WELCOME TO THE

MultiplEYE

DATA COLLECTION

Nora Hollenstein, Action Chair
nora.hollenstein@ktu.lt

Lena Jäger, Vice-Chair
jaeger@cl.uzh.ch

MultiplEYE

Agenda

1. **What is MultiplEYE?**

2. **The MultiplEYE data collection**
   - Stimulus materials
   - Psychometric tests
   - Procedure
   - How to contribute?

3. **Your questions & discussion**

The MultiplEYE COST Action

MultiplEYE is an EU COST Action funded by COST and the European Union.

It provides funding for **networking activities:**

- Working group meetings
- Training schools
- Short-term scientific missions
- Conference grants & publications costs

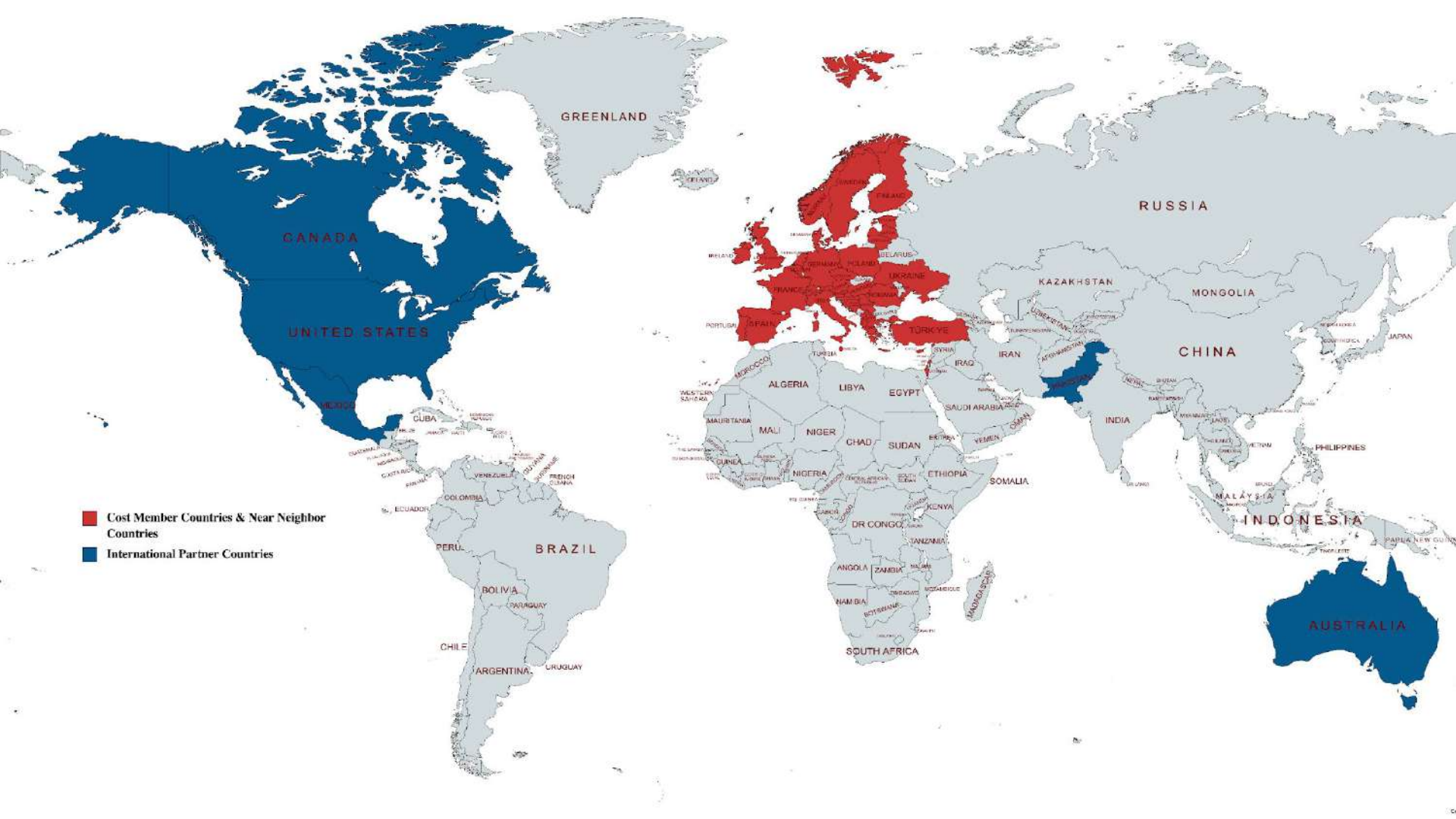
Funded by the European Union

cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

CA21131

MultiplEYE aims to foster an **interdisciplinary network** of research groups working on **collecting eye tracking data** from **reading in many languages**.

WG 1: Data collection
WG 2: Eye-tracking methodology
WG 3: Psycholinguistics
WG 4: Natural language processing
WG 5: Dissemination


GREENLAND
CANADA
UNITED STATES
RUSSIA
KAZAKHSTAN
MONGOLIA
CHINA
JAPAN
IRAN
IRAQ
TÜRKIYE
SYRIA
SAUDI ARABIA
INDIA
PHILIPPINES
MALAYSIA
INDONESIA
AUSTRALIA
ALGERIA
LIBYA
EGYPT
MALI
NIGER
CHAD
SUDAN
ETHIOPIA
SOMALIA
NIGERIA
KENYA
DR CONGO
TANZANIA
ANGOLA
ZAMBIA
NAMIBIA
SOUTH AFRICA
CUBA
BRAZIL
COLOMBIA
VENEZUELA
PERU
BOLIVIA
PARAGUAY
CHILE
ARGENTINA
URUGUAY
Cost Member Countries & Near Neighbor Countries
International Partner Countries

The MultiplEYE data collection

A large-scale multi-lingual multi-lab eye-tracking-while reading experiment

Unique features:

- Diversity of the original languages included in the stimulus texts
- Diversity in text genres included in the stimulus texts
- Diversity in participating languages
- Rigor of comprehension questions design & evaluation
- Acceptability of a wide range of eye-trackers & low-cost alternatives
- Comprehensive psychometric assessment of individual differences
- Use cases in different fields (psycholinguistics, natural language processing, …)
- User-friendly dedicated data repository
