INP FROM IDEA TO PROTOTYPE

BLITZ and WOLKE

A lightweight framework for fast, scalable image data exploration

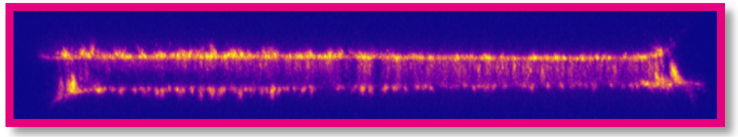
P. Mattern, R. Krieg, H. Höft, **T. Gerling**, M. Becker

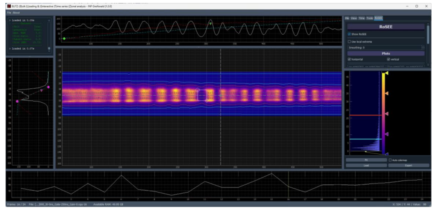
Leibniz Association

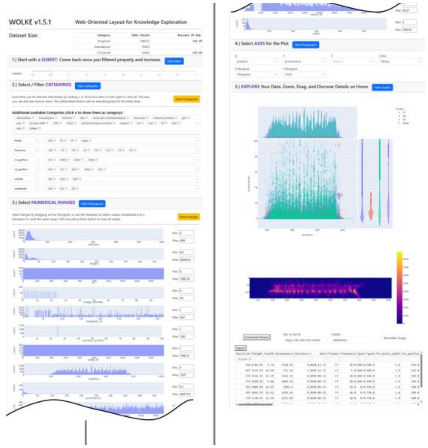
FROM IDEA TO PROTOTYPE

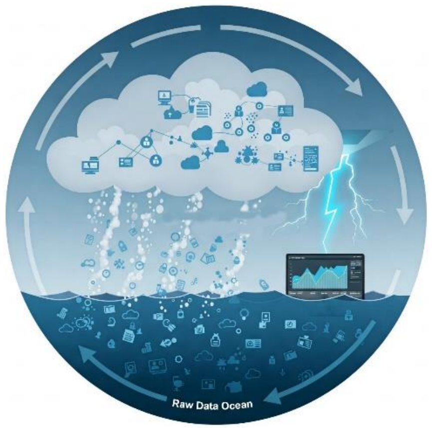
Our Journey Today

1. Challenges in Image-Based Experiments
2. Current Tools and Interfaces
3. BLITZ – High-speed image inspection
4. Integration via DAMPF
5. WOLKE – Metadata filtering & dashboards
6. Vision: The WETTER Framework

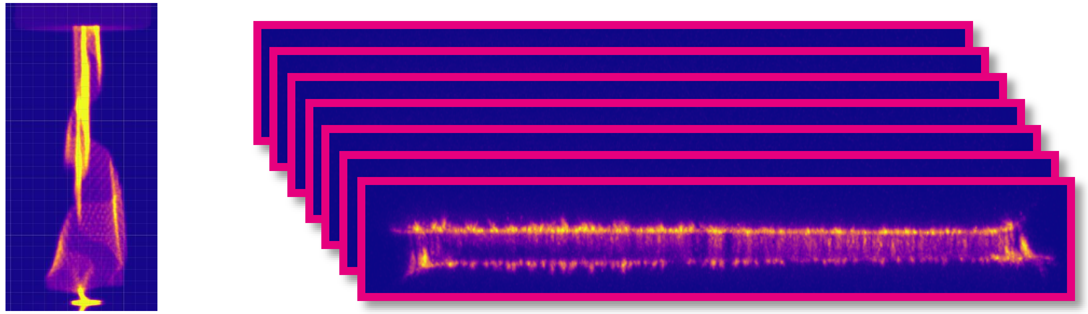




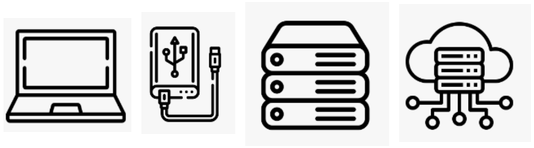




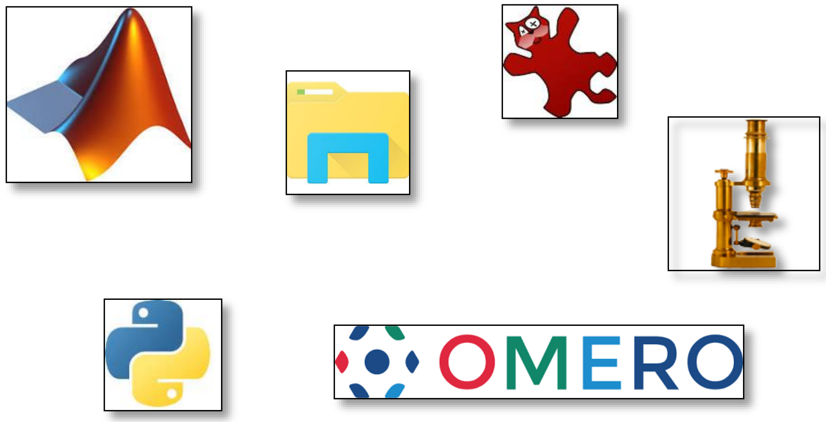
Why Handling Experimental Image Data Is Hard



- Massive volume of data



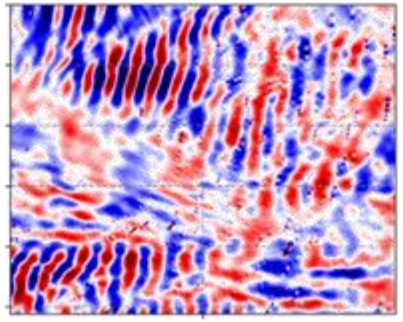
- Limited resources for handling large image datasets (computing power and storage)



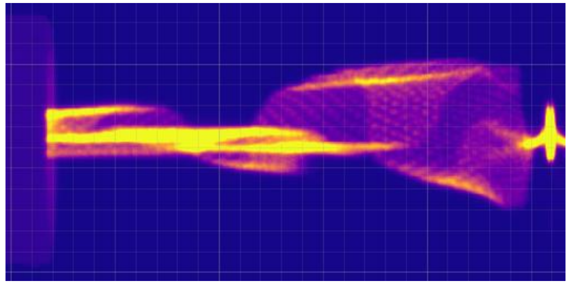
- Need for fast preselection and filtering methods
- Visualization and metadata analysis often inefficient
- Tooling is fragmented and rarely user-friendly

What Are We Really Dealing With?

