scalable Data Science
Lectures 2024

May 21st 2024

Slide 51

TECHNISCHE
UNIVERSITÄT
DRESDEN

UNIVERSITÄT
LEIPZIG

Scalable Data Science
Lectures 2024

May 21st 2024

Slide 52

TECHNISCHE
UNIVERSITÄT
DRESDEN

UNIVERSITÄT
LEIPZIG

Scalable Data Science
Lectures 2024

May 21st 2024

Slide 53

TECHNISCHE
UNIVERSITÄT
DRESDEN

UNIVERSITÄT
LEIPZIG

Scalable Data Science
Lectures 2024

May 21st 2024

Slide 54

TECHNISCHE
UNIVERSITÄT
DRESDEN

UNIVERSITÄT
LEIPZIG

Scalable Data Science
Lectures 2024

May 21st 2024

Slide 55

TECHNISCHE
UNIVERSITÄT
DRESDEN

UNIVERSITÄT
LEIPZIG

Scalable Data Science
Lectures 2024

May 21st 2024

Slide 56

TECHNISCHE
UNIVERSITÄT
DRESDEN

UNIVERSITÄT
LEIPZIG

Scalable Data Science
Lectures 2024

May 21st 2024

Slide 57

TECHNISCHE
UNIVERSITÄT
DRESDEN

UNIVERSITÄT
LEIPZIG

Scalable Data Science
Lectures 2024

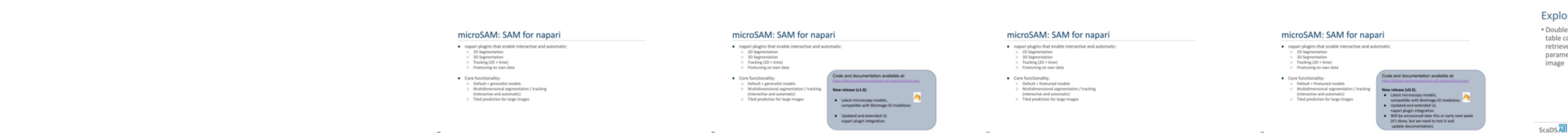
May 21st 2024

Slide 58

<p

Query: What is Napari and when should I use it?

Approach B



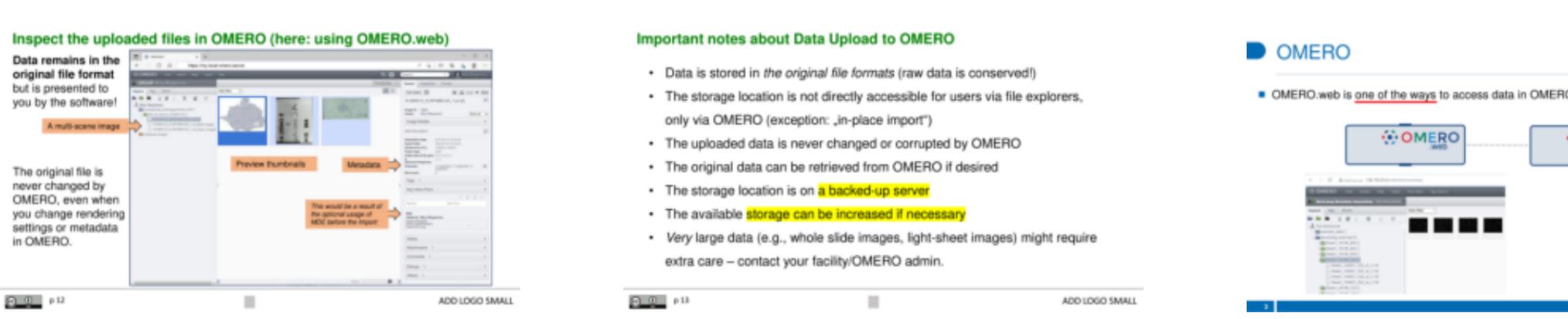
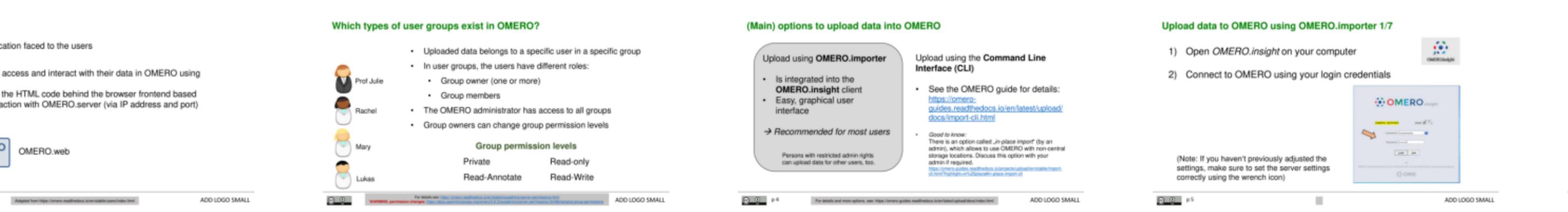
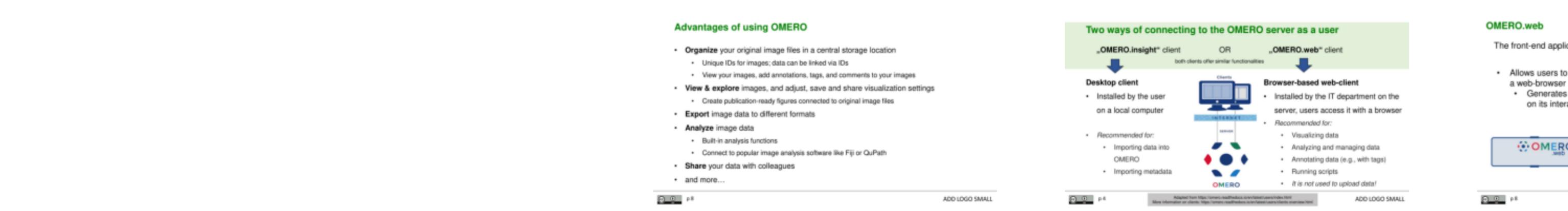
Deconstruction of napari-chatGPT

* Napari-chatGPT defines a list of "tools"

* Double-click on table column to retrieve a parametric map image

Query: How does OMERO help with data storage and sharing?

Approach B



Data from Survey on Research Data Management on MC&PUL Campus, October 2023

Session 3: Managing and sharing images using OMERO ROM CAMPUS SURVEY 2023

hhu

How familiar are you with the usage of tools such as OMERO?

Platforms

1 - not familiar at all
2
3
4
5 - very familiar

10 20 30 40 50

0

ADD LOGO SMALL

p.12

For details see <https://omero.readthedocs.io/en/latest/share/share.html>

14 OMERO

ADD LOGO SMALL

p.13

Added from <https://omero.readthedocs.io/en/latest/share/share.html>

Quiz

• When you shared materials over the internet, which license did you use?

	None	Public Domain	Creative Commons	BSD/GPL/MIT/...

