FAIRSpectra

Enabling the FAIRification of Spectroscopy and Spectrometry

Alex Henderson

University of Manchester

The University of Manchester
Office for Open Research

https://alexhenderson.info

https://fairspectra.net

Thanks…

- For financial support
    - University of Manchester's Office for Open Research
    - SurfaceSpectra Ltd.
- For in-kind support (free exhibition space)
    - UK Surface Analysis Users Forum (UKSAF)
    - SIMS Europe
    - SpringSciX 2024
    - 101st IUVSTA Workshop (The International Union for Vacuum Science, Technique and Applications)
    - Zulip (free upgrade)

What is FAIR?

## The FAIR Guiding Principles

- **F**indable
- **A**ccessible
- **I**nteroperable
- **R**eusable

https://www.go-fair.org/
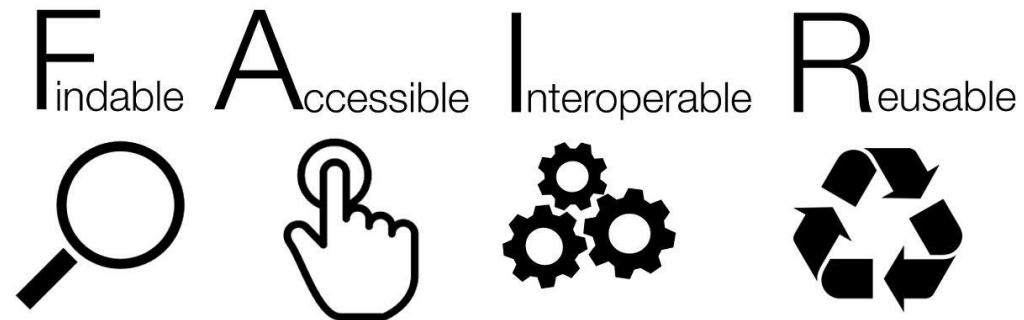
What are Spectroscopy and Spectrometry?

- Instrumentation-based chemical analysis
  - mass spectrometry (MALDI, SIMS, DESI)
  - UV-vis / infrared / Raman spectroscopies
  - NMR
  - X-ray diffraction
  - …
- Variants of each technique have own requirements
- Need to consider combination of techniques → data fusion
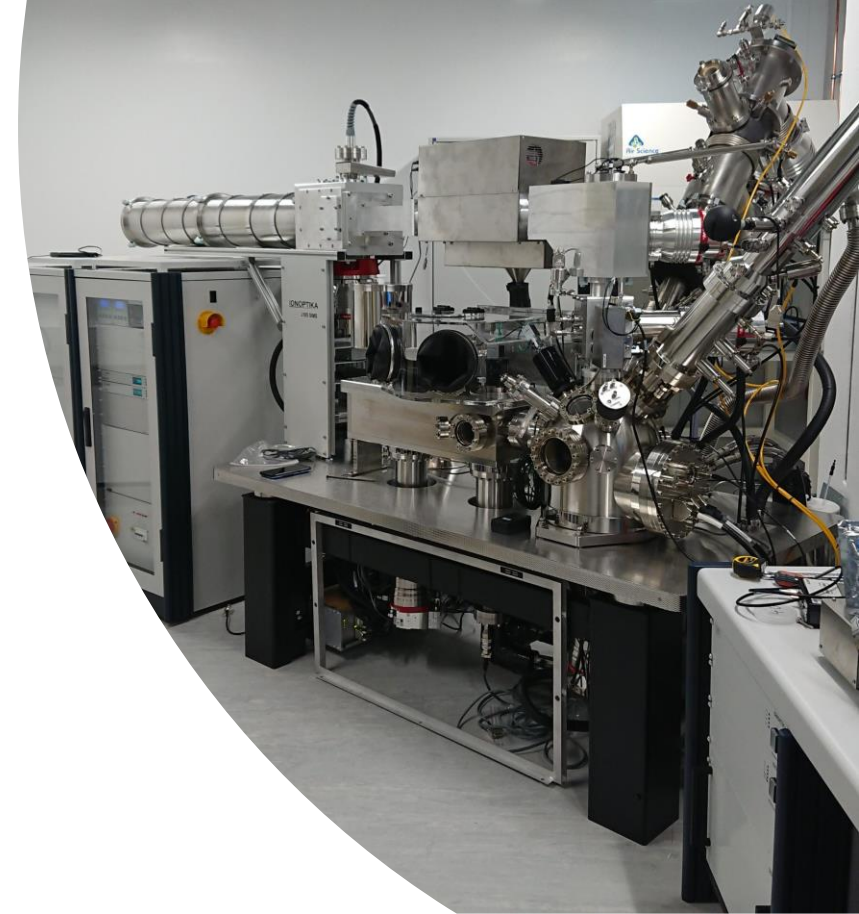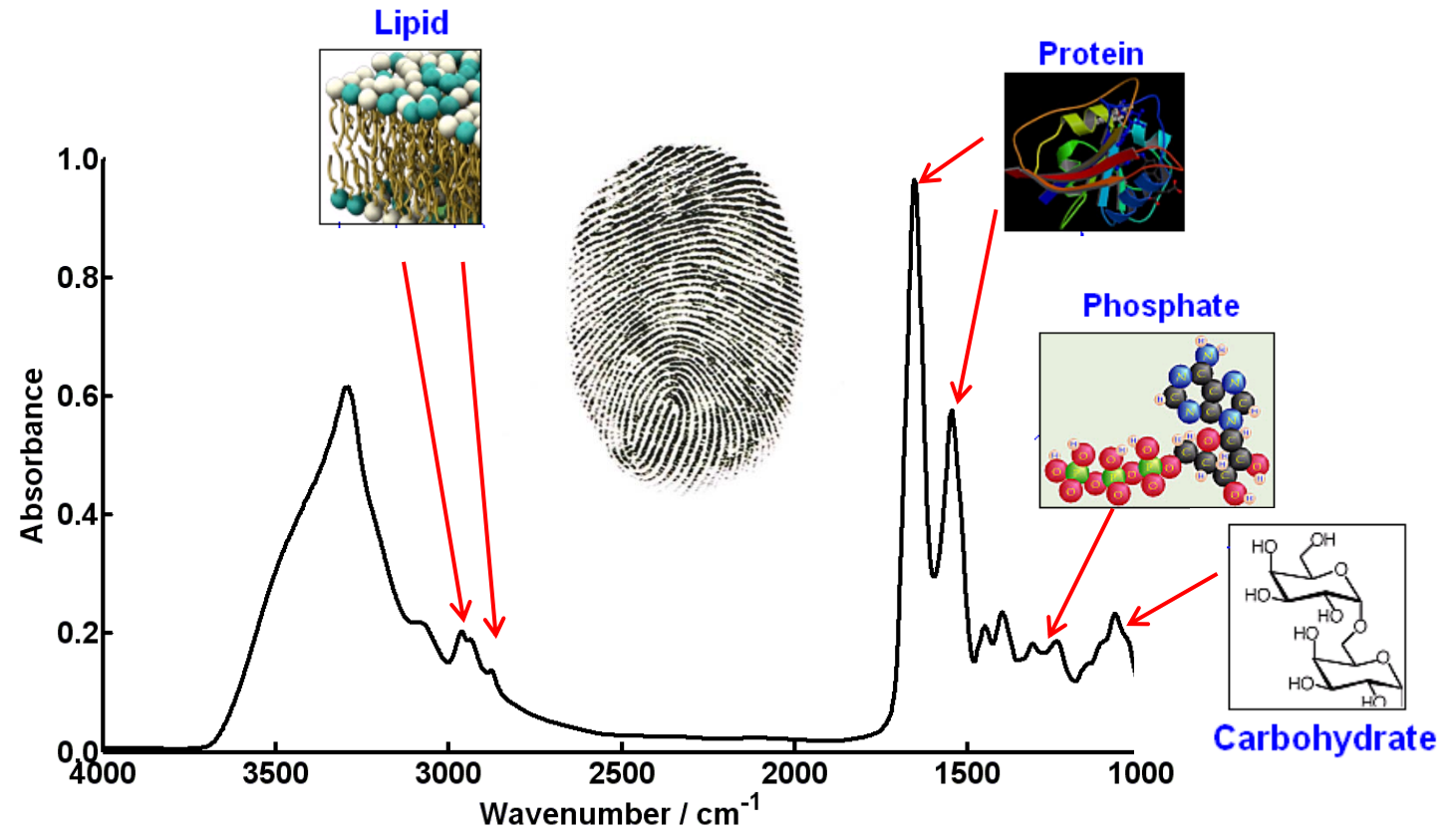
Modalities

