Funded by the
European Union

Project Acronym: **GAIN**

Project Full Title: **Georgian Artificial Intelligence Networking and Twinning Initiative**

Grant Agreement Number: 101078950

HORIZON EUROPE - WIDERA-2021-ACCESS-03 (Twinning)

WEBSITE: www.gain-twinning.eu

Project Start Date: 1st October 2022

Funded by the
European Union

Deliverable D.6.4. Data Management Plan, Version 2

Dissemination level:	Sensitive
Due Date:	March 31, 2024
Deliverable leader:	MICM
Author:	George Giorgobiani
Contributors:	All
Reviewers:	J. Alexandersson, P. Müller, B. E. Wirth, F. Bremond
Type:	Report
WP / Task responsible:	MICM
Key words:	FAIR, GDPR, ELSI, data, management, plan

Contents

1. Executive Summary

This document describes the **Data Management Plan, Version 2 (D.6.4.)** of the project GAIN.

Document History			
Version	**Date**	**Contributors**	**Description**
1	27/01/2024	J. Alexandersson, P. Müller, B. E. Wirth – DFKI. F. Bremond– INRIA. G. Giorgobiani – MICM.	Data Management Plan, Version 2
2	18/03/2024	J. Alexandersson, P. Müller, B. E. Wirth – DFKI. F. Bremond– INRIA. G. Giorgobiani – MICM.	Data Management Plan, Version 2
Final	28/03/2024	J. Alexandersson, P. Müller, B. E. Wirth – DFKI. F. Bremond– INRIA. G. Giorgobiani – MICM.	Data Management Plan, Version 2

2. Introduction

The first version of the Data Management Plan (DMP) of the project was given in D.6.3, Data Management Plan, Version 1 (WP6, Coordination and Management). The DMP is based on the requirements of the General Data Protection Regulation, on the 4 principles of FAIR Data (Findability, Accessibility, Interoperability and Reusability) and Open Science principles.

This document, DMP version 2, is the update of the DMP version 1, at the middle of the project. It includes information about the scientific results and datasets available at this point of time.

For sharing the results "Project's Online Management System", the project website and the Zenodo Platform[1] are used.

The GAIN capacity building programme includes the creation of a multimodal corpora (speech, video, biosensors) of social patient-clinician interactions. These corpora are subject to the aspects of ethical, legal and social implications (ELSI), GDPR requirements compliance and to the Consortium Agreement of GAIN as well as the Ethics Protocol (Deliverables 1.1, 1.2 and 1.3) created within the project. This kind of data is sensitive and thus restricted.

According to the DoA of the project, the reasons for updates might be significant changes in the project, such as changes in the project consortium. During the reporting period there were no such changes in the project.

[1] https://zenodo.org/me/uploads?q=&l=list&p=1&s=10&sort=newest

3. Research results

According to the FAIR principles:

- Findability is ensured by sharing data through the Project Online Management Platform. Access to the platform is provided to all participants of the project. Non-restricted and non-sensitive data may be shared to external stakeholders. Public data is shared on the project website, social media (Facebook and LinkedIn), Zenodo platform and via the dissemination activities.
- Accessibility is ensured by making the public (anonymized) data of the project available to all parties that request access.
- Interoperability is ensured by storing project data in an appropriate format that makes data accessible to all professionals who are interested in exploiting.
- Reusability will be ensured by maintaining the project data in a reusable way at the project website for as long as possible after the end of the project.

3.1. Publications and conference talks.

Scientific results of the project are handled in accordance with the Open Science and FAIR principles. During the reporting period, the project has produced:

- 5 research papers, published in peer reviewed, open access journals, 1 paper is published in conference materials, 2 will appear in 2024, 1 is submitted and 3 papers prepared. Electronic copies of each published paper (see Appendix, Publications [1 – 6]), at the time of publication were available at the websites of the journals and repositories and immediate open access was provided. All publications were free of charge.
- 17 conference talks have been delivered at international scientific conferences. Georgian researchers have provided 5 presentations at the First Summer School (See Appendix).

