

**Toward an Implementable Framework of the FAIR Data Principles for
Earth Science Data Management and Stewardship**

A Dissertation

Presented in Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

with a

Major in Computer Science

in the

College of Graduate Studies

University of Idaho

by

Abdullah N. Alowairdhi

Major Professor: Xiaogang Ma, Ph.D.


Committee Members: Clint Jeffery, Ph.D.; Paul Gessler, Ph.D.; Jia Song, Ph.D.

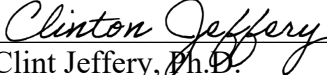
Department Administrator: Terence Soule, Ph.D.

December 2020

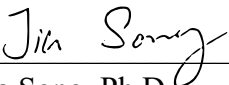
Authorization to Submit Dissertation


This dissertation of Abdullah N. Alowairdhi, submitted for the degree of Doctor of Philosophy with a Major in Computer Science and titled "Toward an Implementable Framework of The FAIR Data Principles for Earth Science Data Management and Stewardship," has been reviewed in final form. Permission, as indicated by the signatures and dates below, is now granted to submit final copies to the College of Graduate Studies for approval.

Major Professor:  Date: 12/03/2020
Xiaogang Ma, Ph.D.

Committee Members:  Date: 12/3/2020
Clint Jeffery, Ph.D.

 Date: 12/4/2020
Paul Gessler, Ph.D.

 Date: 12/04/2020
Jia Song, Ph.D.

Department Administrator:  Date: Dec. 7 2020
Terence Soule, Ph.D.

Abstract

This Ph.D. research aims to improve the data management and stewardship for Earth science digital resources by developing an implementable framework for the FAIR data principles. The lack of a pragmatic framework to facilitate the translation of FAIR data principles into the digital world has led to a gap between the theories and implementation of those principles for Earth science data stewardship. To overcome this challenge, we ask four themed questions to guide the research activities: First, how can we verify the validity and tautology of FAIR data principles? Second, how can we theoretically address FAIR data principles? Third, how can we technically approach FAIR data principles? Fourth, how can we efficiently evaluate the FAIRness of a digital resource? The two formal logical methods used in this work are the Truth Table and the Natural Deduction. The development method used is semantic web technologies supported by a FAIR ontology. Furthermore, the FAIRness level evaluation method used is a Fuzzy logic method. We show that FAIR data principles are valid and tautological, which resulted in the formulation of FAIR theorems. This research is the first research that implements formal logic to verify the FAIR data principles and uses fuzzy logic to assess the FAIRness level, which helps set up a bridge between the human conceptualization and the machine implementation of the FAIR data principles. We also show the prototype of FAIRtool.org, a semantic web application that adopts FAIR data principles, and the creation of the Fuzzy FAIR Assessment Framework (FFAF). The development of the FAIR theorems establishes rules to translate the FAIR data principles into machine-readable formats, which are necessary for the implementation of FAIR in the cyberinfrastructure. Using the FFAF model to assess the

FAIRness of a digital resource led to an efficient FAIRness level evaluation. We demonstrated the outputs of this research with two examples from Earth science, the “NCDC Storm Events Database” use case and the “Data for Building an Open Science Framework to Model Soil Organic Carbon” use case. The Earth science community is actively promoting the adoption and implementation of FAIR principles. This Ph.D. research provides evidence about the logic validity of FAIR principles. The pilot system and examples show the implementability of FAIR principles in the cyberinfrastructure for various datasets and other digital resources. With more work and community of practice, this advancement in cyberinfrastructure will eventually promote the precision of Earth science data management and stewardship to a new level.

Acknowledgements

I sincerely appreciate and thank the Almighty الله for his graces, his might, his sustenance and, above all, his love from the beginning of my life to this level and beyond. His care helped me excel and succeed in all my life's endeavors.

I would like to express my special appreciation and my thanks to my advisor Major Professor Dr. Xiaogang Ma; you have been a tremendous mentor for me. I would like to thank you for encouraging my research and for allowing me to grow as a research scientist. I would also like to thank my committee members Professor Clint Jeffery, Professor Paul Gessler, and Professor Jia Song for serving as my committee members and for their valuable comments and suggestions. I also want to thank all who supported my research, notably ESIP for encouraging me with their seed grant.

A special thanks to my family. Words cannot express how grateful I am to my family for all the sacrifices that they have made on my behalf. Special thanks to my parents, whose love and guidance are with me in whatever I pursue. They are the ultimate role models. Most importantly, I wish to thank my loving and supportive wife, Amani, and my wonderful children Dema, Nasser, Mohammad, Raghad, Lamia, and Omar, who provide continual inspiration.

Dedication

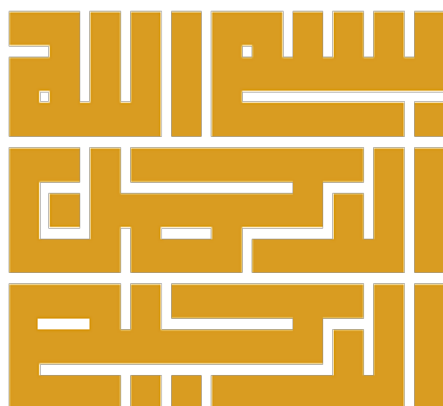


Table of Contents

Authorization to Submit Dissertation	ii
Abstract	iii
Acknowledgements	v
Dedication	vi
Table of Contents	vii
List of Figures	ix
List of Tables	x
List of Listings	xi
Chapter 1: Introduction	1
1.1 General description	1
1.2 FAIR data principles	1
1.3 Data stewardship	3
1.4 Thesis statement	3
1.5 Motivation and problem statement.....	4
1.6 Contributions.....	5
1.6.1 Theoretical Contribution	5
1.6.2 Technical Contribution.....	6
1.7 Research Questions	7
1.8 Organization of Chapters	7
Chapter 2: Related Work and Background	10
2.1 Introduction.....	10
2.2 Related Work	10
2.3 Background of the FAIR logical perspective.....	14
2.3.1 Build an argument for FAIR sentences.....	16
2.3.2 Truth Table method.....	17
2.3.3 Natural Deduction method	19
2.4 Fuzzy logic for FAIRness assessment.....	22
2.4.1 Fuzzification of input data	23
2.4.2 Fuzzy inference (Rule-Based) systems.....	24
2.4.3 Aggregating the outputs of Fuzzy Inference System.....	25
2.4.4 Defuzzification of the output.....	26
2.5 Technical concepts for FAIR data principles implementation.....	27
2.6 Conclusions	29
Chapter 3: Logical Perspective on the Implementation of the FAIR Data Principles	30
3.1 Introduction.....	30
3.2 Logical analysis of FAIR data principles.....	31
3.2.1 Paraphrase, Symbolize, and Translate FAIR sentences.....	31

3.2.2 <i>Formal logical analysis evaluation methods</i>	45
3.3 Results of the Formal logical analysis for FAIR data principles	47
3.3.1 <i>Formal proof of the Truth Table method</i>	48
3.3.2 <i>Formal Proof of Natural Deduction method</i>	53
3.3.3 <i>Discussion of the findings</i>	56
3.3.4 <i>Construction of FAIR theorem</i>	58
3.4 Discussion	60
3.5 Conclusions	61
Chapter 4: Technical Design and Implementation of FAIR Framework	62
4.1 Introduction	62
4.2 Technical design.....	63
4.2.1 <i>The design of FAIR semantic web ontology</i>	63
4.2.2 <i>The design of the interface</i>	73
4.2.3 <i>FAIRtool.org triple store Setup</i>	78
4.3 Demonstration of the NCDC Use Case.....	79
4.3.1 <i>NCDC Storm Events Database Use case</i>	79
4.3.2 <i>Result of the use Case</i>	81
4.4 Discussion	89
4.5 Conclusions	89
Chapter 5: Utilizing Fuzzy Logic for Evaluating “FAIRness” of A Digital Resource	91
5.1 Introduction	91
5.2 FAIRness level evaluation with an NKN dataset use case	92
5.2.1 <i>Modeling FFAF Inputs</i>	94
5.2.2 <i>FFAF Fuzzification</i>	95
5.2.3 <i>FFAF Inference System</i>	98
5.2.4 <i>Aggregation and Defuzzification of the output of FFAF</i>	99
5.3 Demonstration of NKN use case in R.....	100
5.3.1 <i>Result of FFAF system</i>	102
5.4 Discussion	104
5.5 Conclusions	105
Chapter 6: Conclusions	106
6.1 Summary of results	106
6.2 Main Conclusions.....	108
6.3 Recommendations for future research	110
References	112
List Of Publications	120
Refereed Papers in Conferences.....	120
Refereed Papers in Edited Books.....	120
Presentations in Conferences and Workshops	121
Refereed Papers in Progress.....	121
Appendix A: FFAF R Code	123

List of Figures

Fig. 1.1: The FAIR data principles [1].....	2
Fig. 2.1: Fuzzy logic system.....	22
Fig. 2.3: Trapezoidal fuzzifier.....	23
Fig. 2.2: Triangular fuzzifier.....	23
Fig. 3.1: Flow diagram: a process of logical analysis for FAIR.....	31
Fig. 4.1: Classes of FAIR Ontology and their Relations – (Generated using Protégé).....	64
Fig. 4.2: Classes and properties of FAIR Ontology – (Generated using VOWL).....	65
Fig. 4.3: Provenance ontology of a digital recourse (Adapted from W3C PROV).....	71
Fig. 4.4: Screenshot of the Findable entry interface of FAIRtool.org.....	74
Fig. 4.5: Screenshot of the Accessible entry interface of FAIRtool.org.....	75
Fig. 4.6: Screenshot of the Interoperable entry interface of FAIRtool.org.....	76
Fig. 4.7: Screenshot of the Reusable entry interface of FAIRtool.org.....	77
Fig. 5.1: Findable membership function.....	96
Fig. 5.2: Accessible membership function.....	97
Fig. 5.3 Interoperable membership function.....	97
Fig. 5.4 Reusable membership function.....	98
Fig. 5.5: FAIRness fuzzy set.....	103
Fig. 5.6: FAIRness level crisp output.....	104

List of Tables

Table 2.1: Sentential Logic (SL) symbols	14
Table 2.2: Logical connectives	15
Table 2.3: Logical connectives truth-values	18
Table 2.4: Rules of inference of valid argument forms.	19
Table 2.5: Rules of inference of Logically Equivalent Expressions.....	21
Table 3.1: FAIR Principles, FAIR's <i>wffs</i> , Equivalent Rule of Inference, and its Tautology form.....	46
Table 3.2: Proof of the Hypothetical Syllogism (H.S.) using Truth Table.....	49
Table 3.3: Proof of the Modus Ponens (M.P.) using Truth Table.	51
Table 3.4: Proof of Absorption (Abs.) using Truth Table.	53
Table 3.5: Natural Deduction method formal Proof of (H.S.) $((p \rightarrow q) \wedge (q \rightarrow r)) \Rightarrow (p \rightarrow r)$	54
Table 3.6: Proof of Modus Ponens using Natural Deduction method $(p \wedge (p \rightarrow q)) \Rightarrow q$	55
Table 3.7: Natural Deduction method for Absorption (Abs.): $(p \rightarrow q) \Rightarrow (p \rightarrow (p \wedge q))$	55
Table 3.8: Summary of FAIR principles validity and tautology proofs.	56
Table 3.9: Findable, Accessible, Interoperable, Reusable Theorems.....	57
Table 3.10 FAIR Theorem.....	58
Table 4.1: The NCDC Storm Events Database use case metadata inputs for FAIRtool.org....	79
Table 5.1: The NKN dataset use case metadata inputs for FAIRtool.org	93
Table 5.2: FFAF Input Variables	95
Table 5.3: FFAF Output Variables	96

List of Listings

Listing 4.1: F1 principle’s RDF: The Identifier.....	81
Listing 4.2: F2 principle’s RDF: Rich Metadata	82
Listing 4.3: F3 principle’s RDF: Data Identifier	82
Listing 4.4: F4 principle’s RDF: Registered in Searchable Data Catalog	83
Listing 4.5: A1 principle’s RDF: Communication Protocol Standard	83
Listing 4.6: A1.1 principle’s RDF: Communication Protocol Characteristics	84
Listing 4.7: A1.2 principle’s RDF: Communication Protocol- Authorization-Authentication Rules	84
Listing 4.8: A2 principle’s RDF: Persistence Policy.....	85
Listing 4.9: I1 principle’s RDF: Knowledge Representation Format.....	85
Listing 4.10: I2 principle’s RDF: Vocabulary Used.....	86
Listing 4.11: I3 principle’s RDF: Metadata Cross-Reference	86
Listing 4.12: R1 principle’s RDF: Richly Accurate Relevant Metadata Attributes	87
Listing 4.13: R1.1 principle’s RDF: Data Usage License	87
Listing 4.14: R1.2 principle’s RDF: Digital Resource Provenance.....	88
Listing 4.15: R1.3 principle’s RDF: Domain-Relevant Community Standard.....	88

Chapter 1: Introduction

1.1 GENERAL DESCRIPTION

In this dissertation, we present an implementable framework of FAIR data principles for Earth science. First, we prove the logical validity and tautology of these FAIR principles, which contributes to the development of FAIR theorems. Then, based on this solid logical foundation, we build a pilot system that adopts the FAIR data principles. This pilot system is a platform that allows users to describe digital resources following the guidelines of FAIR data principles. The FAIR principles center primarily on improving the capacity of computers to locate and utilize the digital resource (e.g., dataset, software code, and workflow) efficiently, then facilitating their reuse by individuals and machines [1]. The main aim is to make digital resources machine-readable. We believe that the implementation of FAIR data principles leads to good data management and stewardship. Consequently, the benefits of successful data management and stewardship are high-quality metadata for digital resources that promote and improve the continuing process of exploration, evaluation, and reuse in subsequent research.

1.2 FAIR DATA PRINCIPLES

FAIR refers to the four foundational pillars Findability, Accessibility, Interoperability, and Reusability. The aim is to improve the ability of computers to find, access, interoperate, and reuse data. As shown in Fig. 1.1, FAIR data principles, in brief, are 1) Findable, discoverable with metadata, locatable, and identifiable through a standard identification mechanism; 2)

Accessible, obtainable, and available all the time; even if the data are restricted, the metadata is open; 3) Interoperable, both syntactically explainable and semantically understandable, enabling data exchange and allowing data reuse among researchers, institutions, organizations, and countries; and 4) Reusable, adequately described, and shared with the permitted licenses, supported by provenance empowering the broadest reuse possible and minimal tedious integration with other data sources.

<p>Textbox 1: The FAIR principles as stated in the study by Wilkinson <i>et al.</i> (1)</p> <p>To be Findable:</p> <ul style="list-style-type: none"> F1. (meta)data are assigned a globally unique and persistent identifier F2. data are described with rich metadata (defined by R1 below) F3. metadata clearly and explicitly include the identifier of the data it describes F4. (meta)data are registered or indexed in a searchable resource <p>To be Accessible:</p> <ul style="list-style-type: none"> A1. (meta)data are retrievable by their identifier using a standardized communications protocol <ul style="list-style-type: none"> A1.1 the protocol is open, free, and universally implementable A1.2 the protocol allows for an authentication and authorization procedure, where necessary A2. metadata are accessible, even when the data are no longer available <p>To be Interoperable:</p> <ul style="list-style-type: none"> I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. I2. (meta)data use vocabularies that follow FAIR principles I3. (meta)data include qualified references to other (meta)data <p>To be Reusable:</p> <ul style="list-style-type: none"> R1. meta(data) are richly described with a plurality of accurate and relevant attributes <ul style="list-style-type: none"> R1.1. (meta)data are released with a clear and accessible data usage license R1.2. (meta)data are associated with detailed provenance R1.3. (meta)data meet domain-relevant community standards

Fig. 1.1: The FAIR data principles [1]

Findable in FAIR data principles requires that each dataset has a specific global identifier and combined with accurate, searchable metadata. Furthermore, to be Accessible, those data and metadata should all be addressed using an accessible, standard protocol; also, they should use a structured, widely understood descriptive language and use popular and widely used

frameworks of related vocabulary and ontologies to make the data and metadata Interoperable. Finally, excess cross-references and a simple, well-defined procedure for obtaining licensing and provenance data should be richly represented by the data owner to make the data Reusable.

1.3 DATA STEWARDSHIP

Data stewardship is the management and oversight of data assets according to established best practices in data governance [2]. The entire process deals responsibly with data throughout and after the scientific discovery process. Data stewardship involves the idea of long-term maintenance and sustainability of these precious data assets, to discover and reuse them for follow-up studies [3]. Currently, Earth science is confronted with several challenges such as lack of metadata standard, struggle to locate needed datasets, ruinous metadata authoring process, and longtime spent on data curation. Therefore, the solution to rectify these issues is to have good data stewardship that relies on tools and technologies adopting FAIR data principles. Besides, publishers and journals like Nature Geoscience to enhance good data stewardship required the FAIR metadata to be part of all articles and datasets submissions from January 2020 onwards [15].

1.4 THESIS STATEMENT

The implementation of FAIR data principles based on concrete logical analysis can improve data management and stewardship for digital resources of Earth science. This data infrastructure advancement promotes knowledge exploration and intellectual innovation for the Earth science community.

1.5 MOTIVATION AND PROBLEM STATEMENT

To ensure the quality of data over the internet, we must have metadata standards to achieve rich annotations. For this reason, a considerable number of metadata standards have been created [4]. However, the result of using these standards of metadata deposited in public metadata repositories is often insufficient for the following reasons [5]. First, the metadata authoring process can be highly onerous for scientists because typical submissions involve a metadata spreadsheet-based entry, which is sometimes distributed through several spreadsheets, accompanied by a manual collection of several spreadsheets and files of raw data aggregated into an overall submission package [6]. Second, standards of metadata are generally written by specialists at a broad abstraction level. For example, while a standard may require an entity related to a biological sample to be collected, it usually does not specify how to provide organism value because there is minimal use of the significant number of structured terminologies currently existing in biomedicine [7]. Third, the submission repositories have limited or non-existent frameworks to link standardized terminologies. Because of this lack of standardization, users often provide ad hoc values or omit many values in metadata input [8]. Fourth, 50% of users searching data through web search engines and database archives struggle to locate the datasets they need [9][10]; Furthermore, discovering, reformatting, and cleansing data (just from 36% of reusable data!) consumes 80% of the data scientists' time [11]. Also, new ecology and evolution research has discovered that 64% of public dataset collections are unusable [12]. Because of this, new global programs, such as the European Open Science Cloud [EOSC] [13] and the NIH Big Data to Knowledge [BD2K][14] and publishers such as Nature Geoscience began to use FAIR data principles in their workflows [15]. Finally, the tools and

technologies available for the Earth science community to facilitate FAIR data principles' implementation are currently limited [16]. Based on the above analyses, we conclude that there is an immediate demand for tools and technologies that implement FAIR data principles to enhance the data management and stewardship for digital resources of Earth science.

1.6 CONTRIBUTIONS

Our contribution to the scientific community lies in two parts: theorem development and tangible solution. For the first part, in theorem development we originated new theorems for data science specifically for the data management and stewardship field by transforming FAIR data principles into FAIR theorems. Furthermore, we utilized fuzzy logic to measure the FAIRness level of digital resources. For the second part, we created a tangible solution, a novel tool to allow the publication of semantic descriptions of Earth science digital resources on the web following the FAIR Data Principles. In the following sections, we will elaborate on these contributions.

1.6.1 THEORETICAL CONTRIBUTION

The first theoretical contribution is a formal logical evaluation of the FAIR data principles. Our objective is to examine the elements of FAIR for logical validity and tautology. We use two logic methods: The Truth Table and the Natural Deduction. Furthermore, Sentential Logic was used as the formal logic language. The design of the study is based on four consecutive processes: 1) Build an argument for the FAIR sentences; 2) Paraphrase, symbolize,

and translate the FAIR arguments into logical symbols and prepare for the formal logical analysis; 3) Conduct formal logical analysis of FAIR arguments using the Truth Table method to discover the truth value of the FAIR argument symbols; and 4) Conduct formal logical analysis of FAIR arguments using the Natural Deduction method to deduct the FAIR argument validity using the inference rules. The results of the study are the proofs of the logical validity and tautology of the FAIR sentences. Ultimately, the proofs led to the development of the FAIR theorems, which culminate in the development of the FAIR theorem. This theoretical part will help establish rules to translate the FAIR data principles into machine-readable formats, which are necessary for the implementation of FAIR in the cyberinfrastructure.

The second theoretical contribution is utilizing fuzzy logic to evaluate the uncertainty of the FAIRness level of a digital resource. To date, there are no FAIRness evaluation studies based on fuzzy logic. We thus argue that fuzzy logic is an efficient method for evaluating the level of FAIRness of a digital resource. Therefore, we built a Fuzzy FAIR Assessment Framework (FFAF) to measure this uncertainty; the three major components of fuzzy logic are the foundation of FFAF: fuzzification, inferencing, and defuzzification. As a result, applying the FFAF model on the FAIR data principles led to a specific FAIRness level degree.

1.6.2 TECHNICAL CONTRIBUTION

We propose a framework that provides an intuitive and principled semantic approach to metadata acquisition. The framework uses integrated semantic ontology-based metadata standards to help guide users to annotate their metadata quickly and precisely. The goal is to provide the ability for researchers to easily create metadata that are comprehensive and

standardized and make the corresponding digital resource conform to FAIR data principles. This tangible solution will be built upon advanced semantic web technologies through a combination of community standards and ontology developed in house and will adopt the FAIR data principles. The deliverable is a semantic web application (i.e., FAIRtool.org) that incubates the Earth science community research outputs and ultimately accumulates FAIR metadata and generates FAIR metadata datasets.

1.7 RESEARCH QUESTIONS

There are many remaining open-ended research questions about the future of FAIR data principles' implementation, especially regarding logical analysis, tooling, and infrastructure. In this dissertation, we address four research questions, which can enhance the management and stewardship of the Earth science digital resources.

- (1) How can we verify the validity and tautology of FAIR data principles?
- (2) How can we theoretically address FAIR data principles?
- (3) How can we technically approach FAIR data principles?
- (4) How can we efficiently evaluate the FAIRness of a digital resource?

1.8 ORGANIZATION OF CHAPTERS

The dissertation consists of six chapters, three of which (3-5) contain primary contributions and focus on the above-stated four research questions. We summarize these chapters and highlight their objectives in the following paragraphs.

Chapter 2: Background and Related Work. This chapter presents related work and background to the three areas of knowledge that compose this dissertation: 1) FAIR logical perspective; 2) FAIR data principal technical implementation; and 3) Fuzzy logic utilization for FAIRness assessment. For each area, we introduced all the necessary definitions, methods, and other relevant information.

Chapter 3: Logical perspective on the implementation of the FAIR data principles. This chapter presents the details of logical analysis methods. It explores the methods in four consecutive processes: 1) Build an argument for the FAIR sentences; 2) Paraphrase, symbolize, and translate the FAIR arguments into logical symbols and prepare for the formal logical analysis; 3) Conduct formal logical analysis of FAIR arguments using the Truth Table method to discover the truth value of the FAIR argument symbols; and 4) Conduct formal logical analysis of FAIR arguments using the Natural Deduction method to deduct the FAIR argument validity using the inference rules.

Chapter 4: Technical design and implementation of the FAIR framework. This chapter describes the processes and identifies the components that build this tangible solution, i.e., FAIRtool.org. Furthermore, it illustrates the solution framework based on a semantic web platform called Vitro; this platform enables specialists to build a semantic web application for the desired task. Moreover, it shows that this semantic web application is depending on an in-house-developed ontology called FAIR ontology and relies on a triple store to store the input and output of the system. It explains the three steps that involve the design and implementation

of FAIR data principles. The first step is to design the ontology; the second step is to build the entry interface and setup the triple store, the final step is to display the result of the use case example of Earth science digital resource through simulating the 15 FAIR data principles.

Chapter 5: Utilizing fuzzy logic for assessing the FAIRness of a digital resource. This chapter presents the possibility of utilizing fuzzy logic to evaluate the uncertainty of the FAIRness level of a digital resource. It also demonstrates how to measure this uncertainty, build a fuzzy FAIR assessment framework model (FFAF) based on fuzzy logic. Further, it describes how the FFAF model is constructed based on four crisp inputs and one crisp output to evaluate and measure the FAIRness level of a digital resource. This chapter explains the three main processes of FFAF, which are fuzzification, inferencing, and defuzzification. It also shows how to achieve an efficient FAIRness level result by applying the FFAF on the FAIR. Finally, this chapter demonstrates FAIRness level evaluation for a dataset from NKN using the FFAF model through R code.

Chapter 6: Conclusions. This chapter reiterates the summary of results discussed in chapters 3-5 and explains their significance, draws the main conclusion, and provides recommendations for future research.

Chapter 2: Related Work and Background

2.1 INTRODUCTION

This chapter provides the related work and background of the three areas of knowledge that compose the foundation for this dissertation: 1) FAIR logical perspective; 2) FAIR data principal technical implementation; and 3) Fuzzy logic FAIRness level assessment. We first highlight the relevant work to identify the gap. Then, we will review the logical perspective of FAIR. After that, we will explain the Fuzzy logic system for the FAIRness level assessment. Lastly, we will describe the technical components that will contribute to the technical implementation of FAIR data principles in chapter 4. Furthermore, we will identify all of the needed definitions.

2.2 RELATED WORK

First, in recent years, several projects have been developed to evaluate and analyze FAIR data principles. The following are highlights of the novel projects: 1) OpenPVSignal presents a qualitative evaluation of the FAIR principles [34]; 2) The traffic-light rating system includes statistical and descriptive analysis evaluation [35]; and 3) OpenPREDICT shows how to build a machine learning FAIR workflow and introduces how a conventional ontology validation method addresses the FAIR principle [37]. The crucial drawback of these efforts is the absence of logic analysis and evaluation of the FAIR data principles. To overcome this drawback, we performed research mostly oriented towards logic analysis and validation of the

FAIR sentences. Therefore, we intend to construct formal proofs for validity and tautology for these sentences using well-known logic methods such as Truth Tables and Natural Deduction. Furthermore, this logical evaluation of the FAIR sentences is essential for the scientific community to ensure that these principles are logically valid and tautology before adopting them.