Licensing: Creative Commons (CC)

- Public domain (CC0)
- Attribution International (CC-BY)
- Attribution ShareAlike Int. (CC-BY-SA)
- Attribution Non-Commercial Int. (CC-BY-NC)
- Attribution NoDerivatives Int. (CC-BY-ND)

+ Combinations, e.g. CC-BY-NC-ND

ScraDS AI Robert Hasen
Technische Universität Berlin
Universität Regensburg

ScraDS AI Robert Hasen
Technische Universität Berlin
Universität Regensburg

Management Plans (DMPs)

Define responsibilities and procedures early!

- Imaging / data acquisition
- Data Analysis
- Paper writing
- Attendee + trainer acquisition
- Training material preparation
- Conduct workshop

Are we going to publish data / materials / code?

What license can we use?

Only if procedures are defined early, everyone can follow them.
Licenses are important when assembling materials (> Copyright)
Meta-data might have higher quality if the person responsible for publishing the data is aware of their duties.

Deciding by the end of the project is too late!

Data Management Plans (DMPs)

• Define responsibilities and procedures early!

Open Science

Experiment design → Imaging / data acquisition → Data Analysis → Paper writing

"Data / materials we produce will be published under CC-BY 4.0!"

"Robert will do this by end of 2025!"

Only if procedures are defined early, everyone can follow them.
Licenses are important when assembling materials (> Copyright)
Meta-data might have higher quality if the person responsible for publishing the data is aware of their duties.

Are we going to publish data / materials / code?

What license can we use?

Deciding by the end of the project is too late!

DMP

Permissive versus restrictive

Use a license which

- is compatible to other projects you're collaborating with and
- fits to your needs / role.

Our project → **not ok** → **Their project**

BSD3 or CC-BY licensed project

ok → **GPL or CC-BY-SA licensed project**

e.g.

e.g.

Licensing: Permissive versus restrictive

	Download and share for free	Reuse parts, e.g. Figures	Reuse parts, e.g. in paid training
CC-BY	✓	✓	✓
CC-BY-SA	✓	Only under CC-BY-SA	Only under CC-BY-SA
CC-BY-NC	✓	Only under CC-BY-NC	Only under CC-BY-NC
CC-BY-ND	✓	Only under CC-BY-ND	Only under CC-BY-ND
CC-BY-NC-ND	✓	Only under CC-BY-NC-ND	Only under CC-BY-NC-ND

Bad for the progress of science
In particular in the training context

The first screenshot shows a bioRxiv preprint page for a paper titled "Bringing the Gap Integrating Existing Big Data Techniques into Biological Imaging with DeepLearning". A green callout box contains the text: "I would love to show you a Figure from this paper!". A red line points from this box to a red-bordered area on the page where it says "But I'm not allowed!".

The second screenshot shows a cropped figure from a CC-BY license. It features a green box with the text: "Look at this great figure! It's cropped from / licensed CC-BY 4.0 by W. Lei et al." Below this is a diagram of a network of devices connected to a central cloud icon labeled "Data Management".

The third screenshot is a comparison chart titled "Licensing: Permissive versus restrictive". It lists five types of licenses under two main categories: "Download and share for free" and "Reuse parts, e.g. Figures".

Licenses	Download and share for free	Reuse parts, e.g. Figures
CC-BY	✓	✓
CC-BY-SA	✓ (Only under CC-BY SA)	✓ (Only under CC-BY SA)
CC-BY-NC	✓	✓ (Only under CC-BY NC)
CC-BY-ND	✓	✗
CC-BY-NC-ND	✓	✗

Licensing: Permissive versus restrictive

Example

I would love to show you a Figure from this paper!

But I'm not allowed!

Query: What are different license types that can be used for sharing data?
Approach B

Quiz

- When you shared materials over the internet, which license did you use?

None Public Domain Creative Commons BSD/GPL/MIT/...

Take home message

- If you share material (openly or not)

license it

and it'll be harder to steal it

Example: Restrictive licensing

- If you share material (openly or not)

license it

and it'll be harder to steal it

Take home message

- Can I build a commercial product on the basis of GPL-licensed code?

Quiz (recap)

- If I combine two works licensed CC-BY and CC-BY-SA, what license do I have to use?

Yes No

Survey

Which open-source license might be the least popular in companies?

CC-BY CC-BY-SA CC-BY-ND (not possible)

Do I have to release the code openly for this commercial product?

Yes No

Quiz (recap)

- If I combine two works licensed CC-BY and CC-BY-SA, what license do I have to use?

Yes No

Survey

Which open-source license might be the least popular in companies?

CC-BY-ND-SA CC-BY-SA CC-BY-ND (not possible)

Query: What does open science mean for data sharing?
Approach A

FOSTER Open Science

The slide features a grid of icons representing open science principles: Data Protection, Best Practice, Data Archiving, Open Access Publishing, Peer Review, Data Licensing, Data Reuse, Software & Materials, Imaging & Sharing Research Data, Open Licensing, Data Protection & Ethics, Open Source Software, and Open Access Publishing.

Open Science, Sharing & Licensing
Robert Haase

The slide illustrates the research data management process: Experiment design, Data / materials, Imaging / data acquisition, Data Analysis, and Paper writing. It also highlights the use of various platforms: Slides, Code, Text, Data, and ...

Collaborative platforms and third-party tools

The slide shows the research data management process: Experiment design, Data / materials, Imaging / data acquisition, Data Analysis, and Paper writing. It lists platforms: OSF, Figshare, F1000, Github, Zenodo, Focalplane, and Institutional servers (e.g., Dropbox, G Drive, or open source e.g., ownCloud).

Open Science
• Define responsibilities and procedures early!

The slide emphasizes defining responsibilities and procedures early. It lists: Vorgaben von Forschungsförderern erfüllen, Often tailored towards general audience (science communication), Transfer of domain-specific knowledge, and Earliest at the time a manuscript is published (e.g. as preprint).

Data Management Plans (DMPs)
• Define responsibilities and procedures early!

The slide emphasizes defining responsibilities and procedures early. It lists: Only if procedures are defined early, everyone can follow them, Licenses are important when assembling materials (> Copyright), Meta-data might have higher quality if the person responsible for publishing the data is aware of their duties, and Deciding by the end of the project is too late!

Open Training
• Routine tasks (colder topics)

The slide emphasizes routine tasks. It lists: Open science related content, OSF - open platform for sharing data in a secure place with fellow researchers and others in a secure environment, Often tailored towards general audience (science communication), Third-party - commercial (e.g. Dropbox, G Drive, or open source (e.g. ownCloud)), and Open Science: neue Erkenntnisse und wissenschaftliche Kooperationen fördern.

Lizenierung
• Session gestern @ DataWeek Leipzig

The slide emphasizes licensing. It lists: Data / materials we produce will be published under CC-BY 4.0, Only if procedures are defined early, everyone can follow them, Licenses are important when assembling materials (> Copyright), Meta-data might have higher quality if the person responsible for publishing the data is aware of their duties, and Deciding by the end of the project is too late!

Query: What does open science mean for data sharing? Approach B

FOSTER Open Science

What is „metadata“?

In general, metadata is often called „**data about the data**“.

