

METABOLATOR: Establishing a Citable Web Application for Automated Metabolic Load Analysis

David Pape, Mani Lokamani, Ayush Seal, Oliver Knodel, Jeffrey Kelling,
Karim Fahmy, Guido Juckeland



The METABOLATOR Service

A Collaboration Between Two Departments

Institute of Resource Ecology –
Biophysics Department

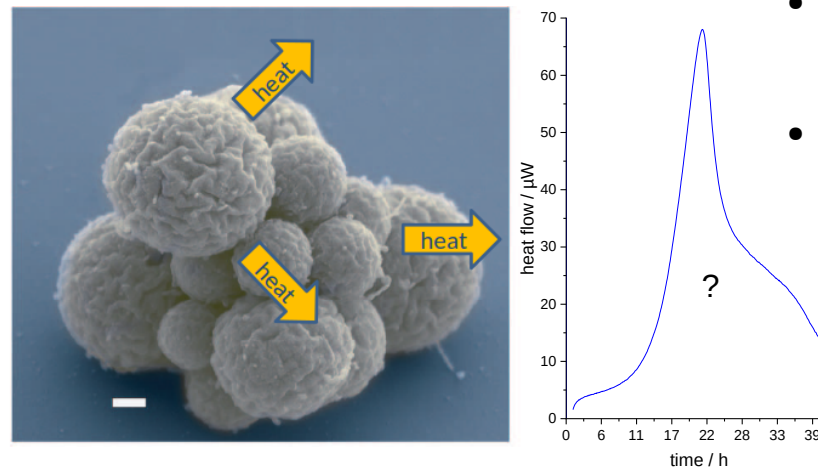
Central Department of Information
Services and Computing –
Computational Science Department

- Publication: Analysis method for growth-metabolism relations in microorganisms based on heat flow patterns
- Idea: Build automated analysis tool and offer as a service for the whole community!
- Motivated student assistants

Mycoplasma JCVI-syn3.0

Genome is smaller than that of any autonomously replicating cell in nature.

Mycoplasma mycoides JCVI Syn 3.0, SEM Stock Image - C029/3010 - Science Photo, John Craig Venter Institute



- Computing infrastructure
- Expertise in tools, frameworks, and approaches for automating analyses
- Research Software Engineering (RSE) → [FAIR4RS](#)

Analysis

Examination of Growth-Metabolism Relations in Microorganisms Using Monod's Equation

