Machine learning supported image analysis of microfluidic droplets: Using Random Forest classifiers and Bayesian inference for identification of experimental conditions

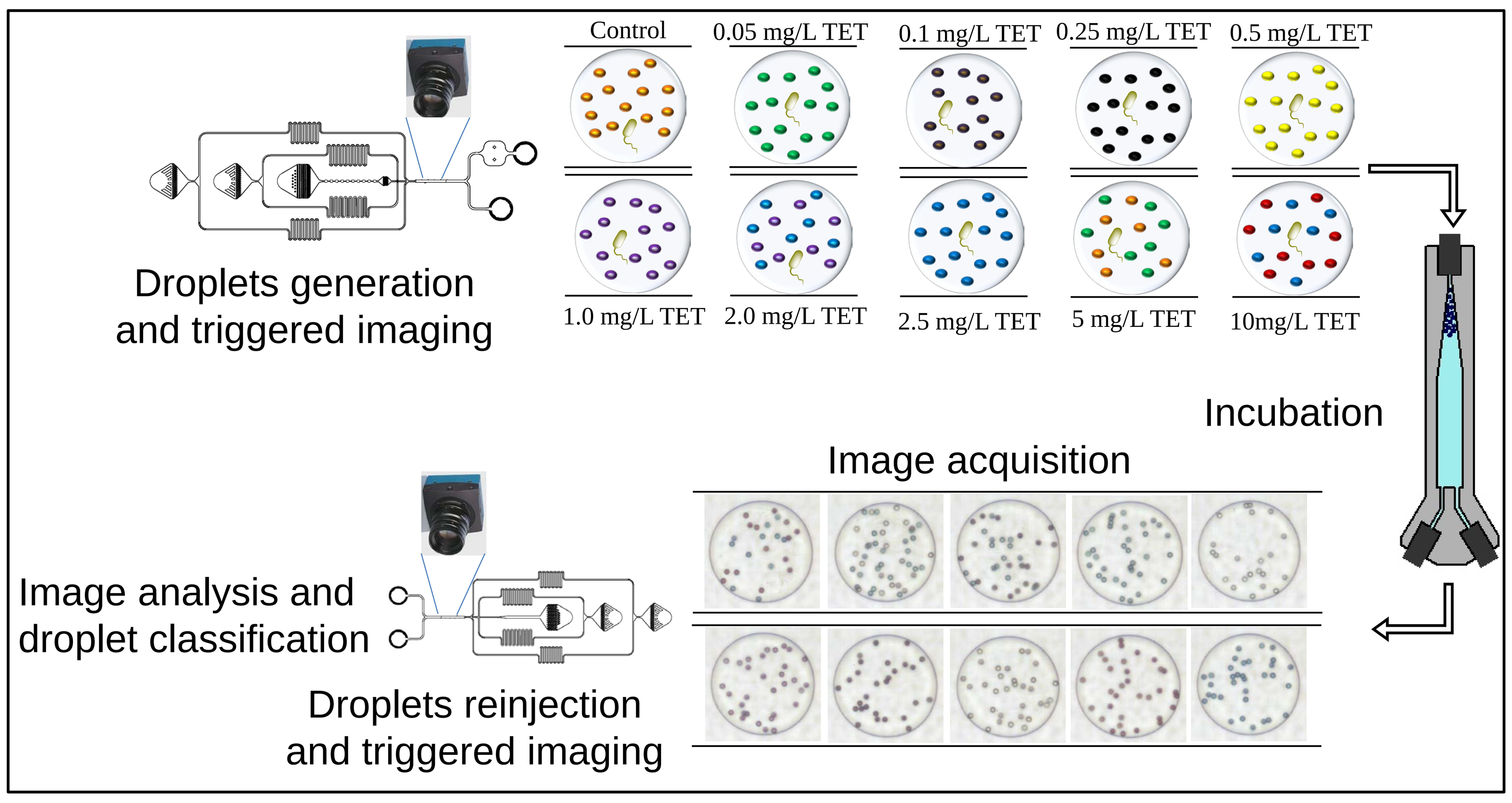
Carl-Magnus Svensson[1],Oksana Shvydkiv[2], Lisa Mahler[2,3], Stefanie Dietrich[1,3],Thomas Weber[2,3], Mahipal Choudhary[2], Miguel Tovar[2,3], Martin Roth[2] and Marc Thilo Figge[1,3]