In bioimaging, metadata accompanies the actual image (pixel) data:

- Technical metadata (*automatically recorded or added manually*)
 - Information about the instrument, including hardware components, filter settings, etc.
- Sample metadata (*researcher's documentation or published protocol*)
 - Information about the specimen, organ, cell type, sample type, test group, etc., and the experimental procedures during sample preparation, e.g., sample fixation, staining, use of antibodies, etc.
- Analysis metadata

Where to find the metadata?

Image file header Sidecar / additional file Electronic labbook Paper notebook In the data organization ...

Image file header Sidecar / additional file Electronic labbook Paper notebook In the data organization ...

ADD LOGO SMALL

What is „metadata“?

In general, metadata is often called „**data about the data**“.

In bioimaging, metadata accompanies the actual image (pixel) data:

- Technical metadata (*automatically recorded or added manually*)
 - Information about the instrument, including hardware components, filter settings, etc.
- Sample metadata (*researcher's documentation or published protocol*)
 - Information about the specimen, organ, cell type, sample type, test group, etc., and the experimental procedures during sample preparation, e.g., sample fixation, staining, use of antibodies, etc.
- Analysis metadata

Where to find the metadata?

Image file header Sidecar / additional file Electronic labbook Paper notebook In the data organization ...

Image file header Sidecar / additional file Electronic labbook Paper notebook In the data organization ...

ADD LOGO SMALL

Closed science

- Research related (hot topics)
- Routine tasks (colder topics)

Open Science

- Why are some science-related materials/code not shared?

Closed science

- Risk of being scooped
- Often tailored towards general audience (imposter syndrome)
- Transfer of domain-specific knowledge

Open Training

- Often tailored towards general audience (science communication)
- Lack of awareness (who is allowed to publish my work?)
- Earliest at the time a manuscript is published (e.g. as preprint)
- Assumption: it's not worth the effort.

Closed science

- Transfer of domain-specific knowledge

Open Training

- Assumption: it's not worth the effort.

Quiz

• What is the role of the OpenData Portal of Leipzig in the context of publishing data?

Open Science, Sharing & Licensing

Robert Haase

Open Science, Sharing & Licensing

Robert Haase

GHGA

Do you have personal data?

Closed science

- Research related (hot topics)
- Routine tasks (colder topics)

Open Science

- Why are some science-related materials/code not shared?

Closed science

- Risk of being scooped
- Often tailored towards general audience (imposter syndrome)
- Transfer of domain-specific knowledge

Open Training

- Often tailored towards general audience (science communication)
- Lack of awareness (who is allowed to publish my work?)
- Earliest at the time a manuscript is published (e.g. as preprint)
- Assumption: it's not worth the effort.

Closed science

- Transfer of domain-specific knowledge

Open Training

- Assumption: it's not worth the effort.

Quiz

• What is the role of the OpenData Portal of Leipzig in the context of publishing data?

Open Science, Sharing & Licensing

Robert Haase

Open Science, Sharing & Licensing

Robert Haase

GHGA

Do you have personal data?

These slides can be reused under the terms of the CC-BY 4.0 license unless mentioned otherwise.
<https://doi.org/10.5281/recode.1096230>

These slides can be reused under the terms of the CC-BY 4.0 license unless mentioned otherwise.
<https://doi.org/10.5281/recode.1096230>

Query: Why is open source software important in scientific research?

Approach A

Crashkurs Forschungsdatenmanagement

WARUM FORSCHUNGSDATEN VERÖFFENTLICHEN?

- Vorgaben von Forschungsförderern erfüllen
- Anderen Wissenschaftler*innen die Arbeit mit qualitativ hochwertigen Daten ermöglichen: Veröffentlichen von besonders aufwändig erhobenen, einzigartigen Forschungsdaten
- Open Science: neue Erkenntnisse und wissenschaftliche Kooperationen fördern

Types of data

- Openly accessible data
 - „open data“
 - „open source“ software
- Business data
- Research data
 - Hot / cold
- Personal data
- Secret data

In need of protection (schutzbedürftig)



ScaDS.ARI ScaDS.ARI Institut für Arbeitsweltwissenschaften April 2024

Types of data

- Openly accessible data
 - „open data“
 - „open source“ software
- Business data
- Research data
 - Hot / cold
- Personal data
- Secret data

In need of protection (schutzbedürftig)



FOSTER Open Science Data Schools

WOZU FORSCHUNGSDATENMANAGEMENT?

- OSF - open platform for sharing data in all steps of analysis
- Glaubwürdigkeit, Nachvollziehbarkeit der eigenen Forschung durch Dritte
- Reproduzierbarkeit von Forschungsergebnissen
- Erfüllen der Vorgaben von Forschungsförderern und Verlagen
- Minimieren des Datenverlustrisikos
- Vereinfachung zukünftiger Nachnutzung der eigenen Daten und Interpretierbarkeit der Daten langfristig sicherstellen
- Unterstützung von Open Science, ermöglichen neuer Erkenntnisse, Metaanalysen und Kooperationen

Collaborative platforms and third-party tools

- Git server an der Universität
- Zenodo.org
- Github.com
- Firmen- / Instituts-Website

In need of protection (schutzbedürftig)



ownCloud ownCloud

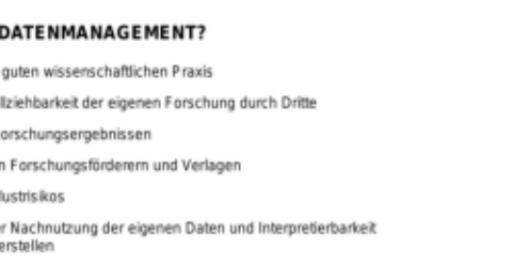
Quiz

Open Science **Open Training**

- Wo ist Open-Source Software am besten sichtbar?

Scientists' needs and wishes for FLIM software

- How can I access and export data?
- How can I chose a colorblind-friendly look&feel?
- Do I need learned software to analyze FLIM data?
- How to apply downstream workflow?
- Are there open source software solutions?
- Why is it so time consuming?



SCIENTIFIC REPORTS **QuPath** exists to fill this gap: Open source software for whole slide analysis (and more)

Open Science

- Research related (hot topics)
- Often tailored towards general audience (science communication)
- Transfer of domain-specific knowledge
- Earliest at the time a manuscript is published (e.g. as preprint)

Open Training

- Routine tasks (colder topics)



ownCloud ownCloud

Open Science

- Grundvoraussetzung der guten wissenschaftlichen Praxis
- Glaubwürdigkeit, Nachvollziehbarkeit der eigenen Forschung durch Dritte
- Reproduzierbarkeit von Forschungsergebnissen
- Erfüllen der Vorgaben von Forschungsförderern und Verlagen
- Minimieren des Datenverlustrisikos
- Vereinfachung zukünftiger Nachnutzung der eigenen Daten und Interpretierbarkeit der Daten langfristig sicherstellen
- Unterstützung von Open Science, ermöglichen neuer Erkenntnisse, Metaanalysen und Kooperationen

Open Training

- Routine tasks (colder topics)



ownCloud ownCloud

Open Science

- Wo ist Open-Source Software am besten sichtbar?

WOZU FORSCHUNGSDATENMANAGEMENT?

