Special Issue, Volume 4, Article 22

Curriculum Development for FAIR Data Stewardship

Francisca Oladipo[1, 2, 3], Sakinat Folorunso[4, 2], Ezekiel Ogundepo[5, 2], Obinna Osigwe[1, 2], Akinyinka Akindele[6,2]

[1] Directorate of Research, Innovations, Consultancy and Extension, Kampala International University, Uganda

[2] Virus Over Data Network (VODAN)-Africa and Asia,

[3] Federal University Lokoja, Nigeria

[4] Department of Mathematical Sciences, Olabisi Onabanjo University, Nigeria

[5] Data Science Nigeria, Lagos, Nigeria

[6] School of eLearning Projects, Kampala International University, Uganda

Correspondence:

Sakinat Folorunso, Olabisi Onabanjo University

sakinat.folorunso@oouagoiwoye.edu.ng

Cite as: Oladipo, F., Folorunso, S., Ogundepo, E. A., Osigwe, E., Akindele, A.T.: Curriculum development for FAIR data stewardship. *Data Intelligence* Vol. 4 (2022)

Abstract

The FAIR Guidelines attempts to make digital data Findable, Accessible, Interoperable, and Reusable (FAIR). To prepare FAIR data, a new data science discipline known as data stewardship is emerging and, as the FAIR Guidelines gain more acceptance, an increase in the demand for data stewards is expected. Consequently, there is a need to develop curricula to foster professional skills in data stewardship through effective knowledge communication. There have been a number of initiatives aimed at bridging the gap in FAIR data management training through both formal and informal programmes. This article describes the experience of developing a digital initiative for FAIR data management training under the Digital Innovations and Skills Hub (DISH) project. The FAIR Data Management course offers 6 short on-demand certificate modules over 12 weeks. The modules are divided into two sets: FAIR data and data science. The core subjects cover elementary topics in data science, regulatory frameworks, FAIR data management, intermediate to advanced topics in FAIR Data Point installation, and FAIR data in the management of healthcare and semantic data. Each week, participants are required to devote 7–8 hours of self-study to the modules, based on the resources provided. Once they have satisfied all requirements, students are certified as FAIR data scientists and qualified to serve as both FAIR data stewards and analysts. It is expected that in-depth and focused curricula development with diverse participants will build a core of FAIR data scientists for Data Competence Centres and encourage the rapid adoption of the FAIR Guidelines for research and development.

Keywords: data steward, data science, FAIR Guidelines, FAIR, digital technology, FDP installation, FAIR Data Trains, semantic web, Personal Health Train

Acronyms

CF-DS	Data Science Competence Framework
DMP	data management plan
DS-BoK	Data Science Body of Knowledge
DSDA	Data Science Analytics
DSDK	Data Science Domain Knowledge
DSDM	Data Management and Governance
DSENG	Data Science Engineering
DSP	Data Science Professional
DSPP	Data Science Professional Profiles
EDSF	EDISON Data Science Framework
ESCO	European Skills, Competences, Qualifications and Occupations
FAIR	Findable, Accessible, Interoperable, Reusable
ICT	information and communications technology
ITSM	information technology service management
MC-DS	Data Science Model Curriculum

1. Introduction

In 2019, the World Economic Forum estimated that, by 2025, an average of 463 exabytes of data (Tweets, email messages, Facebook posts, WhatsApp messages, clinical data, and music files, etc.) will be created every day [1]. This data will be in different formats, like images, text, or audio, and from different domains. In response, the big data landscape is redefining requirements for data curation infrastructure, which is evolving to meet the challenges [2]. By employing data analytics, the metadata of curated health data can provide insights into solving health problems, gearing the industry toward value-based healthcare and opening doors to remarkable advancements, while reducing costs. However, constraints, such as the misrepresentation of data, privacy issues, siloed data, security, and data not being machine-readable, among other things, can lead to false inferences being drawn from data analytics. While the FAIR Guidelines [3] – that data should be Findable, Accessible, Interoperable and Reusable (FAIR) – tend to mitigate some of these constraints, these principles are foreign to most of the stakeholders whose devices, infrastructures and research generate such data. Thus, there is a need to train data stewards using customised training to equip them with the skills required to implement the FAIR Guidelines. Accordingly, an appropriate curriculum needs to be designed, validated and deployed, which is the subject of this article.