- Novel standards for transparency, reproducibility and data quality reporting in eye-tracking research

The MultiplEYE data collection

Core principles:

- Consistency
- Open Access
- Diversity
- Quality Reporting

The MultiplEYE data collection

- **Consistency across languages and labs**
  - Same experiment design and protocol, natural reading corpus
  - Same population: 100 healthy, L1 adults per language
  - Same pre-processing

Consistency

→ Keep the **data as consistent as possible**

→ But **allow for as much flexibility as possible** when it comes to devices

- Different languages, scripts, eye trackers, sampling rates, etc.

Diversity

The MultiplEYE data collection

Session 1

- 10 stimulus texts + comprehension questions
- At least 1 obligatory break
- Participant questionnaire

approx. 90 minutes

At least 30 minutes in between, better on two different days

Session 2

- Psychometric tests (optional, but strongly encouraged)
- Stroop, Flanker, Lewandowsky Working Memory Battery, RAN, PLAB, WikiVocab

approx. 60 minutes

For details see the Data Collection Guidelines

Genre	Title	Short Description	Orig. Lang.
Popular Science	Welcome to MultiplEYE!	Popular science text describing the MultiplEYE project	EN
	Caveman	Swarthy blue-eyed caveman revealed	EN
Institutional	Universal Declaration of Human Rights - Preamble	Preamble of the Universal Human Rights Declaration	-
	Report from the Commission to the Council: Progress report on a Learning Mobility Benchmark	Progress report on a Learning Mobility Benchmark	-
Literary	The Alchemist - Chapter 1	Brazilian fantasy novel by Paulo Coelho	PT
	The Magic Mountain - Foreword	Fictional novel by Thomas Mann	DE
	Solaris - Chapter 2: The Solarists	Science fiction novel by Stanisław Lem	PL
	Broken April - Chapter 3	Fictional novel by Ismail Kadare	SQ
Argumentative	Rapa Nui	Text used in the PISA 2018 test	EN/FR
	Cow's Milk	Text used in the PISA 2018 test	EN/FR
Practice text: encyclopedic	Wikipedia – The Moon	Excerpts from the Wikipedia entry about the moon	EN
Practice text: Literary	The North Wind and the Sun	One of Aesop's fables from ancient Greece	EL

