

**UNIVERSIDADE FEDERAL DE PELOTAS**  
**Centro de Desenvolvimento Tecnológico**  
**Programa de Pós-Graduação em Computação**



Dissertação

**Integração de Dados Educacionais: Learning Record Warehouse como Base  
para Learning Analytics**

**Juary Costa-Rocha**

Pelotas, 2025

**Juary Costa-Rocha**

**Integração de Dados Educacionais: Learning Record Warehouse como Base  
para Learning Analytics**

Dissertação apresentada ao Programa de Pós-Graduação em Computação do Centro de Desenvolvimento Tecnológico da Universidade Federal de Pelotas, como requisito parcial à obtenção do título de Mestre em Ciência da Computação.

Orientador: Prof. Dr. Cristian Cechinel

Pelotas, 2025

Universidade Federal de Pelotas / Sistema de Bibliotecas  
Catalogação da Publicação

R672i Rocha, Juary Costa

Integração de dados educacionais [recurso eletrônico] : Learning Record Warehouse como base para Learning Analytics / Juary Costa Rocha ; Cristian Cechinel, orientador. — Pelotas, 2025.  
110 f.

Dissertação (Mestrado) — Programa de Pós-Graduação em Computação, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, 2025.

1. Análise de aprendizagem. 2. Intercâmbio de dados. 3. Caliper Analyzer. 4. Learning Record Warehouse. 5. Interoperabilidade. I. Cechinel, Cristian, orient. II. Título.

CDD 005

**Juary Costa-Rocha**

**Integração de Dados Educacionais: Learning Record Warehouse como Base para Learning Analytics**

Dissertação aprovada, como requisito parcial, para obtenção do grau de Mestre em Ciência da Computação, Programa de Pós-Graduação em Computação, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas.

**Data da Defesa:** 21 de Março de 2025

**Banca Examinadora:**

Prof. Dr. Cristian Cechinel (orientador)

Doutor em Ciência da Computação pela Universidad de Alcalá.

Prof. Dr. Emanuel Marques Queiroga.

Doutor em Ciência da Computação pela Universidade Federal de Pelotas.

Prof. Dr. Tiago Thompsen Primo.

Doutor em Ciência da Computação pela Universidade Federal do Rio Grande do Sul.

Prof. Dr. Vinicius Ramos Faria Culmant

Doutor em Engenharia de Sistemas e Computação pela Universidade Federal do Rio de Janeiro.

## **AGRADECIMENTOS**

Agradeço aos meus pais, Erineu Rocha e Margarida Costa, por todo o apoio incondicional e por me permitirem ser a pessoa que sou hoje – sem eles, nada disso seria possível. Sou igualmente grato aos meus irmãos, Kelly, Mario e Mira, pela força e incentivo ao longo dessa jornada. Agradeço ao meu Orientador Prof. Cristian pelo suporte prestado durante este percurso.

Um agradecimento especial à Emilcy, cuja ajuda e apoio foram fundamentais durante todo o processo de desenvolvimento desta dissertação. Sem ela, esta conquista não teria sido possível.

Agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (auxílios 409633/2022-4, 305731/2021-1) e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) por meio do Programa de Pós-Graduação em Computação da Universidade Federal de Pelotas (UFPel - Chamada Interna PPGC nº 01/2022) pelo financiamento deste trabalho.

*Lança na vida, um dia de cada vez...* — NEU E GUIDA

## RESUMO

COSTA-ROCHA, Juary. **Integração de Dados Educacionais: Learning Record Warehouse como Base para Learning Analytics**. Orientador: Cristian Cechinel. 2025. 111 f. Dissertação (Mestrado em Ciência da Computação) – Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, Pelotas, 2025.

Esta dissertação de mestrado apresenta uma proposta para a integração de dados educacionais usando um Learning Record Warehouse (LRW) como base para o Learning Analytics (LA). O objetivo principal é desenvolver uma solução que colete e armazene dados de aprendizagem de forma padronizada, usando especificações de interoperabilidade, buscando melhorar a compatibilidade entre sistemas de aprendizagem digital para facilitar a análise de dados. A pesquisa é justificada pela crescente adoção do e-learning e pela necessidade de rastrear e analisar dados educacionais. Ela busca implementar um LRW para coletar e armazenar atividades de aprendizagem de forma padronizada, com objetivos específicos, como estudar as tecnologias disponíveis, caracterizar as plataformas educacionais, desenvolver um modelo de Learning Record Warehouse e estabelecer um modo de validação. O documento explora conceitos-chave como Learning Analytics (LA), interoperabilidade e especificações de e-learning, como xAPI e Caliper Analytics. xAPI é apresentada como uma especificação para rastrear qualquer atividade de aprendizagem, enquanto o Caliper Analytics se concentra na coleta de dados com foco em métricas. O Learning Record Storage (LRS) como repositórios para xAPI e processos de extrair, transformar e carregar (ETL) para integração de dados também são abordados. Foi realizada uma análise sistemática da literatura, identificando desafios e tendências na interoperabilidade de dados educacionais, e constatou-se que, embora existam especificações como xAPI e Caliper, seu uso não é generalizado. A metodologia usada é a adaptada da Design Science Research (DSR), com experimentos iterativos para criar e avaliar a arquitetura proposta. No primeiro experimento, o OpenLRW é instalado, mas são encontrados problemas com a conversão de dados. No segundo experimento, a capacidade de enviar dados do Caliper para o LRS é testada e um conversor do Caliper para xAPI é desenvolvido. A arquitetura final inclui esse conversor para armazenar dados em um LRS, usando o PostgreSQL como banco de dados. Destaca-se a importância de manter a carga útil original do Caliper para evitar a perda de dados. Os resultados mostram a capacidade de consultar e visualizar dados no formato JSON, usando ferramentas como o Apache Superset. Conclui-se que a interoperabilidade é essencial para a integração de dados de várias fontes, e o LRW proposto visa ser um modelo de transição entre os sistemas estabelecidos e a xAPI, que é um padrão da IEEE. Como

trabalho futuro, propõe-se a integração de um LRW para alimentar sistemas capazes de implementar modelos preditivos, como a previsão do risco de evasão escolar.

Palavras-chave: Análise de Aprendizagem; Caliper Analyzer; Interoperabilidade; Intercâmbio de dados; Learning Record Warehouse; xAPI.

## ABSTRACT

COSTA-ROCHA, Juary. **Educational Data Integration: Learning Record Warehouse as the Basis for Learning Analytics**. Advisor: Cristian Cechinel. 2025. 111 f. Dissertation (Masters in Computer Science) – Technology Development Center, Federal University of Pelotas, Pelotas, 2025.

This master's dissertation presents a proposal for the integration of educational data using a Learning Record Warehouse (LRW) as the basis for Learning Analytics (LA). The main objective is to develop a solution that collects and stores learning data in a standardized way, using interoperability specifications to improve compatibility between digital learning systems and facilitate data analysis. The research is justified by the growing adoption of e-learning and the need to track and analyze educational data. It aims to implement an LRW to collect and store learning activities in a standardized way, with specific objectives such as studying available technologies, characterizing educational platforms, developing a Learning Record Warehouse model, and establishing a validation method. The document explores key concepts such as Learning Analytics (LA), interoperability, and e-learning specifications, including xAPI and Caliper Analytics. xAPI is presented as a specification for tracking any learning activity, while Caliper Analytics focuses on data collection with an emphasis on metrics. The Learning Record Store (LRS) as a repository for xAPI and Extract, Transform, Load (ETL) processes for data integration are also discussed. A systematic literature review was conducted to identify challenges and trends in educational data interoperability, revealing that although specifications such as xAPI and Caliper exist, their use is not widespread. The methodology follows an adapted Design Science Research (DSR) approach, with iterative experiments to create and evaluate the proposed architecture. In the first experiment, OpenLRW is installed, but issues arise with data conversion. In the second experiment, the ability to send Caliper data to the LRS is tested, and a Caliper-to-xAPI converter is developed. The final architecture includes this converter to store data in an LRS, using PostgreSQL as the database. Emphasis is placed on preserving the original Caliper payload to prevent data loss. The results demonstrate the ability to query and visualize data in JSON format using tools such as Apache Superset. The study concludes that interoperability is essential for integrating data from multiple sources, and the proposed LRW aims to serve as a transition model between established systems and xAPI, which is an IEEE standard. As future work, the integration of an LRW is proposed to feed systems capable of implementing predictive models, such as forecasting the risk of student dropout.

Keywords: Caliper Analyzer; Data Interchange; Interoperability; Learning Analytics; Learning Record Warehouse; xAPI.

## LISTA DE FIGURAS

Figura 1	Modelo sofisticado de LA. Fonte: (Siemens; Dawson; Lynch, 2013)	23
Figura 2	Padrões e Organizações de Sistemas de E-Learning Fonte: Bakhouyi et al. . . . . .	27
Figura 3	xAPI Statement - Declaração básica. Fonte: Autor . . . . .	29
Figura 4	Representação json de xAPI statement. Fonte: Autor . . . . .	30
Figura 5	Estrutura de xAPI. Fonte: (ADL; Johnson, 2024) . . . . .	31
Figura 6	Rastreamento de Experiencias com xAPI. Fonte: (ADL; Johnson, 2024) . . . . .	31
Figura 7	Evolução do xAPI: Do Conceito ao Padrão. Fonte: Adaptado de (Bakhouyi et al., 2017) . . . . .	33
Figura 8	Estrutura do Caliper Analytics. Fonte: (1EdTech, 2023) . . . . .	34
Figura 9	Coleta dados xapi. Fonte: Adaptado de (Väkevã, 2023) . . . . .	41
Figura 10	Etapas do processo de revisão. Fonte: Autor . . . . .	45
Figura 11	Palavras-chave e sinônimos. Fonte: Autor . . . . .	45
Figura 12	Documentos encontrado no Scopus por área temática. Fonte: Autor	48
Figura 13	Documentos por país ou território na Scopus . Fonte: Autor . . . . .	49
Figura 14	Categorias da Web of Science. Fonte: Autor . . . . .	49
Figura 15	Citações e publicações ao longo do tempo da Web of Science. Fonte: Autor . . . . .	49
Figura 16	Estudos selecionados de publicação do ano da linha de tendência da Web of Science. Fonte: Autor . . . . .	50
Figura 17	Arquiterura LRW. Fonte: Autor. . . . .	59
Figura 18	Passos de desenvolvimento do trabalho. Fonte: Autor. . . . .	60
Figura 19	Configurando o Caliper plugin. Fonte: Autor . . . . .	62
Figura 20	Credencial no LRSQL. Fonte: Autor. . . . .	64
Figura 21	Fluxograma do Conversor. Fonte: Autor. . . . .	66
Figura 22	Envio do Caliper payload para o LRSQL. Fonte: Autor. . . . .	67
Figura 23	Payloads Original Caliper . Fonte: Autor . . . . .	68
Figura 24	Payloads Convertido Caliper . Fonte: Autor . . . . .	69
Figura 25	Configurando o Caliper plugin. Fonte: Autor . . . . .	70
Figura 26	Consulta ao banco pelo payload xAPI. Fonte: Autor . . . . .	70
Figura 27	Configuração plugin xAPI no Moodle. Fonte: Autor . . . . .	77
Figura 28	Dados coletados do Moodle armazenados no LRW. Fonte: Autor . .	78
Figura 29	Consulta ao banco utilizando o Apache Superset. Fonte: Autor . . .	78

Figura 30	Dashboard criado a partir dos dados armazenados no LRW. Fonte: Autor . . . . .	79
Figura 31	Visualização da query no Superset . . . . .	101
Figura 32	Visualização no SuperSet das Atividades mais acessadas. . . . .	103

## LISTA DE TABELAS

Tabela 1	Perfis de Métricas do Caliper. Fonte: (1EdTech, 2023). . . . .	36
Tabela 2	Comparação entre Caliper Analytics e xAPI . . . . .	39
Tabela 3	Lista de Verificação para Avaliação da Qualidade . . . . .	46
Tabela 4	Formulário de Extração de Dados . . . . .	47
Tabela 5	Estudos Extraídos por Banco de Dados . . . . .	48
Tabela 6	Comparação de estrutura entre Caliper Event e xAPI Statement . .	65
Tabela 7	Resumo dos eventos suportados pelos plugins Caliper e xAPI. Para a lista completa, consultar o Apêndice C . . . . .	76
Tabela 8	Eventos suportados pelos plugins Caliper e xAPI . . . . .	104

## LISTA DE ABREVIATURAS E SIGLAS

AVA	Ambientes Virtuais de Aprendizagem
ADL	Advanced Distributed Learning Initiative
ABNT	Associação Brasileira de Normas Técnicas
LA	Learning Analytics
SPMD	Single Program Multiple Data
SCORM	Sharable Content Object Reference Model
API	Application Programming Interface
xAPI	Experience API
UFPeI	Universidade Federal de Pelotas
LMS	Learning Management Systems
LRS	Learning Record Storage
LRW	Learning Record Warehouse
AICC	Aviation Industry CBT Committee
IEEE	Institute of Electrical and Electronics Engineers
LTI	Learning Tools Interoperability
CMI	Computer Managed Instruction
CC	Common Cartridge
LRP	Learning Record Provider
LRC	Learning Record Consumer
TIC	Tecnologia da Informação e Comunicação
ETL	Extract, Transform, Load

# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	16
1.1	Justificativa e Relevância	17
1.2	Objectivo Geral	18
1.3	Objectivos Específicos	18
1.4	Pergunta de Pesquisa	19
<b>2</b>	<b>REFERENCIAL TEÓRICO E TECNOLÓGICO</b>	20
2.1	Learning Analytics	20
2.2	Interoperabilidade de dados educacionais	21
2.3	Especificações e Padrões de E-Learning	23
2.3.1	E-learning	24
2.3.2	Especificações e Padrões	25
2.3.3	Experience API	28
2.3.4	Caliper Analytics	33
2.3.5	Diferença entre Xapi e Caliper Analytics	37
2.4	Learning Record Storage	40
2.5	ETL	42
<b>3</b>	<b>REVISÃO DA LITERATURA</b>	44
3.1	Metodologia da revisão sistemática	44
3.2	Resultados: Análise Bibliométrica	47
3.3	Síntese Dos Trabalhos Encontrados	50
<b>4</b>	<b>CONCEPÇÃO DO TRABALHO</b>	58
4.1	Metodologia	58
4.2	Experimento I	60
4.2.1	Softwares Adicionais Necessários	61
4.2.2	Processo de Instalação e Configuração	61
4.2.3	Resultados	62
4.3	Experimento II	63
4.3.1	Softwares Adicionais Necessários	64
4.3.2	Processo de Instalação, Configuração e Desenvolvimento	64
4.3.3	Resultados	66
<b>5</b>	<b>ARQUITETURA PROPOSTA</b>	71
5.1	Descrição da arquitetura	71
5.1.1	Componentes Principais	72
5.2	Implementação do LRS	73

<b>5.3</b>	<b>Eventos e dados Reais coletados</b>	74
<b>6</b>	<b>RESULTADOS E DISCUSSÃO</b>	75
<b>6.1</b>	<b>Resultados</b>	75
6.1.1	Capacidade de coleta, armazenamento e recuperação	75
<b>6.2</b>	<b>Discussão</b>	80
<b>7</b>	<b>CONSIDERAÇÃO FINAIS</b>	82
<b>7.1</b>	<b>Conclusões</b>	82
<b>7.2</b>	<b>Recomendações</b>	83
<b>7.3</b>	<b>Trabalhos Futuros</b>	84
<b>7.4</b>	<b>Publicações Acadêmicas</b>	84
7.4.1	Artigos Publicados	85
7.4.2	Participação de Intercambio Acadêmico	85
	<b>REFERÊNCIAS</b>	86
	<b>APÊNDICE A ARQUIVOS TÉCNICOS</b>	95
<b>A.1</b>	<b>docker-compose para configurar o LRW</b>	95
<b>A.2</b>	<b>docker-compose para configurar o Apache SuperSet</b>	96
	<b>APÊNDICE B QUERYS DE CONSULTA</b>	100
<b>B.1</b>	<b>Atividades dos estudantes</b>	100
<b>B.2</b>	<b>Atividades Visualizadas</b>	101
	<b>APÊNDICE C TABELAS DE EVENTOS</b>	104

# 1 INTRODUÇÃO

A crescente integração de tecnologias em ambientes educacionais tem transformado significativamente a forma como a aprendizagem é gerenciada, monitorada e analisada. Nesse contexto, as Learning Analytics (LA) emergem como um conjunto de métodos que possibilitam a coleta, medição e análise de dados sobre estudantes e seus contextos, com o objetivo de compreender e melhorar os processos educacionais (Dondorf, 2022). Uma das questões centrais em sistemas de LA é a capacidade de integração e interoperabilidade entre diferentes fontes de dados, permitindo que informações provenientes de diversos sistemas sejam analisadas de forma padronizada e sem perda de significado (Comissão Europeia, 2010).

A interoperabilidade, definida como a capacidade de sistemas compartilharem informações e aplicações entre si (Bishr, 1997; Sheth, 1999; Sayão; Marcondes, 2008), é essencial para a implementação eficaz de Learning Analytics. Em particular, a interoperabilidade semântica ganha destaque, pois garante que os dados mantenham seu significado original ao serem transferidos entre sistemas. Para alcançar esse nível de interoperabilidade, é necessário padronizar os dados por meio de especificações como a xAPI ou o Caliper Analytics, que permitem a tradução ou integração de diferentes vocabulários em um formato comum compreensível por todos os sistemas envolvidos (Farinelli; Almeida, 2014).

A interoperabilidade tornou-se uma questão-chave na comunidade de LA desde a proposta de Siemens et al. (2011), pois a interoperabilidade é a capacidade de um sistema trocar informações com outros, garantindo compatibilidade entre os diferentes sistemas utilizados no processo, independentemente de quem os desenvolveu (Furniel; Mendonça; Silva, 2016; Sayão; Marcondes, 2008; Bishr, 1997; Sheth, 1999). No entanto, como sugerido por Siemens; Dawson; Lynch (2013), este é um processo gradual, que envolve a progressão por diferentes estágios de maturidade.

As plataformas de e-learning atuais, embora eficientes em certos aspectos, são limitadas na coleta de dados granulares e no acompanhamento de atividades de aprendizagem mais complexas, especialmente aquelas que ocorrem fora do ambiente tradicional dos Ambientes Virtuais de Aprendizagem (Väkevã, 2023).

As especificações como a *Experience API* (xAPI) e o *Caliper Analytics*, são reconhecidas no domínio educacional por possibilitarem a representação padronizada das ações realizadas pelos usuários em diferentes contextos. Cada ação é registrada como uma "*activity*", composta por atributos como "*actor*"(quem realizou a ação) e "*verb*"(a ação realizada), formando um "*stream*"que descreve a sequência de atividades do usuário (Serrano-laguna et al., 2017). Essas especificações promovem não apenas a interoperabilidade semântica, mas também o armazenamento estruturado de dados por meio de soluções como o Learning Record Storage (LRS), definido no âmbito do xAPI, ou Learning Record Warehouse (LRW) como o OpenLRW, que suportava múltiplas especificações (Apereo, 2019).

Apesar das vantagens da adoção dessas especificações, os sistemas educacionais ainda enfrentam desafios significativos, como a integração de dados de fontes heterogêneas, requisitos de armazenamento, transmissão, processamento e rastreabilidade (Samuelsen; Chen; Wasson, 2021a). Tais desafios destacam a necessidade de sistemas robustos, planejados especificamente para o contexto educacional, que suportem não apenas a interoperabilidade semântica, mas também ofereçam infraestrutura para armazenamento e análise eficiente de dados.

Portanto, este trabalho apresenta a implementação de um Learning Record Warehouse (LRW) como solução para coleta e armazenamento padronizado de atividades de aprendizagem no ambiente virtual de aprendizagem, utilizando especificações de interoperabilidade de dados conhecidas no domínio das Learning Analytics.

## 1.1 Justificativa e Relevância

As motivações para a execução deste trabalho estão vinculadas à crescente adoção do ensino digital, que pode ocorrer de forma totalmente online (Educação a Distância) ou no formato híbrido, também conhecido como aprendizagem combinada. Essa tendência tem transformado o ambiente de ensino digital e os sistemas de gestão de aprendizagem em componentes essenciais das atividades educacionais diárias.

A pandemia de COVID-19, em 2020, acelerou significativamente essa transição, deslocando quase toda a educação para o formato online (Dondorf, 2022). Esse contexto impulsionou o interesse por ferramentas de apoio ao ensino e aprendizagem digital (Dondorf, 2022), gerando um grande volume de dados que necessitam de sistemas de armazenamento projetados para integrar informações provenientes de diversas fontes (Samuelsen; Chen; Wasson, 2021a).

Em ambientes de aprendizagem, Learning Analytics (LA) oferece métodos para medir, coletar e analisar dados sobre estudantes e seus contextos (Dondorf, 2022), auxiliando as instituições a compreender e aprimorar o processo educacional. Um aspecto crucial na coleta de dados em sistemas de LA é garantir a interoperabilidade

semântica entre diferentes sistemas de informação, permitindo que dados de fontes internas e externas sejam processados por uma aplicação central, preservando seu sentido original.

Muitas instituições educacionais, seja por iniciativa própria ou necessidade, incorporaram o e-learning como um componente essencial de seus processos de ensino. Essas transformações abriram a oportunidade de rastrear e coletar volumes crescentes de dados relacionados à aprendizagem, proporcionando uma compreensão mais abrangente dos alunos e de seus processos de aprendizagem. Entretanto, para que os diversos conteúdos de e-learning sejam rastreados, armazenados e analisados de forma eficaz, é fundamental que os sistemas e aplicativos ofereçam flexibilidade e interoperabilidade.

Para atender aos objetivos e expectativas futuras no âmbito educacional, torna-se essencial rastrear, coletar e analisar, de maneira eficiente, volumes cada vez maiores de dados educacionais, e neste trabalho se visa explorar ferramentas que permitem realizar essas ações.

## **1.2 Objectivo Geral**

O objetivo geral deste trabalho é desenvolver e implementar uma solução baseada em um Learning Record Warehouse (LRW) para coletar e armazenar atividades de aprendizagem de forma padronizada, utilizando especificações de interoperabilidade de dados.

Essa solução busca aprimorar a compatibilidade entre sistemas de aprendizagem digital e facilitar a análise e interpretação dos dados educacionais.

## **1.3 Objectivos Específicos**

Este trabalho busca alcançar os seguintes objetivos específicos:

1. Realizar um estudo detalhado das tecnologias disponíveis por meio de revisão de literatura e benchmarking, com o intuito de identificar elementos relevantes e potenciais melhorias que contribuam para o suporte à tomada de decisão em ambientes educacionais.
2. Caracterizar as plataformas e sistemas de informação educacionais que poderão ser modelados no Learning Data Warehouse, definindo o escopo e o alcance deste trabalho com base nas necessidades e possibilidades do contexto analisado.
3. Desenvolver um modelo de Learning Data Warehouse baseado na análise de tecnologias e fontes de informação educacionais. O modelo deve permitir a pa-

dronização e centralização de dados, possibilitando a aplicação de técnicas de Learning Analytics de forma eficiente e consistente.

4. Integrar o Learning Data Warehouse ao serviço de Learning Analytics, explorando as vantagens oferecidas pela utilização, como escalabilidade, flexibilidade e eficiência no processamento e armazenamento de dados.
5. Estabelecer parâmetros de validação para o Learning Data Warehouse proposto, utilizando métricas específicas que permitam avaliar o impacto, a eficácia e os benefícios da solução em cenários reais de aplicação.

Considerando os aspectos apresentados, este trabalho busca fornecer uma solução que integre tecnologias, de modo a promover avanços na área de Learning Analytics e auxiliar instituições educacionais a otimizar processos e resultados.

## 1.4 Pergunta de Pesquisa

Este trabalho busca responder à seguinte pergunta de pesquisa:

**"Como um Learning Record Warehouse pode ser integrado a uma Sistema de Learning Analytics de modo suportar o recebimento de dados provenientes de diferentes Ambientes Virtuais de Aprendizagem e diversas outras fontes?"**

A resposta a essa pergunta permitirá entender como integrar soluções de interoperabilidade para Learning Analytics em ambientes digitais de aprendizagem, promovendo maior compatibilidade entre sistemas e eficiência na análise de dados educacionais. A implementação de um Learning Record Warehouse, integrada a uma solução de Learning Analytics, poderá suportar a coleta e análise padronizada de dados provenientes de múltiplas fontes. Essa abordagem tem o potencial de aprimorar a interoperabilidade dos sistemas, proporcionando uma visão holística do desempenho acadêmico e possibilitando intervenções mais eficazes. Para explorar essa questão, seguiremos os objetivos destacados anteriormente.

## **2 REFERENCIAL TEÓRICO E TECNOLÓGICO**

Neste capítulo, são discutidos os principais conceitos que fundamentam este estudo, abrangendo Learning Analytics, Interoperabilidade, Especificações e Padrões de E-Learning, bem como Learning Record Storage e ETL.

### **2.1 Learning Analytics**

Learning Analytics (LA) é uma área de pesquisa emergente que começou a se desenvolver no início dos anos 2000 e se dedica à coleta, medição, análise e relato de dados relacionados ao aprendizado e aos contextos nos quais ele ocorre (Dondorf et al., 2019). Com o avanço das tecnologias digitais e a crescente disponibilidade de dados educacionais, LA surge como uma abordagem promissora para aprimorar a eficácia do ensino e personalizar a experiência de aprendizado dos estudantes. A definição mais aceita de Learning Analytics considera “a medição, coleta, análise e relato de dados sobre os estudantes e seus contextos, para propósitos de compreender e otimizar a aprendizagem no ambiente em que ocorre” (Siemens et al., 2011). Esta prática tem como fundamento o uso de técnicas avançadas de análise de dados para fornecer insights sobre o desempenho dos estudantes, identificar padrões de comportamento e prever resultados acadêmicos. O objetivo principal é entender e otimizar os processos educacionais, proporcionando uma base sólida para a tomada de decisões fundamentada por dados e no desenvolvimento de estratégias pedagógicas mais eficazes. LA não se refere apenas à interpretação de um grande volume de dados produzidos pelos estudantes, mas ao processo, concentrando-se em como ele ocorre, que é um ciclo com um fim indeterminado, geralmente composto de coleta e pré-processamento de dados, aplicação de métodos e avaliação, intervenção e aprendizado sobre como o processo ocorreu (Romero; Ventura, 2020).

Nos sistemas que integram LA, os dados de interação dos estudantes poderão ser coletados de fontes heterogêneas e normalmente mostrados em painéis para a visualização de indicadores (Joarder et al., 2023; Dondorf; Nacken, 2017). Esses painéis permitem que educadores e administradores acompanhem o progresso dos estudan-

tes, e desta forma identifiquem áreas que necessitam de intervenção e ajustem suas abordagens pedagógicas em tempo real. A visualização de indicadores é uma ferramenta poderosa que transforma dados brutos em informações compreensíveis e acionáveis, apoiando a criação de ambientes de aprendizagem mais adaptativos e responsivos.

Além disso, LA concentra-se na análise do desempenho, dos recursos de aprendizagem e das atividades dentro e fora dos ambientes virtuais de aprendizagem (AVA). Os resultados obtidos com essa análise permitem o aprimoramento do processo de ensino e aprendizagem, oferecendo insights valiosos que podem ser utilizados para melhorar os materiais didáticos, ajustar as metodologias de ensino e fornecer suporte personalizado aos estudantes (Dondorf; Nacken, 2017; Paik et al., 2024a). Com as informações coletadas aplicando LA, instituições e instrutores podem planejar intervenções necessárias de modo a ajudar os estudantes, redesenhar cursos, adaptar conteúdos e métodos de aprendizagem, entre outros ajustes (Conde; Hernandez-garcia, 2015). Dessa forma, LA não apenas facilita a compreensão do comportamento dos estudantes, mas também promove um ambiente educacional mais eficiente e eficaz, contribuindo de modo a enfrentar desafios educacionais, como a retenção de estudantes, a prevenção de risco acadêmico e a melhoria das abordagens de estudo, entre outros (Paik et al., 2024a; Queiroga et al., 2022).

O progresso realizado no ecossistema de ciência de dados, nos últimos anos, permitiu não apenas se falar sobre learning analytics e mineração de dados educacionais, mas também sobre Academic Analytics, Institutional Analytics, Teaching Analytics, Data-Driven Education, Data-Driven Decision-Making in Education, Big Data in Education e Educational Data Science (Romero; Ventura, 2020; Dondorf, 2022). No entanto, para esta dissertação, a LA será considerada como um termo genérico que engloba os esforços concentrados em dar valor aos dados educacionais produzidos a partir da interação dos estudantes em diferentes ambientes educacionais, inclusive aqueles mediados pela tecnologia, como ambientes tradicionais ou salas de aula presenciais.

## **2.2 Interoperabilidade de dados educacionais**

O conceito de interoperabilidade no contexto educacional refere-se à capacidade de diferentes sistemas e plataformas de aprendizagem se comunicarem e compartilharem informações de forma eficiente. Isso permite que os estudantes acessem e utilizem recursos de aprendizagem de diversas fontes, e que as instituições capturem dados de aprendizagem de forma mais abrangente.

A interoperabilidade refere-se à capacidade de diferentes sistemas e aplicativos de software trabalharem juntos de forma eficaz, garantindo que o compartilhamento

de conteúdo, a troca de dados e a utilização das informações ocorram de maneira harmoniosa e integrada. (Bakhouyi et al., 2017). Em termos simples, a interoperabilidade permite que diferentes sistemas compartilhem e troquem informações entre si, garantindo a compatibilidade do conteúdo utilizado no processo, independentemente de quem o tenha desenvolvido (Bishr, 1997; Sheth, 1999; Furniel; Mendonça; Silva, 2016). A interoperabilidade assegura que as informações possam ser compartilhadas e utilizadas em diversas plataformas e sistemas, possibilitando uma comunicação e colaboração eficazes (Cooper, 2014). Para Cooper (2013) podemos pensar na interoperabilidade como uma redução de energia desperdiçada na transformação ou integração manual de dados .

Para expandir o alcance de LA, especialmente ao integrar dados de fontes heterogêneas, é crucial que os dados sejam descritos de maneira consistente. Nesse contexto, a interoperabilidade torna-se um aspecto essencial, permitindo a troca coerente de informações entre diferentes sistemas e plataformas (Samuelsen; Chen; Wasson, 2021a). A interoperabilidade tornou-se uma questão central na comunidade de LA desde a proposta de Siemens et al., (2011). No entanto, como destaca Siemens et al., (2011), esse é um processo gradual que passa por diferentes estágios de maturidade. Ele começa com a conscientização, onde são gerados relatórios básicos e coletados registros, e avança até o estágio final, onde o setor é transformado por meio de recursos avançados de compartilhamento de dados. Somente ao atingir esse estágio final dos processos (veja a Fig. 1), conforme Siemens et al., (2011), veremos uma transformação significativa no setor, caracterizada por uma alta interoperabilidade.