The design of any curriculum has four critical components that address four questions:

- Why is instruction initiated?
- What needs to be taught to achieve the set intent and objectives?
- How can we connect all target learning outcomes?
- What has been realised and what other actions need to be taken in relation to the instructional programme, learners, and teachers?

Worldwide, these components are usually addressed differently depending on the philosophy of the domain curriculum and model on which a design is based [4]. The goal of curriculum development is to communicate knowledge effectively to learners. This article explores the frameworks implemented in data stewardship programmes, towards designing a curriculum for training data stewards, in an effort to equip them with the relevant skills.

2. Literature review

2.1 Data stewardship: Description, roles and goals

Data stewardship is a concept that is deeply rooted in the sciences and should be considered in any funded research. It relates to the procedure for gathering, sharing, and analysing data and reflects the values underpinning fair information practices [5]. Principally, data stewardship involves all activities related to research data management over the research lifecycle. It has the potential to improve research, as it improves data management approaches for the collection, storage, aggregation, and de-identification of data, as well as procedures for data release and use [6]. In 2020, Wildgaard [7] posited that the position of a data steward is trust-based. Data stewards are responsible for the administration, management and manipulation of data belonging to researchers or enterprises. However, the professionalization of data stewardship can only progress with improved data steward education opportunities [7]. Therefore, as an activity that is part of performing creative research, data stewardship encompasses the design of all activities to do with (digital) data throughout the research project lifecycle, with the aim of optimising the usability, reusability, and reproducibility of the resulting data [8]. The study and practice of data stewardship is necessary for FAIR and open research. The European Open Science Cloud for Research Pilot Project [9] explains data stewardship as the shared responsibility of the professional groups involved in data management: data management and curation, data science and analytics, data services engineering

and domain research [9]. Competences, skills groups, and organisational roles are defined around typical processes and stages in data management: planning and design, capture and processing, integration and analysis, evaluation and presentation, publishing and release, exposure and discovery, governance and assessment, scope and resources, advice and enabling.

Collins et al. [9] point out that transitioning to FAIR data stewardship requires education programmes for both data scientists and data stewards. In fact, both pedagogy and curricula are needed. Some of the popular existing curricular frameworks for digital curation and data science are EDISON [10], EOSCPilot [9] and DigCurV [11]. These curricular frameworks could be implemented as postgraduate degree programmes in universities [12] to increase the accessibility of professional data science and stewardship programmes.

Wildgaard et al. [12] explain that the major roles for a data steward are administrator, analyst, developer and agent of change. Like the roles of the data system developer, the role of the data steward is to optimise the data through good project management, advise on FAIR Guidelines, create a data plan, facilitate collaboration and knowledge sharing to raise business intelligence, innovate, and develop procedures and guidelines. These authors also propose three models for data steward education: The first model is for students with bachelor degrees. This model spans one year for students with programming skills and two years for students without. The second model consists of PhD students or equivalent from any university faculty. And the third model is for students with professional studies or technical education and vocational training. Some of the other training options that could be explored for data stewardships as part of continuing professional development are: summer schools, on-the-job training, workshops, training-of-trainers, and online learning [9]. FAIR-themed programmes, like workshops, conference sessions, lectures, webinars, hackathons, workshops, visiting scholar programmes and so forth, could also be adopted to enhance FAIR data stewardship. All of these methods have proven to be effective in training

students from all disciplines on the foundational data skills they need to be professional data stewards. For examples, CODATA-RDA [13] organised a short course programme in the form of a summer school in 2019 to upskill the research community for professional FAIR data stewardship. Some of the subjects taught were research data science, research data management, software and data carpentry, machine learning, visualisation and computational infrastructure.