- OSF - open platform for sharing data in all steps of analysis
- Glaubwürdigkeit, Nachvollziehbarkeit der eigenen Forschung durch Dritte
- Reproduzierbarkeit von Forschungsergebnissen
- Erfüllen der Vorgaben von Forschungsförderern und Verlagen
- Minimieren des Datenverlustrisikos
- Vereinfachung zukünftiger Nachnutzung der eigenen Daten und Interpretierbarkeit der Daten langfristig sicherstellen
- Unterstützung von Open Science, ermöglichen neuer Erkenntnisse, Metaanalysen und Kooperationen

Collaborative platforms and third-party tools

- Git server an der Universität
- Zenodo.org
- Github.com
- Firmen- / Instituts-Website

In need of protection (schutzbedürftig)



SCIENTIFIC REPORTS **QuPath** exists to fill this gap: Open source software for whole slide analysis (and more)

Open Science

- Research related (hot topics)
- Often tailored towards general audience (science communication)
- Transfer of domain-specific knowledge
- Earliest at the time a manuscript is published (e.g. as preprint)

Open Training

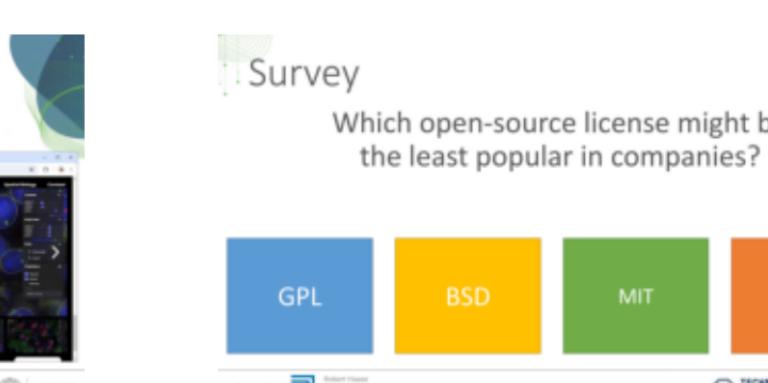
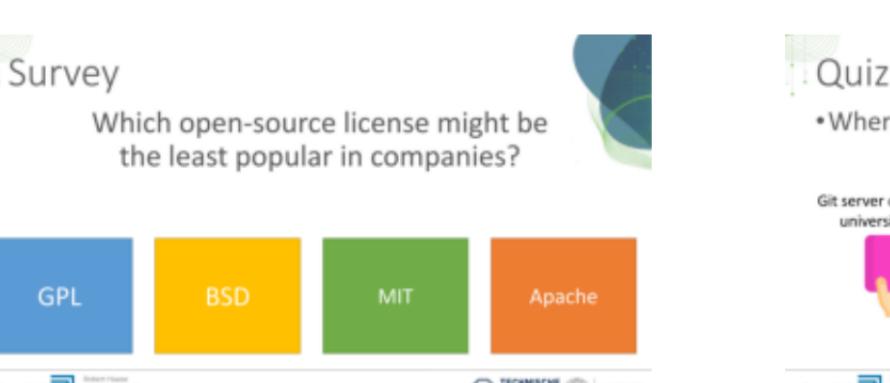
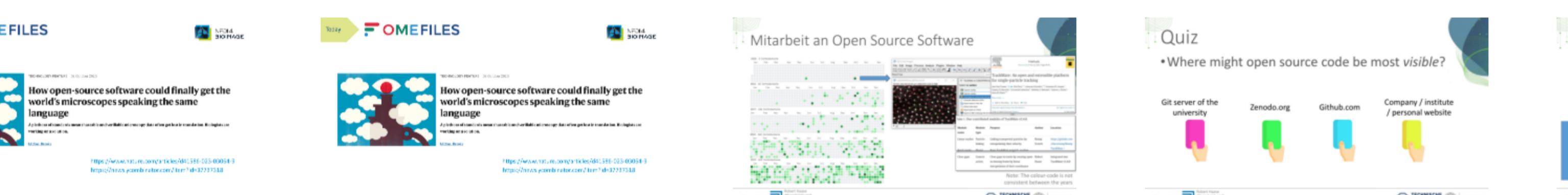
- Routine tasks (colder topics)



ownCloud ownCloud

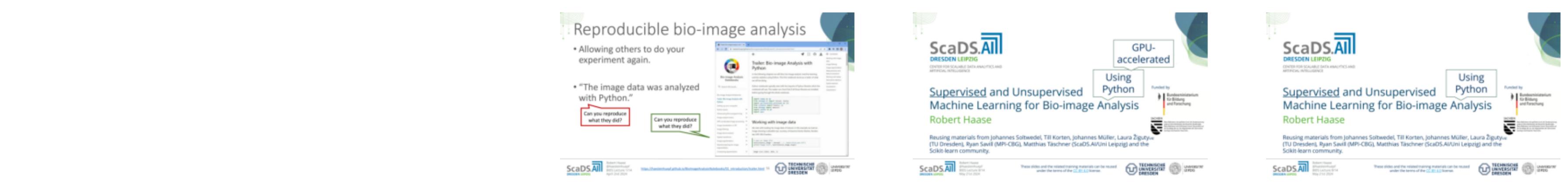
Query: Why is open source software important in scientific research?

Approach B



Query: How can I use Python for bioimage analysis?