As plataformas de e-learning atuais, embora eficazes em certos aspectos, são limitadas na coleta de dados granulares e no acompanhamento de atividades de aprendizagem mais avançadas (Väkevä, 2023). A Caliper Analytics e XAPI permitem rastrear atividades de aprendizagem dentro e fora dos AVA's, fornecendo uma visão mais completa do comportamento do aluno e, assim, contribuir para uma análise mais profunda da aprendizagem. As plataformas de e-learning atuais, embora eficientes em certos aspectos, são limitadas na coleta de dados granulares e no acompanhamento de atividades de aprendizagem mais complexas, especialmente aquelas que ocorrem fora do ambiente tradicional dos AVAs (Väkevä, 2023).

A integração de dados provenientes de fontes heterogêneas possibilita uma análise abrangente, que pode levar a decisões baseadas em dados na área da educação (Cooper, 2014). No entanto, apesar da coleta de dados de múltiplas fontes ser viável, ainda existem desafios significativos relacionados à transmissão, processamento, armazenamento e rastreabilidade desses dados. Para superar esses desafios, é essencial desenvolver sistemas projetados especificamente para o contexto em questão.

A obtenção da interoperabilidade pode ser alcançada por meio de acordos e diretrizes comuns aplicáveis ao conteúdo, sistemas e aplicativos. Padronizar os processos

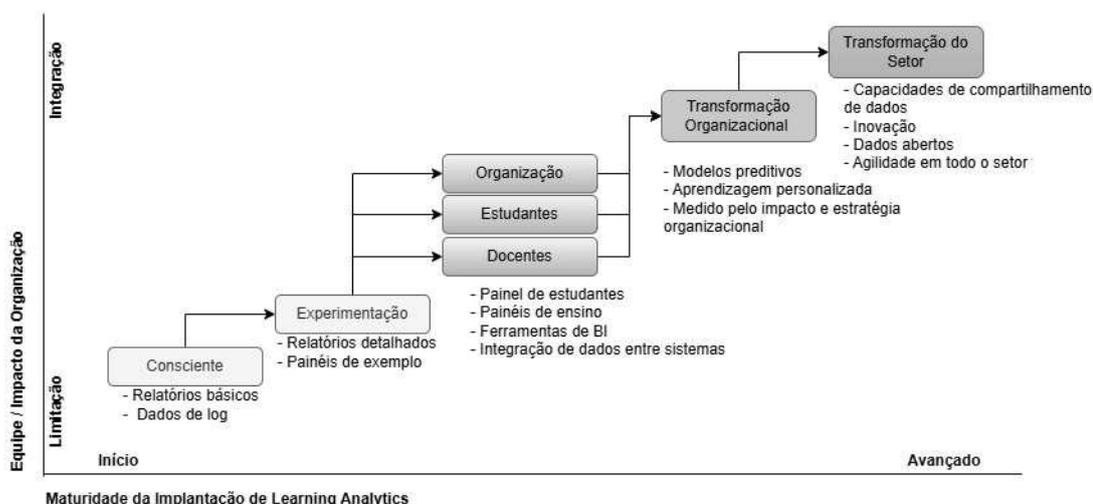


Figura 1 – Modelo sofisticado de LA. Fonte: (Siemens; Dawson; Lynch, 2013)

de troca de dados por meio da interoperabilidade promove o uso de linguagens e definições compartilhadas, facilitando o intercâmbio e a integração de dados entre diferentes plataformas (Cooper, 2014). As especificações e os padrões são considerados elementos centrais no processo de padronização (Bakhouyi et al., 2017).

Na interoperabilidade, se torna necessário adotar um vocabulário comum para garantir que todas as atividades de aprendizagem em diferentes comunidades sejam descritas de forma precisa. Ao formalizar esse vocabulário, estabelece-se um conjunto de atributos e regras sobre os dados, determinando, por exemplo, como devem ser armazenados, recuperados e acessados por outros componentes, sistemas ou atividades (Llc, 2019).

A integração de dados está profundamente ligada à interoperabilidade, que abrange aspectos semânticos, técnicos, legais e organizacionais. No que diz respeito à interoperabilidade técnica e semântica, duas especificações de dados amplamente reconhecidas como padrões de fato na indústria educacional e na LA são o Experience API (xAPI) e o Caliper Analytics (Samuelson; Chen; Wasson, 2021a), que iremos explorar ao longo das próximas seções.

A interoperabilidade entre diferentes sistemas e plataformas de e-learning é fundamental para garantir a fluidez do processo de aprendizagem e a reutilização de conteúdo. Padrões e especificações, como SCORM, AICC, Common Cartridge, xAPI e Caliper Analytics, desempenham um papel crucial nesse contexto.

## 2.3 Especificações e Padrões de E-Learning

No contexto de e-learning, especificações e padrões desempenham um papel fundamental na garantia da interoperabilidade entre diferentes sistemas e plataformas de aprendizagem, além de facilitar a coleta e análise de dados de aprendizagem.

Eles definem um conjunto de regras, formatos e protocolos que permitem que diferentes softwares e conteúdos educacionais se comuniquem e trabalhem em conjunto de forma eficaz (Griffiths; Hoel, 2016; Baker; Inventado, 2014).

Essas diretrizes ajudam a estabelecer como rastrear as atividades e a experiência dos estudantes, integrar multimídia no conteúdo, melhorar a interoperabilidade entre diferentes sistemas de aprendizagem e seguir critérios para criar cursos acessíveis a todos os estudantes (Bayer, 2024). A seguir, iremos explorar alguns destes conceitos e especificações.

### **2.3.1 E-learning**

E-learning é um conceito amplo que engloba a utilização da Tecnologia da Informação e Comunicação (TIC) no processo de aprendizagem. É um guarda-chuva que abrange diversas abordagens, metodologias e ferramentas digitais para a criação e entrega de conteúdo educacional, interação entre estudantes e instrutores e avaliação do progresso do estudante (Väkevä, 2023; Miller et al., 2021).

Ao longo dos anos, o processo de aprendizagem e treinamento evoluiu de cursos e conteúdos distribuídos em mídias físicas, como CDs e disquetes, para o treinamento online baseado na Web. Com essa mudança, surgiram desafios relacionados ao planejamento de conteúdos e à interoperabilidade entre sistemas (Miller et al., 2021). A introdução de novos sistemas no processo de aprendizagem frequentemente exigia que as organizações adaptassem ou reconstruíssem o conteúdo para cada contexto, resultando em processos ineficientes e dispendiosos. Com o avanço constante da tecnologia, as instituições educacionais passaram a dedicar mais atenção à busca por maneiras eficazes de promover a aprendizagem (Chicu, 2018).

Nesse contexto, o e-learning desempenha um papel fundamental na transição do aprendizado presencial tradicional para o uso de ferramentas digitais. Embora o e-learning apresente diferenças em relação à aprendizagem tradicional em sala de aula, ambos compartilham o objetivo comum de promover a aquisição de conhecimento e o desenvolvimento de habilidades (Väkevä, 2023).

Segundo Fallon e Brown (apud Padiadpu, 2008), os padrões para e-learning são divididos em duas categorias principais. Primeiro, os Padrões de Interoperabilidade, que definem como o software de curso se comunica com sistemas administrativos, como um AVA, para a troca de dados sobre os estudantes e seu progresso. Em segundo lugar, os Padrões de Empacotamento de Conteúdo, que especificam como os objetos de aprendizagem e conjuntos de objetos, incluindo cursos completos, devem ser organizados para facilitar a importação em sistemas administrativos, o transporte entre sistemas e a classificação em repositórios de conteúdo, permitindo uma pesquisa, acesso e reutilização mais eficientes.

### 2.3.2 Especificações e Padrões

No contexto de e-learning, as especificações e padrões são geralmente desenvolvidos para garantir a interoperabilidade e a reutilização em sistemas, bem como para a gestão de conteúdo e metadados (Padiadpu, 2008).

Independentemente do modelo de aprendizagem adotado, do conteúdo apresentado ou dos meios utilizados pelos estudantes para acessar um material, é certo que, em algum momento, será necessário registrar dados para análise posterior. Esses registros podem servir tanto para confirmar a conclusão de um curso quanto para armazenar a pontuação obtida pelo estudante durante sua formação. Para realizar esse registro, precisamos de um modelo de dados, que define a estrutura dos dados, mas não necessariamente o que eles significam. Esses dados podem representar pontuação, conclusão de curso ou até interações em um fórum. No entanto, o modelo de dados por si só não atribui significado a essas informações. Além disso, é comum a necessidade de transferir esses dados para outros sistemas. Para isso, utiliza-se um protocolo de transferência, que descreve como os dados devem ser comunicados entre diferentes sistemas. Provavelmente, essa comunicação ocorre de um AVA para uma plataforma de LA, ou, em alguns casos, no sentido inverso. Em outros cenários, pode ser a transferência de dados entre dois AVA diferentes. Para que essa comunicação seja eficaz, é essencial que se tenha especificações e protocolos bem definidos. Essas especificações ou padrões garantem que os dados sejam compreendidos de maneira consistente entre os sistemas, permitindo uma troca de informações eficiente e sem perda de significado ou contexto.

Algumas especificações e padrões têm vindo a ser desenvolvidos por consórcios internacionais envolvidos com a educação, como a ADL (Advanced Distributed Learning), o IMS GLC (Institute of Management System – Global Learning Consortium, agora conhecido como 1EdTech), AICC (Aviation Industry Computer-Based Training Committee) e o IEEE (Institute of Electrical and Electronics Engineers) (Tarouco et al., 2014). Entre essas especificações, incluem-se algumas voltadas especificamente para o uso educacional, como o CMI (Computer Managed Instruction), SCORM (Sharable Content Object Reference Model), CC (Common Cartridge), LTI (Learning Tools Interoperability), xAPI (Experience API), cmi5 e Caliper Analytics (Furniel; Mendonça; Silva, 2016; Quiroz; Muñoz gonzález, 2023).

Dentre essas especificações mencionadas, o CMI é a mais antiga, tendo existido entre 1984 e 2014. Ele utilizava o protocolo HTTP, e embora alguns AVA e ferramentas de criação ainda o utilizem, o CMI não evoluiu muito ao longo dos anos, sendo descontinuado e seus conceitos incorporados ao SCORM (Sharable Content Object Reference Model) e posteriormente no xAPI.

Em 2000, a Advanced Distributed Learning Initiative (ADL) desenvolveu um dos primeiros padrões de Empacotamento de Conteúdo, o SCORM, para abordar ques-

tões de interoperabilidade, reutilização e durabilidade no aprendizado eletrônico (Miller et al., 2021). O SCORM, que possui quatro versões, das quais duas estão em uso atualmente: 1.2 e 2004 (Qgate, 2021), porém, SCORM não é flexível nem extensível o suficiente para dar suporte a um ecossistema de aprendizagem moderno e distribuído (Miller et al., 2021; Qgate, 2021). Embora o SCORM tenha sido um padrão importante no e-learning, apresenta diversas limitações que impulsionaram o desenvolvimento de novas especificações, tais como o xAPI e a especificação cmi5, foram introduzidos para trazer mais flexibilidade e extensibilidade, permitindo um rastreamento robusto de dados e integração com tecnologias emergentes, atendendo melhor ao ambiente de aprendizagem distribuída e em constante evolução (Miller et al., 2021).

Muitas vezes, comenta-se equivocadamente que a xAPI é a “próxima geração do SCORM”, mas a xAPI fornece um formato padronizado para comunicar dados de experiências de aprendizagem, enquanto o SCORM se concentra principalmente na interoperabilidade de conteúdo dentro de um AVA e no empacotamento de cursos para esses sistemas (ADL, 2021). Em 2010, a AICC iniciou o desenvolvimento do projeto cmi5, que tinha o objetivo de substituir as especificações CMI e SCORM. Em 2012, o mecanismo de comunicação do cmi5, baseado em SOAP, estava próximo de ser finalizado, mas, nesse mesmo período, a ADL concluiu a pesquisa sobre a xAPI. Diante da grande sobreposição entre as duas especificações, a ADL e a AICC decidiram colaborar para desenvolver um “perfil xAPI” que atendesse aos casos de uso específicos do cmi5. Assim, em 2012, o projeto cmi5 foi “reiniciado” e sua arquitetura SOAP foi substituída pela xAPI, posicionando o cmi5 como a especificação que realmente busca substituir o SCORM e o CMI (Quiroz; Muñoz gonzález, 2023).

Paralelamente ao apresentado anteriormente, o 1EdTech Consortium (anteriormente conhecido como IMS Global Learning Consortium) também, ao longo dos anos, foi desenvolvendo diversas especificações para o setor de e-learning que se tornaram, de fato, padrões globais. A 1EdTech desenvolveu e apoiou cerca de 50 especificações/padrões, com foco em metadados, empacotamento de conteúdo, portfólio eletrônico, informações do aprendiz, lista de recursos, entre outros (Bakhouyi et al., 2017; 1EdTech, 2023).

As especificações 1EdTech visam garantir a interoperabilidade dos sistemas que se apoiam em e-learning, educação e treinamentos aprimorados pela tecnologia (Bakhouyi et al., 2017). Uma dessas especificações é a LTI, que permite a integração entre plataformas de aprendizagem e ferramentas externas de aprendizagem (Perez-colado et al., 2022; Conde, 2018). O LTI integra conteúdo, aplicativos digitais, ferramentas e aplicações educacionais em sistemas de e-learning (Bakhouyi et al., 2017) mas não tem enfoque em rastreabilidade da aprendizagem, sendo a principal função do LTI facilitar a comunicação entre o AVA e a ferramenta externa, permitindo que os usuários acessem e utilizem recursos externos sem a necessidade de logins separa-

dos ou processos de configuração complexos (Perez-colado et al., 2022). Já a CC, que também foi desenvolvida pela 1EdTech, é uma especificação que busca padronizar a estrutura e o empacotamento de conteúdo de aprendizagem digital, tornando-o portátil e interoperável entre diferentes plataformas de aprendizagem (Behringer; Group, 2013; Våkevå, 2023). O CC visa solucionar alguns dos problemas encontrados em padrões anteriores como o SCORM, especialmente no que diz respeito à organização e reutilização de conteúdo. Ele define um formato de arquivo padronizado que permite que o conteúdo seja facilmente importado e exportado entre diferentes AVAs. Porém, o CC também não almeja a rastreabilidade da aprendizagem. Assim, para suprir essa necessidade, em outubro de 2015, a 1EdTech lançou a especificação Caliper Analytics, que permite às instituições coletar dados de aprendizagem a partir de recursos digitais, apresentando muitas semelhanças com o xAPI, desenvolvido pela ADL (Griffiths; Hoel, 2016).

Na Figura 2 apresenta algumas organizações e as especificações que elas criaram no âmbito do e-learning.

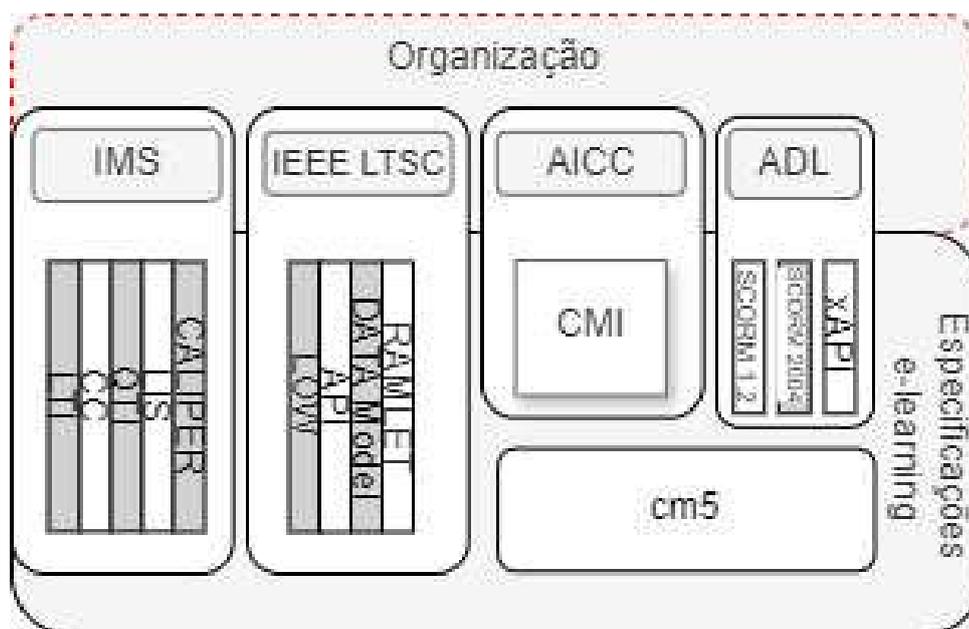


Figura 2 – Padrões e Organizações de Sistemas de E-Learning Fonte: Bakhouyi et al.

Essas especificações possibilitam tanto a troca de conteúdo quanto a integração de dados de atividades de aprendizagem provenientes de diversas ferramentas e fontes, maximizando a interoperabilidade entre plataformas e serviços que produzem, coletam, armazenam e analisam dados de aprendizagem. Dessa forma, dados individuais das atividades representam a interação de um estudante com um objeto de aprendizagem em um ambiente educacional (Samuelson; Chen; Wasson, 2021a; Miller et al., 2021).

Embora existam diversas especificações, elas não são mutuamente exclusivas e podem ser combinadas para criar soluções mais completas. Por exemplo, o cmi5

utiliza a xAPI como base, adicionando uma estrutura compatível com o SCORM, enquanto o Common Cartridge pode ser usado em conjunto com o LTI para integrar atividades externas a um pacote de conteúdo (Conde, 2018).

A seguir, iremos explorar em mais detalhes da xAPI e Caliper Analytics, que são especificações voltadas para a coleta de dados de aprendizagem, com o objetivo de proporcionar uma análise mais aprofundada do aprendizado em ambientes digitais. Embora compartilhem objetivos semelhantes, eles se diferenciam em sua abordagem de desenvolvimento, governança e foco.

### **2.3.3 Experience API**

A Experience Application Programming Interface (xAPI), anteriormente conhecida como Tin Can API, é uma especificação técnica desenvolvida pela ADL que facilita o envio, armazenamento e recuperação de uma ampla gama de atividades de aprendizagem e desempenho, permitindo o compartilhamento dessas informações entre plataformas (Väkevä, 2023; Bakhouyi et al., 2017). A xAPI permite o rastreamento de qualquer atividade de aprendizagem, seja online ou offline, incluindo interações em jogos, simulações, plataformas de mídia social e atividades do mundo real. Essa capacidade de rastrear experiências além do AVA tradicional diferencia a xAPI de padrões anteriores, como o SCORM (Väkevä, 2023).

Seu principal objetivo é criar uma tecnologia educacional que possibilite a comunicação entre conteúdos e sistemas de aprendizagem, permitindo também o registro e rastreamento de uma ampla gama de experiências de aprendizagem. A xAPI define um meio interoperável de documentar e comunicar informações sobre essas experiências, especificando uma estrutura para descrevê-las e definindo como essas descrições podem ser trocadas eletronicamente (IEEE, 2023; Miller et al., 2021; Väkevä, 2023).

A xAPI surgiu como um dos primeiros serviços da ADL direcionados ao compartilhamento de dados sobre experiências de aprendizagem (ADL; Johnson, 2024; Furniel; Mendonça; Silva, 2016; Pedro et al., 2018). Em conjunto com a LA, a xAPI tem o propósito de transformar significativamente a forma como a educação e o treinamento são estruturados, gerenciados e avaliados. Ao promover a interoperabilidade entre diferentes sistemas e plataformas de aprendizagem, a xAPI possibilita que dados educacionais sejam compartilhados e agrupados de maneira consistente, contribuindo para a formação de ecossistemas de aprendizagem mais integrados e flexíveis (Väkevä, 2023; ADL; Johnson, 2024; Furniel; Mendonça; Silva, 2016; Pedro et al., 2018).

Sua arquitetura é bastante maleável e expansível, permitindo aos usuários criar seus próprios vocabulários, extensões e perfis a fim de rastrear atividades específicas. Essa adaptabilidade faz com que a xAPI possa atender a diversos contextos de aprendizagem e evoluir em consonância com novas tecnologias (Väkevä, 2023;

Perez-colado et al., 2022).

No xAPI, as experiências de aprendizagem são registradas em uma estrutura flexível de “ator-verbo-objeto”, que possibilita rastrear diversos tipos de interações e atividades educativas, superando as métricas tradicionais de conclusão de cursos e notas em avaliações. Essa estrutura segue o formato de uma afirmação do tipo “Eu fiz isso”, coletada por provedores de registro de aprendizagem (LRP) ou provedores de atividade e enviada, via HTTP ou HTTPS, em ordem cronológica, para um repositório especializado denominado Learning Record Store (LRS) (Väkevä, 2023). As declarações de atividade adotam uma sintaxe legível por humanos. Nesse formato, a interação do estudante com o objeto de aprendizagem é registrada por meio de uma gramática composta por substantivo (ator), verbo e objeto, conferindo clareza e objetividade ao monitoramento das atividades (Behringer; Group, 2013).

Os principais elementos dessa afirmação são:

**Ator:** Representa quem realizou a ação e se refere ao “Eu” em “Eu fiz isso”.

**Verbo:** Descreve a ação realizada pelo ator e representa o “fiz” em “Eu fiz isso”.

**Objeto:** Define o elemento sobre o qual a ação foi realizada e forma a parte “isso” em “Eu fiz isso”.



Figura 3 – xAPI Statement - Declaração básica. Fonte: Autor

Além dos elementos principais, uma afirmação xAPI pode incluir informações adicionais por meio de elementos opcionais, como:

**Resultado:** Descreve o resultado da atividade, como pontuação, sucesso ou conclusão. Contexto: Fornece informações contextuais sobre a atividade, como o local, a hora ou a plataforma em que ocorreu.

**Extensões:** Permitem adicionar dados personalizados à afirmação, expandindo sua capacidade de capturar informações específicas.

A estrutura responsável pelo armazenamento dos componentes da xAPI é denominada *statements*, e o processo de guardar e recuperar esses dados ocorre em um LRS. Esse LRS possibilita a comunicação com rastreabilidade, pois permite armaze-

nar, analisar, agregar e consumir as informações geradas e compartilhadas (Costa, 2021).

A xAPI estabelece requisitos tanto para a forma como os dados são estruturados quanto para as funcionalidades do LRS. As declarações (*statements*) possuem três propriedades principais (Figura 3): ator, verbo e objeto, que descrevem uma experiência de aprendizagem, sendo todas construídas em formato JSON (*JavaScript Object Notation*) (Figura 4) (Llc, 2019; Samuelsen; Chen; Wasson, 2021a).

```

{
  "id": "fd41c918-b88b-4b20-a0a5-a4c32391aaa0",
  "timestamp": "2015-11-18T12:17:00+00:00",
  "actor": {
    "objectType": "Agent",
    "name": "Project Tin Can API",
    "mbox": "mailto:user@example.com"
  },
  "verb": {
    "id": "http://example.com/xapi/verbs#sent-a-statement",
    "display": {
      "en-US": "sent"
    }
  },
  "object": {
    "id": "http://example.com/xapi/activity/simplestatement",
    "definition": {
      "name": {
        "en-US": "simple statement"
      },
      "description": {
        "en-US": "A simple Experience API statement. Note that the LRS
          does not need to have any prior information about the Actor (learner), the
          verb, or the Activity/object."
      }
    }
  }
}

```

Figura 4 – Representação json de xAPI statement. Fonte: Autor

Como dados textuais estruturados em formato JSON com vocabulário predefinido, as declarações xAPI não são apenas interoperáveis, mas também podem ser rapidamente processadas por computadores. Além das propriedades predefinidas para incluir informações adicionais, como dados de contexto e de avaliação (resultados), foi projetada para ser extensível a fim de atender a necessidades imprevistas de coleta de dados (Kevan; Ryan, 2016).

As atividades realizadas por um estudante podem ser rastreadas por essas declarações, possibilitando a verificação do comportamento do estudante durante o uso de recursos educacionais, incluindo diferentes tipos de recursos, sendo estes conceitos básicos ilustrados na Figura 5

A Figura 6 ilustra o rastreamento de experiências de aprendizagem que um estudante vivencia, podendo ocorrer em um curso online, no ambiente de trabalho ou até em uma atividade recreativa – ou seja, pode ser qualquer tipo de experiência. Essa experiência é registrada em nome do estudante por um LRP confiável, que também

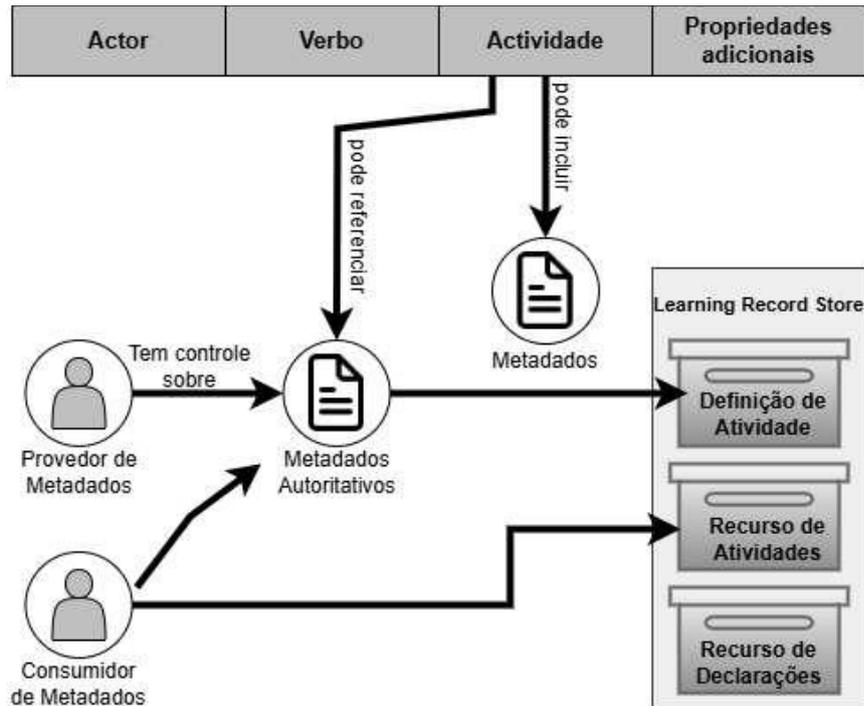


Figura 5 – Estrutura de xAPI. Fonte: (ADL; Johnson, 2024)

pode ser responsável por assegurar a relação confiável entre a experiência e o estudante. O LRP pode ser um AVA como Moodle ou Canvas, mas também pode ser qualquer site onde ocorrem atividades de aprendizagem. O LRP envia esses registros de Aprendizagem a um ou mais LRS. O LRS armazena os registros de Aprendizagem e os disponibiliza a qualquer cliente autorizado. Um consumidor de registros de aprendizagem (LRC) é um tipo de cliente que acessa os registros de aprendizagem e faz uso deles (ADL; Johnson, 2024; IEEE, 2023).

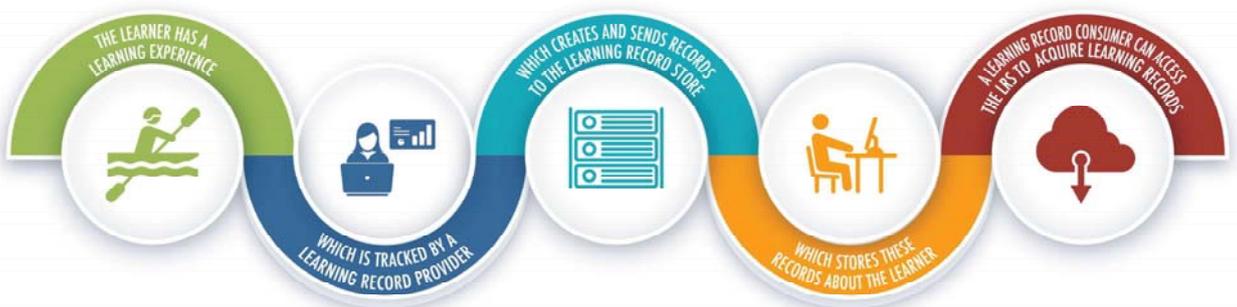


Figura 6 – Rastreamento de Experiências com xAPI. Fonte: (ADL; Johnson, 2024)

A xAPI representa uma mudança significativa na forma como os dados de aprendizagem são rastreados, coletados e analisados, oferecendo um conjunto de recursos mais amplo e flexível em comparação com os padrões tradicionais de e-learning. Sua capacidade de coletar dados detalhados sobre uma ampla variedade de experiências de aprendizagem, tanto online quanto offline, abre novas possibilidades para análise

de aprendizagem e personalização do ensino (Väkevä, 2023).

Ao utilizar a xAPI, é possível realizar análises de aprendizagem mais abrangentes e personalizadas, permitindo que as instituições educacionais compreendam melhor o comportamento dos alunos por meio dos dados granulares coletados. Esses dados oferecem *insights* sobre como os alunos interagem com o conteúdo, quais recursos utilizam e quais dificuldades enfrentam. Com base nessas informações, as instituições podem personalizar a experiência de aprendizagem, adaptando o conteúdo e as atividades às necessidades individuais dos alunos, oferecendo percursos de aprendizagem mais personalizados.

Além disso, a xAPI possibilita uma avaliação mais eficaz de programas de aprendizagem. Ela permite rastrear o progresso dos alunos em relação aos objetivos educacionais, fornecendo dados valiosos para avaliar a eficácia de cursos e programas.

Essa capacidade de rastrear atividades fora dos AVAs tradicionais, como o uso de aplicativos móveis, simulações e aprendizagem informal, oferece uma visão mais completa do comportamento do aluno e enriquece a análise de aprendizagem (Väkevä, 2023).

Em 2018, membros do Instituto de Engenheiros Elétricos e Eletrônicos (IEEE) formaram um grupo de trabalho para padronizar a especificação xAPI. O processo de padronização do IEEE inclui rigor adicional e envolveu diversas partes interessadas de modo a garantir um padrão tecnológico de aplicação geral mais robusto [ADL xRAP Final Project Report]. Em 2023, a especificação xAPI finalmente se tornou um padrão oficial da Institute of Electrical and Electronics Engineers - Learning Technology Standards Committee (IEEE-LTSC) (IEEE, 2023), sendo o primeiro padrão de código aberto na história do IEEE [xAPI Published as IEEE Standard]. Na Figura 7 podemos ver a evolução da xAPI até se tornar o padrão que é hoje em dia, fora 10 anos de trabalho feitos da ADL e comunidade opensource.

O xAPI, ao longo dos anos, evoluiu até tornar-se um padrão IEEE (IEEE 192..1.1.1) (Behringer; Group, 2013; IEEE, 2023), consolidando-se atualmente como o principal padrão aberto para a coleta de dados de aprendizagem. A versão mais recente, xAPI 2.0.0, utiliza o formato JSON, proporcionando grande flexibilidade. Por exemplo, é possível definir novos verbos e tipos de atividade para as declarações, idealmente compartilhando esses vocabulários em perfis junto a comunidades de práticas relevantes (Samuelsen; Chen; Wasson, 2021a).