Second, a considerable volume of literature has been published on the evaluations of FAIRness to help improve the stewardship of digital resources. For example, the FAIRness evaluation framework proposed by Mark D. Wilkinson focuses on a series of FAIR indicators that can be identified by an automated agent in a digital object [38]. Furthermore, FAIRshake is a toolkit for assessing digital resources' findability, accessibility, interoperability, and reusability [39]. Besides, FAIR Evaluator is a framework that enables all involved parties to evaluate the FAIRness of a digital resource [36]. Moreover, the FAIR metrics community published 14 general-purpose maturity indicators for describing FAIRness, these indicators incorporated into a program that can collect metadata automatically from metadata suppliers and produce a principle-specific assessment for FAIRness [40]. These efforts did not solve the FAIRness vagueness problem because the usage of the Fuzzy logic to evaluate the FAIRness level was certainly not discussed in these evaluation efforts, and none of the literature utilized the Fuzzy logic method to measure the FAIRness level. Therefore, to overcome this FAIRness vagueness problem, we study the possibility of using Fuzzy logic to evaluate the uncertainty of FAIRness of a digital resource. Thus, we proposed a Fuzzy FAIRness Assessments Framework (FFAF); the goal of FFAF is to solve the problem of FAIRness vagueness by producing a specific FAIRness level.

Third, there are some existing tools and methods for helping to implement FAIR data principles. We describe a variety of efforts most related to our work: 1) FAIR Data Point (FDP), in brief, is a RESTful web service that allows data providers to display their datasets utilizing rich machine-readable metadata [41]. FDP mainly depends on FAIRifier based on OpenRefine from google [43]. FAIRifier has all the capabilities of OpenRefine to boost data quality, but it is a difficult task for non-technical users to use; furthermore, in [42] the author discussed the two main components of FDP (i.e., FAIR Accessor and FAIR Projector) of the FAIR infrastructure proposal and stated that this proposed FAIR infrastructure seems difficult to achieve. 2) The Center for Expanded Data Annotation and Retrieval (CEDAR) was founded in 2014 to establish computational ecosystem for the creation, analysis, use, and enhancement of biomedical metadata. Their methodology relies on using metadata models that identify the data elements required to represent specific types of biomedical experiments. These models limit standardized words and synonyms for biomedical data items only. CEDAR utilizes these models as a repository to assist biomedical scientists with adding annotated datasets to suitable online data repositories [44]. 3) UniProt is a thorough protein sequence and annotation data tool. A secure URL defines all records individually and provides access in various formats such as web pages, basic text, and Resource Description Framework (RDF) to the repository. The repository provides rich metadata, which is human-readable (HTML) or machine-readable (CSV, RDF), where popular vocabulary and ontology like UniProt Core, ECO, and FALDO can be used for RDF encoded responses. Each UniProt record has global URL links to over 150 different databases for example PubMed that allow rich citations for medical data items. These URL links are machine-actionable using RDF format. Lastly, with the RDF format, UniProt

Core Ontology specifically categorizes all the records, leaving no doubt about what the metadata represents. This enables the fully automatic extraction of records and cross-reference details [45]. 4) The EarthCube Project 418 is a technical implementation attempt of Schema.org that is intended to illustrate common publication methods for data facilities using Schema.org and extensions (e.g., geolink:, datacite:, gdx:, and earthcollab:). Project 418 scales the F in the FAIR so that professionals can consider and adopt this method [49, 50], but it did not include the complete FAIR data principles elements. 5) The Force 11 Data Citation Implementation Group has also given practical recommendations on the implementation of several FAIR data principles to help groups and organizations that have already pursued FAIR goals [11]. 6) Other repositories such as Zenodo, Figshare, and the Open Science Framework also offer useful tools that help researchers make their uploaded data FAIR data principles partially compliant by, for example, generating a DOI and populating the metadata. 7) There are a variety of new initiatives [1] for which FAIR data principles compliance is one of their main goals; however, these initiatives only deliver useful guidance and recommendation to those interested in complying with the FAIR data principles, but they are not involved in technical implementations.

In conclusion, these data systems and tools lack the use of formal logical modeling and fuzzy logic in the analysis and evaluation of FAIR data principles. Furthermore, they also have a deficiency in the comprehension and ease of use of their solutions. However, to conquer these gaps, we developed three novel methods: 1) Perform a logical analysis for the FAIR data principles to examine their validity and tautology; 2) Utilize fuzzy logic to assess the FAIRness level of the digital resource; and 3) Provide concrete technical implementation which focuses on ease of use and comprehensiveness (i.e., contains all the 15 FAIR data principles) of FAIR

data principles using semantic web technologies. In the following sections of this chapter, we will set the background and defines the needed definitions for these methods.

2.3 BACKGROUND OF THE FAIR LOGICAL PERSPECTIVE

In this section, we will set the background for the formal logic method that we will apply to address the identified logical gap in section 2.2. chapter 3 will demonstrate the complete implementation. The intent of the formal logic method is to utilize modern-symbolic logic to build arguments out of the FAIR sentences and then analyze them logically to prove their validity and tautology. Thus, we employed a logical language called Sentential Logic (further referred to as SL), also called modern symbolic logic, propositional Logic-Calculus, and Truth-Functional Logic [17]. Since SL is a fundamental building block of formal logic, syntactic letters symbolize the atomic sentences. For example, we can use the atom S to denote the proposition “Socrates is a man.” Furthermore, capital letters represent core-atomic sentences, and logical connectives create highly sophisticated sentences [18].

Table 2.1: Sentential Logic (SL) symbols

Capital letters	A, B, C, . . . , Z
Subscripts if needed	A1, B1, C1, A2,, K217, . . .
Logical connectives	$\neg, \wedge, \vee, \rightarrow, \leftrightarrow$
Parentheses and comma	(,)

To illustrate, Table 2.1 shows SL symbols that represent symbolization keys. For instance, the letters “A, B, C, ...” could mean any atomic sentence, and the logical connectives \neg , \wedge , \vee , \rightarrow , \leftrightarrow (as shown in Table 2.2) combine atomic sentences [18].

Table 2.2: Logical connectives

Symbol	Name	Meaning
\neg	Negation	“It is not the case that . . .”
\wedge	Conjunction	“Both . . . and . . .”
\vee	Disjunction	“Either . . . or . . .”
\rightarrow	Conditional	“If . . . then . . .”
\leftrightarrow	Biconditional	“. . . if and only if . . .”

To demonstrate SL, consider this famous logical argument that consists of three sentences: “Since Socrates is a man and all men are mortal, then Socrates is mortal” can be symbolized, using Table 2.1, as such:

S: “Socrates is a man.”
M: “All men are mortal.”
C: “Socrates is mortal.”

Then, after applying logical connectives, using Table 2.2, we obtain:

$$(\mathbf{S} \wedge \mathbf{M}) \rightarrow \mathbf{C} \quad (\text{wff } 1)$$

It is therefore essential to have a symbolism key when translating from a natural language such as English to SL. For every sentence letter that has been used in a symbolization, the key contains an English language sentence. Because every symbol series is an expression, only a valid expression is called the well-formed formula, using the acronym *wff* (plural *wffs*) [18]. In

this context, $(S \wedge M) \rightarrow C$ is *wff*. Magnus [18] devised the following formal definition for the well-formed SL formula, which will be used in our work for analyzing FAIR data principles.

Definition 2.1. *well-formed formulas (wffs)* in the SL are defined as follows:

1. “Every atomic sentence is a *wff*.”
2. If \mathcal{A} is a *wff*, then $\neg \mathcal{A}$ is a *wff* of SL.
3. If \mathcal{A} and \mathcal{B} are *wffs*, then $(\mathcal{A} \wedge \mathcal{B})$ is a *wff*.
4. If \mathcal{A} and \mathcal{B} are *wffs*, then $(\mathcal{A} \vee \mathcal{B})$ is a *wff*.
5. If \mathcal{A} and \mathcal{B} are *wffs*, then $(\mathcal{A} \rightarrow \mathcal{B})$ is a *wff*.
6. If \mathcal{A} and \mathcal{B} are *wffs*, then $(\mathcal{A} \leftrightarrow \mathcal{B})$ is a *wff*.
7. All and only *wffs* of SL can be generated by applications of these rules.”

Here \mathcal{A} and \mathcal{B} are not the sentence letter A and B; they are variables that represents any *wff*.

2.3.1 BUILD AN ARGUMENT FOR FAIR SENTENCES

An argument is a group of propositions or sentences in the form of premises and a single conclusion. There are two types of arguments regarding the premises and a single conclusion relationship: a deductive argument and an inductive argument. A deductive argument claims to give substantial grounds for its conclusion. If it guarantees such grounds, then they are valid; if not, they are invalid. An example of this logic is SL [17]. The inductive argument argues that its premises offer a certain degree of possibility, although still not a certainty, to its conclusion. An example of such logic is fuzzy logic [19]. We are interested in adopting a deductive

argument, which can be described as a sequence of sentences consisting of premise sentences at the beginning and a conclusion sentence at the end. If the premise sentences are correct and have a valid claim, then we believe the truth of the conclusion sentence. Validity can never apply to any single proposition by itself because the needed relation cannot possibly be found within any single proposition. To illustrate, consider the following examples:

1. Proposition/sentence example, (attributes are: True, or False):

Socrates is a man. (True)

2. Argument example, (attributes are: Valid, or Invalid):

Socrates is a man. (True)
All men are mortal. (True)

∴ Socrates is mortal. (True)
 Therefore, this argument is valid

Thus, the attributes of an individual proposition or sentence are true and false, while the attributes of arguments are valid and invalid. The combination of propositions values decides the validity of an argument. Next, we will introduce the truth table that will help us understand the true value of combined propositions which form an argument like that argument in (*wff 1*).

2.3.2 TRUTH TABLE METHOD

The method of Truth Table introduces a semantic way to evaluate SL's sentences and arguments. The composite sentence's truth value, either (True = 1, or False = 0), relies only on its atomic sentences' truth values [18]. For example, we should first identify the truth value of A and the truth value of B to realize the truth value of $(A \leftrightarrow B)$ and then consider the truth value

of the logical connection " \leftrightarrow " that links them. Table 2.3 illustrates the truth value for each logical connective:

Table 2.3: Logical connectives truth-values

$\neg A$	A	B	$A \wedge B$	$A \vee B$	$A \rightarrow B$	$A \leftrightarrow B$
0	1	1	1	1	1	1
0	1	0	0	1	0	0
1	0	1	0	1	1	0
1	0	0	0	0	1	1

Accordingly, we apply the Truth Table method on the FAIR arguments and obtain satisfactory results, which will be explained in chapter 3. It is important to realize some essential definitions which we will utilize in the Truth Table and Natural Deduction methods for our proofs, as listed below:

Definition 2.2: A deductive argument is **valid**, if, and only if, the premises are true, then, the conclusion ought to be true. Otherwise, it is invalid.

Definition 2.3: A sentence is a **theorem** if, and only if, it is a **tautology**.

Definition 2.4: A **theorem** is a conclusion shown to be true by writing a proof.

Definition 2.5: a *wff* is a **tautology** if, and only if it is true for all possible truth-value assignments to the statement letters that form it.

Definition 2.6: A **formal proof** is a valid argument consisting of a series of propositions such that the last proposition in the series is the conclusion of the claim, and each

proposition in the sequence is either a hypothesis of the argument or results by a logical inference or a logical equivalence from a prior proposition in the sequence.

2.3.3 NATURAL DEDUCTION METHOD

Natural Deduction is a process of utilizing the laws of inference to show the truth of a deductive argument. The objective of the Natural Deduction method is to show that specific arguments are valid in a manner that helps one to accept the logic that might involve those arguments. Inference rules are a logical toolbox from which, if necessary, the tools can be taken to prove validity. With this in mind, we use Natural Deduction to manipulate sentences in conformance with the well-known rules of inference [20] listed in Table 2.4 and Table 2.5.

Table 2.4: Rules of inference of valid argument forms.

	Rules of logical inference	Valid argument forms
1	Absorption (Abs.)	If p implies q, absorption permits the inference that p implies both p and q. Symbolized as: $p \rightarrow q \therefore p \rightarrow (p \wedge q)$
2	Addition (Add.)	Given any proposition p, addition permits the inference that p or q. Also called “logical addition”. Symbolized as: $p, \therefore p \vee q$
3	Conjunction (Conj.)	A truth-functional connective meaning “and” symbolized by the “ \wedge ” symbol. A statement $p \wedge q$ is true if and only if p is true and q is true. Symbolized as: $p, q, \therefore p \wedge q$

4	Constructive Dilemma (C.D.)	Constructive Dilemma permits the inference that if $(p \rightarrow q) \wedge (r \rightarrow s)$ is true, and $(p \vee r)$ is also true, then $(q \vee s)$ must be true. Symbolized as: $(p \rightarrow q) \wedge (r \rightarrow s), (p \vee r), \therefore (q \vee s)$
5	Disjunctive Syllogism (D.S.)	One premise is a disjunction, another premise is the denial of one of the two disjuncts, and the conclusion is the truth of the other disjunct. Symbolized as: $p \vee q, \sim p, \therefore q$
6	Hypothetical Syllogism (H.S.)	If the premises $(p \rightarrow q)$, and $(q \rightarrow r)$ are assumed to be true, permits the conclusion that $(p \rightarrow r)$ is true. Symbolized as: $p \rightarrow q, q \rightarrow r, \therefore p \rightarrow r$.
7	Modus Ponens (M.P.)	If the truth of a hypothetical premise is assumed, and the truth of the antecedent of that premise is also assumed, we may conclude that the consequent of that premise is true. Symbolized as: $p \rightarrow q, p, \therefore q$
8	Modus Tollens (M.T.)	If the truth of a hypothetical premise is assumed, and the falsity of the consequent of that premise is also assumed, we may conclude that the antecedent of that premise is false. Symbolized as $p \rightarrow q, \sim q, \therefore \sim p$

9	Simplification (Simp.)	It permits the separation of conjoined statements. If the conjunction of p and q is given, simplification permits the inference that p. Symbolized as: $p \wedge q, \therefore p$
---	------------------------	---

Table 2.5: Rules of inference of Logically Equivalent Expressions.

	The Rules of Inference	Logically Equivalent Expressions
10	De Morgan's theorems (De M.)	$\sim(p \wedge q) \equiv (\sim p \vee \sim q)$ $\sim(p \vee q) \equiv (\sim p \wedge \sim q)$
11	Commutation (Com.)	$(p \vee q) \equiv (q \vee p)$ $(p \wedge q) \equiv (q \wedge p)$
12	Association (Assoc.)	$[p \vee (q \vee r)] \equiv [(p \vee q) \vee r]$ $[p \wedge (q \wedge r)] \equiv [(p \wedge q) \wedge r]$
13	Distribution (Dist.)	$[p \wedge (q \vee r)] \equiv [(p \wedge q) \vee (p \wedge r)]$ $[p \vee (q \wedge r)] \equiv [(p \vee q) \wedge (p \vee r)]$
14	Double Negation (D.N.)	$p \equiv \sim \sim p$
15	Transposition (Trans.)	$(p \rightarrow q) \equiv (\sim q \rightarrow \sim p)$
16	Material Implication (Impl.)	$(p \rightarrow q) \equiv (\sim p \vee q)$

17	Material Equivalence (Equiv.)	$(p \equiv q) \equiv [(p \rightarrow q) \wedge (q \rightarrow p)]$ $(p \equiv q) \equiv [(p \wedge q) \vee (\sim p \wedge \sim q)]$
18	Exportation (Exp.)	$[(p \wedge q) \rightarrow r] \equiv [p \rightarrow (q \rightarrow r)]$
19	Tautology (Taut.)	$p \equiv (p \vee p)$ $p \equiv (p \wedge p)$

2.4 FUZZY LOGIC FOR FAIRNESS ASSESSMENT

In this section, we will set the background for the fuzzy logic method that we will apply to address the gap that was identified in 2.2. Fuzzification, inferencing, and defuzzification are the three main steps of the Fuzzy logic system. First, Fuzzification involves the transformation of classical crisp data into fuzzy data. Second, the Fuzzy inference mechanism links membership functions to fuzzy rules in order to extract the output fuzzy sets. Third, Defuzzification computes each related fuzzy output and produces classical single crisp output data [46]. The complete implementation is illustrated in chapter 5. Fig. 2.1 depicts these three steps, in which a Fuzzy logic system transforms crisp inputs into crisp outputs utilizing fuzzy inference and rules [19]. In the next sections, we will describe in detail these three main steps.

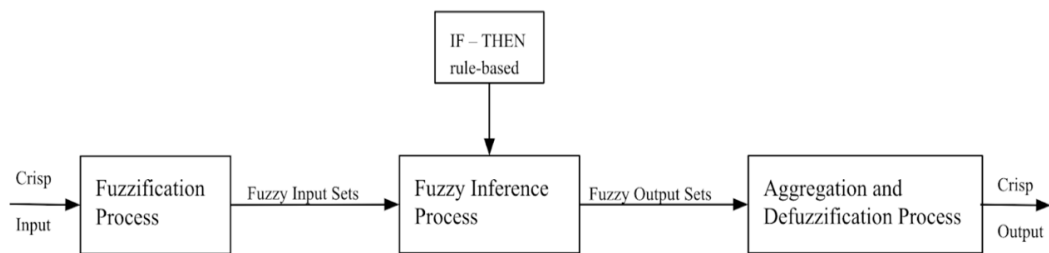


Fig. 2.1: Fuzzy logic system

2.4.1 FUZZIFICATION OF INPUT DATA

In the fuzzification step, a fuzzifier's task is to transform the external specific (crisp) input data into convenient semantic fuzzy data [47]. There are three forms of fuzzifiers: Gaussian fuzzifier, Triangular fuzzifier, and Trapezoidal fuzzifier [49]. For our purposes, we will use the Triangular fuzzifier and the Trapezoidal fuzzifier. These fuzzifiers assign crisp data input x to a fuzzy set A with distinct membership functions $\mu_A(x)$, as shown in Fig. 2.2 and Fig. 2.3 below.

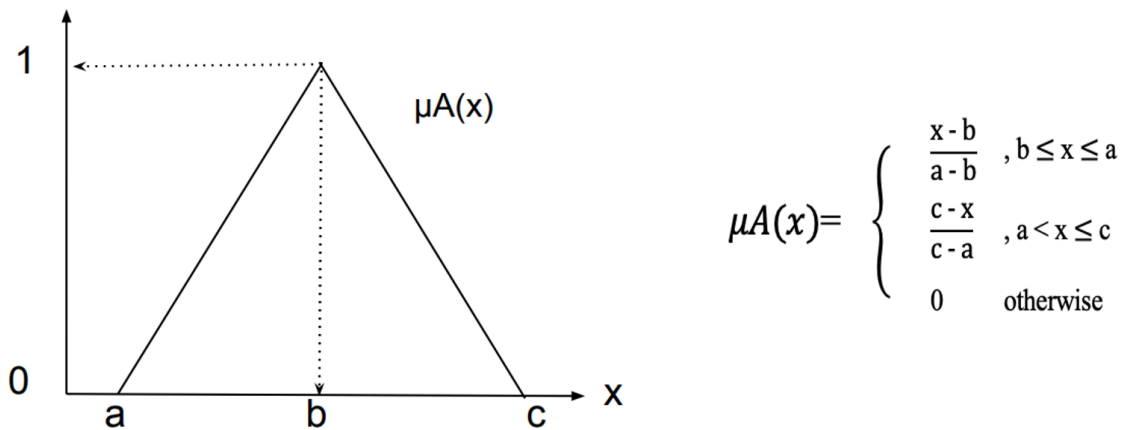


Fig. 2.2: Triangular fuzzifier

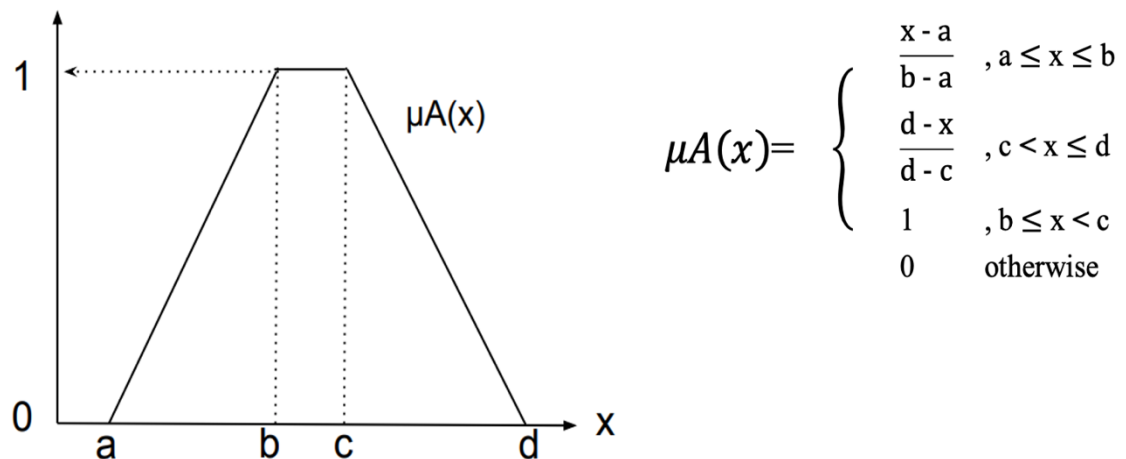


Fig. 2.3: Trapezoidal fuzzifier

2.4.2 FUZZY INFERENCE (RULE-BASED) SYSTEMS

Fuzzy Inference Systems (FIS) are three models: the Mamdani model [55], the Takagi-Sugeno model [56], and the Tsukamoto model [57]. The Mamdani approach is the most commonly employed FIS due in large part to its simplistic form and mostly used to address all general decision-making issues [58]. In this research, we comply with the Mamdani model, which follows the following rule-based formula:

“IF premise antecedent, THEN conclusion consequent.” (1)

The rule-based expression formula (1) is defined technically as the IF-THEN rule-based structure; typically, described as the logical deductive form. Usually, an inference is formulated, in such a manner, that if we know the truth of (premise, antecedent, or hypotheses), we then conclude or draw another reality called a conclusion (consequential) [59].

In the first step, the crisp data input x is to identify the degree to which the data x belongs to any of the relevant fuzzy sets. Once the input data fuzzified and the membership values are collected; then, the next step is to add them to the context of the fuzzy rules. When a particular fuzzy rule has several antecedents, a Fuzzy logical operator (AND or OR) is used to produce a specific number that represents the result of the antecedent evaluation. That number then be applied to a corresponding membership function [49]. To assess the conjunction of antecedents, the rule logical operator AND is used (aka T-Norms). Usually, the Fuzzy logic model uses the traditional fuzzy intersection method to execute this operation. For example, the intersection of fuzzy sets \tilde{A} and \tilde{E} is denoted by $\tilde{A} \cap \tilde{E}$ and defined by formula (2):

$$\mu_{\tilde{A}}(x) \cap \mu_{\tilde{E}}(x) = \min(\mu_{\tilde{A}}(x), \mu_{\tilde{E}}(x)) \quad (2)$$

Assume $\mu_A(x) = 0.17$, $\mu_B(y) = 0.83$, then we have $\mu_C(z) = \min[\mu_A(x), \mu_B(y)] = 0.17$.

Similarly,

$$x \text{ AND } y = \min(\text{truth}(x), \text{truth}(y))$$

If $x < y$, $\min(x, y) = x$. For instance, $\min(0.17, 0.83) = 0.17$.

These concepts formulate fuzzy formula (2):

“If x is A AND y is B, then z is C.”

On the other hand, for the evaluation of the disjunction of rule antecedents, the logical operator OR is used, which is introduced by the Union fuzzy operation in Fuzzy logic systems (aka T-Co-Norms). For example, the Union of fuzzy sets \tilde{A} or \tilde{E} is denoted by $\tilde{A} \cup \tilde{E}$ and defined by formula (3):

$$\mu_{\tilde{A}}(x) \cup \mu_{\tilde{E}}(x) = \max(\mu_{\tilde{A}}(x), \mu_{\tilde{E}}(x)) \quad (3)$$

Assume $\mu_A(x) = 0.17$, $\mu_B(y) = 0.83$, then we have $\mu_C(z) = \max[\mu_A(x), \mu_B(y)] = 0.83$.

Similarly,

$$x \text{ OR } y = \max(\text{truth}(x), \text{truth}(y))$$

If $x < y$, $\max(x, y) = y$. For instance, $\max(0.17, 0.83) = 0.83$.

Then, the fuzzy formula (3) is formulated as:

“If x is A OR y is B, then z is C.”

2.4.3 AGGREGATING THE OUTPUTS OF FUZZY INFERENCE SYSTEM

Aggregation is the unification of the outputs of all fuzzy rules. Thus, the aggregation step combines the membership functions outputs of all the rules and incorporates them into a

single fuzzy set [60]. There are many composition approaches, such as the max-product, the max-average, and max-min methods [47]. For our research purposes, we will use the aggregation max-min composition method, as described in equation (4).

$$\mu_{A \circ B}(x, z) = \max[\min(\mu_A(x, y), \mu_B(y, z))] \quad (4)$$

2.4.4 DEFUZZIFICATION OF THE OUTPUT

The last step in a Fuzzy logic system is defuzzification, which is the reverse of fuzzification. Fundamentally, this step generates a crisp single output data for a Fuzzy logic system from an aggregated fuzzy set. For that reason, a variety of defuzzification methods have been established, such as centroid defuzzifier, maximum of maximum defuzzifier, minimum of maximum defuzzifier, and means of maxima defuzzifier. The most common one is the centroid defuzzifier [61]. The centroid of area (COA)/center of gravity (COG) is defined mathematically by equation (5):

$$z^* = COG = \frac{\sum_{x=a}^b \mu_A(x)x}{\sum_{x=a}^b \mu_A(x)} \quad (5)$$

The COG of the fuzzy set delivers a crisp data value depending on the centroid defuzzifier method; in our research, we used equation (5). Then, the combined area for the membership function is split into a variety of sub-areas. The COG of each sub-area is then

calculated and the quantities of all those sub-areas are translated to the defuzzified crisp data value [62].

