A Framework and Practical Guidelines for Sharing Open Benchmark Datasets in Cartographic User Research Utilizing Neuroscientific Methods

Merve Keskin*, Tong Qin**, Bing Liu**

* The Department of Geoinformatics, ZGIS, The PLUS, Austria
** Geography Department, Ghent University, Belgium
*** BYD Co. Ltd, China

Using Eye Tracking and AI to Personalize Map Reading: The importance of open benchmark datasets

Merve Keskin

International Cartographic Association

Keywords: eye tracking, geoAI, benchmark datasets, open eye tracking datasets
EYE TRACKING

a non-invasive method of measuring the gaze location, where one is looking

$x, y, t \rightarrow$ fixations, saccades, blinks & pupil metrics

good for visual attention, perception, learning, HCI interaction
EEG

*electroencephalogram*

a non-invasive method of recording electrical activity of the brain

measures the voltage fluctuations occurring between different regions of the scalp

a valuable tool to extract cognitive load

Keskin, 2020
fMRI

functional magnetic resonance imaging

measures brain activity by detecting changes in blood flow.

measures magnetic differences between oxygen-rich hemoglobin and deoxygenated hemoglobin: active brain region = more blood flow

a valuable tool to extract sensory response

Lobben et al., 2005
Relevant case study examples

Perceptions of space and time of public transport travel associated with human brain activities: A case study of bus travel in Beijing

Tong Qin, a,b *, Weihua Dong, a *, Haosheng Huang, a

travel recall

Using fMRI to Explore the Influence of Road Network Patterns on Geospatial Cognition

Bing Liu, a,b *, Weihua Dong, a *, Lin Zhu, a, Huiping Liu, a, Liqiu Meng

orientation shortest-route-selection

Visual attention and neuro-cognitive processes in map use

Tong Qin a,b, *, Haosheng Huang a,b

b CaroGIS, Ghent University, Ghent, Belgium, tong.qin@ugent.be, haosheng.huang@ugent.be
global search distance comparison route following route planning

Differences in the Gaze Behaviours of Pedestrians Navigating between Regular and Irregular Road Patterns

by Bing Liu 1, 2, *, Weihua Dong 1, *, Zhicheng Zhan 1, Shengkai Wang 1 and Liqiu Meng 2

road pattern orientation shortest route selection

Exploring the Cognitive Load of Expert and Novice Map Users Using EEG and Eye Tracking

by Merve Keskin 1, 2, *, Kristien Ooms 1, Ahmet Ozgur Dogru 2 and Philippe De Maeyer 1

memorability recognition landmarks
With the increasing amount of neuroscientific/physiological data collected, there is a growing need for further analytical and methodological research in map reading.

Future efforts should focus on developing models to automatize the analyses of existing and large volumes of user data, to predict user behavior and personalize maps (in terms of design, content, context) anticipating users’ intention providing the most suitable map design adapting to the context of use

To develop predictive models, we need data generated during the observation of maps.
A comprehensive structure for benchmark datasets & guidelines for open-data sharing in neuroscientific/physiological user research* for geospatial data and maps

*focusing on eye tracking, EEG, and fMRI
Geospatial data and maps present unique complexities that differentiate them from traditional stimuli subjected to experimental psychology. Therefore, it becomes imperative to adapt existing practices to suit the specific requirements of geospatial tasks.

Karim et al., 2017

Why?

He et al., 2023, GeoEye dataset
Traditional saliency models are designed to predict visual attention during the observation of natural images, free viewing and cannot perform as well when it comes to maps.

Research Gap 1.1: There are not many open (eye-tracking) datasets available for map-reading tasks.