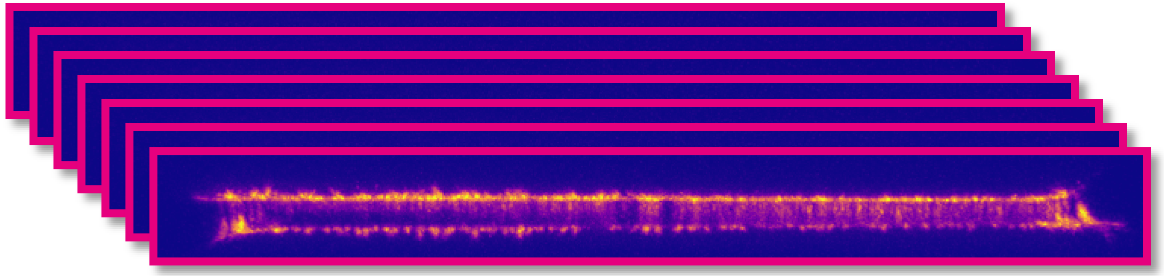
- Massive volume of data (Examples **per Set**!)
- Limited resources for handling large image datasets (computing as well as storage)
- Effective preselection and filtering methods
- Efficient visualization and analysis of metadata
- Interactivity and user-friendly interfaces



Atmospheric Airglow:
- 47 GB
- 30k files



CAPPJ (kINPen):
- 113 GB
- 36k files



Multifilament iCCD:
- 52 GB
- 32k files

Each dataset is a small universe — containing tens of thousands of frames and multiple GBs of raw data.

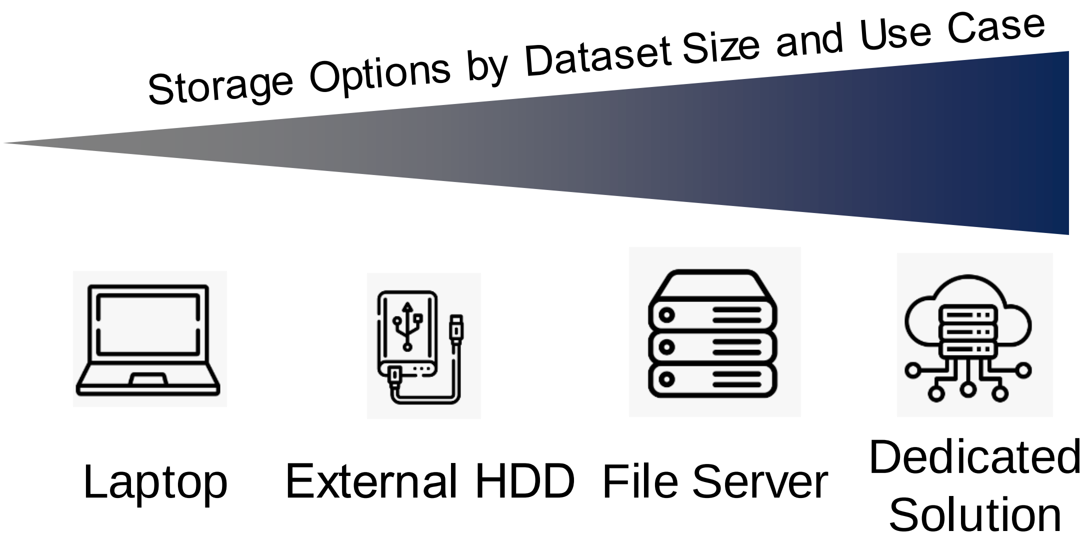
Where Does All That Data Go?

- Massive volume of data
- Limited resources for handling large image datasets (computing as well as storage)
- Effective preselection and filtering methods
- Efficient visualization and analysis of metadata
- Interactivity and user-friendly interfaces

Scalability of Image Dataset Storage:

- Storage should scale with dataset size and complexity
- Metadata systems already help — even for single sets
- Folder structures remain the most common solution

Storage Type	Typical Dataset Range	Use Case
Laptop	**10 – 1,000 images**	Local experiments, quick previews
External HDD	**1,000 – 30,000 images**	Archiving, manual browsing
File Server	**30,000 – 500,000 images**	Shared access, project-wide datasets
Dedicated Solution	**500,000 – 10+ million**	Large-scale research, searchable archives



INP FROM IDEA TO PROTOTYPE

How People Explore Image Data Today

- Massive volume of data
- Limited resources for handling large image datasets (computing as well as storage)
- Effective preselection and filtering methods
- Efficient visualization and analysis of metadata
- Interactivity and user-friendly interfaces

File Explorer: Windows Mac Ubuntu

Database: Google Photos OMERO Open Microscopy Environment

Traditional Image Viewers:

Microsoft Fotos ImageJ IrfanView

Programmatic Interfaces:

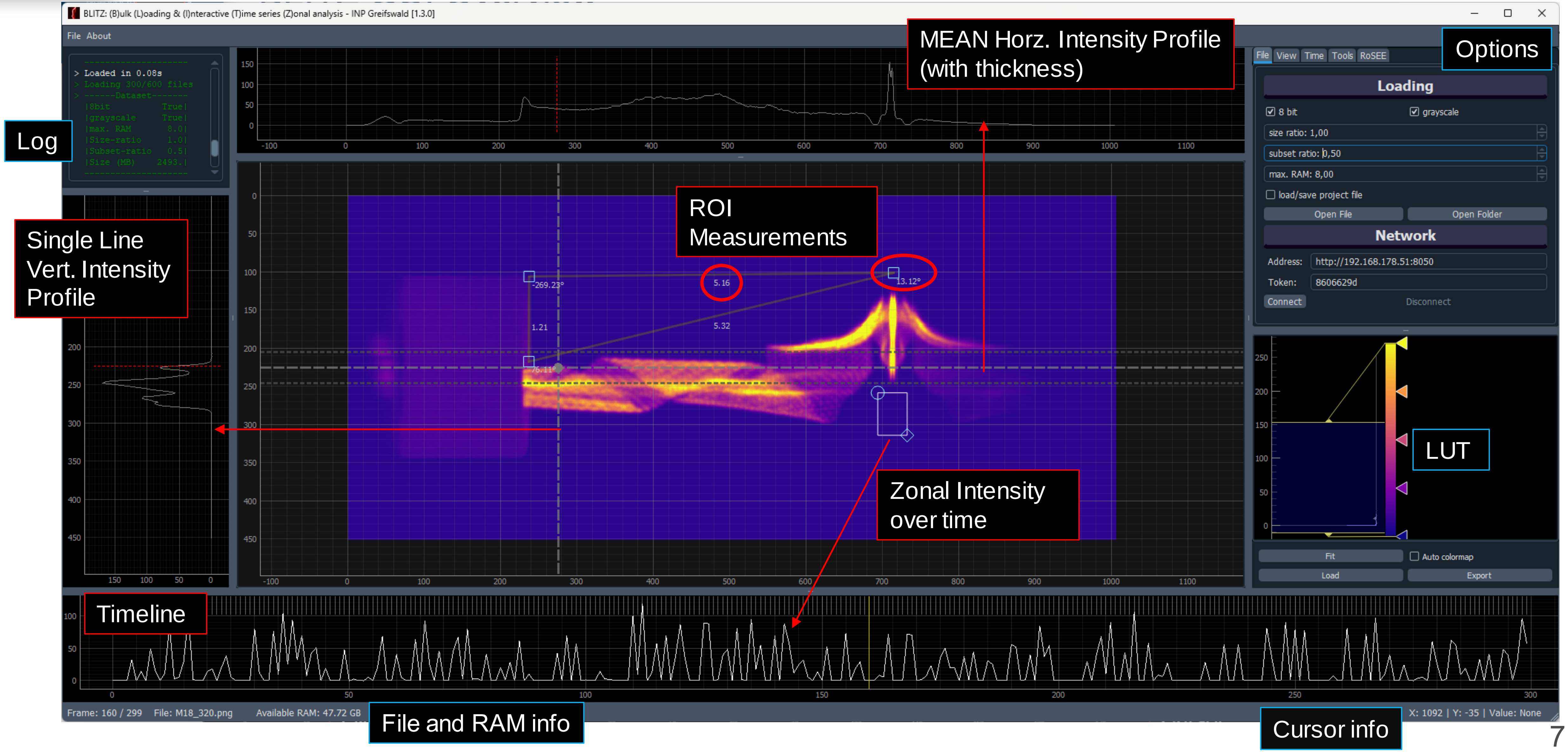
JS Javascript Python Matlab

DIY Exotic Flavours:

Unmaintained scripts, forgotten web GUIs, or that one tool only Peter can run...

"Most of us juggle multiple tools to get the job done — but what if filtering, interaction, and inspection were all part of a single flow?"

BLITZ (Bulk Loading & Interactive Time-series Zonal analysis)



INP FROM IDEA TO PROTOTYPE

BLITZ (**B**ulk **L**oading & **I**nteractive **T**ime-series **Z**onal analysis)

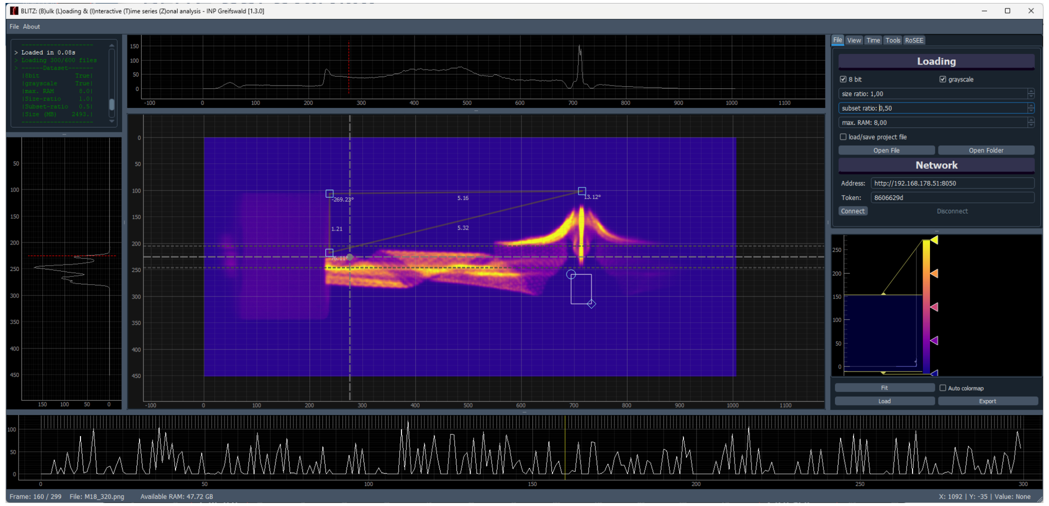
Definition and Purpose:

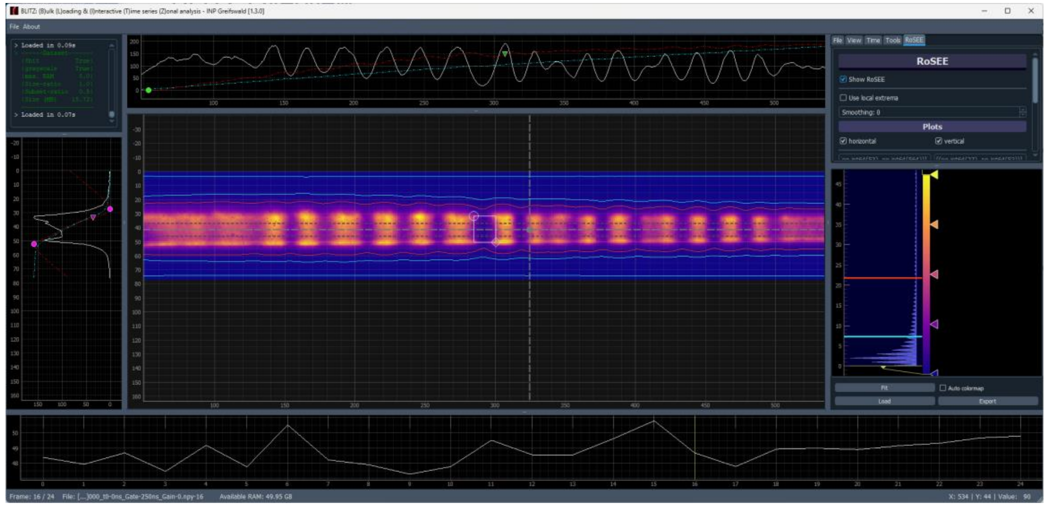
- Matrix-based interactive image viewer
- Designed for efficient large-scale image datasets

Key Features:

- Extremely fast loading (e.g. 21,000 images, 25 GB in 35 sec)
- Real-time interactive exploration and analysis
- Robust tools for measuring, profiling, and zonal analysis
- Drag & Drop support for quick dataset loading

When traditional tools choke on 10,000+ images, BLITZ opens them in seconds - ready to analyze.