O xAPI representa um avanço significativo na padronização e interoperabilidade em sistemas de e-learning, abrindo caminho para experiências de aprendizagem mais personalizadas, eficazes e orientadas por dados.

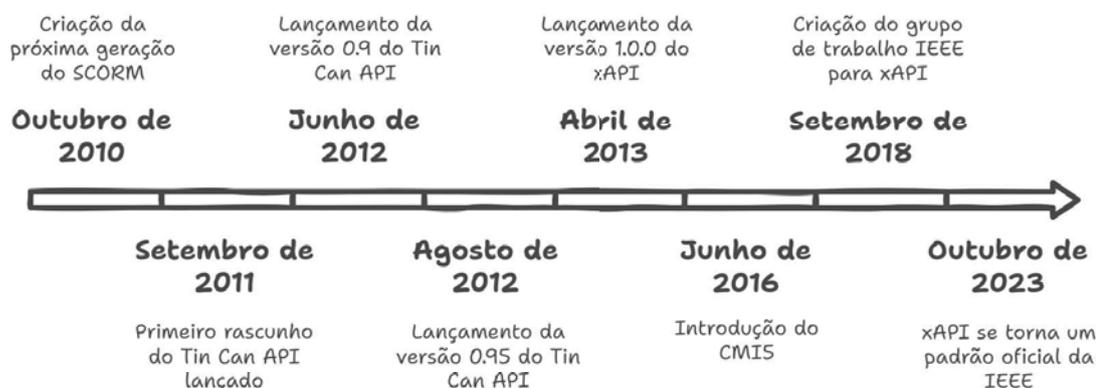


Figura 7 – Evolução do xAPI: Do Conceito ao Padrão. Fonte: Adaptado de (Bakhoui et al., 2017)

### 2.3.4 Caliper Analytics

Caliper Analytics é uma especificação desenvolvida pela 1EdTech, uma organização sem fins lucrativos fundada em 1999, constituída por uma comunidade internacional de universidades e empresas que promove a padronização e a definição de padrões globais em e-learning e TIC educacionais. Ela foi lançada em outubro de 2015 e estabelece uma abordagem estruturada para descrever, coletar e compartilhar dados de atividades de aprendizagem, buscando criar um vocabulário comum para representar as interações de aprendizagem, além de fomentar a interoperabilidade de dados, o compartilhamento de informações (1EdTech, 2023; Krumova, 2023). O Caliper Analytics é uma especificação técnica que descreve um conjunto estruturado de vocabulário para ajudar as instituições a coletar dados de aprendizagem e uso de recursos digitais e ferramentas de aprendizagem.

O Caliper Analytics tem como base o conceito de evento (*event*), que descreve um contexto de uso composto por um ator (*actor*) — estudante, professor ou qualquer outro agente educacional —, uma ação (*action* ou *intervention*), que indica o que esse ator está fazendo, e um objeto (*object*), que corresponde ao recurso utilizado na ação. Um exemplo seria um estudante de ensino médio encarregado de ler um livro para a disciplina de Inglês: Bruce (ator) leu (ação) *The Great Gatsby* (objeto).

Assim como o xAPI, o Caliper Analytics objetiva a coleta de dados de aprendizagem, mas se destaca por enfatizar a criação de métricas quantitativas padronizadas voltadas para análises. Ele costuma ser adotado por instituições de grande porte e fornecedores de AVA que buscam uma abordagem mais estruturada de coleta e análise de dados (Väkevã, 2023).

A figura 8 define um fluxo de dados do Caliper Analytics, onde as atividades de aprendizagem são rastreadas por sensores, formatadas em eventos padronizados com vocabulários e estruturas definidas, transmitidas via Sensor para um Sistema de armazenamento.

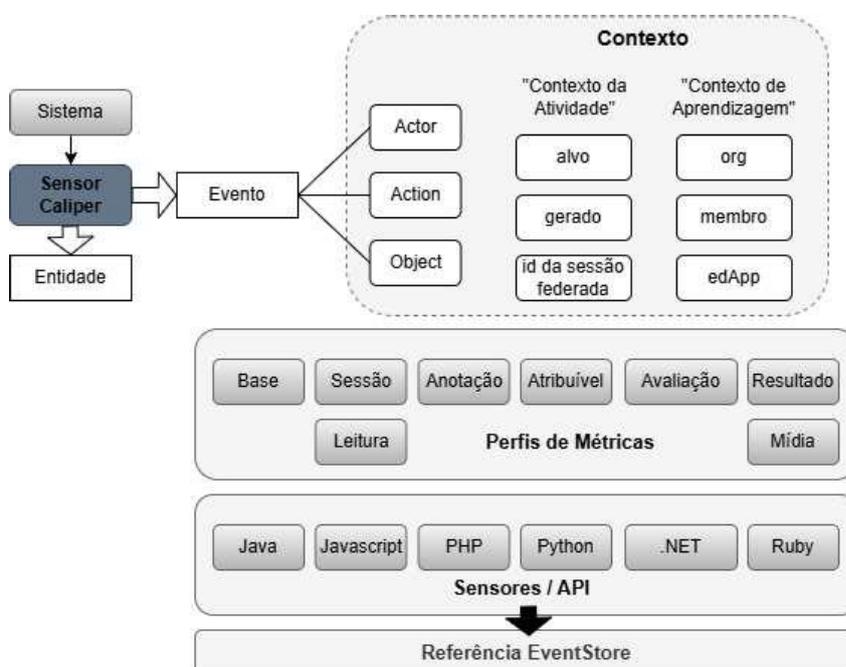


Figura 8 – Estrutura do Caliper Analytics. Fonte: (1EdTech, 2023)

O Caliper Analytics define uma série de perfis de métricas (*metric profiles*), cada um modelando uma atividade de aprendizagem ou uma atividade de suporte que ajuda a facilitar o processo de aprendizagem, conforme mostrado na Tabela 1. Cada perfil fornece um vocabulário compartilhado que os desenvolvedores usam para descrever as interações do usuário de maneira consistente. Cada perfil é composto por um ou mais tipos de Eventos (por exemplo, *AssessmentEvent*, *NavigationEvent*), e cada tipo de evento está associado a um conjunto de ações realizadas por alunos, instrutores e demais participantes. Uma ação em Caliper Analytics é definida estritamente por um perfil de métricas, em contraste com o que ocorre no xAPI (Llc, 2019). Uma *action* no Caliper Analytics é estritamente definida por um perfil métrico, diferentemente do xAPI (Ryunosuke et al., 2016). O Caliper Analytics inclui diversos perfis métricos, modelando atividades de aprendizagem ou atividades de apoio que facilitam a aprendizagem. Os perfis incluem, mas não estão limitados a: Anotação, Avaliação, Atribuível, *Feedback*, *Grading*, Mídia, Leitura, Gerenciamento de Recursos, Busca, Pesquisa, Lançamento de Ferramentas e Uso de Ferramentas.

É importante notar que, diferentemente do xAPI, que depende de um LRS para armazenar dados, o Caliper Analytics não especifica explicitamente um sistema de armazenamento centralizado. As capacidades de armazenamento ou recuperação de dados não são definidas como parte da especificação do Caliper Analytics ; no

entanto, os atributos de dados definidos sobre *Events* e *Entities* são claros, e as regras em torno da especificação promovem a interoperabilidade. Caliper não inclui formalmente um armazenamento de eventos/LRS aberto e baseado em padrões em seu escopo inicial. No entanto, uma implementação de *EventStore* de referência foi fornecida para fins de desenvolvimento, teste e demonstração.

O processo de coleta de dados com o Caliper Analytics inicia por meio de sensores, que são trechos de código implementados dentro de aplicativos de aprendizagem. Esses sensores monitoram as atividades dos usuários, como o acesso a recursos, a realização de avaliações e a participação em fóruns. As atividades dos usuários são registradas como eventos, que descrevem a interação entre um ator (como um aluno ou professor) e um objeto (como um recurso digital ou uma ferramenta). Um evento Caliper Analytics é composto por um triplo (Ator, Ação, Objeto). As ações representam as atividades realizadas para atingir um propósito e são expressas pelo atributo “*action*” em um evento. Cada ação é definida por um perfil métrico e faz parte de um vocabulário controlado. Por exemplo, as ações podem ser “*Started*”(iniciado), “*Completed*” (concluído), “*Commented*” (comentado), “*Searched*” (pesquisado) ou “*Viewed*” (visualizado). Cada atividade de aprendizagem tem um ou mais perfis métricos associados; perfil métrico modela uma atividade de aprendizagem ou qualquer atividade de suporte, fornecendo um vocabulário compartilhado para descrever as interações do usuário de maneira consistente. Os perfis métricos agrupam eventos, ações e entidades para modelar atividades específicas, como avaliações, leitura de textos, uso de ferramentas e participação em fóruns.

O Caliper Analytics possui 14 perfis disponíveis ( *Metric profiles*), que abrangem experiências como avaliações, fóruns e uso de ferramentas. Na tabela 1 são apresentados alguns desses perfis. De forma semelhante ao xAPI, o Caliper também oferece suporte a descrições de dados legíveis por máquina por meio de JSON-LD (Samuelson; Chen; Wasson, 2021a).

Além dos perfis determinados no Caliper Analytics, é possível adicionar outras informações além da experiência de aprendizagem às “extensões” do “ator” e do “objeto”, para que se possa incluir com flexibilidade informações extras que não estejam definidas no perfil métrico (Ryunosuke et al., 2016).

Os perfis métricos estabelecem uma linguagem unificada para caracterizar a atividade dos estudantes, facilitando a transferência de dados. É essencial ressaltar que, apesar de os perfis métricos estabelecerem uma padronização, eles não determinam como se deve criar um produto. Diferentes produtos podem ser desenvolvidos utilizando os mesmos códigos definidos pelo padrão.

Devido à sua capacidade de padronizar a coleta de dados de aprendizagem, Caliper Analytics permite uma compreensão mais profunda e abrangente do processo educacional. Ao estabelecer um vocabulário comum e estruturas consistentes, o Cali-

Tabela 1 – Perfis de Métricas do Caliper. Fonte: (1EdTech, 2023).

<b>Perfil</b>	<b>Descrição</b>
Annotation	Modela atividades relacionadas à anotação de um Recurso Digital.
Assessment	Modela atividades relacionadas à avaliação, incluindo interações com itens individuais de avaliação.
Assignable	Modela atividades associadas à atribuição de conteúdo digital a um aluno para conclusão de acordo com critérios específicos.
Forum	Modela interações entre alunos e outros participantes em comunidades de fóruns online.
Grading	Modela atividades de avaliação realizadas por um Agente, geralmente uma Pessoa ou um Software/Aplicação.
Media	Modela interações entre alunos e conteúdos ricos, como áudio, imagens e vídeos.
Reading	Modela atividades associadas à navegação e visualização de conteúdo textual digital.
Session	Modela a criação e a subsequente finalização de uma sessão de usuário estabelecida por uma Pessoa interagindo com um Aplicativo de Software.
Tool Use	Modela uma interação intencional entre um usuário e uma ferramenta.

per Analytics facilita a interoperabilidade e a análise de dados de diversas ferramentas e plataformas de aprendizagem. O vocabulário comum e estruturas de dados padronizadas (como ações, contextos, entidades e perfis) asseguram que diferentes sistemas interpretem os dados de maneira uniforme. Essa padronização facilita a troca de dados entre plataformas e sistemas de análise, quebrando silos de informação e promovendo a integração de dados. Assim, ao se coletar e analisar dados de diversas ferramentas e plataformas, o Caliper Analytics oferece uma visão mais abrangente da atividade do aluno. Isso permite uma compreensão mais completa do processo de aprendizagem, que vai além de dados isolados de cada ferramenta. Caliper Analytics se concentra na criação de métricas quantitativas padronizadas para análise. Diferentemente de outras abordagens, como o xAPI, que se concentra em capturar a experiência do aluno de forma mais geral, o Caliper Analytics visa fornecer dados

mensuráveis e comparáveis.

A versão mais recente do Caliper Analytics é a 1.2, lançada em 27 de março de 2020 e ao que parece tem vindo a ganhar uma adoção entre fornecedores de ferramentas de aprendizagem, incluindo Blackboard e Canvas. A adoção é verificada por meio de um processo de certificação de conformidade conduzido pelo 1EdTech, garantindo a interoperabilidade entre sistemas. O Caliper Analytics foi desenvolvido para aproveitar as especificações 1EdTech existentes, como LTI (Learning Tool Interoperability), LIS (Learning Information Services) e QTI (Question Test Interoperability). Isso facilita a integração com outras ferramentas e sistemas educacionais. A adoção do Caliper Analytics é verificada por meio de um processo de certificação de conformidade conduzido pelo 1EdTech, garantindo a interoperabilidade entre os sistemas dos membros. A certificação de conformidade com o Caliper Analytics é um processo que pode gerar custos para as instituições, especialmente para as que não são membros do 1EdTech.

Caliper Analytics evoluiu de uma necessidade de padronização para se tornar uma ferramenta crucial para instituições educacionais que desejam coletar, analisar e usar dados para melhorar a qualidade do ensino e da aprendizagem. A especificação Caliper Analytics parece estar ganhando adoção entre os fornecedores de ferramentas de aprendizagem, incluindo Blackboard, Canvas e outros. Como outros padrões 1EdTech são predominantes na educação K-12, espera-se que essa adoção continue a crescer (Llc, 2019).

### **2.3.5 Diferença entre Xapi e Caliper Analytics**

O xAPI e o Caliper Analytics compartilham o objetivo de rastrear e analisar dados de aprendizagem para melhorar a experiência educacional, mas suas abordagens de desenvolvimento e implementação diferem significativamente, influenciando a flexibilidade, padronização e o esforço técnico exigido para sua adoção.

O xAPI, desenvolvido pela ADL sob licença Apache 2.0, é um padrão aberto que promove colaboração e inovação. Ele permite o registro de uma ampla gama de atividades de aprendizagem, tanto online quanto offline, em diversos contextos e plataformas. A estrutura baseada no modelo “ator-verbo-objeto” facilita a captura de dados de maneira granular e contextualizada. Sua flexibilidade permite às instituições criar perfis personalizados e adaptar o padrão a necessidades específicas. Isso inclui desde o rastreamento de interações em jogos sérios e simulações até atividades realizadas fora de ambientes digitais tradicionais (Väkevã, 2023).

No entanto, a flexibilidade do xAPI apresenta desafios. A ausência de vocabulários padronizados pode levar a inconsistências na definição de termos e estruturas de dados, dificultando a interoperabilidade entre sistemas. Além disso, o xAPI exige a implementação de um Learning Record Store (LRS), que serve como repositório

central para o armazenamento de dados. Isso demanda maior investimento técnico e financeiro, especialmente para instituições que buscam implantar o padrão em larga escala (Ryunosuke et al., 2016).

Por outro lado, o Caliper Analytics, criado pela 1EdTech, segue um modelo de desenvolvimento fechado. Sua propriedade intelectual pertence à 1EdTech, e a participação no processo de desenvolvimento exige associação paga. O Caliper Analytics foi projetado para priorizar a padronização e a coleta de métricas quantitativas. Ele define um modelo de dados estruturado baseado em eventos, com atributos como “ator”, “ação” e “objeto”, além de perfis métricos predefinidos. Essa abordagem garante interoperabilidade entre sistemas compatíveis, facilitando a comparação e análise de dados (Griffiths; Hoel, 2016).

O foco do Caliper Analytics na padronização faz com que seja mais atrativo para instituições que buscam soluções prontas e integradas. Sua integração com padrões adicionais da 1EdTech, como LTI, LIS e QTI, cria um ecossistema robusto de ferramentas de gestão da aprendizagem. Grandes fornecedores de AVA, como Blackboard e Moodle, são adotantes proeminentes do Caliper, o que impulsiona sua implementação em instituições que já utilizam essas plataformas (Väkevä, 2023).

No entanto, o modelo fechado do Caliper Analytics limita sua flexibilidade e personalização. A especificação é menos adaptável a contextos educacionais que fogem de ambientes tradicionais de AVA, como o aprendizado em campo ou experiências mais imersivas. Além disso, a falta de um sistema de armazenamento centralizado como o LRS pode ser uma barreira para instituições que precisam consolidar grandes volumes de dados de diversas fontes (Perez-colado et al., 2022).

Apesar de suas diferenças, xAPI e Caliper Analytics não são mutuamente exclusivos. As duas especificações podem ser usadas de forma complementar para atender a diferentes necessidades educacionais. O xAPI é ideal para instituições que valorizam flexibilidade e controle total sobre os dados de aprendizagem, enquanto o Caliper Analytics é mais indicado para aquelas que priorizam padronização e facilidade de implementação. Essa complementaridade abre espaço para a criação de ecossistemas de aprendizagem mais ricos, combinando a capacidade do xAPI de capturar experiências diversificadas com a estrutura padronizada do Caliper Analytics para métricas quantitativas (Samuelsen; Chen; Wasson, 2021a).

Além disso, a escolha entre as duas especificações deve considerar fatores como os objetivos de análise de dados, o nível de expertise técnica disponível e os recursos financeiros da instituição. Por exemplo, para organizações com equipes técnicas experientes e necessidades de personalização, o xAPI oferece maior liberdade para explorar dados de aprendizagem em diferentes contextos. Por outro lado, instituições que buscam soluções de rápida implementação e integração com sistemas de gestão de aprendizagem já existentes podem se beneficiar mais do Caliper Analytics.

A adoção de qualquer uma das especificações também implica desafios. A flexibilidade do xAPI pode resultar em dificuldades de interoperabilidade se práticas unificadas não forem estabelecidas. No caso do Caliper Analytics, a dependência de perfis métricos predefinidos pode limitar sua aplicabilidade em cenários educacionais mais inovadores. Além disso, a associação ao 1EdTech e os custos envolvidos podem restringir sua adoção por instituições menores ou com orçamento limitado (Griffiths; Hoel, 2016; Väkevä, 2023).

A tabela 2 a seguir apresenta um breve resumo das principais diferenças entre o Caliper Analytics e o xAPI; a tabela aborda aspectos como desenvolvedores, objetivos, modelos de dados, formatos, adoção e complexidade.

Tabela 2 – Comparação entre Caliper Analytics e xAPI

<b>Aspecto</b>	<b>Caliper Analytics</b>	<b>xAPI</b>
<b>Desenvolvedor</b>	1EdTech	ADL Initiative
<b>Objetivo</b>	Focado em interoperabilidade em ambientes educacionais	Flexível para capturar dados de aprendizagem em diversos contextos
<b>Modelo de dados</b>	Baseado em eventos predefinidos (e.g., modelos de eventos Caliper)	Estrutura aberta para atividades, contextos e resultados
<b>Formato</b>	JSON com esquema rígido	JSON com estrutura mais flexível
<b>Adoção</b>	Predominantemente em educação e plataformas IMS	Ampla adoção em diversas indústrias
<b>Complexidade</b>	Mais rígido e específico	Mais genérico e flexível

Independentemente da escolha, é essencial garantir que as saídas do xAPI e do Caliper Analytics possam ser mapeadas entre si para facilitar a interoperabilidade e a troca de dados entre sistemas. Essa integração não só enriquece a análise de dados de aprendizagem, como também promove a criação de experiências educacionais mais personalizadas e eficazes. No futuro, a convergência entre as duas especificações pode possibilitar ecossistemas de aprendizagem ainda mais integrados, atendendo às demandas diversificadas de instituições e alunos (Perez-colado et al., 2022; Väkevä, 2023).

Em suma, o xAPI e o Caliper Analytics representam abordagens distintas, mas complementares, no rastreamento de dados de aprendizagem. A escolha entre eles deve ser guiada pelas necessidades específicas da instituição, considerando os benefícios e limitações de cada padrão. Enquanto o xAPI oferece flexibilidade e adaptabilidade, o Caliper Analytics garante padronização e simplicidade de implementação. Usados em conjunto, eles têm o potencial de transformar a coleta, análise e uso de dados educacionais, criando um cenário de aprendizagem mais eficiente e personali-

zado.

## 2.4 Learning Record Storage

O Learning Record Store (LRS) é um componente central na arquitetura do padrão xAPI, desempenhando o papel de repositório para o armazenamento e gestão de dados relacionados a experiências de aprendizagem. Conforme definido pela ADL, o LRS é “um servidor capaz de receber, processar e fornecer acesso a registros de aprendizagem”. Ele é essencial para a implementação do xAPI, garantindo que as declarações (*statements*) enviadas por diferentes sistemas sejam validadas, armazenadas e acessíveis para análise e integração com outras ferramentas (xapi, 2018; IEEE, 2023).

O LRS utiliza uma arquitetura distribuída, permitindo que diferentes repositórios armazenem dados de maneira descentralizada e se comuniquem entre si. Isso possibilita que experiências de aprendizagem acompanhem os indivíduos entre organizações e sistemas, promovendo interoperabilidade. Um LRS pode ser público, privado ou híbrido, compartilhando dados de forma limitada, dependendo das necessidades da instituição (Labba; Roussanaly; Boyer, 2017).

Projetado para atender às especificações do xAPI, o LRS armazena declarações xAPI geradas por Learning Record Providers (LRP) e disponibiliza essas informações para Learning Record Consumers (LRC). Ele é, essencialmente, um serviço web que utiliza *REST* para garantir a transferência eficiente de dados, funcionando como um *hub* central em um ecossistema de aprendizagem empresarial ou educacional. Além disso, sua integração opcional com sistemas como LMS (Learning Management Systems), plataformas de recursos humanos (RH) ou outras aplicações facilita a centralização (Fig. 9) e personalização da análise de dados (Lockwood hills federal, 2019; Miller et al., 2021).

A conformidade com a especificação xAPI é fundamental para o funcionamento de um LRS. Antes de armazenar uma declaração, o LRS valida seu formato para garantir que ela esteja em conformidade com o padrão. Essa validação é crucial para assegurar a qualidade dos dados armazenados e a interoperabilidade com outros sistemas. Por isso, organizações que implementam um LRS devem garantir que ele passe por uma *Conformance Test Suite*, demonstrando sua adesão ao padrão xAPI (Llc, 2019).

Um dos principais benefícios do LRS é sua capacidade de fornecer rastreamento granular de dados de aprendizagem. Isso permite que organizações analisem o progresso de estudantes em detalhes e personalizem ações com base em resultados específicos. Por exemplo, se um aluno falhar em uma avaliação, o LRS pode ser utilizado para realizar análises que possibilitem a recomendação de materiais ou cursos adicionais, contribuindo para o aprimoramento do aprendizado. Essa funcionalidade

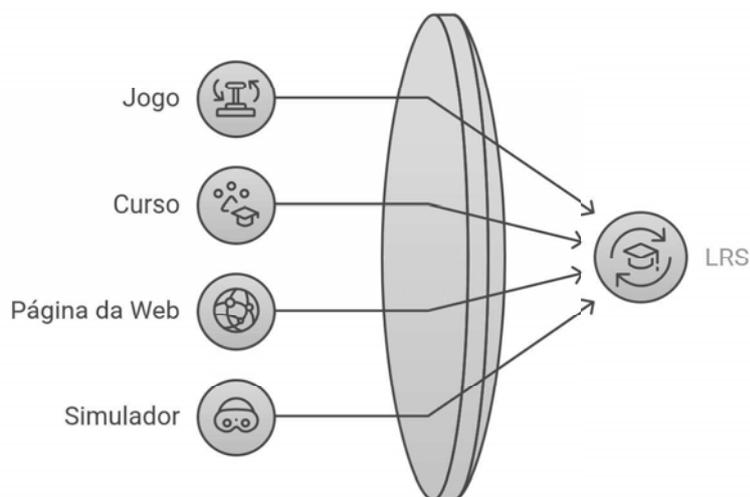


Figura 9 – Coleta dados xapi. Fonte: Adaptado de (Väkevä, 2023)

não apenas aumenta a precisão no rastreamento, mas também favorece a personalização das experiências educacionais (Miller et al., 2021).

Na prática, o LRS serve como elemento central para ferramentas e sistemas que transmitem ou obtêm dados de aprendizagem usando a especificação xAPI. Isso permite que as organizações acompanhem e analisem atividades de aprendizado, desde a inscrição e conclusão em cursos até interações complexas em simulações ou jogos sérios. Este aspecto é especialmente importante em ambientes corporativos, onde o LRS auxilia na mensuração da eficácia dos treinamentos, na identificação de lacunas de aprendizado e na tomada de decisões baseadas em dados para melhorar iniciativas educacionais (learningpool, 2024; Costa, 2021).

Além disso, a interoperabilidade promovida pelo LRS permite que plataformas extensíveis integrem serviços externos, como ferramentas de criação de conteúdo e sistemas de administração de AVA. Essa integração facilita a troca de dados entre sistemas e melhora a eficiência das operações educacionais (Asthagiri; Malik; Zarecor, 2018).

Por último, o LRS assume um papel estratégico na evolução da educação digital, permitindo o rastreamento de dados, a customização de experiências e uma tomada de decisões mais informada. A implementação sólida e em conformidade com o padrão xAPI é crucial para organizações que desejam maximizar os benefícios da aprendizagem orientada por dados.

## 2.5 ETL

Os processos de análise de dados requerem pré-processamento antes da aplicação de técnicas ou algoritmos, a fim de aumentar a qualidade desses dados e adaptá-los aos formatos necessários para seu processamento, principalmente quando os dados vêm de fontes diferentes. A preparação dos dados inclui tarefas de coleta, compressão, limpeza, transformação, integração e carga. Na coleta e compressão de dados, podem surgir problemas como: fontes de dados com estruturas diferentes, dados ausentes, dados atípicos, dados errôneos que geram conflitos e violação de validações, e atualização tediosa do armazenamento (Oliveira; Belo, 2017). Esses problemas precisam ser tratados. Sem uma preparação e pré-processamento adequados, os resultados dos processos de análise e mineração estariam limitados, pois essa etapa permite conhecer a “forma” dos dados, para que se possam fazer julgamentos sobre sua confiabilidade e qualidade (Costello; Blackshear, 2019). Na limpeza e transformação, é necessário detectar inconsistências, ajustar os dados e tomar decisões sobre as medidas e dimensões necessárias. Os resultados obtidos devem focar na qualidade dos dados para garantir que a informação esteja em condições adequadas, ajustada à realidade e ao problema estudado, especialmente quando há um aumento significativo na quantidade de dados. Após a preparação dos dados, considera-se o *design* do esquema de armazenamento, onde eles serão integrados e carregados. Com isso concluído, deve-se garantir um conjunto de dados (*dataset*) confiável e pronto para o processamento (Oliveira; Belo, 2017). O tratamento de dados produzidos em ambientes educacionais, assim como em outras áreas de estudo, exige reunir os requisitos, definir as fontes de dados e escolher a ferramenta e as técnicas adequadas para posterior análise. A aplicação de diferentes técnicas e a realização de experimentos dependem do conhecimento e domínio dos dados educacionais a serem analisados. É importante fazer um diagnóstico inicial desses dados e do problema abordado, levando em consideração a experiência dos interessados e dos especialistas (Samuelson; Chen; Wasson, 2019a). A organização unificada dos dados é uma alternativa interessante para fornecer informações que permitam análises específicas por parte dos gestores educacionais. Além disso, ter os dados sob um único esquema dá a capacidade de visualizar de forma precisa os dados, que muitas vezes estão distribuídos em vários softwares dentro das instituições educacionais, facilitando a busca por informações específicas, fator decisivo para a tomada de decisões nos momentos apropriados e uma diferença importante para os administradores da área educacional (De sá mussa et al., 2018). Por isso, é necessário explorar uma forma precisa e eficaz de coleta e pré-processamento de dados educacionais que permita integrar as múltiplas fontes, escalas e granularidades; levando em conta também que as necessidades de análise de dados acadêmicos são crescentes e exigem melhorias para otimizar o

aprendizado dos estudantes e a eficácia institucional (Mhon; Kham, 2020).

O processo de ETL (Extração, Transformação e Carga) é fundamental no gerenciamento de dados educacionais, pois permite integrar informações provenientes de diversas fontes de dados, como sistemas acadêmicos, plataformas de gestão escolar, registros de desempenho de estudantes e outros sistemas externos. A primeira fase do processo, extração, consiste em coletar os dados de diferentes fontes e garantir que todos os dados relevantes sejam recuperados. No contexto educacional, isso pode incluir a extração de informações sobre matrículas, notas, presença, atividades extracurriculares e outros dados relacionados ao desempenho e à progressão dos alunos. Essa fase exige que os dados sejam coletados de forma eficiente, sem perda de informações importantes, e muitas vezes envolve a conexão com vários sistemas que armazenam os dados em diferentes formatos e estruturas (Hugo; Lopes, 2023).

A segunda etapa do processo de ETL é a transformação, que consiste em limpar, validar e padronizar os dados extraídos. No contexto educacional, os dados podem estar desorganizados, apresentar valores ausentes ou até inconsistentes, o que torna a transformação crucial para garantir que as informações sejam precisas e utilizáveis. A transformação pode envolver a conversão de formatos de dados, a remoção de duplicatas, a substituição de valores ausentes por valores padrão ou a correção de dados errôneos. Além disso, é essencial realizar a agregação e a normalização dos dados, de forma que eles possam ser analisados de maneira comparativa e significativa, permitindo uma melhor compreensão do desempenho acadêmico dos alunos e das operações escolares (Barbosa, 2020).

Após a transformação, vem a fase de carga, onde os dados são carregados em um repositório centralizado, como um banco de dados ou *data warehouse*, onde serão armazenados e acessados para análises posteriores. No contexto educacional, o repositório de dados deve ser projetado de forma a facilitar consultas rápidas e seguras, oferecendo uma visão unificada das informações. A carga de dados pode ser realizada de maneira incremental ou total, dependendo da necessidade de atualização das informações. Uma carga bem-sucedida garante que os dados transformados estejam disponíveis para análises avançadas, como a criação de relatórios, *dashboards* e modelos preditivos que ajudam na tomada de decisões, planejamento escolar e melhorias no processo de ensino-aprendizagem (Hernández-leal; Costa-rocha; Duque-méndez, 2022).

O processo de ETL para dados educacionais não se limita à simples movimentação de dados entre sistemas, mas é uma etapa estratégica para garantir a qualidade e a integridade da informação. O sucesso desse processo depende da utilização de ferramentas e técnicas adequadas para cada fase, bem como de um monitoramento constante para corrigir possíveis falhas e melhorar a eficiência do processo (Hernandez-leal, 2024).

## 3 REVISÃO DA LITERATURA

Uma revisão da literatura é um estudo que tem o objetivo de identificar, analisar e interpretar as evidências disponíveis em estudos ou trabalhos acadêmicos relacionados a uma pergunta de pesquisa específica. De acordo com Kitchenham et al. (2007), a condução de uma revisão sistemática da literatura demanda o planejamento, a execução e o relato da revisão. Esta tarefa implica um esforço significativo, incluindo a criação do protocolo, o ajuste da cadeia de pesquisa, a filtragem dos resultados com critérios de inclusão e exclusão, a seleção de artigos, a compilação do relatório dos resultados e responder às perguntas da pesquisa.