Publications, conference talks and presentations always contain references to the European Union and the GAIN project and relevant acknowledgements.

3.2. Datasets

During the reporting period, the Gain researchers worked on generating the following datasets:

1. **EEG dataset**: Two types of experiments were conducted at MICM, which involve the recording of Event Related Potentials (ERP) with a commercial Unicorn Hybrid Black EEG device purchased by the GAIN project. The device has 8 active electrodes and a recording frequency of 250 Hz, covering 10-20 system positions FZ, C3, CZ, C4, PZ, PO7, OZ and PO8. Both experiments recorded auditory evoked potentials (AEP), which measure the electrical activity of the brain responding to sound stimuli, and both involve target/oddball tasks, although in the second experiment subjects were required to make a certain response depending on the type of stimulus. As planned, the dataset and its analyses will be ready for publication after 20 recordings (each including 2 sessions

per participant). Up to this day, 5 records have been conducted and the experiment is still in progress. The research material and the dataset will be uploaded to the Zenodo platform as well as the website of the project and will be open access.

2. **Georgian language corpus**: it is planned to use the GTU library's electronic documents to create Georgian language corpus. The work is in progress. The dataset will be uploaded to the Zenodo platform and the GTU and the project's websites and will be open access.
3. **Records of doctor-patient interviews**: in the reporting period, 5 interviews with 3 patients have been recorded at the Tbilisi Mental Health Centre. This data, containing medical and personal information, is subject to the Ethics Requirements and, consequently, is restricted. Data is encrypted and stored at the GAIN-MICM server. The work is in progress.

3.3. Open source software

3.3.1. Open source software framework EEGain.

Within the GAIN project, a novel open source software framework EEGain is designed to improve the comparability and generalizability of EEG-based emotion recognition approaches. The framework will be presented at the first workshop on Generalizability of Emotion Recognition from EEG Signals (EmoRec EEG) at ACII'24. It enables the participants to run 4 models – Tsception, EEGNet, DeepConvNet, ShallowConvNet, on 6 datasets – AMIGOS, DEAP, MAHNOB-HCI, SEED-IV, SEED, DREAMER.
Workshop link: https://emotionlab.github.io/EmoRec-EEG/
Framework link: https://github.com/EmotionLab/EEGain

3.3.2. Open source software VideoMAE.

For the problem of emotion recognition from video, scale is the primary ingredient in attaining generalizable video representations. While scaling model capacity and data size for video masked autoencoders, VideoMAE has been explored for large generic datasets of short videos, adapting these models for longer video sequences remains a challenge which we aimed to address. Open source software VideoMAE was used as a feature extractor model, and with an added classifier layer on top, it was fine-tuned on specific emotion recognition datasets to enhance its performance in identifying and categorizing emotions from video data.
The code is available at https://github.com/EmotionLab/EmotionVMAE

4. Data Protection

4.1. Protecting the research Data

Two types of protection to keep the sensitive data safe during sending, receiving and storing are applied.

Sending/Receiving:

- Reliable virtual private network (MICM-VPN) is created at MICM.
- End-to-end encryption is used to keep communication safe for sending/receiving speech, video, photo and other restricted data.

Keeping data safe:

- Dedicated user accounts and read/write -access to the files/folders are only created for authorized user.
- A well set up Access Logging (monitored routinely) with strong passwords is launched.
- The facility hosting the GAIN-MICM server is secured with restricted access.

4.2. Protection of Personal data

Personal data protection follows the Directive 95/46/EC of the GDPR. The personal and business data collected from the partners, deliverables, meeting minutes, technical and other documents, collected in the GAIN project, are securely stored.

5. Appendix.

Publications, conference talks and presentations contain references to the European Union and the GAIN project and relevant acknowledgements[2].