This requires universities and other data-rich facilities to invest in Data Competence Centers (DCCs). In the FAIR Data Science environment, these are called Data Stewardship Competence Centers (DSCCs), which are established to embed professional, institution-wide research data stewardship and its related infrastructure, and which collaborate with the data processors in their institutions to enable better data management and comply with the FAIR Guidelines (Go-FAIR). Rosenbaum [6] agrees that the majority of data stewards have good research data management and domain-specific knowledge, but notes that it would be beneficial to provide pedagogical training to impart the soft skills required to efficiently engage with researchers and meet their needs [6]. Accordingly, this article proposes designing a digital skills curriculum for FAIR data stewardship. The proposed curriculum is divided into three main courses: computing and information technology, analytics, and FAIR data.

2.2 EDISON Data Science Framework

The EDISON Data Science Framework (EDSF) provides a basis for the definition of data science and enables the definition of other components related to data science education, training, organisational roles, and skills management, as well as professional certification. This framework contains five main components:

- Data Science Competence Framework (CF-DS) [10]
- Data Science Body of Knowledge (DS-BoK) [14]
- Data Science Model Curriculum (MC-DS) [15]

- Data Science Professional Profiles (DSPP) and occupations taxonomy [16]
- Data Science Taxonomy and Scientific Disciplines Classification

The CF-DS provides the overall basis for the EDSF. The core CF-DS competences and skills groups identified by the EDISON Community [10] as essential for data scientists in different workplaces include:

- Data Science Analytics (DSDA) – which uses suitable statistical methods and predictive analytics (such as statistical analysis, machine learning, data mining, and business analytics, etc.) on presented data to deliver insights and discover new relations.
- Data Science Engineering (DSENG) – which uses engineering principles to research, design, develop and implement new instruments and applications for data collection, analysis and management.
- Data Management and Governance (DSDM) – which relates to the development and implementation of a data management approach (using techniques such as software and applications engineering, data warehousing, big data infrastructure and tools for data stewardship, curation, and preservation) for data collection, storage, preservation, and availability for further processing.
- Data Science Research Methods and Project Management (DSRMP) – which relates to the research domain, and Data Science Business Process Management (DSBPM), which creates new understandings and capabilities by using scientific methods (such as hypothesis, test/artefact, and evaluation) or similar engineering methods to discover new approaches to create new knowledge and achieve research or organisational goals.
- Data Science Domain Knowledge (DSDK) – which uses the domain knowledge (scientific or business) to develop relevant data analytics applications and adopt general data science methods

for domain specific data types and presentations, data and process models, organisational roles and relations.

The DS-BoK defines the knowledge areas (KA) required for building a data science curriculum that supports identified data science competences. The DS-BoK is organised by knowledge area groups (KAG) that correspond to the CF-DS competence groups. These are Data Science Analytics, Data Science Engineering, Data Management, Research Methods and Project Management, and Business Analytics [14]

The MC-DS is built based on CF-DS and DS-BoK, for which learning outcomes are defined based on CF-DS competences and learning units are mapped to knowledge units in DS-BoK. Three mastery (or proficiency) levels are defined for each learning outcome to allow for flexible curricula development and profiling for different data science professional profiles.

The DSPP is defined as an extension of the European Skills, Competences, Qualifications and Occupations (ESCO) to the ESCO occupations taxonomy, using the ESCO top classification groups. The definition of DSPP provides an important instrument for defining effective organisational structures and roles related to data science positions – and can be also used for building individual career paths and corresponding competences and skills transferability between organisations and sectors.

The Data Science Taxonomy and Scientific Disciplines Classification serves to maintain consistency and links between the four core components of EDSF (CF-DS, DS-BoK, MC-DS, and DSPP).