### MIT/Tuebingen Saliency Benchmark

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Citation</th>
<th>Images</th>
<th>Observers</th>
<th>Tasks</th>
<th>Durations</th>
<th>Extra Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIT300</td>
<td>Tilke Iudd, Fredo Durand, Antonio Torralba. A Benchmark of Computational Models of Saliency to Predict Human Fixations [MIT tech report 2012]</td>
<td>300 natural indoor and outdoor scenes size: max dim: 1024px, other dim: 457-1024px 1 dva* ~ 35px</td>
<td>39 ages: 18-50</td>
<td>free viewing</td>
<td>3 sec</td>
<td>This was the first data set with held-out human eye movements, and is used as benchmark test set in the MIT/Tübingen Saliency Benchmark. eyetracker: ETL 400 ISCAN (240Hz) Download 300 test images.</td>
</tr>
<tr>
<td>CAT2000</td>
<td>Ali Borji, Laurent Itti. CAT2000: A Large Scale Fixation Dataset for Boosting Saliency Research [CVPR 2015 workshop on “Future of Datasets”]</td>
<td>4000 images from 20 different categories size: 1920x1080px 1 dva* ~ 38px</td>
<td>24 per image (120 in total) ages: 18-27</td>
<td>free viewing</td>
<td>5 sec</td>
<td>This dataset contains two sets of images: train and test. Train images (100 from each category) and fixations of 18 observers are shared but 6 observers are held-out. Test images are available but fixations of all 24 observers are held-out. eyetracker: EyeLink1000 (1000Hz) Download 2000 test images. Download 2000 train images (with fixations of 18 observers).</td>
</tr>
</tbody>
</table>
Research Gap 1.2: There are not many open (physiological) datasets available for map-reading tasks.

<table>
<thead>
<tr>
<th><strong>EEG datasets (github):</strong></th>
<th><a href="https://github.com/meagmohit/EEG-Datasets">https://github.com/meagmohit/EEG-Datasets</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>visual perception, memory, and motor control (Agarwal, 2023)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Donders Data Repository:</strong></th>
<th><a href="https://github.com/Donders-Institute/meg-hackathon">https://github.com/Donders-Institute/meg-hackathon</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Brain imaging (fMRI, MEG, EEG) datasets:</td>
<td></td>
</tr>
<tr>
<td>language, attention, and emotion</td>
<td></td>
</tr>
<tr>
<td>FAIR (Findable, Accessible, Interoperable, and Reusable) principles</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>The SJTU Emotion EEG Dataset (SEED):</strong></th>
<th><a href="https://bcmi.sjtu.edu.cn/home/seed/">https://bcmi.sjtu.edu.cn/home/seed/</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>BCMI laboratory datasets for emotion recognition</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Radboud Coregistration Corpus of Narrative Sentences (RaCCooNS):</strong></th>
<th><a href="https://repository.ubn.ru.nl/handle/2066/296635">https://repository.ubn.ru.nl/handle/2066/296635</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>eye-tracking-with-EEG data during reading:</td>
<td></td>
</tr>
<tr>
<td>human sentence comprehension and for evaluating the cognitive validity of computational language models</td>
<td></td>
</tr>
</tbody>
</table>
Research Gap 2: Benchmarks should be aimed at **specific scientific issues** rather than just data standards.

We know that eye movements and brain activities show different patterns for different tasks.

Yarbus, 1967

Not free viewing, task-specific

Lobben et al., 2019
Research Gap 3: There exists no standards for reporting multi-modal physiological user data.

The multi-modal information fusion and analyses methods are well-developed.