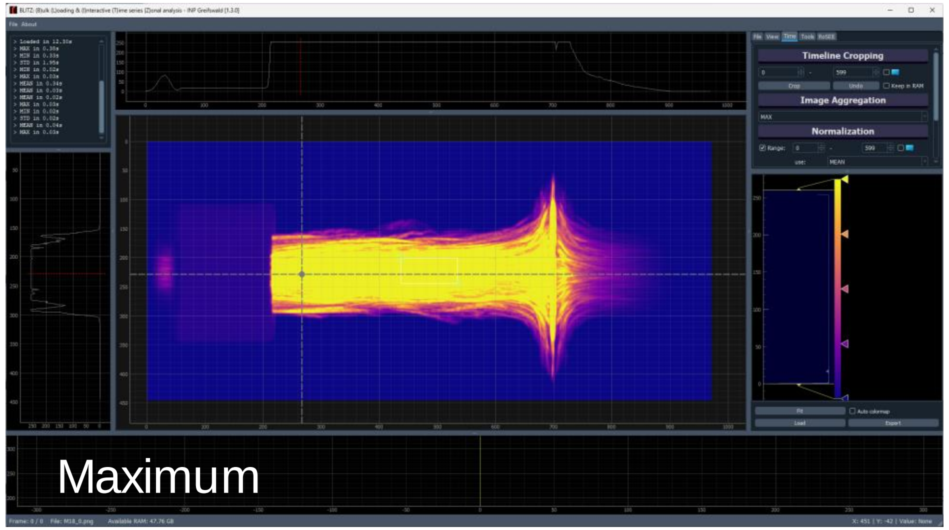
BLITZ - Performance

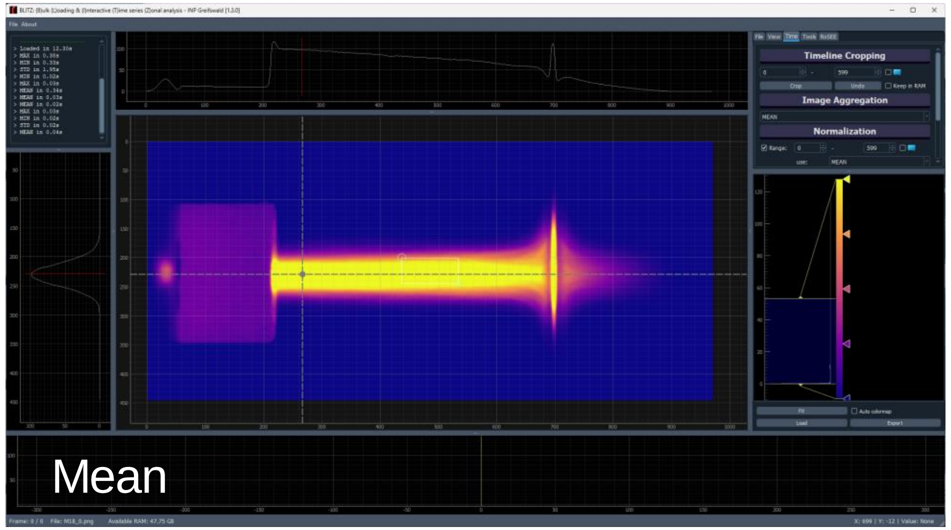
```
> STD in 0.03s
> Loading 600/600 files
> ------Dataset-------
  |8bit          True|
  |grayscale     True|
  |max. RAM       8.0|
  |Size-ratio     1.0|
  |Subset-ratio   1.0|
  |Size (MB)   4987.|
  --------------------
> Loaded in 12.75s
> MEAN in 0.33s
> MAX in 0.35s
> STD in 1.97s
```

Matrix based approach:

- 600 images (5 GB) loaded in 12 sec
- MAX / MEAN in < 0.35 sec
- STD in < 2 sec

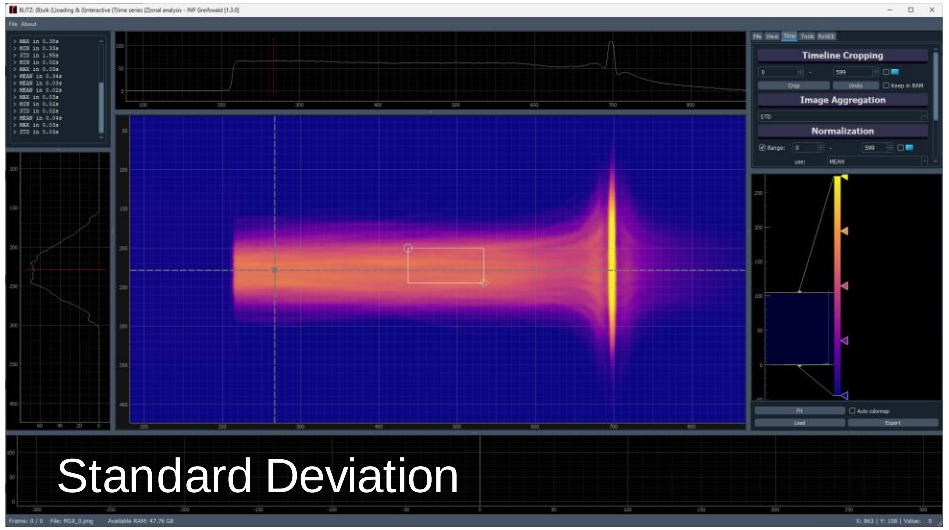
(Tested on a gaming laptop)





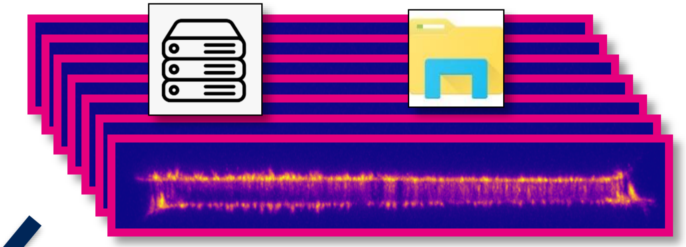
Intensity Values per Pixel over Time

Built for users who just don't want to wait …



DAMPF (Data Aggregation & Modular Processing Framework)

Image Folder Structure



Non- Domain Specific Metrics

i.e. via OpenCV:

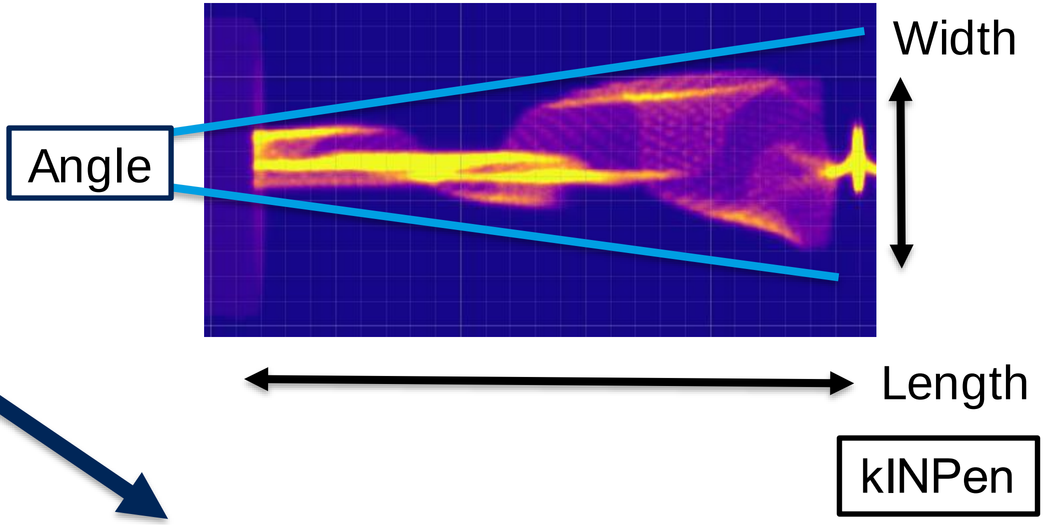
- Min, Max, Mean Intensity
- Laplacian („Noise“)
- Self-similarity (auto-correlation)
- Date / Time