2.5 TECHNICAL CONCEPTS FOR FAIR DATA PRINCIPLES IMPLEMENTATION

In this section, we explain the concepts and the technologies used in the FAIR data principles implementation. Furthermore, chapter 4 describes in detail the technical implementation of FAIR data principles. The outcome is a semantic web application (i.e., FAIRtool.org).

The Semantic Web is an extension of the World Wide Web through standards set by the World Wide Web Consortium (W3C) [51]. The goal of the Semantic Web is to make data of Internet machine-readable. To encode metadata with semantics technologies such as Resource Description Framework (RDF) [21] and Web Ontology Language (OWL) [22] are used to represent metadata. For example, ontology can describe concepts, relationships between entities, and categories of things. These embedded semantics offer significant advantages, such as reasoning over data and operating with heterogeneous data sources [52].

Resource Description Framework (RDF) considers the basis of the principles and recommendations of the semantic web and linked data [21]. The RDF utilizes Unified Resource Identifiers (URIs) [53] to distinctly identify resources like Web pages, data items, concepts, persons, processes. URIs enable data to be distinctly identified and, therefore, findable and reachable via the World Wide Web.

OWL [22] and RDF Schema [21] describe both concepts and high-level semantic connections (for example, hierarchies between concepts that include objects properties, data, classes, cardinal restrictions on subjects' properties. Description Logics constructs the semantics of OWL [48]. Hence, which can be used by applications named reasoners, that allow automatic inferences.

Vitro is a web browser-based instance and ontology editor for general purposes besides a custom navigation system [23]. Furthermore, Vitro is an interactive ontology editor and software-based semantic platform deployed in a Tomcat servlet container as a Java web application. Vitro was primarily developed by Cornell University and used as the core of the popular scholarship and research portal VIVO [68].

Protégé is an open-source online ontology creator plus a knowledge management system. Protégé offers a graphical user interface to describe ontologies. It also provides deductive classifiers to verify the models are valid and to predict new information based on an ontology inference using reasoners. Protégé is a product of Stanford University [24].

The definition of ontology started to evolve with Gruber (1993) [25]. He initially defined it as: "an explicit specification for conceptualization" [27]. Similarly, Borst (1997), defined it as: "a formal specification of a shared conceptualization" [29]. That interpretation allows the conceptualization to reflect a common opinion among various groups. This conceptual framework should also be described in terms of a formal machine-readable

structure. Accordingly, Studer (1998) combines these two views, arguing that "An ontology is a formal, explicit specification of a shared conceptualization" [30].

A triple-store is storage for the RDF network [26]. It is like a relational database for data tables. Triple-stores are also known as "RDF store", "RDF database", "graph store", and "semantic repository". Various proprietary and open-source triple stores are available. SPARQL is the main triple store interface [54]; SPARQL is like the SQL in a "relational database query language". Nevertheless, SPARQL is likely applicable for database design used in logical languages like Prolog [28].

2.6 CONCLUSIONS

In this chapter, we summarized the relevant work and set up the background for the rest of this dissertation. We reviewed the logical perspective of FAIR; then, we explained the Fuzzy logic system for FAIRness level evaluation. After that, we described the technical components that contribute to the technical implementation of FAIR data principles. Furthermore, we identified all the needed definitions and discussed our approaches. In the following chapter 3, we will look in detail at the formal logical analysis of FAIR data principles.

Chapter 3: Logical Perspective on the Implementation of the FAIR Data Principles

3.1 INTRODUCTION

In this chapter, we examine the FAIR data principles (hereafter referred to as FAIR) [1] from a formal logical perspective. Formal logic is the techniques and rules used to differentiate between right and wrong claims [20]. We build and analyze FAIR arguments in formal logic. The arguments are constructed with propositions, which will support the logical reasoning to determine the level of FAIR. We can confirm a proposition if it is true or rejects it if it is false [20]. For elements in FAIR, it is important to understand clearly what an argument is and what it means to validate an argument. Therefore, we will essentially transform the 15 FAIR sentences into a formal language called symbolic logic; then, we will use formal logic methods and logical inference rules to analyze the propositions of the 15 FAIR sentences for validity. Besides, we want the formal validity to include at least some of the essential characteristics of natural language.

We also emphasize the significant features of our methods in the subsequent sections. Section 3.2 describes the formal logic methods used to analyze FAIR. Section 3.3 illustrates the results of the Truth Table and the Natural Deduction methods used to validate FAIR. Section 3.4 discusses the results and Section 3.5 states the conclusions.

3.2 LOGICAL ANALYSIS OF FAIR DATA PRINCIPLES

Our methods consist of four consecutive steps: 1) Build an argument for each FAIR sentence; 2) Paraphrase, symbolize, and translate FAIR sentences; 3) Analyze FAIR argument validity using the Truth Table method; and 4) Analyze FAIR argument validity using the Natural Deduction method. Fig. 3.1 summarizes the steps used in our logical analysis of the FAIR arguments.

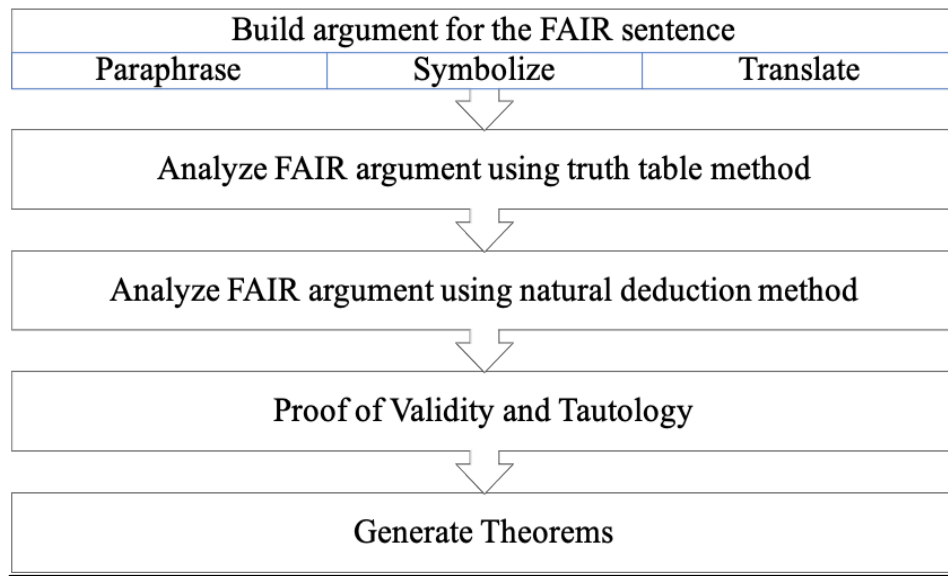


Fig. 3.1: Flow diagram: a process of logical analysis for FAIR

3.2.1 PARAPHRASE, SYMBOLIZE, AND TRANSLATE FAIR SENTENCES

The paraphrase is the most effective technique for starting the analysis of FAIR arguments [67]. Thus, we paraphrase FAIR arguments by setting out its sentences in plain language and logical order. This process would entail the reformulation of FAIR sentences; therefore, we take special attention to ensure that the paraphrase brought forward reflects

accurately and completely the argument of FAIR that needs to be analyzed. Thus, we followed Hardegree subsequent steps [66] to paraphrase and translate FAIR sentences:

1. Identify and abbreviate atomic sentences by capital letters.
2. Identify all the logical connectives.
3. Identify the main connective.
4. Note down the first hybrid formula to ensure that internal punctuation is maintained.
5. Symbolize the major connective; attach brackets, if necessary; and return to step 3 and operate on the resulting hybrid formula.
6. Work independently on the constitutive formula, applying steps 3-5 to each constitutive formula.
7. Substitute the symbolization of the constitutes back into the original hybrid formula.

It worth mentioning that in the original FAIR paper [1] that digital resources refer to (Meta)data and other objects, for instance, software code and workflow. Therefore, we will use “digital resource” instead of “(Meta)data” and “UD” in place of “Universe of Discourse.” Now we are ready to paraphrase, symbolize, and translate the FAIR sentences.

3.2.1.1 FINDABLE

F1 Principle

“F1. (Meta)data are assigned a globally unique and persistent identifier”

F1 Paraphrased as:

“If the digital resource is assigned a globally unique identifier and the digital resource is assigned a persistent identifier, then the digital resource would be found and resolved.”

F1 Symbolized as:

UD: Findable
 D: digital resource
 G: global unique identifier
 P: persistent identifier
 F: found
 R: resolved

Which yields the following hybrid formula:

“If D is assigned (G \wedge P), then (G \wedge P) leads to (F \wedge R), therefore D is (F \wedge R)”

F1 Translated as *wffs*:

$$D \rightarrow (G \wedge P)$$

$$(G \wedge P) \rightarrow (F \wedge R)$$

$$\therefore D \rightarrow (F \wedge R)$$

$$\equiv \text{Hypothetical Syllogism (H.S.)}$$

F2 Principle

“F2. Data are described with rich metadata”

F2 Paraphrased as:

“If the digital resource is described with rich metadata, then the digital resource is more findable.”

F2 Symbolized as:

UD: Findable
 D: digital resource is described with rich metadata

M: the digital resource is more findable.

Which yields the following hybrid formula:

“If D, then M. D therefore, M”

F2 Translated as *wffs*:

$D \rightarrow M$

D

$\therefore M$

\equiv Modus Ponens (M.P.)

F3 Principle

“F3. Metadata clearly and explicitly include the identifier of the data they describe”

F3 Paraphrased as:

“If the identifier of the digital resource is clearly and explicitly included in metadata, then the digital resource is more findable.”

F3 Symbolized as:

UD: Findable

I: identified of digital resource

C: clearly included in metadata

E: explicitly included in metadata

F: more findable digital resource

Which yields the following hybrid formula:

“If I is C and E in metadata, and $C \wedge F$ leads to F, then I is F.”

F3 Translated as *wffs*:

$$\begin{aligned}
 &I \rightarrow (C \wedge E) \\
 &(C \wedge E) \rightarrow F \\
 &\therefore I \rightarrow F \\
 &\equiv \text{Hypothetical Syllogism (H.S.)}
 \end{aligned}$$

F4 Principle

“F4. (Meta)data are registered or indexed in a searchable resource”

F4 Paraphrased as:

“If the digital resource is registered in a searchable resource or the digital resource indexed in a searchable resource, then the digital resource is more findable.”

F4 Symbolized as:

UD: Findable
 D: the digital resource
 R: registered in a searchable resource
 I: indexed in a searchable resource
 F: more findable digital resource

Which yields the following hybrid formula:

If D is R or D is I, therefore D is F.

F4 Translated as *wffs*:

$$\begin{aligned}
 &D \rightarrow (R \vee I) \\
 &(R \vee I) \rightarrow F \\
 &\therefore D \rightarrow F \\
 &\equiv \text{Hypothetical Syllogism (H.S.)}
 \end{aligned}$$

3.2.1.2 ACCESSIBLE

A1 Principle

“A1. (Meta)data are retrievable by their identifier using a standardized communications protocol”

A1. Paraphrased as:

“If an identifier and standardized communications protocol (e.g. Http and doi) are used, then the digital resource is more accessible.”

A1 Symbolized as:

UD: Accessible

I: identifier

S: standardized communications protocol

R: digital resource is more accessible

Which yields the following hybrid formula:

“If I and S are used, then R.”

A1 Translated as *wffs*:

$(I \wedge S) \rightarrow R$

$(I \wedge S)$

$\therefore R$

\equiv Modus Ponens (M.P.)

A1.1 Principle

“A1.1 The protocol is open, free, and universally implementable”

A1.1 Paraphrased as:

“If the protocol is open, and the protocol is free, and the protocol is universally implementable, then the digital resource is more accessible.”

A1.1 Symbolized as:

UD: Accessible

O: the protocol is open

F: the protocol is free

U: the protocol is universally implementable

A: the digital resource is more accessible

Which yields the following hybrid formula:

If O and F and U, then A

A1.1 Translated as *wffs*:

$(O \wedge F \wedge U) \rightarrow A$

$\therefore (O \wedge F \wedge U) \rightarrow ((O \wedge F \wedge U) \wedge A)$

\equiv Absorption (Abs.)

A1.2 Principle

“A1.2 The protocol allows for an authentication and authorization procedure, where necessary”

A1.2 Paraphrased as:

“If the protocol allows for an authentication procedure and the protocol allows for an authorization procedure, then the digital resource is more accessible.”

A1.2 Symbolized as:

UD: Accessible

C: the protocol allows for an authentication procedure

Z: the protocol allows for an authorization procedure

A: the protocol is more accessible

Which yields the following hybrid formula:

“If C and Z, then A.”

A1.2 Translated as *wffs*:

$$(C \wedge Z) \rightarrow A$$

$$\therefore (C \wedge Z) \rightarrow ((C \wedge Z) \wedge A)$$

$$\equiv \text{Absorption (Abs.)}$$

A2 Principle

“A2. Metadata are accessible, even when the data are no longer available”

A2 Paraphrased as:

“Even if it is not the case that the digital resource is available,
metadata is accessible.”

A2 Symbolized as:

UD: Accessible

D: digital resource is no longer available

M: metadata is accessible

Which yields the following hybrid formula:

“Even If it is not the case that D, then M”

A2 Translated as *wffs*:

$D \rightarrow M$

$\therefore D \rightarrow (D \wedge M)$

\equiv Absorption (Abs.)

3.2.1.3 INTEROPERABLE

I1 Principle

“I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.”

I1. Paraphrased as:

“If the digital resource uses a formal, accessible, shared, and broadly applicable language for knowledge representation, then the digital resource is more interoperable.”

I1 Symbolized as:

UD: Interoperable

F: digital resource uses a formal language for knowledge representation

A: digital resource uses an accessible language for knowledge representation

S: digital resource uses a shared language for knowledge representation

B: digital resource uses a broadly applicable language for knowledge representation

E: the digital resource is more interoperable

Which yields the following hybrid formula:

“If F and A and S and B, then E.”

I1 Translated as *wffs*:

$$(F \wedge A \wedge S \wedge B) \rightarrow E$$

$$\therefore (F \wedge A \wedge S \wedge B) \rightarrow ((F \wedge A \wedge S \wedge B) \wedge E)$$

$$\equiv \text{Absorption (Abs.)}$$

I2 Principle

“I2. (Meta)data use vocabularies that follow FAIR principles”

I2. Paraphrased as:

“If the digital resource uses vocabulary and that vocabulary follows FAIR, then the digital resource is more interoperable.”

I2. Symbolized as:

UD: Interoperable

U: the digital resource uses vocabulary

F: the vocabulary follows FAIR

I: the digital resource is more interoperable

Which yields the following hybrid formula:

If U and F, then I.

I2. Translated as *wffs*:

$$(U \wedge F) \rightarrow I$$

$$\begin{aligned} &\therefore (U \wedge F) \rightarrow ((U \wedge F) \wedge I) \\ &\equiv \text{Absorption (Abs.)} \end{aligned}$$

I3 Principle

“I3. (Meta)data include qualified references to other (Meta)data”

I3. Paraphrased as:

“If the digital resource includes qualified references to other (Meta)data, then the digital resource is more interoperable.”

I3. Symbolized as:

UD: Interoperable

Q: the digital resource includes qualified references to other (Meta)data

M: the digital resource is more interoperable

Which yields the following hybrid formula:

“If Q, then M.”

I3. Translated as *wffs*:

$$Q \rightarrow M$$

$$\therefore Q \rightarrow (Q \wedge M)$$

$$\equiv \text{Absorption (Abs.)}$$

3.2.1.4 REUSABLE

R1 Principle

“R1. Meta(data) are richly described with a plurality of accurate and relevant attributes”

R1. Paraphrased as:

“If the digital resource is richly described with a plurality of accurate attributes and the digital resource is richly described with a plurality of relevant attributes, then the digital resource is more reusable.”

R1 Symbolized as:

UD: Reusable

A: the digital resource is richly described with a plurality of accurate attributes

R: the digital resource is richly described with a plurality of relevant attributes

M: the digital resource is more interoperable

Which yields the following hybrid formula:

If A and R, then M.

R1 Translated as *wffs*:

$(A \wedge R) \rightarrow M$

$\therefore (A \wedge R) \rightarrow ((A \wedge R) \wedge M)$

\equiv Absorption (Abs.)

R1.1 Principle

“R1.1 (Meta)data are released with a clear and accessible data usage license”

R1.1 Paraphrased as:

“If the digital resource is released with a clear data usage license and the digital resource is released with an accessible data usage license, then the digital resource is more reusable.”

R1.1 Symbolized as:

UD: Reusable

C: the digital resource is richly described with a plurality of accurate attributes

A: the digital resource is richly described with a plurality of relevant attributes

R: the digital resource is more reusable

Which yields the following hybrid formula:

If C and A, then R.

R1.1 Translated as *wffs*:

$(C \wedge A) \rightarrow R$

$\therefore (C \wedge A) \rightarrow ((C \wedge A) \wedge R)$

\equiv Absorption (Abs.)

R1.2 Principle

“R1.2 (Meta)data are associated with detailed provenance”

R1.2 Paraphrased as:

“If the digital resource is associated with detailed provenance, then the digital resource is more reusable.”

R1.2 Symbolized as:

UD: Reusable

P: the digital resource is associated with detailed provenance

R: the digital resource is more reusable

Which yields the following hybrid formula:

“If P, then R.”

R1.2 Translated as *wffs*:

$P \rightarrow R$

$\therefore P \rightarrow (P \wedge R)$

\equiv Absorption (Abs.)

R1.3 Principle

“R1.3 (Meta)data meet domain-relevant community standards”

R1.3 Paraphrased as:

“If the digital resource meets domain-relevant community standards, then the digital resource is more reusable.”

R1.3 Symbolized as:

UD: Reusable

D: the digital resource meets domain-relevant community standards

I: the digital resource is more reusable

Which yields the following hybrid formula:

“If D, then I.”

R1.3 Translated as *wffs*:

$D \rightarrow I$

$$\begin{aligned} & \therefore D \rightarrow (D \wedge I) \\ & \equiv \text{Absorption (Abs)}. \end{aligned}$$

3.2.2 FORMAL LOGICAL ANALYSIS EVALUATION METHODS

After translating the FAIR sentences into arguments in a logical order, and we obtain the *wffs* for all the 15 FAIR principles, as defined by *definition 2.1*, we can apply logic methods to analyze them. We used the rules of logical inference described in Table 2.4 and Table 2.5 of chapter 2. In the following sections, we will delve into the logical analysis of the *wffs* of FAIR; first, we will explicate the replacement property of the logical inference rules, then, we will describe the formal proof of both the Truth Table method and then the Natural Deduction method.

3.2.2.1 REPLACEMENT PROPERTY OF THE LOGICAL INFERENCE RULES

Firstly, we must implement logical inference rules strictly. For instance, an argument that one proves valid using Modus Ponens (M.P.) must have the exact form “ $p, p \rightarrow q$, then q .” Then, we must consistently and correctly substitute the variable of every statement by any assertion (simple or compound). Afterward, we must precisely fit the primary argument form to the argument we work with; this is necessary because we want to learn with confidence that the outcome of our logic is valid. We can only verify it if we can show that any component in our chain of reasoning is concrete. Table 3.1 describes the well-formed formula *wffs* (i.e., created in section 3.2), the equivalent rule of inference, and the corresponding tautology forms for the FAIR data principles. The replacement property of the logical inference rules allows us

to use the tautology forms, as described in table 3.1, to prove the validity and tautology of the FAIR arguments [7].

Table 3.1: FAIR Principles, FAIR's *wffs*, Equivalent Rule of Inference, and its Tautology form.

FAIR Principles	Well-Formed Formulas (<i>wffs</i>) of FAIR	Equivalent Rule of Inference	Tautology form
F1	$D \rightarrow (G \wedge P)$ $(G \wedge P) \rightarrow (F \wedge R)$ $\therefore D \rightarrow (F \wedge R)$	Hypothetical Syllogism (H.S.)	$((p \rightarrow q) \wedge (q \rightarrow r)) \Rightarrow (p \rightarrow r)$
F2	$D \rightarrow M, D, \therefore M$	Modus Ponens (M.P.)	$(p \wedge (p \rightarrow q)) \Rightarrow q$
F3	$I \rightarrow (C \wedge E), (C \wedge E) \rightarrow F$ $\therefore I \rightarrow F$	Hypothetical Syllogism (H.S.)	$((p \rightarrow q) \wedge (q \rightarrow r)) \Rightarrow (p \rightarrow r)$
F4	$D \rightarrow (R \vee I), (R \vee I) \rightarrow F$ $\therefore D \rightarrow F$	Hypothetical Syllogism (H.S.)	$((p \rightarrow q) \wedge (q \rightarrow r)) \Rightarrow (p \rightarrow r)$
A1	$(I \wedge S) \rightarrow R, (I \wedge S), \therefore R$	Modus Ponens (M.P.)	$(p \wedge (p \rightarrow q)) \Rightarrow q$
A1.1	$(O \wedge F \wedge U) \rightarrow A$ $\therefore (O \wedge F \wedge U) \rightarrow ((O \wedge F \wedge U) \wedge A)$	Absorption (Abs.)	$(p \rightarrow q) \Rightarrow (p \rightarrow (p \wedge q))$
A1.2	$(C \wedge Z) \rightarrow A$ $\therefore (C \wedge Z) \rightarrow ((C \wedge Z) \wedge A)$	Absorption (Abs.)	$(p \rightarrow q) \Rightarrow (p \rightarrow (p \wedge q))$
A2	$D \rightarrow M, \therefore D \rightarrow (D \wedge M)$	Absorption (Abs.)	$(p \rightarrow q) \Rightarrow (p \rightarrow (p \wedge q))$

I1	$(F \wedge A \wedge S \wedge B) \rightarrow E$ $\therefore (F \wedge A \wedge S \wedge B) \rightarrow ((F \wedge A \wedge S \wedge B) \wedge E)$	Absorption (Abs.)	$(p \rightarrow q) \Rightarrow (p \rightarrow (p \wedge q))$
I2	$(U \wedge F) \rightarrow I$ $\therefore (U \wedge F) \rightarrow ((U \wedge F) \wedge I)$	Absorption (Abs.)	$(p \rightarrow q) \Rightarrow (p \rightarrow (p \wedge q))$
I3	$Q \rightarrow M, \therefore Q \rightarrow (Q \wedge M)$	Absorption (Abs.)	$(p \rightarrow q) \Rightarrow (p \rightarrow (p \wedge q))$
R1	$(A \wedge R) \rightarrow M$ $\therefore (A \wedge R) \rightarrow ((A \wedge R) \wedge M)$	Absorption (Abs.)	$(p \rightarrow q) \Rightarrow (p \rightarrow (p \wedge q))$
R1.1	$(C \wedge A) \rightarrow R$ $\therefore (C \wedge A) \rightarrow ((C \wedge A) \wedge R)$	Absorption (Abs.)	$(p \rightarrow q) \Rightarrow (p \rightarrow (p \wedge q))$
R1.2	$P \rightarrow R, \therefore P \rightarrow (P \wedge R)$	Absorption (Abs.)	$(p \rightarrow q) \Rightarrow (p \rightarrow (p \wedge q))$
R1.3	$D \rightarrow I, \therefore D \rightarrow (D \wedge I)$	Absorption (Abs.)	$(p \rightarrow q) \Rightarrow (p \rightarrow (p \wedge q))$

3.3 RESULTS OF THE FORMAL LOGICAL ANALYSIS FOR FAIR DATA PRINCIPLES

For a given argument, we have defined a formal proof of validity as a series of statements, as described in *definition 2.6*. Each statement being either a premise of that argument or results from the previous statements of the series by a preliminary valid argument or by a logical symmetry so that the last statement in the series is the conclusion of the argument whose validity has been proven [6]. We examined the *wffs* of the FAIR sentences that were

generated in section 3.2 and found that they are identical to the three types of rules of inference: 1) Hypothetical Syllogism (H.S.); 2) Modus Ponens (M.P.); and 3) Absorption (Abs.), as illustrated in Table 3.1. If we prove the corresponding tautology form of these three rules to be valid and tautology, then these proofs are also generalized for all the FAIR arguments since they are identical. We used two methods of proof the Truth Table method and the Natural Deduction method, as elucidated below.

3.3.1 FORMAL PROOF OF THE TRUTH TABLE METHOD

We have three groups: 1) Hypothetical Syllogism (H.S.) group that include F1, F3, and F4; 2) Modus Ponens (M.P.) group that includes F2 and A1; and 3) Absorption (Abs.) group that includes A1.1, A1.2, A2, I1, I2, I3, R1, R1.1, R1.2, and R1.3. Next, we will construct formal proofs for these three groups using the Truth Table method.

