Report with the analysis carried out on technical specifications of collection of data from heterogeneous sources.

"Improving the quality and sustainability of learning using early intervention methods based on learning analytics"

Project No. 2023-1-FI01-KA220-HED-000159757

Project ref. number	2023-1-FI01-KA220-HED-000159757
Project title	**ISILA - Improving the quality and sustainability of learning using early intervention methods based on learning analytics**
Document title	Report with the analysis carried out on technical specifications of collection of data from heterogeneous sources
Document Type	Report
Document version	v1
Previous version(s)	<Noto Serif 10>
Planned date of delivery	March 1, 2024
Language	English
Dissemination level	Public
Number of pages	6
Partner responsible	UEF
Author(s)	Ramy Elmoazen
With contributions by:	Sonsoles López Pernas
Revised by:	Sonsoles López Pernas Kamila Misiejuk (UiB)
Abstract	<Noto Serif 10>
Keywords	<Noto Serif 10>

Table of Content

Introduction

The ISILA project aims to combine data from multiple sources to provide teachers with relevant information on their students, enabling timely interventions to improve learning outcomes. In the first stage, each partner investigates relevant data sources for their teaching implementations. This report details the technical specifications for collecting data from these heterogeneous sources and outlines the technical requirements for integrating learner data into a common online platform.

1 Choice of Courses and Possible Data Sources

Each partner in the ISILA project investigates relevant data sources for their teaching implementations.

1.1 University of Eastern Finland (UEF)

1.1.1 Web Programming (early semester 1)

The Web Programming course, conducted in the Fall semester is characterized by face-to-face practical exercises. Data sources for this course include Moodle, GitHub, Audience Response Systems, surveys, LLMs, and videos.

1.1.2 Data Management Systems (early semester 2)

The Data Management Systems course, held early in the second semester, also involves face-to-face practical exercises about databases, with data sourced from Moodle, ARS, games, exercises, surveys, LLMs, and AI.

1.1.3 Social Network Analysis (late semester 2)

The Social Network Analysis course, conducted online late in the second semester, utilizes Moodle, LLMs, Discord, and surveys as data sources.

1.2 University of León (ULe)

1.2.1 Computer Architecture (first semester)

The Computer Architecture course in the first semester employs a flipped classroom model, with data collected from surveys, Moodle logs, and teamwork analysis via Telegram or Discord.

1.2.2 Web Applications

The Web Applications course, characterized by project-based learning, gathers data from surveys and Moodle logs

1.3 University of Bergen (UiB)

1.3.1 DIGI110: Fantastic Data

The Fantastic Data course will be conducted online, with data sources including Canvas logs, navigation of learning materials on external websites, and surveys.

1.3.2 STAT110: Basic Course in Statistics

The Basic Course in Statistics is a blended course, with data collected from Canvas logs (including video data), and surveys.

1.4 Belgrade Metropolitan University (BMU)

1.4.1 Face-to-Face Courses with Teaching Materials on LAMS

Courses will be conducted physically with teaching materials available on LAMS, with data sourced from the educational management system, surveys, LAMS data, chat/Discord, and videos. Planned courses for the piloting are Object oriented programming 1 and Foundations of Web Development.

1.5 Sofia University (SU)

1.5.1 Human-Computer Interaction

The human-computer interaction course is a project-based learning course, with data collected from surveys, Moodle logs, and videos Digital Design and Multimedia - Digital MediaDigital Design and Multimedia.

The digital design and multimedia course is a project-based learning course, with data collected from surveys, Moodle logs, and videos.

1.6 Summary

Combining all requirements together, the data sources relevant to the project ISILA as a whole are as follows:

	LMS	Video	Survey	Discord	Games	External websites	LLMs
UEF	Moodle	✕	✕	✕	✕		✕
ULe	Moodle	✕	✕	✕			
UiB	Canvas	✕	✕			✕	
BMU	LAMS	✕	✕	✕			
SU	Moodle	✕	✕				

2 Technical Requirements for Data Sources

2.1 LMS logs

LMS logs are automatically collected by the vendors used by the participating institutions (Moodle, LAMS, Canvas). There are plugins to convert the logs to xAPI and send them to an LRS.

2.2 Video

Videos are embedded on the LRS from YouTube and Google Drive. Detection of section clicks is possible, but no data on video duration from LAMS. Further development is needed for data extraction of fine-grained video watching data.

2.3 Escape Room games

Logs can be downloaded from the Escapp platform (https://github.com/ging/escapp) but they need to be converted to xAPI format and uploaded to an LRS.

2.4 Discord/Telegram

Discord data can be collected using a Discord bot, with an alternative being the Matrix protocol, an open protocol for decentralized, secure communications that can be connected with an API to most instant messaging apps. Telegram can be used instead, with the data gathered using Telegram bots, with the Matrix protocol as an alternative. The conversation data needs to be forwarded to an LRS.

2.5 Survey

There is potential to use a Discord bot for courses using Discord. The LMS records survey completion but not the content in xAPI format. The survey response data can be downloaded but needs to be converted to xAPI beforehand.

2.6 External websites

Navigation data can be collected from websites that present students with learning materials. After a student has finished reading a section, they can check that they have finished this task and solve an optional quiz. This data can be collected using APIs and uploaded to an LRS.

2.7 LLMs

Conversation data between students and LLMs cannot currently be easily accessed or retrieved, much less in xAPI format, needed for storing it in the lRS.