Publications

1. **Philipp Müller, Michal Balazia** et. Al. MultiMediate'23: Engagement Estimation and Bodily Behaviour Recognition in Social Interactions. 10.1145/3581783.3613851, https://arxiv.org/pdf/2308.08256.pdf
2. **T. Saghinadze**. Constructing Convolutional Neural Networks with 90 Degree Rotational Equivariance and Invariance. Georgian Electronic Scientific Journal: Computer Science and Telecommunications 2023|No.1(63), p. 39 – 43. ISSN 1512-1232. https://gesj.internet-academy.org.ge/en/list_artic_en.php?b_sec=comp
3. S. Chobanyan, L. Chobanyan, Z. Gorgadze and **G. Ghlonti**. An Algorithm for Finding a Near-Optimal Rearrangement in the Steinitz Functional. Bulletin of TICMI. Vol. 27, No. 1, 2023, 21–27. http://www.viam.science.tsu.ge/others/ticmi/blt/vol27_1/3.pdf
4. N. Abzianidze, N. Dogonadze, **G. Ghlonti**, Z. Kipshidze. About knowledge delivery strategies for intelligent tutoring systems in mathematics and computer science. Bulletin of TICMI. Vol. 27, No. 1, 2023, 29–37, ISSN 1512-0082 http://www.viam.science.tsu.ge/others/ticmi/blt/vol27_1/4.pdf
5. **V. Kvaratskhelia, G. Giorgobiani, M. Menteshashvili.** On one connection between the moments of random variables. Computer Science and Information Technologies CSIT 2023, September 25 - 30, 2023, Yerevan, Armenia. Book of abstracts https://csit.am/2023/proceedings/DMCA/DMCA_5.pdf
6. G. Baghaturia, **M. Menteshashvili**. Application of general integral of quasi-linear equation to solving of non-linear Cauchy problem. Bulletin of TICMI. Vol. 27, No. 2, 2023, 59–65. ISSN 1512-0082. https://www.emis.de/journals/TICMI/vol27_2/2%20Bagaturia_Menteshashvili_23_TICMI.pdf
7. G. Chelidze, S. Chobanyan, **G. Giorgobiani**, V. Tarieladze. Trigonometric series and the permutation sign convergence condition. *To appear in Analysis Mathematica, 2024.*
8. **Kachiashvili K.J., Kachiashvili J.K., Kalandadze R. M., Kvaratskhelia V.V**. Automatic diagnosis of lung diseases (pneumonia, cancer) with given reliabilities on the basis of an irradiation images of patients. *Submitted to the journal "Cancer Investigation".*
9. **T. Agrawal.** Robust and Efficient Multimodal Multi-dataset Multitask Learning. *Prepared to submit.*
10. **N. Kukhilava, T. Tsmindashvili, R. Kalandadze, L. Ferrari, V. Strizhkova**. Multimodal emotion recognition with physiological signals and video. *Prepared to submit.*

[2] The authors with bold letters are the GAIN consortium members.

11. **T. Tsmindashvili, N. Kukhilava, S. Katamadze, R. Kalandadze, L. M. Ferrari, P. Müller, B. E. Wirth** Evaluation in EEG Emotion Recognition: State-of-the-Art Review and Unified Framework. *Prepared to submit*.
12. **G. Giorgobiani, V. Kvaratskheliaa, and M. Menteshashvili**. Unconditional Convergence of Sub-Gaussian Random Series. *To appear in Pattern Recognition and Image Analysis, 2024.*

Conference Talks

The conference talks were supported by the GAIN project. Presentations contain the logos of the EU and the project and all relevant acknowledgements (see https://www.gain-twinning.eu/?page_id=300)