Reporting results and sharing datasets and methods?
Existing standardization work for reporting eye tracking

<table>
<thead>
<tr>
<th>Category</th>
<th>Item no.</th>
<th>Checklist item</th>
<th>Reported on page no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>Manufacturer and model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A2</td>
<td>Software and firmware versions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A3</td>
<td>Eye tracking technology</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A4</td>
<td>Sampling frequency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A5</td>
<td>Head movement restrictions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A6</td>
<td>Eye(s) recorded</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A7</td>
<td>Parameters recorded</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A8</td>
<td>Environment lighting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A9</td>
<td>Calibration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A10</td>
<td>Measurement uncertainty</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A11</td>
<td>Data processing steps</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A12</td>
<td>Data loss</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B1</td>
<td>Signal latencies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>Participant to display monitor distance</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Additional item to be reported by studies of eye movement dynamics

Additional item to be reported by studies reporting screen-based gaze coordinates

---

Table 1 Checklist of information to include when reporting an eye tracking study

<table>
<thead>
<tr>
<th>Items to be reported by all eye tracking studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1 Manufacturer and model</td>
</tr>
<tr>
<td>A2 Software and firmware versions</td>
</tr>
<tr>
<td>A3 Eye tracking technology</td>
</tr>
<tr>
<td>A4 Sampling frequency</td>
</tr>
<tr>
<td>A5 Head movement restrictions</td>
</tr>
<tr>
<td>A6 Eye(s) recorded</td>
</tr>
<tr>
<td>A7 Parameters recorded</td>
</tr>
<tr>
<td>A8 Environment lighting</td>
</tr>
<tr>
<td>A9 Calibration</td>
</tr>
<tr>
<td>A10 Measurement uncertainty</td>
</tr>
<tr>
<td>A11 Data processing steps</td>
</tr>
<tr>
<td>A12 Data loss</td>
</tr>
<tr>
<td>B1 Signal latencies</td>
</tr>
<tr>
<td>C1 Participant to display monitor distance</td>
</tr>
</tbody>
</table>

---

Minimal reporting guideline for research involving eye tracking (2023 edition)

- Matt J. Dunn
- Robert G. Alexander
- Obyekachukwu M. Amiebenomo
- Gemma Arblaster
- Denize Atan
- Jonathan T. Erichsen
- Ulrich Ettinger
- Mario E. Giardini
- Jain D. Gilchrist
- Ruth Hamilton
- Roy S. Hessels
- Scott Hodgins
- Ignace T. C. Hooge
- Brooke S. Jackson
- Helena Lee
- Stephen L. Macknik
- Susana Martinez-Conde
- Lee McIlveen
- Lisa M. Muratori
- Diedrich C. Niehorster
- Marcus Nyström
- Jorge Otero-Millan
- Michael M. Schlüssel
- Jay E. Self
- Tarkeshwar Singh
- Nikolaos Smyrnitis
- Andreas Sprenger

---

Table 1 Checklist of information to include when reporting an eye tracking study

<table>
<thead>
<tr>
<th>Category</th>
<th>Item no.</th>
<th>Checklist item</th>
<th>Reported on page no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>Manufacturer and model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A2</td>
<td>Software and firmware versions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A3</td>
<td>Eye tracking technology</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A4</td>
<td>Sampling frequency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A5</td>
<td>Head movement restrictions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A6</td>
<td>Eye(s) recorded</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A7</td>
<td>Parameters recorded</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A8</td>
<td>Environment lighting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A9</td>
<td>Calibration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A10</td>
<td>Measurement uncertainty</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A11</td>
<td>Data processing steps</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A12</td>
<td>Data loss</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B1</td>
<td>Signal latencies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>Participant to display monitor distance</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---
A framework to facilitate advanced mixed methods studies for investigating interventions in road space for cycling (Werner et al., 2022)

The Knowledge Graph available on Zotero: GraphML editor to adapt and extend the graph for specific use cases
WHAT WE HAVE DONE