3.3.1.1 TRUTH TABLE METHOD FOR FAIR PRINCIPLES: F1, F3, AND F4

As specified in Table 3.1, FAIR data principles F1, F3, and F4 are logically equivalent to the Hypothetical Syllogism (H.S.) rule of inference. By *definition 2.2*, the Hypothetical Syllogism (H.S.) is a valid logical argument consisting of two premises and one conclusion, which altogether are constructed from three propositions p , q , and r , as shown in Table 3.1. By applying the replacement property of inference rules, we have:

$$\mathbf{F1} = [(D \rightarrow (G \wedge P)) \wedge ((G \wedge P) \rightarrow (F \wedge R))] \Rightarrow [(D \rightarrow (F \wedge R))]$$

$p=D$, $q=(G \wedge P)$, and $r=(F \wedge R)$. By substitution, we get: F1 is logically equivalent to the H.S.: $((p \rightarrow q) \wedge (q \rightarrow r)) \Rightarrow (p \rightarrow r)$

$\therefore F1 \equiv H.S.$

F3 = $[(I \rightarrow (C \wedge E)) \wedge ((C \wedge E) \rightarrow (F))] \Rightarrow [(I \rightarrow F)]$

$p=I$, $q=(G \wedge E)$, and $r=F$. By substitution, we get: F3 is logically equivalent to the H.S.: $((p \rightarrow q) \wedge (q \rightarrow r)) \Rightarrow (p \rightarrow r)$

$\therefore F3 \equiv H.S.$

F4 = $[(D \rightarrow (R \vee I)) \wedge ((R \vee I) \rightarrow F)] \Rightarrow [(D \rightarrow F)]$

$p=D$, $q=(R \vee I)$, and $r=F$. By substitution, we get: F4 is logically equivalent to the H.S.: $((p \rightarrow q) \wedge (q \rightarrow r)) \Rightarrow (p \rightarrow r)$

$\therefore F4 \equiv H.S.$

In Table 3.2, we prove that the (H.S.) is a valid logical argument using the Truth Table method, as illustrated in Table 3.1 (H.S.), can be formulated as $((p \rightarrow q) \wedge (q \rightarrow r)) \Rightarrow (p \rightarrow r)$, which is corresponding to its tautology form, as shown in Table 3.1.

Table 3.2: Proof of the Hypothetical Syllogism (H.S.) using Truth Table.

p	q	r	$(p \rightarrow q)$	\wedge	$(q \rightarrow r)$	\Rightarrow	$(p \rightarrow r)$
1	1	1	1	1	1	1	1
1	1	0	1	1	1	1	1
1	0	1	1	0	0	1	1
1	0	0	1	1	1	1	1
0	1	1	0	0	1	1	0
0	1	0	0	0	1	1	1

0	0	1	1	0	0	1	0
0	0	0	1	1	1	1	1

Validity: by observing Table 3.2, in the first row we can observe that $(p \rightarrow r)$ is true, whenever $(p \rightarrow q) \wedge (q \rightarrow r)$ are both true. As a result, by *definition 2.2* of a valid logical argument, the H.S. $((p \rightarrow q) \wedge (q \rightarrow r)) \Rightarrow (p \rightarrow r)$ is valid, and since the *wffs* of F1, F3, and F4 are identical to the H.S., as illustrated in Table 3.1, then the F1, F3, and F4 arguments are also valid.

Tautology: by calculating the Truth Table of the whole expression of H.S. (Table 3.2), we can observe the truth values (**bolded 1's**) of the main connective " \Rightarrow " of the H.S. expression: $((p \rightarrow q) \wedge (q \rightarrow r)) \Rightarrow (p \rightarrow r)$ are all "true, **1's**", as a result, by *definition 2.5*, which confirms that F1, F3, and F4 are tautology.

3.3.1.2 TRUTH TABLE METHOD FOR FAIR PRINCIPLES: F2 AND A1

As identified in Table 3.1, the FAIR data principles F2 and A1 are logically equivalent to the Modus Ponens (M.P.) rule of inference. By *definition 2.2*, the M.P. is a valid logical argument consisting of one premise and one conclusion, which altogether are constructed from two propositions p and q , as shown in Table 3.3. By applying the replacement property of inference rules, we have:

$$\mathbf{F2} = (D \wedge D \rightarrow M) \Rightarrow M$$

$p=D, q=M$. By substitution, we get F2 is logically equivalent to M.P.:

$$\mathbf{(p \wedge (p \rightarrow q)) \Rightarrow q}$$

$\therefore F2 \equiv M.P.$

$$A1 = ((I \wedge S) \wedge ((I \wedge S) \rightarrow R)) \Rightarrow (R)$$

$p = (I \wedge S)$, $q = R$. By substitution, we get A1 is logically equivalent to M.P.:

$$(p \wedge (p \rightarrow q)) \Rightarrow q$$

$\therefore A1 \equiv M.P.$

We prove that the M.P. is a valid logical argument using the Truth Table method, as in Table 3.3. The M.P. can be reformulated as $(p \wedge (p \rightarrow q)) \Rightarrow q$, which is corresponding to its tautology form, as shown in Table 3.1.

Table 3.3: Proof of the Modus Ponens (M.P.) using Truth Table.

p	q	$((p)$	\wedge	$(p \rightarrow q))$	\Rightarrow	(q)
1	1	1	1	1	1	1
1	0	1	0	0	1	0
0	1	0	0	1	1	1
0	0	0	0	1	1	0

Validity: by observing Table 3.3, in the first row, we can see that (q) is true, whenever $(p) \wedge (p \rightarrow q)$ are both true. As a result, by *definition 2.2* of a valid logical argument, the M.P., $(p \wedge (p \rightarrow q)) \Rightarrow q$ is valid, and since the *wffs* of F2 and A1 are identical to the M.P., then the F2 and A1 arguments are also valid.

Tautology: by calculating the Truth Table for the whole expression of M.P. (Table 3.3), we can observe the true values (**bolded 1's**) of the main connective “ \Rightarrow ” of the M.P. expression: $(p \wedge$

$(p \rightarrow q) \Rightarrow q$ are all “true, 1’s”, as a result, by *definition 2.5*, which confirms that F2 and A1 are tautology.

3.3.1.3 TRUTH TABLE METHOD FOR FAIR PRINCIPLES: A1.1, A1.2, A2, I1, I2, I3, R1, R1.1, R1.2, AND R1.3

FAIR principles A1.1, A1.2, A2, I1, I2, I3, R1, R1.1, R1.2, and R1.3 are all logically equivalent to the Absorption (Abs.) inference rule $(p \rightarrow q) \Rightarrow (p \rightarrow (p \wedge q))$, as shown in Table 3.1. By *definition 2.2*, the Absorption (Abs.) inference rule is a valid logical argument. As an example, applying the replacement property of Absorption (Abs.) inference rules to FAIR principle A1.1, we get:

$$\mathbf{A1.1} = ((O \wedge F \wedge U) \rightarrow A) \Rightarrow ((O \wedge F \wedge U) \rightarrow ((O \wedge F \wedge U) \wedge A))$$

$p = (O \wedge F \wedge U)$, and $q = A$. Then, by substitution, we get A1.1 is logically equivalent to the Absorption (Abs.): $(p \rightarrow q) \Rightarrow (p \rightarrow (p \wedge q))$

$$\therefore \mathbf{A1.1} \equiv \mathbf{Abs.}$$

Similarly, we apply the same method for A1.2, A2, I1, I2, I3, R1, R1.1, R1.2, and R1.3; then, we have them all being logically equivalent to the Absorption (Abs.) inference rule.

We prove that the Absorption (Abs.) inference rule is a valid logical argument using the Truth Table method, as illustrated in Table 3.4. The Absorption (Abs.) inference rule can be formulated as $(p \rightarrow q) \Rightarrow (p \rightarrow (p \wedge q))$, which is corresponding to its tautology form, as shown in Table 3.1.

Table 3.4: Proof of Absorption (Abs.) using Truth Table.

p	q	$(p \rightarrow q)$	\Rightarrow	(p)	\rightarrow	$(p \wedge q)$
1	1	1	1	1	1	1
1	0	0	1	1	0	0
0	1	1	1	0	1	0
0	0	1	1	0	1	0

Validity: by observing Table 3.4, in the first row, we can see that $(p \rightarrow (p \wedge q))$ is true, whenever $(p \rightarrow q)$ is true. As a result, by *definition 2.2* of a valid logical argument, the Abs.: $(p \rightarrow q) \Rightarrow (p \rightarrow (p \wedge q))$ is valid, and since the *wffs* of A1.1, A1.2, A2, I1, I2, I3, R1, R1.1, R1.2, and R1.3 are identical to the M.P., then the A1.1, A1.2, A2, I1, I2, I3, R1, R1.1, R1.2, and R1.3 arguments are also valid.

Tautology: by calculating the Truth Table for the whole expression of Absorption (Abs.), as shown in Table 3.4, we can observe the true value (**bolded 1's**) of the main connective “ \Rightarrow ” of the Absorption (Abs.) expression $(p \rightarrow q) \Rightarrow (p \rightarrow (p \wedge q))$ are all “true, 1's”, as a result, by *definition 2.5*, which confirms that A1.1, A1.2, A2, I1, I2, I3, R1, R1.1, R1.2, and R1.3 are tautology.

3.3.2 FORMAL PROOF OF NATURAL DEDUCTION METHOD

In this section, we discuss another logical method called the Natural Deduction method. Same as the previous section we have three groups: 1) Hypothetical Syllogism (H.S.) group that includes F1, F3, and F4; 2) Modus Ponens (M.P.) group that include F2 and A1; and 3) Absorption (Abs.) group that includes A1.1, A1.2, A2, I1, I2, I3, R1, R1.1, R1.2, and R1.3.

Next, we will construct formal proofs for these three groups using the Natural Deduction method.

3.3.2.1. NATURAL DEDUCTION METHOD FOR FAIR PRINCIPLES: F1, F3, AND F4

As seen in section 3.2.3 Hypothetical Syllogism (H.S.) is logically identical to F1, F3, and F4; therefore, we need only to show the Natural Deduction method proof of (H.S.) which is also the proof of validity and tautology for F1, F3, and F4 principles. Table 3.5 illustrates the steps and reasons of the proof in which the conclusion was driven from the premises.

Table 3.5: Natural Deduction method formal Proof of (H.S.) $((p \rightarrow q) \wedge (q \rightarrow r)) \Rightarrow (p \rightarrow r)$

<i>Step</i>	<i>Reason</i>
1. p	Assumption
2. $p \rightarrow q$	Premise
3. q	Modus ponens from (1) and (2)
4. $q \rightarrow r$	Premise
5. r	Modus ponens from (3) and (4)
6. $p \rightarrow r$	Direct method proof (\rightarrow I) from (1) and (5)

3.3.2.2 NATURAL DEDUCTION METHOD FOR FAIR PRINCIPLES: F2 AND A1

As shown in section 3.2.3, Modus Ponens (M.P.) is logically equivalent to F2 and A1, so we only need to demonstrate the Natural Deduction method proof of M.P. This is also verification of validity and tautology for the principles F2 and A1. Table 3.6 outlines the steps and explanations provided by the facts under which the inference was derived from the premises.

Table 3.6: Proof of Modus Ponens using Natural Deduction method $(p \wedge (p \rightarrow q)) \Rightarrow q$

<i>Step</i>	<i>Reason</i>
1. $p \wedge (p \rightarrow q)$	Premise
2. p	Conjunction elimination from (1) ($\wedge E$)
3. $p \rightarrow q$	Conjunction elimination from (1) ($\wedge E$)
4. q	Modus Ponens from (2) and (3)

3.3.2.3 NATURAL DEDUCTION METHOD FOR FAIR PRINCIPLES: A1.1, A1.2, A2, I1, I2, I3, R1, R1.1, R1.2, AND R1.3

Section 3.2.3 showed the Absorption (Abs.) inference rule is logically equivalent to A1.1, A1.2, A2, I1, I2, I3, R1, R1.1, R1.2, and R1.3; we only need to demonstrate the Natural Deduction method proof of Absorption (Abs.), which is also proof of validity and tautology for A1.1, A1.2, A2, I1, I2, I3, R1, R1.1, R1.2, and R1.3 principles. Table 3.7 describes the steps and the explanations behind the circumstances under which the inference was driven from the premises.

Table 3.7: Natural Deduction method for Absorption (Abs.): $(p \rightarrow q) \Rightarrow (p \rightarrow (p \wedge q))$.

<i>Step</i>	<i>Reason</i>
1. $p \rightarrow q$	Premise
2. $\neg p \vee q$	Material Implication (Table 2.5)
3. $\neg p \vee p$	Law of excluded middle (LEM) (Table 2.5)
4. $(\neg p \vee p) \wedge (\neg p \vee q)$	Conjunction Introduction (2), (3) ($\wedge I$)
5. $\neg p \vee (p \wedge q)$	Reverse Distribution (Table 2.5)
6. $p \rightarrow (p \wedge q)$	Material Implication (Table 2.5)

3.3.3 DISCUSSION OF THE FINDINGS

To summarize the findings, we illustrated in Table 3.8 the inference rules, the identical FAIR principles, and their proven validity and tautology.

Table 3.8: Summary of FAIR principles validity and tautology proofs.

Inference Rule	Identical FAIR principle	Proven Valid	Proven Tautology
1. Hypothetical Syllogism (H.S.)	F1, F3, F4	Yes	Yes
2. Modus Ponens (M.P.)	F2, A1	Yes	Yes
3. Absorption (Abs.)	A1.1, A1.2, A2, I1, I2, I3, R1, R1.1, R1.2, R1.3	Yes	Yes

The logical analysis of the FAIR data principles resulted in the following Lemmas:

Lemma 1: We show the result of the Truth Table method and the Natural Deduction method for F1, F3, and F4; we conclude that the F1, F3, and F4 arguments are both valid and tautology. Therefore, by *definition 2.3*, since they are tautology, they are theorems.

Lemma 2: As a result of the proof of both the Truth Table method and the Natural Deduction method, we conclude that the FAIR's F2 and A1 arguments are both valid and tautology. Therefore, by *definition 2.3*, since they are tautology, they are theorems.

Lemma 3: As a result of the proof of both the Truth Table method and the Natural Deduction method for the Abs., we conclude that the FAIR's A1.1, A1.2, A2, I1, I2, I3, R1, R1.1, R1.2, and R1.3 arguments are all valid and tautology. By *definition 2.3*, since they are tautology, they are theorems.

As outlined in Table 3.9, the theorems of Findable, Accessible, Interoperable, Reusable are described. Consequently, these theorems are used to construct the FAIR theorem.

Table 3.9: Findable, Accessible, Interoperable, Reusable Theorems.

Theorem Name	Justification	Theorem Statement
Findable theorem	By using Lemma 1,2 and 3, if F1, F2, F3, and F4 are tautology, then Findable is a theorem.	<i>For every digital resource in the domain, if, and only if, the digital resource has satisfied F1, F2, F3, and F4, then that digital resource is Findable.</i>
Accessible theorem	By Lemma 1, 2, and 3, if A1, A1.1, A1.2, and A2 are tautology, then Accessible is a theorem.	<i>For every digital resource in the domain, if, and only if, the digital resource has satisfied A1, A1.1, A1.2, and A2, then that digital resource is Accessible.</i>
Interoperable theorem	By Lemma 1, 2, and 3, if I1, I2, and I3 are tautology, then Interoperable is a theorem.	<i>For every digital resource in the domain, if, and only if, the digital resource has satisfied I1, I2, and I3, then that digital resource is Interoperable.</i>
Reusable theorem	By Lemma 1, 2, and 3, if R1, R1.1, R1.2, and R1.3 are tautology, then Reusable is a theorem.	<i>For every digital resource in the domain, if, and only if, the digital resource has satisfied R1, R1.1, R1.2, and R1.3, then that digital resource is Reusable.</i>

In the next section, we advance the discussion on the construction of FAIR theorem using modus ponens and universal modus ponens inference rules.

3.3.4 CONSTRUCTION OF FAIR THEOREM

If we apply the Modus Ponens (M.P.) inference rule on the above theorems we get:

$p \rightarrow q = (\text{Findable, Accessible, Interoperable, and Reusable}) \rightarrow (\text{FAIR})$ *premise 1*

$p = (\text{Findable, Accessible, Interoperable, and Reusable})$ *premise 2 (Fact, Table 3.9)*

$\therefore q = \therefore (\text{FAIR})$ *conclusion*

Therefore, we affirm the following statement:

If (Findable, Accessible, Interoperable, and Reusable) are theorems, then (FAIR) is a theorem.

Table 3.10 FAIR Theorem

Theorem Name	Justification	Theorem Statement
FAIR theorem	If (Findable, Accessible, Interoperable, and Reusable) are theorems, then (FAIR) is a theorem.	<i>For every digital resource in the domain, if, and only if, the digital resource is Findable, Accessible, Interoperable, and Reusable, then that digital resource is FAIR.</i>

To verify the proof for the FAIR theorem, consider the following model:

Domain: FAIR digital resources

F(x): x is Findable

$A(x)$: x is Accessible

$I(x)$: x is Interoperable

$R(x)$: x is Reusable

$Q(x)$: x is FAIR

According to this model we construct the following formula:

$\forall x((F(x) \wedge A(x) \wedge I(x) \wedge R(x)) \rightarrow Q(x))$ which reads as follow: “for every x in the domain, if x is Findable, and x is Accessible, and x is Interoperable, and x is Reusable, then x is FAIR.”

Furthermore, to help prove the validity and tautology of $Q(x)$ we introduce a new inference rule called **universal modus ponens** [7], which combines universal instantiation rule “ $(\forall x P(x) \rightarrow P(c))$ for any c in the domain)” and modus ponens rule “ $((P \wedge (P \rightarrow Q)) \rightarrow Q)$ ”. This combined rule functions as follows:

If $\forall x(P(x) \rightarrow Q(x))$ is true, and if $P(c)$ is true in the universal quantifier domain for a particular element, then the conclusion $Q(c)$ must also be true. Note that $P(c) \rightarrow Q(c)$ is true by universal instantiation. Then, $Q(c)$ must also be true by means of modus ponens. Therefore, universal modus ponens can be described as:

$$\forall x(P(x) \rightarrow Q(x))$$

$P(c)$, where c is a particular element in the domain

$$\therefore Q(c)$$

By understanding the previous discussion then we can show that the premises “All digital resources in the domain are FAIR digital resources” and “Iris dataset is a digital resource in this domain” imply the conclusion “Iris dataset is a FAIR digital resource”.

To prove this argument: Let $P(x)$ denote “ x is a digital resource in the FAIR digital resources’ domain,” and let $Q(x)$ denote “ x is FAIR.” Then the premises are $\forall x(P(x) \rightarrow Q(x))$ and $P(\text{Iris})$, where Iris dataset is a specific digital resource in the FAIR digital resources’ domain. Therefore, the conclusion is $Q(\text{Iris})$. The following steps show the Natural Deduction method proof of universal modus ponens to deduce the conclusion from the premises.

<i>Step</i>	<i>Reason</i>
1. $\forall x(P(x) \rightarrow Q(x))$	Premise
2. $P(\text{Iris}) \rightarrow Q(\text{Iris})$	Universal instantiation from (1)
3. $P(\text{Iris})$	Premise
4. $Q(\text{Iris})$	Modus ponens from (2) and (3)

Lastly, *definition 2.4* implies that a theorem is a conclusion shown to be true by writing a proof; section 3.3.4 showed the proof of the FAIR theorem; therefore, we can affirm the FAIR theorem, as depicted in Table 3.10.

3.4 DISCUSSION

To enhance the data management and stewardship of digital resources of Earth science, we seek to implement FAIR data principles. To concrete the road for this implementation first, we need to logically-analyze the validity and tautology of FAIR data principles. For that reason, we proposed the following research question: 1) How can we verify the validity and tautology of FAIR data principles? To answer this question, section 3.2 explained the process for verifying the FAIR data principles validity and tautology; in brief, the work performed by using two well-known logical analysis methods the Truth Table and the Natural Deduction. Therefore, we affirm the validity and tautology for the FAIR data principles. On the other front,

the equally important matter is the generation of FAIR theorems; for this purpose, we form the second research question: 2) How can we theoretically address FAIR data principles? To answer this question, section 3.3 described the steps of the metamorphosis of the FAIR principles to FAIR theorems. We find that the Truth Table and the Natural Deduction methods associated with arguments inference rules such as Hypothetical Syllogism (H.S.), Modus Ponens (M.P.), and Absorption (Abs.) are rational research methods; thus, we used them to deduce FAIR theorems out of the FAIR principles. Finally, to the best of our knowledge, this is novel research with no work in the literature to compare.

3.5 CONCLUSIONS

Our new focus on utilizing the formal logic with the natural language to analyze FAIR data principles opens a door for scientists to contribute to this direction and benefit the scientific community. The prove of the validity and tautology of the FAIR principles besides the resulting FAIR theorems denote a significant contribution to the data science scientific community. These findings help pave the way for the technical implementation of FAIR data principles. In turn, the work will improve the stewardship of Earth science digital resources in the cyberinfrastructure. In chapter 4, we will delve into the technical implementation of these FAIR theorems.

Chapter 4: Technical Design and Implementation of FAIR Framework

4.1 INTRODUCTION

We have seen in chapter 3 the inference rules govern the validity and tautology of an argument. To uncover these rules, we analyzed FAIR data principles arguments to examine their validity and tautology. The analysis went through four steps: 1) Build an argument for the FAIR sentence; 2) Paraphrase, Symbolize, Translate of FAIR sentences; 3) Analyze FAIR argument validity and tautology using the Truth Table method; and 4) Analyze FAIR argument validity and tautology using the Natural Deduction method. The outcome of these processes was the generation of FAIR theorems.

In this chapter, we will show the implementation of these FAIR data principles in a state-of-the-art semantic web technology based on the newly generated FAIR theorems. In the following sections, we will describe the three steps of the technical implementation of the FAIR data principles: 1) designing the FAIR Ontology; 2) designing the interface; and 3) setting up the system and demonstrating a use case of a dataset called “NCDC Storm Events Database [74]” to verify the functionalities of FAIRtool.org in compliance with the 15 FAIR data principles. The outcome of this use case is provided in RDF format, which is the only encoding format permit publishing structured data on the Web. We argue that using FAIR with semantic web technologies in Earth science will make their datasets better Findable, Accessible, Interoperable, and Reusable (i.e., better FAIR).

4.2 TECHNICAL DESIGN

Our technical solution framework is a semantic web platform based on Vitro [23], which is also the base for VIVO [68]. This platform allows building a semantic web application for the desired solution. Furthermore, this semantic web application depends on an ontology called FAIR ontology, which was built in-house using a tool called protégé, as shown in Fig. 4.1, which was generated by protégé, the classes of FAIR ontology and the relations between them. We built a semantic web application called FAIRtool.org; furthermore, we used a triple store to store the FAIRtool.org application triples. To verify FAIRtool.org functionalities, we employed a use case from the Earth science domain. The following sections describe in detail the technical design of FAIRtool.org.

4.2.1 THE DESIGN OF FAIR SEMANTIC WEB ONTOLOGY

To build a FAIR ontology, we reused several classes and properties from a wide range of third-party ontologies. Furthermore, these classes and properties come from multiple well-known standard vocabularies such as Schema.org, Data Catalog Vocabulary (dct), FOAF, Provenance vocabulary (PROV), Data Quality Vocabulary (dqv), and ComputerNetworks (cn), which will be used to construct the FAIRtool.org. As shown in Fig. 4.2, the FAIR ontology consists of OWL classes, OWL object properties, and OWL data properties. The object properties represent the relationship between the classes. The data properties represent the relation between the classes and the data values of certain classes. In the following sections, we will distribute these classes and properties on the FAIR data principles and illustrate how each principle will fit with its related classes and properties.

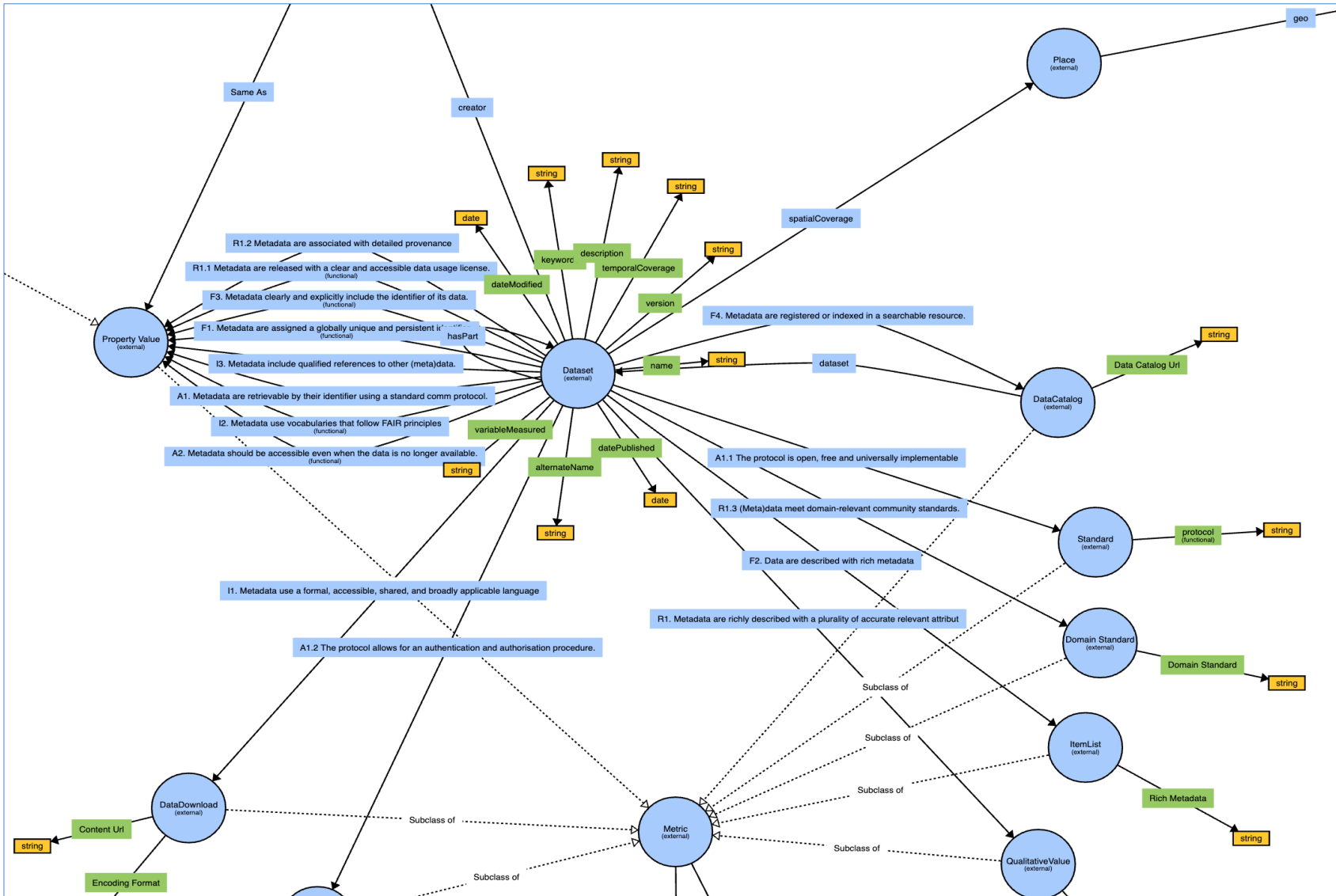


Fig. 4.2: Classes and properties of FAIR Ontology – (Generated using VOWL)

4.2.1.1 FAIR ONTOLOGY COMPONENTS

Findable: F1 principle is represented by the following classes and properties: classes are `schema:Dataset`, `schema:PropertyValue`, and `dqv:Metric`; object property is `schema:identifier`; and data properties are `schema:text`, `schema:url`, and `dqv:value`. The object property `schema:identifier` is dedicated to assign Persistent Identifiers (PIDs), a globally unique identifier, to the digital objects located in `schema:Dataset`. These PIDs can be acquired from identifier registrar agencies such as DOIs, Handles, and ARK identifiers [50]. Further, the class `schema:PropertyValue` is responsible for holding the identifiers for each digital object including its text and URL values. Besides, the `dqv:Metric` is responsible for recording the rating value of each FAIR principle. The text and URL values of the identifier and the metric are handled by the data properties `schema:text` and `schema:url`; the data type of the data properties `schema:text` is `xsd:string`, and the data type of the data properties `schema:url` is `xsd:anyURI`.