1 Applied Systems Biology, Hans Knöll Institute, Jena, Germany. 2 Bio Pilot Plant, Hans Knöll Institute, Jena, Germany. 3 Friedrich Schiller University, Jena, Germany

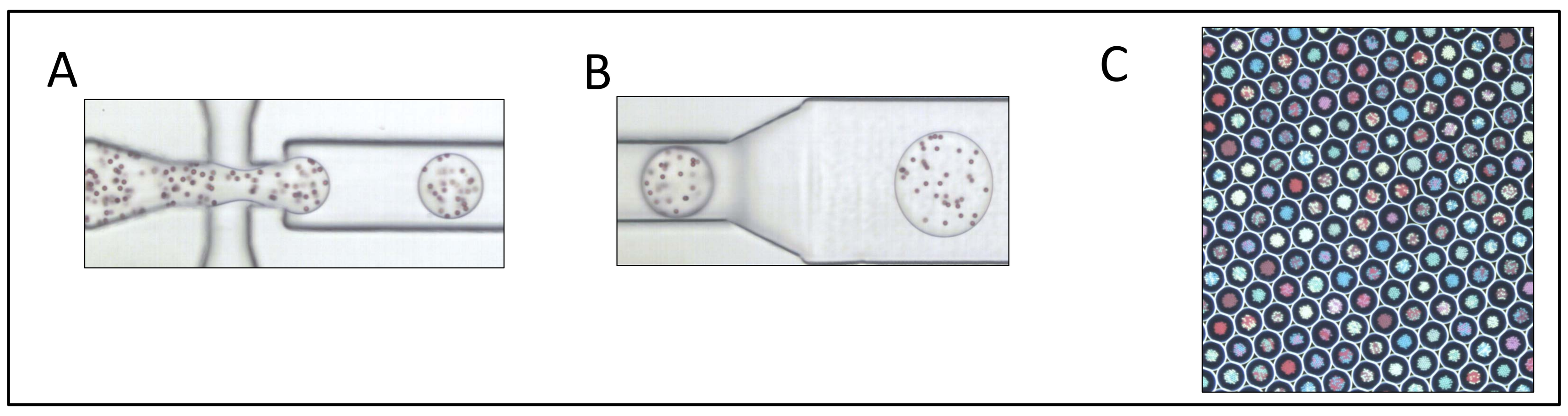
Leibniz-Gemeinschaft

1. Experimental setup

- Coincubation of *Escherichia coli* and antibiotics in 300 pL droplets
- Encoding experimental conditions by combinations of bead colors
- Analysis of droplet triggered images
- Up to 20 codes per experiment tested so far, theoretically 72 codes possible



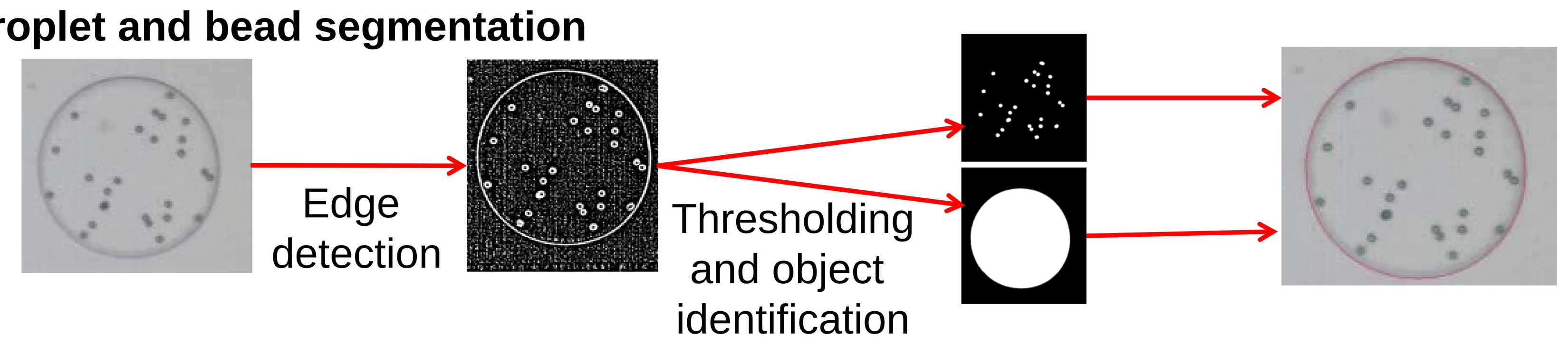
Overview of the experimental setup exemplified by ten codes encoding ten concentrations of the antibiotic Tetracycline hydrochloride (TET).