Approach A



ScaDS.AI
DRESDEN LEIPZIG
CENTER FOR SCALABLE DATA ANALYTICS AND
ARTIFICIAL INTELLIGENCE

Supervised and Unsupervised Machine Learning for Bio-image Analysis
Robert Haase

GPU-accelerated
Using Python

Bio-image Analysis GPT

ChatGPT for analyzing scientific image data

Prompt engineering

ChatGPT for analyzing scientific image data

Prompt engineering

ChatGPT for analyzing scientific image data

Summary: Deep Learning for Bio-image Analysis

More resources

ChatGPT can solve simple image analysis tasks

OpenAI allows customization of ChatGPT

Just submit a pull-request to the Bio-image Analysis Notebooks

ChatGPT does not know how to do this and hallucinates code that does not work

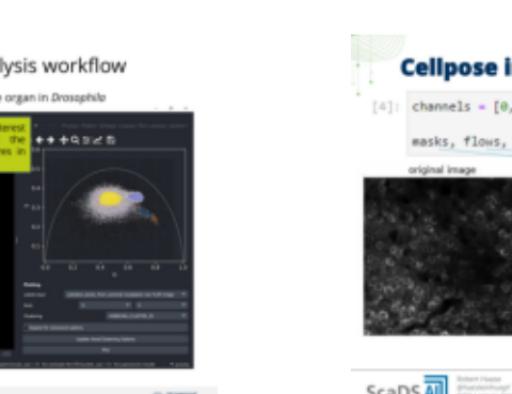
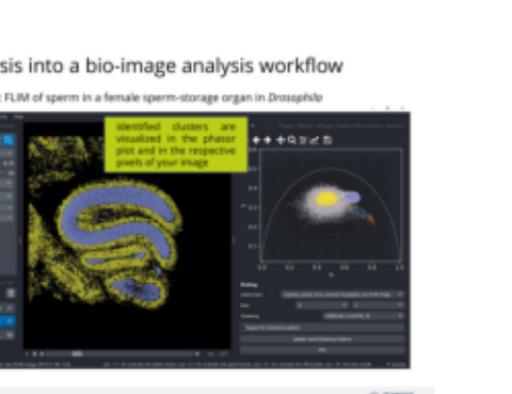
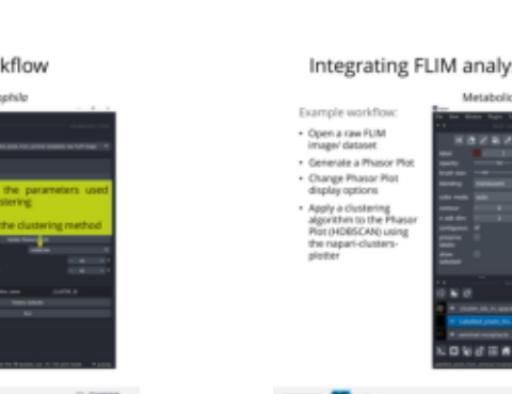
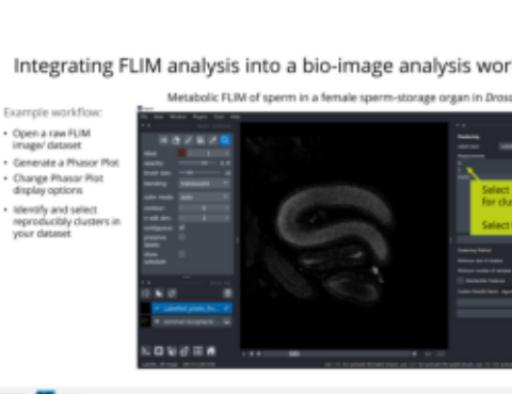
ChatGPT for analyzing scientific image data

Prompt engineering

ChatGPT for analyzing scientific image data

Summary: Deep Learning for Bio-image Analysis

More resources



A screenshot of a Jupyter Notebook interface. The title bar reads "Bio-image Analysis Notebooks". The notebook contains several cells of Python code, some of which have been run and show their output. A callout box on the right side of the screen contains the text "Just submit a pull-request to the Bio-image Analysis Notebooks".

Machine learning

Finds more and more applications, also in life sciences.

Artificial intelligence

Machine learning

Deep Learning

Trainable Weka Segmentation
<https://imagej.net/plugins/tws/>

LabKit
<https://imagej.net/plugins/labkit/>

Python / scikit-learn / napari / apoc

FLUTE

GUI Development and Implementation

Python programming language

Qt

Libraries

NumPy

SciPy

FLUTE

GLB

Image data source: BBPCellLine, available from the Broad Biomed.

Testing executability

- LLMs can solve simple image analysis tasks

Prompt	Response	Result
<pre>simile_question = """ Write Python code only and no additional explanatory text. write a python program, that # takes image from 'image1.jpg' # labels objects in this image, and # visualizes results. Assume this program could be executed in a Jupyter notebook. It is not necessary to save the results. Show the results in Jupyter. """ simile_question</pre>	<pre>code = preprocess(simile_question) print(code) def main(): from skimage import io from skimage.filters import threshold_otsu from skimage.measure import label from skimage.morphology import closing import matplotlib.pyplot as plt image = io.imread('image1.jpg') binary = threshold_otsu(image) binary = image > binary labels = label(binary) image_label_overlay = label_img(labels, image=image) fig, ax = plt.subplots(figsize=(10, 10)) ax.imshow(image_label_overlay) plt.show()</pre>	

cupy

CUDA-based GPU-accelerated [image] data processing in

<https://sciphi.readthedocs.io/en/latest/notebooks/107.ipynb>

Python

- Github
 - Python Notebooks online lesbar
 - Fehlersuche einfach, ohne etwas herunter laden zu müssen

Reproducible bio-image analysis

- Allowing others to do your experiment again.
- “The image data was analyzed with Python.”

Can you reproduce what they did?

Can you reproduce what they did?

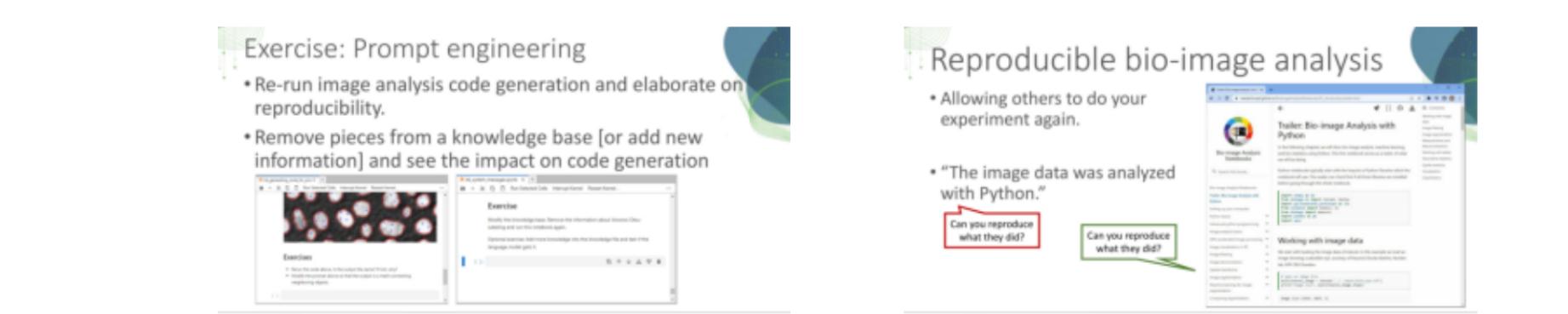
Future Developments

- Direct import of the common FLIM file formats (std, fbd and .ptu) inside the Python code by using already available Python libraries and open-source codes.
- Intermediary file format which encompasses matrices for intensity, g and s coordinates.
- Increase FLUTE speed with parallel computing capabilities of Graphics processing units (GPUs) by specialized Python libraries such as Numba and CuPy for real time phasor analysis and representation during experiments.
- Adapt phasor analysis to typical time-gated sampling limitations by taking in account the effect of decay truncation and gate shape.
- Integrate a fully automated calculation and mapping of fraction of molecules (with 2 known molecular species).
- Advanced analysis tools such as, different filters, freehand cluster drawing, cluster analysis with Machine Learning, FRET trajectory estimation and calculation of absolute concentration of NADH.
- Napari

Query: How can I ensure reproducibility in my image analysis pipeline? Approach A

Exercise: Prompt engineering

- Re-run image analysis code generation and elaborate on reproducibility.
- Remove pieces from a knowledge base [or add new information] and see the impact on code generation