- Single spectrum
- Collections of spectra
- Spectral maps
- Multispectral images
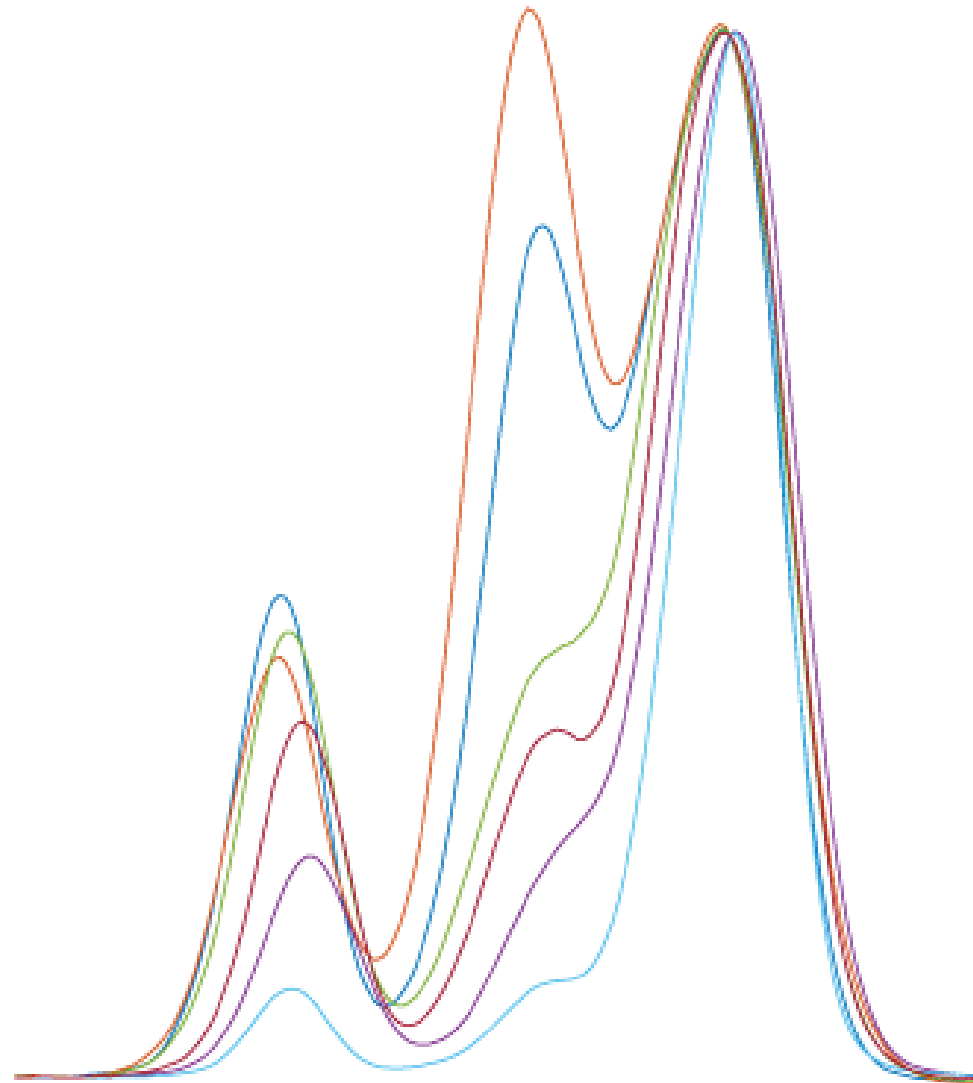- Hyperspectral images
- 3D images

Modalities

- **Single spectrum**
- Collections of spectra
- Spectral maps
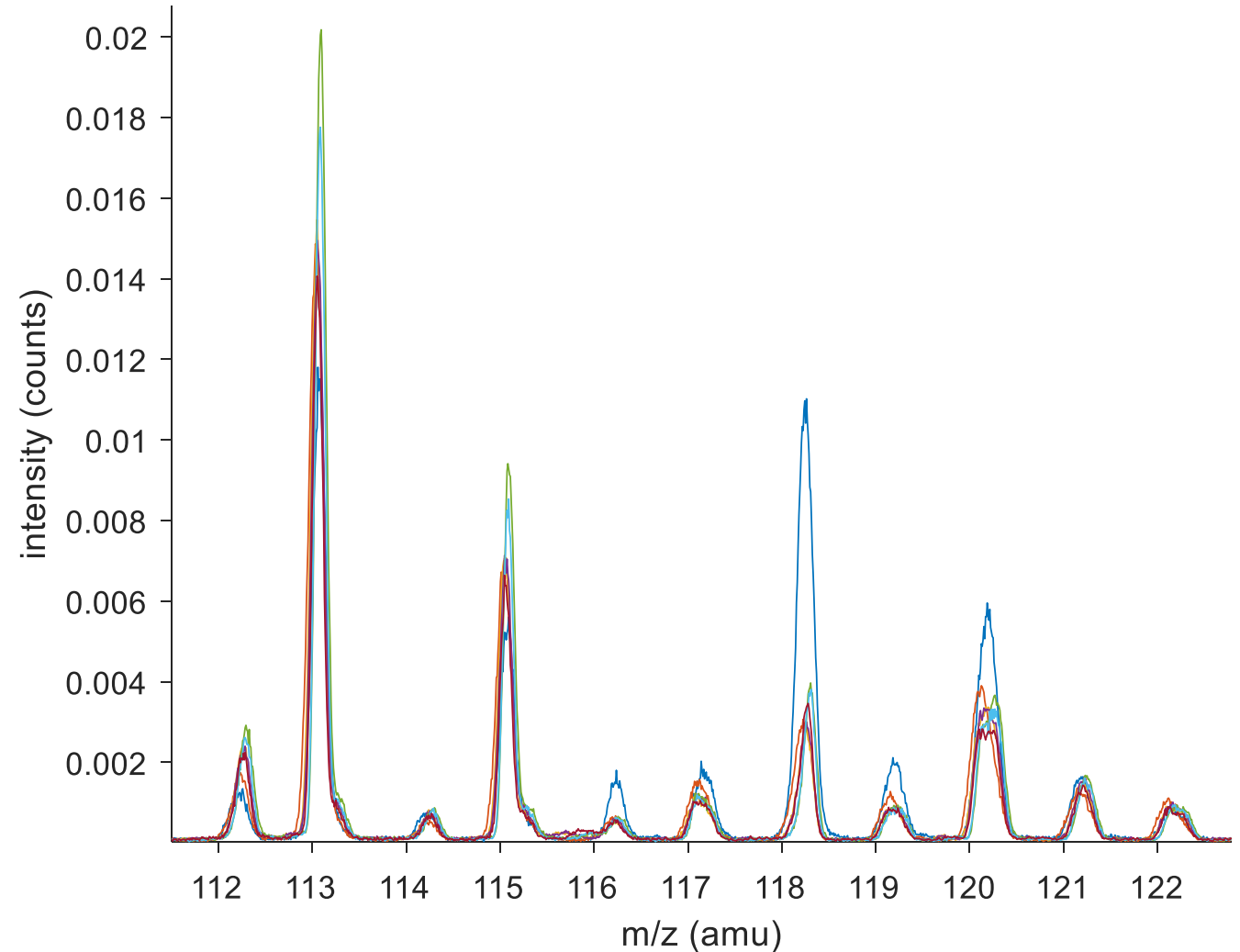- Multispectral images
- Hyperspectral images
- 3D images

