Documentation Linked Data Education Overview

Final Report

06.09.2024

Bern University of Applied Sciences
Business School
Institute Public Sector Transformation (IPST)

Table of Contents

1 Project Objective

The primary objective of this project was to provide a comprehensive overview of Linked Data education and training programs, encompassing both Swiss and international offerings. Additionally, programs focusing on Linked Geo Data were also considered.

As the use of Linked Data expands across various sectors, there is an increasing need for education and training in this field. This project aims to assist individuals who want to increase their knowledge about Linked Data by helping them find appropriate educational resources. Furthermore, it will support the development of future educational programs by highlighting areas that need further attention.

The outcome is a detailed overview of Linked Data training resources, highlighting the most relevant training opportunities. The overview was documented in a separate spreadsheet file. This report summarizes the findings.

2 Approach

To create a comprehensive overview and gather extensive information on Linked Data education, a structured approach was used. The first step involved specifying appropriate search keywords. To ensure transparency in the data collection process, the keywords used are listed in Table 1.

Language	Category	Keywords
English	General Linked Data education	– Linked (Geo) Data courses – Linked (Geo) Data training – Linked (Geo) Data certificate programs – Online courses in Linked (Geo) Data" – Linked (Geo) Data MOOCs" (Massive Open Online Courses) – (Geo)SPARQL courses – Linked (Geo) Data tutorials – Linked (Geo) Data documentations – Linked (Geo) Data communities
	Regional Linked Data: Countries (Switzerland, Germany, Austria, Italy, and France)	– Linked Data education (country) – Linked Data training/courses (country) – Linked Data certificate programs (country)
	Related terms	– Semantic Web – RDF – Ontologies – Knowledge Graphs
German	General Linked Data education	– Kurse zu Linked Data – Linked Data Schulungen – Linked Data Zertifikatsprogramme – Online Kurse in Linked Data – SPARQL Kurse – Linked Data Tutorials/Dokumentationen/Communities
	Related terms	– Semantisches Web – Ontologien – Wissensgraphen
Italian	General Linked Data education	– Corsi Linked Data – Formazione su Linked Data – Programmi di certificazione Linked Data – Corsi online su Linked Data – MOOC su Linked Data (corsi online aperti su larga scala) – Corsi SPARQL – Tutorial/documentazioni/comunità su Linked Data

	Related terms	– Web semantico – Ontologie – Grafi della conoscenza
French	General Linked Data education	– Cours sur les Linked Data – Formation sur les Linked Data – Programmes de certificat en Linked Data – Cours en ligne sur les Linked Data – MOOCs sur les Linked Data – Cours SPARQL – Tutoriels/documentations/communautés sur les Linked Data
	Related terms	– Web sémantique – Graphes de connaissances

Table 1: Search keywords

After the initial step of specifying the keywords, desk research was conducted. This involved searching various Massive Open Online Course (MOOC) sites, using Google for internet searches and YouTube to find videos. Courses were carefully selected, excluding those where Linked Data was merely a chapter or just a small part of the course, as well as academic programs requiring enrollment in an entire study program.

Although scientific articles are valuable, they were excluded since the overview is aimed at practitioners. The search was limited to the first three pages of Google search results to focus on the most relevant content.

To refine the search for relevant YouTube content, filters were applied for video length (over 20 minutes) and age (less than 8 years old). However, even with these criteria, evaluating and categorizing YouTube content by quality and relevance remained challenging.

On-site programs were narrowed down to those accessible within a two-hour train ride beyond the Swiss border. When both on-site and online options were available, the online option was chosen because it is more easily accessible. Regarding course fees, standard prices were used due to the unpredictability of special offers. This structured approach ensured a thorough and practical overview of Linked Data education opportunities.

To help users find suitable training programs, the resources were categorized. A predefined list of keywords was used, attempting to assign them as accurately as possible to each resource. These keywords, however, do not represent the content comprehensively but indicate the general direction of the course. It is crucial to choose the right course based on more than just the keywords.

3 Attributes to Describe Training Resources

Table 2 outlines the attributes used to classify the identified resources. Efforts were made to specify all attributes for each resource, with empty fields indicating a lack of available information.