F2 principle is represented by the following classes and properties: classes are `schema:Dataset`, `schema:ItemList`, and `dqv:Metric`; object property is `schema:itemListElement`; and data properties from `schema.org` vocabulary are `Creator`, `Date Modified`, `Date Published`, `Keywords`, `Spatial Coverage`, `Temporal Coverage`, `Variable Measured`, `Version`, `Dataset Name`, `Description`, `AlternateName`. The object property `schema:itemListElement` is dedicated to establish a relation between `schema:Dataset` class and `schema:ItemList` class. All the rich metadata such as `Creator`, `Date Modified`, `Date Published`, `Keywords`, `Spatial Coverage`, `Temporal Coverage`, `Variable Measured`, `Version`, `Dataset Name`, `Description`, `AlternateName`, and `Rating` are entered and stored under `schema:ItemList` class.

F3 principle is represented by the following classes and properties: classes are schema:Dataset, schema:PropertyValue, and dqv:Metric; object property is schema:identifier; data properties are schema:text, schema:url, and dqv:value. This principle is responsible for holding the identifier of the digital object (i.e., data) that was assigned to the data in F1 and storing it as part of the metadata. The object property schema:identifier is dedicated to draw the relationship between the digital object (i.e. data) in the schema:Dataset class and its values located in the class schema:PropertyValue. The text and URL values of the identifier are handled by the data properties schema:text and schema:url; the data type of the data properties schema:text is xsd:string, and the data type of the data properties schema:url is xsd:anyURI.

F4 principle is represented by the following classes and properties: classes are schema:Dataset, schema:DataCatalog, and dqv:Metric; object property is schema:includInDataCatalog; data properties are schema:text, schema:url, and dqv:value. This principle is responsible for recording the registration information for the digital object (i.e., data) in a searchable resource like data.gov. The object property schema:includInDataCatalog is dedicated to draw the relationship between the digital object (i.e. data) in the schema:Dataset class and its values located in the class schema:DataCatalog. The value and URL of the searchable resource are handled by the data properties schema:text and schema:url; the data type of the data properties schema:text is xsd:string, and the data type of the data properties schema:url is xsd:anyURI.

Accessible: A1 principle is represented by the following classes and properties: classes are schema:Dataset, cn:Protocol, and dqv:Metric; object property is cn:standard_of; data

properties are `schema:text`, `schema:url`, and `dqv:value`. This principle means that the identifier of the digital object must use standard communication protocols such as HTTP. The class `cn:Protocol` records the type of the protocol. The Object property: `cn:standard_of` establishes the relation between the class `schema:Dataset`, which holds the digital resource, and the class `cn:Protocol`, which holds its text and URL values. The value and URL of the protocol are handled by the data properties `schema:text` and `schema:url`; the data type of the data properties `schema:text` is `xsd:string`, and the data type of the data properties `schema:url` is `xsd:anyURI`.

A1.1 principle is represented by the following classes and properties: classes are `cn:Protocol`, `dct:Standard`, and `dqv:Metric`; object property is `dct:conformsTo`; data properties are `schema:text`, `schema:url`, and `dqv:value`. A1.1 principle is an extension to principle A1; it gives more description to the identifying protocol. The class `cn:Protocol` records the type of that protocol. The Object properties: `dct:conformsTo` establishes the relationship between the class `cn:Protocol`, which holds the protocol type of the digital resource, and the class `dct:Standard`, which holds its text and URL values. The value and URL of the protocol are handled by the data properties `schema:text` and `schema:url`; the data type of the data properties `schema:text` is `xsd:string`, and the data type of the data properties `schema:url` is `xsd:anyURI`.

A1.2 principle is represented by the following classes and properties: classes are `cn:Protocol`, `dct:RightsStatement`, and `dqv:Metric`; object property is `dct:accessRights`; data properties are `schema:text`, `schema:url`, and `dqv:value`. A1.2 is another extension to principle A.1. The Object property `dct:accessRights` establishes the relationship between the class `cn:Protocol`, which holds the protocol type of the digital resource, and the class

dct:RightsStatement, which holds the rights statement that states the authorization and authentication procedure for that protocol. The text and URL values of the rights statement are handled by the data properties schema:text and schema:url; the data type of the data properties schema:text is xsd:string, and the data type of the data properties schema:url is xsd:anyURI.

A2 principle is represented by the following classes and properties: classes are schema:Dataset, Schema:PropertyValue, and dqv:Metric; object property is schema:isPartOf; data properties are schema:text, schema:url, and dqv:value. A2 principle is devoted to ensuring the availability of the metadata in a persistence repository like data.gov. The object properties schema:isPartOf establish the relationship between the class schema:Dataset, which holds the digital object, and the class schema:PropertyValue, which holds the descriptive text and URL of the metadata in a persistence repository, e.g. data.gov. The descriptive text and URL value of the metadata are handled by the data properties schema:text and schema:url; the data type of the data properties schema:text is xsd:string, and the data type of the data properties schema:url is xsd:anyURI.

Interoperable: I1 principle is represented by the following classes and properties: classes are schema:Dataset, Schema:DataDownload, and dqv:Metric; object property is schema:distribution; data properties are schema:encodingFormat, schema:contentUrl, and dqv:value. The object property schema:distribution is dedicated to establish the relationship between the class schema:Dataset, which holds the digital object, and the class schema:DataDownload, which holds the formal, accessible, shared, encoding Format and the content URL of the digital object. The descriptive text and URL value of the digital object are

handled by the data properties `schema:encodingFormat` and `schema:contentUrl`; the data type of the data property `schema:encodingFormat` is `xsd:string`, and the data type of the data property `schema:contentUrl` is `xsd:anyURI`.

I2 principle is represented by the following classes and properties: classes are `schema:Dataset`, `dct:Standard`, and `dqv:Metric`; object property is `void:vocabulary`; data properties are `schema:text`, `schema:url`, and `dqv:value`. The object property `dct:conformsTo` establish the relationship between the class `schema:Dataset`, which holds the digital resource, and the class `dct:Standard`, which holds the ontology that the cataloged digital resource content conforms to. The meaning of conformance is determined by provisions in the target standard, i.e., FAIR ontology. The text and url values of the FAIR ontology are handled by the data properties `schema:text` and `schema:url`; the data type of the data properties `schema:text` is `xsd:string`, and the data type of the data properties `schema:url` is `xsd:anyURI`.

I3 principle is represented by the following classes and properties: classes are `schema:Dataset`, `schema:PropertyValue`, and `dqv:Metric`; object property is `dct:references`; data properties are `schema:text`, `schema:url`, and `dqv:value`. The object property `dct:references` establish the relationship between the class `schema:Dataset`, which holds the metadata of the digital resource, and the class `schema:PropertyValue`, which holds the qualified references to other metadata in a persistent repository. The text and URL values of the qualified references is handled by the data properties `schema:text` and `schema:url`; the data type of the data properties `schema:text` is `xsd:string` and the data type of the data properties `schema:url` is `xsd:anyURI`.

Reusable: R1 principle is represented by the following classes and properties: classes are `schema:Dataset`, `dqv:QualityMetadata`, and `dqv:Metric`; object property is `dqv:hasQualityMetadata`; data properties are `schema:text`, `schema:url`, and `dqv:value`. The object property `dqv:hasQualityMetadata` establishes the relationship between the class `schema:Dataset`, which holds the metadata of the digital resource, and the class `dqv:QualityMetadata`, which holds the accurate relevant attributes of the digital object. The text and URL values of the quality metadata are handled by the data properties `schema:text` and `schema:url`; the data type of the data properties `schema:text` is `xsd:string`, and the data type of the data properties `schema:url` is `xsd:anyURI`.

R1.1 principle is represented by the following classes and properties: classes are `schema:Dataset`, `dct:LicenseDocument`, and `dqv:Metric`; object property is `schema:license`; data properties are `schema:text`, `schema:url`, and `dqv:value`. The object property `schema:license` establishes the relationship between the class `schema:Dataset`, which holds the metadata of the

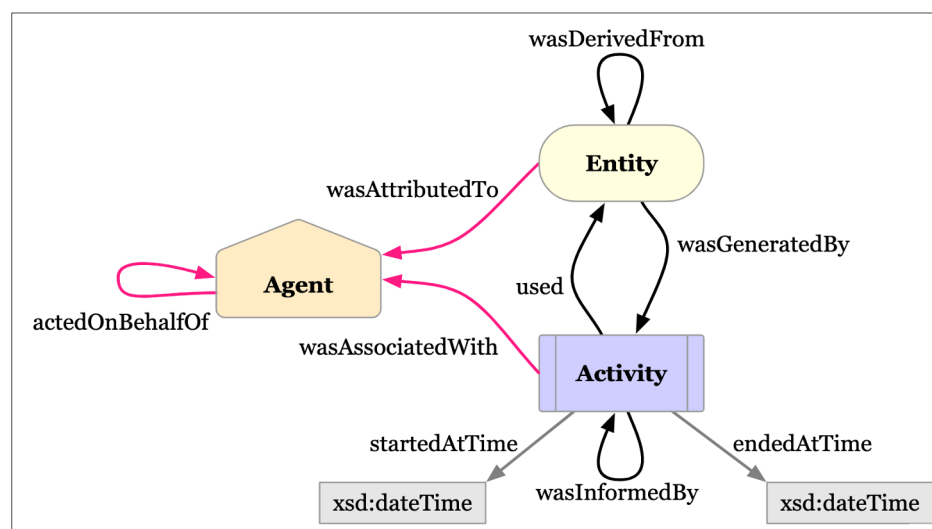


Fig. 4.3: Provenance ontology of a digital recourse (Adapted from W3C PROV)

digital resource, and the class `dct:LicenseDocument`, which holds the license document of that digital object, e.g. Creative Commons (CC) document. The text and URL values of the license document are handled by the data properties `schema:text` and `schema:url`; the data type of the data properties `schema:text` is `xsd:string`, and the data type of the data properties `schema:url` is `xsd:anyURI`.

R1.2 principle is the provenance and represented by the following classes and properties: classes are `prov:Entity`, `prov:Activity`, `prov:Agent`, and `dqv:Metric`; object properties are `prov:wasGeneratedBy`, `wasAssociatedWith`, and `wasDrivenFrom`; data properties are `schema:text`, `schema:url`, and `dqv:value`. As shown in Fig. 4.3, the object property `prov:wasGeneratedBy` is intended to identify the activity, which involves collecting the data, that creates an entity, the `wasAssociatedWith` is intended to identify the agent (e.g. person or software) that performs this activity, and `prov:wasDrivenFrom` is intended to identify the other entity involved in generating this entity [33]. The classes `prov:Entity`, `prov:Activity`, `prov:Agent` hold the provenance of that digital object. The text and URL value are handled by the data properties `schema:text` and `schema:url`; the data type of the data properties `schema:text` is `xsd:string`, and the data type of the data properties `schema:url` is `xsd:anyURI`.

R1.3 principle is represented by the following classes and properties: classes are `schema:Dataset`, `dct:Standard`, and `dqv:Metric`. Object property is `FAIR:domainRelevant`; Data properties are `schema:text`, `schema:url`, and `dqv:value`. The object property `FAIR:domainRelevant` establishes the relationship between the class `schema:Dataset`, which holds the digital resource, and the class `dct:Standard`, which holds the community-standard

ontology that the digital resource content conforms to. The meaning of conformity is determined in accordance with the provisions of the target standard. For example, the domain-relevant community standard of Earth science (ISO-19115 Geographic Metadata Information) is different from the domain-relevant community standard of biological science (i.e., the attributes will have a different meaning); therefore, the ontology must be different. The text and URL values of the R1.3 principle are handled by the data properties `schema:text` and `schema:url`; the data type of the data properties `schema:text` is `xsd:string`, and the data type of the data properties `schema:url` is `xsd:anyURI`.

Finally, for all the 15 FAIR data principles, the class `dqv:Metric` from Data Quality Vocabulary (`dqv`) is responsible for recording the rating value, and the data properties `dqv:value` is used to record the rating. This is very important because we will use them in the FAIRness level evaluation in chapter 5.

4.2.2 THE DESIGN OF THE INTERFACE

We took into consideration the ease of use when designing the interface for the FAIRtool.org, which consists of four tabs representing the four pillars (i.e., Findable, Accessible, Interoperable, and Reusable) of FAIR data principles. The details of each tab of FAIR data principles are described in the consequence sections:

First, the Findable tab includes four data-entry sections representing the four findable components, i.e., F1, F2, F3, and F4, as shown in Fig. 4.4. The metadata that can be entered for F1 is the identifier of the digital resource. Well-known identifiers can be inserted such as doi, ARK, handle, and ISBN. For F2, the rich metadata of the digital resource can be entered through

a list of already identified fields representing rich metadata for Earth science. The rich metadata that can be entered into the system are 1. Name of the digital resource; 2. Creator; 3. Date Published; 4. Date Modified; 5. Description; 6. AlternateName; 7. Spatial Coverage, i.e., City and Geo Shape; 8. Temporal Coverage; 9. Variable Measured; 10. Version; and 11. Keywords. For F3, the user can enter the identifier of the digital resource to be part of its metadata. For F4, the system allows the user to enter the name and web-link of the repository that the metadata is indexed or registered in.

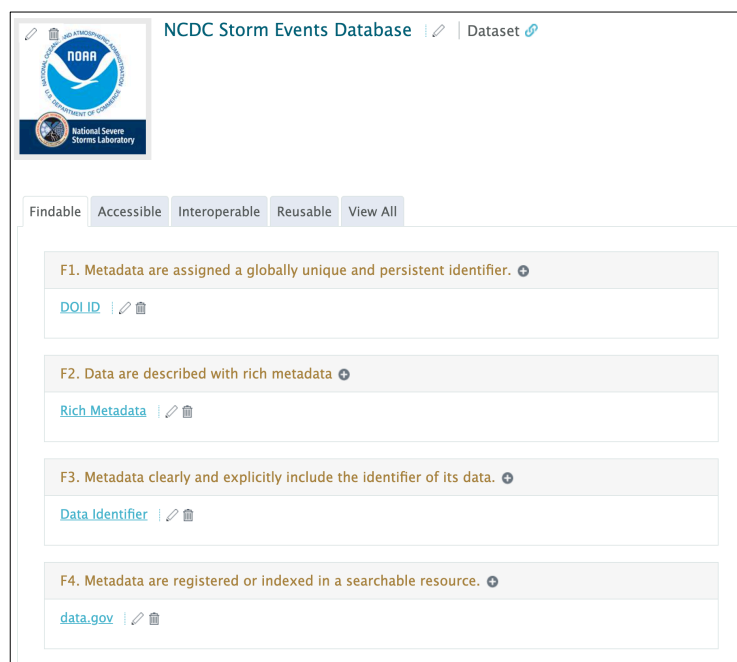


Fig. 4.4: Screenshot of the Findable entry interface of FAIRtool.org

Second, the Accessible tab includes four data-entry sections representing the four accessible components, i.e. A1, A1.1, A1.2, and A2. as shown in Fig. 4.5. For A1, the metadata element that can be entered is the hyperlink of the identifier via a standard communication protocol such as HTTP. So, the metadata is accessible by their identifier through a standard

communication protocol, e.g., <https://doi.org/10.1575/1912/bco-dmo.665253>. A1.1 is extension to principle A1 that indicates that the standard communication protocol used should be open, free, and universally implementable. Therefore, a document containing this information can be entered in this field. Also, A1.2 is another extension of A1 principle that

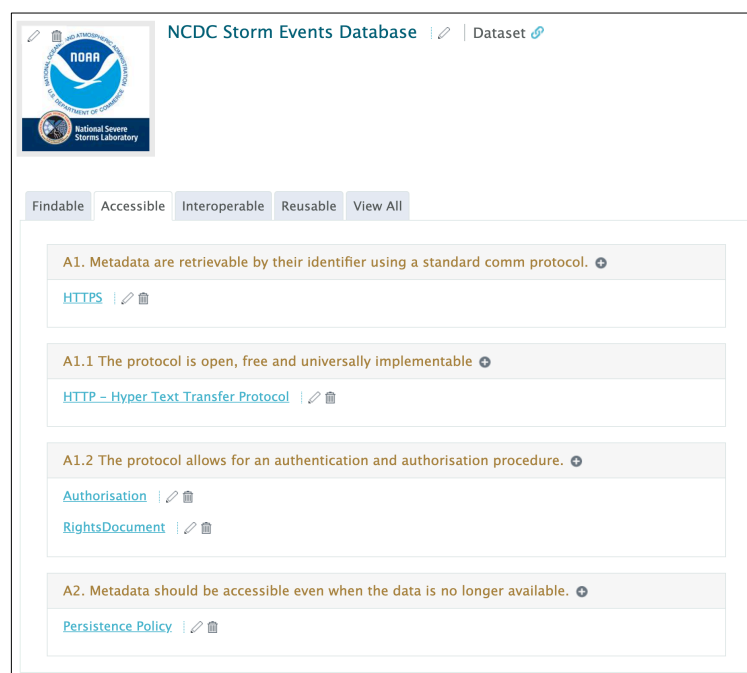


Fig. 4.5: Screenshot of the Accessible entry interface of FAIRtool.org

provides documents indicating evidence that standard communication protocol allows for an authentication and authorization procedure. The last field of the Accessible tab is A2. The metadata element entered in this field is the hyperlink to persistence repository like data.gov to ensure the availability of the metadata even if the data are no longer available.

Third, as shown in Fig. 4.6, the Interoperable tab allows for recording the metadata that concerns the interoperability with other external digital entities. The first principle in this category is I1, which specifies the type of digital resource format. The format-type that can be entered in this field should be a formal, accessible, shared and broadly applicable language. The Resource Description Framework (RDF), which is the specification of the W3C on how to represent information on the internet in a machine-accessible format, is the most commonly agreed alternative to comply with this concept at present. Thus, the only format-type that satisfies these criteria is the RDF format [31]. I2 FAIR data principle indicates that the metadata should use vocabularies that adhere to FAIR data principles. The entry of the I2 FAIR data principle allows for recording the information of the used vocabulary include the URL, which mainly will be FAIR vocabulary. The I3 FAIR data principle enables the user to enter an identifier such as CrossRef and DOI of qualified references to other metadata that relate to this metadata.

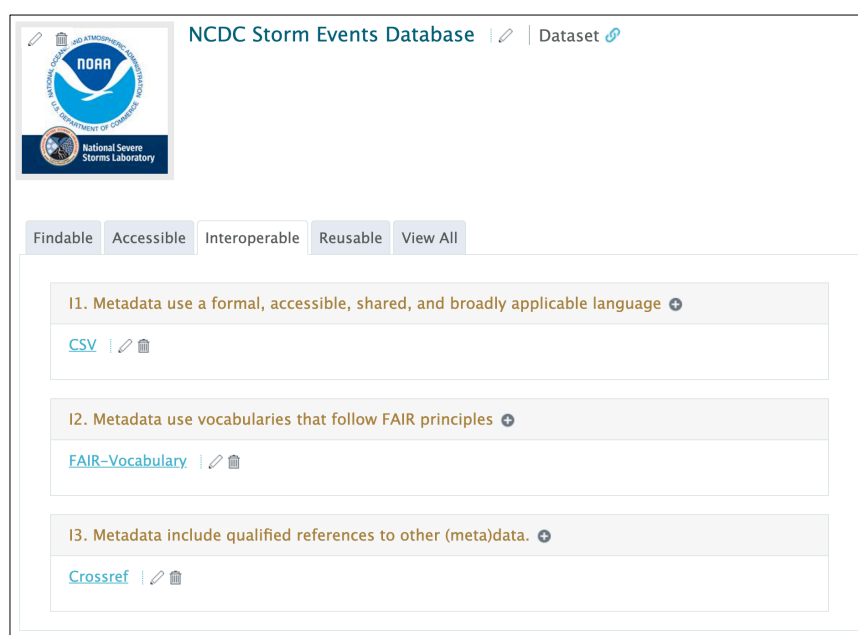


Fig. 4.6: Screenshot of the Interoperable entry interface of FAIRtool.org

Fourth, Fig. 4.7 depicts a screenshot of the reusable tab of the system. It contains four sections, R1, R1.1, R1.2, and R1.3. The R1 is responsible for recording metadata that is richly described with a plurality of accurate relevant attributes. In some sense, this principle is identical to the F2 principle; therefore, we can refer to F2 rich metadata. R1.1 principle is concerned with the type of license for reusing this digital resource. For example, the user can enter the license type that the data owner decided to give to this digital resource, such as Creative Commons (CC), MIT, or Apache. In the field of R1.2, there is a place to record the provenance of the digital resource; properties like `wasGeneratedBy`, `wasAssociatedWith`, and `wasDrivenFrom` create the relationship between classes such as Entity, Agent, and Activity to establish provenance. Lastly is the R1.3 principle, which allows for recording the domain-relevant community standards. In our case, the Earth science metadata community standard is ISO 19115 [33].

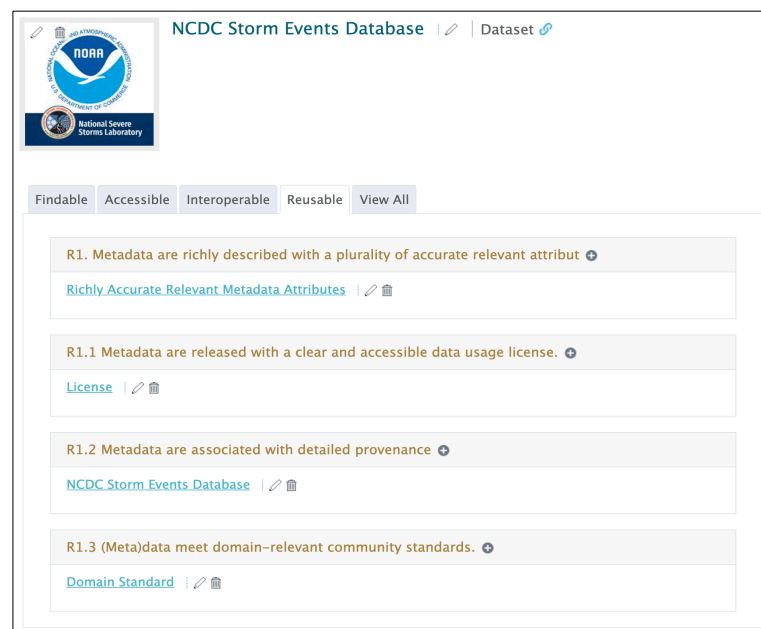


Fig. 4.7: Screenshot of the Reusable entry interface of FAIRtool.org

Finally, FAIRtool.org enables users to download the metadata entered the system as an RDF format. Also, it allows other software agents to link and pull metadata in RDF format via a SPARQL endpoint. This functionality is applicable for the whole items of a digital resource as one RDF file or partial items (for example, RDF for only F1-the identifier) through the chain symbol that appears at the top of every screen. We will illustrate this functionality in more detail in section 4.3.

4.2.3 FAIRTOOL.ORG TRIPLE STORE SETUP

The configuration of FAIRtool.org uses the Apache Jena SDB triple-store to hold its triples contents [23]. SDB is a Java Loader that takes incoming triples and breaks them down into components ready for the database. Furthermore, SDB requires a relational database, and the default configuration specifies MySQL as the storage for SDB. However, we can configure FAIRtool.org to use a different database as the basis for Apache Jena SDB; we can also configure FAIRtool.org to use a different triple store such as AllegroGraph, GraphDB, and SparkleDB. Further, FAIRtool.org can work with databases like Oracle, Microsoft SQL Server, DB2, PostgreSQL, MySQL, Apache Derby, H2, HSQLDB, or with other types of triple stores [69]. In FAIRtool.org, we used Apache Jena SDB triple-store to hold the triple network of our system, and we connected the Apache Jena SDB triple-store to MySQL RDBMS through a secure embedded username and password.