Typical infrared spectrum of tissue. Courtesy Peter Gardner @ Manchester

Modalities

- Single spectrum
- **Collections of spectra**
- Spectral maps
- Multispectral images
- Hyperspectral images
- 3D images

X–ray photoelectron spectra. Range of carbon functionalities in polymers. SurfaceSpectra Ltd.

Modalities

- Single spectrum
- **Collections of spectra**
- Spectral maps
- Multispectral images
- Hyperspectral images
- 3D images

Secondary ion mass spectrometry spectra from a range of urinary-tract infection bacteria. John Fletcher @ Manchester now Gothenburg
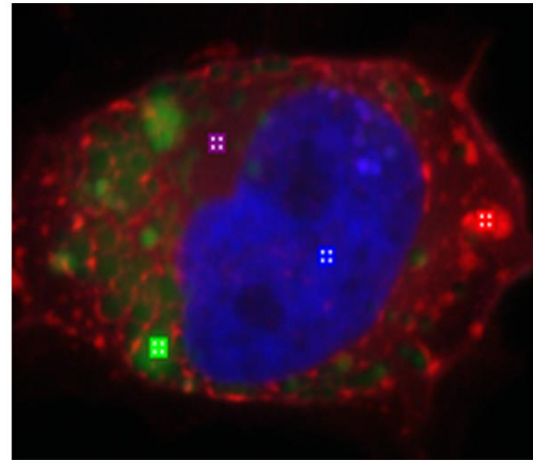
Modalities
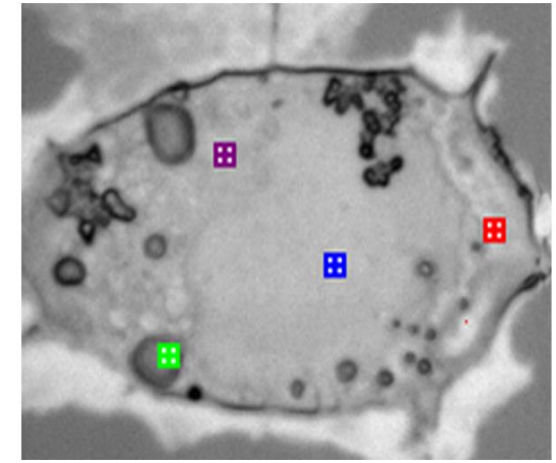
- Single spectrum
- Collections of spectra
- **Spectral maps**
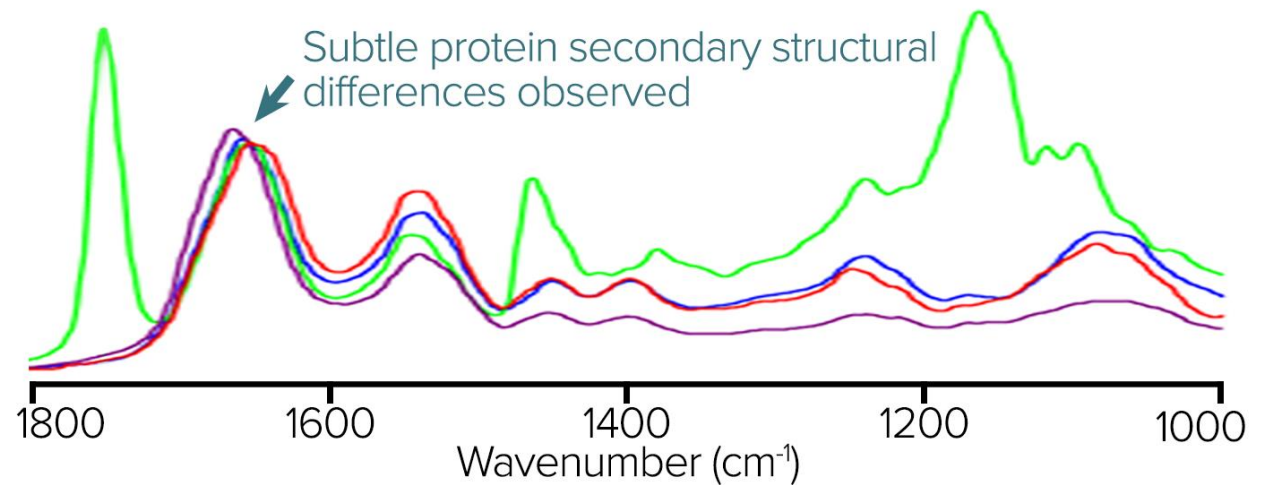- Multispectral images
- Hyperspectral images
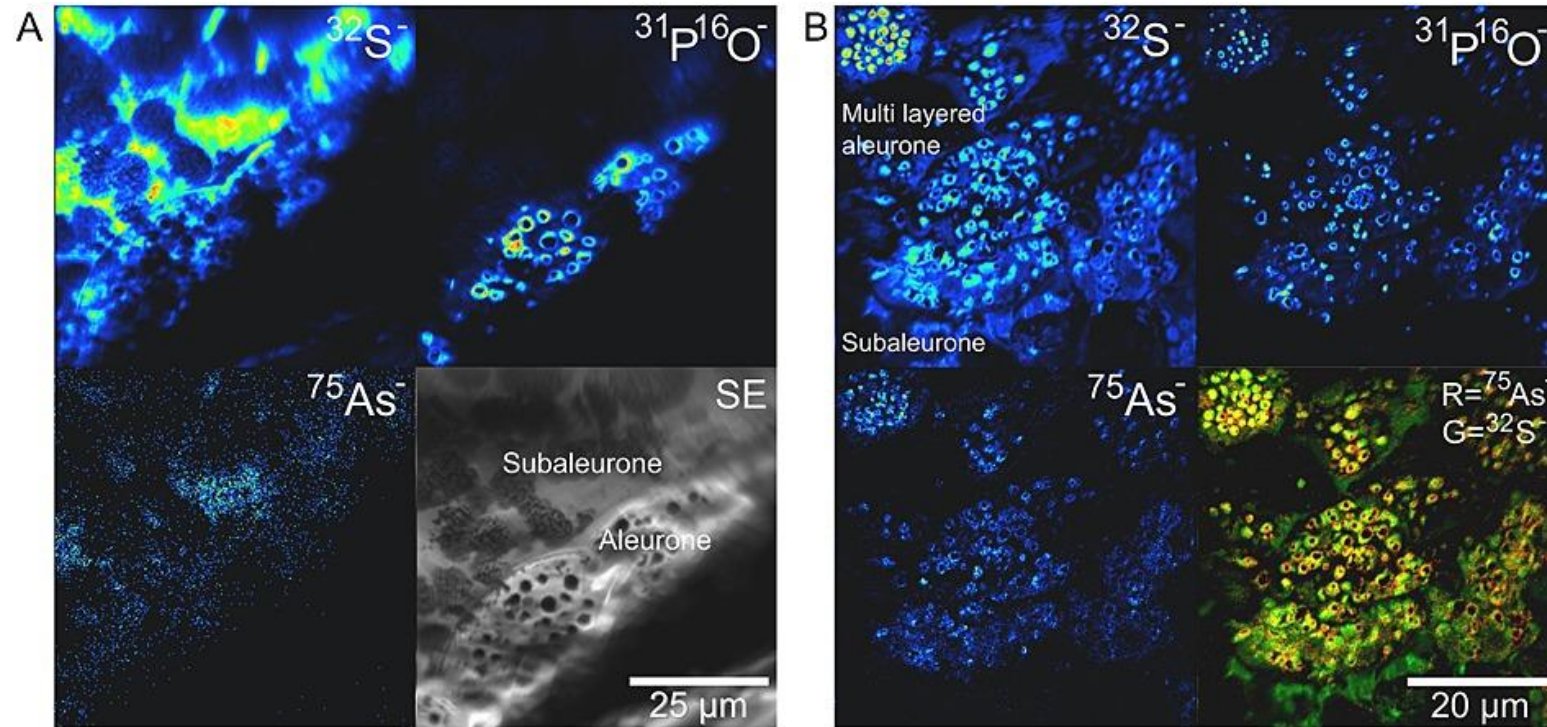- 3D images