2.3 ESCO framework and platform

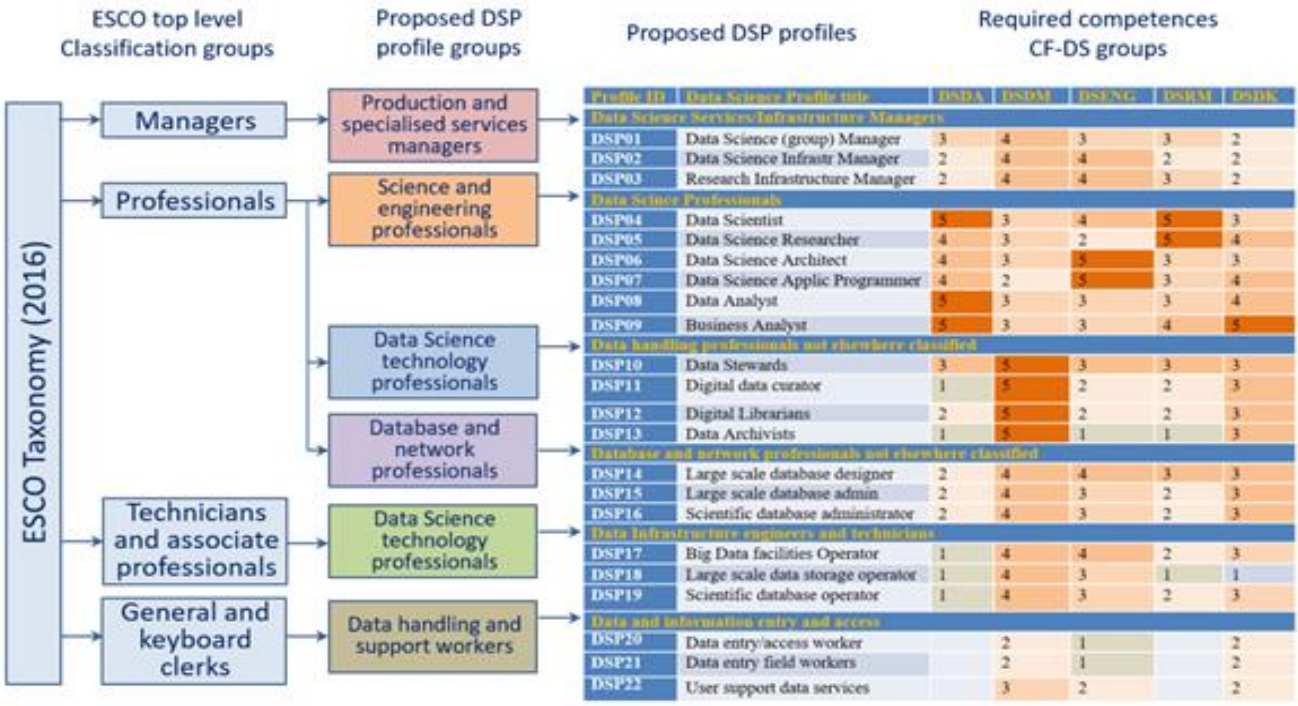
The ESCO classification identifies and categorises skills, competences, qualifications and occupations relevant for the European Union labour market, education and training. It systematically shows the relationships between the different concepts [17]. The ESCO Data Science Professional Profiles (DSPP) occupation hierarchy is: managers, professionals, technicians, and associate professionals, and clerical

support workers. The ESCO DSPP taxonomy can be extended to situations where proposed profile competences and organisational roles are similar to CEN Workshop Agreement (CWA) 16458 ICT profile definitions, such as to:

- Managers who are production and specialised services managers (data science/big data infrastructure managers) whose role spans DSP01–DSP03
- Professionals from three major groups:
    - Science and engineering professionals (data science professionals) whose roles span DSP04–DSP09)
    - Information and communication technology (ICT) professionals (data science technology professionals) whose roles span DSP10–DSP13
    - Science and engineering professionals (database and network professionals) whose roles span DSP14–DSP16
- Technicians and associate professionals, such as science and engineering associate professionals (data science technology professionals) whose roles span DSP17–DSP19
- Clerical support workers, such as general and keyboard clerks (data handling and support workers) whose roles span DSP20–DSP22

Figure 1 illustrates the existing ESCO hierarchy and the proposed new data science classification groups and corresponding new data science related profiles. The table in this figure shows competence groups relevant to each profile by indicating competence relevance from 0 to 5 (0 – not relevant, 5 – very important). The profile definitions for specific roles for DSP01–DSP22 are detailed on the EDISON Community website [16]. For example, the profile for data steward is DSP10 under the hierarchy of data science technology professionals. Mapping ‘data steward’ with CF-DS competences and skills groups, the

relevance level with DSDA, DSENG, DSRM and DSDK is 3. Data steward is most relevant to DSDM. Data steward is well mapped with the CF-DS competency groups with an average value of 3.



Profile ID	Data Science Profile title	DSDA	DSDM	DSENG	DSRM	DSDK
Data Science Services/Infrastructure Managers						
DSP01	Data Science (group) Manager	3	4	3	3	2
DSP02	Data Science Infrastr Manager	2	4	4	2	2
DSP03	Research Infrastructure Manager	2	4	4	3	2
Data Science Professionals						
DSP04	Data Scientist	5	3	4	5	3
DSP05	Data Science Researcher	4	3	2	5	4
DSP06	Data Science Architect	4	3	5	3	3
DSP07	Data Science Applic Programmer	4	2	5	3	4
DSP08	Data Analyst	5	3	3	3	4
DSP09	Business Analyst	5	3	3	4	5
Data handling professionals not elsewhere classified						
DSP10	Data Stewards	3	5	3	3	3
DSP11	Digital data curator	1	5	2	2	3
DSP12	Digital Librarians	2	5	2	2	3
DSP13	Data Archivists	1	5	1	1	3
Database and network professionals not elsewhere classified						
DSP14	Large scale database designer	2	4	4	3	3
DSP15	Large scale database admin	2	4	3	2	3
DSP16	Scientific database administrator	2	4	3	2	3
Data Infrastructure engineers and technicians						
DSP17	Big Data facilities Operator	1	4	4	2	3
DSP18	Large scale data storage operator	1	4	3	1	1
DSP19	Scientific database operator	1	4	3	2	3
Data and Information entry and access						
DSP20	Data entry/access worker		2	1		2
DSP21	Data entry field workers		2	1		2
DSP22	User support data services		3	2		2

Figure 1. Proposed data science related extensions to the ESCO classification hierarchy and corresponding DSPP by classification groups [16]

Table 1. DSP profiles definition [16]

Profile title	**Data steward (DSPP10)**		
Mission	Plans, implements and manages (research) data input, storage, search, presentation; creates data model for domain specific data; supports and advises domain scientists/researchers; interacts with the data analytics team; does data preparation, inspection, visualisation; prepares data for archiving and publication		
Deliverables	**Accountable** ▪ Data model ▪ Data management plan	**Responsible** ▪ Data collection/ingest	**Contributor** ▪ Domain related models ▪ Data analytics result inspection
Main tasks	▪ Define/build/optimise data model and schemas ▪ Use existing or define new metadata framework ▪ Publish research data to existing scientific data archives ▪ Manage organisational or project-related data ▪ Search and promote research data ▪ Assist main domain researcher/scientist in selecting right data analytics methods ▪ Monitor application of FAIR (Findable, Accessible, Interoperable, Reusable) and open data principles to data created by organisation or project		