guidelines for open-data sharing

- CONTROLLED CONDITIONS
- WELL-DEFINED TASKS & RESEARCH QUESTIONS
- WELL-DEFINED DATA
- WELL-DEFINED METRICS
- WELL-DEFINED PARTICIPANT CHARACTERISTICS
- WELL-DEFINED STIMULUS PROPERTIES
- ETHICS
<table>
<thead>
<tr>
<th>CONTROLLED CONDITIONS</th>
<th>WELL-DEFINED TASKS &amp; RESEARCH QUESTIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medium/display</td>
<td>Purpose with keywords defining the study</td>
</tr>
<tr>
<td>Performance for data collection and system specification</td>
<td>Full procedural details (e.g., standard flowchart)</td>
</tr>
<tr>
<td>Input modality</td>
<td>Free viewing vs. task-specific</td>
</tr>
<tr>
<td>Recording devices</td>
<td>In labs or in real-world environments</td>
</tr>
<tr>
<td>Extraneous variables</td>
<td>Visuospatial or perceptual tasks</td>
</tr>
<tr>
<td></td>
<td>Trial tasks, orientation and instructions</td>
</tr>
<tr>
<td></td>
<td>Task design (e.g., randomized block design, event-related design)</td>
</tr>
<tr>
<td></td>
<td>Task duration and total recording length</td>
</tr>
<tr>
<td>WELL-DEFINED DATA</td>
<td>WELL-DEFINED METRICS</td>
</tr>
<tr>
<td>----------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Artifact-free or raw data</td>
<td>Behavioral metrics: response time, response accuracy</td>
</tr>
<tr>
<td>Data quality</td>
<td>Eye tracking: fixation- or saccade-related, AOI- (area of interest) specific metrics, scanpaths, heatmaps</td>
</tr>
<tr>
<td>Sufficiently large data samples to ensure the generalizability of the results</td>
<td></td>
</tr>
<tr>
<td>Data format and compatibility</td>
<td>EEG: time-domain: Event-Related Potentials (ERP)</td>
</tr>
<tr>
<td>Detailed documentation data collection, pre-processing and analysis protocols, and open codes for such analysis</td>
<td>frequency-domain: Power Spectral Density (PSD)</td>
</tr>
<tr>
<td>Relevant scientific research and/or other relevant references</td>
<td>time frequency-domain: Event-Related Synchronization &amp; Desynchronization (ERS/ERD)</td>
</tr>
<tr>
<td>Data specific descriptions:</td>
<td>fMRI: Blood Oxygen Level Dependent (BOLD) Signal Functional Connectivity</td>
</tr>
</tbody>
</table>

Data specific descriptions:

- Eye tracking: dominant eye, resolution, fixation recognition algorithm/parameters
- EEG: resolution, the number, type, and spatial distribution of electrodes
- fMRI: the number of channel head coils, repetition and echo time (for functional/structural images), layer scan
### WELL-DEFINED PARTICIPANT CHARACTERISTICS

- Sample size: often large for EEG and fMRI
- Individual participant characteristics (age, gender, education)
- Additional tests to classify participants based on spatial abilities
- Color blindness, users with other disabilities
- Self-reports, pre- or post-test questionnaires, and structured verbal interviews

### WELL-DEFINED STIMULUS PROPERTIES

- Screen map, animation, web-service
- 2D, 3D or XR
- Static, dynamic, interactive
- Size, position, and format of the media
- Visual or task-related manipulation
- Experimental stimuli preparation details

### ETHICS

- Asking local ethics committees for permission if needed
- Adhering to ethical standards, i.e., consent from participants
- Anonymization of participants’ data
**CartoGAZE**

an open eye tracking dataset from a map memorability experiment

40+ downloads since Jan 2023

---

**Visual Attention and Recognition Differences Based on Expertise in a Map Reading and Memorability Study**

by Merve Keskin 1, Vassilios Krassanakis 2 and Arzu Çöltekin 3

1 Finnish Geospatial Research Institute, National Land Survey of Finland (FGI/NLS), Vuorinlehterntie 5, 02160 Espoo, Finland
2 Department of Surveying and Geoinformatics Engineering, Egaleo Park Campus, University of West Attica, Ag. Smyrhdonos Str., 12243 Egaleo, Greece
3 Institute of Interactive Technologies, School of Engineering, University of Applied Sciences and Arts Northwestern Switzerland, Bahnhofstrasse 6, 5210 Windisch, Switzerland