1. **G. Giorgobiani, V. Kvaratskhelia, T. Saghinadze.** Mathematics of Artificial Intelligence. 2nd Int. Conf.: Science, Education, Innovations and Chemical Technologies – From Idea to Implementation. Tbilisi, Georgia, 23 – 24 November 2023. https://conference23iice.ge/
2. **N. Kukhilava, T. Tsmindashvili, R. Kalandadze, L. Ferrari, V. Strizhkova**. VideoMAE for Emotion Recognition. Second CERN-GTU collaboration meeting PMBC2023, 6 – 10 November 2023, GTU, Tbilisi, Georgia. https://indico.cern.ch/event/1334518/timetable/
3. **T. Tsmindashvili, N. Kukhilava, S. Katamadze, R. Kalandadze, L. M. Ferrari, P. Müller, B. E. Wirth**. Evaluation in EEG Emotion Recognition: State-of-the-Art Review and Unified Framework. Second CERN-GTU collaboration meeting PMBC2023, 6 – 10 November 2023, GTU, Tbilisi, Georgia. https://indico.cern.ch/event/1334518/timetable/
4. **S. Katamadze, T. Tsmindashvili, N. Kukhilava, R. Kalandadze, L. M. Ferrari, P. Müller, B. E. Wirth**. Enhancing Emotion Recognition: EEG Evaluation and AI Models. DataFest Tbilisi 2023. 9 – 11 November 2023, Tbilisi, Georgia. www.datafest.ge
5. M. Kublashvili, **Z. Sanikidze, T. Saghinadze**, M. Kublashvili. On the Mathematical Aspects of the Numerical Calculation of Engineering Constructions Weakened by Cracks. Second CERN-GTU collaboration meeting PMBC2023, 6 – 10 November, 2023, GTU, Tbilisi, Georgia. https://indico.cern.ch/event/1334518/timetable/
6. M. Zakradze, **Z. Tabagari**. Numerical analysis of some problems related to the calculation of electrostatic fields. Second CERN-GTU collaboration meeting PMBC2023, 6 – 10 November 2023, GTU, Tbilisi, Georgia. https://indico.cern.ch/event/1334518/timetable/
7. **V. Kvaratskhelia, G. Giorgobiani, M. Menteshashvili.** On one connection between the moments of random variables. Computer Science and Information Technologies CSIT 2023, September 25 - 30, 2023, Yerevan, Armenia. https://www.csit.am/2023/schedule.php
8. **Kachiashvili K., Kachiashvili J., Kalandadze R., Kvaratskhelia V**. Automatic Diagnosis of Lung Disease on the Basis of an X-Ray Images of a Patient with Given Reliability. XIII International Conference of the Georgian Mathematical Union, Batumi, September 4-9, 2023. p. 140. http://gmu.gtu.ge/conferences/wp-content/uploads/2023/09/Conference_GMU_2023_01.09.pdf
9. **Kachiashvili K., Kachiashvili J., Kalandadze R., Kvaratskhelia V**. Automatic Diagnosis of Diseases on the Basis of an Irradiation Images of a Patient with Restrictions Both Type of Errors. The 4th International Conference on Modern Management based on Big Data (MMBD2023), August 1-4, 2023, Seoul, South Korea. p. 19 http://www.mmbdconf.org/
10. **Kachiashvili K., Kachiashvili J., Kalandadze R., Kvaratskhelia V**. The automatization of the medical diagnosis on the basis of an X-ray image of a patient with the restrictions of both possible errors on the desired levels. The International Conference "Distributed Computing and Grid Technologies in Science and Education", 3-7 July, 2023, Dubna, Russia. JINR Meshcheryakov Laboratory of Information Technologies. https://indico.jinr.ru/event/3505/attachments/16120/27954/program_GRID2023.pdf
11. **G. Giorgobiani**, G. Chelidze, V. Tarieladze. Rearrangement universality of the Dirichlet type series in a complex field. 14th ISAAC Congress. Ribeirão Preto, University of Sao Paulo, Brazil, 17 – 21 July, 2023. https://dcm.ffclrp.usp.br/isaac/abstracts.pdf