Category	Description
Name	Title of the resource
Location	Course location (primarily on-site or online)
Type	Type of training resource: – Course: In-depth/comprehensive explanation of the topic – Tutorial: Hands-on, step by step explanation/implementation – Documentation: Mostly technical documentation for specific implementations – Multiple resources: Combination of multiple resources (e.g. blogs, presentations, exercises etc.) which cannot be easily assigned to one of the above categories.
Price	Course fee

Currency	Currency in which the price is quoted
Payment	Frequency of payment (in case of subscription models)
Time Effort	Total duration of the program in hours
Date	The date, when the course takes place. The information differs depending on whether it refers to a static resource (e.g., an online video) or an offering that occurs at specific times or intervals (e.g., weekly courses). – Static = Last updated or creation date – Periodic = Periodicity (e.g. daily, weekly, monthly)
Duration	Specifies the time required to complete the course. – Online courses: Flexible – In-person courses: e.g., 3 months
Keywords	Predefined set of terms assigned to each program to indicate the main contents
Content	Brief description of the topics covered
Preferred/required knowledge/materials to participate	Specifies any preferred or required knowledge or materials needed for the program
Certification	Indicates if a certificate is provided for attending or completing the program (Yes / No)
Teaching format	Specifies how the content is presented to the end-user. – Classroom teaching: Mostly in-person and synchronous – Self-study: Mostly online and asynchronous – Blended learning: Combination of classroom teaching and self-study
Provider	The platform or organization offering the program
Language	The primary language used for presenting the content
Target audience	The primary group intended to benefit from the program
URL	Direct link to the corresponding program
Summary	A brief description of the program

Table 2: Explanation attributes

4 Key Findings

This section is divided into two parts. The first part presents general insights gained from the research. The second part lists and describes specific insights based on an analysis of the gathered data.

4.1 General insights

– The overview contains 100 entries on various educational resources related to Linked Data.

– The concept of Linked Data encompasses a broad spectrum of technologies and methodologies, such as knowledge graphs, ontologies, SPARQL, OWL, data management, data reuse etc. While numerous resources are available on individual Linked Data topics, comprehensive and complete materials covering most topics are scarce.

– There are only few resources focusing on Linked Geo Data. Most of the content in this area focuses on Geospatial Data Analytics or Geospatial Information Systems, which are outside the scope of this overview. Resources on GeoSPARQL are limited, with the majority being technical and scientific documentations.

– There is a lack of on-site classroom teaching opportunities.

– Academic resources remain limited. While some universities offer Linked Data courses, they often require full program enrollment, are geographically inaccessible, or lack clear information regarding standalone participation options.

– Although a wealth of scientific articles exists, they were excluded from this review due to their specificity and unsuitability for a general audience. Similarly, many PowerPoint presentations on related topics are available, but their quality and reliability couldn't be consistently assessed. Therefore, they were excluded.

– Several relevant resources were either one-time offerings with no clear indication of their future availability, or they were already concluded and not available anymore.

– Categorizing YouTube content and assessing it based on quality and relevance remains very difficult.

4.2 Key descriptive statistics

70% of the listed resources are free of charge, with most being available online, offering greater accessibility and location independence.

Price (in %)

70	26	4
Free	Paid	N/A

More than 80% of the resources are designed for self-study, while blended learning and classroom teaching make up a much smaller share.

Teaching Format (in %)

|			
88	6	4	2
Self-study	Blended learning	Classroom teaching	N/A
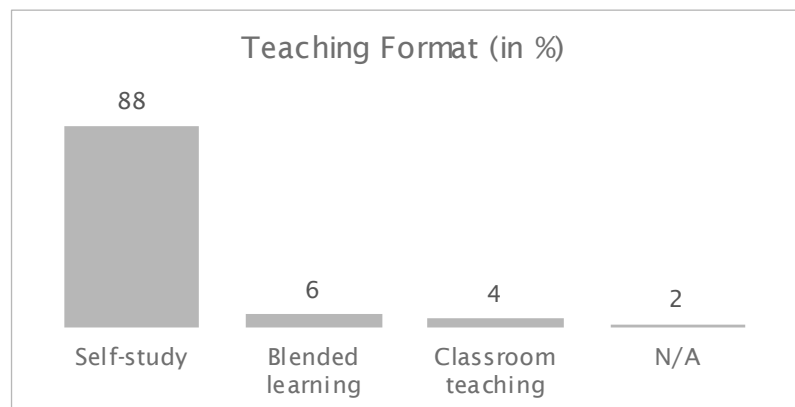
Courses are the most common resource type, emphasizing structured, in-depth learning over other formats. Many resources combine various teaching elements. Documentations and tutorials are less common.