### 3.1 Metodologia da revisão sistemática

Para facilitar a organização dos elementos a serem incluídos no protocolo de revisão da literatura, foi usada a ferramenta Parsifal <sup>1</sup>. Essa ferramenta facilita o desenvolvimento das três etapas fundamentais definidas na metodologia de Kitchenham et al. (2007) para revisões de literatura no domínio da engenharia de software, que é apresentada na Figura 10.

Nessa ordem de ideias, desenvolvendo a metodologia, na primeira etapa foram seguidos os três passos definidos. Como primeiro passo, a definição das questões de pesquisa: RQ1. Quais são os principais desafios para a modelagem de uma arquitetura de suporte à interoperabilidade de dados educacionais para a aplicação da análise de aprendizagem?, RQ2. Quais fontes de dados educacionais usam a interoperabilidade na análise de aprendizagem?, RQ3. Quais são as especificações de interoperabilidade usadas para apoiar a aplicação da análise de aprendizagem em dados provenientes de diferentes fontes e níveis de granularidade?

Durante as segunda e terceira etapas, o desenvolvimento do protocolo de análise e sua validação foram conduzidos de forma iterativa. Inicialmente, foi determinado o objetivo da revisão, que consistia em realizar uma revisão da literatura para investigar o estado da arte relacionado à interoperabilidade em LA. Posteriormente, foram

---

<sup>1</sup><https://parsif.al/>

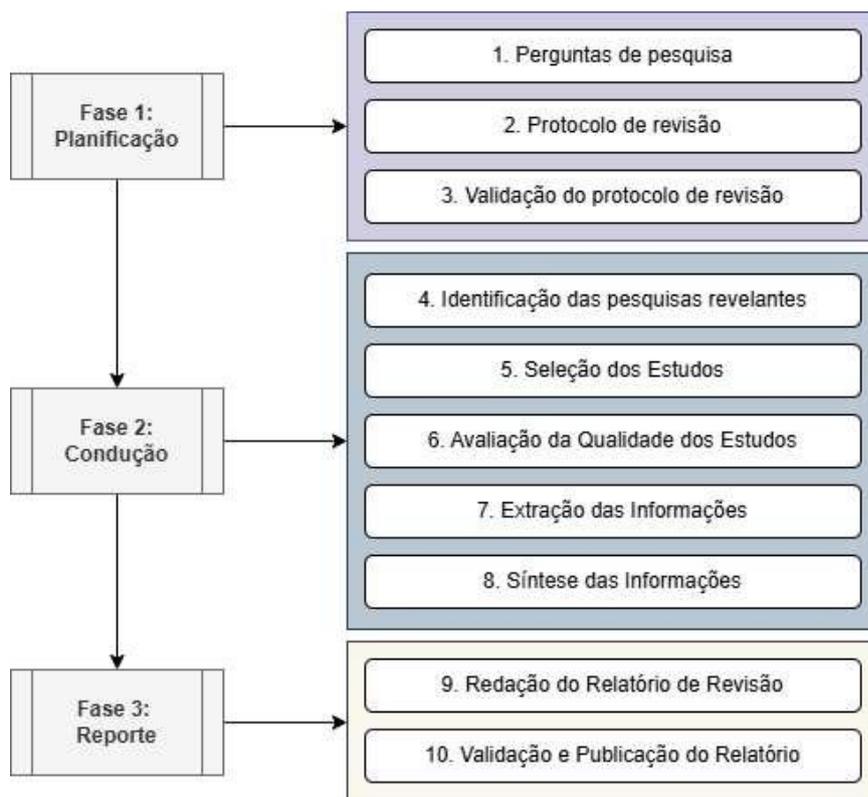


Figura 10 – Etapas do processo de revisão. Fonte: Autor

estabelecidos os critérios PICOC (População, Intervenção, Comparação, Resultado e Contexto). Após isso, definiram-se as palavras-chave juntamente com seus sinônimos (veja a Figura 11).

Keywords and Synonyms		
To edit or remove a certain keyword or synonym you may click on it's description to enable the field.		
Keyword	Synonyms	Related to
data source	Data collection Data store Information source	Population
interoperability	Data interchange Educational data specifications Integration architecture	Intervention
learning analytics	Academic analytics Educational data mining	Outcome

Figura 11 – Palavras-chave e sinônimos. Fonte: Autor

Conforme abordado anteriormente, a ferramenta cria a *string* de pesquisa utilizando operadores booleanos; a equação está apresentada a seguir em (1).

("data source" OR "Data collection" OR "Data store" OR "Information source")

AND ("interoperability" OR "Data interchange" OR "Educational data specifications" OR "Integration architecture")

AND ("learning analytics" OR "Academic analytics" OR "Educational data mining")

(1)

Durante a fase inicial de planejamento, foram escolhidos os bancos de dados a serem consultados, como Scopus, Web of Science e Google Acadêmico. Em seguida, elaborou-se uma Lista de Verificação de Avaliação de Qualidade, que inclui perguntas a serem utilizadas posteriormente para avaliar a qualidade dos artigos e integrar os critérios para seleção dos estudos. A Tabela 3 oferece um resumo dessas perguntas, as respostas possíveis e as respectivas pontuações.

Tabela 3 – Lista de Verificação para Avaliação da Qualidade

Perguntas	Respostas		
	Sim	Parcialmente	Não
O estudo utiliza interoperabilidade?	1.0	0.5	0.0
São utilizadas fontes de dados educacionais?	1.0	0.5	0.0
Menciona técnicas de Learning Analytics?	1.0	0.5	0.0
O estudo utiliza algum padrão ou especificação para interoperabilidade?	1.0	0.5	0.0
<b>Pontuação</b>	<b>4.0</b>	<b>2.0</b>	<b>0.0</b>

Durante a etapa 6, essas perguntas são empregadas para avaliar a qualidade dos estudos. Um estudo recebe 1,0 de pontuação se a resposta for 'Sim', 0,5 se for parcialmente concordante, e 0,0 se não estiver de acordo ou faltar algo. Assim, o máximo que um estudo pode receber é 4,0, com a pontuação de corte estabelecida em 2,0.

Caso o estudo alcance uma pontuação inferior a 2,0 na avaliação das quatro questões, ele será excluído da análise. A ferramenta possibilita uma análise de qualidade após a primeira triagem dos estudos, apresentando as opções de resposta para cada caso e atribuindo automaticamente a nota de acordo com a parametrização prévia. Finalmente, a fase de planejamento terminou com a definição do Formulário de Extração de Dados (consulte a Tabela 4).

Tabela 4 – Formulário de Extração de Dados

Descrição	Tipo	Valores
Armazenamento de Dados	Selecionar um campo	DWH Excel - CSV LRS LRW NAN NRDB RDBS
Fonte de Dados	Selecionar vários campos	Provedor de autenticação Repositório de documentos Sistema de Gerenciamento de Aprendizagem NAN Outro Ferramenta de relatórios/análises Sistema de Informações do Estudante
Learning Analytics (técnica)	Campo de texto	n/a
Caliper (utilizado)	Campo booleano	Sim/Não
xAPI (utilizado)	Campo booleano	Sim/Não

Na segunda fase, denominada condução, utilizamos a string de pesquisa para coletar os artigos de cada um dos bancos de dados selecionados, que foram posteriormente importados para a ferramenta Parsifal através de arquivos BibTeX. Inicialmente, uma seleção foi realizada com base na correspondência dos títulos com o objetivo do estudo, procedendo-se à avaliação da qualidade (etapa 6) e à extração das informações (etapa 7). Ao término dessa segunda fase, procedemos à análise e síntese das informações que serão detalhadas nas seções seguintes deste documento.

A terceira fase, referente ao relatório, abrangeu a elaboração deste capítulo, que constitui o relatório inicial do processo de revisão efetuado. É denominado 'relatório inicial', pois se planeja iterar no processo, formulando novas perguntas para avaliar a qualidade dos estudos, ajustando os componentes utilizados para a extração de dados e, por fim, ampliando os bancos de dados revisados.

### 3.2 Resultados: Análise Bibliométrica

Nas fases dois e três da metodologia, a Tabela 5 exibe os resultados da pesquisa nos bancos de dados escolhidos, incluindo os achados iniciais obtidos pela equação (1). No caso do Google Acadêmico, foi utilizado um critério de exclusão, e, finalmente, é mostrado o total levado à ferramenta para realizar a etapa 5 (seleção de estudos) e

a etapa 6 (avaliação da qualidade dos estudos).

Tabela 5 – Estudos Extraídos por Banco de Dados

Resultados	Bases de Dados		
	Google Acadêmico	Scopus	Web of Science
Equação	481	16	31
O critério de exclusão foi aplicado?	Sim	Não	Não
Critério de exclusão	Ano >= 2019	n/a	n/a
Total	244	16	3

A seguir, é apresentado um extrato da análise bibliométrica, realizada para os resultados dos bancos de dados e, posteriormente, para os estudos que foram selecionados para o relatório. A Figura 12 mostra a distribuição dos documentos encontrados por área no banco de dados Scopus. Para esse mesmo banco de dados, a Figura 13 apresenta a contagem de documentos por país ou território de publicação.

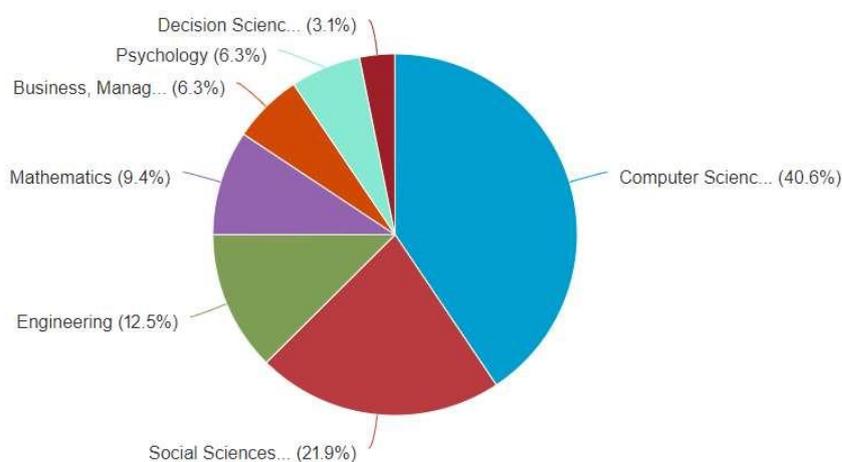


Figura 12 – Documentos encontrado no Scopus por área temática. Fonte: Autor

Agora, a Figura 14 refere-se aos documentos classificados pelas categorias do Web of Sciences, usando um gráfico TreeMap. Enquanto isso, a Figura 15 mostra a relação entre o ano de publicação e as citações dos estudos revisados da WoS.

Uma vez concluído o processo de seleção e avaliação da qualidade dos estudos, para a etapa 7 (extração de informações), trabalhou-se com um total de vinte estudos primários. A Figura 16 mostra a lista desses estudos selecionados e seus anos de publicação. O resultado permite concluir que o tópico discutido neste estudo é relativamente novo, datando dos últimos dez anos. Além disso, os anos de 2019 e 2021 mostram os maiores aumentos na tendência, indicando um crescimento contínuo desde 2018.

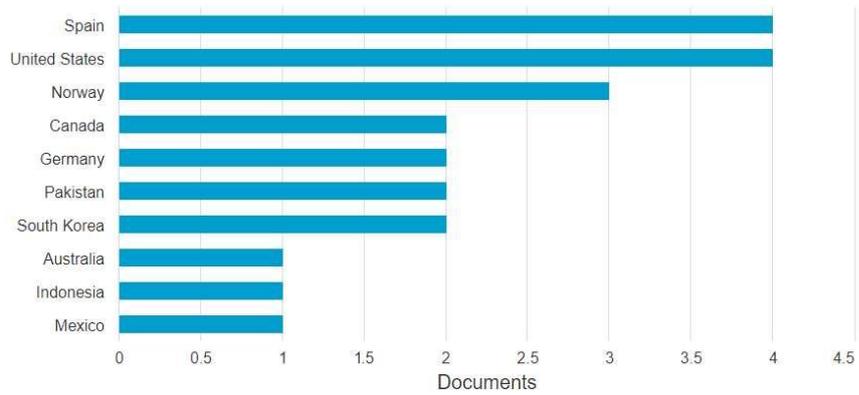


Figura 13 – Documentos por país ou território na Scopus . Fonte: Autor



Figura 14 – Categorias da Web of Science. Fonte: Autor

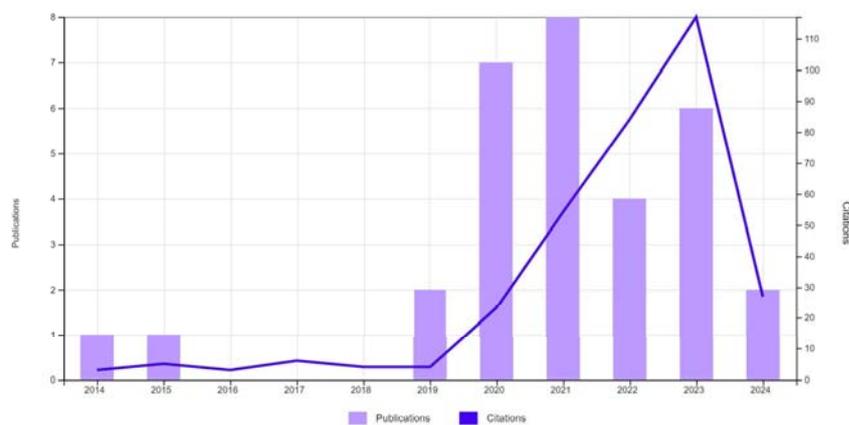


Figura 15 – Citações e publicações ao longo do tempo da Web of Science. Fonte: Autor

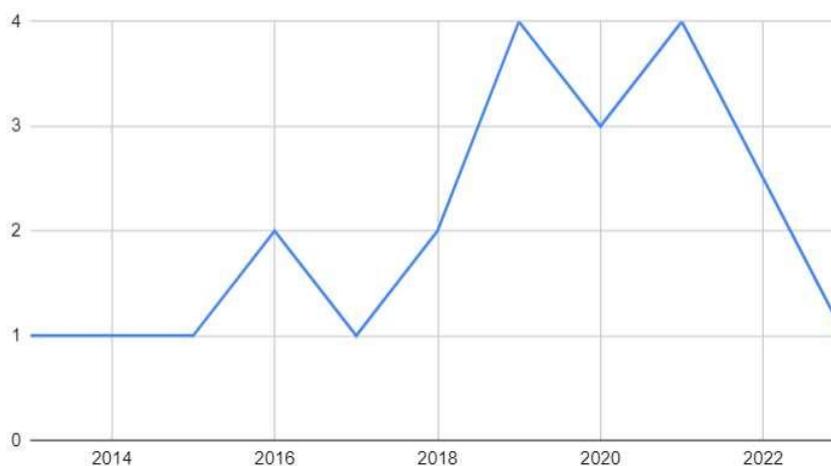


Figura 16 – Estudos selecionados de publicação do ano da linha de tendência da Web of Science. Fonte: Autor

As informações extraídas e a síntese realizada (etapa 8) são apresentadas na seção a seguir.

### 3.3 Síntese Dos Trabalhos Encontrados

Nos últimos anos, os sistemas educacionais testemunharam um rápido crescimento na integração de tecnologias on-line como parte da experiência de aprendizado do aluno. Essa integração de tecnologias educacionais, juntamente com a coleta de dados dos estudantes, levou ao aumento da aplicação de LA para a tomada de decisões orientadas por dados para melhorar o aprendizado dos estudantes (Paik et al., 2024b).

Nesse contexto, à medida que as tecnologias são cada vez mais adotadas, os sistemas educacionais deixam de ser isolados e precisam ser capazes de incorporar componentes e dados educacionais produzidos em um ambiente, utilizando um conjunto específico de ferramentas e plataformas, e aplicá-los em outro ambiente com ferramentas ou plataformas distintas. Em sistemas colaborativos, o manejo de fontes de dados variadas representa um grande desafio. Contudo, a integração de dados heterogêneos por meio da LA possibilita uma análise abrangente, potencializando uma tomada de decisão mais precisa no domínio educacional.

A integração de dados está intimamente relacionada à interoperabilidade, que envolve os níveis semântico, técnico, legal e organizacional (European commission, 2017). Com relação à interoperabilidade técnica e semântica, há duas especificações de dados amplamente conhecidas (padrões do setor) que visam ao domínio educacional e à LA, a saber, a Experience API (xAPI) e o Caliper Analytics (Samuelsen; Chen; Wasson, 2019b).

Ao adotar essas especificações, um vocabulário comum é estabelecido, o que é

necessário para garantir que todas as atividades de aprendizagem entre todos os diferentes sistemas descrevam a mesma experiência. Ao formalizar o vocabulário usado, um conjunto de atributos e regras sobre os dados é estabelecido, o que determina como eles serão armazenados, recuperados e acessados por outros componentes, sistemas ou atividades (Lockwood hills federal, 2019). Esses benefícios contribuem para um intercâmbio mais eficaz e contínuo de dados para fins analíticos. Podemos pensar na interoperabilidade como uma forma de reduzir a energia gasta na transformação e integração de dados manualmente. Nesse sentido, podemos usar especificações já estabelecidas no mercado, como xAPI ou Caliper Analytics.

No estudo realizado por Whyte; Nayak; Johnston (2016), a especificação Caliper Analytics é discutida para oferecer um modelo de informações extensível, vocabulários controlados e uma API para registrar eventos de aprendizagem. Entretanto, reconhece-se que muitas atividades de aprendizagem ainda não foram modeladas pelo grupo de trabalho do Caliper Analytics. Para superar essa deficiência, o documento sugere a participação da comunidade SoLAR no desenvolvimento de novos perfis de atividades de aprendizagem conforme a especificação Caliper.

Em Samuelsen; Chen; Wasson (2019b) examina-se o estado atual da análise de aprendizagem (LA) com foco na integração de dados em contextos de ensino superior. O documento descreve a escalabilidade na LA como um desafio a ser alcançado. A revisão resultou em 20 estudos analisados. As descobertas revelam que algumas fontes de dados, principalmente os Sistemas de Gerenciamento de Aprendizagem, são utilizadas com frequência em estudos de LA. No entanto, o número de fontes de dados integradas tende a ser limitado, e formatos de dados semelhantes são frequentemente combinados, o que representa um desafio técnico relativamente simples. Além disso, a literatura geralmente carece de informações detalhadas sobre os processos de integração de dados nos sistemas implementados, e as especificações de dados educacionais, como a xAPI, são subutilizadas, apesar de seus possíveis benefícios.

Por outro lado, Samuelsen; Chen; Wasson (2021b) discutem os desafios ligados ao uso dos padrões de dados de atividades de aprendizagem disponíveis atualmente, como a xAPI, na descrição do contexto educacional e na promoção da interoperabilidade e integração de dados no campo de LA. Embora padrões como o xAPI possam aumentar a integração de dados provenientes de diferentes fontes, estudos indicam que eles ainda são pouco utilizados na LA. Para abordar essa lacuna, o artigo apresenta os resultados de um estudo de caso exploratório que envolveu a realização de entrevistas aprofundadas com as partes interessadas que utilizaram a xAPI em projetos práticos. Por meio da análise temática dessas entrevistas e de um exame da própria xAPI, o estudo identifica desafios e limitações na descrição dos dados do contexto de aprendizagem. Além disso, o documento fornece recomendações concisas para enriquecer as descrições de contexto e aumentar a expressividade da xAPI.

Baseando a pesquisa em cenários reais, este estudo visa conectar a pesquisa acadêmica às aplicações práticas, especialmente focando nos padrões e na escalabilidade dos dados de atividades de aprendizagem, enfatizando a descrição do contexto de aprendizagem.

Em Rayon; Guenaga; Nuñez (2015), é abordado o desafio de avaliar competências em um contexto educacional, no qual as universidades priorizam cada vez mais as habilidades e competências em detrimento da mera aquisição de conhecimento. Aproveitando os dados gerados pelos estudantes em aplicativos educacionais, há potencial para estudar como ocorre a aprendizagem e reunir evidências para a avaliação de competências. No entanto, a descentralização e a falta de interoperabilidade entre esses aplicativos educacionais apresentam obstáculos à utilização eficaz desses dados. Para enfrentar esses problemas, os pesquisadores criaram o SCALA (Avaliação de Competências Escalável através de uma abordagem de Análise de Aprendizagem), um sistema analítico que combina dados de utilização e de rastreamento social a fim de auxiliar na avaliação de competências. O artigo analisa o processo de extração, transformação e carregamento (ETL) no contexto da integração de diversas fontes de dados em um modelo de dados unificado, seguindo a estrutura Caliper Analytics para assegurar a integridade e interoperabilidade dos dados.

Por outro lado, Doderó et al. (2017) discutem os desafios impostos pela heterogeneidade dos sistemas externos em ambientes de e-learning, especialmente em relação aos requisitos de interoperabilidade e desempenho para registrar e armazenar dados de aprendizagem. O autor enfatiza a criação de protocolos baseados na Web voltados para melhorar a interoperabilidade e permitir análises significativas em sistemas de e-learning. O artigo é centrado em um ambiente de aprendizagem online dedicado ao treinamento de pessoas no comando e controle de veículos autônomos não tripulados, incorporando funcionalidades analíticas. O ambiente une um sistema externo de gestão de conteúdo para a Web e um motor de simulação, ambos possuindo requisitos de desempenho diversos para o registro de eventos importantes no decorrer do processo de aprendizagem.

Já em Samuelsen (2018), o objetivo foi abordar a complexa tarefa de implementar LA a nível institucional por meio da construção de uma arquitetura de LA escalável. Foi realizada uma revisão sistemática para examinar o cenário atual de pesquisas e práticas relacionadas ao uso e à combinação de várias fontes de dados na análise da aprendizagem. Essa análise forneceu percepções valiosas que informaram o desenvolvimento de um modelo conceitual inicial para a arquitetura. Na construção da arquitetura, as tecnologias da Web semântica desempenharão um papel fundamental no gerenciamento do significado dos dados compartilhados e na facilitação da interoperabilidade entre diversas fontes de dados.

No estudo de Del blanco et al. (2013), os autores exploram detalhadamente o

campo de LA, focando na análise de dados educacionais que emergem das interações dos estudantes com recursos *online*. Embora grande parte desses dados se origine de Sistemas de Gerenciamento de Aprendizagem implantados em instituições educacionais tradicionais, uma parte significativa também vem de outras plataformas de aprendizagem *online*, como Cursos Online Abertos e Massivos (MOOCs), como Udacity e Coursera, e iniciativas educacionais como a Khan Academy. No entanto, não existe um modelo de dados universalmente aceito para as interações dos estudantes, resultando em uma situação em que as ferramentas de análise necessitam ser adaptadas à estrutura de dados específica de cada sistema. Essa fragmentação reduz a interoperabilidade e aumenta os custos de desenvolvimento. Para enfrentar esse desafio, o documento examina como as principais ferramentas de LA coletam dados e correlacionam suas abordagens com os modelos de dados definidos por dois padrões de e-learning: o Padrão IEEE para Tecnologia de Aprendizagem e o xAPI. Ao analisar essas relações, o documento destaca os possíveis benefícios da adoção desses padrões de e-learning sob a perspectiva da Análise de Aprendizagem. Esta análise comparativa destaca os benefícios da padronização de modelos de dados no âmbito de LA, o que pode fomentar uma melhor interoperabilidade e diminuir a carga de desenvolvimento na área.

A pesquisa apresentada em Hoel; Chen (2020) ilustra um estudo de caso sobre um projeto norueguês de padronização voltado para lidar com os desafios enfrentados ao expandir LA. O estudo enfatiza os desafios conceituais que emergem quando partes interessadas com interesses diversos colaboram para alcançar um consenso sobre compartilhamento de dados e padrões de interoperabilidade. Com base nesse estudo de caso, o documento propõe construções conceituais e um processo estruturado para facilitar a construção de consenso sobre vários aspectos relacionados ao acesso e ao intercâmbio de dados de diversas fontes relevantes para a análise da aprendizagem e seus fatores contextuais.

No estudo conduzido por Nwachukwu et al. (2021), é apresentada uma implementação de um método através de um caso prático, que tem como foco simplificar a importação e comparação de registros de atividades provenientes de diversas fontes de dados, auxiliando os professores em seu processo instrucional. Além disso, são discutidas as implicações dessa abordagem para a análise de aprendizagem entre plataformas, juntamente com possíveis caminhos para pesquisa e desenvolvimento futuros. De modo geral, o documento oferece uma solução prática para o desafio da integração e agregação de dados na análise de aprendizagem, enfatizando a capacitação e a personalização do usuário para apoiar a análise e a tomada de decisões significativas em ambientes educacionais.

Mangaroska et al. (2021) argumentam a favor da utilização de dados provenientes de diversos sistemas de e-learning para adquirir um entendimento mais abrangente

sobre o comportamento dos estudantes e as sutilezas do processo de aprendizagem. Sustentam que, ao combinar análises de várias plataformas, é possível melhorar a avaliação de projetos educacionais e maximizar as oportunidades de aprendizagem em ambientes distribuídos. Para avançar nesse objetivo, o estudo apresenta uma arquitetura de plataforma cruzada projetada para capturar, integrar e armazenar dados relacionados à aprendizagem de três sistemas distintos de e-learning. Os resultados indicam que integrar dados de diferentes sistemas de aprendizado digital aumenta notavelmente a precisão na classificação em relação a usar dados de um único sistema, destacando a importância da análise de aprendizagem entre plataformas.

Da mesma forma, Hodrob; Ewais; Maree (2019) propõe um sistema de indicadores analíticos baseado em aprendizagem projetado usando linguagens de modelagem específicas do domínio, com foco no aprimoramento do processo de aprendizagem para estudantes de engenharia de sistemas de computação. O sistema inclui um componente de gerenciamento de desempenho destinado a monitorar e melhorar os resultados da aprendizagem. Os instrutores e formuladores de políticas podem definir indicadores, tanto qualitativos quanto quantitativos, e o sistema calcula automaticamente seus valores, fornecendo recomendações de ações para aprimorar o processo de ensino-aprendizagem. Um aspecto essencial do sistema é a aquisição de fontes de aprendizagem de diversas fontes de dados e informações disponíveis na Web. Diferentemente das abordagens tradicionais que dependem de uma única fonte de dados, o sistema aproveita várias fontes de aprendizado de dados baseadas na Web, aprimorando a riqueza e a diversidade dos materiais de aprendizagem disponíveis para os estudantes.

A pesquisa de Nguyen; Khánh (2019) procura explorar o desenvolvimento e a implementação de Sistemas de Informações Analíticas de Aprendizagem (LAIS) no âmbito do ensino superior. O estudo começa com a análise de vários conceitos relacionados à análise de dados no ensino superior, como a análise de aprendizagem (LA), a análise acadêmica (AA) e a mineração de dados educacionais (EDM). Depois, foca-se particularmente em LA, que está voltada para o aprendizado e o ensino. As principais dimensões da análise da aprendizagem são reconhecidas para criar uma taxonomia em múltiplos níveis de suas aplicações e desenvolver uma ontologia específica para o domínio. Além disso, são propostas estruturas conceituais e roteiros para a aplicação de LA em diversos ambientes de aprendizagem. O estudo estabelece bases para o desenvolvimento e a implementação do LAIS, delineando um conjunto de princípios de design e propondo uma arquitetura para o gerenciamento de processos e tecnologias dentro do LAIS.

Por outro lado, os desafios associados à implementação de LA em cursos on-line abertos e massivos (MOOCs) são discutidos por Rohloff et al. (2019), pois esses cursos produzem um grande volume de dados sobre as atividades de aprendizagem de

numerosos usuários. O artigo apresenta uma arquitetura para implementar a análise de aprendizagem em uma plataforma MOOC orientada a serviços, com foco específico na plataforma MOOC HPI. A arquitetura apresenta um serviço baseado em pipelines de processamento extensíveis e independentes de esquema, o que permite o manuseio flexível de diversas fontes e formatos de dados. A abordagem é avaliada quanto à escalabilidade, à extensibilidade e à versatilidade por meio de casos de uso do mundo real, com uma consideração cuidadosa das preocupações com a privacidade dos dados.

Hecking et al. (2014) abordam o cenário atual das infraestruturas de análise de aprendizagem, destacando a necessidade de soluções mais flexíveis e gerais que possam acomodar diferentes ambientes de aprendizagem e tarefas de análise. Embora as abordagens existentes possam ser adaptadas a tipos específicos de ambientes de aprendizagem ou limitadas em seus recursos analíticos, há uma lacuna na pesquisa de soluções mais abrangentes. O artigo descreve o projeto de uma arquitetura flexível e extensível para uma infraestrutura de análise de aprendizagem que engloba vários aspectos analíticos, inclusive armazenamento de dados, mecanismos de *feedback* e algoritmos de análise. Essa infraestrutura é construída com base em agentes de software de acoplamento frouxo, capazes de executar diferentes tarefas de análise de forma independente. Como resultado, a arquitetura permite a fácil extensão da funcionalidade analítica por meio da adição de novos componentes de agente, promovendo a adaptabilidade e o dimensionamento. Além disso, o documento enfatiza a interoperabilidade dessa infraestrutura, permitindo que os sistemas de análise existentes acessem os dados e utilizem os componentes da infraestrutura como um serviço.

Da mesma forma, Arroyo-figueroa et al. (2019) apresentam um projeto em andamento que visa ao desenvolvimento de um ecossistema de aprendizagem para o treinamento de eletricitas em manobras de manutenção, aproveitando tecnologias eletrônicas como realidade virtual e sistemas de gerenciamento de aprendizagem. O objetivo é enfrentar os desafios de interoperabilidade, escalabilidade e utilização de dados de rastreamento gerados pelos usuários nesses sistemas. O ecossistema proposto foi desenvolvido com base na especificação xAPI, incorporando componentes de Big Data e técnicas de análise de aprendizado. O artigo descreve a arquitetura do ecossistema e o processo de coleta de experiências de aprendizagem e estados emocionais dos usuários. O documento apresenta uma nova abordagem para enfrentar os desafios do treinamento de eletricitas por meio da integração de tecnologias eletrônicas, análise de dados e caminhos de aprendizagem personalizados.

Além disso, Tlili et al. (2018) explora LA dentro do contexto da educação *online*, ressaltando seu potencial para melhorar as experiências de estudo. Contudo, ele aponta para um grande obstáculo: a ausência de diretrizes claras para a implemen-

tação eficiente da LA, particularmente no que diz respeito à preparação de dados. O artigo tem como objetivo abordar esse desafio investigando as principais questões que devem ser consideradas ao projetar experiências de LA, com foco na preparação de dados. Por meio dessa investigação, os autores pretendem fornecer uma estrutura abrangente, composta por vinte e duas questões de design que as partes interessadas devem considerar em vários contextos. Além disso, o artigo oferece um conjunto de diretrizes para aprimorar o projeto de experiências de AL, facilitando assim sua implementação eficaz em ambientes educacionais.