The MultiplEYE comprehension questions

- Each text is followed by **6 single-choice comprehension questions with 4 answer options**
- They are developed in English

→ **Need to be translated into your language**

Development and evaluation of comprehension questions

- Designed according to Gehrer et al. (2013) and Berzak et al. (2020)
- Three different **questions types** targeting different linguistic levels (local, bridging, global)
- Each question has different **distractor types**
- Annotation of **target** and **distractor spans** in the stimulus texts

  → allows for analysing the link between eye-movements and text comprehension
- Several iterations of **manual human review**
- Evaluation of the comprehension questions via **online** and **in-lab experiments**

Your involvement

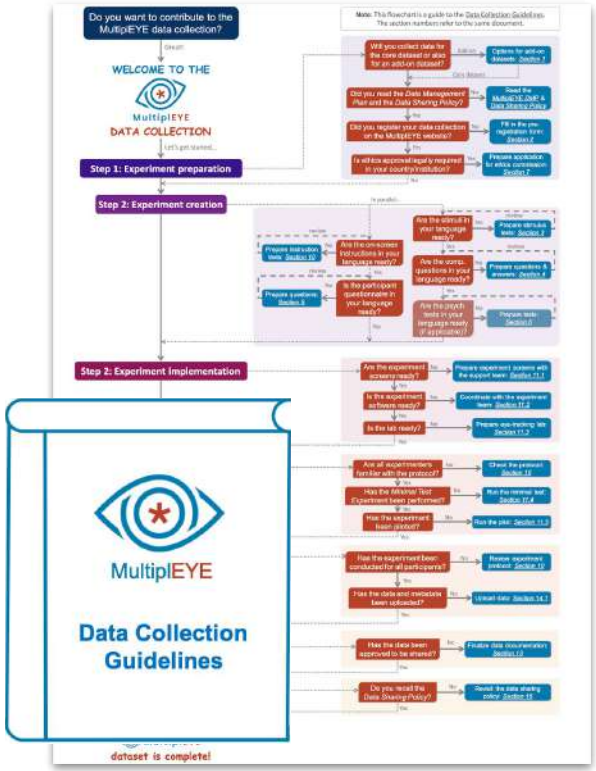
- ❑ Join the data collection → pre-register your dataset
- ❑ Prepare experiment in your language
- ❑ Prepare the local team & lab
- ❑ Check out the Data Collection Guidelines
- ❑ Join the MultiplEYE COST Action (see www.multipleye.eu)
- ❑ Apply for external funding for lab assistants & participant reimbursement
- ❑ Get in touch!

Benefits

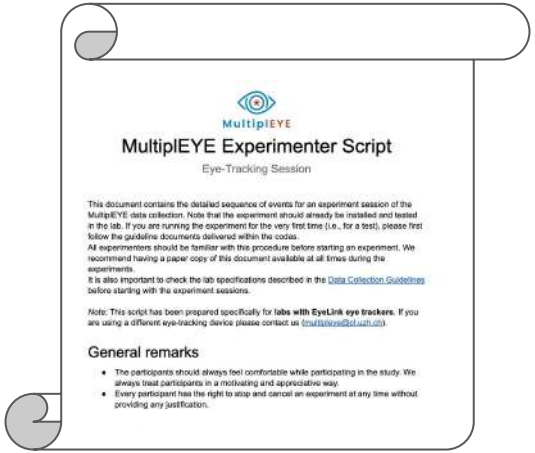
- ★ Become part of our large international multilingual data collection
- ★ Become part of the MultiplEYE network
- ★ Authorship in dataset publication
- ★ Ownership of your data (see Data Sharing Policy)
- ★ Visibility of your data collection and your language