Experimental Conditions:

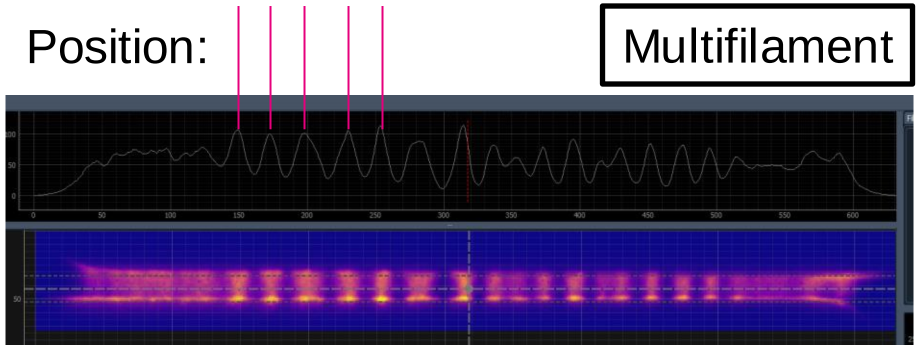
- Voltage
- Current
- Gas Mixture
- …

All into a SQL DB

SQL



Domain Specific Metrics



Creating Metadata from Raw Image Folders
“Think of DAMPF as your metadata generator — turning folders into searchable databases.”

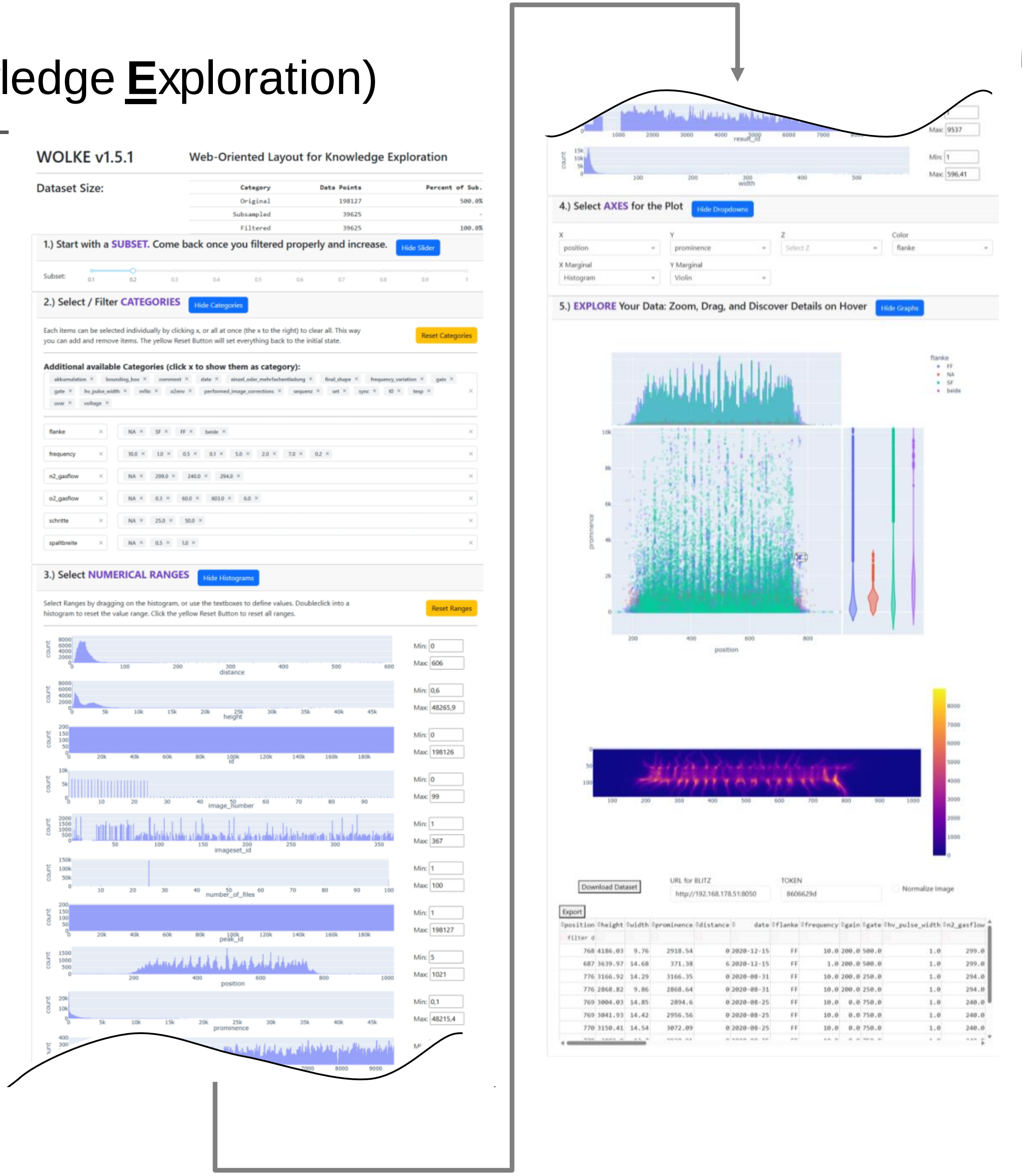
WOLKE (Web-Oriented Layout for Knowledge Exploration)

An interactive web dashboard built with Dash-Plotly for efficient metadata filtering and visualization.

Functionality:

- Explore metadata by type: numeric or categorical
- Create interactive x/y/(z) scatter plots with color-coded metadata

WOLKE helps researchers reduce complexity and focus on relevant subsets – no coding required.