(A) Snapshot of droplet generation. (B) Droplets entering the 20 µm high imaging channel. (C) Phase contrast image of droplets in the incubator.

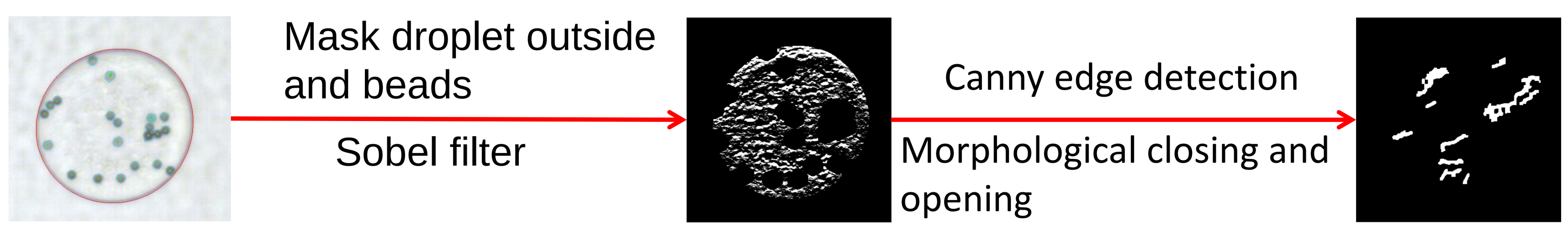
2. Segmentation

Droplet and bead segmentation



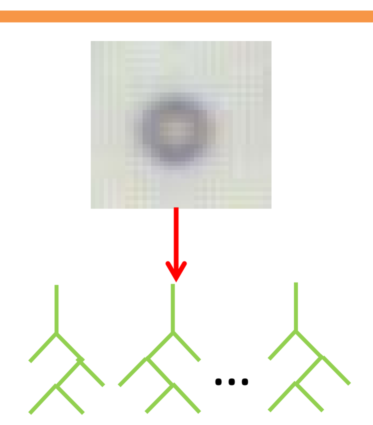
1989 beads: $Acc = 0.979$, $Rec = 0.983$ and $Pre = 0.996$

Growth detection



3000 droplets: $Acc = 0.925$, $Rec = 0.788$ and $Pre = 0.992$

3. Bead classification



C_T: True bead color
C_D: Bead color detected by RF
RF are resistant to faulty labels in training data [1]

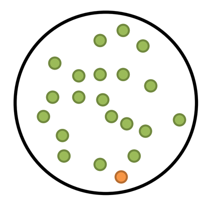
- 8 bead colors
- 11x11 pixels
- Lab colour space
- Random forest (RF) with 5000 trees
- Implemented in Python using *scikit-learn*

C_D \ C_T	black	blue	red	orange	green	yellow	dark purple	purple
black	0.95	0.0028	1e-06	0.00034	0.023	0.015	0.00014	0.00011
blue	0.00029	0.98	1.9e-05	2.8e-05	0.0047	5.4e-05	1e-06	6.7e-05
red	1e-06	5.4e-05	0.96	0.00012	0.00028	1e-06	0.012	0.0022
orange	0.00026	5.4e-05	0.0013	0.99	0.00036	0.011	0.0039	0.0014
green	0.019	0.016	1.9e-05	0.0016	0.97	0.0015	2.8e-05	0.00011
yellow	0.034	0.00011	0.00015	0.0077	0.0012	0.97	0.00041	0.00021
dark purple	3.4e-05	0.00011	0.035	0.00094	7e-05	0.00011	0.98	0.0028
purple	0.00012	3.6e-05	0.00088	0.00015	1e-06	1e-06	0.0026	0.99