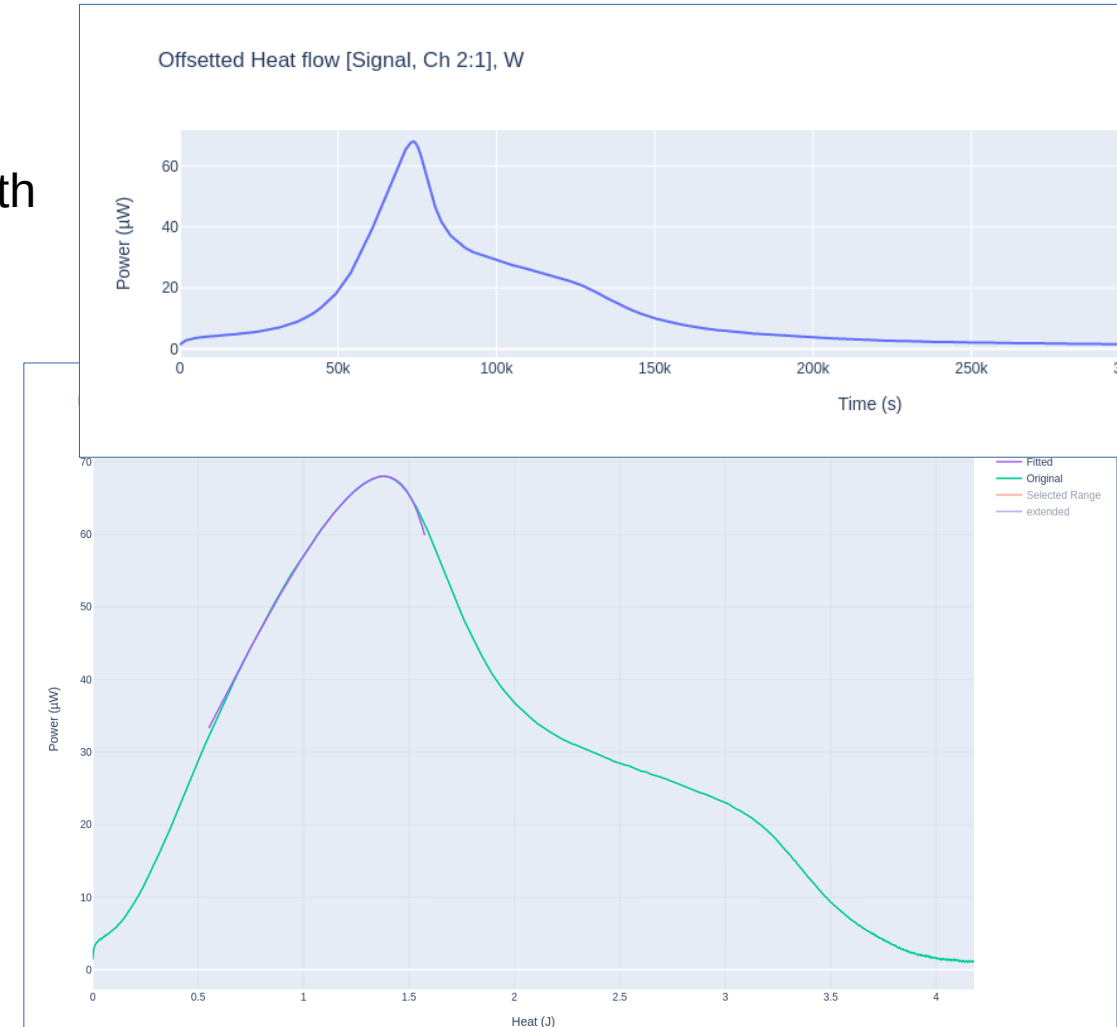
Paper: *Simple Growth–Metabolism Relations Are Revealed by Conserved Patterns of Heat Flow from Cultured Microorganisms*

- Calorimetric measurements of heat flow $\mathbf{P(t)}$ during growth
- Interpretation in enthalpy domain (released heat, \mathbf{H})
 - Thermal metabolic power of the culture $\mathbf{P(H)}$
- Model: "Metabolic Load" Θ

$$\Theta(t) = \frac{H_0 - H(t)}{H_0 - H(t) + H_f} \quad P(H) = r_0 \cdot H \cdot \Theta(H)$$

Find characterizing values H_0 , H_f , r_0 !

→ Automated fitting procedure



The METABOLATOR Tool

Required Features

- Web-based for ease of use
- Users can upload own datasets
- Offsets and fitting interval can be applied (manually and based on heuristics)
- Automated analysis
 - Find parameters **H0**, **Hf**, **r0**, standard errors, goodness of fit **R²**
 - Robust method to fit around pole
 - Optimization for good initial values
- Download of fitting results