* Author to whom correspondence should be addressed

*SPRS Int. J. Geo-Inf. 2023, 12(1), 21; https://doi.org/10.3390/ijgi12010021
Received: 30 October 2022 / Revised: 26 December 2022 / Accepted: 6 January 2023 /
Published: 12 January 2023

---

Automated AOI-based fixation analysis of large eye tracking data
**Tasks**

- \( t_1 \) = Hard (Block 1 & 2) 74 trials
- \( t_2 \) = Moderate (Block 3 & 4) 74 trials
- \( t_3 \) = Easy (Block 5, 6 & 7) 111 trials

**Map features**

- roads
- road junctions
- green areas
- hydro areas
- hydro lines

**Participants**

- Novices (N=21) (age and gender balanced)
- Experts (N=17)

**Block Information**

- **Block 1**: road & hydrographic features, green areas
- **Block 2**: road & hydrographic features
- **Block 3**: road & green areas
- **Block 4**: green areas & hydrographic features
- **Block 5**: green areas
- **Block 6**: hydrographic features
- **Block 7**: road features

**Presentation Timeline**

- Stimulus onset
- Presentation of fixation cross 2 s
- Presentation of the map stimulus 7 s
- Presentation of graphical response screen
- Presentation of response screen with letters
  - Pressing the space bar
  - Selecting the correct answer
The transregional Collaborative Research Center 161 (SFB-TRR 161) is an interdisciplinary research center of the University of Stuttgart, University of Konstanz, Ulm University, and the LMU Munich.

the Dimensionality Reduction Special Interest Group (DR SIG) Hackathon (in 2024):

DR for Eye-tracking (DR4ET)

User Performance and Reading Strategies for Map Users: An Evaluation of Eye Tracking Study

Merve Keskin and Kun-Ting Chen

1Department of Geoinformatics - Z-GIS, University of Salzburg (PLUS), Salzburg, Austria.
2Centre for Research on Engineering Software Technologies (CREST), University of Adelaide

From Hackathon, we expect

- automatic or semi-automatic user grouping based on AOI-gaze features
- guiding statistics framework for analyzing gaze data and visualizing outcomes (e.g., statistical grayscale heatmaps)
- automatic classification of participants and recommending suitable visualization of map designs.

Relation to Dimension Reduction: find the most critical spatiotemporal eye tracking data characteristics and visual variables of the map features, relating to task difficulty, expertise group, and spatial memory strategies of human operators.
What can our geo-community do?

Creating a platform disseminating open datasets and repositories

– similar to MIT/Tuebingen Saliency Benchmark but with a “geo” focus and compatible with FAIR data sharing standards

Encouraging ourselves and everyone else to share experimental data within and outside our communities (datavis, infovis, UX/UI, ET, AI, etc.)
– benefits: interdisciplinary collaborations, esp. with researchers with similar problems
WEBCAM-BASED EYE TRACKING: current capabilities and future opportunities

reviving Remote Eye Tracking Benchmarking exercise – initiated by Amy Griffin, 2021

Eye tracking experts from Geoscience, Computer Science, Psychology, Optometry
Datasets capable of addressing specific spatial problems hold the potential to become benchmarks in developing complexity algorithms & predictive models that consider the visual attention and map use capabilities of users.

Benchmarks should be aimed at specific scientific issues rather than just data standards.

Researchers across domains bear the responsibility of actively seeking concrete methods to encourage the open sharing of experimental data, complemented by high-quality metadata.

By fostering the creation of benchmark datasets and promoting open-data sharing, collaboration is enhanced, geospatial research advances, and the scientific community is empowered to effectively address cartographic challenges.
«IT'S OK TO BE A CARTOGRAPHER»

Georg Gärtner, resident of International Cartographic Association (ICA)

drmervekeskin.com