Challenge BLITZ WOLKE Outlook

Subset and Filtering

1.) Start with a Subset:

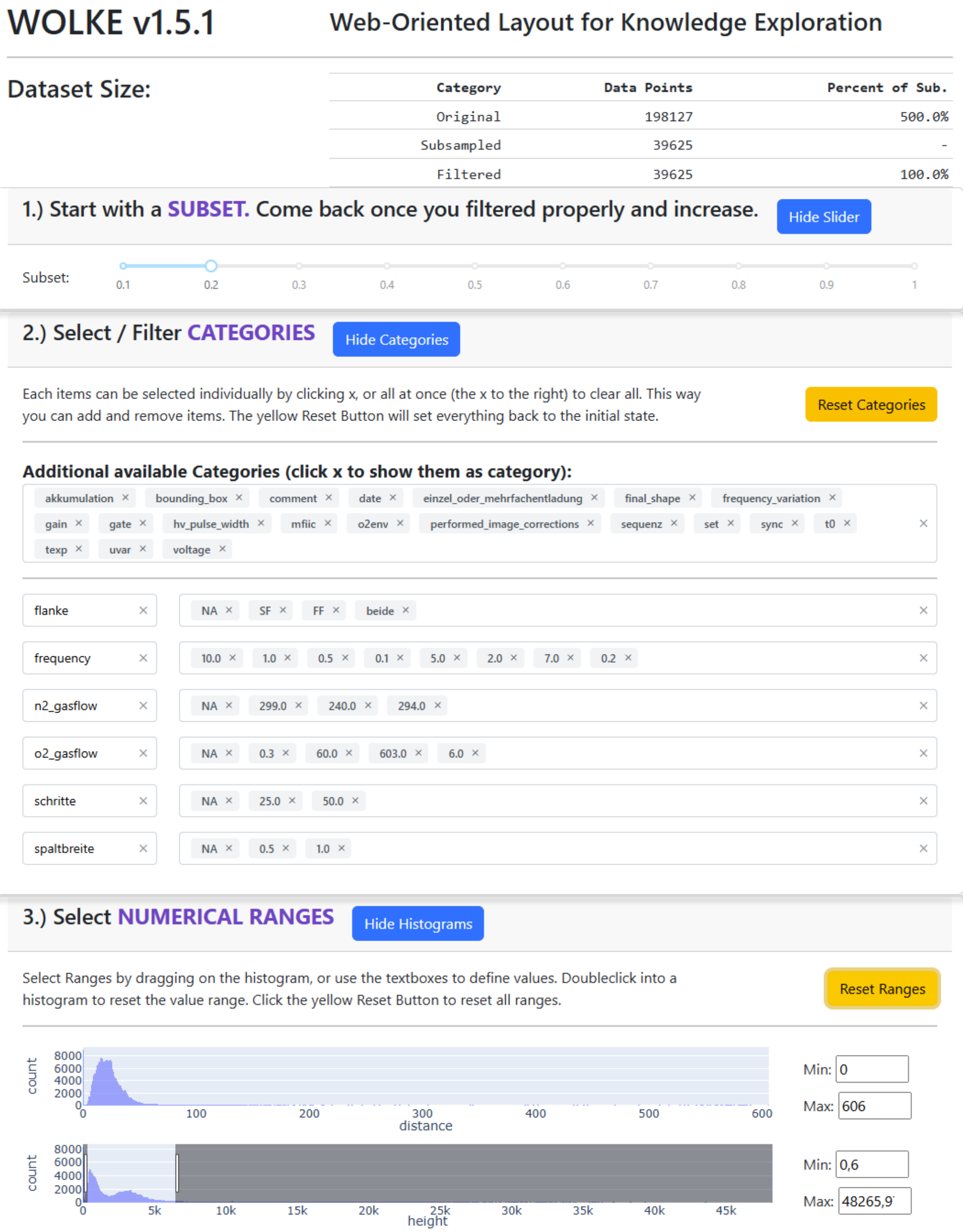
Improves performance on large datasets

2.) Filter Categories:

- Categories = all non-numeric data
- Numeric data with low variance (e.g. frequency, n=8) *can* also be treated as categories

3.) Filter Numerical Ranges:

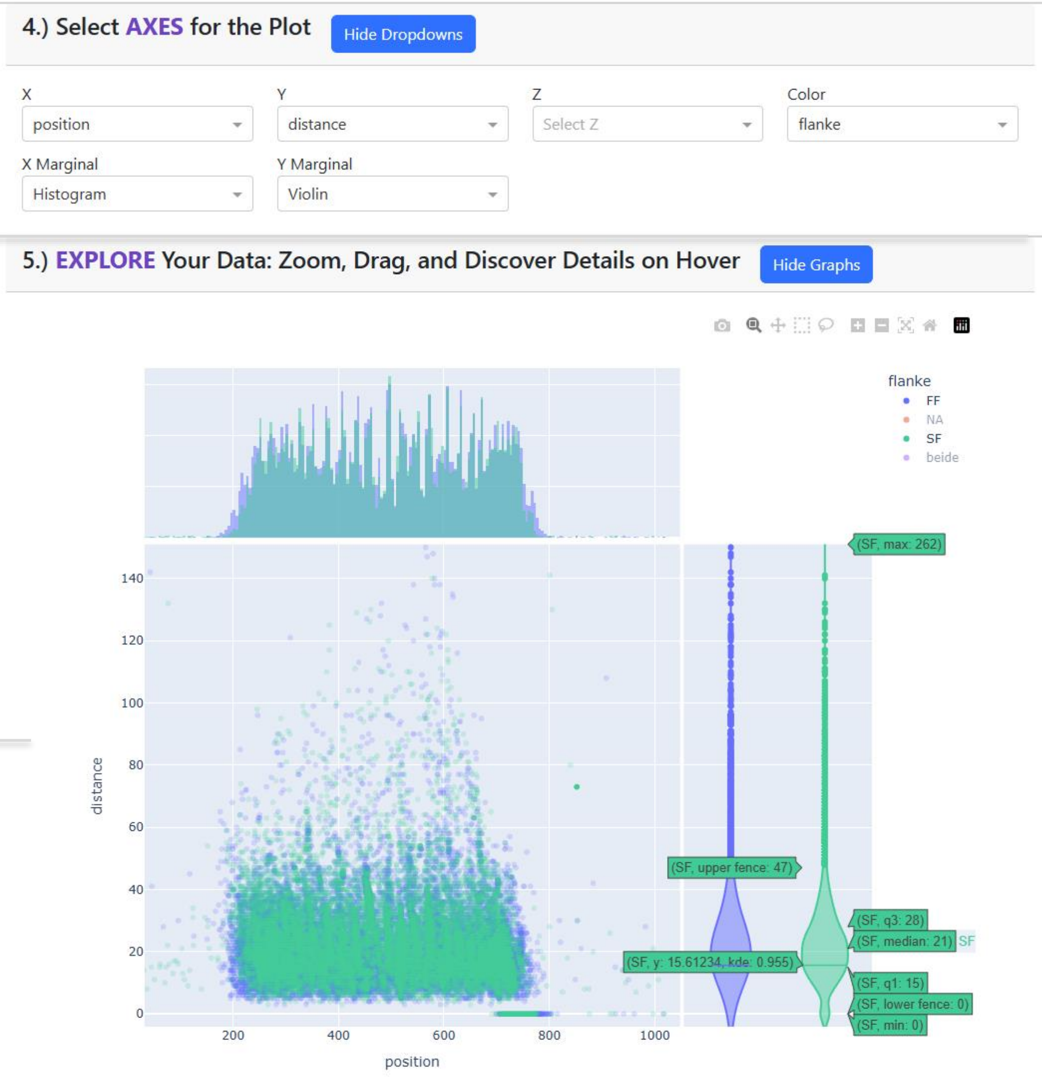
- View histogram of value distribution
- Filter using range sliders or input fields