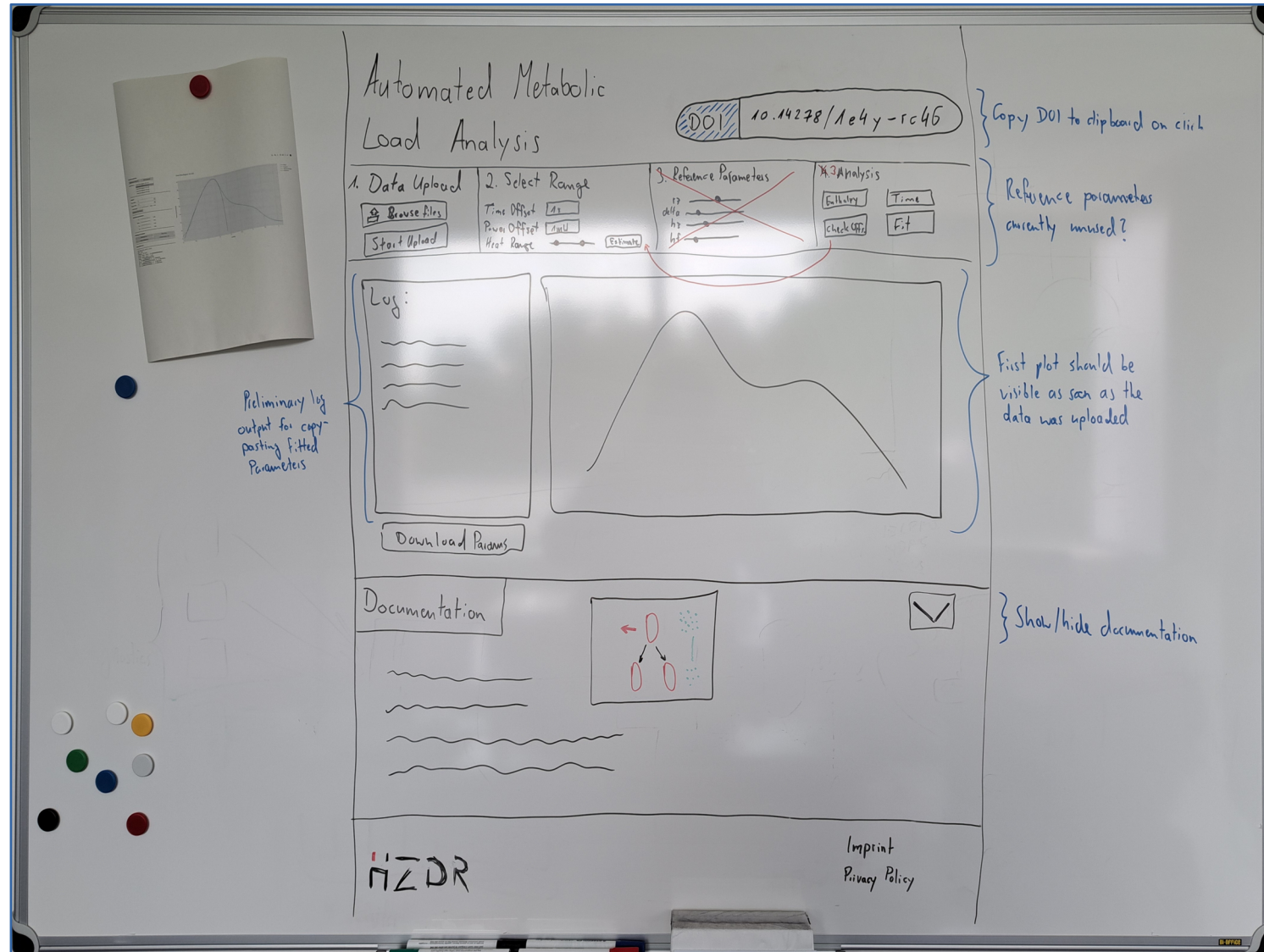
Typical Dataset:

- Column-wise time series data of heatflow
- Duplicate and triplicate of each measurement
- 1 measurement/minute, over days or weeks (e.g. 7500 points in 5 days)
- Stored in Excel file

