

FAIR DATA ENGINEERING – 2024-2025

# FAIR LOANS

## FAIRIFICATION OF LOAN DATA

KIONA BIJKER (S1979418)  
RAGUNATH SEENIVASAGAN ( S3482820)

07-11-2025

UNIVERSITY OF TWENTE.



# TABLE OF CONTENTS

1.	Introduction .....	3
2.	FAIRification planning .....	4
2.1	FAIRification goals .....	4
2.2	Target resource(s) .....	4
2.3	Target stakeholders .....	4
2.4	Reuse stakeholders .....	4
2.5	FAIRification requirements .....	5
2.6	Semantic types .....	5
2.7	Competency questions .....	6
3.	FAIRification process .....	6
3.1	initial fairness assessment .....	6
3.2	Semantic data model .....	7
3.3	Data triplification .....	9
3.4	Semantic metadata model .....	11
3.5	Metadata and data publication .....	12
3.6	Final FAIRness assessment .....	13
4.	discussions and Conclusions .....	14
5.	annexes .....	15
5.1	The metadata shacl .....	15
5.2	DATA SHACL .....	20

# 1. INTRODUCTION

Within this project we will be FAIRifying the ‘Lending Club’ dataset. This is a financial dataset of various loans containing 27 columns and a little over a million rows in a csv format. The dataset is available at <https://datasetsearch.research.google.com/search?query=lending%20club%20-site%3Akaggle.com&docid=L2cvMTFzand2ZHJzYg%3D%3D> and published under the CCO Public domain license.

We have some financial expertise in the team as Ragunath Seenivasagan used to be a Data Steward in the banking sector. To reduce workload we will filter to records including unique/rare values (such as ‘Joint App’) and add from the remaining rows at random. In this project we will be using 1000 rows and the following columns:

Name column	Description	Type
issue_d	Issue date of loan	Date
sub_grade	Grade awarded to person requesting loan, seems to determine interest rate	String, may be enumerable
term	Time to pay back loan	Integer “months”
home_ownership	Seems to be MORTGAGE or RENT	String, may be enumerable
fico_range_low	Low end of the credit score	Integer
pub_rec	Public records on person (criminal, bankruptcy, Integer etc. anything filed in courts)	
revol_util	Revolving credit line utilisation rate (use loan as source for other loan application or as a balance to take from as needed)	Decimal number
int_rate	Interest rate	Decimal number
application_type	Mainly ‘Individual’ some ‘Joint App’, describes number of people applying	String, may be enumerable
verification_status	Verified, Not Verified, or Source Verified. ID and other truth checks in loan application	String, may be enumerable
addr_state	State the registered address of applicant is located in	String, may be enumerable
fico_range_high	High end of the credit score	Integer
id	ID of loan application	Integer
emp_length	Duration of total employment	Integer
loan_status	Loan approved or not	1 or 0, boolean

## 2. FAIRIFICATION PLANNING

### 2.1 FAIRIFICATION GOALS

- **G1 (Findable):** Assign **W3ID persistent identifiers (PIPs)** for the dataset, its versions, and key entities such as loans and applicants. Publish indexable metadata with globally resolvable URIs to improve discoverability.
- **G2 (Accessible):** Release both **CSV** and **RDF/JSON-LD** representations via **HTTPS**, ensuring permanent open access to metadata. Maintain versioned snapshots for traceability and reproducibility.
- **G3 (Interoperable):** Convert the dataset into **RDF** using **OpenRefine (RDF extension)** and align its schema in **Protégé** with **FIBO-Loan** and **Schema.org** vocabularies for semantic interoperability.
- **G4 (Reusable):** Provide a **clear license (CC0)**, data card, and documentation of data quality and fairness metrics. Validate structural integrity and constraints using **SHACL** and publish the validation reports.

### 2.2 TARGET RESOURCE(S)

Resource	Description
train-lending-club-filtered1000.csv	Main dataset with loan information (issue date, sub-grade, term, etc.)
train-lending-club-filtered1000-csv.rdf	RDF/JSON-LD exports created with <b>OpenRefine RDF extension</b>
Protégé FAIR Data Group12.owx	Ontology designed in <b>Protégé</b> linking dataset fields to FIBO and schema.org
Shacl-train-lending-club-filtered1000-csv.ttl	<b>SHACL</b> constraints and validation results
FAIR_Loan_Dataset.ttl	FAIR metadata record describing the dataset
/.htaccess (w3id repo)	Redirect rules mapping w3id URIs to live dataset locations

### 2.3 TARGET STAKEHOLDERS

- **Data owners or stewards:** Course/research data stewards; dataset maintainers.
- **Data engineers:** People building features, pipelines, and models on Lending Club data.
- **Subject Matter Experts:** Credit risk analysts, economists.
- **Governance/compliance:** Research ethics boards, DPOs, EU AI Act/GDPR compliance teams
- **Educators or students:** Using the corpus for teaching ML/fairness

### 2.4 REUSE STAKEHOLDERS

- **Academia:** Fairness/bias researchers; reproducibility studies
- **Industry:** FinTech risk teams, explainability tooling vendors

- **Policy & audit:** Auditors/regulators evaluating model fairness & documentation
- **Open-source community:** Benchmark creators
- **Civic tech & NGOs:** Consumer-rights orgs studying credit access equity

## 2.5 FAIRIFICATION REQUIREMENTS

- Ensure that any columns containing personal or identifying information remain **pseudonymized**.
- Use **controlled vocabularies**, particularly **FIBO**, to describe financial and credit-related concepts.
- Apply **SHACL constraints** to validate required fields and datatype/range restrictions.
- Release all resources under the **CC0 license**, including clear provenance metadata.
- Maintain data accessibility through **HTTPS** and verify file integrity using checksums.
- Document all **OpenRefine transformation steps** in a provenance file for transparency and reproducibility.