Por fim, Perez-berenguer; Kessler; Garcia-molina (2020) introduz uma arquitetura de análise de aprendizagem desenvolvida no Centro de Produção de Conteúdo Digital da Universidade Técnica de Cartagena (Espanha), com o objetivo de aprimorar os processos de ensino e aprendizagem. A arquitetura oferece duas contribuições significativas para o campo de LA: personalização de painéis e maior eficiência na análise dos dados de interação dos estudantes. Os principais recursos da arquitetura incluem a captura e o armazenamento de eventos resultantes das interações dos estudantes no formato padrão Caliper Analytics, facilitando o processamento incremental para a exibição imediata do painel aos professores. A personalização dos painéis é destacada como um requisito essencial para as ferramentas de análise de aprendizagem, e o documento aborda esse aspecto propondo uma linguagem específica de domínio (DSL) que permite aos professores personalizar os painéis de acordo com suas necessidades. A DSL permite que os professores expressem indicadores (expressões lógicas) para classificar os estudantes em diferentes grupos de desempenho.

Como conclusão desta seção e em resposta às perguntas que orientaram a análise do estado da arte, as respostas a cada uma dessas perguntas são apresentadas a seguir.

A integração de dados educacionais enfrenta desafios significativos devido à diversidade de fontes de dados, que têm formatos e estruturas variadas, tornando a integração complexa. A integração de dados está intimamente relacionada à interoperabilidade, envolvendo níveis semânticos, técnicos, legais e organizacionais. Outro problema observado foi a ausência de padronização, uma vez que não havia um modelo de dados universal para as interações dos estudantes. Isso resultou em altos custos de desenvolvimento e pouca interoperabilidade. Apesar das especificações que existiam, ainda não havia um modelo unificado. A interoperabilidade técnica e semântica foi igualmente um desafio, já que era necessário assegurar tanto a comunicação eficaz quanto a preservação do significado dos dados entre os sistemas. Especificações como xAPI e Caliper Analytics existiram, mas sua adoção foi ainda restrita. Ademais, a capacidade de lidar com grandes volumes de dados sem perder desempenho representou um problema de escalabilidade. Por último, as preocupações com privacidade e segurança foram significativas devido aos desafios legais e

éticos na integração e análise dos dados estudantis.

Nos estudos analisados, constatou-se que os dados derivavam principalmente dos Sistemas de Gerenciamento de Aprendizagem, como o Moodle, que registravam as interações e as métricas de engajamento dos estudantes. Outras fontes significativas incluíam os Cursos Online Abertos e Massivos (MOOCs) e os Sistemas de Informações Acadêmicas, que controlavam os registros estudantis e os dados administrativos. Havia também fontes externas, como dados de redes sociais e fóruns on-line, que ofereciam um contexto mais amplo das interações dos estudantes.

Confirmou-se que as duas principais especificações voltadas para o domínio educacional e análise de aprendizagem foram a Experience API (xAPI) e a Caliper Analytics. Apesar de essas especificações de dados para atividades de aprendizagem estarem disponíveis, elas não foram amplamente utilizadas para integrar dados de várias fontes. No caso da xAPI, foram observados poucos exemplos de sua aplicação no contexto deste estudo. Contudo, notamos um crescimento no número de artigos que utilizam o Caliper Analytics, totalizando seis artigos. Isso pode ser atribuído à habilidade do Caliper Analytics de contornar certos problemas decorrentes da flexibilidade da xAPI. Outro fator que pode ter impulsionado o uso do Caliper Analytics é o apoio de grandes empresas de tecnologia educacional através de um consórcio, o que pode representar uma barreira para pequenas instituições e pesquisadores que não conseguem arcar com os custos do suporte do consórcio.

## 4 CONCEPÇÃO DO TRABALHO

Este capítulo apresenta a concepção do trabalho desenvolvido, abordando o contexto dos experimentos, a metodologia proposta e os experimentos realizados. No andamento da pesquisa, foram alcançados resultados iniciais que auxiliaram a delinear a metodologia utilizada nesta dissertação.

### 4.1 Metodologia

O trabalho foi realizado com base em uma metodologia iterativa baseada em experimentos exploratórios, fundamentada no Design Science Research (DSR). Essa abordagem é amplamente utilizada para guiar o desenvolvimento e avaliação de artefatos tecnológicos em ciclos iterativos, permitindo resolver problemas práticos de forma sistemática, enquanto gera contribuições teóricas relevantes (Hevner et al., 2004).

O processo foi estruturado em dois experimentos iniciais, que ajudaram a orientar a criação e o refinamento da arquitetura final do estudo, apresentada na Figura 17 e discutida no próximo capítulo.

No primeiro experimento, foi realizada a instalação do OpenLRW, um software de código aberto desenvolvido pela Apereo Foundation, com capacidade de armazenar diversas especificações, como xAPI e Caliper Analytics. O objetivo foi configurar o LRW e, através dele, enviar dados de um AVA, neste caso, o Moodle, utilizando os plugins de Caliper Analytics e xAPI. Durante o experimento, foram identificadas falhas na abordagem inicial, o que possibilitou a proposição de uma nova arquitetura para o LRW.

No segundo experimento, foi testada a capacidade de envio de especificações Caliper para o LRS. Esse experimento serviu de base para o desenvolvimento de um conversor de Caliper para xAPI, elemento central da nova arquitetura proposta.

A metodologia Design Science Research permitiu que os experimentos fossem realizados de forma iterativa, explorando soluções para desafios relacionados à implementação prática de um Learning Record Warehouse. O trabalho concentrou-se em abordar problemas como a interoperabilidade e o armazenamento de dados educacio-

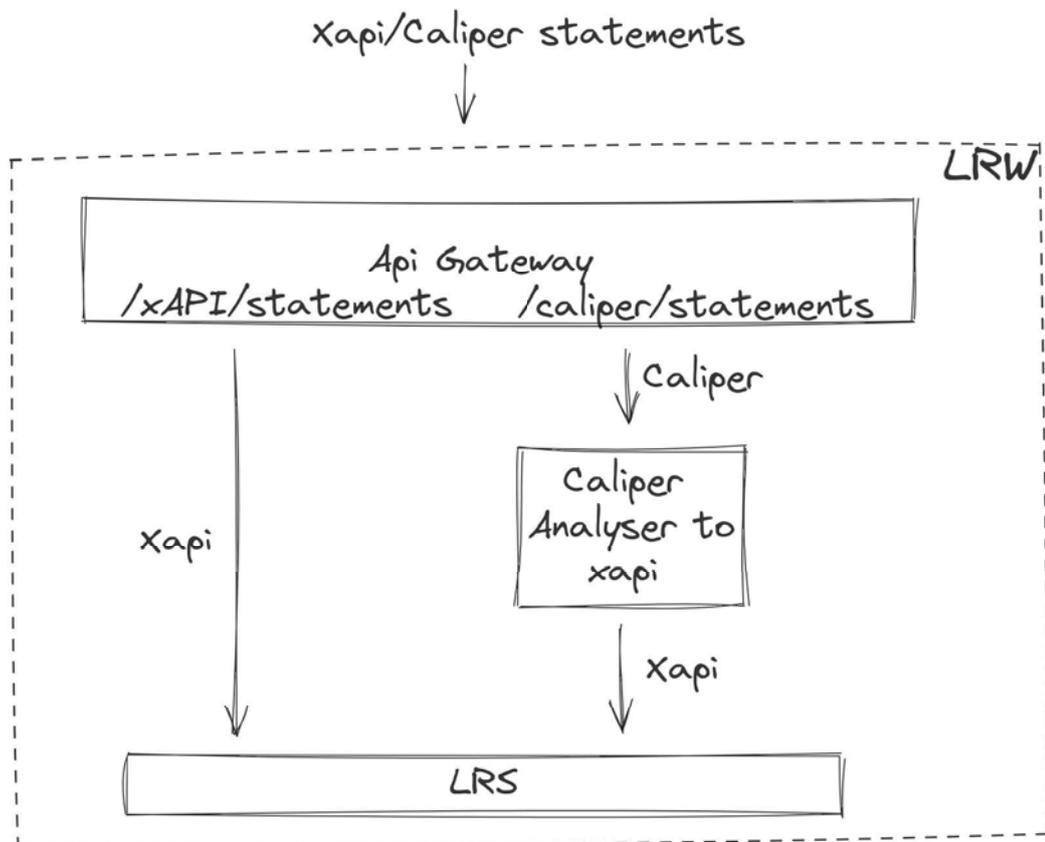


Figura 17 – Arquitetura LRW. Fonte: Autor.

nais provenientes de múltiplas fontes. Ademais, mediante a aplicação da metodologia DSR, garantiu-se que os artefatos criados – a nova arquitetura do LRW e o conversor de Caliper para xAPI – fossem analisados e refinados com base em evidências empíricas obtidas durante os experimentos.

Segundo Hevner et al. (2004), o DSR estrutura-se em ciclos iterativos que incluem a construção, avaliação e refinamento de artefatos, sempre conectando os resultados obtidos à literatura existente. Esse processo garante que os artefatos sejam robustos e tenham impacto prático e acadêmico.

Na Figura 18, é ilustrado o fluxo através do qual os experimentos são conduzidos até se alcançar a arquitetura final.



Figura 18 – Passos de desenvolvimento do trabalho. Fonte: Autor.

Essa abordagem metodológica possibilitou o refinamento progressivo das soluções propostas, conectando os resultados obtidos com a literatura existente e assegurando a relevância prática e acadêmica do trabalho.

## 4.2 Experimento I

Dentro deste experimento, foi configurado um software que possibilita o armazenamento dos eventos educativos gerados por um AVA, especificamente o Moodle: o OpenLRW, que oferece suporte para as tecnologias Caliper Analytics e xAPI.

No estudo conduzido por Syed et al. (2019), apresenta-se um Learning Record Warehouse (LRW), que amplia o conceito de Learning Record Store (LRS) referido no xAPI. Introduce-se uma ferramenta denominada OpenLRW, desenvolvida em Java, para armazenar eventos provenientes de diversas especificações como xAPI e Caliper Analytics. Dessa forma, implementar o conceito de LRW viabiliza o uso de múltiplas especificações, facilitando a interoperabilidade almejada.

OpenLRW é uma ferramenta de software de código aberto, codificada em Java, sob a manutenção da Apereo Foundation e disponível no GitHub<sup>1</sup>. Este software foi desenvolvido para substituir o OpenLRS, que teve seu suporte oficialmente encerrado em 28 de junho de 2017 (Apereo; Gilbert, 2017). O OpenLRS inicialmente suportava as especificações xAPI e Caliper Analytics, adicionando o suporte ao último em 2016. Contudo, com a introdução do OpenLRW, que ampliou suas funcionalidades para incluir outras especificações além de xAPI e Caliper, foi decidido descontinuar o OpenLRS. O OpenLRW utiliza o modelo de dados OneRoster da 1Edtech, com o Caliper Analytics

<sup>1</sup><https://github.com/Apereo-Learning-Analytics-Initiative/OpenLRW>

da mesma entidade atuando como padrão para armazenamento de eventos, refletindo seu design adaptado à educação. Como o OpenLRW era totalmente dedicado a aplicações na educação, optou-se por utilizar a especificação Caliper Analytics para armazenamento, dado que o xAPI tinha um escopo mais abrangente. Além disso, muitos sistemas integrados pela Apereo Foundation, incluindo ambientes de aprendizagem virtual (AVAs) como Moodle e Canvas, já empregavam o Caliper Analytics como o padrão para captura de eventos (Whyte; Nayak; Johnston, 2016).

O Sensor Caliper gera e formata um evento no padrão Caliper partindo das informações previamente tratadas, encaminhando-o ao servidor que aloja OpenLRW, mais precisamente, para o endpoint `/v1/caliper`.

#### **4.2.1 Softwares Adicionais Necessários**

A instalação do OpenLRW exigiu a utilização de softwares adicionais, incluindo Docker, Docker Compose, MongoDB e Spring Boot. Assim como no caso do Moodle, a instalação foi realizada por meio do Docker em conjunto com o Docker Compose, permitindo a criação de um ambiente controlado e replicável. Com o ambiente devidamente configurado, foram instalados e ajustados os plugins de Caliper Analytics e xAPI no Moodle, garantindo a integração adequada ao sistema e a correta autenticação por meio da inserção das credenciais necessárias.

#### **4.2.2 Processo de Instalação e Configuração**

A instalação e configuração do OpenLRW foram realizadas utilizando uma imagem Docker do software. O serviço foi iniciado com o auxílio de um arquivo `docker-compose.yml`, que configurava dois serviços principais: o banco de dados MongoDB (versão 3.9) e o OpenLRW.

Para enviar os eventos ao serviço, foi necessário obter a chave e o segredo da chave, que seriam utilizados como usuário e senha no Moodle. Infelizmente, o OpenLRW não disponibiliza uma API ou interface gráfica para obter essas informações, sendo necessário conectar-se diretamente ao MongoDB para acessá-las.

No Moodle, foram instalados dois plugins capazes de enviar registros gerados pelo sistema. Um dos plugins envia registros na especificação Caliper Analytics, enquanto o outro os envia na especificação xAPI. Ambos os plugins são de código aberto e foram desenvolvidos especificamente para uso no Moodle.

Para enviar eventos ao serviço, é necessário obter uma chave de acesso e o segredo correspondente, que serão usados como nome de usuário e senha no Moodle, conforme mostrado na Figura 19.

## Caliper log store

Event Store URL logstore\_caliper | endpoint  Default: Empty

API key logstore\_caliper | apikey  Default: Empty

Send statements immediately to event store? logstore\_caliper | immediatemode  Default: No  
 This will force Moodle to send the statements to the Event Store as task. This will make the process closer to real time, but could cause of the Event Store.

Batch size logstore\_caliper | batchsize  Default: 1  
 The maximum number of events to send at at time in batch mode.

Figura 19 – Configurando o Caliper plugin. Fonte: Autor

Infelizmente, o OpenLRW não oferece uma API ou interface gráfica para obter essas credenciais diretamente. Assim, é necessário conectar-se ao banco de dados MongoDB para acessar essas informações.

### 4.2.3 Resultados

Quando os plugins foram ativados, os registros de interação dos usuários foram enviados ao OpenLRW na forma de eventos. De acordo com a documentação do OpenLRW, ele pode receber diferentes especificações (como xAPI e Caliper Analytics), mas armazena todos os eventos exclusivamente no formato Caliper Analytics. Essa decisão foi tomada pelos desenvolvedores do OpenLRW, que consideraram o Caliper Analytics a especificação mais robusta para dados educacionais, uma vez que foi projetada especificamente para lidar com eventos provenientes de sistemas educacionais, enquanto o xAPI era visto como uma especificação mais genérica. Assim, ao enviar um evento para o endpoint xAPI do OpenLRW, uma função de conversão era responsável por traduzir o formato xAPI para o formato Caliper Analytics antes do armazenamento do evento. Nos experimentos conduzidos, verificou-se que o OpenLRW executava a conversão com precisão quando o corpo do evento tinha uma configuração básica, composta por apenas um ator (*actor*), um verbo (*verb*) e um objeto (*object*). Contudo, a presença de elementos adicionais no corpo do evento frequentemente resultava em falhas na operação prevista do conversor.

Este fenômeno tornou-se especialmente evidente nos eventos transmitidos pelos plugins xAPI do Moodle, que frequentemente exibiam estruturas de eventos altamente complexas. Estas estruturas incluíam componentes adicionais que não estavam corretamente mapeados pelo conversor, causando falhas no processo de tradução. Consequentemente, o OpenLRW deixava de armazenar os eventos, resultando na perda

dos dados.

Nos experimentos conduzidos, foi possível detectar várias dificuldades, incluindo o potencial para perda de dados ao empregar o OpenLRW e a falta de suporte sustentado para suas atualizações. Estes obstáculos, junto com a evolução persistente de especificações como o xAPI, podem constituir desafios significativos para a interoperabilidade e a fiabilidade do serviço.

Em 2019, ocorreu a última atualização do OpenLRW, e desde então seus desenvolvedores cessaram o suporte. Consequentemente, novas atualizações das especificações não são mais incorporadas. Como exemplo, a especificação xAPI progrediu até a versão 2.0, homologada pela IEEE como um padrão formal, o que, além das deficiências já citadas, limita ainda mais a aplicação do conversor do OpenLRW.

### 4.3 Experimento II

No segundo experimento, foi avaliada a aptidão para transmitir especificações Caliper rumo ao LRS. Utilizou-se o LRSQL, uma Learning Record Store (LRS) persistente que lida com declarações xAPI e seus anexos, além de documentos relacionados. Diferindo de outras LRSs criadas pela Yet Analytics, o LRSQL é fundamentado em SQL e é compatível com variados sistemas de gerenciamento de bancos de dados (DBMSs), incluindo SQLite e PostgreSQL.

O intuito deste experimento foi examinar a compatibilidade do LRSQL no envio de eventos educacionais fundamentados no Caliper, além de avaliar a criação de um conversor de Caliper para xAPI. Este estudo formou a base para o desenvolvimento do conversor, um componente central na arquitetura proposta recentemente.

O experimento consistiu em quatro fases fundamentais:

Fase 1 - Ambiente Configurado: Na fase inicial, o ambiente do LRSQL foi devidamente configurado, o que envolveu tanto a instalação quanto a preparação para integração com dados provenientes do Caliper e do xAPI.

Fase 2 - Validação Inicial: Uma tentativa foi feita para validar a transmissão de dados das especificações Caliper Analyser e xAPI ao LRS. Contudo, dificuldades técnicas surgiram, resultando em um sucesso parcial nesta etapa, revelando restrições na abordagem utilizada inicialmente.

Fase 3 - Desenvolvimento do Conversor Caliper para xAPI: Como resultado das deficiências detectadas na Fase 2, tornou-se imperativo projetar um conversor para a transmutação de dados Caliper para xAPI, assegurando tanto a compatibilidade quanto a capacidade de transmissão para o LRS.

Fase 4 - Validação Conclusiva: Após a execução do conversor, foram conduzidos testes finais para sancionar a compatibilidade do LRSQL com os dados convertidos, confirmando o correto funcionamento e a interoperabilidade pretendida.

### 4.3.1 Softwares Adicionais Necessários

Para a implementação do LRSQL, tornou-se indispensável a utilização de softwares suplementares, tais como Docker, Docker Compose e PostgreSQL, fundamentais para assegurar um ambiente de execução apropriado e armazenamento eficiente dos dados coligidos. Ademais, no que tange ao desenvolvimento do conversor, foi requisitado um conjunto de softwares adicionais que incluíam Docker, Docker Compose, Nginx e Python.

### 4.3.2 Processo de Instalação, Configuração e Desenvolvimento

A implementação e configuração do LRSQL tiveram como base uma imagem Docker do software. O serviço foi ativado com o uso de um arquivo `docker-compose.yml`, que definia a configuração de dois serviços centrais: o banco de dados PostgreSQL (versão 16) e o LRSQL.

Para transmitir os eventos ao serviço, foi essencial adquirir tanto a chave quanto seu segredo, que funcionariam como credenciais de usuário e senha no Moodle. No ambiente LRSQL, essa ação é facilitada por sua interface gráfica, conforme ilustrado na Figura 20

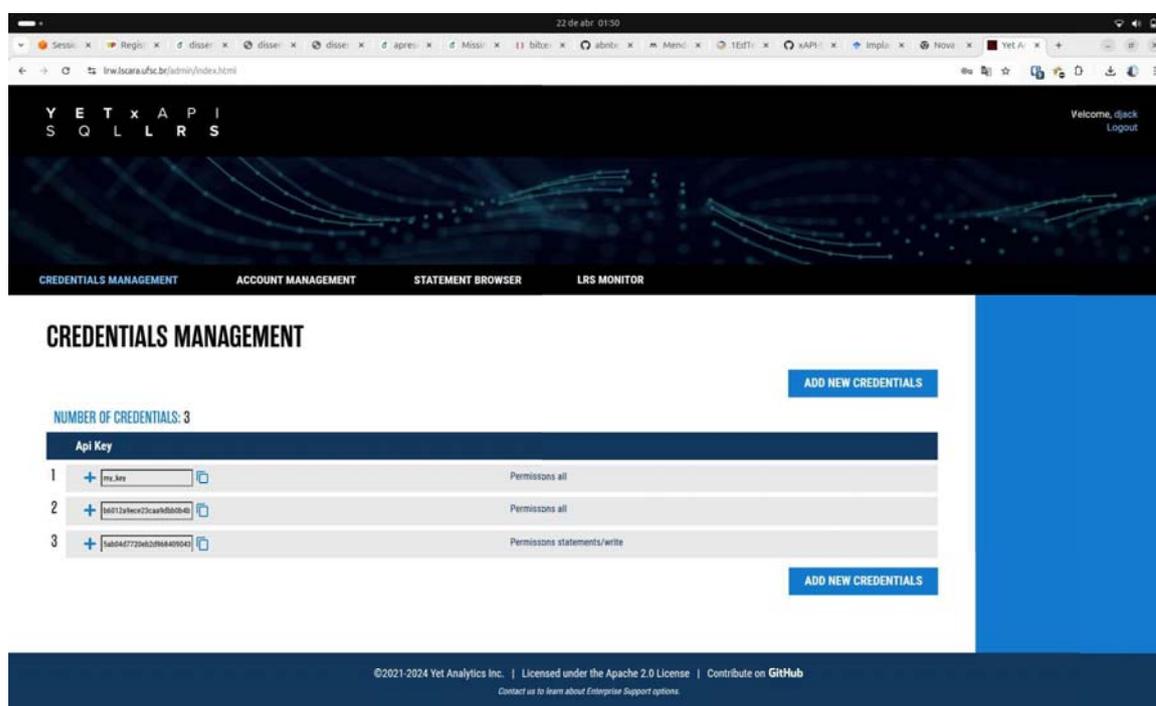


Figura 20 – Credencial no LRSQL. Fonte: Autor.

Após a conclusão das configurações do LRSQL e a execução dos testes relacionados ao envio do caliper para o LRSQL, foi desenvolvido um mecanismo de conversão ou tradução de analisador de Caliper Analytics para xAPI, permitindo assim o armazenamento do payload no LRS.

Caliper Analytics e xAPI possuem uma estrutura conceitual análoga para representar interações de aprendizagem, o que possibilita estabelecer uma equivalência entre esses modelos. Ambos empregam campos principais, como o *'actor'*, que identifica o participante da interação, e incluem referências a uma ação e a um objeto envolvidos no evento. Adicionalmente, ambos permitem extensões por meio de campos opcionais para incluir informações adicionais, como contexto, resultados e metadados associados.

A semelhança estrutural entre as duas especificações permite a elaboração de um conversor entre Caliper Analytics e xAPI. Por exemplo, eventos Caliper, que incluem campos como *'actor'*, *'action'* e *'object'*, podem ser convertidos diretamente em *statements* do xAPI, utilizando campos correspondentes como *'actor'*, *'verb'* e *'object'*. Ainda que variem em terminologia, os campos opcionais podem ser correlacionados de acordo com a semântica dos dados. Essa interoperabilidade facilita a integração entre plataformas com especificações distintas, fomentando um ecossistema mais coeso para a análise de dados educacionais. Tabela 6 proporciona uma análise comparativa dos campos que podem ser correspondentes entre as especificações.

Tabela 6 – Comparação de estrutura entre Caliper Event e xAPI Statement

Caliper Event	xAPI Statement
<b>Campos principais</b>	
actor, action, object, eventTime, id	actor, verb, object, id, stored
<b>Campos opcionais</b>	
target, generated, referrer, edApp, group, membership, session, federatedSession, extensions	timestamp, context, result, authority, attachments, version

O conversor foi concebido para preservar a totalidade do *payload* original do Caliper no campo *extensions* do xAPI, prevenindo a perda de dados durante a conversão. Assim, mesmo os campos que não encontram um mapeamento direto na estrutura xAPI são mantidos, assegurando a integridade e detalhamento dos dados originais. Dessa forma, plataformas que operam com especificações diversas podem ser integradas, promovendo um ambiente de análise de dados educacionais mais unificado e interoperável.

O desenvolvimento do conversor de Caliper Analytics para xAPI foi realizado sob a linguagem de programação Python, fundamentando-se na figura 21.

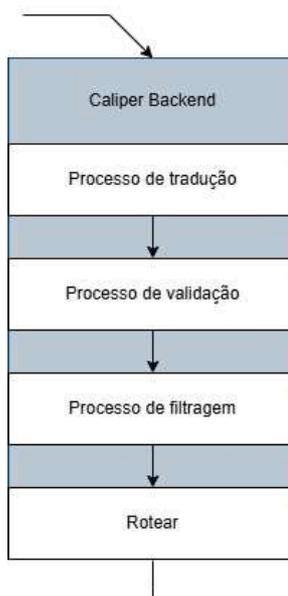


Figura 21 – Fluxograma do Conversor. Fonte: Autor.

Após a criação e a implementação do conversor, conduziram-se testes preliminares para verificar a integração do LRSQL com os dados transformados, garantindo tanto o funcionamento correto quanto a desejada interoperabilidade, cujos resultados são descritos a seguir.

### 4.3.3 Resultados

Após a configuração adequada do sistema, foram feitas tentativas de validar o envio de dados segundo as especificações do Caliper Analytics e xAPI para o LRS. No entanto, devido ao processo de validação realizado por todo LRS para confirmar a compatibilidade do payload com xAPI, o envio de informações no formato Caliper Analytics para o LRS não é viável. Isso ocorre mesmo com a documentação do Caliper Analytics mencionando a possibilidade de integração com o LRS, prescindindo, entretanto, do desenvolvimento de um conversor. A figura 22 ilustra os erros gerados durante o envio do payload do Caliper para o LRSQL sem a presença de um conversor entre as duas especificações.

```

djack@djack-Vostro-15-3510: ~/Documents/Universidade/experimentos/lrw/data$ curl -X POST "http://localhost:8080/xapi/statements" -u "061e4686e2b12fb0350f98051c5d2951a5418b52f3c50551d9f32418a31eb41d:c86442132745305d59b84aeb62783917359787098b1396449a8452da95c47543" -H "Content-Type: application/json; charset=utf-8" -H "X-Experience-API-Version: 1.0.3" -d @caliper.json && echo
{"error":{"message":"Invalid Statement Data","statement-data":{"@context":"http://purl.imsglobal.org/ctx/caliper/v1/Context","@type":"SessionEvent","actor":{"@id":"http://example.edu/user/554433","@type":"Person"},"action":"LoggedIn","object":{"@id":"http://example.edu/system","@type":"SoftwareApplication"},"generated":{"@type":"Session","startedAtTime":"2024-07-31T12:34:56Z"},"spec-error":{"agent/@id \"http://example.edu/user/554433\", :agent/@type \"Person\"} - failed: (or (contains? % :agent/mbox) (contains? % :agent/mbox_sha1sum) (contains? % :agent/openid) (contains? % :agent/account)) in: [:statement/actor] at: [:single :statement/actor :actor/agent] spec: :xapi-schema.spec/agent\n(:activity/@id \"http://example.edu/system\", :activity/@type \"SoftwareApplication\") - failed: (contains? % :activity/id) in: [:statement/object] at: [:single :statement/object :statement/object :activity] spec: :xapi-schema.spec/activity\n[\"@context\" \"http://purl.imsglobal.org/ctx/caliper/v1/Context\"] - failed: (conformer (partial conform-ns-map \"statement\") uniform-ns-map) in: [0] at: [:multiple] spec: :xapi-schema.spec/statement\n[\"@type\" \"SessionEvent\"] - failed: (conformer (partial conform-ns-map \"statement\") uniform-ns-map) in: [1] at: [:multiple] spec: :xapi-schema.spec/statement\n[\"actor\" [\"@id\" \"http://example.edu/user/554433\", \"@type\" \"Person\"]] - failed: (conformer (partial conform-ns-map \"statement\") uniform-ns-map) in: [2] at: [:multiple] spec: :xapi-schema.spec/statement\n[\"action\" \"LoggedIn\"] - failed: (conformer (partial conform-ns-map \"statement\") uniform-ns-map) in: [3] at: [:multiple] spec: :xapi-schema.spec/statement\n[\"object\" {\"@id\" \"http://example.edu/system\", \"@type\" \"SoftwareApplication\"}] - failed: (conformer (partial conform-ns-map \"statement\") uniform-ns-map) in: [4] at: [:multiple] spec: :xapi-schema.spec/statement\n[\"generated\" {\"@type\" \"Session\", \"startedAtTime\" \"2024-07-31T12:34:56Z\"}] - failed: (conformer (partial conform-ns-map \"statement\") uniform-ns-map) in: [5] at: [:multiple] spec: :xapi-schema
0] 0: [tmux]*Z

```

Figura 22 – Envio do Caliper payload para o LRSQL. Fonte: Autor.

Através da ocorrência contínua de erros, o conversor foi sendo validado e aprimorado com base na documentação do Caliper Analytics e do xAPI, a fim de identificar as compatibilidades existentes entre eles.