4. Droplet decoding

- Each droplet has a true code (K_T) and the RF gives a detected code (K_D)

$$K_D = \{C_{D,1}, \dots, C_{D,N_D}\}$$



$$K_D = \{g, o\}$$

- We are looking for the most likely code given:
  - o Bayes theorem

$$p(K_T|K_D) \sim p(K_D|K_T)p(K_T)$$

$$\boldsymbol{v}_{K_D} = [k_1, k_2] = [22,1]$$

$$K_T = \{g,o\}, \{g\} \text{ or } \{o\}$$

N_D: Number of detected colors in a droplet

  - o The misclassification rate of the RF

$$p(K_D|K_T) = \prod_{j=1}^{N_D} p(C_{D,j}|C_{T,j})^{\boxed{k_j}}$$

Number of beads of color j

  - o A multinomial distribution with equal probability for each color

$$p(K_T) = p_{MN}(\boldsymbol{v}_{K_T}) = \frac{\sum_{j=1}^{N_T} k_j}{\prod_{j=1}^{N_T}(k_j!)}\left(\frac{1}{N_T}\right)^{\sum_{j=1}^{N_T} k_j}$$

$$p(\{g\}) = p(g|g)^{22}p(o|g)^{1}p_{MN}([23]) \approx 7.98 \cdot 10^{-5}$$ Most probable code

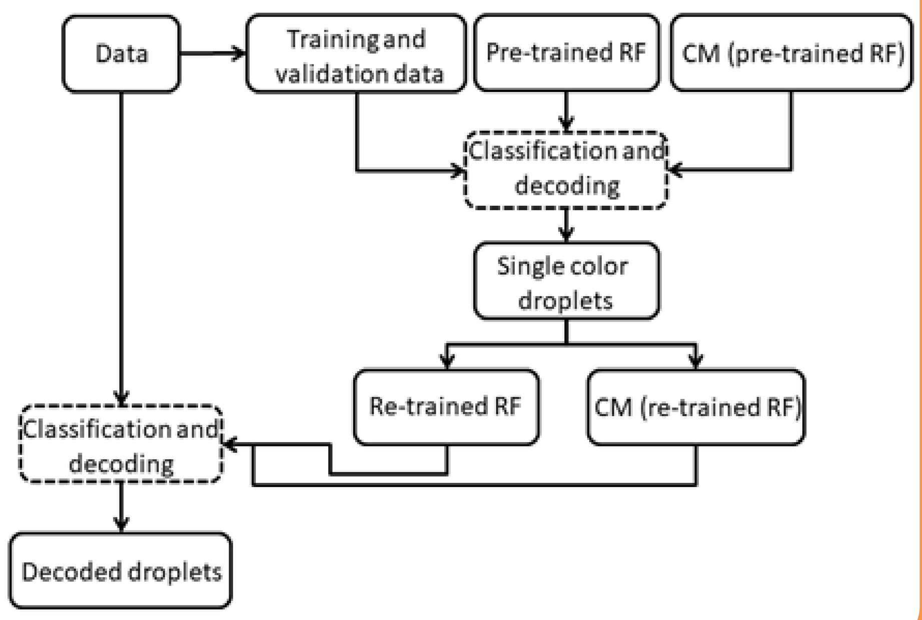
$$p(\{o\}) = p(g|o)^{22}p(o|o)^{1}p_{MN}([23]) \approx 1.82 \cdot 10^{-76}$$

$$p(\{g,o\}) = p(g|g)^{22}p(o|o)^{1}p_{MN}([22,1]) \approx 1.39 \cdot 10^{-6}$$

- Adapting to each new dataset
- Automatic re-training of the RF
- Single color droplets identified