4.3 DEMONSTRATION OF THE NCDC USE CASE

In this section, we explore the utility of FAIRtool.org, which is supported by Semantic Web technologies (as explained in Section 4.2), with the following specific use case:

4.3.1 NCDC STORM EVENTS DATABASE USE CASE

Then apply all the 15 FAIR data principles to it. This metadata must be represented in a machine-readable RDF format along with good quality metadata, which will facilitate the easy retrieval of the data through SPARQL queries. Also, it must be done in an easy and efficient way. Besides, the metadata must adhere to the new standard “FAIR data principles” which ensure that this dataset is findable, accessible, interoperable, and then reusable. Eventually, assure the sustainability of this digital resource on the Internet. We gathered the metadata from various resources, which are fed into FAIRtool.org. Table 4.1 shows the inputs as metadata collection for the FAIR data principles inputs.

Table 4.1: The NCDC Storm Events Database use case metadata inputs for FAIRtool.org

FAIR data principles	Metadata
F1: Identifier:	https://data.globalchange.gov/dataset/noaa-ncdc-c00510
F2: Rich MetaData: list of rich metadata:	
1. Dataset Name:	“NCDC Storm Events Database”
2. AlternateName:	“National Weather Service Storm Data”
3. Creator:	“National Weather Service (NWS)”
4. Date Modified:	” Dec 18, 2013“
5. Date Published:	” January 1, 1996“
6. Keywords:	“ATMOSPHERE, ATMOSPHERIC PHENOMENA, CYCLONES, DROUGHT, FOG, FREEZE”
7. Spatial Coverage:	USA, Geo Shape Spatial Extent: West Bounding Longitude: 172.0

	East Bounding Longitude: -65.0 North Bounding Latitude: 18.0 South Bounding Latitude: 72.0
8. Temporal Coverage:	“January 1, 1950/December 18, 2013”
9. Variable Measured:	“Atmospheric Phenomena”
10. Version:	“1.0”
11. Description:	“Storm Data is provided by the National Weather Service (NWS) and contains statistics”
F3: The metadata should include the identifier of the dataset it describes:	https://data.globalchange.gov/dataset/noaa-ncdc-c00510
F4: Searchable resource That the digital resource registered in:	https://catalog.data.gov/dataset/ncdc-storm-events-database
A1: The standardized communication protocols used:	HTTP
A1.1: Proof of the protocol is open, free, and universally implemented:	“ https://www.w3.org/Protocols/HTTP ”
A1.2: Authentication and authorization procedure of the protocol used:	“ https://tools.ietf.org/html/rfc7235 ”
A2: Accessibility of the metadata even if the data is no longer available:	https://catalog.data.gov/harvest/object/1d35197b-76f9-47a0-aa5a-14beb34f460a/html
I1: The formal, accessible, shared, and broadly applicable language used:	“CSV”
I2: Following FAIR principles vocabulary:	“FAIR-O”
I3: Qualified references to other metadata:	https://www.ncdc.noaa.gov/stormevents/
R1: Accurate relevant attributes:	Same as F2
R1.1: A clear and accessible data usage license:	https://www.usa.gov/government-works
R1.2: Detailed provenance:	Prov-O – See Fig 4.6 for details
R1.3: Domain-relevant community standards:	ISO 19115-2:2019 Geographic Information – Metadata Standard url: https://www.iso.org/standard/67039.html

4.3.2 RESULT OF THE USE CASE

The result of entering the listed metadata in Table 4.1 into FAIRtool.org can be displayed to the public via the FAIRtool.org website and can also be downloaded in RDF format. Listing 4.1 to Listing 4.15 shows the RDF codes of the FAIR data principles. RDF codes are in Turtle serialization, which applies to all principles. Listing 4.1 depicts the F1 principle's RDF showing the identifier's URL and value of the "NCDC Storm Events Database" and the rating value of 25 points.

```

<http://fairtool.org/individual/n2082>
a owl:Thing, <http://fairtool.org/Dataset> ;
rdfs:label "NCDC Storm Events Database"^^rdf:langString ;
ns0:identifier ns0:n1050 .

ns0:n1050
a <https://www.w3.org/TR/vocab-dqv/#dqv:Metric>, <http://fairtool.org/Identifier>, owl:Thing,
<http://fairtool.org/PropertyValue> ;
ns1:rating ns0:n4349 ;
rdfs:label "DOI ID" ;
ns2:mostSpecificType ns1:Identifier ;
dqv:value 25 ;
ns1:url "<a href=\"https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ncdc:C00510\"</a>" ;
ns1:value "<em></em>gov.noaa.ncdc:C00510<br /><br /><br />" .

ns0:n4349
a ns1:Rating, owl:Thing ;
rdfs:label "25 Points Rating Value"^^rdf:langString .

```

Listing 4.1: F1 principle's RDF: The Identifier

```

<http://fairtool.org/individual/n3239>
  a ns0:ItemList, <https://www.w3.org/TR/vocab-dqv/#dqv:Metric>, owl:Thing ;
  ns0:alternateName "National Weather Service Storm Data." ;
  ns0:description "Storm Data is provided by the National Weather Service (NWS) and contains statistics." ;
  ns0:temporalCoverage "January 1, 1950/December 18, 2013" ;
  ns0:dateModified "2016-05-10"^^xsd:date ;
  ns0:keywords "ATMOSPHERE, ATMOSPHERIC PHENOMENA, CYCLONES, DROUGHT, FOG, FREEZE ;
  ns0:datePublished "2013-01-01"^^xsd:date ;
  ns0:rating <http://fairtool.org/individual/n5457> ;
  ns0:creator <http://fairtool.org/individual/n306> ;
  ns0:variableMeasured "Atmospheric Phenomena." ;
  dqv:value 25 ;
  rdfs:label "Rich Metadata" ;
  ns0:spatialCoverage <http://fairtool.org/individual/n3260>, <http://fairtool.org/individual/n7782> ;
  ns1:mostSpecificType ns0:ItemList ;
  ns0:version "1.0" ;
  ns2:name "NCDC Storm Events Database." .

```

Listing 4.2: F2 principle's RDF: Rich Metadata

Listing 4.2 shows F2 the rich metadata in RDF format with a rating value of 25 points.

Listing 4.3 shows the value and URL of the F3 principle in RDF with a rating value of 25 points.

```

<http://fairtool.org/individual/n1570>
  a ns0:Identifier, owl:Thing, ns0:PropertyValue, <https://www.w3.org/TR/vocab-dqv/#dqv:Metric> ;
  ns0:value "gov.noaa.ncdc:C00510" ;
  ns0:url "<a href=\"https://catalog.data.gov/dataset/ncdc-storm-events-database/C00510\">" ;
  ns1:mostSpecificType ns0:Identifier ;
  rdfs:label "Data Identifier" ;
  ns0:rating <http://fairtool.org/individual/n4349> ;
  dqv:value 25 .

<http://fairtool.org/individual/n4349>
  a ns0:Rating, owl:Thing ;
  rdfs:label "25 Points Rating Value"^^rdf:langString .

```

Listing 4.3: F3 principle's RDF: Data Identifier

Listing 4.4 shows the value and URL of the digital resource is registered in the searchable data catalog for F4 principle with a rating value of 25 points.

```
<http://fairtool.org/individual/n4100>
a <https://www.w3.org/TR/vocab-dqv/#dqv:Metric>, owl:Thing, <http://fairtool.org/DataCatalog> ;
rdfs:label "data.gov" ;
ns0:dataCatalogUrl "<a href=\"https://catalog.data.gov/dataset/dd13b9be-9b00-4639-8ba5-e2035bf4f514\"a>" ;
ns0:dataset <http://fairtool.org/individual/n2082> ;
ns0:rating <http://fairtool.org/individual/n7188> ;
ns1:mostSpecificType ns0:DataCatalog ;
dqv:value 25 .

<http://fairtool.org/individual/n2082>
a ns0:Dataset, owl:Thing ;
rdfs:label "NCDc Storm Events Database"^^rdf:langString ;
ns0:includedInDataCatalog <http://fairtool.org/individual/n4100> .

<http://fairtool.org/individual/n7188>
a ns0:Rating, owl:Thing ;
rdfs:label "25 Points Rating Value"^^rdf:langString .
```

Listing 4.4: F4 principle's RDF: Registered in Searchable Data Catalog

```
<http://fairtool.org/individual/n1492>
a <https://www.w3.org/TR/vocab-dqv/#dqv:Metric>, ns0:Standard, owl:Thing ;
rdfs:label "HTTPS" ;
ns0:protocol "<a href=\"https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ncdc:C00510\" title=\"Storm
Events Database\">https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ncdc:C00510</a>" ;
ns0:rating <http://fairtool.org/individual/n5457> ;
ns1:mostSpecificType ns0:Standard ;
dqv:value 25 .

<http://fairtool.org/individual/n5457>
a ns0:Rating, owl:Thing ;
rdfs:label "25 Points Rating Value"^^rdf:langString .
```

Listing 4.5: A1 principle's RDF: Communication Protocol Standard

Listing 4.5 shows the A1 principle's RDF code that describes the protocol used and a rating value of 25 points.

```

<http://fairtool.org/individual/n4314>
  a <https://www.w3.org/TR/vocab-dqv/#dqv:Metric>, ns0:Standard, owl:Thing ;
  rdfs:label "HTTP - Hyper Text Transfer Protocol" ;
  ns0:protocol "<a href=\"https://www.w3.org/Protocols/HTTP\">https://www.w3.org/Protocols/HTTP</a>" ;
  ns0:rating <http://fairtool.org/individual/n7188> ;
  ns1:mostSpecificType ns0:Standard ;
  dqv:value 25 .

<http://fairtool.org/individual/n7188>
  a ns0:Rating, owl:Thing ;
  rdfs:label "25 Points Rating Value"^^rdf:langString .

```

Listing 4.6: A1.1 principle's RDF: Communication Protocol Characteristics

Listing 4.6 shows the RDF code of A1.1 that describes the characteristics of the protocol used, and a rating value of 25 points.

```

<http://fairtool.org/individual/n909>
  a <https://www.w3.org/TR/vocab-dqv/#dqv:Metric>, ns0:RightsStatement, owl:Thing ;
  rdfs:label "Authorization" ;
  ns0:rating <http://fairtool.org/individual/n7188> ;
  ns0:rightsStatements "<a href=\"https://tools.ietf.org/html/rfc7235\"></a>" ;
  ns1:mostSpecificType ns0:RightsStatement ;
  dqv:value 25 .

<http://fairtool.org/individual/n7188>
  a ns0:Rating, owl:Thing ;
  rdfs:label "25 Points Rating Value"^^rdf:langString .

```

Listing 4.7: A1.2 principle's RDF: Communication Protocol- Authorization-Authentication Rules

Listing 4.7 shows the RDF code of the A1.2 principle that illustrate the authorization and authentication rules of the protocol used, and a rating value of 25 points.

```

<http://fairtool.org/individual/n1297>
  a ns0:Identifier, ns0:PropertyValue, <https://www.w3.org/TR/vocab-dqv/#dqv:Metric>, owl:Thing ;
  rdfs:label "Persistence Policy" ;
  ns0:url "<a href=\"https://catalog.data.gov/harvest/object/1d35197b-76f9-47a0-aa5a-14beb34f460a/html\"
title=\"Storm Events Database Metadata\"></a>" ;
  ns0:value "<p>This is to insure the availability of the metadata in a persistence repository like
data.gov.</p>" ;
  ns0:rating <http://fairtool.org/individual/n8183> ;
  ns1:mostSpecificType ns0:Identifier ;
  dqv:value 25 .

<http://fairtool.org/individual/n8183>
  a ns0:Rating, owl:Thing ;
  rdfs:label "25 Points Rating Value"^^rdf:langString .

```

Listing 4.8: A2 principle's RDF: Persistence Policy

Listing 4.8 shows the RDF code of the A2 principle that illustrate the availability of the metadata in a persistence repository like data.gov and a rating value of 25 points.

```

<http://fairtool.org/individual/n639>
  a <https://www.w3.org/TR/vocab-dqv/#dqv:Metric>, ns0:DataDownload, owl:Thing ;
  rdfs:label "CSV" ;
  ns0:contentType "<a href=\"https://www1.ncdc.noaa.gov/pub/data/swdi/stormevents/csvfiles/\" title=\"Storm
Events Database\">https://www1.ncdc.noaa.gov/pub/data/swdi/stormevents/csvfiles/</a><br /><br />" ;
  ns0:encodingFormat "CSV" ;
  ns0:rating <http://fairtool.org/individual/n4349> ;
  ns1:mostSpecificType ns0:DataDownload ;
  dqv:value 25 .

<http://fairtool.org/individual/n4349>
  a ns0:Rating, owl:Thing ;
  rdfs:label "25 Points Rating Value"^^rdf:langString .

```

Listing 4.9: I1 principle's RDF: Knowledge Representation Format

Listing 4.9 shows the RDF code of the I1 principle that illustrates the type of the knowledge representation format of the data which is CSV in this use case and the rating value is 25 points.


```

<http://fairtool.org/individual/n1677>
  a <https://www.w3.org/TR/vocab-dqv/#dqv:Metric>, owl:Thing, ns0:PropertyValue, ns0:Identifier ;
  ns0:rating <http://fairtool.org/individual/n2696> ;
  rdfs:label "FAIR-Vocabulary" ;
  ns0:value "FAIR-0" ;
  ns0:url "<a href='\"http://fairtool.org/\"' title='\"fairtool.org\">https://fairtool.org/</a>" ;
  ns1:mostSpecificType ns0:Identifier ;
  dqv:value 25 .
<http://fairtool.org/individual/n2696>
  a ns0:Rating, owl:Thing ;
  rdfs:label "25 Points Rating Value"^^rdf:langString .

```

Listing 4.10: I2 principle's RDF: Vocabulary Used

Listing 4.10 shows the value and URL of the RDF code of I2 which represent the vocabulary used in describing this dataset associated with a rating value of 25 points.

Listing 4.11 shows an RDF code of the I3 principle that illustrates the metadata cross-reference associated with a rating value of 25 points.

```

<http://fairtool.org/individual/n739>
  a owl:Thing, ns0:PropertyValue, ns0:Identifier, <https://www.w3.org/TR/vocab-dqv/#dqv:Metric> ;
  rdfs:label "Crossref" ;
  ns0:value "<p><a
href='\"https://www.ncdc.noaa.gov/stormevents/\">https://www.ncdc.noaa.gov/stormevents/</a></p>" ;
  ns1:mostSpecificType ns0:Identifier ;
  dqv:value 25 ;
  ns0:rating <http://fairtool.org/individual/n8183> ;
  ns0:url "<p><a href='\"https://www.geoplatform.gov/resources/datasets/a8e78f8b33a2295755e05b95c0e694d6/\"
title='\"Storm Events Database CrossRef\
\">https://www.geoplatform.gov/resources/datasets/a8e78f8b33a2295755e05b95c0e694d6/</a></p>" .

<http://fairtool.org/individual/n8183>
  a ns0:Rating, owl:Thing ;
  rdfs:label "25 Points Rating Value"^^rdf:langString .

```

Listing 4.11: I3 principle's RDF: Metadata Cross-Reference

Listing 4.12 shows the RDF code of R1 that indicates sameAs relation with F2 which is the rich metadata of the dataset associated with a rating value of 25 points.

```
<http://fairtool.org/individual/n702>
  a <https://www.w3.org/TR/vocab-dqv/#dqv:Metric>, owl:Thing, ns0:QualitativeValue ;
  rdfs:label "Richly Accurate Relevant Metadata Attributes" ;
  ns0:rating <http://fairtool.org/individual/n2696> ;
  ns0:sameAs <http://fairtool.org/individual/n3239> ;
  ns1:mostSpecificType ns0:QualitativeValue ;
  dqv:value 25 .

<http://fairtool.org/individual/n2696>
  a ns0:Rating, owl:Thing ;
  rdfs:label "25 Points Rating Value"^^rdf:langString .

<http://fairtool.org/individual/n3239>
  a <https://www.w3.org/TR/vocab-dqv/#dqv:Metric>, ns0:ItemList, owl:Thing ;
  rdfs:label "Rich Metadata" .
```

Listing 4.12: R1 principle's RDF: Richly Accurate Relevant Metadata Attributes

Listing 4.13 shows the RDF code of the R1.1 principle that illustrates the data usage license type associated with a rating value of 25 points.

```
<http://fairtool.org/individual/n2082>
  a owl:Thing, <http://fairtool.org/Dataset> ;
  rdfs:label "NCDC Storm Events Database"^^rdf:langString ;
  ns0:license <http://fairtool.org/individual/n3697> .

<http://fairtool.org/individual/n3697>
  a ns0:Identifier, <https://www.w3.org/TR/vocab-dqv/#dqv:Metric>, owl:Thing, ns0:PropertyValue ;
  ns0:url "<a href=\"http://www.usa.gov/publicdomain/label/1.0/\">U.S. Government Work</a>" ;
  ns0:rating <http://fairtool.org/individual/n7188> ;
  dqv:value 25 ;
  ns1:mostSpecificType ns0:Identifier ;
  ns0:value "Copyright applies to U.S. government works" ;
  rdfs:label "License" .

<http://fairtool.org/individual/n7188>
  a ns0:Rating, owl:Thing ;
  rdfs:label "25 Points Rating Value"^^rdf:langString .
```

Listing 4.13: R1.1 principle's RDF: Data Usage License

Listing 4.14 shows the RDF code of the R1.2 principle that illustrates the provenance of the dataset associated with a rating of 25 points.

```
<http://fairtool.org/individual/n8133>
  a prov:Activity, owl:Thing ;
  rdfs:label "Storm Events Data Collection " ;
  ns0:mostSpecificType prov:Activity ;
  prov:endedAtTime "2016-05-10T17:01:01"^^xsd:dateTime ;
  prov:startedAtTime "2013-01-01T09:01:01"^^xsd:dateTime ;
  prov:used <http://fairtool.org/individual/n4501> ;
  prov:wasAssociatedWith <http://fairtool.org/individual/n4544> ;
  prov:wasInformedBy <http://fairtool.org/individual/n8133> .

<http://fairtool.org/individual/n4501>
  a prov:Entity, owl:Thing ;
  rdfs:label "NCDC Storm Events Database" ;
  prov:wasGeneratedBy <http://fairtool.org/individual/n8133> .

<http://fairtool.org/individual/n4544>
  a owl:Thing, prov:Agent ;
  rdfs:label "National Centers for Environmental Information" .
```

Listing 4.14: R1.2 principle's RDF: Digital Resource Provenance

Listing 4.15 shows an RDF code of the R1.3 principle that illustrates the domain-relevant community standard used associated with a rating value of 25 points.

```
<http://fairtool.org/individual/n2082>
  a owl:Thing, <http://fairtool.org/Dataset> ;
  rdfs:label "NCDC Storm Events Database"^^rdf:langString ;
  ns0:conformsTo <http://fairtool.org/individual/n5804> .

<http://fairtool.org/individual/n5804>
  a ns0:DomainStandard, <https://www.w3.org/TR/vocab-dqv/#dqv:Metric>, owl:Thing ;
  rdfs:label "Domain Standard" ;
  ns0:domainStandard "<a href=\"https://earthdata.nasa.gov/esdis/eso/standards-and-references/iso-19115</a>"
  ;
  ns0:rating <http://fairtool.org/individual/n4349> ;
  ns1:mostSpecificType ns0:DomainStandard ;
  dqv:value 25 .

<http://fairtool.org/individual/n4349>
  a ns0:Rating, owl:Thing ;
  rdfs:label "25 Points Rating Value"^^rdf:langString .
```

Listing 4.15: R1.3 principle's RDF: Domain-Relevant Community Standard

4.4 DISCUSSION

The objective of this research is to improve the stewardship of Earth science digital resources through the implementation of FAIR data principles. To meet this objective, we address the following research question: 1) How can we technically approach FAIR data principles? To answer this question, section 4.2 illustrates the detailed process of designing the FAIR ontology, designing the semantic web application's interfaces, and setting up the triple store database. Concisely, the only technology that supports the web of data currently is the semantic web (SW) technology; SW must include ontology and linked to a triple store to hold the application's data. Therefore, we use SW and build FAIR ontology to support the creation of FAIRtool.org semantic web application. Besides, we designed interfaces and the database for FAIRtool.org. Then, as proof of concept and to show the feasibility of the FAIRtool.org, we applied the 15 FAIR data principles on the dataset "NCDC Storm Events Dataset" from the Earth science domain; the findings is an online semantic web application utilized to enable FAIR data principles and produces an RDF file contains the complete set of FAIR data principles, as shown in section 4.3. Therefore, we argue that utilizing FAIRtool.org supported by FAIR ontology in Earth science will enable their digital resources to be Findable, Accessible, Interoperable, and Reusable (i.e., FAIR).

4.5 CONCLUSIONS

We have introduced the emerging technology of SW based on FAIR ontology and showed how FAIRtool.org used to implement FAIR data principles. As a result, FAIRtool.org

become FAIR data principles semantic web application. Therefore, FAIRtool.org can become the base of international standards for annotating digital resources for a variety of scientific domains; we anticipate this conclusion because FAIRtool.org has two main characteristics the ease of use and the FAIR data principles applicability. Besides, FAIRtool.org can contribute to improving data stewardship of other scientific fields as well, such as biomedical and natural resources; ontology engineers can perform little modifications on the F2 principle, the rich metadata in FAIR ontology, to fit the structured terminologies of other scientific domains. Finally, FAIRtool.org will have a magnificent effect on resolving data stewardship problems of Earth science as well as other science domains. In the following chapter, we will describe an intuitive method for evaluating the FAIRness level of a digital resource like dataset by utilizing Fuzzy logic.

Chapter 5: Utilizing Fuzzy Logic for Evaluating “FAIRness” of A Digital Resource

5.1 INTRODUCTION

In chapter 4, we accomplished the technical implementation of the FAIR data principles. We adopted the emerging semantic web technology based on FAIR ontology and demonstrated how FAIRtool.org used to solve main issues in the management and stewardship of Earth science data. Then, we utilized a use case from the Earth science domain to show the practicality of deploying the FAIR data principles. Also, during the deployment process, we collected rating points for each present FAIR data principle to measure the FAIRness level of this use case dataset. In this chapter, we propose an intuitive method for FAIRness level evaluation by utilizing Fuzzy logic based on the collected rating points. FAIRness denotes the level a digital resource is “Findable, Accessible, Interoperable, and Reusable.” Besides, FAIRness is the level of FAIR maturity of the digital resource [64]. Therefore, it is significant to emphasize that the FAIRness of a digital resource should be measured through a controlled framework to reflect the exact level of FAIR maturity. Since the four FAIR pillars have both quantitative and qualitative elements, Fuzzy logic can evaluate their FAIRness level. There are two types of logic Boolean logic (two-valued logic) and Fuzzy logic (multi-valued logic) [65]. Boolean logic membership degree is either truth value 1, which denotes the actual true value, or 0, which represents the real false value. In contrast, Fuzzy logic is a range of membership degrees between 1 and 0; in Fuzzy logic, there is a present intermediate value, which is partly true and

partly false; for example, in the real world, the digital resource can be Findable to an imprecise degree such as 0.6 (e.g., 60% true Findable); that is the degree of membership. In the following sections, we will illustrate how we utilized Fuzzy logic to solve the imprecision issue of the FAIRness level of a digital resource through the introduction of the Fuzzy FAIRness Assessments Framework (FFAF). To demonstrate the usage of FFAF, we exemplify one dataset from the Earth science domain; the dataset is “Data for Building an Open Science Framework to Model Soil Organic Carbon, [71]” obtained from the Northwest Knowledge Network (NKN) [72]. Table 5.1 shows the metadata of the dataset “Data for Building an Open Science Framework to Model Soil Organic Carbon” that entered in FAIRtool.org. The rating points for the “Data for Building an Open Science Framework to Model Soil Organic Carbon” dataset is acquired from FAIRtool.org and used as an input to the fuzzy logic method. The statistical programming language R used to implement the Fuzzy logic method [73].

5.2 FAIRNESS LEVEL EVALUATION WITH AN NKN DATASET USE CASE

Following the structure of the Fuzzy logic system described in chapter 2 section 2.4, the design for the Fuzzy FAIRness Assessments Framework (FFAF) consists of four steps: 1) Modeling FFAF inputs; 2) Fuzzifying inputs; 3) Inferencing fuzzified inputs; 4) Aggregating and defuzzifying the fuzzy outputs. The design of FFAF is four-inputs (i.e., Findable, Accessible, Interoperable, and Reusable) and one-output (i.e., FAIRness). In the following sections, we will demonstrate the performance of the FFAF system to evaluate the FAIRness level of the “Data for Building an Open Science Framework to Model Soil Organic Carbon” dataset [71].

Table 5.1: The NKN dataset use case metadata inputs for FAIRtool.org

FAIR data principles	Metadata
F1: Identifier:	https://doi.org/10.7923/g4xp72zb
F2: Rich MetaData: list of rich metadata:	
1. Dataset Name:	“Data for Building an Open Science Framework to Model Soil Organic Carbon (data)”
2. AlternateName:	
3. Creator:	Edward Flathers, Paul Gessler
4. Date Modified:	2019-03-19
5. Date Published:	2018-03-07
6. Keywords:	Carbon, Soil, Agriculture, REACCH
7. Spatial Coverage:	POLYGON ((-121.83837890625 43.948536571678, -121.83837890625 49.055150393383, -115.51025390625 49.055150393383, -115.51025390625 43.948536571678))
8. Temporal Coverage:	Monday, January 1, 1923 - 00:00 to Saturday, December 31, 2016 - 00:00
9. Variable Measured:	Climatology, Meteorology, Atmosphere
10. Version:	
11. Description:	Framework to model and map soil organic carbon (SOC) in the cereal grains production region of the northwestern United States. Primarily associated with soil organic matter, SOC relates to many soil properties that influence resiliency and soil health for agriculture.
F3: The metadata should include the identifier of the dataset it describes:	Missing
F4: Searchable resource That the digital resource registered in:	Northwest Knowledge Network (NKN)
A1: The standardized communication protocols used:	HTTP
A1.1: Proof of the protocol is open, free, and universally implemented:	Missing
A1.2: Authentication and authorization procedure of the protocol used:	Missing
A2: Accessibility of the metadata even if the data is no longer available:	https://data.nkn.uidaho.edu/dataset/data-building-open-science-framework-model-soil-organic-carbon
I1: The formal, accessible, shared, and broadly applicable language used:	RDF, Json
I2: Following FAIR principles vocabulary:	Missing
I3: Qualified references to other metadata:	Missing

R1: Accurate relevant attributes:	Same as F2: Accurate rich metadata
R1.1: A clear and accessible data usage license:	Creative Commons Attribution Non-Commercial Share-Alike
R1.2: Detailed provenance:	Missing
R1.3: Domain-relevant community standards:	Missing

5.2.1 MODELING FFAF INPUTS

The input of FFAF would be coming from the rating points of the FAIRtool.org system. This tool enables a user to describe a digital resource according to the FAIR data principles. It has four tabs, each of which will hold data for one of the four FAIR data principles pillars (Findable, Accessible, Interoperable, and Reusable). Furthermore, each principle has its own designated data entry point. Once the data is entered into this entry point, the system will assign a rating value for that entry as points. For example, Findable consists of F1, F2, F3, and F4; when the value of F1 (i.e., identifier) is entered, the system will assign 25 points and store as rating points; this is also the case with F2, F3, F4. Therefore, Findable can be either 25, 50, 75, or 100. The same concept applies to Accessible, Interoperable, and Reusable. Accordingly, the metadata of the dataset “Data for Building an Open Science Framework to Model Soil Organic Carbon” inserted into FAIRtool.org, it produced the following rating points and are used as inputs for FFAF:

Findable = 75 points because F1, F2, and F4 are filled with metadata.