## 2.6 SEMANTIC TYPES

**Classes:**

Class	Description
Loan_Application	A loan request/record containing terms, status, subgrade, interest rate, FICO, etc.
Applicant	The person associated with the loan application.
Subgrade	Letter–number credit grade used for pricing (e.g., A1, B3).
Fico_Range	Numeric FICO score bounds (low/high) tied to a loan/application.

**Data Properties:**

Property	Domain	Datatype	Description
ID	Loan_Application	string	Unique identifier for the loan/application.
issue_date	Loan_Application	date	Date the loan/application was issued.
term	Loan_Application	int	Term in months (e.g., 36, 60).
interest_rate	Loan_Application	decimal	Percentage rate (e.g., 13.49).
loan_status	Loan_Application	boolean	1 = approved/current; 0 = rejected/closed (as defined).
application_type	Loan_Application	string	“Individual”, “Joint”, etc.
revolving_utilisation	Loan_Application	decimal	Revolving utilization ratio/percent.
public_record	Loan_Application	int	Count of public records.
sub_grade	Subgrade	string	Code like A1, B3.
fico_range_low	Fico_Range	int	Lower FICO bound.

<b>fico_range_high</b>	Fico_Range	int	Upper FICO bound.
<b>employment_length</b>	Applicant	int	Years of employment.
<b>home_ownership_status</b>	Applicant	string	“MORTGAGE”, “RENT”, “OWN”, etc.
<b>verification_status</b>	Applicant	string	“Verified”, “Source Verified”, “Not Verified”.
<b>state</b>	Applicant	string	Two-letter state code (or region).

#### Object properties:

Property	Domain	Range	Description
<b>hasApplicant</b>	Loan_Application	Applicant	Link each loan/application to its applicant.
<b>hasEmployment</b>	Applicant	Record/concept	Use if you model employment as a node; else keep as data property <code>employment_length</code> .
<b>hasFicoRange</b>	Loan_Application	Fico_Range	Connects a loan/application to its FICO bounds.
<b>hasHomeOwnershipStatus</b>	Applicant	record	Use if modeling status as a node; else use data property <code>home_ownership_status</code> .
<b>hasInterestRate</b>	Loan_Application	<i>subgrade</i>	Optional conceptual link; numeric value is <code>interest_rate</code> .
<b>hasSubgrade</b>	Loan_Application	Subgrade	Assigns the subgrade (A1...G5).
<b>influencedBy</b>	Subgrade	Fico_Range	Optional: capture that subgrade is influenced by FICO.

## 2.7 COMPETENCY QUESTIONS

1. How can a user query for all **approved loans** (`loan_status = 1`) issued after a specific date?
2. What is the **average interest rate** grouped by `sub_grade` or `application_type`?
3. How can we retrieve the **FICO score range** and associated `interestRate` for a given loan?

## 3. FAIRIFICATION PROCESS

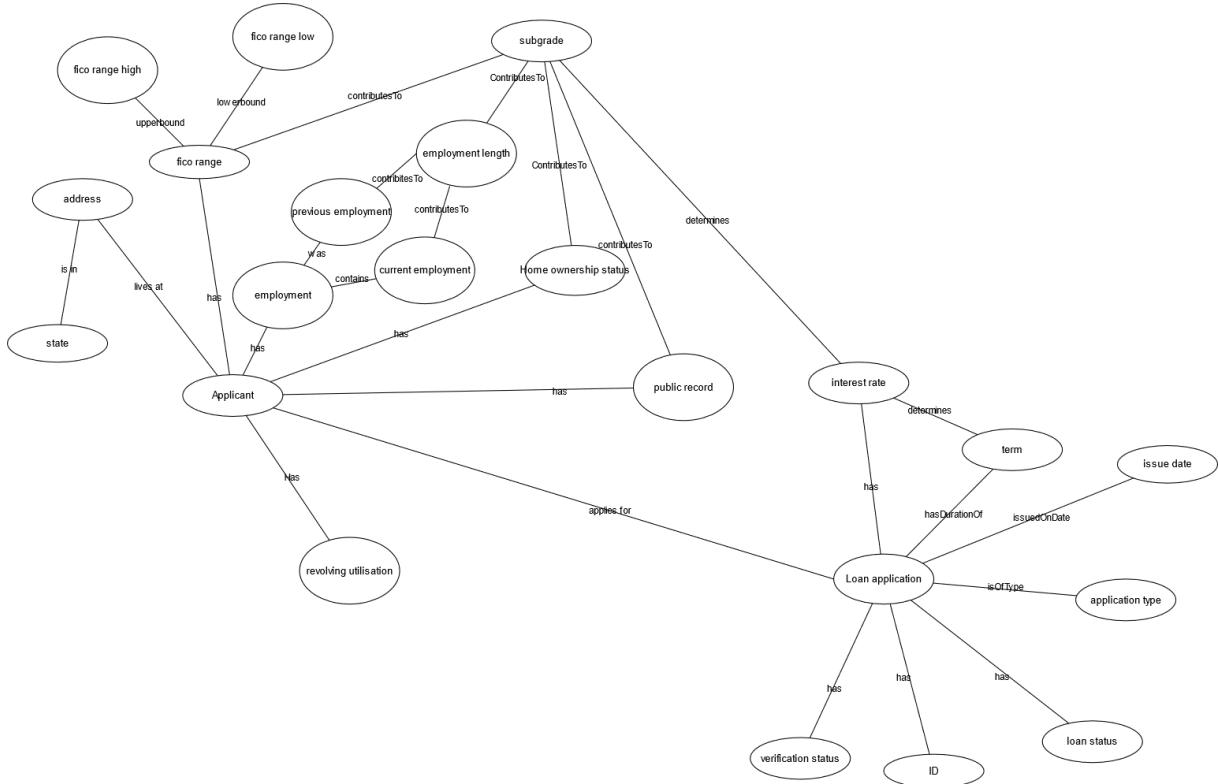
### 3.1 INITIAL FAIRNESS ASSESSMENT

Principle #	Criteria	Conformance	Note
F1	Globally unique identifier	partially	Identifier is unique within dataset but likely