FL image

Brightfield image

0   (μm)   53   0   (μm)   53

O-PTIR spectra of stained cells in seconds with <500 nm spatial resolution

Subtle protein secondary structural differences observed

Wavenumber (cm⁻¹)

🟥 Protein stress granule, 🟦 Nucleus, 🟪 Cytoplasm, 🟩 Lipid droplets

Co-located infrared and Raman spectroscopy spectra of neuroglioma cells. Photothermal Inc.

Modalities

- Single spectrum
- Collections of spectra
- Spectral maps
- **Multispectral images**
- Hyperspectral images
- 3D images

Secondary ion mass spectrometry image of arsenic, sulfur and phosphate abundance in rice. Katie Moore @ Manchester

What is hyperspectral imaging?

- Operates more like a camera, with multiple image elements
- 128 × 128 pixels, liquid nitrogen cooled
- Mosaic these 'pictures' to cover large areas

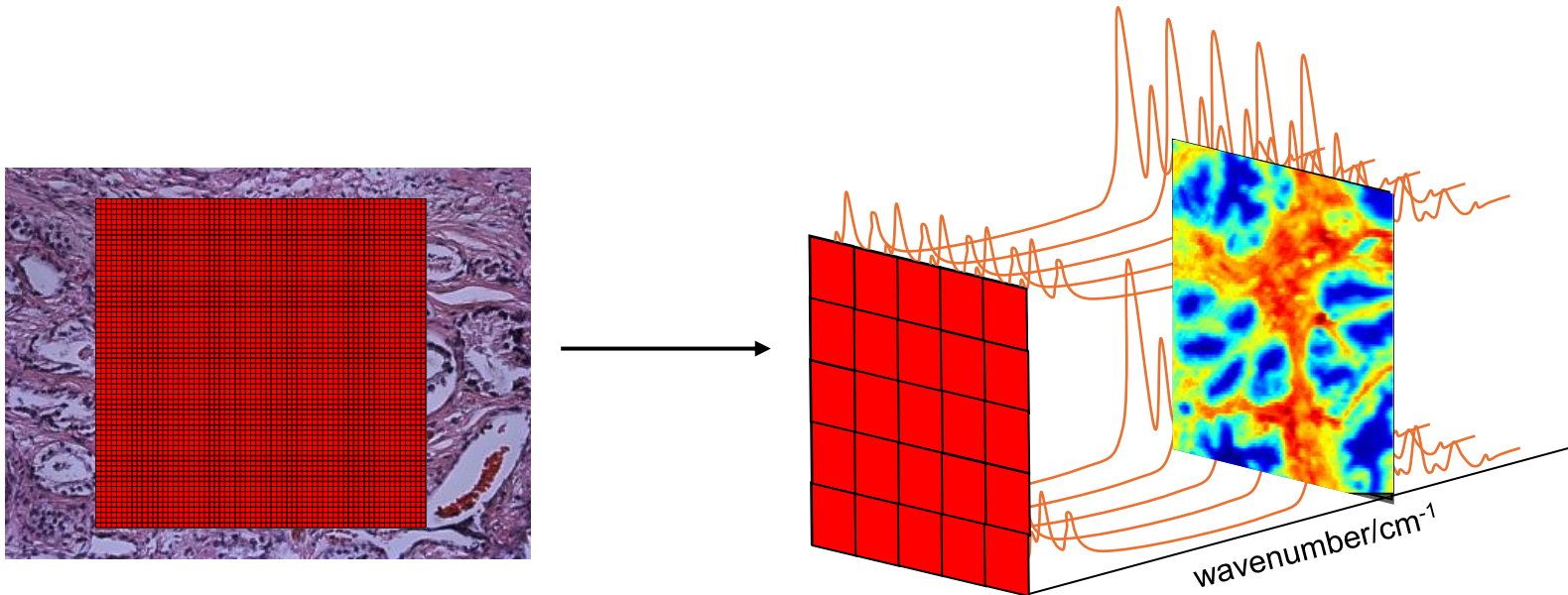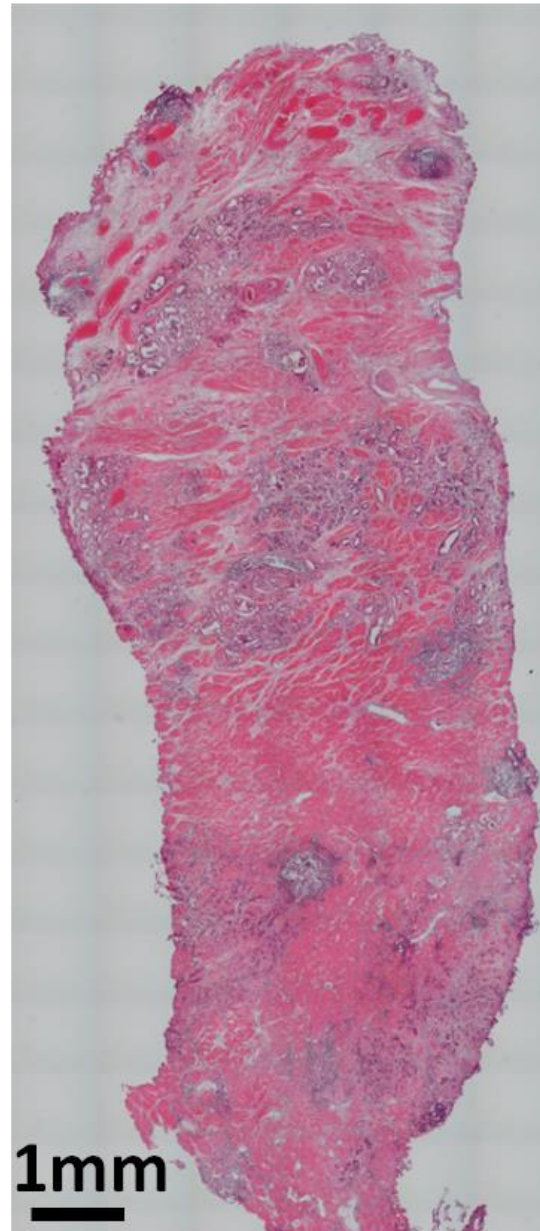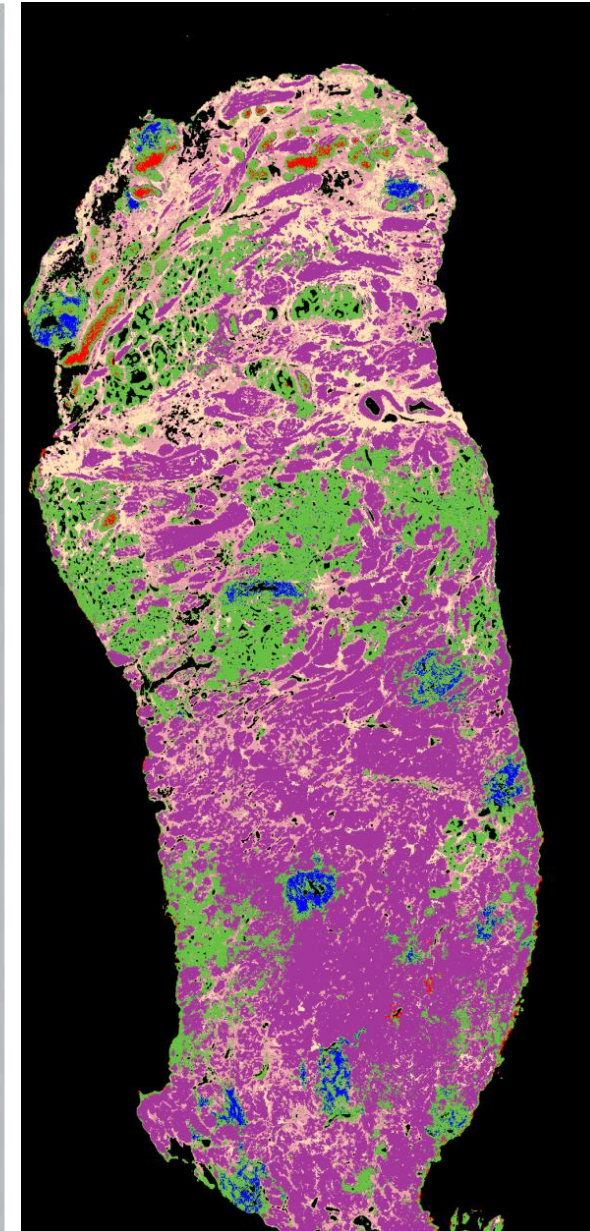## Modalities

- Single spectrum
- Collections of spectra
- Spectral maps
- Multispectral images
- **Hyperspectral images**
- 3D images