Accessible = 50 points because A1 and A2 are filled with metadata.

Interoperable = 25 points because only I1 is filled with metadata.

Reusable = 50 points because R1 and R1.1 are filled with metadata.

Therefore, the four crisp inputs of the “Data for Building an Open Science Framework to Model Soil Organic Carbon” dataset for the FFAF system that can be loaded in the “R program” are (75, 50, 25, 50). In the next section 5.2.2, we will explain the fuzzification step.

5.2.2 FFAF FUZZIFICATION

Fuzzification is the process by which the crisp input value is converted to the fuzzy value by projecting the crisp input x to the fuzzy set A . This is accomplished by using Triangular and Trapezoidal fuzzifiers. We construct Table 5.2 to define the input variables and their value range. Furthermore, we constructed Table 5.3 to define the output variables and their value range. For example, to map a score of 75 to the triangular fuzzifiers, the fuzzifier task is to transform clear (crisp) external input data into suitable semantic fuzzy data. The fuzzy theory uses the numerical region between $[0, 1]$ of the membership function to reflect the fuzzy set in the numerical region.

Table 5.2: FFAF Input Variables

Input Variables			
	<i>Linguistic variables syntax</i>	<i>Semantics</i>	<i>Range</i>
x1	Low Findable	LF	0-50
	Medium Findable	MF	20-80
	High Findable	HF	50-100
x2	Low Accessible	LA	0-50
	Medium Accessible	MA	20-80
	High Accessible	HA	50-100
x3	Low Interoperable	LI	0-50
	Medium Interoperable	MI	20-80
	High Interoperable	HI	50-100
x4	Low Reusable	LR	0-50
	Medium Reusable	MR	20-80
	High Reusable	HR	50-100

Table 5.3: FFAF Output Variables

Output Variables			
<i>Linguistic variables syntax</i>		<i>Semantics</i>	<i>Range</i>
y1	Poor FAIRness	PF	0-50
	Ok FAIRness	OF	20-80
	Good FAIRness	GF	50-100

Findable: As shown in Fig. 5.1, Findable crisp input of 75 maps to 0.0 Low Findable (LF), 0.17 Med Findable (MF), and 0.83 High Findable (HF). Thus, the Findable fuzzy set becomes: $\mu_{A_{\text{Findable}}} = \{0.0, 0.17, 0.83\}$.

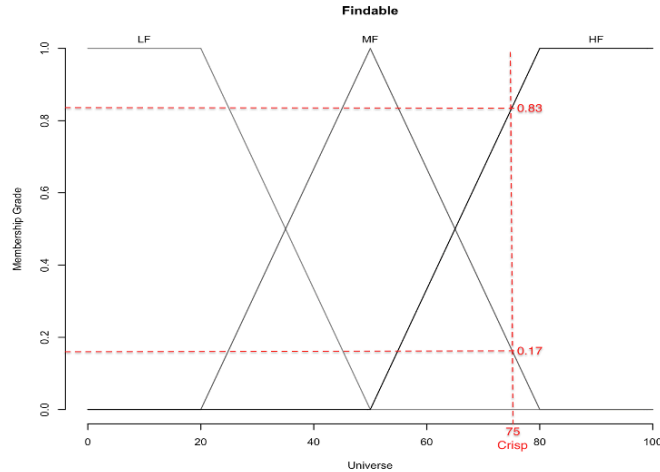


Fig. 5.1: Findable membership function

Accessible: Fig. 5.2 shows that the Accessible crisp input of 50 maps to 0.0 Low Accessible (LA), 1.0 Med Accessible (MA), and 0.0 High Accessible (HA). Thus, the Accessible fuzzy set becomes: $\mu_{A_{\text{Accessible}}} = \{0.0, 1.0, 0.0\}$.

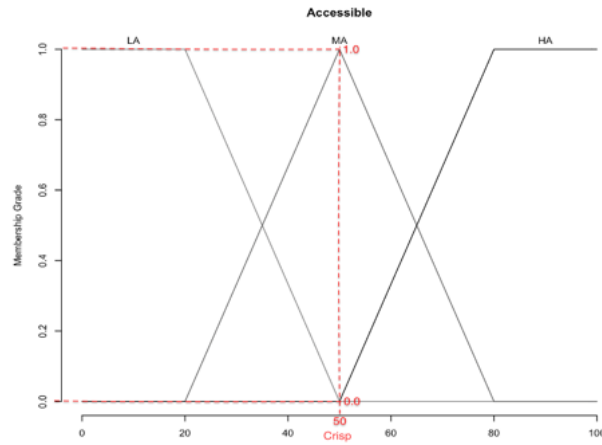


Fig. 5.2: Accessible membership function

Interoperable: Fig. 5.3 shows that the Interoperable crisp input of 25 maps to 0.83 Low Interoperable (LI), 0.17 Med Interoperable (MI), and 0.0 High Interoperable (HI). Thus, the Interoperable fuzzy set becomes: $\mu_{A_{Interoperable}} = \{0.83, 0.17, 0.0\}$.

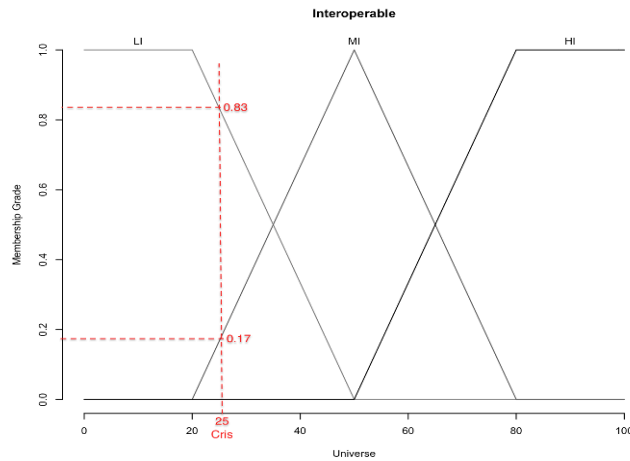


Fig. 5.3 Interoperable membership function

Reusable: Fig. 5.4 illustrates that the Reusable crisp input of 50 maps to 0.0 Low Reusable (LR), 1.0 Med Reusable (MR), and 0.0 High Reusable (HR). Thus, the Reusable fuzzy set becomes: $\mu_{A_{\text{Reusable}}} = \{0.0, 1.0, 0.0\}$.

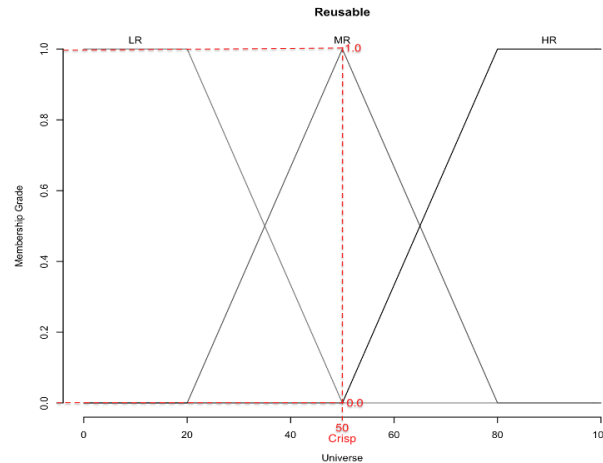


Fig. 5.4 Reusable membership function

5.2.3 FFAF INFERENCE SYSTEM

Once we have the fuzzy sets for all four FAIR pillars, we can build the fuzzy inference system and apply it to these fuzzy sets. To know how many rules to build, we need to calculate the number of variable terms (three triangular and trapezoidal membership functions) {Low, Med, High} to the power of four {Findable, Accessible, Interoperable, Reusable}, which means 3^4 leads to 81 rules. In a fuzzy model, there is no limit to the number of variable terms, but increasing the number of variable terms raises the number of rules exponentially and thus increases the system complexity, e.g., 4^4 leads to 256 rules, 5^4 leads to 625 rules, etc. However, increasing the number of variable terms gives a more precise result. Occasionally, some of the rules fire some do not, depending on the combination of the crisp input's value of the variables.

For example, if all the crisp inputs are on the LR range, the rules of the HR range will not fire, i.e., the rules cannot be applied.

The FAIRness degree can be {PF, OF, GF}. Thus, by following the IF-THEN rule-based expression formula (1) and the Mamdani-type rules evaluation [55], we get 81 rules

$$\begin{aligned}
 &1. \text{ IF } x1=LF \text{ AND } x2=LA \text{ AND } x3=LI \text{ AND } x4=LR, \text{ THEN } y1=PF \\
 &\quad \vdots \\
 &\text{The complete list of the 81 rules can be found at Appendix A:} \\
 &\quad \vdots \\
 &81. \text{ IF } x1=HF \text{ AND } x2=HA \text{ AND } x3=HI \text{ AND } x4=HR, \text{ THEN } y1=GF
 \end{aligned}$$

These 81 rules construct the fuzzy associative memory (FAM) for our FFAF Inference System.

5.2.4 AGGREGATION AND DEFUZZIFICATION OF THE OUTPUT OF FFAF

Since all the relations between our variables are ANDs logical operators, the intersection formula (2) is used to assess the conjunction of rule antecedents.

The formula we used for our aggregated output is defined by formula (6):

$$\mu_{A \circ B}(x, z) = \max [\min(\mu_A(x, y), \mu_B(y, z))] \quad (6)$$

And the defuzzifier we used for our crisp output is centroid defuzzifier COG/COA, which is defined by $\text{COG} = \frac{\sum_{x=a}^b \mu_A(x) * x}{\sum_{x=a}^b \mu_A(x)}$ where $\mu_A(x)$ denotes the area of the sub-area and x indicates the centroid of the sub-area. We have six sub-areas the crisp output point for these sub-areas of our FFAF Inference System is:

$$C_{\text{FAIRness}} = (0.83 * 12.5 + \dots + 0.17 * 2.5) / (10 + \dots + 0.425) = (1755.225 / 32.225) = 54.467$$

In the next section, we will demonstrate our FFAF Inference System using the R programming language.

5.3 DEMONSTRATION OF NKN USE CASE IN R

To demonstrate the FFAF (4 inputs, 1 output) system, we have employed the Fuzzy logic programming in R. There are several different R packages available for Fuzzy logic programming. This method uses the “sets” package [63].

In the beginning, we defined the range and granularity of our universe, so our universe range is between 0 and 100 with a granularity of 0.1. The crisp inputs of all variables must be in this range, and the granularity is used to specify the accuracy of the fuzzy inference. Furthermore, the range defines the x-axis of our system plot. The command for this process is shown in this R code, `sets_options(“universe”, seq(from = 0, to = 100, by = 0.1))`. After this step, we defined our linguistic variables, which we use to describe our numeric variables. For example, our linguistic variables of the FAIR principles can have three levels of rating, as described by Table 5.1. We defined “LF” to represent Low Findable, “MF” to represent Medium Findable, and “HF” to represent High Findable.

Out of the variety of fuzzy membership functions available to define variables, we used Triangular and Trapezoidal membership functions. This is illustrated in the following R code: `fuzzy_trapezoid(corners = c(-1, 1, 20, 50))`, `fuzzy_triangular(corners = c(20, 50, 80))`, and `fuzzy_trapezoid(corners = c(50, 80, 100, 101))`. After defining all the variables, the membership function was assigned to each of them, as shown in the following R code:

```
“variables <- set(
```

```

Findable = fuzzy_variable(LF = fuzzy_trapezoid(corners = c(-1, 0, 20, 50)),
MF = fuzzy_triangular(corners = c(20, 50, 80)), HF = fuzzy_trapezoid(corners = c(50, 80, 100, 101))),

Accessible = fuzzy_variable(LA = fuzzy_trapezoid(corners = c(-1, 0, 20, 50)),
MA = fuzzy_triangular(corners = c(20, 50, 80)), HA = fuzzy_trapezoid(corners = c(50, 80, 100, 101))),

Interoperable = fuzzy_variable(LI = fuzzy_trapezoid(corners = c(-1, 0, 20, 50)),
MI = fuzzy_triangular(corners = c(20, 50, 80)), HI = fuzzy_trapezoid(corners = c(50, 80, 100, 101))),

Reusable = fuzzy_variable(LR = fuzzy_trapezoid(corners = c(-1, 0, 20, 50)),
MR = fuzzy_triangular(corners = c(20, 50, 80)), HR = fuzzy_trapezoid(corners = c(50, 80, 100, 101))),

FAIRness = fuzzy_variable(PF = fuzzy_trapezoid(corners = c(-1, 0, 20, 50)),
OF = fuzzy_triangular(corners = c(20, 50, 80)), GF = fuzzy_trapezoid(corners = c(50, 80, 100,
101))))”

```

Now that the linguistic variables have been defined, we move to the creation of rules. Fuzzy rules were used to link the linguistic variables of Findable, Accessible, Interoperable, and Reusable to the linguistic variable FAIRness. We had four variables each consist of three categories that resulted in a total of 81 rules ($3^4=81$ Rules). Below is one rule example:

```

“rules <- set(fuzzy_rule(Findable %is% LF && Accessible %is% LA && Interoperable
%is% LI && Reusable %is% LR, FAIRness %is%PF)”

```

The rest of the 81 rules’ R code can be found in Appendix A.

Now that the linguistic variables and rules are defined, we build our model with the following R code:

```

“model <- fuzzy_system(variables, rules)”

```

Next, the result of all rules are fuzzy sets that must be aggregated, and the max-min composition method is used as the aggregation method, which is denoted by “implication = c("minimum")”

in the following R code. Then, the next step is the fuzzy inference, which allows us to input the values for Findable, Accessible, Interoperable, and Reusable. The following command stores the inferred value into the variable “NKNuseCase”.

```
“NKNuseCase <- fuzzy_inference(model, list(Findable = 75, Accessible = 50, Interoperable = 25, Reusable = 50) , implication = c("minimum"))”
```

The inferred FAIRness values of all membership functions were stored in “NKNuseCase” as a composition of fuzzy sets.

Finally, we reached the defuzzification step to get the crisp output. As stated earlier, there are several algorithms for defuzzification, and we employed the centroid defuzzification method.

The following command performs a centroid defuzzification:

```
“gset_defuzzify(NKNuseCase, "centroid")”
```

The result of the experiment is presented below.

5.3.1 RESULT OF FFAF SYSTEM

In this section, we will describe the result of the FFAF system. We will go through the Fuzzy logic process as reflected in the FFAF steps and show the corresponding result.

First, in the fuzzification step, we converted crisp data into fuzzy data. We had four crisp data points, each of which represented a FAIR pillar (Findable, Accessible, Interoperable, and Reusable). We had 75 as crisp data input for Findable, 50 as crisp data input for Accessible, 25 as crisp data input for Interoperable, and 50 as crisp data input for Reusable. These crisp data

inputs were converted to fuzzy data using trapezoidal and triangular fuzzifiers. The resultant fuzzy sets were as follows:

As shown in Fig. 5.1, fuzzy set of 75 crisp data point is Findable(LF,MF,HF)={0.0,0.17,0.83}.

As shown in Fig. 5.2, fuzzy set of 50 crisp data point is Accessible(LA,MA,HA)={ 0.0,1.0,0.0}.

As shown in Fig. 5.3, fuzzy set of 25 crisp data point is Interoperable(LI,MI,HI)={0.83, 0.17,0.0}.

As shown in Fig. 5.4, fuzzy set of 50 crisp data point is Reusable(LR,MR,HR)={0.0,1.0,0.0}.

Next, the fuzzy inference bonded the membership functions to the fuzzy rules to generate the fuzzy output. For that reason, we built 81 Fuzzy rules and ran the fuzzy sets against these rules. Then, we obtained an aggregated fuzzy set combining all the original fuzzy sets, as shown in Fig. 5.5.

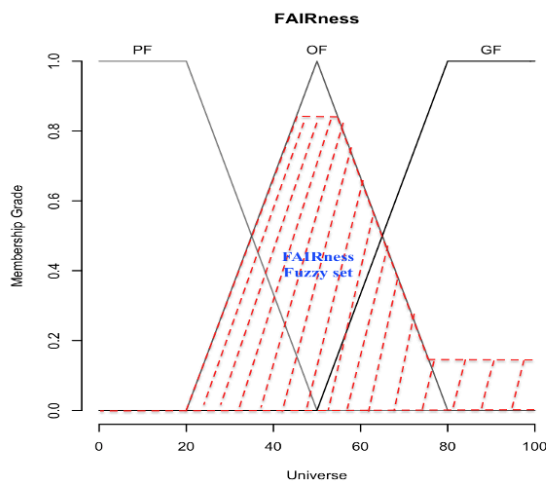


Fig. 5.5: FAIRness fuzzy set

Finally, the Defuzzification step received the resulting single FAIRness fuzzy set and computed a single crisp output using the COG centroid method, as shown in Fig. 5.6.

Fig. 5.6 shows the result of FFAF, which is the crisp output of %54.467, and that signifies the FAIRness level of the “Data for Building an Open Science Framework to Model Soil Organic Carbon” dataset.

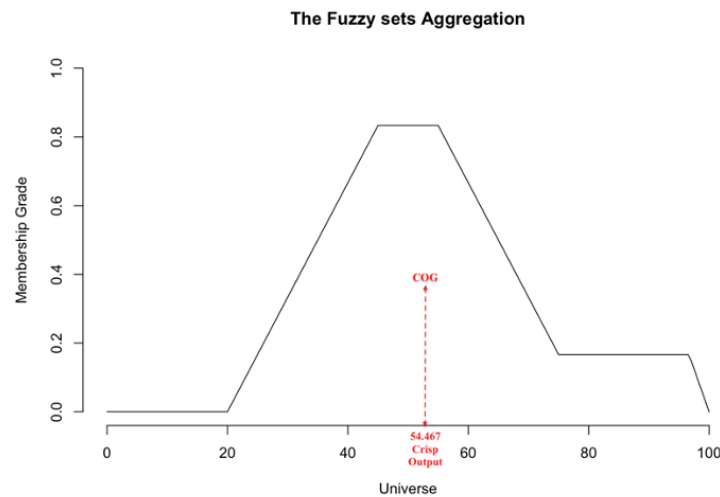


Fig. 5.6: FAIRness level crisp output

5.4 DISCUSSION

Accurately measuring the FAIRness level of a digital resource like a dataset is a daunting problem. For that reason, we initiated the following research question:(4) How can we efficiently evaluate the FAIRness of a digital resource? To answer this research question, we in chapter 5 introduced a novel method for efficiently evaluating the FAIRness level of datasets adopting FAIR data principles; we have shown the possibility of utilizing Fuzzy logic to evaluate the FAIRness level of a digital resource. To evince this possibility, we built a Fuzzy FAIR Assessment Framework (FFAF) to assess the FAIRness of digital resources. Furthermore, to demonstrate the usefulness of our method, we used it to measure the FAIRness

level for one of the NKN datasets called “Data for Building an Open Science Framework to Model Soil Organic Carbon,” this implementation of the FFAF model ultimately led to a specific FAIRness level crisp value. We thus assert that fuzzy logic is an efficient method for evaluating the FAIRness level of a digital resource. And this, therefore, certainly addresses research question number (4) of this dissertation.

5.5 CONCLUSIONS

The potential of utilizing the FFAF method to assess FAIRness uncertainty is an effective method to evaluate the FAIRness level. FFAF method can be reused several times for the same dataset to measure the improvement of the FAIRness level as more metadata introduce to the FAIRtool.org system. When the missing metadata is fulfilled with time by the dataset owner, the rating points will increase, and then the FAIRness level will rise. This rise in FAIRness level, in turn, will encourage the dataset owners to provide more controlled metadata, which will result in a higher degree of FAIRness level of the dataset. Consequently, this will contribute to improving the data stewardship for Earth science. This novel method is promising because the paper reviewers stated the paper will be well-cited. We can utilize the proposed Fuzzy logic-based FAIRness evaluation method FFAF to demonstrate more use cases; to do so, Appendix A includes the full R code that builds this FFAF system.

Chapters 3-5 presented and explained approaches to address the four research questions specified in chapter 1. Chapter 6 below articulates the outcomes of the studies mentioned in the previous chapters, present conclusions, and recommendations for future research.

Chapter 6: Conclusions

The studies conducted in chapters 3-5 explored methods and solutions to answer the four research questions specified in chapter 1. The subsequent sections in this chapter summarize the results of chapters 3-5, provide conclusions, and include recommendations for future research.

6.1 SUMMARY OF RESULTS

Chapter 3, the improvement of data stewardship through the utilization of an implementable framework for the FAIR data principles for Earth science data management and stewardship is the objective of this dissertation. Rationally, we strongly believe that formal logic must be the base of any successful applied work; therefore, we started by utilizing modern symbolic logic methods. First, we prepared the FAIR data principles for logical analysis by performing several steps, as illustrated in section 3.2.2. After the FAIR data principles sentences were converted to arguments and then symbolized; then, inference rules, as described in Table 2.4 and Table 3.1, were compared with the symbols of FAIR data principles arguments to deduce the equivalent inference rules (e.g., Absorption (Abs.), Hypothetical Syllogism (H.S.), and Modus Ponens (M.P.)) to be used by logical analysis. After that, we used two well-known logic analysis methods, the Truth Table method, and the Natural Deduction method, respectively, to prove the validity and tautology of the FAIR data principles. Finally, we concluded with the proof of the Tautology of FAIR data principles; therefore, by *definition 2.3*, we devised FAIR theorems, as shown in Table 3.9 and Table 3.10.

Chapter 4, after we affirmed the validity and tautology of FAIR data principles and generated FAIR theorems, we were ready to perform the FAIR technical implementation. We started by identifying the needed technologies. Then, we adopted semantic web technology because it is the only web of data technology, as explained in section 2.5. The semantic web application was built based on FAIR ontology to accommodate the logically validated FAIR data principles, as illustrated in section 4.2. FAIRtool.org was designed with user-friendliness and an intuitive web interface to allow scientists of the Earth science community to fill in their metadata according to FAIR data principles guidelines. This semantic web application is supported by in-house built FAIR ontology, which consists of the required set of semantic classes, semantic object properties, and semantic data properties. As described in section 4.2.1, these classes and properties were collected from a variety of well-known ontologies to fulfill the FAIR data principles characteristics. One of the main advantages of this semantic web application was the generation of an RDF file for the metadata of the FAIR data principles; the RDF file format constitutes the web of data because it depends on URIs to define digital resources. As described in section 2.5, RDF distinctly identifies digital resources; therefore, URIs enable data to be Findable, Accessible, Interoperable, and then Reusable via the World Wide Web. Section 4.2 demonstrated a use case of the Earth science community to verify the compliance of the recently constructed semantic web application (i.e., FAIRtool.org) with the FAIR data principles.

Chapter 5, as we have seen in chapter 4, FAIRtool.org produces FAIR metadata for a digital resource, e.g., dataset; through this procedure, we logged rating points for every present FAIR data principle; these rating points are to evaluate the FAIRness level of that dataset. Thus,

to measure the FAIRness level efficiently, in this Chapter, we delivered a method that produces a specific FAIRness level because it is important to emphasize that the FAIRness of digital resources should be measured through a controlled framework to reflect the right level of FAIR maturity. In section 5.2, we described an intuitive method using fuzzy logic to perform FAIRness evaluation for the dataset “Data for Building an Open Science Framework to Model Soil Organic Carbon; [71]” this method is called the Fuzzy FAIRness Assessments Framework (FFAF). FFAF consists of four steps: 1) Modeling FFAF inputs; 2) Fuzzifying inputs; 3) Inferencing the fuzzified inputs; 4) Aggregating and defuzzifying the fuzzy output. The design of FFAF is four-crisp-inputs and one-crisp-output. To demonstrate the FFAF (four inputs, one output) system, we have implemented FFAF with the Fuzzy logic set package library in R programming language; then, demonstrated four inputs rating points (75, 50, 25,50) representing Findable, Accessible, Interoperable, and Reusable of the dataset “Data for Building an Open Science Framework to Model Soil Organic Carbon; [71]” and one output representing the FAIRness level. The result of this demonstration, as shown in Fig. 5.6, was % 54.476 FAIRness level for the dataset “Data for Building an Open Science Framework to Model Soil Organic Carbon. [71]”

6.2 MAIN CONCLUSIONS

Earth science suffers from sparse data stewardship due to several challenges such as lack of metadata standards, locating needed datasets, ruinous metadata authoring process, and longtime data curation. Therefore, the objective of this research is to improve the stewardship

of Earth science digital resources through the implementation of FAIR data principles. In order to satisfy this objective, we addressed three essential aspects: theoretical analysis, technical implementations, and FAIRness level assessment. Firstly, the significance of the theoretical part is to logically validate the FAIR data principles before the technical deployment takes place. Because coherent technical implementation relies on verifying the validity and tautology of FAIR data principles, the proof of validity and tautology of the FAIR data principles was performed by chapter 3; besides, the theoretical contribution was addressed by the formulation of the FAIR theorems, as depicted in Table 3.9, and Table 3.10. These theoretical aspects paved the way for technical implementations. Secondly, the technical represented by the technical implementation of FAIR data principles, the emerging technology of Semantic Web grounded by FAIR ontology constructed the FAIRtool.org to implement FAIR data principles. Furthermore, the significance of FAIRtool.org is to facilitate the adoption of FAIR data principles; in turn, this improves the data stewardship of Earth science. Thirdly, once we have the metadata of the dataset processed by FAIRtool.org, then, how can we efficiently assess its FAIRness level. Adequately, chapter 5 presented a method called the Fuzzy FAIR Assessment Framework (FFAF) based on Fuzzy logic to evaluate the FAIRness level; the demonstration of a use case from the Earth science domain showed an efficient FAIRness level measurement for a dataset. Finally, we are nowadays in the age of data, to promote data management and stewardship of the scientific community including Earth science, must act now and move toward the web of structured data utilizing web of data tool like FAIRtool.org to have their data take a position on the web of data and enhance their data stewardship. Our work in this dissertation concreted the road for the Earth science community to improve their data

stewardship and open the door for other scientific communities to follow this path to enhance their data stewardship as well.