ScaDS.Ai
ScaDS.Dresden
Technische Universität Dresden
Universität Leipzig

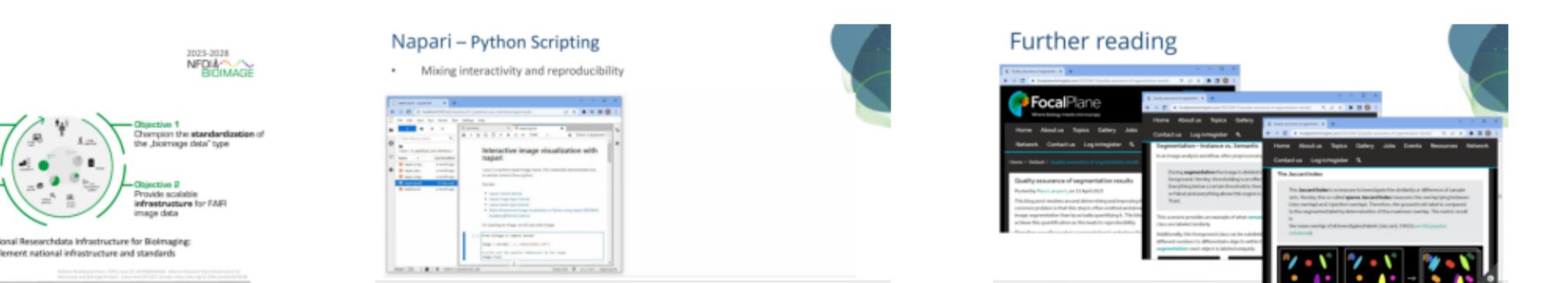
Reproducible bio-image analysis

- Allowing others to do your experiment again.
- "The image data was analyzed with Python."



NFDI4BIOIMAGE
Ki67 scoring in breast cancer biopsies using HALO, QuPath, QuantPath
Acs et al. Lab Invest (2018)
FocalPlane
ScaDS.Ai
Technische Universität Dresden
Universität Leipzig

Independent comparison of digital pathology software

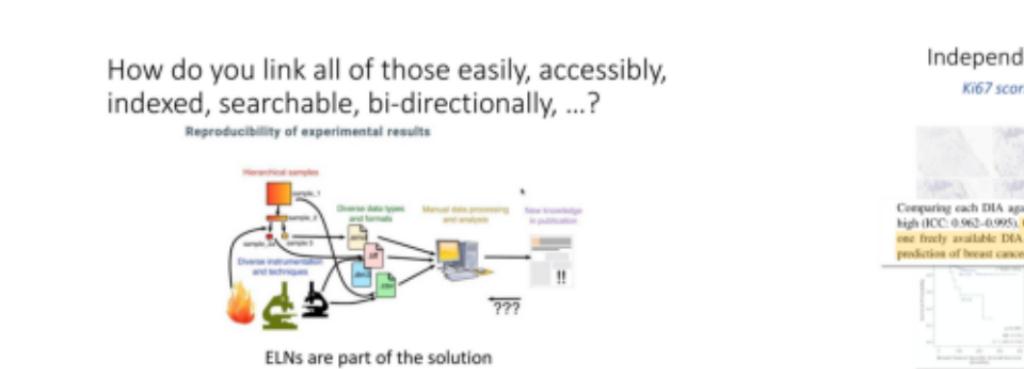


NFDI4BIOIMAGE
Ki67 scoring in breast cancer biopsies using HALO, QuPath, QuantPath
Acs et al. Lab Invest (2018)
International Ki67 in Breast Cancer Working Group
National Research Data Infrastructure for Biomedicine: Implement national infrastructure and standards
nfdi
ScaDS.Ai
Technische Universität Dresden
Universität Leipzig

Napari – Python Scripting

Further reading

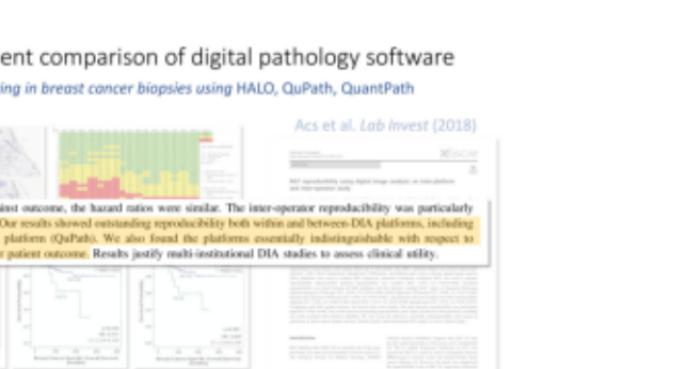
- Accurate & reproducible Ki67 scoring is hard!
- Mixing interactivity and reproducibility



NFDI4BIOIMAGE
Ki67 scoring in breast cancer biopsies using HALO, QuPath, QuantPath
Acs et al. Lab Invest (2018)
International Ki67 in Breast Cancer Working Group
National Research Data Infrastructure for Biomedicine: Implement national infrastructure and standards
nfdi
ScaDS.Ai
Technische Universität Dresden
Universität Leipzig

How do you link all of those easily, accessibly, indexed, searchable, bi-directionally, ...?

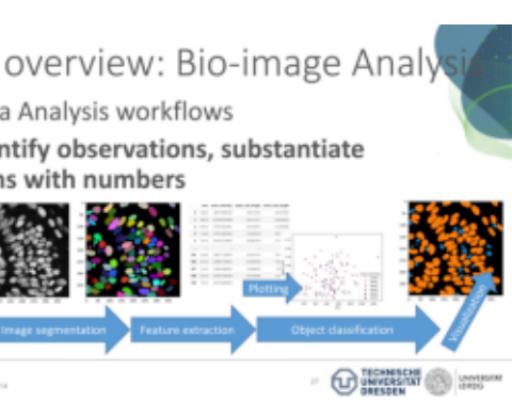
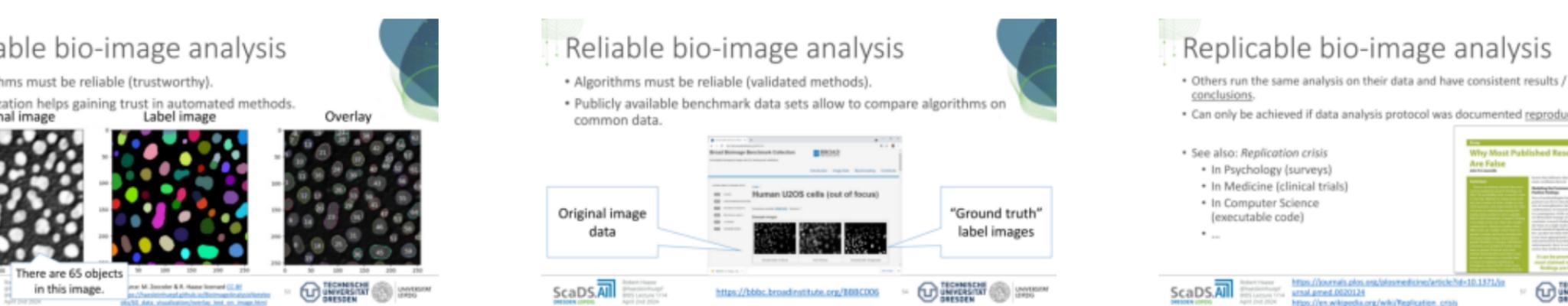
Independent comparison of digital pathology software



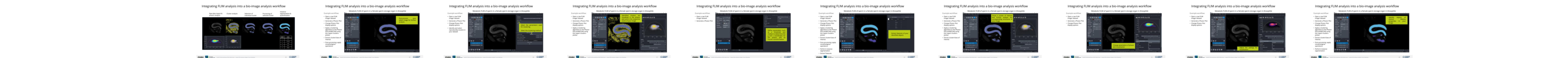
NFDI4BIOIMAGE
Ki67 scoring in breast cancer biopsies using HALO, QuPath, QuantPath
Acs et al. Lab Invest (2018)
International Ki67 in Breast Cancer Working Group
National Research Data Infrastructure for Biomedicine: Implement national infrastructure and standards
nfdi
ScaDS.Ai
Technische Universität Dresden
Universität Leipzig

Query: How can I ensure reproducibility in my image analysis pipeline?