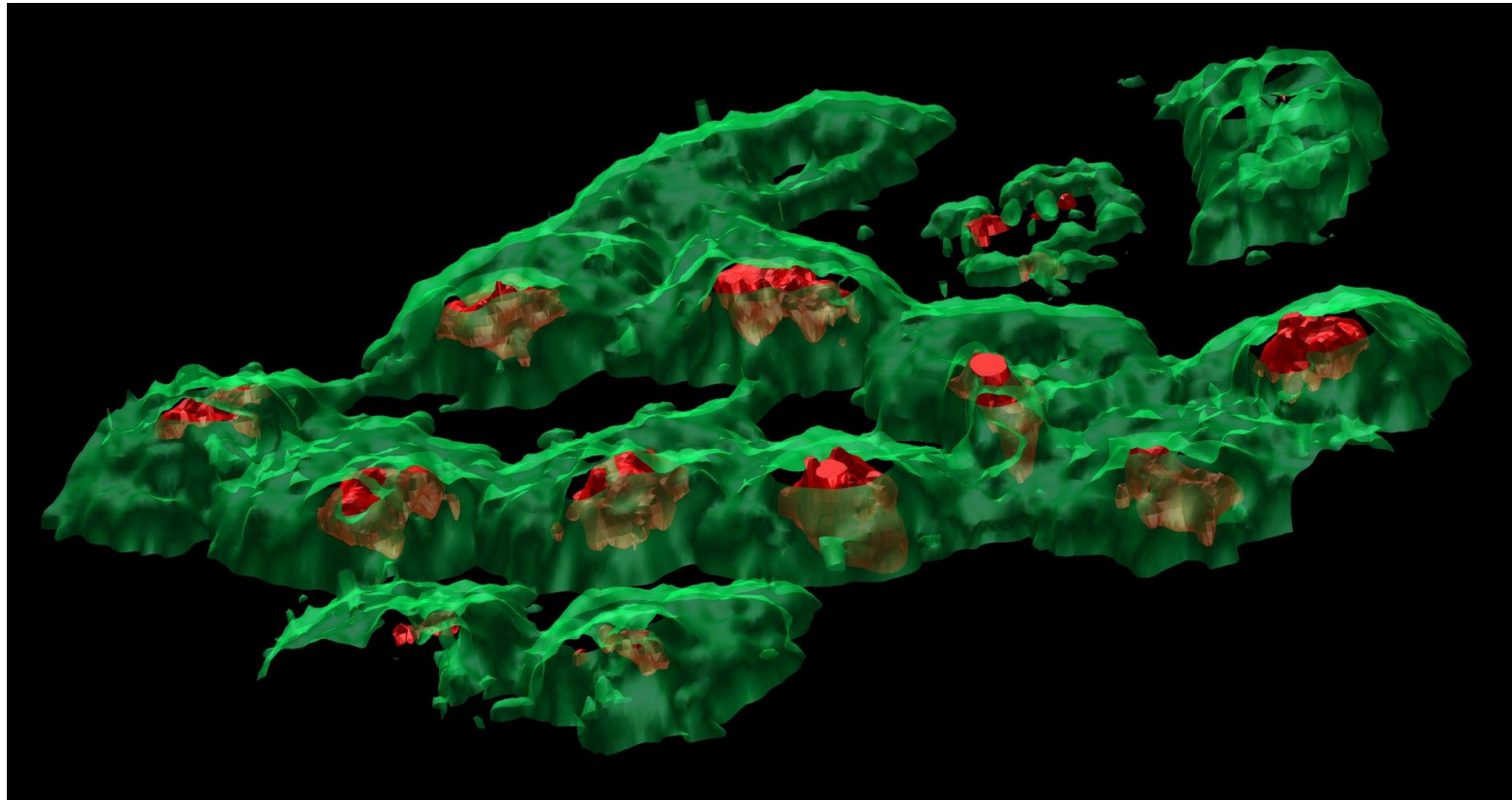H&E optical   FTIR

Epithelium
Smooth Muscle
Lymphocytes
Blood
ECM
Concretion
Fibrous Stroma

1mm

Infrared spectroscopy hyperspectral image of prostate cancer tissue. False coloured Random Forests classification of spectra. Peter Gardner @ Manchester

Modalities

- Single spectrum
- Collections of spectra