As Figuras 23) e 24) ilustram exemplos de JSON payloads em Caliper Analytics e sua posterior conversão para o formato xAPI. O primeiro representa o formato Caliper Analytics original, enquanto o segundo demonstra a conversão para xAPI, agora compatível com LRSQL.

```
"@context": "http://purl.imsglobal.org/ctx/caliper/v1/Context",
"@type": "SessionEvent",
"actor":
"@id": "http://example.edu/user/554433",
"@type": "Person"
,
"action": "LoggedIn",
"object":
"@id": "http://example.edu/system",
"@type": "SoftwareApplication"
,
"generated":
"@type": "Session",
"startedAtTime": "2024-07-31T12:34:56Z"
```

Figura 23 – Payloads Original Caliper . Fonte: Autor

```

"actor":
  "objectType": "Agent",
  "account":
    "homePage": "http://example.edu",
    "name": "554433"
  ,
  "verb":
    "id": "http://adlnet.gov/expapi/verbs/loggedin",
    "display":
      "en-US": "LoggedIn"
    ,
    "object":
      "id": "http://example.edu/system",
      "definition":
        "type": "http://adlnet.gov/expapi/activities/software-application",
        "name":
          "en-US": "SoftwareApplication"
        ,
        "timestamp": "2024-07-31T12:34:56Z",
        "context":
          "extensions":
            "http://example.edu/extensions/caliperEvent":
              "eyJAY29udGV4dCI6ICJodHRwOi8vcHVybC5pbXNnbG9iYWwub3JnL2N0eC9jYWxp
              cGVyL3YxL0NvbnRleHQiLCAiQHR5cGUiOiAiU2Vzc2lrbkV2ZW50IiwgImFjdG9yIj
              ogeyJAaWQiOiAiaHR0cDovL2V4YW1wbGUuZWR1L3VzZXIvNTUONDMzIiwgIkB0eXB
              lIjogIlBlcnNvbiJ9LCAiYWN0aW9u
              IjogIkxvZ2dlZEluIiwgIm9iamVjdCI6IHsiQGlkIjogImh0dHA6Ly9leGFtcGxlLmVkdS
              9zeXNOZW0iLCAiQHR5cGUiOiAiU29mdHdhcmVBChBSaW9lbnRlbiJ9LCAiZ2VuZXJh
              dGVkIjogeyJAdHlwZSI6ICJlZ2VzZXNzaW9uIiwgInNOYXJ0ZWRBdFRpbWUiOiAiMjAyNC0
              wNyOzMVQzMjozND01NloifX0="

```

Figura 24 – Payloads Convertido Caliper . Fonte: Autor

A Figura 25 ilustra o procedimento para transmissão de um payload formatado em Caliper Analytics ao LRS após conversão. Inicialmente, o payload no formato Caliper Analytics é submetido ao conversor, que o adapta ao formato xAPI correspondente. Posteriormente, o payload já convertido é transmitido ao LRSQL, que então fornece um ID, confirmando que a operação foi executada com sucesso.

```

djack@djack-Vostro-15-3510:~/Documents/Universidade/experimentos/lrw/data$ python3 converter.py > caliper-to-xapi.json
djack@djack-Vostro-15-3510:~/Documents/Universidade/experimentos/lrw/data$ curl -X POST "http://localhost:8080/xapi/statements" -u "061e4686e2b12fb0350f98051c5d2951a5418b52f3c50551d9f32418a31eb41d:c86442132745305d59b84aeb62703917359707090b1396449a8452da95c47543" -H "Content-Type: application/json;charset=utf-8" -H "X-Experience-API-Version: 1.0.3" -d @caliper-to-xapi.json && echo
["7a34a631-b15b-4b67-90da-c2abeacb8b79"]
djack@djack-Vostro-15-3510:~/Documents/Universidade/experimentos/lrw/data$ █

[0] 0:bash*Z "djack-Vostro-15-3510" 08:51 31-Jul-24

```

Figura 25 – Configurando o Caliper plugin. Fonte: Autor

Como mostrado na Figura 26, os dados enviados ao LRS são consultados. Com a utilização do LRSQL, é possível consultar diretamente o banco de dados PostgreSQL por meio de uma instrução SQL, possibilitando a construção de queries avançadas de análise de aprendizado.

```

novobanco=# select payload#>>'[actor,name]' actor_name, COUNT(DISTINCT id) FROM xapi_statement GROUP BY actor_name;

```

actor_name	count
Example Admin	1
Example Learner	1
	6

```

(3 rows)
novobanco=# █

```

Figura 26 – Consulta ao banco pelo payload xAPI. Fonte: Autor

A validação do experimento confirmou o desenvolvimento de um conversor de Caliper Analytics para xAPI, o qual desempenhou um papel crucial na arquitetura final descrita neste a seguir.

## 5 ARQUITETURA PROPOSTA

Neste capítulo, é apresentada uma descrição da arquitetura proposta, que já foi apoiada no capítulo anterior pelos experimentos e seus resultados. Em seguida, há uma breve descrição da arquitetura em termos dos principais componentes, seguida de uma descrição das *queries* implementadas e uma descrição dos eventos e dados coletados.

### 5.1 Descrição da arquitetura

A arquitetura proposta é estruturada para atender às necessidades de integração entre os padrões xAPI e Caliper Analytics, garantindo a interoperabilidade e processamento eficiente dos *statements* de aprendizagem. A solução está organizada em múltiplos componentes, cada um desempenhando uma função específica para o fluxo de dados, desde a recepção até o armazenamento. Ao juntar todos esses componentes, obtemos o **Learning Record Warehouse (LRW)**, uma solução integrada e robusta para análise e armazenamento de dados de aprendizagem. Devido à sua capacidade de expansão para receber diversas especificações, a arquitetura pode ser utilizada no futuro para adição de novos conversores, sem alterar o esqueleto base, proporcionando escalabilidade e expansibilidade à proposta.

As especificações xAPI (Experience API) e Caliper Analytics foram escolhidas por permitirem a representação padronizada das ações dos usuários em diversos contextos de aprendizagem, promovendo a interoperabilidade semântica e o armazenamento estruturado dos dados. A adoção de um Learning Record Store (LRS) como base do Learning Record Warehouse (LRW) foi motivada, principalmente, pelo fato de o xAPI ter se tornado um padrão da IEEE em 2023. Esse reconhecimento implica que a especificação passou por um processo rigoroso de revisão técnica, garantindo sua solidez, qualidade e maturidade. A IEEE oferece uma estrutura de governança clara e estável, o que assegura a manutenção contínua, atualizações planejadas e a longevidade do padrão. Em contraste, especificações desenvolvidas por terceiros podem depender de iniciativas isoladas ou de grupos informais, o que representa um risco maior de

descontinuidade. Outro critério importante para a escolha do LRS foi sua compatibilidade com bancos de dados relacionais, como o PostgreSQL. Essa compatibilidade facilita consultas complexas via SQL e possibilita a adoção de soluções robustas e escaláveis para o armazenamento de dados. Em síntese, a seleção dessas tecnologias foi orientada pela necessidade de enfrentar os desafios de integração e interoperabilidade de dados educacionais heterogêneos, com o objetivo de estabelecer uma base sólida para a aplicação de Learning Analytics e a melhoria contínua dos processos de aprendizagem.

### 5.1.1 Componentes Principais

A arquitetura apresentada é composta por três componentes principais que operam de forma integrada para garantir o fluxo eficiente e o processamento adequado dos dados: o API Gateway, o Conversor Caliper-xAPI e o LRS .

O Gateway de API atua como intermediário na comunicação entre aplicativos dos clientes e os serviços internos, disponibilizando pontos de extremidade para a recepção de dados. A estrutura exibida na Figura 17 administra dois tipos principais de *endpoints*: */xAPI/statements*, para dados compatíveis com o padrão xAPI, e */caliper/statements*, destinados a dados no formato Caliper Analytics. Estes endpoints encaminham as informações para processamento posterior, e, no futuro, a arquitetura pode incluir outros *endpoints* à medida que novos conversores forem testados e integrados.

O componente Conversor Caliper-xAPI serve como um intermediário, realizando a conversão de *statements* do formato Caliper Analytics para o padrão xAPI. Essa conversão é crucial para garantir a interoperabilidade entre sistemas que empregam padrões distintos, possibilitando que dados estruturados inicialmente em Caliper Analytics sejam integrados e utilizados por sistemas que operam no padrão xAPI. Este componente manipula os dados recebidos via *endpoint* Caliper Analytics, ajustando-os para se adequarem ao formato xAPI exigido nas etapas subsequentes do processo.

Já o LRS é o repositório central de armazenamento dos dados no formato xAPI. Ele armazena tanto os dados enviados diretamente ao *endpoint* */xAPI/statements* quanto os processados pelo Caliper Analytics to xAPI, garantindo a persistência e acessibilidade das informações para futuras análises e consultas.

Na arquitetura examinada, o caminho dos dados obedece a uma ordem claramente estabelecida. Inicialmente, os dados são recebidos no API Gateway, nos *endpoints* */xAPI/statements* ou */caliper/statements*. Seguindo esse recebimento, o módulo Caliper-xAPI converte os dados do formato Caliper para o formato xAPI. Finalmente, os dados transformados são direcionados ao LRS, onde são guardados de maneira segura e de fácil acesso, permitindo sua posterior recuperação para análises. Essa configuração foi desenvolvida para facilitar a interconexão entre padrões diversos, garantindo um fluxo de dados eficaz e promovendo a interoperabilidade, além de permitir

a realização de análises baseadas em dados armazenados de forma uniforme.

A decisão acerca do LRS é flexível durante a implementação arquitetural, no entanto, para este estudo, optou-se por um LRS compatível com o PostgreSQL. Este é um dos sistemas de gerenciamento de banco de dados mais robustos e amplamente aceitos, uma vez que a familiaridade com o padrão SQL facilita a elaboração de consultas e análises complexas. Esta estratégia assegura a continuidade e a eficiência no fluxo de dados, desde a recepção até o armazenamento, promovendo a interoperabilidade e proporcionando suporte à evolução contínua do sistema.

O LRW surge como o produto final dessa integração, fornecendo uma solução centralizada para o gerenciamento e análise de dados de aprendizagem. A estrutura proposta também é expansível, permitindo que novos conversores sejam adicionados no futuro sem alterações na base existente.

## 5.2 Implementação do LRS

Durante a implementação da arquitetura, optou-se por utilizar o LRSQL como base de armazenamento, devido à sua compatibilidade com o PostgreSQL. Esse banco de dados destaca-se pela aderência aos padrões SQL e pela extensibilidade, incluindo suporte avançado aos tipos de dados 'JSON' e 'JSONB'. Em especial, o 'JSONB' permite armazenar documentos JSON em uma estrutura binária otimizada, combinando a flexibilidade do JSON com a eficiência de um banco relacional. O PostgreSQL oferece ainda um ambiente familiar para a criação de consultas em SQL, o que reduz a curva de aprendizado para desenvolvedores e analistas de dados, facilitando o uso imediato da ferramenta. Além disso, conta com operadores e índices avançados, como GIN e B-tree, que possibilitam consultas rápidas e análises eficientes, mesmo em cenários complexos, garantindo maior precisão e flexibilidade na manipulação de dados JSON.

A implementação das queries no PostgreSQL aproveita a flexibilidade e eficiência proporcionadas pelo suporte para dados estruturados e semi-estruturados. Com o uso do tipo de dado `jsonb`, é possível realizar consultas complexas em dados semi-estruturados, como filtrar documentos, verificar valores específicos e manipular chaves e atributos. Operadores como `->`, `->`, e `@>` permitem um acesso intuitivo aos dados armazenados.

Por exemplo, uma consulta pode ser realizada para isolar *statements* que possuam um campo específico no formato `jsonb`; a query adequada para tal seria:

```
SELECT * FROM statements
WHERE payload @> '{"actor": {"name": "John Doe"}}';
```

A presente abordagem oferece flexibilidade para manipular diversas estruturas de dados sem demandar a redefinição de esquemas, o que é uma vantagem crucial em

aplicações como xAPI e Caliper, onde a diversidade de dados é frequente. Ademais, o PostgreSQL permite integrar consultas estruturadas com semi-estruturadas, possibilitando análises mais complexas utilizando comandos SQL convencionais, como JOIN e agregações, juntamente com dados armazenados em JSONB. Um estudo aprofundado dos dados Caliper é apresentado no Apêndice 31, oferecendo análises sofisticadas.

O Conversor proporciona um benefício adicional notável: o *payload* original do Caliper é preservado no campo de extensão do xAPI em formato Base64, mantendo a integridade das informações originais. Este método assegura que o *payload* possa ser rapidamente restaurado quando necessário, pois o PostgreSQL tem suporte nativo para a decodificação de Base64. Após a decodificação, os dados podem ser convertidos diretamente para JSON, permitindo a recuperação da estrutura original e simplificando o uso para análises detalhadas.

Esta função acentua a aptidão do PostgreSQL em manejar cenários de dados diversos e intrincados, assegurando simultaneamente a eficiência no armazenamento e na recuperação de dados essenciais.

### 5.3 Eventos e dados Reais coletados

Para avaliar a eficácia da arquitetura, um terceiro teste foi conduzido empregando o Moodle como meio de transmissão de logs nos formatos xAPI e Caliper. Para isso, dois plugins de código aberto, criados exclusivamente para essa plataforma, foram instalados no Moodle, com o objetivo de criar e transmitir dados de registro. Um plugin é baseado na especificação Caliper Analytics, enquanto o outro segue a especificação xAPI.

Em pormenor, os dados foram obtidos através da plataforma Moodle, acessível em . Nela, foram configurados os plugins moodle-logstore<sup>1</sup> e Caliper-logStore<sup>2</sup>. Tais dados foram então transmitidos para o LRW, com hospedagem em . Subsequentemente, a análise e a visualização dos dados foram realizadas na interface do Apache Superset, instalada em . Assim, certos registros de atividades conduzidas pelos usuários na plataforma são transferidos do Moodle para o LRSQL, permitindo, através do Superset, a criação de visualizações interativas referentes às atividades dos alunos, incluindo frequência, tempo de participação, engajamento com atividades online e desempenho em avaliações.

A subsequente seção de resultados elucida a aplicação da arquitetura proposta, oferecendo uma descrição detalhada do que foi previamente mencionado.

---

<sup>1</sup>[https://moodle.org/plugins/logstore\\_xapi](https://moodle.org/plugins/logstore_xapi)

<sup>2</sup>[https://moodle.org/plugins/logstore\\_caliper](https://moodle.org/plugins/logstore_caliper)

## 6 RESULTADOS E DISCUSSÃO

Neste capítulo, analisamos os resultados decorrentes da implementação da arquitetura em um contexto prático, com foco nos processos de coleta, armazenamento e recuperação de dados. O capítulo é subdividido em duas seções principais: uma dedicada aos resultados observados e outra à sua análise.

### 6.1 Resultados

Esta seção discute a habilidade de coletar, armazenar e recuperar dados.

#### 6.1.1 Capacidade de coleta, armazenamento e recuperação

Para testar a integração do Learning Record Warehouse (LRW) com um sistema de Análise de Aprendizagem (LA), utilizou-se um ambiente educacional que incluía dois cursos ativos, totalizando 57 alunos que participavam de atividades na plataforma. Para a coleta de dados, foram implementados plugins compatíveis com a especificação xAPI e o Caliper Analyzer, possibilitando rastrear as interações dos estudantes.

Foram adequadamente configuradas as estruturas essenciais para hospedar o LRW no servidor, abrangendo serviços críticos como LRS (LRSQL), o conversor xAPI-Caliper, o banco de dados PostgreSQL, além do NGINX, que serve como API Gateway. Os plugins que são compatíveis com a versão instalada do Moodle foram descarregados e configurados usando as chaves de autenticação correspondentes para o envio de dados. O Docker Compose foi empregado para orquestrar toda a configuração, e detalhes completos dos arquivos de configuração necessários para iniciar cada serviço podem ser encontrados no Apêndice A.1.

Os plugins no Moodle não coletam todos os eventos gerados, pois cada um é projetado para lidar apenas com um conjunto específico de eventos. A versão do plugin `logstore_xapi` instalada era capaz de registrar 62 tipos de eventos, abrangendo atividades desde o login até a entrega de avaliações. Em contraste, o plugin `Caliper-logStore` capturava apenas 30 eventos. A principal divergência na coleta de dados por esses plugins ocorre nos eventos do `mod_bigbluebuttonbn`, que não são suporta-

dos pelo Caliper-logStore, bem como em eventos associados a módulos como *core*, *mod\_book*, *mod\_facetoface*, *mod\_forum*, *mod\_quiz*, *mod\_scorm* e *totara\_program*.

Tabela 7 – Resumo dos eventos suportados pelos plugins Caliper e xAPI. Para a lista completa, consultar o Apêndice C

<b>Evento</b>	<b>Plugin Caliper</b>	<b>Plugin xAPI</b>
\core\event\user_loggedin	✓	✓
\core\event\user_loggedout	✓	✓
\core\event\course_viewed	✓	✓
\mod_forum\event\course_module_viewed	✓	✓
\mod_quiz\event\attempt_started	✓	✓
\core\event\course_completed	X	✓
\core\event\user_enrolment_created	X	✓
\mod_bigbluebuttonbn\event\meeting_created	X	✓
\mod_scorm\event\sco_launched	X	✓
\totara_program\event\program_assigned	X	✓

O plugin *logstore\_xapi* proporciona um benefício notável ao possibilitar a escolha dos eventos a serem transmitidos; em contraste, o plugin Caliper transmite todos os eventos sem opções configuráveis. Ademais, o xAPI permite a inspeção de logs com falhas e o acesso a um registro histórico de logs transmitidos, conforme exemplificado na Figura 27. É importante também observar que o plugin xAPI recebe novos recursos de maneira mais frequente.

The screenshot shows the Moodle administration interface for the Logstore xAPI plugin. The page is titled "Logstore xAPI" and contains the following configuration options:

- Your LRS endpoint for the xAPI:** `https://localhost:8080/lrs/xapi` (highlighted with a blue box). Default: `http://example.com/endpoint/location/`
- Username:** `xxxxxx`. Default: `username`
- Your LRS basic auth secret/password for the xAPI:** `xxxxxxx`. Default: `password`
- Send statements by scheduled task?** . Default: `Sim`. Description: "This will force Moodle to send the statements to the LRS in the background, via a cron task to avoid blocking page responses. This will make the process less close to real time, but will help to prevent unpredictable Moodle performance linked to the performance of the LRS."
- Maximum batch size:** `30`. Default: `30`. Description: "Statements are sent to the LRS in batches. This setting controls the maximum number of statements that will be sent in a single operation. Setting this to zero will cause all available statements to be sent at once, although this is not recommended."
- Maximum batch size for failed requests:** `15`. Default: `15`. Description: "Statements are sent to the LRS in batches. This setting controls the maximum number of statements that will be sent in a single operation for failed requests. Setting this to zero will cause all available statements to be sent at once, although this is not recommended."
- Maximum batch size for historical requests:** `30`. Default: `30`. Description: "Statements are sent to the LRS in batches. This setting controls the maximum number of statements that will be sent in a single operation for historical requests. Setting this to zero will cause all available statements to be sent at once, although this is not recommended."
- Resend failed batches:** . Default: `Não`. Description: "When processing events in batches, try re-sending events in smaller batches if a batch fails. If not selected, the whole batch will not be sent in the event of a failed event."
- Identify users by email:** . Default: `Não`. Description: "Statements will identify users with their email (mbox) when this box is ticked."
- Send short course name:** . Default: `Não`. Description: "Statements will contain the shortname for a course as a short course id extension"

Figura 27 – Configuração plugin xAPI no Moodle. Fonte: Autor

A autenticação dos plugins adere a diferentes normas. O plugin xAPI requer a autenticação no LRS usando credenciais de nome de usuário e senha, em conformidade com o padrão xAPI. O plugin Caliper, em contrapartida, utiliza uma chave secreta. A autenticação no Irsql é realizada através do conversor Caliper-xAPI, criado no contexto deste estudo. Este conversor valida a chave no LRS e recupera a senha associada para transmitir os dados, eventualmente.

Uma vez configurados os plugins, ambos iniciaram a transmissão de dados para o LRW, possibilitando a análise dos dados coletados.

**DATA BROWSER**

Credentials to Use: Credential to Browse [Refresh](#)

Current Query: /xapi/statements?limit=10

#	Statement ID	Actor	Verb	Object	Timestamp
1	8319f596-8ab0-4c61-841d-a99fab0b148	Juany Rocha	viewed	YouTube - Video	17/03/2025, 14:55:41
2	158800f2-662b-4b23-ae2b-9179fdab8c33	Juany Rocha	Viewed	YouTube - Video	17/03/2025, 14:55:41
3	5096bacf23c4ba1-913f-fb0b7125c423	Juany Rocha	viewed	Plano de Ensino - 2021.1	26/02/2025, 09:51:09
4	aa2fc623-086c-4be1-93a4-b688439548ff	Juany Rocha	Viewed	Plano de Ensino - 2021.1	26/02/2025, 09:51:09
5	8730f805-a154-4aa4-b685-fac695151b5	Juany Rocha	viewed	Demo - Estrutura de Dados e Algoritmos (2021)	26/02/2025, 09:50:58
6	8730f805-a154-4aa4-b685-fac695151b4	Juany Rocha	viewed	Demo - Estrutura de Dados e Algoritmos (2021)	26/02/2025, 09:50:42
7	d7668e4f-9d07-49bd-af04-ac2d3f971f9	Juany Rocha	NavigatedTo	Demo - Estrutura de Dados e Algoritmos (2021)	26/02/2025, 09:50:58
8	c7474b67-fbe5-4d37-bad7-af47f63e6ed	Juany Rocha	NavigatedTo	Demo - Estrutura de Dados e Algoritmos (2021)	26/02/2025, 09:50:42
9	8e32ee26-c90e-4d17-a609-49880847888f	Juany Rocha	logged into	Moodle Analytics Dashboard 2	26/02/2025, 09:48:55
10	a1ea16be-91b8-4474-a1af-9451c1f647aa	Juany Rocha	LoggedIn	Moodle Analytics Dashboard 2	26/02/2025, 09:48:55

Figura 28 – Dados coletados do Moodle armazenados no LRW. Fonte: Autor

Para a visualização dos dados obtidos, foi efetuada a integração da LRW com a ferramenta de visualização Apache Superset. A configuração do Superset foi implementada via Docker Compose, e o documento de configuração pode ser encontrado no Apêndice A.2.

**Superset** Dashboards Charts Datasets SQL

Query: `SELECT payload -> 'object' -> 'definition' -> 'name' -> 'en-US' AS objeto_nome, -- Obtém o nome do objeto  
COUNT(*) AS total_viewed, -- Conta quantas vezes o objeto foi visualizado (convert...  
(convert_from(decode(payload -> 'context' -> 'extensions' -> 'http://example.edu/extensions/caliperEvent', 'base64'), 'utf-8'))::json -> 'membership' -> 'organization' AS organization`

objeto_nome	total_viewed	organization
Trabalho A2 - Entrega em 13.12.2024	107	https://mad2.cne.ufsc.br/course/23
Trabalho A1 - Entrega em 22.10.2024	48	https://mad2.cne.ufsc.br/course/23
2024.2 - CI77159 - Programacao em Computadores	45	https://mad2.cne.ufsc.br/course/23
Slides Polimorfismo	36	https://mad2.cne.ufsc.br/course/23
Slides ArrayList	32	https://mad2.cne.ufsc.br/course/23
Avísls	29	https://mad2.cne.ufsc.br/course/23
Aula 13 - Arquivos - Slides	22	https://mad2.cne.ufsc.br/course/23
Slides Agregação	18	https://mad2.cne.ufsc.br/course/23

Figura 29 – Consulta ao banco utilizando o Apache Superset. Fonte: Autor

Além de realizar consultas, o Superset facilita a transformação instantânea dos

dados em representações gráficas através de sua interface interativa. Uma vez criados os gráficos, eles puderam ser organizados em painéis personalizados.

Na Figura 30, um dashboard utilizando Apache Superset, denominado "Mad2", foi apresentado em um formato rascunho para fins ilustrativos. O objetivo era mostrar exemplos de uso dos dados reunidos a partir de eventos de aprendizado extraídos do Learning Record Store (LRS). A visualização estava estruturada em cinco componentes principais. No canto superior esquerdo, havia uma nuvem de palavras representando os identificadores dos estudantes que realizaram login. O tamanho das palavras indicava a frequência de acesso. No centro superior, havia uma tabela que apresentava os objetos de aprendizagem mais visualizados, junto com suas respectivas quantidades de visualizações. A maioria dos itens consistia em tarefas e descrições de trabalhos, com destaque para o "Trabalho A2 – Entrega em 13.12.2024", que teve 107 visualizações. À direita, um calendário em formato de heatmap foi gerado, demonstrando a distribuição temporal dos eventos coletados, com uma concentração mais elevada registrada nos meses de fevereiro e março. No canto inferior esquerdo, um gráfico de rosca mostrou a distribuição dos dados por curso, destacando dois cursos principais: CIT7139-02652 (20242) – Programação em Computadores e CIT7584-03652 (20242). Este painel ilustrou como foi possível obter uma visão geral sobre o engajamento dos estudantes usando dados coletados via Caliper Analytics e armazenados em um LRS baseado em SQL (Irsql). Os gráficos permitiram identificar padrões de acesso (logins), interação com conteúdos (visualizações), variações temporais (eventos por dia) e distribuição por curso.

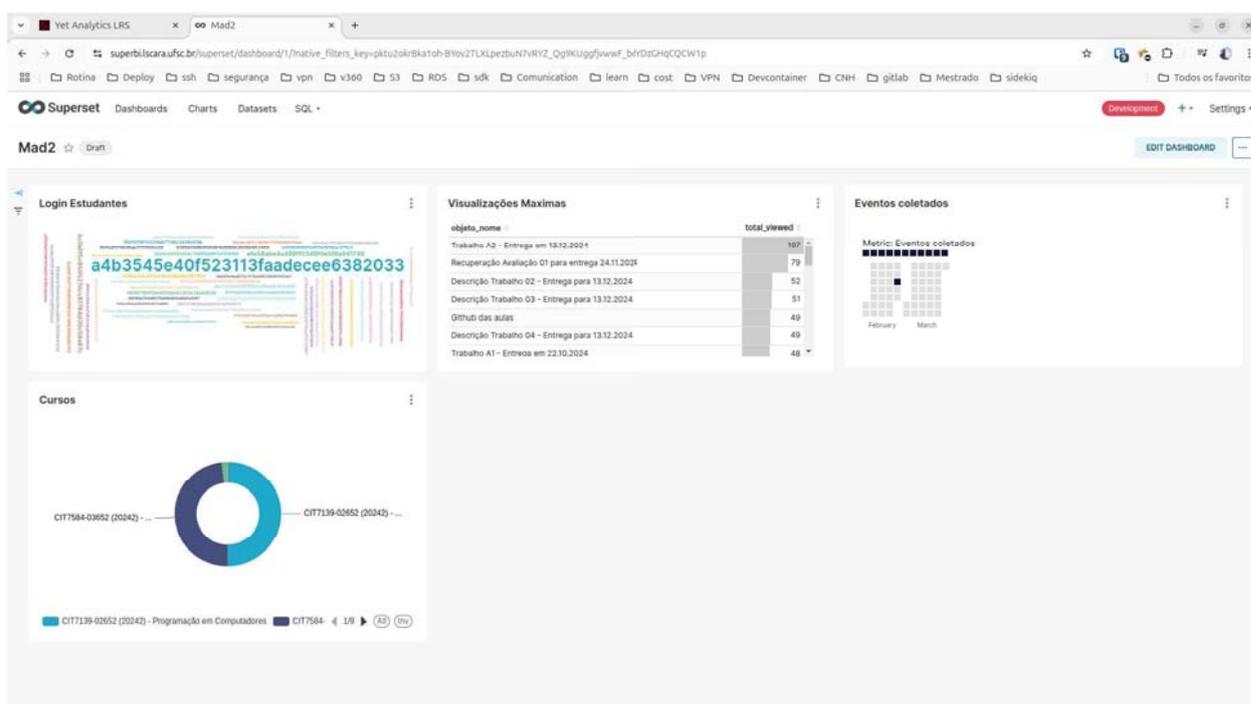


Figura 30 – Dashboard criado a partir dos dados armazenados no LRW. Fonte: Autor

## 6.2 Discussão

Integrar e configurar os serviços facilitou a captura, armazenamento e análise de eventos produzidos no Moodle por meio do Learning Record Warehouse (LRW). A análise comparativa entre os plugins xAPI e Caliper revelou variações consideráveis na quantidade e diversidade de eventos recolhidos. Tal distinção afeta diretamente a abrangência e a qualidade das análises subsequentes, uma vez que diferentes especificações proporcionam perspectivas variadas sobre as interações dos estudantes na plataforma.

O emprego do Apache Superset permitiu a criação de visualizações dinâmicas dos dados coletados no LRW, estabelecendo uma base sólida para futuras pesquisas sobre o comportamento discente.

O *payload* inicial é armazenado no campo "context" -> 'extensions' -> 'http://example.edu/extensions/caliperEvent', e está codificado em base64. A utilidade do PostgreSQL em suportar a decodificação nativa de base64 permitiu-nos transformar o arquivo original para o formato JSON, facilitando a extração das informações exigidas. As informações de interesse eram encontradas no campo "membership" -> 'roles', que enumera os papéis associados ao usuário no Moodle. Utilizando essas informações, filtramos os usuários que tinham a função de estudante, garantindo que o gráfico de logins representasse exclusivamente o padrão de comportamento deste grupo.

Por fim, logramos desenvolver um painel detalhado com os dados obtidos. A aplicação de consultas SQL sobre informações contidas em arquivos JSON, combinada com a extração de dados adicionais do payload original do Caliper Analytics, codificado em base64, facilitou a criação de um gráfico preciso dos logins dos estudantes. Este método sublinha a relevância da flexibilidade na manipulação de dados em Learning Analytics, possibilitando a integração e análise conjunta de diferentes fontes e formatos.

As especificações Caliper e xAPI apresentam discrepâncias marcantes em relação à modelagem de eventos e à granularidade das informações capturadas. A interoperabilidade entre esses formatos também é um desafio relevante, já que muitas ferramentas de Learning Record Store (LRS) carecem de suporte nativo para o Caliper Analytics e outras especificações alternativas. Nesse cenário, o uso de um conversor Caliper-xAPI revelou-se uma opção prática, viabilizando o armazenamento desses dados no LRS e posterior extração para análise. Este procedimento ilustra que a convergência entre diferentes especificações pode ser alcançada por meio de soluções adaptáveis, otimizando o uso de dados educacionais para decisões fundamentadas em evidências.

Nossa pergunta de pesquisa "Como um Learning Record Warehouse pode ser in-

tegrado a um sistema de Learning Analytics de modo a suportar o recebimento de dados provenientes de diferentes Ambientes Virtuais de Aprendizagem e diversas outras fontes?” pode ser respondida da seguinte forma: este trabalho demonstra a viabilidade dessa integração por meio da adoção de padrões consolidados de interoperabilidade. A solução proposta fundamenta-se na implementação de um Learning Record Store (LRS) como repositório central, utilizando a especificação xAPI como formato padrão de armazenamento. Para viabilizar a incorporação de dados provenientes de outras especificações, como o Caliper Analytics, foi desenvolvido um conversor Caliper-xAPI, responsável por transformar e padronizar os dados para inserção no LRS. Esse conversor foi projetado com uma arquitetura modular e expansível, o que permite sua extensão para futuras especificações, ampliando a aplicabilidade da solução em cenários educacionais com múltiplas fontes de dados e especificações. Adotou-se o PostgreSQL como sistema de gerenciamento de banco de dados, permitindo o armazenamento não estruturado em formato JSON, mas utilizando consultas dinâmicas e flexíveis em SQL.

## 7 CONSIDERAÇÃO FINAIS

O presente capítulo finaliza o estudo oferecendo um resumo das principais conclusões alcançadas ao longo da pesquisa. Além disso, propõe algumas recomendações e sugere direções de investigação que continuam abertas para futuras pesquisas nesta área.

### 7.1 Conclusões

A proposta apresentada nesta dissertação corresponde à integração de dados educacionais usando um Learning Record Warehouse (LRW) como base para o Learning Analytics (LA). O principal objetivo foi desenvolver uma solução que coletasse e armazenasse dados de aprendizagem de forma padronizada, usando especificações de interoperabilidade para melhorar a compatibilidade entre sistemas de aprendizagem digital e facilitar a análise de dados. A pesquisa é justificada pela crescente adoção do e-learning e pela necessidade de rastrear e analisar dados educacionais.