For more information, see Data Collection Guidelines

Our support
Guidelines & Tutorials



Dominant eye test



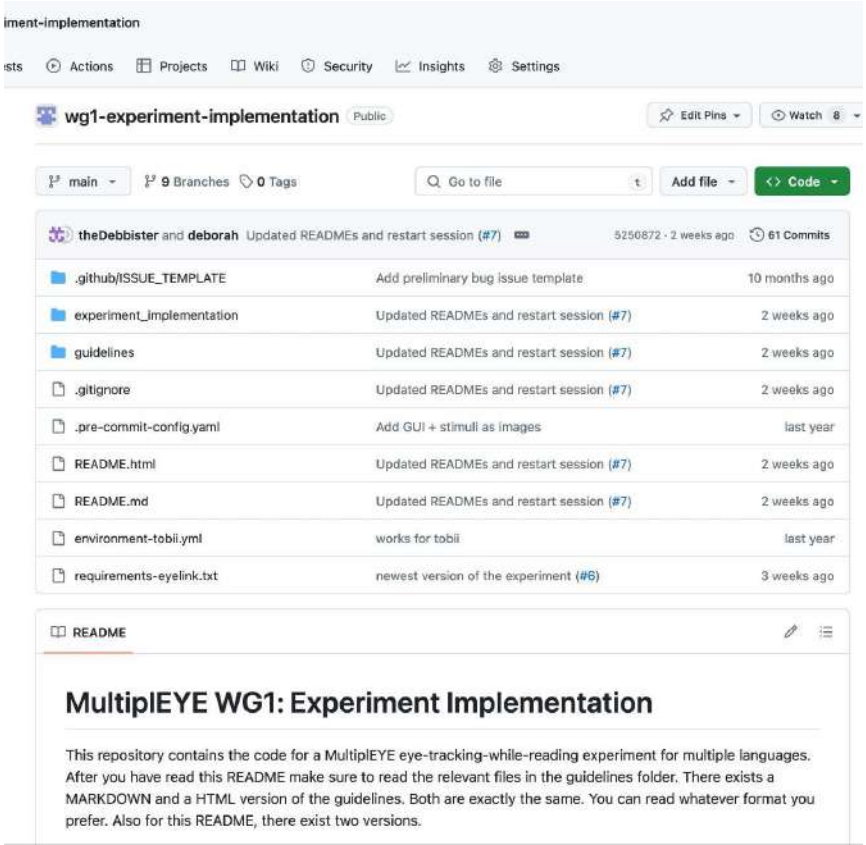
Camera set-up & calibration

How to run the psychometric tests

Our support

Eye-tracking experiment implementation

- Open-source implementation in Python
- Ensures identical procedure and stimulus layout across labs and languages

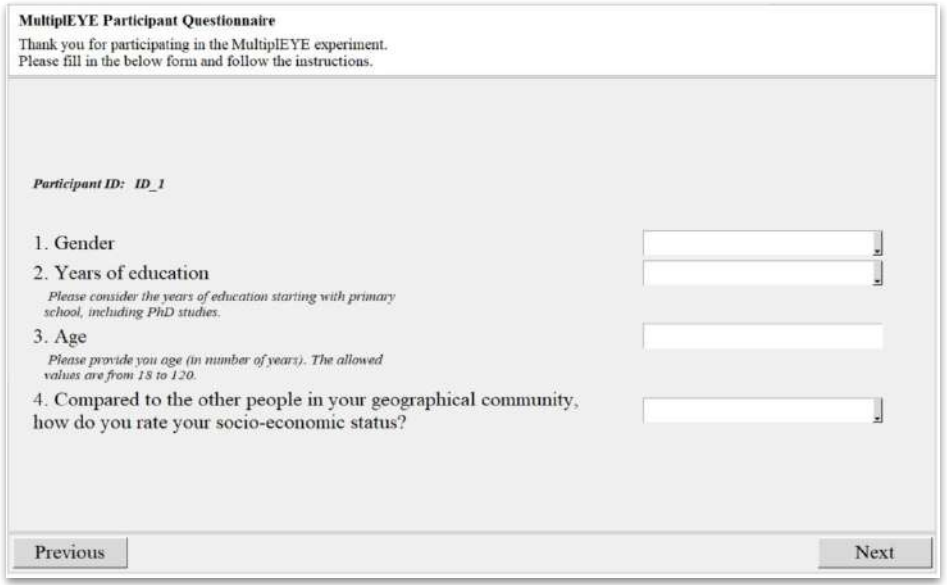


Our support

Participant questionnaire

- Demographic and linguistic background
- Computer-based implementation to ensure compatibility of data across labs