Visual Data Exploration

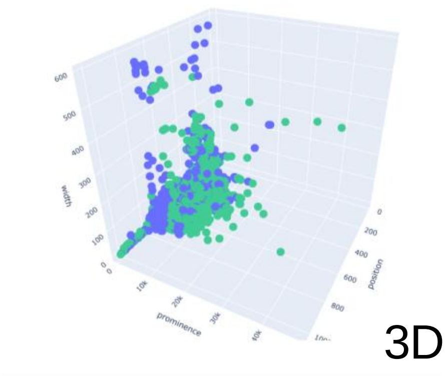
4.) Select Plot Axis:

- 2D or 3D scatter plots
- Optional: Add side plots for statistics (e.g. histogram, violin)

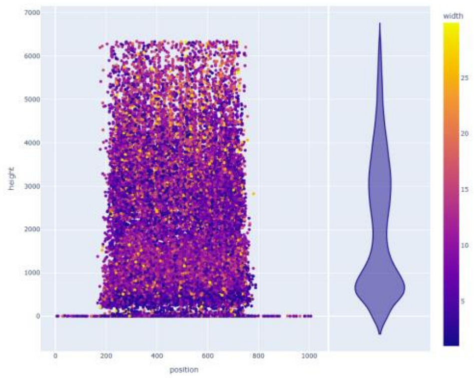


5.) Explore Your Data Visually:

- Zoom, drag, and hover with your mouse
- Show/hide categories to focus the view
- Select single or multiple data points interactively



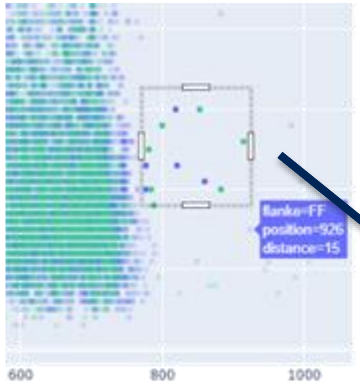
3D Scatter



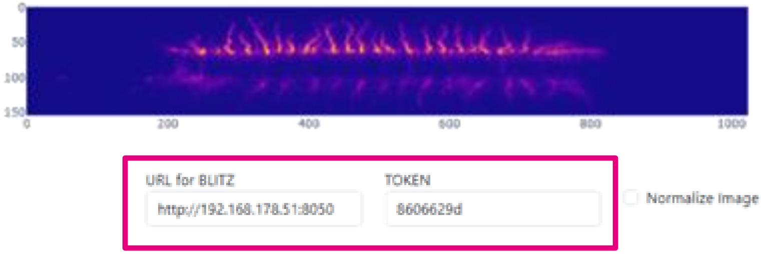
Continous Colorscale

From filtered metadata to deep image analysis — seamlessly.

Select

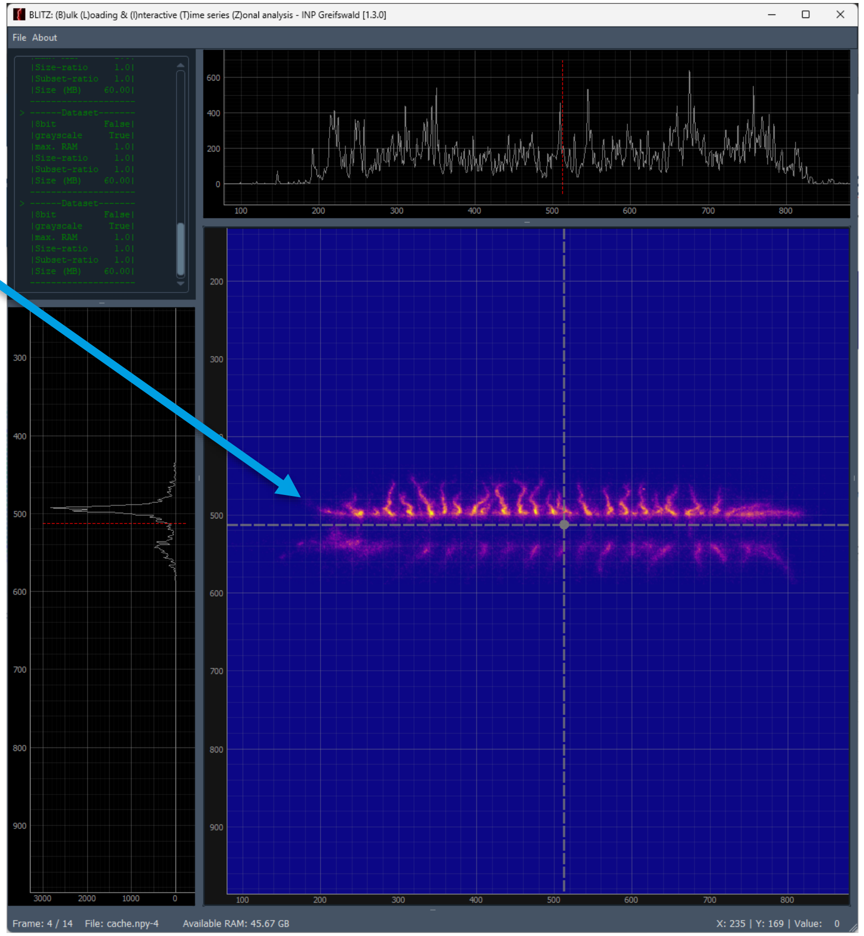


Inspect



position	height	width	prominence	distance	date	flanke	frequency	gain	gate	hv_pulse_width	n2_gasflo
filter d											
669	3077.64	5.73	868.91	18	2020-08-13	FF	10.0	200.0	250.0	50.0	299.
331	2326.88	13.02	2253.48	40	2020-08-25	FF	10.0	0.0	250.0	1.0	240.
854	2.02	4.95	0.24	30	2021-01-13	FF	10.0	200.0	2.0	1.0	299.
778	929.48	3.46	32.51	20	2021-07-14	FF	10.0	200.0	250.0	0.2	299.
778	723.83	4.74	21.58	23	2021-07-14	FF	10.0	200.0	250.0	0.2	299.
524	903.93	22.43	836.69	32	2020-08-25	FF	10.0	200.0	750.0	50.0	240.
362	2357.59	8.26	1876.06	16	2020-08-25	FF	10.0	200.0	750.0	10.0	240.