Type (in %)

63	25	6	6
Course	Multiple resources	Documentation	Tutorial
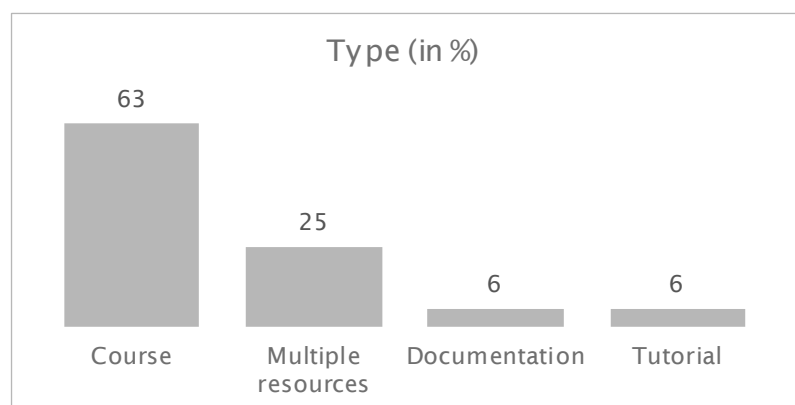
A majority of the resources do not offer any type of certification.

Certification (in %)

No	Yes	N/A
65	23	12
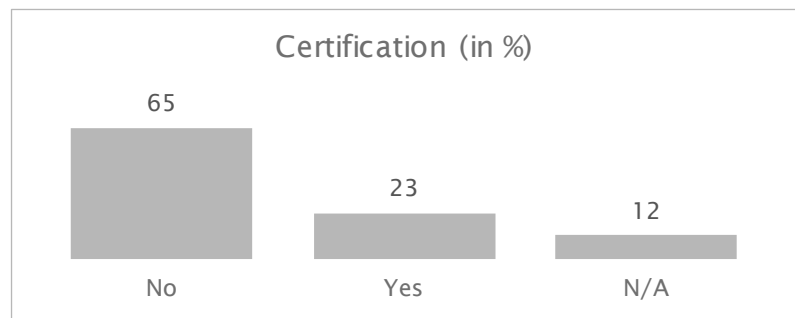
Most resources are in English, with French, German, and Italian being less represented.

Language (in %)

English	French	German	Italian
77	12	8	3
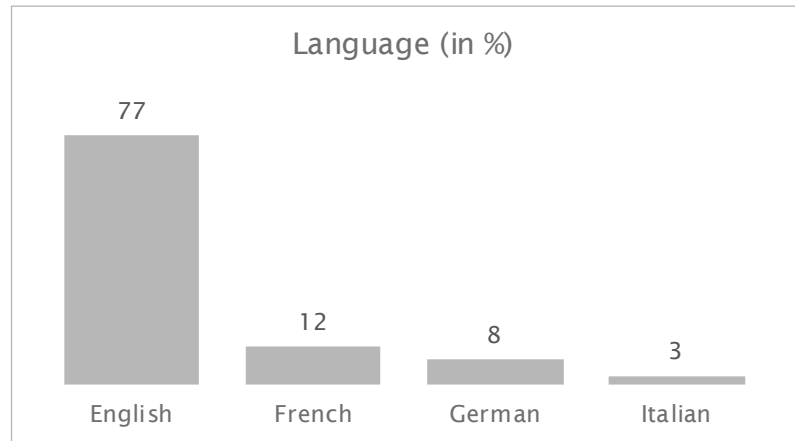
5 Limitations

This project encountered several limitations, primarily due to the diverse nature of available resources. This made it challenging to categorize and evaluate their content and quality accurately. To tackle this challenge, we defined clear search parameters (see section *2 Approach*) and relied on personal judgment only when it was unavoidable. If the quality or relevance of a resource could not be clearly assessed, it was included to arrive at a list of teaching resources that is as comprehensive as possible. Based on the gathered data, future work could be conducted to assess the relevance and quality of each resource in greater detail.

6 Conclusion & Recommendation

The project successfully managed to gather data about a wide range of teaching resources and courses. The resulting overview provides a comprehensive resource list that supports learning about Linked Data and related topics, helping individuals expand their knowledge and refine their skills. However, the search and subsequent analysis also revealed gaps. The lack of classroom courses in Switzerland disadvantages all that favor a more structured and cooperative learning approach together with peers. The lack of resources focusing on Geo Linked Data poses a challenge for the wider adoption of this technology. Based on the gathered data we thus recommend the development of additional education resources focusing on Geo Linked Data, jointly by Linked Data and Geodata experts. Finally, it is important to recognize that this list, while extensive, is not exhaustive; new resources will continue to emerge as the field evolves over time.