			no longer unique when combining data across banks
F1	Persistent identifier	no	Unknown, no persistence policy given
F2	Has metadata	yes	Found on kaggle and figshare
F2	Has rich metadata	partially	Provides license and some provenance
F3	Has data identifier in metadata	yes	Link provided
F3	Has data identifier with clear predicate in metadata	no	Just DOI under "unique identifier" predicate
F4	Indexed/registered in search resource	yes	Google dataset search
A1	Has resolution protocol	yes	HTTP
A1.1	Protocol is open, free and universally implementable	yes	
A1.2	Protocol has authentication and authorisation mechanisms	yes	HTTPS has these
A2	Metadata has longer persistency than data	no	There is no clear policy to retain metadata after the dataset is no longer available
I1	(meta)data	no	Column names are unclear, do seem to be somewhat common use

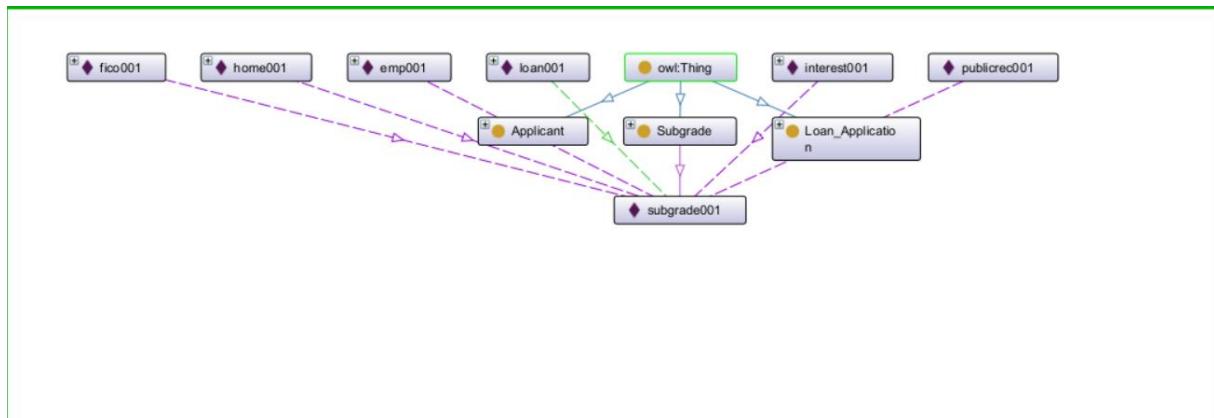
### 3.2 SEMANTIC DATA MODEL

Based on our joint knowledge of the domain we came up with the following initial sketched model (made with draw.io).

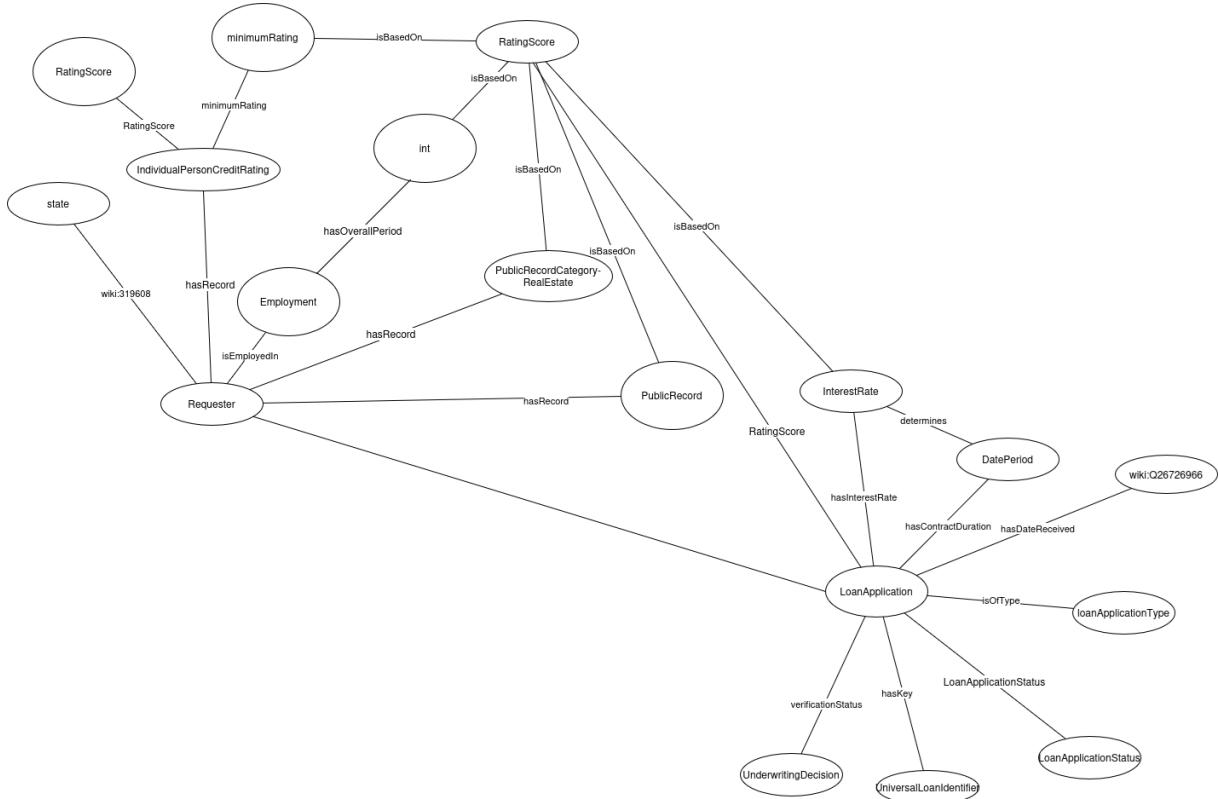


As existing vocabularies we used FIBO, especially the loan applications module (<https://spec.edmcouncil.org/fibo/ontology/LOAN/LoansGeneral/LoanApplications/>) and wikidata. The exact links can be found in the data shacl in the annexes as well as the rdf.