12. **V. Kvaratskhelia, G. Giorgobiani, M. Menteshashvili.** On one connection between the moments of random variables. Int. Conf. Applications of Stochastic Processes and Mathematical Statistics to Financial Economics and Social Sciences. November 15 –16, 2023, Tbilisi, Georgia
13. **V. Kvaratskhelia, G. Giorgobiani,** V. Tarieladze. Subgaussian Random Elements in Infinite Dimensional Spaces. XIII International Conference of the Georgian Mathematical Union. September 4 – 9, 2023, Batumi, Georgia. Book of Abstracts, p. 116.
http://gmu.gtu.ge/conferences/wp-content/uploads/2023/10/Conference_GMU_2023_7.10_last.pdf
14. **V. Kvaratskhelia, G. Giorgobiani, M. Menteshashvili.** On one connection between the moments of random variables. Ninth International Conference on Statistics for Twenty-first Century - 2023 (ICSTC-2023). 15-18 December 2023, Kerala University, India. https://sites.google.com/view/icstc2023/home
15. **V. Kvaratskhelia, G. Giorgobiani,** V. Tarieladze. Subgaussian Random Elements in Infinite Dimensional Spaces. The Fourth International Conference "Modern Problems in Applied Mathematics" dedicated to the 105th Anniversary of I. Javakhishvili Tbilisi State University (TSU) & 55th Anniversary of I. Vekua Institute of Applied Mathematics (VIAM). September 13-15, 2023, Tbilisi, Georgia.
16. **S. Katamadze, T. Tsmindashvili, N. Kukhilava, R. Kalandadze, L. M. Ferrari, P. Müller, B. E. Wirth**. Enhancing Emotion Recognition: EEG Evaluation and AI Models. Online I En, Data Zen Community. December 20, 2023, Tbilisi, Georgia. https://wearecommunity.io/events/enhancing-emotion-recognition-eeg-evaluation-and-ai-models
17. S. Katamadze, T. Tsmindashvili, N. Kukhilava, R. Kalandadze, L. M. Ferrari, P. Müller, B. E. Wirth. Enhancing Emotion Recognition: EEG Evaluation and AI Models. Online I Georgian language. March 22th, 2024, Tbilisi, Georgia. https://wearecommunity.io/events/enhancing-emotion-recognition-eeg-evaluation-and-ai-models-ge?utm_source=facebook&utm_medium=social&utm_campaign=ta?fbclid=IwAR0-mmpPbgQOhrvfV1T3y6awSz3Z9xFadOpKyYg0dS1k0M-rBYFcGai7cAM

Presentations at the First Summer School

(available at https://www.gain-twinning.eu/?page_id=302)

1. Evaluation in EEG Emotion Recognition: State-of-the-Art Review and Unified Framework.
Sub-task 1: AI Technologies for Human Behaviour Understanding, Emotions (face crop video, biosignals). Group 1: R. Kalandadze, N. Kukhilava, T. Tsmindashvili, S. Katamadze. Supervised by: F. Bremond, L. Ferrari (INRIA), P. Muller, B. Wirth (DFKI). Presented by R. Kalandadze and S. Katamadze.
2. The newest achievements of AI in Emotion Recognition from Human Body Movements. *Sub-task 2: AI Technologies for Human Behaviour Understanding, Emotions (full body video)*
Group 2: I. Katchiasvhili, L. Tabagari. Supervised by: P. Muller, B. Wirth (DFKI). Presented by I. Kachashvili.
3. Exploring Image Captioning with Parameter-Efficient Transfer Learning for Vision Transformers. *Sub-task 3: Action detection/recognition (AI Technologies for Human Behaviour Understanding). Group 3: T. Saghinadze. Supervised by: F. Bremond, T. Agraval (INRIA). Presented by T. Saghinadze.*
4. Georgian Pretrained Language Understanding Model. *Sub-task 5: AI Methods for Deep Speech Analysis in Health, NLP. Group: B. Mikaberidze, B. Tepnadze. Supervised by: P. Muller, H. Lindsay (DFKI). Presented by B. Mikaberidze.*
5. Studying neural correlates of speech production using fMRI and NLP. *Sub-task 6: AI Methods for Deep Speech Analysis in Health - NLP-fMRI. Group 6: T. Giorgobiani, S. Tsagareishvili. Supervised by: P. Muller, H. Lindsay, B. Wirth (DFKI). Presented by T. Giorgobiani.*