- Spectral maps
- Multispectral images
- Hyperspectral images
- **3D images**

Secondary ion mass spectrometry 3D image depth profile, topography corrected, false coloured green=lipid (DPPC), R=nucleic acid (Adenine). John Fletcher & Alex Henderson @ Manchester
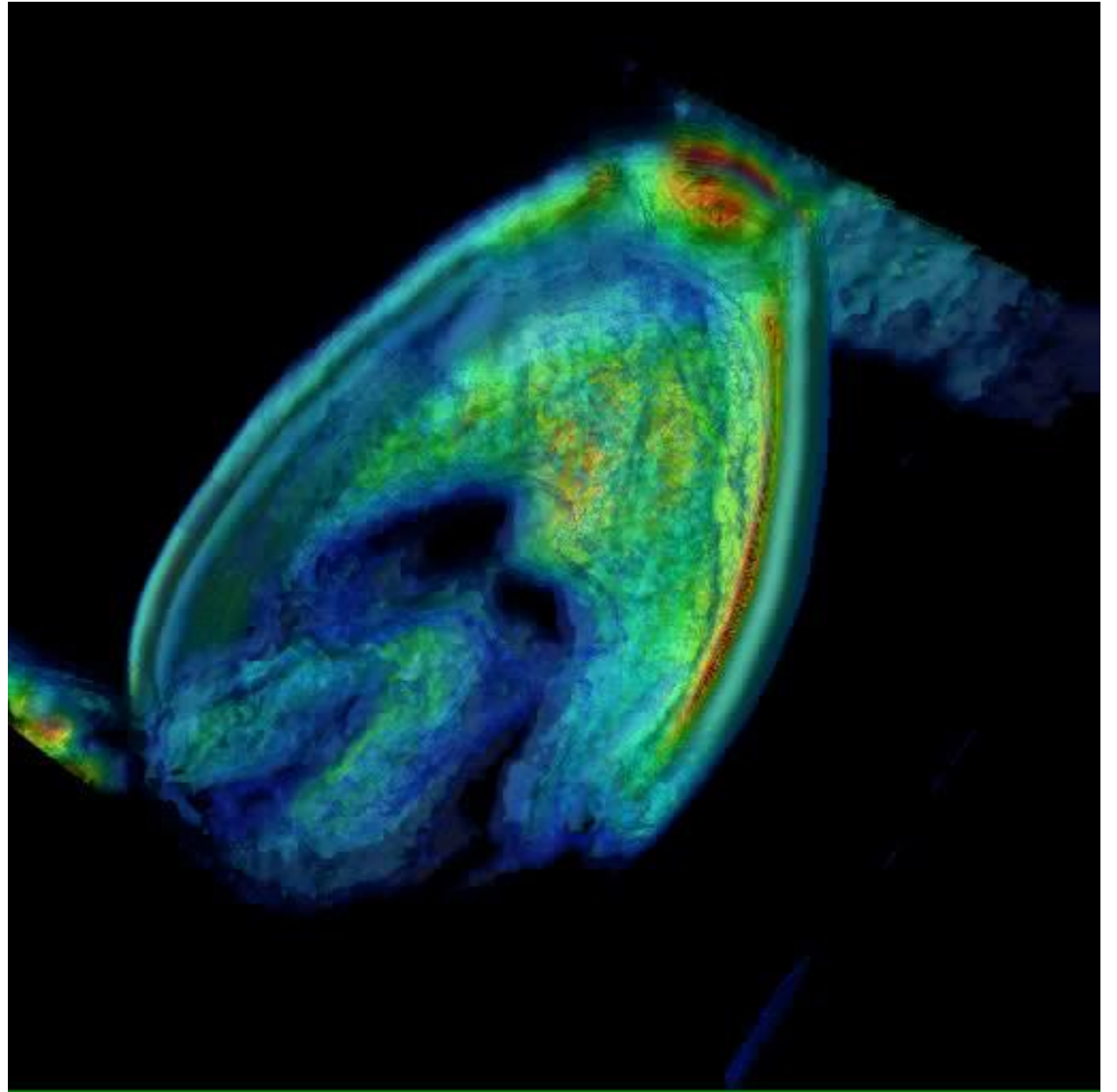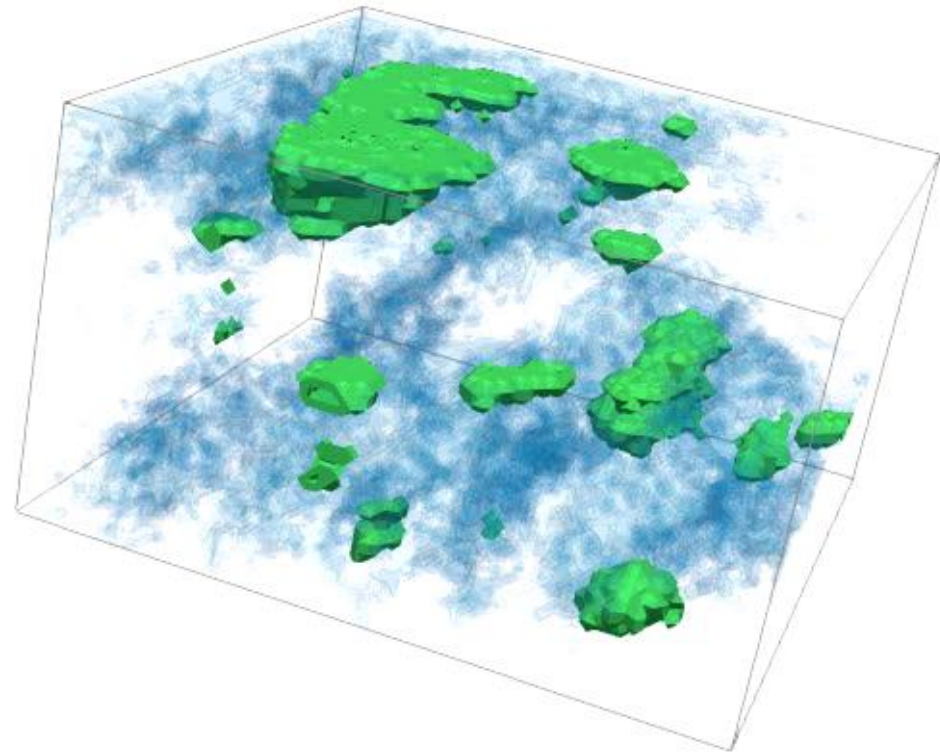
Modalities