We added the following structure using Protege to extend the existing vocabularies.



Using this, below is the updated semantic model (again made with draw.io) using the ontology names. For readability the hasValue fields and the full links have been left out.



### 3.3 DATA TRIPLIFICATION

Within the triplification we did not include the current and previous employment as there we could not find URIs for such relations. Instead employment directly connects to employment length.

See the OpenRefine screenshots below for triplification.

Available prefixes: rdf owl fibo\_loan wiki rdft tibo [+Add](#) [Manage](#)

(Row index) URI	Operations	Description	Type
<a href="#">X fibo_loan:LoanApplication</a>	<a href="#">X ➞ owl:hasKey ➔</a>	<a href="#">fibo:LOAN/RealEstateLoans/HomeMo... Add type</a>	<a href="#">id Cell</a>
	<a href="#">X ➞ fibo:FND/DatesAndTimes/Financial... ➔</a>	<a href="#">wiki:Q26726966 Add type</a>	<a href="#">issue_d Cell</a>
	<a href="#">X ➞ :verificationStatus ➔</a>	<a href="#">fibo:LOAN/LoansGeneral/LoanAppli... Add type</a>	<a href="#">Add property verification_status owl:hasValue Cell</a>
	<a href="#">X ➞ fibo_loan:LoanApplicationStatus ➔</a>	<a href="#">fibo:LOAN/LoansGeneral/LoanAppli... Add type</a>	<a href="#">loan_status Cell</a>

Available prefixes:	rdt owl tibo_loan wkl1 rdts tibo	+Add	Manage
		Add type	Cell
X → fibo_loan:LoanApplicationStatus →	□ fibo:LOAN/ LoansGeneral/ LoanAppli... Add type Add type	□ X → owl:hasValue → □ loan_status Cell	
X → :isOfType →	□ loanApplicationType Add type	□ X → owl:hasValue → □ application_type Cell → Add property	
X → fibo:FND/Agreements/Contracts/ha... →	□ https:// www.omg.org/ spec/Commons... Add type Add type	□ X → owl:hasValue → □ term Cell	
X → fibo:FBC/DebtAndEquities/Debt/ha... →	□ fibo:FND/ Accounting/ CurrencyAmou... Add type Add type	□ X → owl:hasValue → □ int_rate Cell	



### 3.4 SEMANTIC METADATA MODEL

See the metadata SHACL in the annex.

For the metadata we assumed that each dataset for loans will have differing amounts of information available.<sup>1</sup> Next to the standard UI language, license and dates available we added the following fields:

Contact person (name & email)	This information allows a future user to contact the data steward for this data set with questions and issues regarding the dataset. We assume no dataset metadata form will be perfect for each use of the dataset. A way to contact a person who may have or know how to find the information for such uses, or can say it does not exist, saves future users time and guesswork.
Semantic model (URI)	Here the semantic model of the dataset can be linked. This improves accessibility and interoperability. By linking the model it does not have to be a

<sup>1</sup>We also endeavored to make the metadata form accessible to differing levels of technical know-how. While there is space for github links, official vocabularies, shacl files, and so on; it is also possible to upload a new dataset with text explanations and a public google drive link to the model.

	Turtle or XML file but can also be a picture of a hand drawn graph.
Data source (URI)	An optional field for datasets based on other datasets. This is the link to the dataset pre-modification
Data provenance (text)	An explanation of how and why the dataset was created or modified. A text field allows for links without making a link to an outside project necessary.
Vocabularies used (URI)	This field adds context to the dataset and its semantic model. Adding the vocabularies used by the dataset also makes possible connections to other datasets (interoperability) more clear.
Data sample (text)	A simple excerpt from a record in the dataset. This gives a user a case to work from when programming.
Sample queries (text)	Example queries the database could answer.
Data management policy (text)	Indicates how long data and metadata will be available and how they may be used or changed over time. This can be a plain text explanation or a link to an external document
AI flag (Boolean)	If true the data is synthetic data
Language(s) of the data (URI)	The language(s) used within the data restrict who can understand both the records and query results.
Extra information (text)	A complete form is rare. This text space allows the dataset uploader to add any notable information that did not fit other categories.

## 3.5 METADATA AND DATA PUBLICATION

### Metadata Record:

The metadata describing the dataset is published in RDF (Turtle) format and is accessible through the FAIR Data Point and GitHub repository.

- GitHub metadata file:  
 <https://github.com/KionaB/FAIRscripts/tree/785e84beccf7b95cef719e16d544e0e0fdbf318c/FAIR%20Data%20Point>

### Dataset Publication:

The FAIRified dataset is openly available on GitHub and can be accessed using the persistent W3ID link.

- Dataset file on GitHub:  
 <https://github.com/KionaB/FAIRscripts/tree/785e84beccf7b95cef719e16d544e0e0fdbf318c/fairified/data>

### Notes:

- The .htaccess file in the W3ID namespace ensures persistent redirection to these resources.
- Metadata and data are publicly accessible and linked through globally unique identifiers.
- Both are released under the **CC0 license**, ensuring unrestricted reuse.