→ **Needs to be translated into your language**



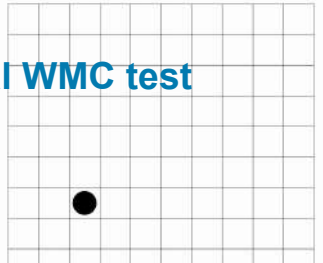
Our support

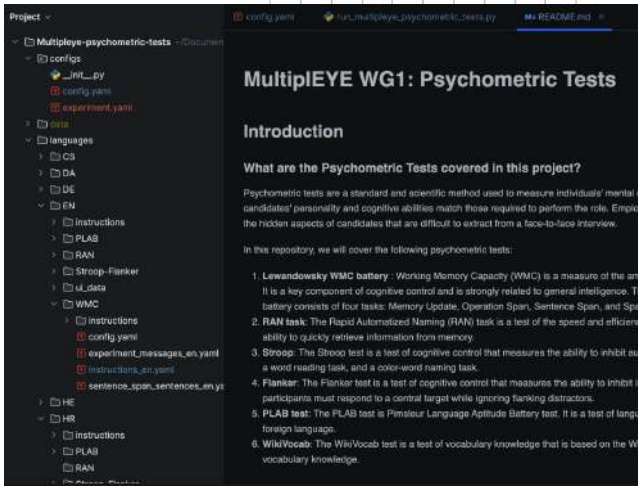
Implementation of psychometric tests

- Open source implementation of **computer-based versions** of psychometric tests
- Most tests are not language-specific

→ Tests (instructions) need to be translated to your language.