- Single spectrum
- Collections of spectra
- Spectral maps
- Multispectral images
- Hyperspectral images
- **3D images**

Secondary ion mass spectrometry 3D depth profile image showing cholesterol distribution in surface of frog oocyte. John Fletcher @ Manchester, now Gothenburg

Modalities

- Single spectrum
- Collections of spectra
- Spectral maps
- Multispectral images
- Hyperspectral images
- **3D images**

X-ray photoelectron spectroscopy image of nitrogen and specific carbon functionality in caffeine tablet. Alex Henderson @ Manchester with Kratos Analytical.

What are the issues?

Academia

Funders require 'data' to be deposited in (open) repositories

But…

- No dedicated repositories

- Metadata terms are patchy

- Instrument data in proprietary file formats

- Many software packages not compatible with open formats

Researchers willing to share, but don't know how

What are the issues?

Commercial activity needs to be considered

Barriers
- FAIR often confused with Open
- In-house processes considered good enough
- Worry about certain metadata usage giving secrets away

Benefits
- Easier to share data in-house, between labs and (overseas) sites
- FAIR practises lead to better records retention
- Acquisitions and mergers become more straightforward
- Third-party (open source) software becomes easily accessible
- Incoming staff already familiar with systems

Instrument manufacturer buy-in is vital
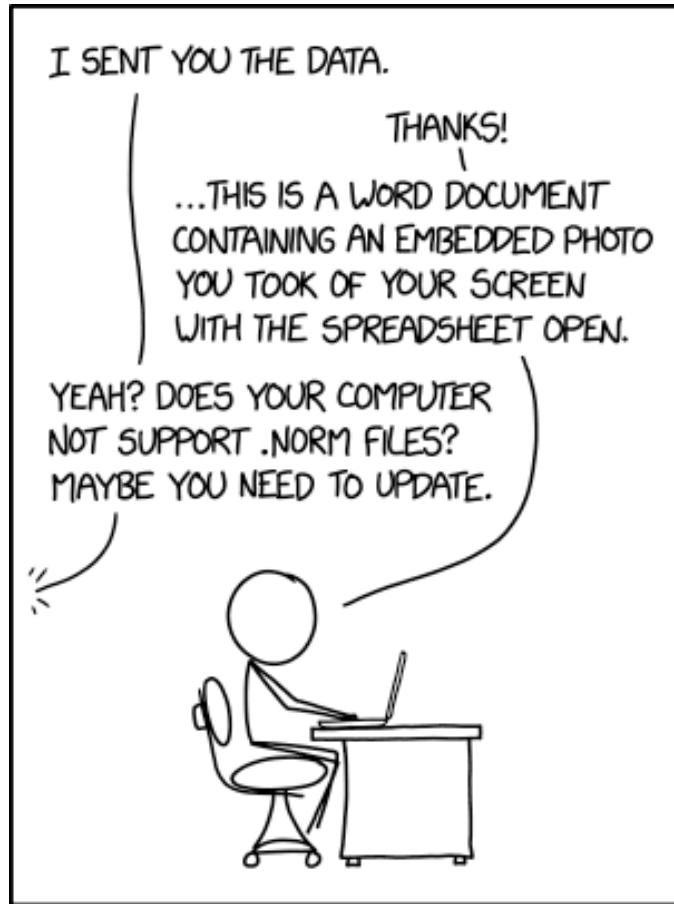
Moving forward

What is FAIRSpectra?
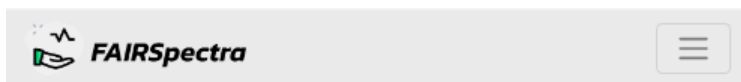
Community driven initiative

Focus on hyperspectral imaging techniques

- File formats for hyperspectral imaging
  - No standards exist right now
  - Software tools to support these
- Metadata requirements
- Education and training
- Raising awareness

What is FAIRSpectra?

https://fairspectra.net

https://fairspectra.zulipchat.com

https://github.com/FAIRSpectra

Which metadata are required?

Upstream sample provenance

Sample
- Sampling method
- Storage conditions
- Chemical modifications
- Physical state
- Pre-treatment
- …

Experiment
- Experiment plan
- Substrate material
- Mounting method
- Region analysed
- Instrument params
- …

Analysis
- Artifact removal
- Pre-processing
- Algorithm choice
- Hyperparameters
- Validation method
- …

Downstream reporting

Where do we start?

Born-digital metadata in data files
Limited/common options

Upstream sample provenance

- Sampling method
- Storage conditions
- Chemical modifications
- Physical state
- Pre-treatment
- …

Sample

- Experiment plan
- Substrate material
- Mounting method
- Region analysed
- Instrument params
- …

Experiment

- Artifact removal
- Pre-processing
- Algorithm choice
- Hyperparameters
- Validation method
- …

Analysis

Downstream reporting

Where do we start?

Upstream sample provenance

Many workflows have
common steps
Default hyperparameters

Sample
- Sampling method
- Storage conditions
- Chemical modifications
- Physical state
- Pre-treatment
- …