### 3.6 FINAL FAIRNESS ASSESSMENT

Principle #	Criteria	Conformance pre FAIRification	Conformance post FAIRification	Note
F1	Globally unique identifier	partially	partially	The data uses identifiers unique within the dataset. The dataset has a globally unique identifier
F1	Persistent identifier	no	yes	W3id was added to provide globally resolvable, persistent identifiers that remain valid even if the data's location changes
F2	Has metadata	yes	yes	Metadata is available and maintained; it provided descriptive information for discovery and reuse
F2	Has rich metadata	partially	yes	Rich metadata was created including context, schema and FAIR-compliant descriptive properties, improving findability
F3	Has data identifier in metadata	yes	yes	The metadata explicitly includes the dataset's unique identifier, ensuring proper linkage between data and metadata
F3	Has data identifier with clear predicate in metadata	no	yes	Added htaccess file and a readme explaining the namespace and linking predicates for identifiers
F4	Indexed/registered in search resource	yes	yes	The dataset is published in Github, which allows indexing and discovery via search engines and metadata registries
A1	Has resolution protocol	yes	yes	The dataset can be accessed over HTTP/HTTPS, providing a standardized resolution protocol
A1.1	Protocol is open, free and universally implementable	yes	yes	HTTP/HTTPS protocol is open and universally supported across platforms and clients
A1.2	Protocol has authentication and authorisation mechanisms	yes	yes	HTTPS supports authentication and authorization where

				required, ensuring controlled data access
A2	Metadata has longer persistency than data	no	yes	Metadata is preserved and accessible independently (via GitHub/W3iD) even if data is removed or relocated
I1	(meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation	no	yes	Metadata and data were represented using RDF/Turtle and validated via SHACL, ensuring interoperability
I2	(meta)data use vocabularies that follow FAIR principles;	no	partially	Controlled vocabularies from FIBO and Wikidata were reused, improving semantic alignment though not to completion. Whether the vocabularies fully follow FAIR is unclear.
I3	(meta)data include qualified references to other (meta)data;	no	partially	The vocabularies used allow for easy referencing and linking to data within the same domain
R1	(meta)data are richly described with a plurality of accurate and relevant attributes	partially	yes	SHACL shapes and extended metadata descriptions were added to provide more accurate and machine-readable attributes
R1.1	(meta)data are released with a clear and accessible data usage license	Yes (CC0)	Yes (CC0)	The dataset and metadata were released under the CC0 license, ensuring open reuse without restrictions
R1.2	(meta)data are associated with detailed provenance;	partially	yes	Provenance information (eg., source repositories from kaggle and Figshare) was included to trace data origin and transformations

## 4. DISCUSSIONS AND CONCLUSIONS

In this project we FAIRified the Lending Club Loan dataset of loans. This dataset had a CC0 license which we preserved. The existing provenance was also preserved. We added semantic structure to the data using the Financial Industry Business Ontology (FIBO) and WikiData vocabularies. Since the dataset's relations and entities were not all fully covered by these vocabularies some concepts were added using Protégé. Finding these vocabularies and deciding when to define our own concepts was one of our most significant challenges. While loan concepts, especially mortgages, are covered in many ontologies, concepts for loan applications are less so. When we found FIBO we found an

ontology that covered loan applications but did not fully cover the specific way Lending Club set up their database. Especially the loan application types (joint or individual) and the lack of requester ID made finding a fitting vocabulary difficult. By using FIBO and WikiData the FAIRified dataset has improved interoperability and semantic clarity.

With the combined vocabulary the dataset could be triplified, creating a linked dataset. Linking the data has made the relations between entities machine readable. To improve the findability for future use we implemented globally unique and persistent identifiers using a W3ID namespace (<https://w3id.org/FAIR-course-UT/2025/group12/main>). The dataset had internally unique identifiers for loan applications, not for requesters. Since the requesters were not clearly identified we could not add unique identifiers for them. We did add metadata identifiers linking the data and metadata together so the data can be found from the metadata. The W3ID namespace guarantees stable long-lasting identifiers that can be redirected to where the data is. We did so through an htaccess file meaning that even if the location of the data or metadata changes the W3ID identifiers will still work (simply redirecting to a new location).

To ensure the structure of both the data and the metadata we created a shacl definition for each. The metadata shacl ensures that the metadata for the dataset is consistent with each following version of it. The form defined by the metadata shacl provides a structure for rich metadata which results in a machine-readable Turtle file of the metadata without a need to write such a file. The shacl of the dataset allows for adding records to the dataset that follow the FAIRified structure of the current dataset. This way the current 1000 records can be extended.

The FAIRification process for this dataset was challenging in ways we did not foresee when we started. We learned that structuring the data was relatively easy once we had the vocabularies we needed. Trying to find those vocabularies showed us why so many datasets use a new vocabulary specific to that dataset or no (official) vocabulary at all. We also saw how much work goes into FAIR data maintenance including updating redirects and keeping the metadata current enough. Through our FAIRification process we gained a better understanding of Linked Data technologies and how to create a machine-readable and usable dataset. We also learned that a FAIRified dataset does not need to be perfectly FAIR, it rarely can be if based on an existing dataset. Even though our FAIRified dataset does not have globally unique identifiers on each level, nor references other datasets; we still improved this dataset significantly.