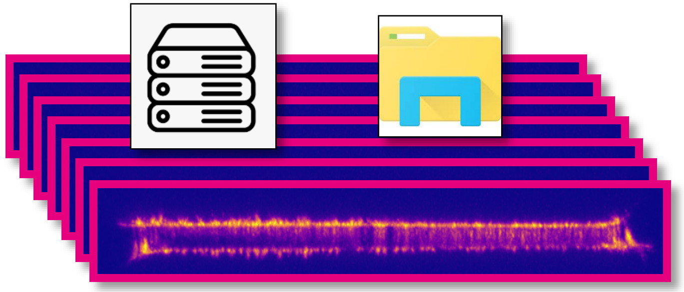
Synchronize with the BLITZ!



WOLKE provides filtered data via HTTP as *.npy arrays,
which can be instantly opened in BLITZ —
or integrated into your own analysis framework.

From Folder to Insight:
Your Workflow with DAMPF, WOLKE & BLITZ

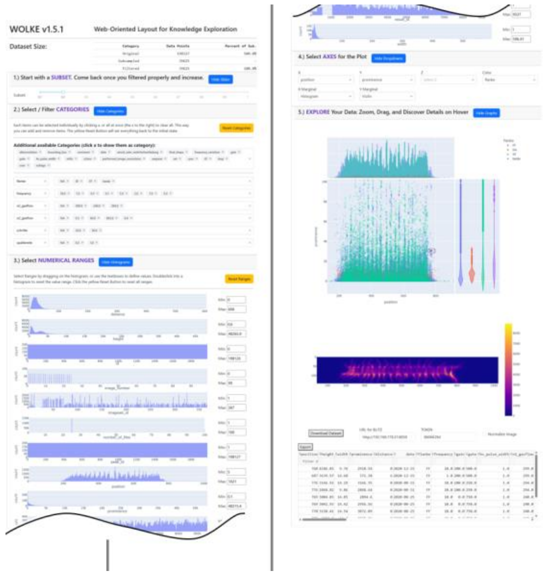
Keep your image data in folders
…just like you're used to!



Augment metadata
with DAMPF

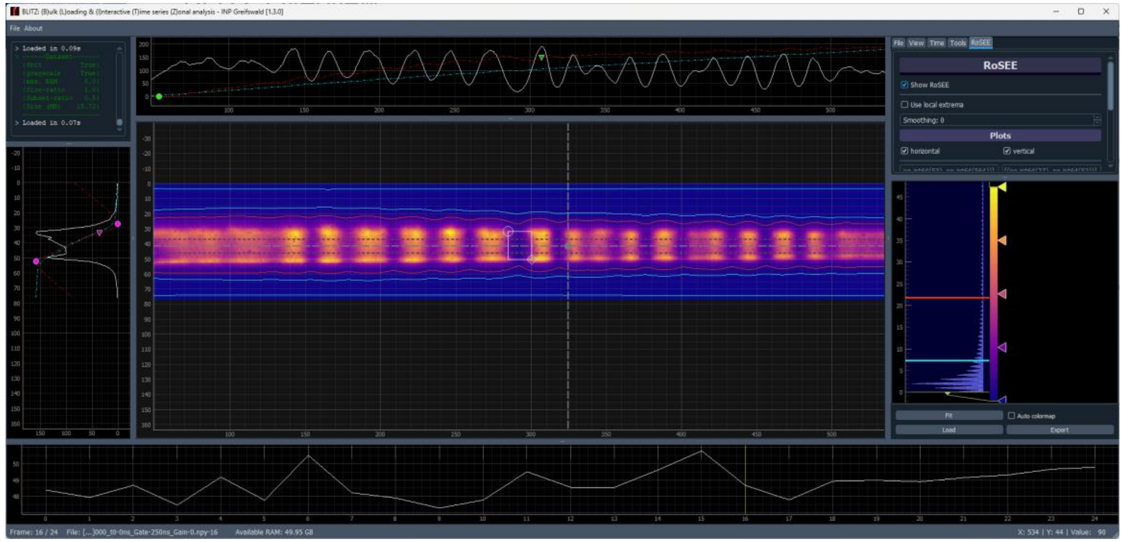
(and store in SQL)

Filter and preselect
using WOLKE



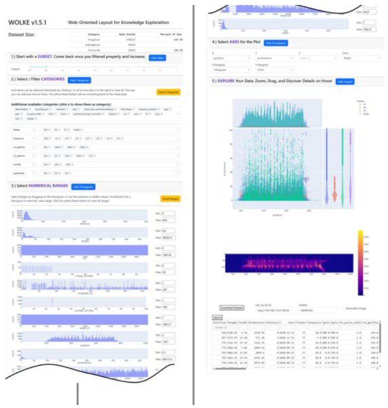
Preview directly
with BLITZ

(Drag & Drop
sneak peek)

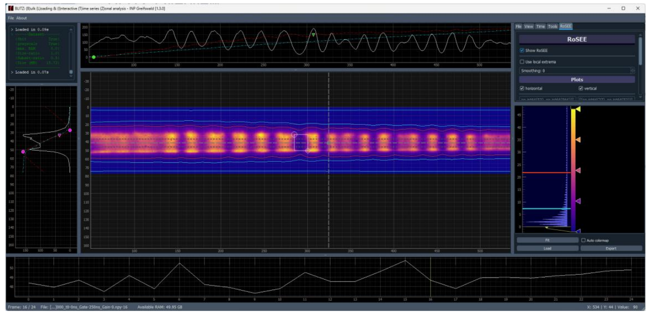


Synchronize filtered
sets with BLITZ for
detailed inspection

Outlook: WETTER Framework

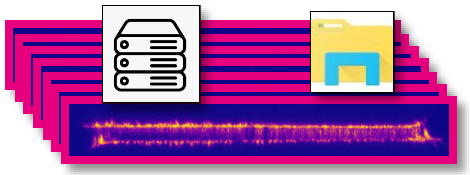


WOLKE



BLITZ

DAMPF



Your Contribution?

HAGEL Get rid of heavy outliers

SCHNEE Solidify trustworthy data

NEBEL Reveal Patterns in the fog of data

REGEN Gather new Data by learning from the old