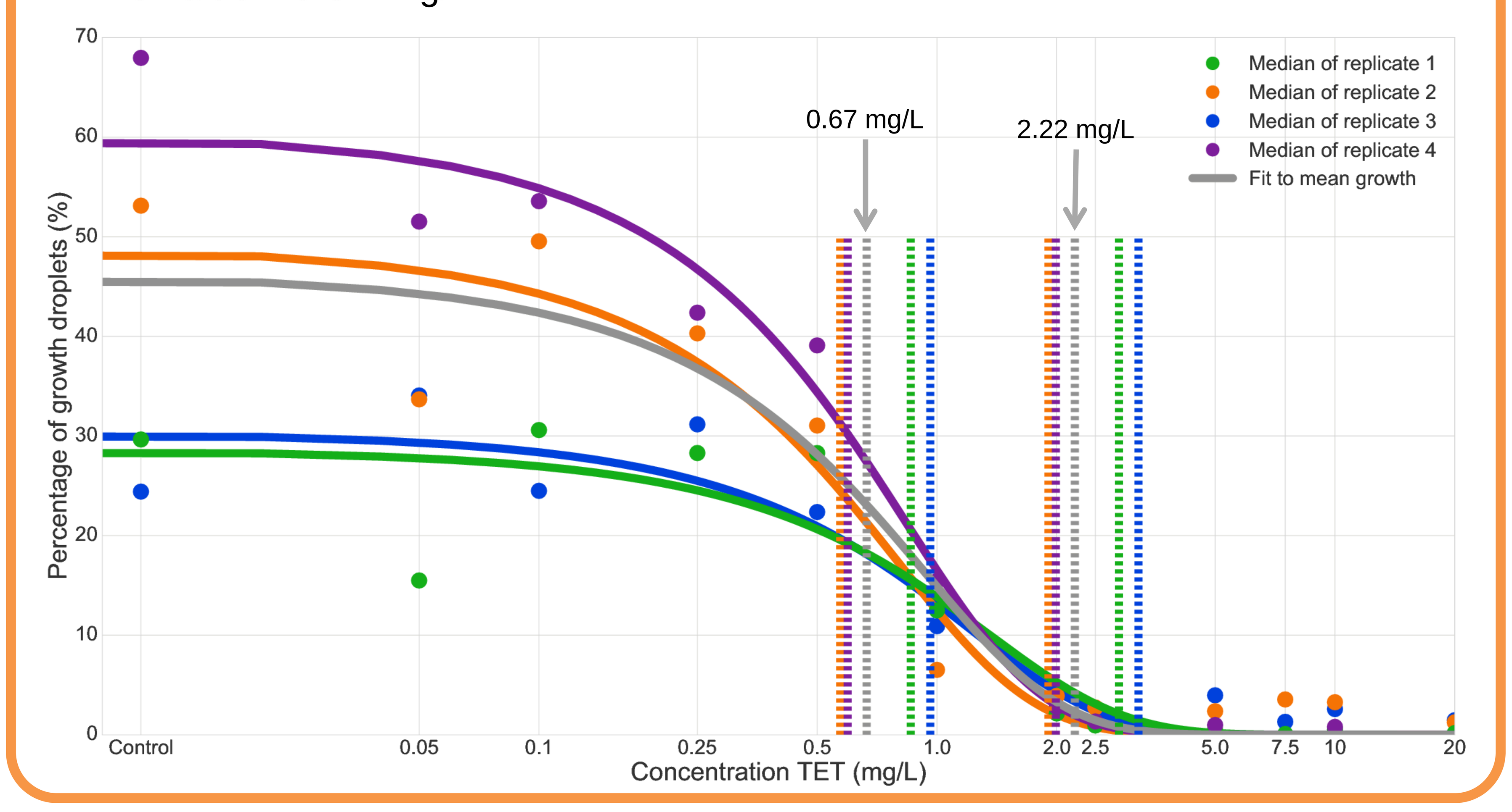
	Acc, pre-trained RF	*Acc*, re-trained RF
Generation	0.94	0.97
8 h incubation	0.92	0.99

- 4259 droplets manually validated
- 18 were wrongly decoded (<0.5%)



5. Minimum inhibitory concentration

- Known MIC of TET versus *E. coli* is ~1-2 mg/L [3]
- Three replicates with 12 concentrations, one with 2x10 concentrations
- Percentage of droplets with growth as antimicrobial susceptibility indication
- Fit $G([C]) = 2A\left(1 - \frac{1}{1+e^{-\alpha[C]}}\right)$ to the growth data, the 50% and 5% levels are used as MIC limits
- In all individual replicates lower MIC value were in range 0.5-1.0 mgL and the upper MIC value 1.9-3.2 mg/L



6. Conclusions

- Method able to encode and decode at least 20 experimental conditions
- No need for fluorescent dyes and laser illumination that may interfere with biological activity
- To be merged with current growth detection and sorting algorithm [3]

References
[1] Svensson *et al.*, (2015) Automated classification of circulating tumor cells and the impact of interobserver variability on classifier training and performance. *Journal of Immunology Research,* ID 573165.
[2] Andrews, (2001) Determination of minimum inhibitory concentrations. *Journal of Antimicrobial Chemistry,* 48:5-16.
[3] Zhang, Brandes *et al.*, (2013) Real-time image processing for label-free enrichment of Actinobacteria cultivated in picolitre droplets. *Lab on a Chip,* 13(18):3707-13.

Email: carl-magnus.svensson@leibniz-hki.de

Funded by Thüringer Aufbaubank through the project Dropcode

Thüringer Aufbaubank