Foi explorado o uso da xAPI e do Caliper Analytics, duas especificações importantes para o rastreamento de atividades de aprendizagem. Enquanto a xAPI permite o rastreamento de qualquer atividade de aprendizagem, o Caliper Analytics se concentra na coleta de dados com foco em métricas. A proposta introduz uma arquitetura de Learning Record Warehouse, fundamentada no Learning Record Storage (LRS) como um repositório especializado em xAPI. No entanto, através deste trabalho, também se acomodam outras especificações, focando inicialmente no Caliper, com potencial para futuras expansões. Adicionalmente, exploraram-se processos de ETL para integrar dados referentes às especificações discutidas.

A abordagem metodológica empregada foi a Design Science Research (DSR), conduzindo experimentos iterativos para o desenvolvimento e avaliação da arquitetura. O experimento inicial revelou dificuldades na transformação de dados ao empregar o OpenLRW. No segundo experimento, testou-se a habilidade de transferir dados do Caliper para o LRSQL e foi elaborado um conversor para xAPI. A arquitetura resultante integra esse conversor para o armazenamento de dados em um LRS, utilizando

o PostgreSQL como base de dados, ressaltando a crucialidade de preservar o *payload* original do Caliper para evitar perda de informações. Os achados demonstram a capacidade de consultar e visualizar dados no formato JSON com o uso de ferramentas como o Apache Superset.

A conclusão alcançada destaca a importância crucial da interoperabilidade para a unificação de dados oriundos de diferentes fontes. O modelo LRW introduzido pretende funcionar como um intermediário entre os sistemas tradicionais e a xAPI. O desenvolvimento e implementação de um sistema xAPI e LRS necessitam de uma análise minuciosa de custo-benefício. A arquitetura sugerida oferece uma integração dos dados educacionais que é ao mesmo tempo mais eficiente e flexível, promovendo o uso da LA para otimizar os processos de aprendizagem.

## 7.2 Recomendações

De acordo com a investigação realizada, são propostas as recomendações subsequentes para futuros desenvolvimentos e implementações no domínio da integração de dados educacionais e análise de aprendizagem (LA):

O estabelecimento de padrões de interoperabilidade, nomeadamente xAPI e Caliper Analytics, pelas instituições educacionais revela-se crucial. Isto propicia a troca eficaz de dados entre variadas plataformas e sistemas, promovendo, portanto, uma perspectiva holística sobre o processo educacional e possibilitando análises mais extensivas. Tal iniciativa sublinha a necessidade de priorização da interoperabilidade nos dados educacionais.

Dado que a adoção da xAPI ainda não está amplamente difundida, é aconselhável o desenvolvimento de conversores que facilitem a migração dos formatos de dados vigentes para a xAPI, assegurando a integridade das informações originais. O conversor Caliper para xAPI desenvolvido nesta pesquisa exemplifica uma estratégia prática para abordar esse desafio.

A implementação de um LRW é crucial para unificar e normalizar a armazenagem de dados educacionais, o que simplifica tanto a análise como a criação de relatórios. É aconselhável empregar bancos de dados robustos, como o PostgreSQL com suporte a JSON, para uma flexibilidade e eficiência aumentadas.

Recomenda-se se aprofundar a investigação da integração entre a xAPI e o Caliper Analytics, utilizando a adaptabilidade da xAPI para monitorar uma ampla gama de atividades educacionais e empregando a arquitetura do Caliper para medições quantitativas.

É imperativo que a comunidade educacional persista na exploração do avanço dos LRWs, que retêm informações em sua forma original, empregando JSON e tecnologias análogas para mitigar a perda de dados. Adicionalmente, a integração dos LRWs

com abordagens de LA baseadas em nuvem deve ser promovida, tendo em vista prevenir o risco de abandono escolar e ajustar a experiência educacional aos estudantes.

### **7.3 Trabalhos Futuros**

Este estudo se aprofunda nas bases para a integração dos dados educacionais utilizando um Learning Record Warehouse (LRW). Todavia, emergem diversas direções para pesquisas futuras a partir deste trabalho, visando ampliar o impacto e a eficácia desta solução. Assim, sugere-se:

A criação de um LRW que armazene dados no seu formato inicial utilizando PostgreSQL e JSON pode prevenir a perda de informação frequentemente associada à conversão entre diversos formatos de dados, sendo compatível com as capacidades dos bancos de dados contemporâneos.

Integração com soluções de Aprendizado de Máquina na nuvem, o que tornaria o acesso a sofisticadas ferramentas analíticas mais acessível e possibilitaria a implementação de modelos de previsão, tais como a predição do risco de abandono escolar.

Elaborar modelos para a predição do risco de desistência dos estudantes é vital para as instituições de ensino. A criação e execução de modelos que se aproveitem dos dados mantidos no LRW podem facilitar a identificação de estudantes em situação de risco, permitindo intervenções antecipadas.

A consideração de outras especificações é crucial porque, apesar deste estudo se centrar na xAPI e no Caliper Analytics, investigar a incorporação de outras normas educacionais pertinentes poderia aumentar a habilidade do LRW em interoperar com um conjunto mais amplo de plataformas.

A realização de testes extensivos de desempenho e escalabilidade é fundamental para confirmar que o LRW tem a capacidade de processar grandes volumes de dados e atender a um elevado número de usuários simultâneos, garantindo dessa forma tanto a estabilidade quanto a eficiência do sistema em condições reais de operação.

### **7.4 Publicações Acadêmicas**

A seguir são descritos os produtos acadêmicos desenvolvidos e divulgados para discussão com a comunidade acadêmica ao longo deste trabalho de dissertação. Essa produção inclui artigos apresentados em forma de palestras em conferências internacionais e publicações em periódicos.

### 7.4.1 Artigos Publicados

Hernández-Leal, E. J., Rocha, J. C., & Méndez, N. D. D. (2022). Pre-procesamiento de datos educativos desde un enfoque de dominio específico. *Respuestas*, 27(1), 22-37.

Cechinel, C., Queiroga, E. M., Primo, T. T., Ramos, V. F. C., Muñoz, R., Machado, M., & Almeida, J. (2023, November). Learning analytics para Moodle em uma arquitetura na nuvem: uma solução escalável para predição de risco acadêmico. In *Workshop de Aplicações Práticas de Learning Analytics em Instituições de Ensino no Brasil (WA-PLA)* (pp. 128-137). SBC.

Rocha, J. C., Hernández-Leal, E. J., Ramos, V., Munoz, R., Cechinel, C., Primo, T. T., & Queiroga, E. M. (2024, June). Implementing a Learning Record Warehouse for Different Interoperability Specifications in Moodle LMS. In *2024 International Symposium on Computers in Education (SIIE)* (pp. 1-6). IEEE.

Rocha, J. C., Hernández-Leal, E. J., Ramos, V., Munoz, R., Queiroga, E., Primo, T., & Cechinel, C. (2024, junho 18). Implementação de um Learning Record Warehouse para diferentes especificações de interoperabilidade em um Learning Management System. In M. Á. Conde González, M. Rosário Rodrigues, & F. J. García-Peñalvo (Eds.), *Book of Abstracts. SIIE 2024. XXVI International Symposium on Computers in Education*. Zenodo. <https://doi.org/10.5281/zenodo.12084382>

Rocha, J. C., Ramos, V., Cechinel, C., Hernández-Leal, E. J., Munoz, R., & Primo, T. T. (2024, August). Data Interoperability in Learning Analytics-Review of Literature. In *2024 L Latin American Computer Conference (CLEI)* (pp. 1-8). IEEE. Disponível em <https://conferencia2024.clei.org/inicio/pre-proceedings/>

Costa-Rocha, J., & Cechinel, C. (2024). Integração de uma solução Learning Record Warehouse em um serviço de predição de risco acadêmico. In *Proceedings do 18º Congresso Colombiano de Computación (18CCC'24)*.

### 7.4.2 Participação de Intercambio Acadêmico

Instituição: Universidad Nacional de Colombia, sede Medellín

Período: 09 de setembro a 16 de novembro de 2024

## REFERÊNCIAS

1EdTech. **Caliper Analytics 1.2 Specification** | **IMS Global Learning Consortium.** 2023. Disponível em: <<https://www.imsglobal.org/spec/caliper/v1p2#assessmentpassed>>.

**ADL. An Introduction to cmi5: Next-generation of e-Learning Interoperability** | **ADL Initiative.** 2021. Disponível em: <<https://adlnet.gov/news/2021/09/09/cmi5-Resources/>>.

ADL; JOHNSON, A. **xAPI-Spec.** 2024. Disponível em: <<https://github.com/adlnet/xAPI-Spec/blob/master/xAPI-About.md#introduction-partone>>.

Apereo. **GitHub - Apereo-Learning-Analytics-Initiative/OpenLRW: Standards-based learning record warehouse built for a scalable learning analytics environment.** 2019. Disponível em: <<https://github.com/Apereo-Learning-Analytics-Initiative/OpenLRW>>.

APEREO, L. A. I.; GILBERT, G. **Deprecation of OpenLRS.** 2017. Disponível em: <<https://github.com/Apereo-Learning-Analytics-Initiative/OpenLRS/commit/600784400dd02600719f8084f1fc2b02a3288e2b>>.

ARROYO-FIGUEROA, G. et al. A Model of Affect and Learning for Intelligent Tutors A Learning Ecosystem for Linemen Training based on Big Data Components and Learning Analytics. **Journal of Universal Computer Science**, [S.l.], v.25, n.5, p.541–568, 2019.

ASTHAGIRI, N.; MALIK, D.; ZARECOR, E. **OEP-26: Real-time Events — Open edX Proposals 1.0 documentation.** 2018. Disponível em: <<https://open-edx-proposals.readthedocs.io/en/latest/architectural-decisions/oep-0026-arch-realtime-events.html>>.

BAKER, R. S.; INVENTADO, P. S. Educational Data Mining and Learning Analytics. In:

LARUSSON, J. A.; WHITE, B. (Ed.). **Learning Analytics: From Research to practice**. New York, NY: Springer New York, 2014. p.61–75.

BAKHOUYI, A. et al. Evolution of standardization and interoperability on E-learning systems: An overview. **2017 16th International Conference on Information Technology Based Higher Education and Training, ITHET 2017**, [S.l.], 10 2017.

BARBOSA, G. A. **Transformação de dados no processo ETL: a importância da qualidade de dados**. 2020. Disponível em: <<https://ric.cps.sp.gov.br/handle/123456789/25378>>.

BAYER, N. **Introduction to e-learning standards | isEazy**. 2024. Disponível em: <<https://www.iseazy.com/blog/introduction-to-e-learning-standards/>>.

BEHRINGER, R.; GROUP, K.-B. **Interoperability Standards for MicroLearning**. 2013. 26–27p. Disponível em: <<https://www.researchgate.net/publication/258449941>>.

BISHR, Y. A. **Semantic aspects of interoperable GIS**. 1997. Disponível em: <<https://api.semanticscholar.org/CorpusID:61010815>>.

BRERETON, P. et al. Lessons from applying the systematic literature review process within the software engineering domain. **Journal of Systems and Software**, [S.l.], v.80, n.4, p.571–583, 4 2007.

CHICU, N. THE ROLE OF DIGITALIZATION IN THE PROCESS OF EDUCATION-E-LEARNING, E-YOUTH APPROACH. **Euromentor Journal - Studies about education**, [S.l.], v.IX, n.01, p.102–111, 2018.

Comissão Europeia. **COMUNICAÇÃO DA COMISSÃO AO PARLAMENTO EUROPEU, AO CONSELHO, AO COMITÉ ECONÓMICO E SOCIAL EUROPEU E AO COMITÉ DAS REGIÕES**. 2010. Disponível em: <<https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:52010DC0744&from=EN>>.

CONDE, M. A. Assessing What Students Learn Beyond the Institution – Learning Analytics in Personal Learning Environments. **www.metropolitan.ac.rs**, [S.l.], p.7, 9 2018.

CONDE, M. A.; HERNANDEZ-GARCIA, A. Learning analytics for educational decision making. **Computers in Human Behavior**, [S.l.], v.47, p.1–3, 6 2015.

COOPER, A. **Learning Analytics Community Exchange Learning Analytics Interoperability-The Big Picture in Brief an introductory briefing**. 2014.

COOPER, A. R. Learning Analytics Interoperability-a survey of current literature and candidate standards. **Retrieved on**, [S.l.], v.12, n.07, p.2017, 2013.

COSTA, V. P. T. d. **Implantando um provedor de serviços educacionais interoperáveis segundo o padrão IMS-LTI**. 2021. Tese (Doutorado em Ciência da Computação) — .

COSTELLO, T.; BLACKSHEAR, L. Prepare Your Data for Tableau: A Practical Guide to the Tableau Data Prep Tool. **Prepare Your Data for Tableau: A Practical Guide to the Tableau Data Prep Tool**, [S.l.], p.1–202, 1 2019.

DE SÁ MUSSA, M. et al. BUSINESS INTELLIGENCE NA EDUCAÇÃO: UMA APLICAÇÃO DO SOFTWARE PENTAHO. **Revista Produção e Desenvolvimento**, [S.l.], v.4, n.3, p.29–41, 7 2018.

DEL BLANCO, A. et al. E-Learning standards and learning analytics. Can data collection be improved by using standard data models? **IEEE Global Engineering Education Conference, EDUCON**, [S.l.], p.1255–1261, 2013.

DODERO, J. M. et al. Trade-off between interoperability and data collection performance when designing an architecture for learning analytics. **Future Generation Computer Systems**, [S.l.], v.68, p.31–37, 3 2017.

DONDORF, T. **Learning analytics for moodle: facilitating the adoption of data privacy friendly learning analytics in higher education**. 2022. Tese (Doutorado em Ciência da Computação) — Dissertation, Rheinisch-Westfälische Technische Hochschule Aachen, 2022.

DONDORF, T. et al. LEARNING ANALYTICS SOFTWARE IMPLEMENTATION FOR THE MOODLE LEARNING MANAGEMENT SYSTEM. **ICERI2019 Proceedings**, [S.l.], v.1, p.6957–6964, 12 2019.

DONDORF, T.; NACKEN, H. A DATA MODEL FOR LEARNING ANALYTICS IN MOODLE. **EDULEARN17 Proceedings**, [S.l.], v.1, p.2402–2408, 7 2017.

EUROPEAN COMMISSION, D.-G. f. D. S. **New European interoperability framework – Promoting seamless services and data flows for European public administrations**. Belgium: [s.n.], 2017.

FARINELLI, F. .; ALMEIDA, M. B. **INTEROPERABILIDADE SEMÂNTICA EM SISTEMAS DE INFORMAÇÃO DE SAÚDE POR MEIO DE ONTOLOGIAS FORMAIS E INFORMAIS: UM ESTUDO DA NORMA OPENEHR**. 2014.

FURNIEL, A. C.; MENDONÇA, A. P.; SILVA, R. Recursos Educacionais Abertos: conceitos e princípios. **Rio de Janeiro: Fiocruz**, [S.l.], 2016.

GRIFFITHS, D.; HOEL, T. **Comparing xapi and Caliper**. 2016. Disponível em: <<https://docplayer.net/33687583-Comparing-xapi-and-caliper.html>>.

HECKING, T. et al. A flexible and extendable learning analytics infrastructure. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, [S.l.], v.8613 LNCS, p.123–132, 2014.

HERNANDEZ-LEAL, E. J. **Modelo de dominio específico para análisis y minería de datos educativos**. 2024. Tese (Doutorado em Ciência da Computação) — Universidad Nacional de Colombia.

HERNÁNDEZ-LEAL, E. J.; COSTA-ROCHA, J.; DUQUE-MÉNDEZ, N. D. Pre-procesamiento de datos educativos desde un enfoque de dominio específico. **Respuestas, ISSN 0122-820X, ISSN-e 2422-5053, Vol. 27, Nº. 1, 2022 (Ejemplar dedicado a: ENERO - ABRIL 2022), págs. 22-37**, [S.l.], v.27, n.1, p.22–37, 2022.

HEVNER, A. R. et al. Design science in information systems research. **MIS Quarterly: Management Information Systems**, [S.l.], v.28, n.1, p.75–105, 2004.

HODROB, R.; EWAS, A.; MAREE, M. On developing a framework for knowledge-based learning indicator system in the context of learning analytics. **Smart Innovation, Systems and Technologies**, [S.l.], v.142, p.27–37, 2019.

HOEL, T.; CHEN, W. Data Sharing for Learning Analytics—designing conceptual artefacts and processes to foster interoperability. In: INTERNATIONAL CONFERENCE ON COMPUTERS IN EDUCATION, 24., 2020. **Anais...** [S.l.: s.n.], 2020.

HUGO, V.; LOPES, E. **Desenvolvimento de uma Ferramenta ETL para conversão de dados semiestruturados e estruturados em JSON para o modelo relacional**. 2023. Tese (Doutorado em Ciência da Computação) — .

IEEE. IEEE Standard for Learning Technology—JavaScript Object Notation (JSON) Data Model Format and Representational State Transfer (RESTful) Web Service for Learner Experience Data Tracking and Access. **IEEE Std 9274.1.1-2023**, [S.l.], p.1–103, 2023.

JOARDER, S. et al. **Towards a Flexible User Interface for 'Quick and Dirty' Learning Analytics Indicator Design**. 2023. Disponível em: <<https://arxiv.org/abs/2304.01711v1>>.

KEVAN, J. M.; RYAN, P. R. Experience API: Flexible, Decentralized and Activity-Centric Data Collection. **Technology, Knowledge and Learning**, [S.l.], v.21, n.1, p.143–149, 4 2016.

KRUMOVA, M. Research on LMS and KPIs for Learning Analysis in Education. **Smart Cities 2023, Vol. 6, Pages 626-638**, [S.l.], v.6, n.1, p.626–638, 2 2023.

LABBA, C.; ROUSSANALY, A.; BOYER, A. Towards an automated Framework for benchmarking Learning Record Stores: Performance Requirements and Scalability. In: ANNUAL LEARNING & STUDENT ANALYTICS CONFERENCE (LSAC 2019),, 3., 2017, France. **Anais...** [S.l.: s.n.], 2017.

learningpool. **Learning Record Store**. 2024. Disponível em: <<https://learningpool.com/learning-record-store>>.

LLC, L. H. F. **Total Learning Architecture: 2018 Reference Implementation Specifications and Standards**. 2019. Disponível em: <<https://apps.dtic.mil/sti/pdfs/AD1077398.pdf>>.

LOCKWOOD HILLS FEDERAL, L. **Total Learning Architecture: 2018 Reference Implementation Specifications and Standards**. [S.l.: s.n.], 2019.

MANGAROSKA, K. et al. Architecting Analytics across Multiple E-Learning Systems to Enhance Learning Design. **IEEE Transactions on Learning Technologies**, [S.l.], v.14, n.2, p.173–188, 4 2021.

MHON, G. G. W.; KHAM, N. S. M. ETL Preprocessing with Multiple Data Sources for Academic Data Analysis. **2020 IEEE Conference on Computer Applications, ICCA 2020**, [S.l.], 2 2020.

MILLER, B. et al. cmi5 Best Practices Guide: From Conception to Conformance. **Advanced Distributed Learning, Rustici Software**, [S.l.], 2021.

NGUYEN, A.; KHÁNH, X. **Learning Analytics Information Systems for Higher Education: Design, Development and Implementation**. 2019. Tese (Doutorado em Ciência da Computação) — University of Auckland.

NWACHUKWU, U. et al. An Integrated Approach to Learning Analytics. **Proceedings of the 7th International Conference on Frontiers in Education: Computer Science & Computer Engineering**, [S.l.], 2021.

OLIVEIRA, B.; BELO, O. On the specification of extract, transform, and load patterns behavior: A domain-specific language approach. **Expert Systems**, [S.l.], v.34, n.1, p.e12168, 2 2017.

PADIADPU, R. Towards mobile learning: A SCORM Player for the Google Android platform. **University of applied science: Hamburg**, [S.l.], 2008.

PAIK, J. H. et al. The relationships among school engagement, students' emotions, and academic performance in an elementary online learning. **ACM International Conference Proceeding Series**, [S.l.], p.219–230, 3 2024.

PAIK, J. H. et al. The relationships among school engagement, students' emotions, and academic performance in an elementary online learning. **LAK '24: Proceedings of the 14th Learning Analytics and Knowledge Conference**, [S.l.], p.219–230, 3 2024.

PEDRO, J. et al. Padrões para interoperabilidade de dados em aplicações educacionais. **researchgate.netJPS Simão, J Pereira, JB da Mota Alves, J BentoAnais do Simpósio Ibero-Americano de Tecnologias Educacionais, 2018•researchgate.net**, [S.l.], 2018.

PEREZ-BERENGUER, D.; KESSLER, M.; GARCIA-MOLINA, J. A customizable and incremental processing approach for learning analytics. **IEEE Access**, [S.l.], v.8, p.36350–36362, 2020.

PEREZ-COLADO, I. J. et al. Using e-Learning Standards to Improve Serious Game Deployment and Evaluation. **IEEE Global Engineering Education Conference, EDUCON**, [S.l.], v.2022-March, p.2077–2083, 2022.

Qgate. **A Brief History of Learning Management Systems | Q Gate**. 2021. Disponível em: <<https://www.qgate.co.uk/blog/crm/brief-history-of-learning-management-systems/>>.

QUEIROGA, E. M. et al. Experimenting Learning Analytics and Educational Data Mining in different educational contexts and levels. In: XVII LATIN AMERICAN CONFERENCE ON LEARNING TECHNOLOGIES (LACLO), 2022., 2022. **Anais...** [S.l.: s.n.], 2022. p.1–9.

QUIROZ, G.; MUÑOZ GONZÁLEZ, L. Platforms for Online Learning: A Product Specification. **European Journal of Education**, [S.l.], v.6, p.33–42, 8 2023.

RAYON, A.; GUENAGA, M.; NUÑEZ, A. Ensuring the integrity and interoperability of educational usage and social data through Caliper framework to support competency-assessment. **Proceedings - Frontiers in Education Conference, FIE**, [S.l.], v.2015-Febru, n.February, 2 2015.

ROHLOFF, T. et al. A ubiquitous learning analytics architecture for a service-oriented MOOC platform. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, [S.l.], v.11475 LNCS, p.162–171, 2019.

ROMERO, C.; VENTURA, S. Educational data mining and learning analytics: An updated survey. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, [S.l.], v.10, n.3, p.e1355, 7 2020.

RYUNOSUKE, M. et al. Development of a logging system for student activities on Mahara according to xAPI standard by using Apache logs and OpenLRS. **IPSJ SIG Technical Report (CLE)**, [S.l.], v.2016, n.2, p.1–4, 2016.

SAMUELSEN, J. Scalable Higher Education Learning Analytics Architecture through Data Integration. In: EC-TEL (DOCTORAL CONSORTIUM), 2018. **Anais...** [S.l.: s.n.], 2018.

SAMUELSEN, J.; CHEN, W.; WASSON, B. Integrating multiple data sources for learning analytics—review of literature. **Research and Practice in Technology Enhanced Learning**, [S.l.], v.14, n.1, p.1–20, 12 2019.

SAMUELSEN, J.; CHEN, W.; WASSON, B. Integrating multiple data sources for learning analytics—review of literature. **Research and Practice in Technology Enhanced Learning**, [S.l.], v.14, n.1, p.1–20, 12 2019.

SAMUELSEN, J.; CHEN, W.; WASSON, B. Enriching context descriptions for enhanced LA scalability: a case study. **Research and Practice in Technology Enhanced Learning**, [S.l.], v.16, n.1, p.1–26, 7 2021.

SAMUELSEN, J.; CHEN, W.; WASSON, B. Enriching context descriptions for enhanced LA scalability: a case study. **Research and Practice in Technology Enhanced Learning**, [S.l.], v.16, n.1, p.1–26, 12 2021.

SAYÃO, L. F.; MARCONDES, C. H. O desafio da interoperabilidade e as novas perspectivas para as bibliotecas digitais. **Transinformação**, [S.l.], v.20, p.133–148, 2008.

SERRANO-LAGUNA, A. et al. Applying standards to systematize learning analytics in serious games. **Computer Standards and Interfaces**, [S.l.], v.50, p.116–123, 2 2017.

SHETH, A. P. Changing focus on interoperability in information systems: from system, syntax, structure to semantics. In: **Interoperating geographic information systems**. [S.l.]: Springer, 1999. p.5–29.

SIEMENS, G.; DAWSON, S.; LYNCH, G. Improving the quality and productivity of the higher education sector. **Policy and Strategy for Systems-Level Deployment of Learning Analytics**. Canberra, Australia: Society for Learning Analytics Research for the Australian Office for Learning and Teaching, [S.l.], v.31, 2013.

SIEMENS, G. et al. **Open Learning Analytics: an integrated & modularized platform Proposal to design, implement and evaluate an open platform to integrate heterogeneous learning analytics techniques Project Overview**. 2011. Disponível em: <[www.solaresearch.org](http://www.solaresearch.org)>.

SYED, M. et al. Integrated closed-loop learning analytics scheme in a first year experience course. **ACM International Conference Proceeding Series**, [S.l.], p.521–530, 3 2019.

TAROUCO, L. M. R. et al. **Objetos de Aprendizagem: teoria e prática**. 2014. [S.l.]: Evangraf, 2014. Disponível em: <<https://www.lume.ufrgs.br/bitstream/handle/10183/102993/000937201.pdf>>.

TLILI, A. et al. A Complete Validated Learning Analytics Framework: Designing Issues from Data Preparation Perspective. **International Journal of Information and Communication Technology Education (IJICTE)**, [S.l.], v.14, n.2, p.1–16, 1 2018.

VÄKEVÄ, V. **Tracking learning experiences with xAPI**. 2023. Tese (Doutorado em Ciência da Computação) — .

WHYTE, A.; NAYAK, P.; JOHNSTON, J. LAK16 Workshop: Extending IMS caliper analyticstm with learning activity profiles. **ACM International Conference Proceeding Series**, [S.l.], v.25-29-April, p.490–491, 4 2016.

xapi. **What is an LRS? Learn more about Learning Record Stores**. 2018. Disponível em: <<https://xapi.com/learning-record-store/>>.

## **Apêndices**

## APÊNDICE A – Arquivos Técnicos

## A.1 docker-compose para configurar o LRW

```
1 services:
2   db:
3     image: postgres:latest
4     container_name: my_postgres
5     environment:
6       POSTGRES_USER: ${POSTGRES_USER}
7       POSTGRES_PASSWORD: ${POSTGRES_PASSWORD}
8       POSTGRES_DB: ${POSTGRES_DB}
9     volumes:
10      - postgres_data:/var/lib/postgresql/data
11
12   lrs:
13     image: yetanalytics/lrsq:latest
14     environment:
15       LRSQ_API_KEY_DEFAULT: ${LRSQ_API_KEY_DEFAULT}
16       LRSQ_API_SECRET_DEFAULT: ${LRSQ_API_SECRET_DEFAULT}
17       LRSQ_ADMIN_USER_DEFAULT: ${LRSQ_ADMIN_USER_DEFAULT}
18       LRSQ_ADMIN_PASS_DEFAULT: ${LRSQ_ADMIN_PASS_DEFAULT}
19       LRSQ_DB_HOST: ${LRSQ_DB_HOST}
20       LRSQ_DB_PORT: ${LRSQ_DB_PORT}
21       LRSQ_DB_NAME: ${LRSQ_DB_NAME}
22       LRSQ_DB_USER: ${LRSQ_DB_USER}
23       LRSQ_DB_PASSWORD: ${LRSQ_DB_PASSWORD}
24       LRSQ_POOL_INITIALIZATION_FAIL_TIMEOUT: ${
25         LRSQ_POOL_INITIALIZATION_FAIL_TIMEOUT}
26       LRSQ_ALLOW_ALL_ORIGINS: ${LRSQ_ALLOW_ALL_ORIGINS}
27       LRSQ_ENABLE_JSONB: true
28     command:
29       - /lrsq/bin/run_postgres.sh
30     depends_on:
31       - db
32     volumes:
```

```

32     - lrsqldb:/lrsqldb
33
34     api:
35         build: .
36         environment:
37             LRSQ_HOST: ${LRSQ_HOST}
38             LRSQ_PORT: ${LRSQ_PORT}
39             LRSQ_ADMIN_USER_DEFAULT: ${LRSQ_ADMIN_USER_DEFAULT}
40             LRSQ_ADMIN_PASS_DEFAULT: ${LRSQ_ADMIN_PASS_DEFAULT}
41         container_name: flask_api
42         depends_on:
43             - db
44             - lrs
45
46     nginx:
47         image: nginx:latest
48         ports:
49             - "${NGINX_PORT}:8080"
50         volumes:
51             - ./nginx.conf:/etc/nginx/nginx.conf
52         depends_on:
53             - api
54             - lrs
55
56     volumes:
57         postgres_data:
58         lrsqldb:

```

## A.2 docker-compose para configurar o Apache SuperSet

```

1 x-superset-image: &superset-image apachesuperset.docker.scarf.sh/
   apache/superset:${TAG:-latest-dev}
2 x-superset-depends-on: &superset-depends-on
3     - db
4     - redis
5 x-superset-volumes:
6     &superset-volumes # /app/pythonpath_docker will be appended to
   the PYTHONPATH in the final container
7     - ./docker:/app/docker
8     - superset_home:/app/superset_home
9 x-superset-networks: &superset-network

```

```
10 - superset
11 services:
12   redis:
13     image: redis:7
14     container_name: superset_cache
15     restart: unless-stopped
16     volumes:
17       - redis:/data
18     networks:
19       - superset
20
21
22   db:
23     env_file:
24       - path: docker/.env # default
25         required: true
26       - path: docker/.env-local # optional override
27         required: false
28     image: postgres:16
29     container_name: superset_db
30     restart: unless-stopped
31     volumes:
32       - db_home:/var/lib/postgresql/data
33       - ./docker/docker-entrypoint-initdb.d:/docker-entrypoint-initdb
34         .d
35     networks:
36       - superset
37
38   superset:
39     env_file:
40       - path: docker/.env # default
41         required: true
42       - path: docker/.env-local # optional override
43         required: false
44     image: *superset-image
45     container_name: superset_app
46     command: ["/app/docker/docker-bootstrap.sh", "app-gunicorn"]
47     user: "root"
48     restart: unless-stopped
49     ports:
50       - 8088:8088
```

```
50 depends_on: *superset -depends -on
51 volumes: *superset -volumes
52 networks:
53   - superset
54   - lrw_default
55
56
57 superset-init:
58 image: *superset-image
59 container_name: superset_init
60 command: ["/app/docker/docker-init.sh"]
61 env_file:
62   - path: docker/.env # default
63     required: true
64   - path: docker/.env-local # optional override
65     required: false
66 depends_on: *superset -depends -on
67 user: "root"
68 volumes: *superset -volumes
69 healthcheck:
70   disable: true
71 networks:
72   - superset
73   - lrw_default
74
75 superset-worker:
76 image: *superset-image
77 container_name: superset_worker
78 command: ["/app/docker/docker-bootstrap.sh", "worker"]
79 env_file:
80   - path: docker/.env # default
81     required: true
82   - path: docker/.env-local # optional override
83     required: false
84 restart: unless-stopped
85 depends_on: *superset -depends -on
86 user: "root"
87 volumes: *superset -volumes
88 networks:
89   - superset
90   - lrw_default
```

```
91 healthcheck:
92   test:
93     [
94       "CMD-SHELL",
95       "celery -A superset.tasks.celery_app:app inspect ping -d
96         celery@$$HOSTNAME",
97     ]
98
99 superset-worker-beat:
100 image: *superset-image
101 container_name: superset_worker_beat
102 command: ["/app/docker/docker-bootstrap.sh", "beat"]
103 env_file:
104   - path: docker/.env # default
105     required: true
106   - path: docker/.env-local # optional override
107     required: false
108 restart: unless-stopped
109 depends_on: *superset-depends-on
110 user: "root"
111 volumes: *superset-volumes
112 networks:
113   - superset
114   - lrw_default
115 healthcheck:
116   disable: true
117
118 volumes:
119   superset_home:
120     external: false
121   db_home:
122     external: false
123   redis:
124     external: false
125 networks:
126   superset:
127     driver: bridge
128   lrw_default:
129     external: true
```