## 5. ANNEXES

### 5.1 THE METADATA SHACL

```
@prefix sh: <http://www.w3.org/ns/shacl#> .  
@prefix schema: <http://schema.org/> .  
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .  
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .  
@prefix dash: <http://datashapes.org/dash#> .  
@prefix foaf: <http://xmlns.com/foaf/0.1/> .  
@prefix vcard: <http://www.w3.org/2006/vcard/ns#> .  
@prefix ex: <http://example.com/> .  
@prefix lexvo: <http://lexvo.org/id/iso639-1/> .  
@prefix fair: <http://fairdatapoint.org/> .  
@prefix dash: <http://datashapes.org/dash#> .  
@prefix dct: <http://purl.org/dc/terms/> .  
@prefix fdp: <https://w3id.org/fdp/fdp-o#> .  
  
fair:FDPShape  
a sh:NodeShape ;
```

```

sh:targetClass fdp:FAIRDataPoint ;
    rdfs:label "Fair Data Point metadata";
sh:property fair:StartDateProperty,
            fair:EndDateProperty,
            fair:UILanguageProperty,
            fair:MetaDatalIdentifierProperty,
            fair:MetaDatalIssuedProperty,
            fair:MetaDatalModifiedProperty,
            fair:ContactPersonProperty,
            fair:SemanticModelProperty,
            fair:DataSourceProperty,
            fair:DataProvenanceProperty,
            fair:DataVocabularyProperty,
            fair:DataSampleProperty,
            fair:DataSampleQueriesProperty,
            fair:DataManagementPolicyProperty,
            fair:DataAlggeneratedFlagProperty,
            fair:DataLanguageProperty,
            fair:LicenseProperty,
            fair:ExtraInformationProperty;

.

#Groups are used to cluster questions together
fair:ContactPropertyGroup
    a sh:PropertyGroup ;
    rdfs:label "Contact";

.

fair:ContactPersonProperty
    a sh:PropertyShape ;
    sh:group fair:ContactPropertyGroup ;
    sh:minCount 0;
    dash:editor dash:DetailsEditor ;
        dash:viewer dash:DetailsViewer ;
        sh:name "Contact Person" ;
        sh:node fair:ContactPersonDeets ;
        sh:nodeKind sh:BlankNode .

fair:ContactPersonDeets
    a sh:NodeShape ;
    rdfs:label "Value with weight" ;
    sh:property fair:NameProperty, fair:EmailProperty;

.

#E-mail of the contact person for the data
fair:EmailProperty
    a sh:PropertyShape ;
    sh:path schema:email;
    rdfs:label "E-mail address";
    sh:datatype xsd:string ;
    dash:editor dash:TextFieldEditor ;
    dash:viewer dash:LiteralViewer ;
        sh:minCount 0 ;
    sh:order 45;

.

#Name of the contact person for the data
fair:NameProperty

```

```

a sh:PropertyShape ;
sh:path schema:name ;
rdfs:label "Name";
sh:datatype xsd:string ;
dash:singleLine true ;
dash:editor dash:TextFieldEditor ;
dash:viewer dash:LiteralViewer ;
sh:maxCount 1 ;
sh:minCount 0 ;
sh:order 44 ;

.

fair:MetaDataPropertyGroup
a sh:PropertyGroup ;
rdfs:label "Metadata";

.

fair:StartDateProperty
sh:path fdp:startDate ;
sh:group fair:MetaDataPropertyGroup;
sh:datatype xsd:date ;
sh:maxCount 1 ;
dash:editor dash:DatePickerEditor ;
dash:viewer dash:LiteralViewer ;
sh:order 40 ;

.

fair:EndDateProperty
sh:path fdp:endDate ;
sh:group fair:MetaDataPropertyGroup;
sh:datatype xsd:date ;
sh:maxCount 1 ;
dash:editor dash:DatePickerEditor ;
dash:viewer dash:LiteralViewer ;
sh:order 41 ;

.

#UI language for FDP
fair:UILanguageProperty
sh:path fdp:uiLanguage ;
sh:group fair:MetaDataPropertyGroup;
sh:nodeKind sh:IRI ;
sh:maxCount 1 ;
sh:defaultValue <http://id.loc.gov/vocabulary/iso639-1/en>;
dash:editor dash:URIEditor ;
dash:viewer dash:LabelViewer ;
sh:order 42 ;

.

#ID that resolves to the metadata
fair:MetaDatalIdentifierProperty
sh:path fdp:metadatalIdentifier ;
sh:group fair:MetaDataPropertyGroup;
sh:nodeKind sh:IRI ;
sh:maxCount 1 ;
dash:editor dash:URIEditor ;
dash:viewer dash:LabelViewer ;

```

```

sh:order 43 ;
.

#When this metadata was uploaded
fair:MetaDatasubmittedProperty
  sh:path fdp:metaDataSubmitted ;
  sh:group fair:MetaDataProviderGroup;
  sh:datatype xsd:dateTime ;
  sh:maxCount 1 ;
  dash:viewer dash:LiteralViewer ;

.

#When the metadata was last modified
fair:MetaDataModifiedProperty
  sh:path fdp:metaDataModified ;
  sh:group fair:MetaDataProviderGroup;
  sh:datatype xsd:dateTime ;
  sh:maxCount 1 ;
  dash:viewer dash:LiteralViewer ;

.

fair:DataPropertyGroup
  a sh:PropertyGroup ;
  rdfs:label "Data properties";

.

#License info
fair:LicenseProperty
  sh:path fair:licenseProperty;
  rdfs:label "Data license";
  rdfs:comment "Add a license to the data";
  sh:group fair:DataPropertyGroup;
  sh:minCount 1;
  sh:datatype xsd:string ;
  dash:editor dash:TextFieldEditor ;
  dash:viewer dash:LiteralViewer ;

.

#link to the semantic model(s)
fair:SematicModelProperty
  sh:path fair:semanticModel;
  rdfs:label "Semantic model link";
  rdfs:comment "The semantic model for the data and/or the domain of the data";
  sh:group fair:DataPropertyGroup;
  sh:minCount 0;
  sh:datatype xsd:anyURI ;
  dash:editor dash:URIEditor ;
  dash:viewer dash:HyperlinkViewer ;

.

#link to the source of the data
fair:DataSourceProperty
  sh:path dct:source;
  rdfs:label "Data source";
  rdfs:comment "Source of the data before current modification";
  sh:group fair:DataPropertyGroup;
  sh:datatype xsd:string ;
  dash:editor dash:URIEditor ;