	A	B	C	D	E	
1	time	Heat flow [Signal, Ch 2:1]	Heat flow [Signal, Ch 2:2]	Heat flow [Signal, Ch 2:3]	Heat flow [Signal, Ch 2:4]	Heat flow [Signal, Ch 2:5]
2	s	W	W	W	W	W
3		Syn3B_2FAs	Syn3B_2FAs_Duplicate	Syn3B_2FAs_Triplicate	Syn3B_2FAs_PC	Syn3B_2FAs
4		37°C / 2ml	37°C / 2ml	37°C / 2ml	37°C / 2ml	37°C / 2ml
5	-6960	-9,86667506292738E-08	-1,69051123323025E-07	-8,1295267672065E-08	-1,77927968441883E-07	-1,342
6	-6900	-1,85743016332254E-07	-1,84220683084626E-07	-1,32847276638315E-07	-2,0608992626825E-07	-1,873
7	-6840	-2,37949186337674E-07	-1,84839289581763E-07	-2,83994289982168E-07	-2,21582614303881E-07	-2,450
8	-6780	-2,1865826951769E-07	-1,52933497187425E-07	-1,50325145614491E-07	-2,06748142321654E-07	-2,63
9	-6720	-1,95811071652537E-07	-1,08222999430804E-07	-7,71643618295661E-08	-1,61642340596241E-07	-2,50
10	-6660	-1,71975871374406E-07	-5,84497756593351E-08	-7,37057995199989E-09	-1,08460427883958E-07	-2,145
11	3720	1,39549505821375E-06	1,61664633472963E-06	3,23753054399768E-06	1,14167490532101E-06	4,734
12	3780	1,44865270109709E-06	1,65725700086494E-06	3,30790416882512E-06	1,21818361122779E-06	5,333
13	3840	1,51951238212344E-06	1,70470546320501E-06	3,36824452585475E-06	1,29637368049027E-06	6,360
14	3900	1,58257619579815E-06	1,7447962256083E-06	3,43607472559441E-06	1,38323982271412E-06	7,178
15	3960	1,65878620247046E-06	1,80086140287304E-06	3,50594332976722E-06	1,46550563854996E-06	7,910
16	4020	1,7352167420061E-06	1,86474723553831E-06	3,57083278069483E-06	1,55111009589155E-06	8,643
17	4080	1,78771105135551E-06	1,91595882399202E-06	3,62380355999714E-06	1,65391611512222E-06	9,433
18	4140	1,84576293477997E-06	1,95984934557402E-06	3,67150451779557E-06	1,7419617723609E-06	1,001
19	4200	1,91348691862888E-06	1,99509438403733E-06	3,68178412508801E-06	1,80709271026174E-06	1,072
20	4260	1,96800272597607E-06	2,0321803614067E-06	3,69884533401322E-06	1,86946259576887E-06	1,142
21	4320	2,01085921903147E-06	2,07682035782222E-06	3,73171186653196E-06	1,9336757970916E-06	1,208
22	4380	2,05644950361146E-06	2,13127322910157E-06	3,77290291926367E-06	1,99714117768973E-06	1,275
23	4440	2,11218356706759E-06	2,17539688960765E-06	3,84289603108203E-06	2,06491085131864E-06	1,331
24	4500	2,16448708779489E-06	2,20305636848728E-06	3,88486574978779E-06	2,13089910813652E-06	1,401
25	4560	2,19901678090044E-06	2,23623664999611E-06	3,91807474679179E-06	2,18611038672889E-06	1,465
26	4620	2,22557410758708E-06	2,27652288531126E-06	3,94033573773269E-06	2,23153380410193E-06	1,516