Competences *(from CF-DS)*	SDSDM02: Use data storage systems, data archive services, digital libraries, and their operational models	Level 1
	SDSDM05: Implement data lifecycle support in organisational workflow, support data provenance and linked data	Level 2
	SDSDM06: Consistently implement data curation and data quality controls, ensure data integration and interoperability	Level 2
	SDSDM08: Use and implement metadata, Persistent Identifier (PID), data registries, data factories, standards and compliance	Level 3
	SDSDM09: Adhere to FAIR Guidelines for open data, open science, open access, use ORCID based services	Level 3
Key performance indicators (KPI) area	Consistent data management workflow Compliance with FAIR Guidelines	

The importance of the role of the data steward is recognised in the European Commission's High Level Expert Group report on European Open Science Cloud (October 2016) [18], which identifies the critical need for core data experts and data stewards in particular. The definition of data steward competences and training in these is an important component of the GO FAIR initiative [19, 20], as well as the Horizon 2020 EOSCPilot project activity [21, 8].

3. Method: NUFFIC data stewardship curriculum

NUFFIC (the Dutch organisation for internationalisation in education) Digital Innovations and Skills Hub (DISH) is a distance education programme sponsored by the Dutch Ministry of Foreign Affairs under the Orange Knowledge Program in conjunction with 12 partners from different countries in East Africa. The project targets learners with low opportunities, such as marginalised youth, including refugees and displaced persons from the Tigray region (Ethiopia), Garowe and Mogadishu (Somalia), Kassala and Khartoum (Sudan), Wau and Juba (South Sudan), and other conflict affected areas from East African region.

3.1 Course curriculum: Topics and description

Given the demography of the targeted learners, the training curriculum for this data stewardship specialisation programme is designed with the assumption that the students have little or no prior computer science skills. Thus, the training curriculum starts from a beginner's perspective and is divided into three courses of five to seven modules, with each course being a prerequisite for the next. The course details are given in tables 2–4.

Table 2. Course I – Introduction to Computer Science I – Communication and Information Technology (CS1)

Course	Module	Module title	Week	Topics	Module description
Course 1 Computer Science I (CS 1) – Communication and Information Technology (CS1)	CS1.1	Peace Building and Conflict Resolution Diplomacy (PBCRD)	Week 1	Introduction to peace, conflict and violence – Conflict analysis – Conflict resolution and peace – Understanding peace building – Peace building diplomacy – Peace building and conflict in the African context	Peace building and conflict resolution are key to building prosperous communities that are stable and at peace. This course focuses on how to resolve conflict and negotiate peacefully when conflict emerges in order to create stability in the community.
		Peace Communication, ICT and Media (PCICTM)		Introduction to peace building and conflict resolution – Conflict resolution and reconciliation – Communication for peace	Resolving conflict requires effective communication. In this course, students learn how to communicate effectively about peace. This includes learning how to write, engage with technology, and communicate with policymakers and the public.

				building – Media and peace building – Peace building communication and attitude change – Peace building process	
	CS1.2	Introduction to Digital Technology	Weeks 2, 3, 4	– Overview of computers and operating systems – The Internet, social media, email and web browsers – Cyber security and cloud computing – Digital literacy – creating, sharing and editing digital content using offline tools – Computer shortcuts	The first unit of this module introduces learners to the basic concepts and gives an overview of computers such as operating systems, the Internet, social media, cloud computing, and cyber security, among other things. Week 2 is on digital literacy, i.e., how to create and edit digital content using both offline and online tools. This includes how to create textual content using word processing software and how to create multimedia – graphics, videos, skits, and animation, etc.