```

```
dash:viewer dash:HyperlinkViewer ;  
.  
  
#description of how dataset came to be / was changed  
fair:DataProvenanceProperty  
    sh:path fair:provenance;  
    rdfs:label "Data provenance";  
    rdfs:comment "How was this data obtained and/or modified";  
    sh:group fair:DataPropertyGroup;  
    sh:datatype xsd:string ;  
    dash:editor dash:TextAreaEditor ;  
    dash:viewer dash:LiteralViewer ;  
.  
  
#List of vocabularies used  
fair:DataVocabularyProperty  
    sh:path fair:vocabularies;  
    rdfs:label "Vocabularies used";  
    rdfs:comment "Links to any vocabularies used in the data or metadata ";  
    sh:group fair:DataPropertyGroup;  
    sh:datatype xsd:anyURI ;  
    sh:minCount 0;  
    dash:editor dash:URIEditor ;  
    dash:viewer dash:URIViewer ;  
.  
  
#Sample of data  
fair:DataSampleProperty  
    sh:path fair:dataSample;  
    rdfs:label "Sample of data";  
    sh:group fair:DataPropertyGroup;  
    sh:data xsd:string ;  
    dash:editor dash:TextAreaEditor ;  
    dash:viewer dash:LiteralViewer ;  
.  
  
#Example queries  
fair:DataSampleQueriesProperty  
    sh:path fair:sampleQueries;  
    rdfs:label "Example queries";  
    rdfs:comment "Examples of queries this dataset answers and/or how to query the dataset";  
    sh:group fair:DataPropertyGroup;  
    sh:datatype xsd:string ;  
    dash:editor dash:TextAreaEditor ;  
    dash:viewer dash:LiteralViewer ;  
.  
  
#Data management policy  
fair:DataManagementPolicyProperty  
    sh:path fair:managementPolicy;  
    rdfs:label "Data management policy";  
    rdfs:comment "Describe or provide a link to the data management policy";  
    sh:group fair:MetaDataTablePropertyGroup;  
    sh:datatype xsd:string ;  
    dash:editor dash:TextAreaEditor ;  
    dash:viewer dash:LiteralViewer ;  
.
```

```

#AI generated flag
fair:DataAlgeneratedFlagProperty
  sh:path fair:aiFlag;
  rdfs:label "AI generated/synthetic data";
  sh:group fair:MetaDataPropertyGroup;
  sh:datatype xsd:boolean ;
  sh:minCount 1;
  sh:maxCount 1;
  dash:editor dash:BooleanSelectEditor ;
  dash:viewer dash:LiteralViewer ;

.

#Language(s) of the data
fair:DataLanguageProperty
  sh:path fair:dataLanguage;
  rdfs:label "Language(s) used in the data";
  sh:group fair:DataPropertyGroup;
  sh:nodeKind sh:IRI ;
  sh:minCount 0;
  sh:defaultValue <http://id.loc.gov/vocabulary/iso639-1/en>;
  dash:editor dash:URIEditor ;
  dash:viewer dash:LabelViewer ;

.

#Space for additional information
fair:ExtraInformationProperty
  sh:path fair:extraInformation;
  rdfs:label "Additional information about the data";
  rdfs:comment "Any additional information that may be useful to future users of this dataset";
  sh:group fair:DataPropertyGroup;
  sh:datatype xsd:string ;
  dash:editor dash:TextAreaEditor ;
  dash:viewer dash:LiteralViewer ;

.

```

## 5.2 DATA SHACL

```

@prefix sh: <http://www.w3.org/ns/shacl#> .
@prefix schema: <http://schema.org/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix ex: <http://example.org/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix fibo_loan: <https://spec.edmcouncil.org/fibo/ontology/LOAN/LoansGeneral/LoanApplications/> .
@prefix wiki: <https://www.wikidata.org/wiki/> .
@prefix fibo: <https://spec.edmcouncil.org/fibo/ontology/> .
@prefix local: <http://localhost:3333/> .

local:0 a fibo_loan:LoanApplication;
ex:isVerified "Not Verified".
fibo:Requester a rdfs:Class, sh:NodeShape ;
  sh:targetClass <fibo:FND/Arrangements/Reporting/Requester> ;
  sh:name "Applicant" ;

```

```

sh:property [
    sh:path
    <https://spec.edmcouncil.org/fibo/ontology/FND/Arrangements/Documents/hasRecord>;
    sh:minCount 3 ;
    sh:maxCount 3 ;
    sh:property [
        sh:path <fibo_loan:PublicRecordCategory-RealEstate>;
        sh:name "Home ownership type" ;
        sh:minCount 1 ;
        sh:maxCount 1 ;
        sh:datatype xsd:string ;
        sh:in (
            "MORTGAGE"
            "OWN"
            "RENT"
        )
    ];
    sh:property [
        sh:path fibo_loan:IndividualPersonCreditRating ;
        sh:minCount 1 ;
        sh:maxCount 1 ;
        sh:property [
            sh:path <fibo:BP/SecuritiesIssuance/AgencyMBSIssuance/minimumRating>;
            sh:name "Minimum fico value" ;
            sh:minCount 1 ;
            sh:maxCount 1 ;
            sh:datatype xsd:int ;
            sh:minInclusive 300 ;
            sh:maxInclusive 850];
        ];
        sh:property [
            sh:path <fibo:FND/Arrangements/Ratings/RatingScore>;
            sh:name "Maximum fico value" ;
            sh:minCount 1 ;
            sh:maxCount 1 ;
            sh:property [
                sh:path owl:hasValue ;
                sh:minCount 1 ;
                sh:maxCount 1 ;
                sh:datatype xsd:int ;
                sh:minInclusive 300 ;
                sh:maxInclusive 850;
            ];
        ];
    ];
], [
    sh:path <fibo:FND/Organizations/FormalOrganizations/Employment>;
    sh:name "Total employment duration in years" ;
    sh:minCount 1 ;
    sh:maxCount 1 ;
    sh:property [

```

```

        sh:path owl:hasValue ;
        sh:minCount 1 ;
        sh:maxCount 1 ;
        sh:datatype xsd:int ;
        sh:minInclusive 0 ;
    ];
```

[, [

```

        sh:path wiki:319608 ;
        sh:name "Address state" ;
        sh:minCount 1 ;
        sh:maxCount 1 ;
        sh:datatype xsd:string ;
```