Spatial WMC test





Tutorial video for experimenters

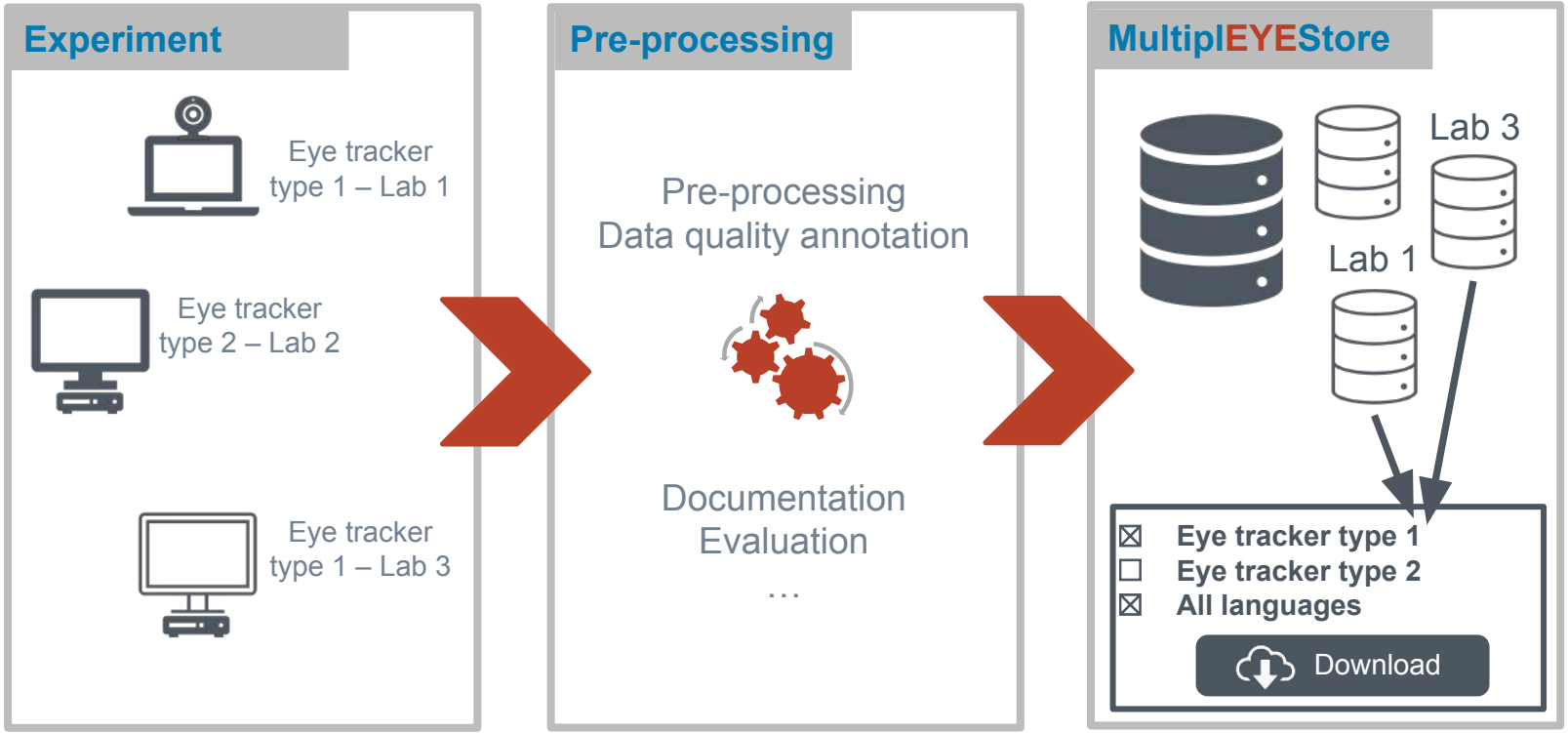
Our support

Participant instructions

- Identical participant instructions across labs ensure comparability of the data

  → **Need to be translated into your language**

Data management
Data diversity while ensuring consistency



Our support

Data management & documentation

Quality Reporting

We provide a standardized documentation pipeline (see Data management plan for details).

- Forms to be filled in by the contributing labs:
  - Pre-Registration Form
  - Meta-Data Form
  - Stimulus Documentation Deviation Form
  - Experimenter Session Documentation Sheet
- Data quality documentation is integrated into our pre-processing pipeline.

MultiplEYE
Meta-Data Form

1. ENTER THE TITLE FOR YOUR DATA COLLECTION

To ensure consistency within the COST Action Project "MultiplEYE", it is essential that the name of your data collection/dataset follows the MultiplEYE naming convention and is consistently applied throughout. The name is composed of the terms "MultiplEYE", the tested language (ISO-639-1, 2-letter language code), the name of your country (ISO- 3166, 2-letter country code), the name of your city, your identifier, and the year when your data collection will end. The name should have already been generated at pre-registration. ***The entire name MUST be identical with the one that has been pre-registered.***

Title: * your pre-registered title

example: MultiplEYE_DE_DE_Berlin_1_2024

2. ENTER THE PERSON(S) RESPONSIBLE FOR THE RESEARCH DATA / FOR CREATING THE DATASET AT THE COLLECTION SITE

Provide the name of any person(s) ***as lead creator(s)*** who contributed to your dataset and how they contributed. A lead contributor to a data collection is anyone who is responsible for the data collection at a lab. This includes those who take on central roles in planning, organizing, and executing the data collection process, ensuring the quality and accuracy of the data collected, and providing documentation. Please name all lead contributors specifically from your collection site / lab / institution.

Note: Add as many names as you need. Use a semicolon to separate the persons. Use full names (titles are omitted) and state the type(s) of contribution.