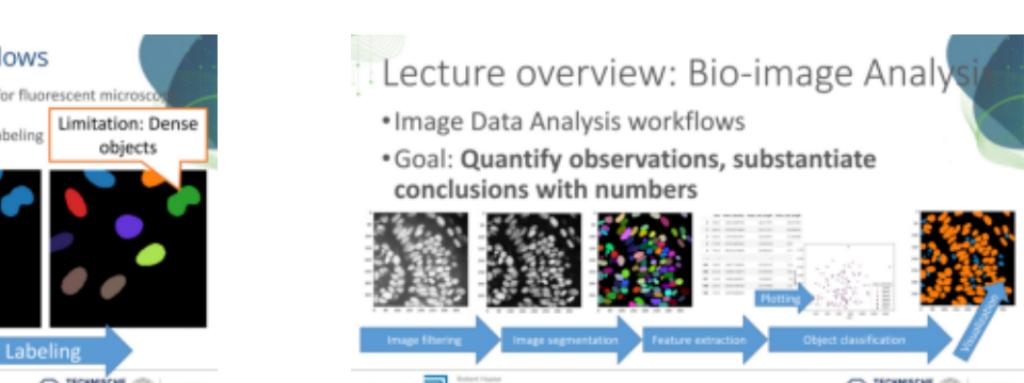
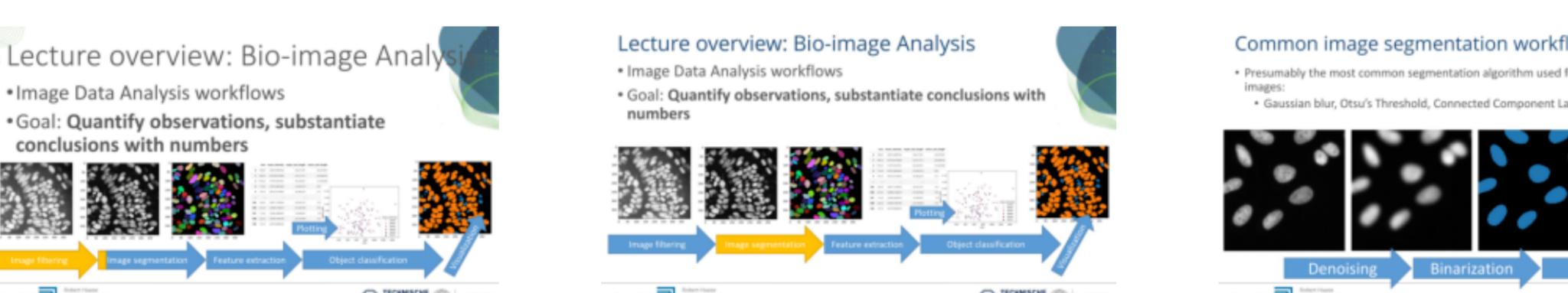
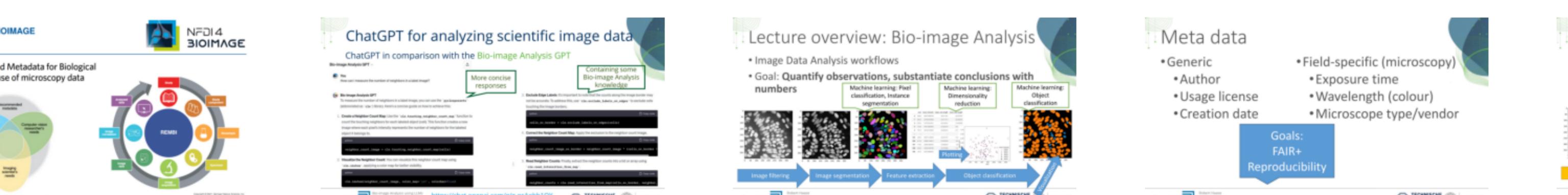
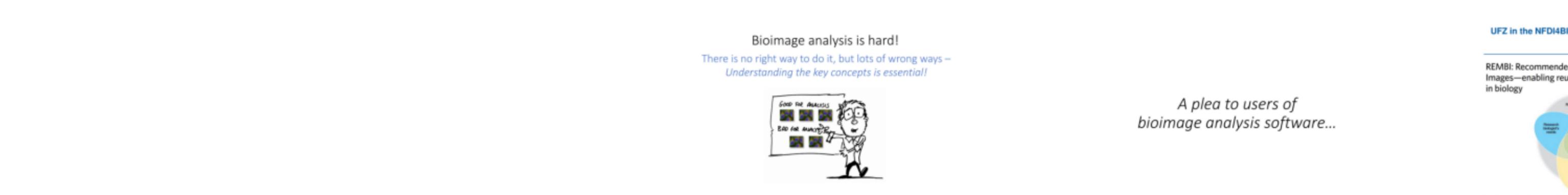
Approach B



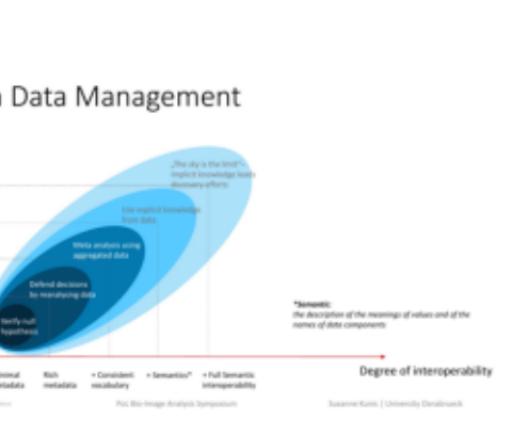
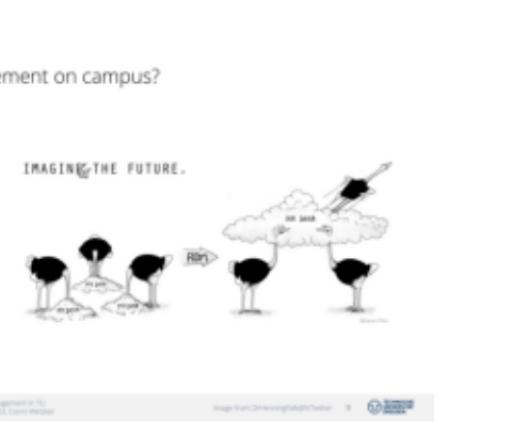
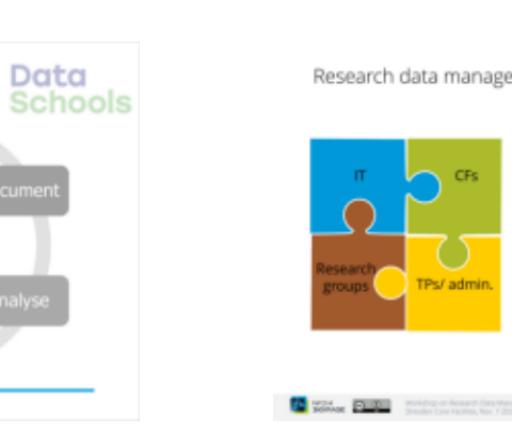
Query: What are common steps in a reproducible bioimage analysis workflow?
Approach A