## APÊNDICE B – Querys de Consulta

### B.1 Atividades dos estudantes

Para identificar estudantes e as respectivas atividades de aprendizagem com maior interação, foi utilizada a seguinte consulta, que calcula o total de visualizações por atividade. A pesquisa filtra eventos onde o verbo é 'Viewed', descartando aqueles que não são relevantes. A análise também é limitada a eventos onde o papel é 'Learner', focando exclusivamente nos estudantes e ignorando interações de professores; esse dado está presente no campo caliperEvent. O nome do estudante é extraído da chave 'name' no campo 'actor' e, para garantir anonimidade, é codificado via função md5. Os dados são então agrupados usando o identificador codificado do estudante, o nome do objeto e a organização correspondente, quantificando o total de visualizações para cada agrupamento.

```

1  SELECT aluno_nome AS aluno_nome ,
2         objeto_nome AS objeto_nome ,
3         total_viewed AS total_viewed
4  FROM
5     (SELECT md5(payload -> 'actor' ->> 'name') AS aluno_nome ,
6         payload -> 'object' -> 'definition' -> 'name' ->> 'en-US'
7         AS objeto_nome ,
8         COUNT(*) AS total_viewed ,
9     (convert_from(decode(payload -> 'context' -> 'extensions' ->> 'http
10      ://example.edu/extensions/caliperEvent', 'base64'), 'utf-8'))::
11      json -> 'membership' ->> 'organization' AS organization
12     FROM xapi_statement
13     WHERE (payload -> 'verb' -> 'display' ->> 'en-US') = 'Viewed'
14     AND TRIM(payload -> 'object' -> 'definition' -> 'name' ->> 'en-US
15         ') != 'Moodle Analytics Dashboard 2'
16     AND (payload -> 'object' -> 'definition' -> 'name' ->> 'en-US')
17         NOT ILIKE '%xAPI%'
18     AND (payload -> 'object' -> 'definition' -> 'name' ->> 'en-US')
19         NOT ILIKE '%Dados%'
20     AND (convert_from(decode(payload -> 'context' -> 'extensions' ->>

```

```

15     'http://example.edu/extensions/caliperEvent', 'base64'), 'utf
16     -8'))::json -> 'membership' ->> 'roles' LIKE '%Learner%'
GROUP BY md5(payload -> 'actor' ->> 'name'),
17     payload -> 'object' ->> 'id', payload -> 'object' -> '
        definition' -> 'name' ->> 'en-US',(convert_from(decode
        (payload -> 'context' -> 'extensions' ->> 'http://
        example.edu/extensions/caliperEvent', 'base64'), 'utf
18     -8'))::json -> 'membership' ->> 'organization'
ORDER BY total_viewed DESC) AS virtual_table
LIMIT 1000;

```

Listing B.1 – Atividades dos Estudantes



aluno_nome	objeto_nome	total_vlewed
efe58abe4a489ff0349f0e5f8a941746	Recuperação 1a - Avaliação 01	11
2cd8f1b3718d60ff0f29368496a8b1f4	A1.2019.01-PC FilaB	9
511a2c3a81d1abd0a8ae420e8adde863	Avisos	9
44588352f7b16845934337af15112e32	Video Árvores - 02/05/2023	6
2cd8f1b3718d60ff0f29368496a8b1f4	Slides Polimorfismo	5
44588352f7b16845934337af15112e32	Árvores AVL (2023.1)	5
854f1e9757fd5d3fadc711723044c202	Avisos	5
354d4a15b9a4d34cc068c7d5f0d4fe34	Descrição Trabalho 01 - Entrega para 22.10.2024	5
08987188f0ae4956be62304e34a44b54	Descrição Trabalho 01 -	4

Figura 31 – Visualização da query no Superset

## B.2 Atividades Visualizadas

A consulta apresentada investiga quais itens na tabela 'xapi\_statement' são mais frequentemente visualizados, classificando-os por nome e a organização relacionada, enquanto acumula o número total de visualizações para cada combinação. Ela seleciona registros nos quais o verbo do evento é 'Viewed', assegurando que

apenas eventos de visualização sejam incluídos, ao mesmo tempo excluindo objetos indesejados. A query também acessa informações adicionais da organização decodificando o campo 'caliperEvent', que está inicialmente em base64, convertendo-o em JSON para extrair o valor correspondente à chave organization no contexto de membership. Com os dados estruturados, os resultados são agrupados, somando o total de visualizações por objeto e organização, e organizados em ordem decrescente conforme o número de visualizações. Isso facilita a identificação dos itens mais visualizados e analisados dentro da plataforma. Conseqüentemente, ela oferece uma visão detalhada da interação dos usuários com vários objetos, ajudando a destacar itens de maior engajamento e a influência das organizações na visibilidade desses conteúdos. Essa abordagem integra as especificações xAPI e Caliper para apresentar uma visualização abrangente dos dados.

```

1 SELECT objeto_nome AS objeto_nome ,
2     total_viewed AS total_viewed
3 FROM
4     (SELECT payload -> 'object' -> 'definition' -> 'name' ->> 'en-US'
5         AS objeto_nome ,
6         COUNT(*) AS total_viewed ,(convert_from(decode(payload -> '
7             context' -> 'extensions' ->> 'http://example.edu/extensions/
8             caliperEvent', 'base64'), 'utf-8'))::json -> 'membership' ->>
9             'organization' AS organization
10        FROM xapi_statement
11        WHERE (payload -> 'verb' -> 'display' ->> 'en-US') = 'Viewed'
12        AND TRIM(payload -> 'object' -> 'definition' -> 'name' ->> 'en-US
13            ') != 'Moodle Analytics Dashboard 2'
14        AND (payload -> 'object' -> 'definition' -> 'name' ->> 'en-US')
15            NOT ILIKE '%xAPI%'
16        AND (payload -> 'object' -> 'definition' -> 'name' ->> 'en-US')
17            NOT ILIKE '%Dados%'
18        GROUP BY payload -> 'object' ->> 'id', payload -> 'object' -> '
19            definition' -> 'name' ->> 'en-US',(convert_from(decode(
20                payload -> 'context' -> 'extensions' ->> 'http://example.edu/
21                extensions/caliperEvent', 'base64'), 'utf-8'))::json -> '
22                membership' ->> 'organization'
23        ORDER BY total_viewed DESC) AS virtual_table
24 LIMIT 1000;

```

Listing B.2 – Atividades Visualizadas

### Atividades ⋮

objeto_nome <span>⌵</span>	total_viewed <span>⌵</span>
2024.2 - CIT7139 - Programacao em Computadores	7 <span>▲</span>
Árvores AVL (2023.1)	7
Avisos	6
Vídeo Árvores - 02/05/2023	6
Fórum de notícias	6
Vídeos Árvores - 10/05/2023	5
09-Arvores	4
Avallação 01 semestre 2018.02	4
Vídeos Árvores - 09/05/2023	4
Vídeos Árvores - 03/05/2023	4
Slides Agregação	4 <span>▼</span>

Figura 32 – Visualização no SuperSet das Atividades mais acessadas.

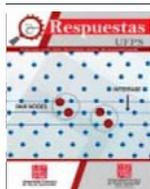
## APÊNDICE C – Tabelas de Eventos

Tabela 8 – Eventos suportados pelos plugins Caliper e xAPI

<b>Evento</b>	<b>Plugin Caliper</b>	<b>Plugin xAPI</b>
\core\event\user_loggedin	✓	✓
\core\event\user_loggedout	✓	✓
\core\event\course_viewed	✓	✓
\mod_book\event\course_module_viewed	✓	✓
\mod_chat\event\course_module_viewed	✓	✓
\mod_choice\event\course_module_viewed	✓	✓
\mod_data\event\course_module_viewed	✓	✓
\mod_facetoface\event\course_module_viewed	✓	✓
\mod_feedback\event\course_module_viewed	✓	✓
\mod_folder\event\course_module_viewed	✓	✓
\mod_forum\event\course_module_viewed	✓	✓
\mod_forum\event\user_report_viewed	✓	✓
\mod_glossary\event\course_module_viewed	✓	✓
\mod_imscc\event\course_module_viewed	✓	✓
\mod_lesson\event\course_module_viewed	✓	✓
\mod_lti\event\course_module_viewed	✓	✓
\mod_quiz\event\attempt_reviewed	✓	✓
\mod_quiz\event\attempt_started	✓	✓
\mod_quiz\event\attempt_submitted	✓	✓
\mod_quiz\event\attempt_viewed	✓	✓
\mod_quiz\event\course_module_viewed	✓	✓
\mod_resource\event\course_module_viewed	✓	✓
\mod_scorm\event\course_module_viewed	✓	✓
\mod_survey\event\course_module_viewed	✓	✓
\mod_url\event\course_module_viewed	✓	✓
\mod_wiki\event\course_module_viewed	✓	✓
\mod_workshop\event\course_module_viewed	✓	✓
\mod_assign\event\assessable_submitted	✓	✓

\mod_assign\event\submission_graded	✓	✓
\mod_quiz\event\attempt_abandoned	✗	✓
\core\event\course_completed	✗	✓
\core\event\user_created	✗	✓
\core\event\user_enrolment_created	✗	✓
\core\event\course_module_completion_updated	✗	✓
\mod_bigbluebuttonbn\event\activity_viewed	✗	✓
\mod_bigbluebuttonbn\event\live_session_event	✗	✓
\mod_bigbluebuttonbn\event\meeting_created	✗	✓
\mod_bigbluebuttonbn\event\meeting_ended	✗	✓
\mod_bigbluebuttonbn\event\meeting_joined	✗	✓
\mod_bigbluebuttonbn\event\meeting_left	✗	✓
\mod_bigbluebuttonbn\event\recording_deleted	✗	✓
\mod_bigbluebuttonbn\event\recording_edited	✗	✓
\mod_bigbluebuttonbn\event\recording_imported	✗	✓
\mod_bigbluebuttonbn\event\recording_protected	✗	✓
\mod_bigbluebuttonbn\event\recording_published	✗	✓
\mod_bigbluebuttonbn\event\recording_unprotected	✗	✓
\mod_bigbluebuttonbn\event\recording_unpublished	✗	✓
\mod_bigbluebuttonbn\event\recording_viewed	✗	✓
\mod_book\event\chapter_viewed	✗	✓
\totara_program\event\program_assigned	✗	✓

## **Anexos**



## Pre-procesamiento de datos educativos desde un enfoque de dominio específico.

Educational data pre-processing from a domain-specific approach.

Emilcy Juliana Hernández-Leal<sup>1\*</sup>; Juary Costa-Rocha<sup>2</sup>; Néstor Darío Duque-Méndez<sup>3</sup>

<sup>1</sup>Doctora(c) en Ingeniería -Industria y Organizaciones-, [ejhernandez@udem.edu.co](mailto:ejhernandez@udem.edu.co), <https://orcid.org/0000-0002-5865-9604>, Universidad de Medellín, Medellín, Colombia.

<sup>2</sup>Ingeniero de Energía, [juary.c.rocha@grad.ufsc.br](mailto:juary.c.rocha@grad.ufsc.br), <https://orcid.org/0000-0001-7028-6524>, Universidade Federal de Santa Catarina, Araranguá, Brasil.

<sup>3</sup>Doctor en Ingeniería, [ndduqueme@unal.edu.co](mailto:ndduqueme@unal.edu.co), <https://orcid.org/0000-0002-4608-281X>, Universidad Nacional de Colombia, Manizales, Colombia.

**Cómo citar:** E. J. Hernández-Leal, J. Costa-Rocha, N. D. Duque-Méndez, “Pre-procesamiento de datos educativos desde un enfoque de dominio específico”, *Respuestas*, vol. 27, no. 1, pp. 22-37, 2022. <https://doi.org/10.22463/0122820X.3113>

Recibido: Agosto 10, 2021; Approved: Diciembre 13, 2021.

### RESUMEN

#### Palabras clave:

Datos Educativos,  
Dominio Específico,  
Minería De Datos  
Educativos (EDM),  
Preparación De Datos, Pre  
procesamiento De Datos.

Los procesos de análisis de datos requieren preprocesamiento antes de la aplicación de técnicas o algoritmos, con el fin de incrementar la calidad de estos y adecuarlos a los formatos necesarios para su procesamiento, principalmente cuando los datos provienen de diferentes fuentes. El presente artículo expone la experiencia en el diseño y construcción de una estrategia con enfoque de dominio específico para el proceso de preparación de datos educativos. La metodología del estudio incluyó tres etapas: (1) diseño y construcción de la estrategia para preparación de los datos educativos, (2) reconocimiento y selección de datos y (3) aplicación de la estrategia y revisión de resultados. El caso de estudio estuvo conformado por datos provenientes del sistema de educación básica y media del departamento Norte de Santander (Colombia). Se contó con datos referentes al proceso de matrícula, los cuales incluyen variables de tipo socioeconómico y familiar y con datos de valoraciones de desempeño académico de estudiantes provenientes de tres instituciones educativas de carácter público. Para las dos fuentes se procesaron datos de los años 2014 a 2018, con un total de más de ochocientos mil registros. Este trabajo aporta valor en tres aspectos principalmente: el alcance respecto al nivel educativo de dónde provienen los datos del caso de estudio, la inclusión del enfoque de dominio específico en la solución y la centralización de los datos de las múltiples fuentes, resultando datos disponibles para posteriores procesos de análisis. En conclusión, este trabajo contribuyó tanto en el ámbito investigativo como en la aplicación del conocimiento en un caso existente y abrió la posibilidad de realizar pruebas posteriores con otro tipo de datos del contexto educativo.

### ABSTRACT

#### Keywords:

Educational Data, Specific  
Domain, Educational Data  
Mining (EDM),  
Data Preparation, Data  
Preprocessing.

The data analysis processes to discover knowledge require pre-processing before applying techniques or algorithms to increase the quality of the data and adapt them to the formats that are best suited for processing, especially when the data comes from different sources. This article presents the experience in designing and constructing a strategy with a specific domain approach for the educational data preparation process. The study methodology included three stages: (1) design and construction of the strategy, (2) recognition and data selection, and (3) application of the strategy and review of results. The study was made up of data from the primary and secondary education system in the Norte de Santander department (Colombia). In addition, there was data referring to the enrollment process, which includes socioeconomic and family variables and data from evaluations of students' academic performance from three public educational institutions. For the two sources, data were processed from 2014 to 2018, with more than eight hundred thousand records. This work adds value in three main aspects: the scope concerning the educational level where the case study data comes from, the inclusion of the domain-specific approach in the solution, and the centralization of the data from multiple sources, resulting in data available in subsequent analysis processes. In conclusion, this work contributed both in the research field and applying knowledge in an existing case. Furthermore, it opened the possibility of carrying out subsequent tests with other data types from the educational context.

\*Corresponding author.

E-mail Address: Emilcy Juliana Hernández-Leal ([ejhernandez@udem.edu.co](mailto:ejhernandez@udem.edu.co))

Peer review is the responsibility of the Universidad Francisco de Paula Santander.



This is an article under the license CC BY-NC 4.0

# Implementação de um Learning Record Warehouse para diferentes especificações de interoperabilidade em um Learning Management System

Juary Costa Rocha  
*Centro de Desenvolvimento Tecnológico*  
*Universidade Federal de Pelotas*  
Pelotas, Brasil  
jcrocha@inf.ufpel.edu.br

Emilcy Juliana Hernández-Leal  
*Facultad de Ingenierías*  
*Universidad de Medellín*  
Medellín, Colombia  
ejhernandez@udemedellin.edu.co

Vinicius Ramos  
*Dept. de Engenharia do Conhecimento*  
*Universidade Federal de Santa Catarina*  
Florianópolis, Brasil  
v.ramos@ufsc.br

Roberto Munoz  
*Escuela de Ingeniería Informática*  
*Universidad de Valparaíso*  
Valparaíso, Chile  
roberto.munoz@uv.cl

Cristian Cechinel  
Coord. Tec. da Informação e Comunicação (CIT)  
*Universidade Federal de Santa Catarina*  
Araranguá, Brasil  
cristian.cechinel@ufsc.br

Tiago Thompsen Primo  
*Centro de Engenharias*  
*Universidade Federal de Pelotas*  
Pelotas, Brasil  
tiago.primo@inf.ufpel.edu.br

Emanuel Marques Queiroga  
*Diretoria de Tec. da Informação*  
*Instituto Federal Sul-rio-grandense*  
Pelotas, Brasil  
emanuelmqueiroga@gmail.com

## I. INTRODUÇÃO

A interoperabilidade semântica desempenha um papel fundamental na eficácia dos sistemas de Análise de Aprendizagem, servindo como o canal pelo qual os dados de diversas fontes podem ser perfeitamente integrados e interpretados por um aplicativo central sem comprometer seu significado inerente.

## II. SOLUÇÕES EXISTENTES PARA INTEROPERABILIDADE

Entre os vários mecanismos que facilitam a interoperabilidade no Learning Analytics, duas especificações se destacam: Caliper Analytics e Experience API (xAPI). Esses padrões fornecem as estruturas necessárias para harmonizar os fluxos de dados entre plataformas e aplicativos educacionais, possibilitando um ecossistema coeso para análise de dados e geração de insights.

O Caliper Analytics é uma especificação liderada pela IEdTech, uma comunidade internacional de universidades e empresas que promove a padronização e as especificações internacionais em e-learning e ICT educacional. O Caliper Analytics adota o conceito de um evento, que aborda um contexto de uso com a seguinte estrutura: um ator, que pode ser um aluno, um professor ou qualquer agente educacional; uma ação (ou intervenção), que representa o que o ator está fazendo; e um objeto, que representa o material que o ator está usando [1].

xAPI, também conhecida como TinCan, é uma especificação técnica aberta para tecnologias educacionais que permite que

o conteúdo de aprendizagem e os sistemas de aprendizagem se comuniquem entre si para registrar e rastrear experiências de aprendizagem. Ela especifica uma estrutura para a descrição de experiências de aprendizagem e define como essas descrições podem ser trocadas, sendo um dos primeiros serviços desenvolvidos com foco no compartilhamento de dados sobre experiências de aprendizagem [2], [3].

A xAPI teve sua primeira versão 1.0 lançada em 2013, enquanto o Caliper Analytics teve sua versão 1.0 lançada em outubro de 2015 [4]. De muitas maneiras, a xAPI e o Caliper executam funções semelhantes, fazendo uso de conceitos semelhantes. No entanto, eles são bastante diferentes em suas abordagens de desenvolvimento e governança. O Caliper foi criado por meio de um processo fechado em uma organização de membros, enquanto a xAPI surgiu de um processo aberto em que todas as partes interessadas são bem-vindas para contribuir. A propriedade intelectual do Caliper é detida pela IEdTech, enquanto a xAPI, por outro lado, é disponibilizada gratuitamente para todos sob uma licença Apache 2. O Caliper se concentra principalmente nos casos de uso coletados das organizações membros da IEdTech, ao mesmo tempo em que fornece extensões, enquanto a xAPI delega grande parte da responsabilidade de definir os vocabulários a serem usados na especificação para as comunidades, ao mesmo tempo em que desempenha um papel na garantia da funcionalidade principal.

# Implementing a Learning Record Warehouse for different interoperability specifications in Moodle LMS

Juary Costa Rocha

*Centro de Desenvolvimento Tecnológico  
Universidade Federal de Pelotas  
Pelotas, Brasil  
jcrocha@inf.ufpel.edu.br*

Emilcy Juliana Hernández-Leal

*Facultad de Ingenierías  
Universidad de Medellín  
Medellín, Colombia  
ejhernandez@udemedellin.edu.co*

Vinicius Ramos

*Depto. de Engenharia do Conhecimento  
Universidade Federal de Santa Catarina  
Florianópolis, Brasil  
v.ramos@ufsc.br*

Roberto Munoz

*Escuela de Ingeniería Informática  
Universidad de Valparaíso  
Valparaíso, Chile  
roberto.munoz@uv.cl*

Cristian Cechinel

*Coord. Tec. da Informação e Comunicação (CIT)  
Universidade Federal de Santa Catarina  
Aranaguá, Brasil  
cristian.cechinel@ufsc.br*

Tiago Thompsen Primo

*Centro de Engenharias  
Universidade Federal de Pelotas  
Pelotas, Brasil  
tiago.primo@inf.ufpel.edu.br*

Emanuel Marques Queiroga

*Diretoria de Tec. da Informação  
Instituto Federal Sul-rio-grandense  
Pelotas, Brasil  
emanuelmqueiroga@gmail.com*

**Abstract**—Semantic interoperability is an important aspect in Learning Analytics systems, allowing data from various sources to be processed by a central application while preserving its original meaning. In this regard, there are two specifications that enable interoperability for Learning Analytics: Caliper Analytics and the Experience API (xAPI). This study presents the implementation of a Learning Record Warehouse (LRW) for storing Moodle logs, using the specifications Caliper Analytics and xAPI. OpenLRW was chosen as the platform to store data collected by plugins responsible for gathering Moodle logs in Caliper Analyzer and xAPI formats. Despite OpenLRW having an internal converter to convert data from xAPI to Caliper Analyzer format, some bottlenecks were identified during the experiments conducted within this study.

**Index Terms**—Learning Analytics, Learning Record Warehouse, Caliper Analyzer, xAPI, Standards, Moodle.

## I. INTRODUCTION

The learning process is currently connected to the digital space, where interaction takes place through different means, such as learning management systems, online training, simulations, digital classes, or games. As a result, there is heterogeneity in learning activities, and keeping track and recording these student learning experiences becomes challenging [1]. In this sense, digital learning environments can be instrumented to allow important educational activities to be collected and stored in a standardized way, both for real-time analysis and for later analysis. Learning Analytics techniques are used for this purpose.

Learning Analytics (LA) is a set of techniques for collecting, measuring, and analyzing data about students in a given context, with the aim of understanding and optimizing learning in educational environments [2]. However, data collection can come from heterogeneous sources, which poses challenges in terms of transmission, processing, storage, and traceability. To address these issues, it is necessary to develop systems designed for these challenges according to the context, so that interoperability can be achieved between systems and standardized collection and storage can be achieved. Interoperability in LA has become a key issue in the LA community since the proposal of [2], because interoperability, is the ability of a system or a digital object to exchange information with each other, guaranteeing compatibility between the different systems used in the process, regardless of who developed them [3]–[6].

Two main specifications aim to address interoperability in LA: Caliper Analytics, from 1EdTech<sup>1</sup>, and the Experience API (xAPI), from Advanced Distributed Learning (ADL<sup>2</sup>).

This article presents an LRW implementation for collecting and storing, in a standardized way, student learning activities in the Moodle virtual learning environment, which were collected using the LA interoperability specifications.

The rest of the document is structured as follows: Section II describes the main concepts covered in this article and section

<sup>1</sup><https://www.1edtech.org/>

<sup>2</sup><https://adlnet.gov/>

# Data Interoperability in Learning Analytics - review of literature

Juary Costa Rocha  
Centro de Desenvolvimento Tecnológico  
Universidade Federal de Pelotas  
Pelotas, Brasil  
jcrocha@inf.ufpel.edu.br

Emiley Juliana Hernández-Leal  
Facultad de Ingenierías  
Universidad de Medellín  
Medellín, Colombia  
ejhernandez@udemedellin.edu.co

Vinicius Ramos  
Depto. de Engenharia do Conhecimento  
Universidade Federal de Santa Catarina  
Florianópolis, Brasil  
v.ramos@ufsc.br

Roberto Munoz  
Escuela de Ingeniería Informática  
Universidad de Valparaíso  
Valparaíso, Chile  
roberto.munoz@uv.cl

Cristian Cechinel  
Coord. Tec. da Informação e  
Comunicação (CIT)  
Universidade Federal de Santa Catarina  
Araranguá, Brasil  
cristian.cechinel@ufsc.br

Tiago Thompsen Primo  
Centro de Engenharias  
Universidade Federal de Pelotas  
Pelotas, Brasil  
tiago.primo@inf.ufpel.edu.br

**Abstract**—Learning analytics (LA) and educational data mining (EDM) are two complementary approaches to modeling and understanding teaching-learning processes and, in general, data from academic environments. LA is applied to data from various sources, which can vary in format, granularity, and structure. Integrating these data is key to addressing the challenge of scalability in LA, a fundamental aspect. To this end, interoperability, understood as the ability of different systems, devices, or applications to connect, interact, and work together effectively, is crucial and generates the need for specifications for the case of academic information systems and Learning Management Systems. According to this context, the objective of this work was to address through a literature review the following main question: What are the main challenges for modeling architecture to support the interoperability of educational data to apply Learning Analytics? To develop the review, the team used Parsifal, an online tool designed to conduct systematic literature reviews in the context of software engineering. The initial search was done in six databases, deciding to include twenty papers in the final report. The results showed that there are still many open spaces for research and development in terms of the design and use of educational data specifications for the subsequent application of LA, to make the transition from models built on data coming from a single source to the construction of models that report results from the integration of several sources using specifications like Caliper Analytics or Experience API.

**Index Terms**—Data Source, Data Interchange, Educational Data Mining, Interoperability, Learning Analytics.

## I. INTRODUCTION

Learning analytics (LA) include different strategies carried out from the application of data analysis techniques in educational environments that seek to address problems associated with the teaching-learning processes [1]. In recent years, the progress made in the data science ecosystem has allowed not only to talk about learning analytics and educational data mining, but also about Academic Analytics, Institutional Analytics, Teaching Analytics, Data-Driven Education, Data-Driven Decision-Making in Education, Big Data in Education, and Educational Data Science.

However, for this paper, learning analytics will be considered as the generic term that encompasses the efforts focused on giving value to the educational data produced from the interaction of students in different educational environments, including those mediated by technology such

as traditional environments or face-to-face classrooms. However, it should be noted that most of the data from higher education that are subject to LA correspond to those produced in Learning Management Systems (LMS) and those that come from the academic information systems of educational institutions [2].

In this order of ideas, considering the diverse sources that can feed an LA study, a characteristic that allows the integration of these sources, interoperability, becomes valuable [3]. Interoperability in LA seeks to allow the integration of data from various sources, but beyond that, to provide a more complete and accurate view of educational processes and student performance [4]. Likewise, by having this factor, an expansion of the analytical scope can be achieved, given that different educational systems and platforms are connected, achieving broader and more varied data sets, which could result in a means to facilitate decision-making, since the possibilities for the discovery of patterns, the generation of models and in general the inputs for the application of different techniques and analytics are expanded. This helps administrators, teachers, and other stakeholders to make informed and data-driven decisions to improve teaching-learning processes and the generation of policies and action plans.

By ensuring that data from various sources, such as learning management systems, educational applications, and academic records, are compatible and can communicate with each other, an enabling environment for LA implementation is created [5]. Interoperability not only optimizes data collection but also enables a holistic view of student progress and performance, which in turn drives informed decision-making and continuous improvement in education [6]. However, achieving interoperability and its benefits requires the application of specifications or standards to support it.

Specifications that support interoperability in LA are key to ensuring the consistency and integrity of the data used in this process. These specifications, such as Caliper [7] and Experience API (xAPI) [8], provide a common framework for communication and data exchange between different educational tools and systems. By adopting these standards, educational institutions can easily integrate various technology solutions and analytics platforms. This standardization not only simplifies the implementation of analytics systems but also fosters collaboration and innovation

# Integração de uma solução Learning Record Warehouse em um Serviço de Predição de Risco Acadêmico

Juary Costa-Rocha<sup>1</sup>[0000-0001-7028-6524] and Cristian Cechinel<sup>1</sup>[0000-0001-6384-409X]

Universidade Federal de Pelotas, Pelotas, Brasil  
{jcrocha,cristian.cechinel}@inf.ufpel.edu.br

**Abstract.** A interoperabilidade semântica é crucial para a eficácia dos sistemas de Análise de Aprendizagem, atuando como o meio pelo qual dados de diversas fontes podem ser integrados e interpretados por um aplicativo central sem perder seu significado original. Entre os diversos mecanismos que promovem a interoperabilidade no Learning Analytics, destacam-se duas especificações: Caliper Analytics e Experience API (xAPI). Esses padrões fornecem as estruturas necessárias para harmonizar os fluxos de dados entre plataformas e aplicativos educacionais, criando um ecossistema coeso para análise de dados e geração de insights.

**Keywords:** Ambientes Virtuais de Aprendizagem · Computação em Nuvem · Data Warehouse · Learning Analytics · Learning Record Store

## 1 Problema e hipótese da pesquisa

As motivações para execução deste projeto são devido à tendência da adoção do ensino digital, que pode acontecer de forma totalmente digital (Educação a Distância) ou de forma híbrida, referidos como aprendizagem combinada [4], tornando, desta forma, o ambiente de ensino digital e sistemas de gestão de aprendizagem, como parte integrante das atividades diárias de ensino. Essa tendência foi intensificada pela pandemia do COVID-19 em 2020, movendo quase toda a educação para o formato online [1], levando a um rápido crescimento no interesse de ferramentas de apoio ao ensino e aprendizagem digital [1]. Com essa adoção, obtêm-se dados em abundância, que necessitam de um sistema de armazenamento pensado para receber dados provenientes de diferentes fontes [6].

Muitas soluções de LA (Learning Analytics) são entregues na forma de SaaS (Software as a Service) [1], permitindo que a mesma solução possa ser utilizada por diferentes instituições. Ou seja, quando se trata de diversas instituições, estas também possuem diferentes tipos de dados tanto com relação ao conteúdo quanto a natureza e utilizam formas de armazenamento e transmissão dos mais variados tipos possíveis. É neste contexto que esta dissertação está inserida e busca responder a seguinte pergunta de pesquisa: Como um Learning Record Warehouse pode ser integrado a uma solução SaaS de Learning Analytics para