Lead Creator(s):

* 1. first name contributor, family name contributor, type of contribution; 2. first name contributor, family name of

Our support

Assistance with lab set-up

- Pre-registration of your data collection via the online Pre-Registration Form

  → We **generate the stimulus files** for your language and lab set-up
- We provide **technical support** with the installation of the experiment

  → **New members in the MultiplEYE support team are welcome!**
- We provide a Checklist for Experiment Preparation
- COST provides funding for Short-Term Scientific Missions that can be used to get on-site help with the set-up of the experiment

Our support

MultiplEYEStore

- All data will be made accessible via the **MultiplEYEStore:** a dedicated data repository hosted by *PsychArchives* (**Leibniz Institute for Psychology ZPID** in Trier)
- FAIR principles, GDPR compliant
- Extensive data documentation & meta-data
- Data quality reports
- User-friendly GUI for filtering the data

Funded by:

swissuniversities

Swiss National Science Foundation

Federal Ministry of Education and Research

… more funding needed

Our support

MultiplEYEStore

Open Access

Search Database

Language

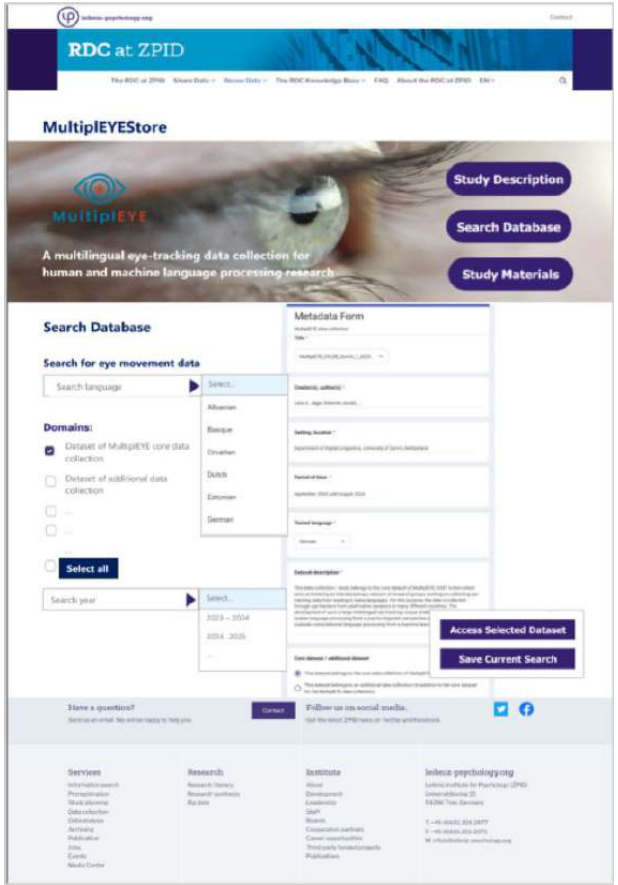
- ❑ Albanian
- ❑ Basque
- ❑ …

Albanian Datasets

- ❑ Lab 1
- ❑ Lab 2
- ❑ …

Quality

- ❑ Avg. validation score above x
- ❑ Trials completed > 6
- ❑ …



Our support

Preprocessing & Data Quality Reporting

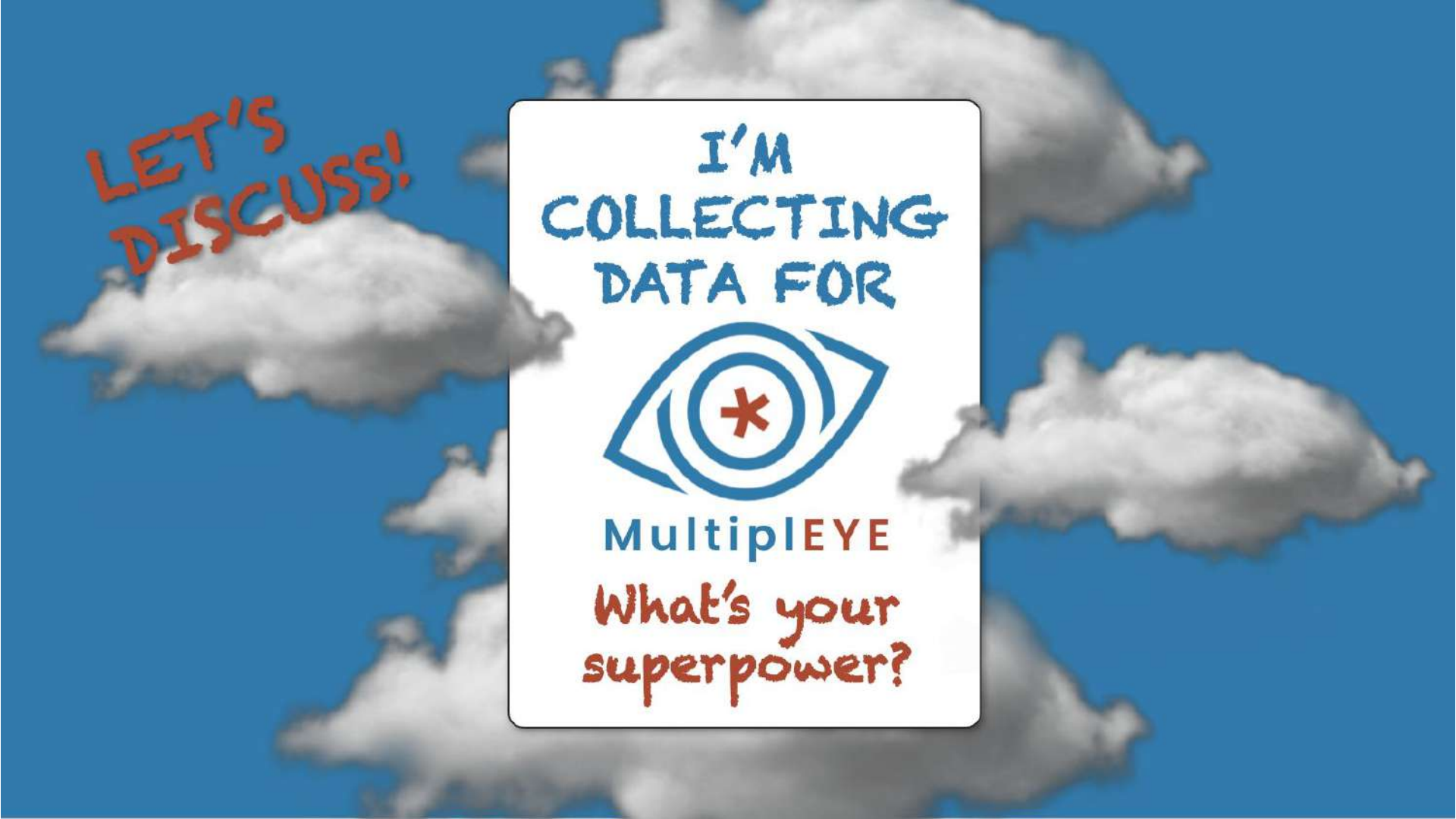