Query: What are common steps in a reproducible bioimage analysis workflow?
Approach B

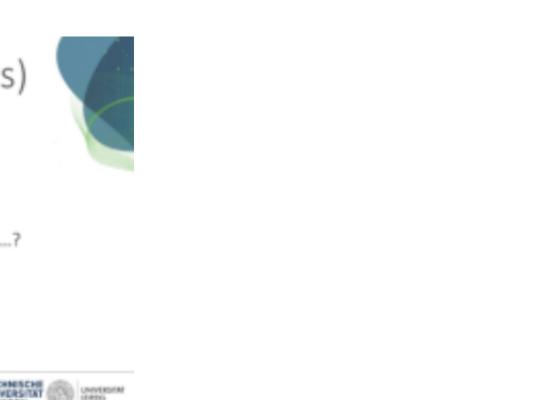
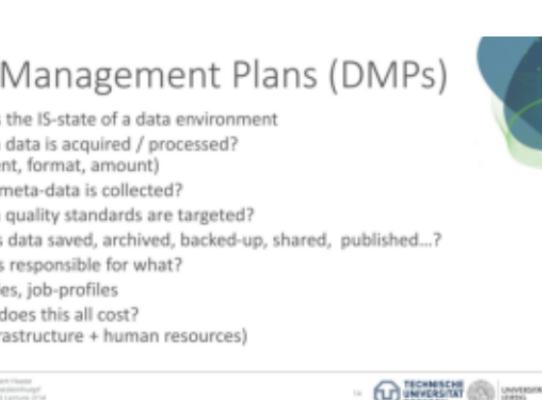
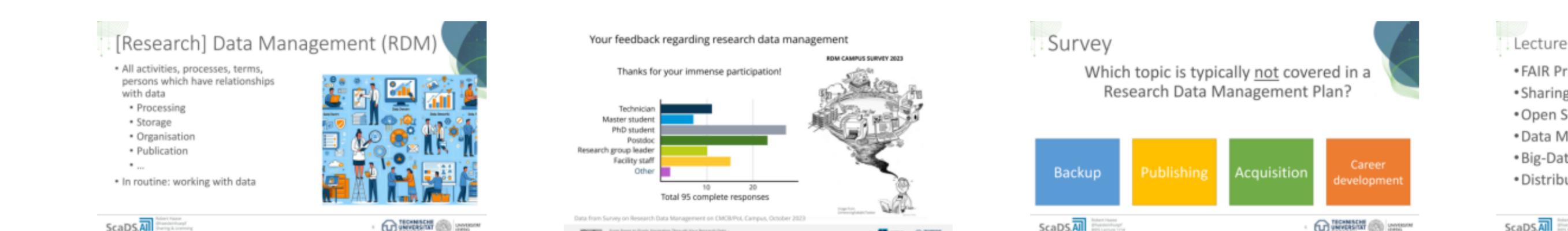
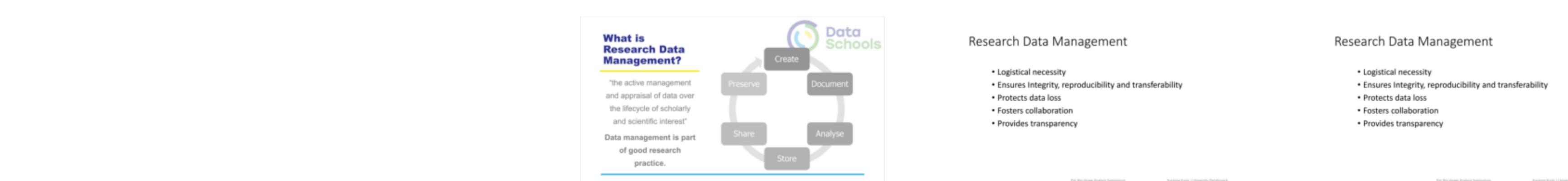


Query: What are the key parts of research data management?
Approach A



Query: What are the key parts of research data management?

Approach B



Query: How can I share my microscopy data with others?

Approach A

Tools – Data Management, metadata handling and sharing
OMERO – Open Microscopy Environment Remote Object

OMERO (Open Microscopy Environment Remote Object) is an open source client/server system for visualizing, managing, and annotating microscope images and metadata.

Key features include:

- Upload (More than 340 file formats supported)
- Sharing (Share and collaborate with colleagues)
- Messaging (Metadata, annotations, filtering)
- Analyze (Access and use images)
- Viewing (Up to 5-dimensional viewing is supported)
- Publish and Export



Where to share?

The I3D:bio knowledge hub www.i3dbio.de

I3D:bio – Information Infrastructure for BioImage Data

Open science related content

- bioRxiv (manuscripts, no reviews)
- Figshare
- F1000
- F1000/NEUBIAS (slides)
- BioImage Archive (data)
- Github (code)
- Zenodo
- FocalPlane
- Image Analysis Training Resource
- Teaching Material & Events
- Guides & Resources
- Help Desk & Contact



Where to share?

The I3D:bio knowledge hub www.i3dbio.de

I3D:bio – Information Infrastructure for BioImage Data

Open training related content

- bioRxiv (manuscripts, no reviews)
- Figshare
- F1000
- F1000/NEUBIAS (slides)
- BioImage Archive (data)
- Github (code)
- Zenodo
- FocalPlane
- Image Analysis Training Resource
- Teaching Material & Events
- Guides & Resources
- Help Desk & Contact



Exercise: Sharing files on Zenodo

Where to share?

Exercise: Sharing files on Zenodo

Advantages of using OMERO

- Organize your original image files in a central storage location
- Open training related content
- bioRxiv (manuscripts, no reviews)
- Figshare
- F1000
- F1000/NEUBIAS (slides)
- BioImage Archive (data)
- Github (code)
- Zenodo
- FocalPlane
- Image Analysis Training Resource
- Teaching Material & Events
- Guides & Resources
- Help Desk & Contact



Exercise: Sharing files on Zenodo

Where to share?

Exercise: Sharing files on Zenodo

Advantages of using OMERO

- Organize your original image files in a central storage location
- Open training related content
- bioRxiv (manuscripts, no reviews)
- Figshare
- F1000
- F1000/NEUBIAS (slides)
- BioImage Archive (data)
- Github (code)
- Zenodo
- FocalPlane
- Image Analysis Training Resource
- Teaching Material & Events
- Guides & Resources
- Help Desk & Contact

Query: How can I share my microscopy data with others?
Approach B