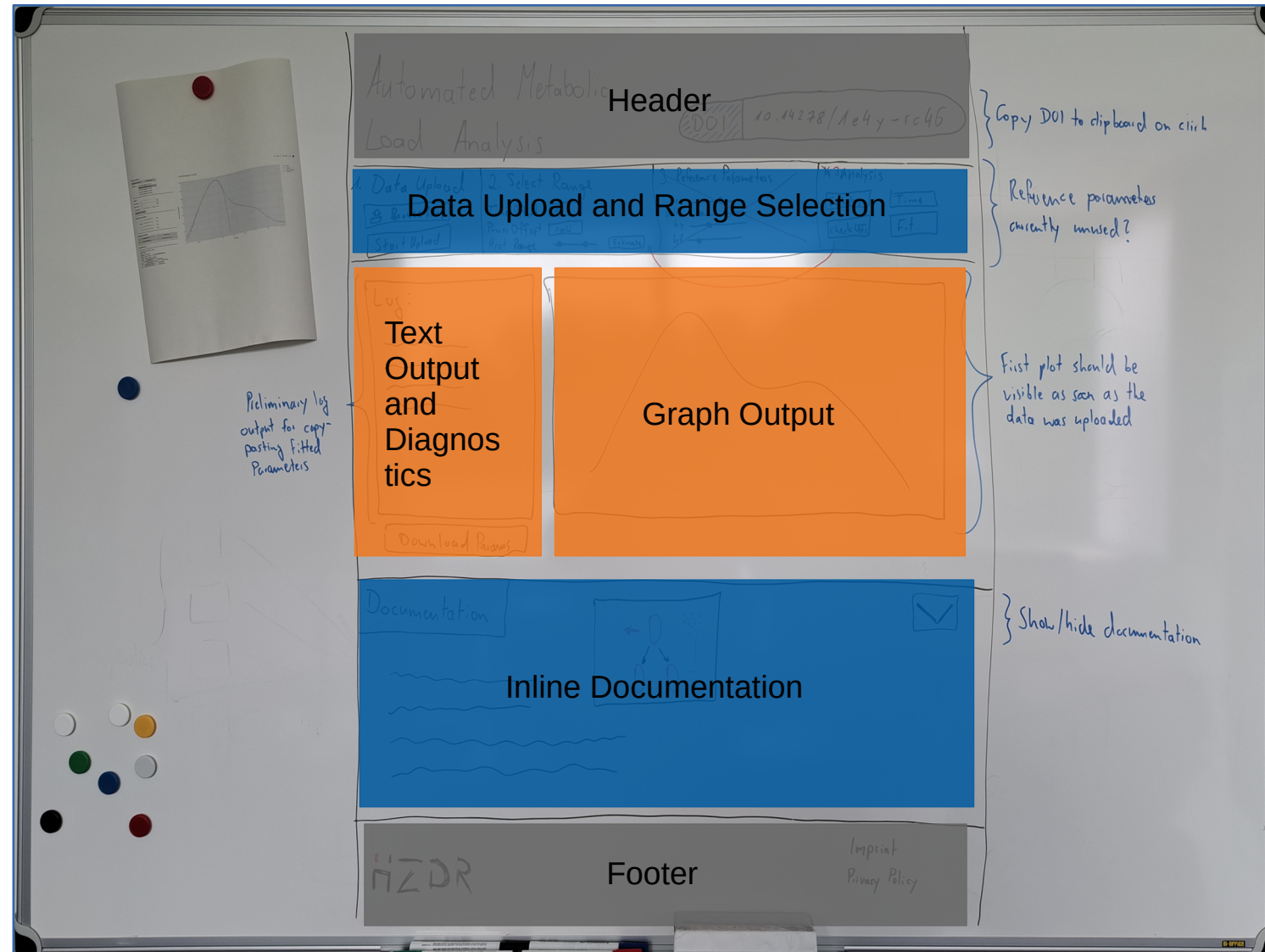
The METABOLATOR Tool

UI Mockup



The METABOLATOR Tool

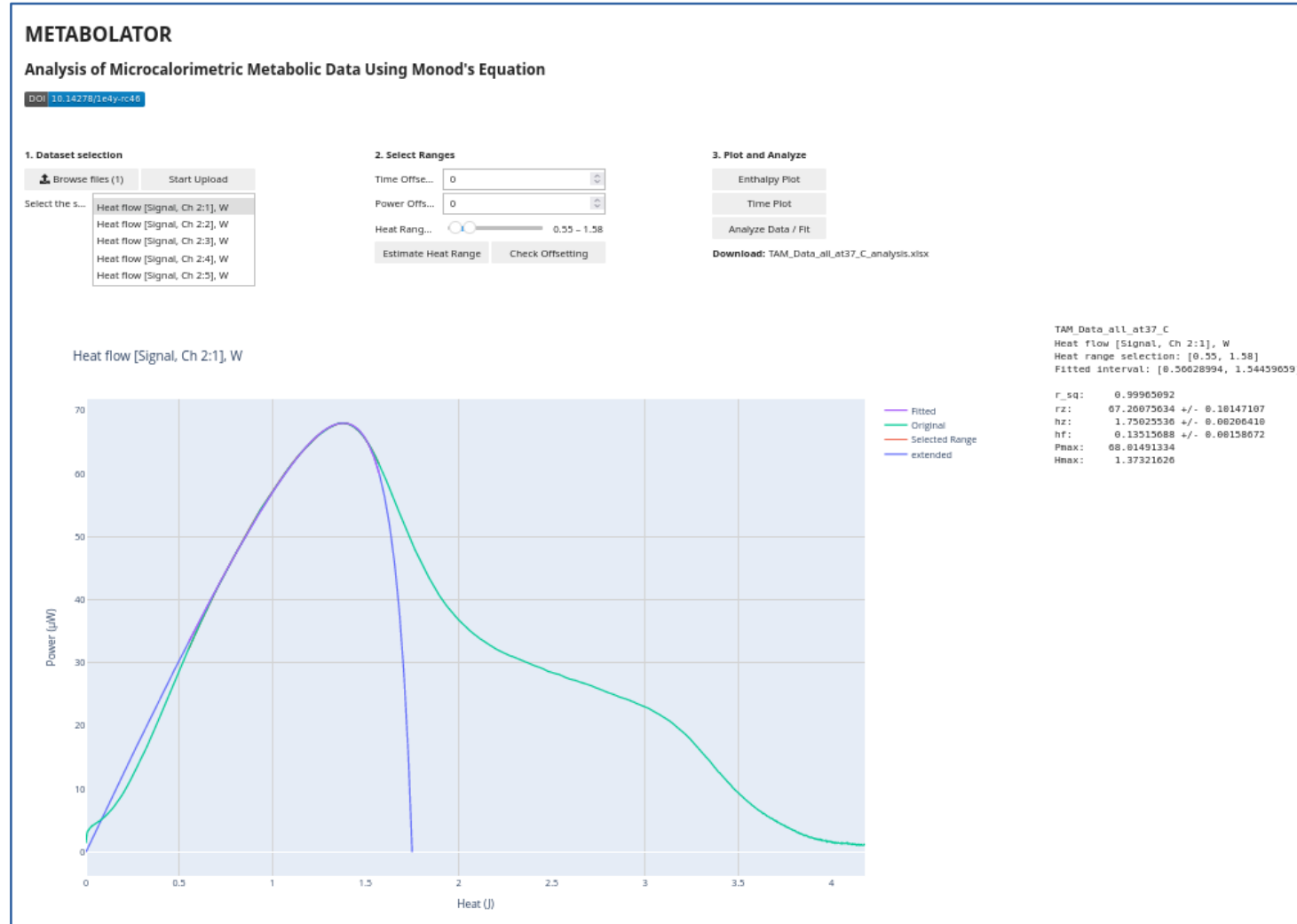
UI Mockup



The METABOLATOR Tool

Current Implementation

NumPy,
pandas,
SciPy,
Imfit



Jupyter Notebook,
Jupyter Widgets
(ipywidgets),
Plotly,
Voilà

The METABOLATOR Tool

Towards a Publication/Production-Ready Software — Application of (R)SE Expertise

- Code cleanup and structuring, defining interfaces (FAIR4RS: R)
 - Modules: `app.py`, `fit.py`, `plots.py`, `files.py`, `METABOLATOR.ipynb`
- Installable Python package (R), reproducible dependencies (R2)
- Repository cleanup
 - Clean environment is more attractive to collaborators
 - Copyright notice, license attached to all files (R1.1), no files or code with unknown authorship!
- CI/CD pipeline runs tests and code analysis to maintain code quality
- More open to-dos (more features, faster fitting, improved data model, ...)

R: *Software is both usable [...] and reusable (can be understood, modified, built upon, or incorporated into other software).*

R1.1: *Software is given a clear and accessible license.*

R2: *Software includes qualified references to other software.*

Clearly Defined Input and Output Formats

Can We Initiate Discussion and Exchange in the Community?

- Community standards are needed for data exchange and common analysis codes (FAIR4RS: I1)
- Currently: Input and output via Excel files; structure “free form“
- Metadata of interest

Dataset

unit of measurement