- Development of a standardized pre-processing pipeline:
  - Open source, transparent, reproducible
  - Make data from different languages and different eyetrackers compatible
  - Provide data in different formats (raw samples, scanpaths, reading measures)
  - Generation of data quality reports to be included in the meta-data
- Development of open-source preprocessing software pymovements (see Krakwczyk et al., 2023)
- Development and implementation of data quality metrics and reporting standards (see Jakobi et al., 2024).

→ Developers are welcome to **join the *pymovements* team!**

Current overview of languages

Albanian, Arabic, Basque, Catalan, Croatian, Chinese, Czech, Danish, Dutch, English, Estonian, French, German, Greek, Italian, Latvian, Lithuanian, Polish, Portuguese, Romanian, Russian, Spanish, Swedish, Turkish, Ukrainian, ...

And your language?


LET'S DISCUSS!
I'M COLLECTING DATA FOR
MultiplEYE
What's your superpower?

Communication & Contact

Communication via Slack:

→ You will receive an invitation to join the **MultiplEYE Slack workspace** once upon acceptance of your application to the Action.

Action Chairs:

Nora Hollenstein nora.hollenstein@ktu.lt
Lena Jäger jaeger@cl.uzh.ch

WG 1 leader:

Ramunė Kasperė ramune.kaspere@ktu.lt

Thank you!

More information:

https://multipleye.eu/

Apply here

Data Collection Guidelines & more:
https://multipleye.eu/forms-and-documentation-hub/

Get started:
https://multipleye.eu/contribute/