6.3 RECOMMENDATIONS FOR FUTURE RESEARCH

- 1) Deploy Machine Learning (ML) methods to suggest values for the missing FAIR data principles elements.
 - Ultimately, FAIRtool.org will accumulate the FAIR dataset; once we have enough FAIR metadata dataset, we are planning to incorporate Machine Learning algorithms to fulfill missing FAIR metadata.

- 2) The integration between FAIRtool.org and FFAF.
 - Establish the work for integrating FAIRtool.org with the FAIRness evaluation assessment tool FFAF. The goal is to incorporate the FFAF tool inside FAIRtool.org so the user can perform a FAIRness level evaluation assessment from within the FAIRtool.org. For this to be done, the SPARQL endpoint of FAIRtool.org must be enabled and then the SPARQL query can be triggered from within the R code of FFAF, which can be embedded inside FAIRtool.org. This will pull the rating values from FAIRtool.org and use it as crisp inputs to the FFAF system.

- 3) Develop an ontology that applies to the standards of other academic disciplines such as life sciences.

- Extend this work to accommodate other disciplines. FAIR ontology can be manipulated to cover other scientific areas. However, domain knowledge is required to create a new ontology and implement the new solution [70].
- 4) Integration with Google dataset search engine.
- Work with the Google dataset search engine team to establish a connection between Google dataset search engine and FAIRtool.org. Google dataset search engine can yield promising results when integrated with tools that reserve metadata of datasets. Google dataset search engine provides API that can be embedded in FAIRtool.org to pull datasets metadata.
- 5) Integration with service providers of Metadata.
- Integrate with metadata providers like DataCite that provide automatic doi and Creative Commons, which provides licensing policies. Also, other useful metadata services are available to enhance the automation of metadata creation. However, these metadata services providers apply charges for providing these services.

References

[1] Wilkinson, M. D., et al. "Comment: the FAIR guiding principles for scientific data management and stewardship. *Sci Data* 3." (2016).

[2] Plotkin, David. *Data stewardship: An actionable guide to effective data management and data governance*. Newnes, 2013.

[3] Mons, Barend. *Data stewardship for open science: Implementing FAIR principles*. CRC Press, 2018.

[4] McQuilton, Peter, et al. "BioSharing: curated and crowd-sourced metadata standards, databases and data policies in the life sciences." *Database* 2016 (2016).

[5] Bui, Yen, and Jung-ran Park. "An assessment of metadata quality: A case study of the national science digital library metadata repository." *Proceedings of the Annual Conference of CAIS/Actes du congrès annuel de l'ACSI*. 2006.

[6] Tenopir, Carol, et al. "Data sharing by scientists: practices and perceptions." *PloS one* 6.6 (2011): e21101.

[7] Rosen, K. H. "Boolean Algebra." *Discrete Mathematics and Its Applications, 8th ed.* New York, NY: McGraw-Hill (2019): 847-883.

[8] Martínez-Romero, Marcos, et al. "Fast and accurate metadata authoring using ontology-based recommendations." *AMIA Annual Symposium Proceedings*. Vol. 2017. American Medical Informatics Association, 2017.

- [9] Kacprzak, Emilia, et al. "A query log analysis of dataset search." *International Conference on Web Engineering*. Springer, Cham, 2017.
- [10] Koesten, Laura M., et al. "The Trials and Tribulations of Working with Structured Data: -a Study on Information Seeking Behaviour." *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 2017.
- [11] Starr, Joan, et al. "Achieving human and machine accessibility of cited data in scholarly publications." *PeerJ Computer Science* 1 (2015): e1.
- [12] Roche, Dominique G., et al. "Public data archiving in ecology and evolution: how well are we doing?." *PLoS Biol* 13.11 (2015): e1002295.
- [13] Hodson, Simon, et al. "Turning FAIR Data into Reality: Interim Report from the European Commission Expert Group on FAIR Data (Version Interim Draft)." *Interim Report from the European Commission Expert Group on FAIR Data*, no. June, 2018, doi:10.5281/zenodo.1285272.
- [14] National Institutes of Health. "NIH Strategic Plan for Data Science. 2018." *Online*. https://commonfund.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf (2019).
- [15] "FAIR Play in Geoscience Data." *Nature Geoscience*, vol. 12, no. 12, Nature Research, 1 Dec. 2019, p. 961, doi:10.1038/s41561-019-0506-4.
- [16] van Reisen, Mirjam, et al. "Towards the Tipping Point for FAIR Implementation." *Data Intelligence* 2.1-2 (2020): 264-275.
- [17] Magnus, P. D., et al. "forall x: Calgary. An Introduction to Formal Logic." (2020).

- [18] Magnus, P. D. "forall x: An introduction to formal logic." (2005).
- [19] Zadeh, Lotfi A. "Fuzzy sets." *Information and control* 8.3 (1965): 338-353.
- [20] Copi, Irving M., Carl Cohen, and Victor Rodych. *Introduction to logic*. Routledge, 2018.
- [21] Brickley, Dan. "RDF vocabulary description language 1.0: RDF schema." <http://www.w3.org/TR/rdf-schema/> (2004).
- [22] McGuinness, Deborah L., and Frank Van Harmelen. "OWL web ontology language overview." *W3C recommendation* 10.10 (2004): 2004.
- [23] Lowe, Brian, et al. "The Vitro Integrated Ontology Editor and Semantic Web Application." *ICBO*. 2011.
- [24] Musen, Mark A. "The protégé project: a look back and a look forward." *AI matters* 1.4 (2015): 4-12.
- [25] Staab, Steffen, and Rudi Studer, eds. *Handbook on ontologies*. Springer Science & Business Media, 2010.
- [26] Rodriguez, Marko A., et al. "Using RDF to model the structure and process of systems." *Unifying Themes in Complex Systems VII*. Springer, Berlin, Heidelberg, 2012. 222-230.
- [27] Gruber, Thomas R. "A translation approach to portable ontology specifications." *Knowledge acquisition* 5.2 (1993): 199-220.

- [28] DuCharme, Bob. *Learning SPARQL: querying and updating with SPARQL 1.1*. "O'Reilly Media, Inc.", 2013.
- [29] Borst, W. "Construction of engineering ontology." *Ph. D. Dissertation, Institute for Telematica and Information Technology, University of Twente, Enschede, the Netherlands* (1997).
- [30] Studer, Rudi, V. Richard Benjamins, and Dieter Fensel. "Knowledge engineering: principles and methods." *Data & knowledge engineering* 25.1-2 (1998): 161-197.
- [31] Jacobsen, Annika, et al. "FAIR principles: interpretations and implementation considerations." (2020): 10-29.
- [32] Di, Liping, Yuanzheng Shao, and Lingjun Kang. "Implementation of geospatial data provenance in a web service workflow environment with ISO 19115 and ISO 19115-2 lineage model." *IEEE transactions on geoscience and remote sensing* 51.11 (2013): 5082-5089.
- [33] Moreau, Luc, and Paul Groth. "Provenance: an introduction to PROV." *Synthesis Lectures on the Semantic Web: Theory and Technology* 3.4 (2013): 1-129.
- [34] Natsiavas, Pantelis, et al. "OpenPVSignal: advancing information search, sharing and reuse on pharmacovigilance signals via FAIR principles and Semantic Web technologies." *Frontiers in pharmacology* 9 (2018): 609.
- [35] Böhmer, Jasmin K. "Are the fair data principles fair?." (2017).
- [36] Wilkinson, Mark D., et al. "Evaluating FAIR-compliance through an objective, automated, community-governed framework." *BioRxiv* (2018): 418376.

[37] Celebi, Remzi, et al. "Towards FAIR protocols and workflows: The OpenPREDICT case study." *arXiv preprint arXiv:1911.09531* (2019).

[38] Mark, D., et al. "Evaluating FAIR Maturity Through a Scalable, Automated, Community-Governed Framework."

[39] Clarke, Daniel JB, et al. "FAIRshake: toolkit to evaluate the findability, accessibility, interoperability, and reusability of research digital resources." *BioRxiv* (2019): 657676.

[40] de Miranda Azevedo, Ricardo, and Michel Dumontier. "Considerations for the conduction and interpretation of FAIRness evaluations." *Data Intelligence 2.1-2* (2020): 285-292.

[41] da Silva Santos, LO Bonino, et al. "FAIR data points supporting big data interoperability." *Enterprise Interoperability in the Digitized and Networked Factory of the Future. ISTE, London* (2016): 270-279.

[42] Wilkinson, Mark D., et al. "Interoperability and FAIRness through a novel combination of Web technologies." *PeerJ Computer Science* 3 (2017): e110.

[43] Kusumasari, Tien Fabrianti. "Data profiling for data quality improvement with OpenRefine." *2016 International Conference on Information Technology Systems and Innovation (ICITSI)*. IEEE, 2016.

[44] Musen, Mark A., et al. "The center for expanded data annotation and retrieval." *Journal of the American Medical Informatics Association* 22.6 (2015): 1148-1152.

[45] UniProt Consortium. "UniProt: a hub for protein information." *Nucleic acids research* 43.D1 (2015): D204-D212.

- [46] Bai, Ying, Hanqi Zhuang, and Dali Wang, eds. *Advanced fuzzy logic technologies in industrial applications*. Springer Science & Business Media, 2007.
- [47] Cheng, Jui-Chuan, Chao-Yuan Chiu, and Te-Jen Su. "Training and Evaluation of Human Cardiorespiratory Endurance Based on a Fuzzy Algorithm." *International Journal of Environmental Research and Public Health* 16.13 (2019): 2390.
- [48] Baader, Franz, Ian Horrocks, and Ulrike Sattler. "Description logics." *Handbook on ontologies*. Springer, Berlin, Heidelberg, 2004. 3-28.
- [49] Wong, Calvin, Z. Xiao Guo, and S. Y. S. Leung. *Optimizing decision making in the apparel supply chain using artificial intelligence (AI): from production to retail*. Elsevier, 2013.
- [50] Van de Sompel, Herbert, et al. "Persistent identifiers for scholarly assets and the web: The need for an unambiguous mapping." (2014).
- [51] Berners-Lee, Tim, James Hendler, and Ora Lassila. "The semantic web." *Scientific american* 284.5 (2001): 34-43.
- [52] Chung, Seung-Hwa, et al. *The MOUSE Approach: Mapping Ontologies Using UML for System Engineers*. Diss. Trinity College Dublin, 2015.
- [53] Berners-Lee, Tim, Roy Fielding, and Larry Masinter. "Uniform Resource Identifier (URI): Generic Syntax (RFC 3986)." *Network Working Group* (2005).
- [54] Pérez, Jorge, Marcelo Arenas, and Claudio Gutierrez. "Semantics and Complexity of SPARQL." *International semantic web conference*. Springer, Berlin, Heidelberg, 2006.

- [55] Mamdani, Ebrahim H., and Sedrak Assilian. "An experiment in linguistic synthesis with a fuzzy logic controller." *International journal of man-machine studies* 7.1 (1975): 1-13.
- [56] Takagi, Tomohiro, and Michio Sugeno. "Fuzzy identification of systems and its applications to modeling and control." *IEEE transactions on systems, man, and cybernetics* 1 (1985): 116-132.
- [57] Shoureshi, Rahmat, and Zhi Hu. "Tsukamoto-type neural fuzzy inference network." *Proceedings of the 2000 American Control Conference. ACC (IEEE Cat. No. 00CH36334)*. Vol. 4. IEEE, 2000.
- [58] Singh, Himanshu, and Yunis Ahmad Lone. "Fuzzy Inference Systems." *Deep Neuro-Fuzzy Systems with Python*. Apress, Berkeley, CA, 2020. 93-127.
- [59] Ross, Timothy J. *Fuzzy logic with engineering applications*. Vol. 2. New York: wiley, 2004.
- [60] Guo, Z., and W. Wong. "Fundamentals of artificial intelligence techniques for apparel management applications." *Optimizing Decision Making in the Apparel Supply Chain Using Artificial Intelligence (AI)*, Woodhead Publishing Series in Textiles (2013): 13-40.
- [61] Saade, Jean J., and Hassan B. Diab. "Defuzzification methods and new techniques for fuzzy controllers." (2004): 161-174.
- [62] Pratihar, Dilip Kumar. *Soft computing: fundamentals and applications*. Alpha Science International, Ltd, 2013.

[63] Hornik, Kurt, and David Meyer. "Generalized and customizable sets in R." *Journal of Statistical Software* 31.2 (2009): 1-27.

[64] Cox, Earl. *Fuzzy modeling and genetic algorithms for data mining and exploration*. Elsevier, 2005.

[65] Kohout, Ladislav J., P. M. Pardalos, and C. A. Floudas. "Boolean and Fuzzy Relations." *Encyclopedia of Optimization* 1 (2009): 189-202.

[66] Hardegree, Gary M. *Symbolic logic: A first course*. McGraw-Hill, 1994.

[67] Forbes, Graeme. "Modern logic: A text in elementary symbolic logic." (1994).

[68] Krafft, Dean B., et al. "Vivo: Enabling national networking of scientists." (2010).

[69] Butt, Anila Sahar. *Analysis of Semantic Web Databases*. Diss. School of Electrical Engineering and Computer Science, National University of Sciences and Technology, 2010.

[70] Wache, Holger, et al. "Ontology-Based Integration of Information-A Survey of Existing Approaches." *Ois@ijcai*. 2001.

[71] Flathers, Edward, and Paul E. Gessler. "Building an open science framework to model soil organic carbon." *Journal of environmental quality* 47.4 (2018): 726-734.

[72] <https://www.northwestknowledge.net>

[73] Grunsky, E. C. "R: a data analysis and statistical programming environment—an emerging tool for the geosciences." *Computers & Geosciences* 28.10 (2002): 1219-1222.

[74] NCDC. "NCDC Storm Events Database." (2007).

LIST OF PUBLICATIONS

REFEREED PAPERS IN CONFERENCES

- Alowairdhi, A., Ma, X., 2020. Utilizing Fuzzy Logic for Assessing "FAIRness" of a Digital Resource. Proceedings of the 2020 International Conference on Computational Science and Computational Intelligence, Las Vegas, VA. In Press

Abstract—The goal of this research is to examine the possibility of utilizing fuzzy logic to evaluate the uncertainty of FAIRness level (findability, accessibility, interoperability, and reusability) of a digital resource. To date, there are no FAIRness evaluation studies based on fuzzy logic. To measure this uncertainty, we built a Fuzzy FAIR Assessment Framework (FFAF). FFAF is based on the three main steps of fuzzy logic: fuzzification, inferencing, and defuzzification. Applying the FFAF model on the FAIR data principles led to a specific FAIRness level result. Overall, this research shows that fuzzy logic is an effective technique for evaluating the FAIRness level of a digital resource.

REFEREED PAPERS IN EDITED BOOKS

- Alowairdhi A., Ma X., 2019. Data Brokers and Data Services. In: Schintler L., McNeely C. (eds.) Encyclopedia of Big Data. Springer, Cham, Switzerland. 4pp. DOI: 10.1007/978-3-319-32001-4_298-1

Abstract—Data brokers are agents that gather individuals' information from a variety of online and offline sources, such as mail, e-mail, social media posts, personal websites, vehicle records, US census records, retailers' system entries, and real estate history records. Data brokers often collect these data without the permission or awareness of the involved individuals. Data are combined and synthesized using sophisticated analytic tools and then offered through data services for sale or rent to other data brokers, businesses, or individuals for different purposes. This entry provides an overview of data brokers, including what data they collect, the sources and methods

they use, the benefits and risks for individuals involved in data collection, and the choices for individuals to opt out the data collection.

PRESENTATIONS IN CONFERENCES AND WORKSHOPS

- Conference Speaker at The ESIP 2019 Summer Meeting, July 16-19, 2019. Tacoma, WA.
- Conference Speaker at The ESIP 2020 Winter Meeting, July 7-9, 2020. Bethesda, MD.
- US2TS Semantic Technologies Symposium 2019, Duke, NC.
- Workshop: Implementing FAIR Data for People and Machines: Impact and Implications- September 11, 2019, Washington, DC.
- Workshop: UI Research Computing and Data Science Symposium, May 15, 2019.
- VIVO 2017 Conference Weill Cornell Medicine, August 2-4, 2017. New York, NY.
- Presenter at the 2020 International Conference on Computational Science and Computational Intelligence (CSCI'20: December 16-18, 2020, Las Vegas, USA)

REFEREED PAPERS IN PROGRESS

- Alowairdhi, A., Ma, X. A logical perspective on the Implementation of the FAIR data principles. Data Science Journal, Under Review.

Abstract— In this paper we present the method for a formal logical evaluation of the FAIR data principles. Our objective is to examine the elements of FAIR for logical validity and tautology. The study uses two formal logical methods, namely, the Truth Table and the Natural Deduction. In addition, the Sentential Logic was used as the formal logic language. The design of the study is based on four consecutive processes: 1) Build argument for the FAIR sentences; 2) Paraphrase, symbolize, and translate the

FAIR arguments into logical symbols and prepare for the formal logical analysis; 3) Conduct formal logical analysis of FAIR arguments using the Truth Table method to discover the truth value of the FAIR argument symbols; and 4) Conduct formal logical analysis of FAIR arguments using the Natural Deduction method in order to deduct the FAIR argument validity using the inference rules. The results of the study are the proof of the logical validity and tautology of the FAIR sentences. Ultimately, the proof leads to the development of the FAIR theorems, which culminate in the development of the FAIR theory. This research will help establish rules to translate the FAIR data principles into machine-readable formats, which are necessary for the implementation of FAIR in the cyberinfrastructure.

- Alowairdhi, A., Ma, X., FAIR Data Principles. Sagar, D.S.D. et al. (eds.) Encyclopedia of Mathematical Geosciences, Cham, Switzerland. Invited Book Chapter, In Progress.

GRANTS

- Earth Science Information Partners (ESIP) small grant recipient through the ESIP Lab program.
- US2TS 2019 travel grant awarded to attend US2TS Semantic Technologies Symposium 2019, Duke, NC.

Appendix A: FFAF R Code

```

# This code is to set up a fuzzy system for finding FAIRness value of a digital resource
# this code is an attachment to a research paper "Utilizing Fuzzy logic for assessing "FAIRness" of a digital resource"

install.packages("sets") ## install package one time only and call sets library
library(sets)

# set universe
sets_options("universal", seq(from = 0, to = 100, by = 0.5))

# set up fuzzy variables
variables <-
set( Findable = fuzzy_variable( LF = fuzzy_trapezoid(corners = c(-1, 0, 20, 50)),
  MF = fuzzy_triangular(corners = c(20, 50, 80)),
  HF = fuzzy_trapezoid(corners = c(50, 80, 100, 101))),
  Accessible = fuzzy_variable( LA = fuzzy_trapezoid(corners = c(-1, 0, 20, 50)),
  MA = fuzzy_triangular(corners = c(20, 50, 80)),
  HA = fuzzy_trapezoid(corners = c(50, 80, 100, 101))),
  Interoperable = fuzzy_variable( LI = fuzzy_trapezoid(corners = c(-1, 0, 20, 50)),
  MI = fuzzy_triangular(corners = c(20, 50, 80)),
  HI = fuzzy_trapezoid(corners = c(50, 80, 100, 101))),
  Reusable = fuzzy_variable( LR = fuzzy_trapezoid(corners = c(-1, 0, 20, 50)),
  MR = fuzzy_triangular(corners = c(20, 50, 80)),
  HR = fuzzy_trapezoid(corners = c(50, 80, 100, 101))),
  FAIRness = fuzzy_variable( PF = fuzzy_trapezoid(corners = c(-1, 0, 20, 50)),
  OF = fuzzy_triangular(corners = c(20, 50, 80)),
  GF = fuzzy_trapezoid(corners = c(50, 80, 100, 101)))
)

# Fuzzy rules

rules <- set(
fuzzy_rule( Findable %is% LF && Accessible %is% LA && Interoperable %is% LI && Reusable %is% LR , FAIRness %is% PF ),
fuzzy_rule( Findable %is% LF && Accessible %is% LA && Interoperable %is% LI && Reusable %is% MR , FAIRness %is% PF ),
fuzzy_rule( Findable %is% LF && Accessible %is% LA && Interoperable %is% LI && Reusable %is% HR , FAIRness %is% OF ),
fuzzy_rule( Findable %is% LF && Accessible %is% LA && Interoperable %is% MI && Reusable %is% LR , FAIRness %is% PF ),
fuzzy_rule( Findable %is% LF && Accessible %is% LA && Interoperable %is% MI && Reusable %is% MR , FAIRness %is% OF ),
fuzzy_rule( Findable %is% LF && Accessible %is% LA && Interoperable %is% MI && Reusable %is% HR , FAIRness %is% OF ),
fuzzy_rule( Findable %is% LF && Accessible %is% LA && Interoperable %is% HI && Reusable %is% LR , FAIRness %is% OF ),
fuzzy_rule( Findable %is% LF && Accessible %is% LA && Interoperable %is% HI && Reusable %is% MR , FAIRness %is% OF ),
fuzzy_rule( Findable %is% LF && Accessible %is% LA && Interoperable %is% HI && Reusable %is% HR , FAIRness %is% OF ),
fuzzy_rule( Findable %is% LF && Accessible %is% MA && Interoperable %is% LI && Reusable %is% LR , FAIRness %is% PF ),
fuzzy_rule( Findable %is% LF && Accessible %is% MA && Interoperable %is% LI && Reusable %is% MR , FAIRness %is% OF ),
fuzzy_rule( Findable %is% LF && Accessible %is% MA && Interoperable %is% LI && Reusable %is% HR , FAIRness %is% OF ),
fuzzy_rule( Findable %is% LF && Accessible %is% MA && Interoperable %is% MI && Reusable %is% LR , FAIRness %is% OF ),
fuzzy_rule( Findable %is% LF && Accessible %is% MA && Interoperable %is% MI && Reusable %is% MR , FAIRness %is% OF ),
fuzzy_rule( Findable %is% LF && Accessible %is% MA && Interoperable %is% MI && Reusable %is% HR , FAIRness %is% OF ),
fuzzy_rule( Findable %is% LF && Accessible %is% MA && Interoperable %is% HI && Reusable %is% LR , FAIRness %is% OF ),
fuzzy_rule( Findable %is% LF && Accessible %is% MA && Interoperable %is% HI && Reusable %is% MR , FAIRness %is% OF ),
fuzzy_rule( Findable %is% LF && Accessible %is% MA && Interoperable %is% HI && Reusable %is% HR , FAIRness %is% OF ),
fuzzy_rule( Findable %is% LF && Accessible %is% HA && Interoperable %is% LI && Reusable %is% MR , FAIRness %is% OF ),
fuzzy_rule( Findable %is% LF && Accessible %is% HA && Interoperable %is% LI && Reusable %is% HR , FAIRness %is% OF ),
fuzzy_rule( Findable %is% LF && Accessible %is% HA && Interoperable %is% MI && Reusable %is% LR , FAIRness %is% OF ),
fuzzy_rule( Findable %is% LF && Accessible %is% HA && Interoperable %is% MI && Reusable %is% MR , FAIRness %is% OF ),
fuzzy_rule( Findable %is% LF && Accessible %is% HA && Interoperable %is% MI && Reusable %is% HR , FAIRness %is% OF ),
fuzzy_rule( Findable %is% LF && Accessible %is% HA && Interoperable %is% HI && Reusable %is% LR , FAIRness %is% OF ),
fuzzy_rule( Findable %is% LF && Accessible %is% HA && Interoperable %is% HI && Reusable %is% MR , FAIRness %is% OF ),
fuzzy_rule( Findable %is% LF && Accessible %is% HA && Interoperable %is% HI && Reusable %is% HR , FAIRness %is% OF ),
fuzzy_rule( Findable %is% MF && Accessible %is% LA && Interoperable %is% LI && Reusable %is% LR , FAIRness %is% PF ),
fuzzy_rule( Findable %is% MF && Accessible %is% LA && Interoperable %is% LI && Reusable %is% MR , FAIRness %is% OF ),

```



```
# centroid defuzzification method to get single point crisp output
gset_defuzzify(NKNuseCase, "centroid")

# Other defuzzification methods

gset_defuzzify(NKNuseCase, "largestofmax")
gset_defuzzify(NKNuseCase, "smallestofmax")
gset_defuzzify(NKNuseCase, "meanofmax")

sets_options("universe", NULL) # always it is a good practice to Reset the universe
```