], [

```

        sh:path fibo_loan:PublicRecord ;
        sh:name "Number of public records" ;
        sh:minCount 1 ;
        sh:maxCount 1 ;
        sh:property [
            sh:path owl:hasValue ;
            sh:minCount 1 ;
            sh:maxCount 1 ;
            sh:datatype xsd:int ;
            sh:minInclusive 0 ;
        ];
```

], [

```

        sh:path fibo_loan:LoanApplication;
        sh:name "Application" ;
        sh:minCount 1 ;
        sh:nodeKind fibo_loan:LoanApplication ;
        sh:class fibo_loan:LoanApplication ;
    ].
```

fibo\_loan:LoanApplication a rdfs:Class, sh:NodeShape ;

```

    sh:targetClass fibo_loan:LoanApplication ;
    sh:name "Loan application" ;
    sh:property [
        sh:path owl:hasKey;
        sh:name "Loan ID" ; #TODO
        sh:minCount 1 ;
        sh:maxCount 1 ;
    ], [
        sh:path
<https://spec.edmcouncil.org/fibo/ontology/FND/DatesAndTimes/FinancialDates/hasDateReceived> ;
        sh:name "Loan issue date" ;
        sh:minCount 1 ;
        sh:maxCount 1 ;
    ], [
        sh:path local:verificationStatus ;
        sh:name "Verification status" ;
        sh:minCount 1 ;
        sh:maxCount 1 ;
    ], [
        sh:path fibo_loan:LoanApplicationStatus ;
        sh:name "Loan approval status" ;
        sh:minCount 1 ;
```

```
    sh:maxCount 1 ;
], [
    sh:path local:isOfType ;
    sh:name "Loan application type" ;
    sh:minCount 1 ;
    sh:maxCount 1 ;
], [
    sh:path
<https://spec.edmcouncil.org/fibo/ontologyFND/Agreements/Contracts/hasContractDuration> ;
    sh:name "Loan term" ;
    sh:minCount 1 ;
    sh:maxCount 1 ;
], [
    sh:path
<https://spec.edmcouncil.org/fibo/ontologyFBC/DebtAndEquities/Debt/hasInterestRate> ;
    sh:name "Loan interest rate" ;
    sh:minCount 1 ;
    sh:maxCount 1 ;
], [
    sh:path <https://spec.edmcouncil.org/fibo/ontologyFND/Arrangements/Ratings/RatingScore>
;
    sh:name "Loan sub-grade" ;
    sh:minCount 1 ;
    sh:maxCount 1 ;
        sh:in (
            local:A1
            local:A2
            local:A3
            local:A4
            local:A5
            local:B1
            local:B2
            local:B3
            local:B4
            local:B5
            local:C1
            local:C2
            local:C3
            local:C4
            local:C5
            local:D1
            local:D2
            local:D3
            local:D4
            local:D5
            local:E1
            local:E2
            local:E3
            local:E4
            local:E5
            local:F1
            local:F2
            local:F3
            local:F4
            local:F5
            local:G1
            local:G2
```

```

        local:G3
        local:G4
        local:G5
    );
.

<fibo:LOAN/RealEstateLoans/HomeMortgageDisclosureActCoveredMortgages/UniversalLoanIdentifier> a
rdfs:Class , sh:NodeShape ;
    sh:property [ sh:path owl:hasValue;
                    sh:minCount 1 ;
                    sh:maxCount 1 ;
                    sh:datatype xsd:int;
                ].

<https://spec.edmcouncil.org/fibo/ontology/FND/DatesAndTimes/FinancialDates/hasDateReceived> a
rdfs:Class , sh:NodeShape ;
    sh:name "Loan issue date";
    sh:property [
                    sh:path <wiki:Q26726966>;
                    sh:minCount 1 ;
                    sh:maxCount 1 ;
                ].

<wiki:Q26726966> a rdfs:Class ,sh:NodeShape;
    sh:property [
                    sh:path owl:hasValue ;
                    sh:minCount 1 ;
                    sh:maxCount 1 ;
                    sh:datatype
xsd:dateTime] .

<fibo:LOAN/LoansGeneral/LoanApplications/UnderwritingDecision> a rdfs:Class , sh:NodeShape ;
    sh:name "Verification status" ;
    sh:property [
                    sh:path owl:hasValue ;
                    sh:minCount 1 ;
                    sh:maxCount 1 ;
                    sh:datatype xsd:string ;
                    sh:in (
                            "Not Verified"
                            "Source Verified"
                            "Verified"
                        )];
.

fibo_loan:LoanApplicationStatus a rdfs:Class , sh:NodeShape ;
    sh:name "Loan approval status" ;
    sh:property [sh:path owl:hasValue ;
                    sh:minCount 1 ;
                    sh:maxCount 1 ;
                    sh:datatype xsd:boolean ;].

local:loanApplicationType a rdfs:Class , sh:NodeShape ;
    sh:name "Loan application type" ;
    sh:property [
                    sh:path owl:hasValue ;
                    sh:minCount 1 ;
                    sh:maxCount 1 ;
                    sh:datatype xsd:string ;
                    sh:in (
                            "Individual"
                            "Joint App"
                        )
                ]
.
```

```
) ;  
].  
  
<https://spec.edmcouncil.org/fibo/ontologyFND/Agreements/Contracts/hasContractDuration> a rdfs:Class ,  
sh:NodeShape ;  
    sh:name "Loan term" ;  
    sh:property [  
        sh:path owl:hasValue ;  
        sh:minCount 1 ;  
        sh:maxCount 1 ;  
        sh:datatype xsd:int ;  
        sh:minInclusive 1 ;  
    ].  
  
<https://spec.edmcouncil.org/fibo/ontologyFBC/DebtAndEquities/Debt/hasInterestRate> a rdfs:Class ,  
sh:NodeShape ;  
    sh:name "Loan interest rate" ;  
    sh:property [  
        sh:path owl:hasValue ;  
        sh:minCount 1 ;  
        sh:maxCount 1 ;  
        sh:datatype xsd:double ;  
        sh:minInclusive "0.0"^^xsd:double ;  
    ].
```