Experiment
- Experiment plan
- Substrate material
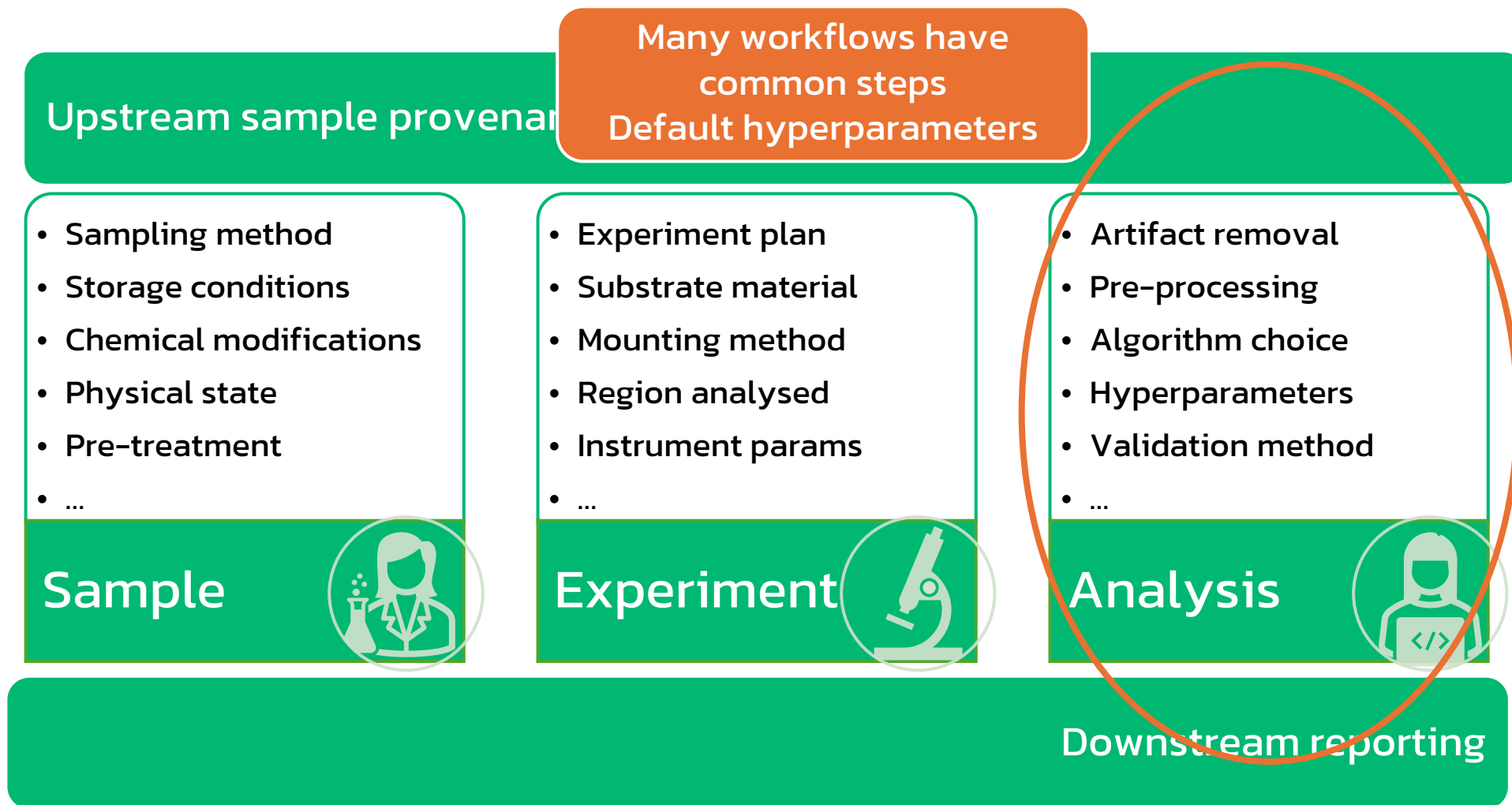- Mounting method
- Region analysed
- Instrument params
- …

Analysis
- Artifact removal
- Pre-processing
- Algorithm choice
- Hyperparameters
- Validation method
- …

Downstream reporting

Where do we start?

Upstream sample provenance

Sample
- Sampling method
- Storage conditions
- Chemical modifications
- Physical state
- Pre-treatment
- …

Experiment
- Experiment plan
- Substrate material
- Mounting method
- Region analysed
- Instrument params
- …

Analysis
- Artifact removal
- Pre-processing
- Algorithm choice
- Hyperparameters
- Validation method
- …

Samples so varied makes this very difficult

Downstream reporting

Where do we start?

Upstream sample provenance

Sample
- Sampling method
- Storage conditions
- Chemical modificati...
- Physical state
- Pre-treatment
- ...

Experiment
- Experiment plan
- ...

Analysis
- Artifact removal
- ...rocessing
- ...ithm choice
- ...rparameters
- ...ation method
- ...

> Workflows also include repeated steps with poorly defined break points

Downstream reporting
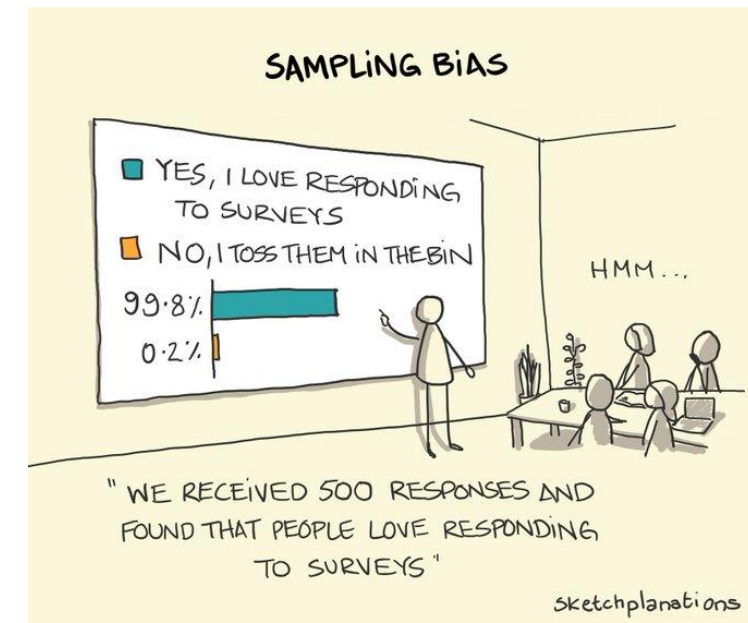
Where are we now – 6 months in?

Exhibition booths at 3 conferences/workshops
- More planned, another 3 before Christmas

Survey
- Positives
  - Everyone wanted to see something done, not sure about how
- Barriers
  - People have difficulty sharing
  - Poor documentation
  - Proprietary file formats – loss of information
  - Raw data versus processed data – large file size
  - Gazumping / IP & prior art / confidentiality
  - Time consuming

SAMPLING BIAS

☐ YES, I LOVE RESPONDING TO SURVEYS
☐ NO, I TOSS THEM IN THE BIN

99.8%
0.2%

HMM...

"WE RECEIVED 500 RESPONSES AND FOUND THAT PEOPLE LOVE RESPONDING TO SURVEYS"

sketchplanations

Where are we now – 6 months in?

Data file formats

- Looking to re-purpose strategies from astronomy, climate science, and microscopy
- Some file convertors ready for testing
- Suitable test files an issue – everyone's a critic

Two instrument vendors interested in getting involved

- Need to be careful not to go too fast
- Only one shot at changing instrument software

Discussions with journals just beginning

- Need minimum reporting requirement
- Need to convince referees this is important

Developing connections with BioFAIR, ELIXIR and ISO

Summary

The researchers are willing, but their resources are weak

- Few solutions currently exist
  - Metadata terms missing
- Proprietary file formats are a barrier
  - Instrument vendor buy-in required
- Lack of awareness persists

But...

- Some low-hanging fruit
- Opportunity to make an impact
- Even Closed FAIR can still have benefits to industry

There's lots to do, but FAIRSpectra is just getting started!

https://fairspectra.net