time scale
end of calibration
association between
replicated measurements

Lab environment

experimentalist/lab
assistant
calorimeter make
and model

Experiment

microbiological
culture
Substrate
temperature

I1: *Software reads, writes and exchanges data in a way that meets domain-relevant community standards.*

The METABOLATOR Service

DOI 10.14278/1e4y-rc46

A Citable Web Service

- Set up in virtual machine in HZDR data center
 - Currently only available internally (final tweaks necessary; legal requirements, e.g. privacy policy)
- Citable via a DOI (FAIR4RS: F1, F2)
 - resourceTypeGeneral: Service (DataCite)
 - Establish relation: Service isSupplementTo software publication
 - Makes publications more easily reproducible
 - Academic credit for providers of software and service

F1: *Software is assigned a globally unique and persistent identifier.*

F2: *Software is described with rich metadata.*

Outlook / Timeline

Planned Publications and Conference Contributions

- Software publication (summer)
- Reference dataset publication (summer)
- Talk/poster at [New Opportunities for Better User Group Software \(NOBUGS\)](#), ESRF (September)
- Hand-over of productive service (before October)
- Talk at [ISBC 2024](#), Mexico City (October), presentation of service to biological calorimetry community
- [JOSS](#) publication or RSE-focused paper
 - More in-depth version of this pitch talk
 - Assessment: "How FAIR is the service?"

Thank you for your attention!