				– Multimedia design – animation, videos and skits – Creating digital content using online tools, e.g., Google apps – Remote work tools and tips	Week 3 focuses on remote work tools such as Google Workspace and Google apps, like G.Slide. G.Doc., G.Form and G.Sheet, among others. This unit also provides tips on how to be productive and manage time in remote work situations.
	CS1.3	Introduction to Computer Networks	Weeks 5, 6, 7	– Introduction to computer networking – Layer architecture (OSI & TCP/IP) – Network hardware, software, and standardisation – Network medium, – IP addressing – Building small to medium level networks including	This module explores the concept of computer networks including their evolution, application, deployment, and standardisation. It focuses on how to set up a computer network and the definition and identification of different types of networks. In subsequent study units, network layer architecture is explored, with emphasis on how to identify different computer networking devices. Learners will be taught about the application of several network protocols, network software, network

				cabling – Configuring TCP/IP – Peer-to-peer networking – Sharing resources – Client-server networking	standards, data transmission media, IP addressing and network protocols.
	CS1.4	Business Administration, Entrepreneurship and Leadership	Week 8	Business and Business Strategies – Entrepreneur mindset, innovation and competitiveness – Business financing: use and sources – Business management practices and marketing	This module covers the introductory part of business strategies, businesses financing and costs, business communication and operating a business, and basic employability skills. It seeks to prepare young people to run their own businesses, be successful at work, and lead healthy and productive lives.

	CS1.5	Information Technology Support Management	Weeks 9, 10	– Basic operational tasks involved in using personal computers – Managing software applications: installing, updating and uninstalling a software application – Managing hardware: assembling or coupling a computer, installing network devices – Personal computer performance, maintenance and diagnostics – External system management tools – use of Team Viewer – Troubleshooting, and	This course is designed to introduce learners to the role of an IT support specialist in an organisation. It intends to prepare them for an entry level role with an IT help desk or support. Learners are introduced to how to identify and verify installed software, and how to update and/or uninstall computer software. Learners are introduced to the hardware components of a computer system. This is followed by an explanation of how the components are arranged and interact within the system. In this module, learners are also introduced to how to resolve slow boot times, device failures, and other machine issues using 'Task Manager', 'Device Manager', 'Windows Defender', and 'System Performance' tools.

				documentation – Ticketing system – Customer service in IT support role – Health and wellbeing of IT users – ITSM processes	Other aspects covered are the roles performed by information technology (IT) help desks such as ticketing systems and customer service, etc. Information technology service management (ITSM) processes and components are explored too.
	CS1.6	Information Technology Project Management	Week 11	– Overview of project management and related terms – Phases and processes of project management – Project methodologies – Importance and advantages of project management – Project management	

				standards ▪ PRINCE2 ▪ PIMBOK – Contemporary issues in project management – Human resources and staffing – IT project risk management – IT project cost management – Change management	

Table 3. Course II – Introduction to Computer Science II – Data Analytics (CS2)

Course	Module	Module title	Week	Topics	Module description
Course 2 Computer Science II – Data Analytics (CS2)	CS2.1	Introduction to Python Programming Language	Weeks 1, 2, 3	Programming in Python – Working with Jupyter Lab – Expression, data type, and variable assignment – String manipulation in Python – Basic string methods – String formatting Data structure in Python – List, tuple, set – Dictionary Python flow control – If else statement	The course context will be contextualised for business and agriculture, i.e., how Python programming can be used to build systems that make it easier for businesses and modern farms to operate efficiently. Examples will be based on different problems that occur within the daily operations of a small business and how to creatively solve these problems with programming. It will also cover examples of how programming can be applied in an agricultural context and give a big picture overview of how technology powered by Python has been able to improve agricultural systems.

				– For loop and while loop – Continue and break statements Python function – Function arguments – Anonymous function – Global and local variables – Python packages	At the end, students should understand how to frame business/process questions and how to solve these problems using Python.
	CS2 2	Introduction to Data Science I	Weeks 4, 5, 6, 7	Introduction to data and data science – Definition, types and sources of data – Lifecycle of data science – Python lists and NumPy arrays – NumPy use cases	The course context will be contextualised for business and agriculture, i.e., how Python programming can be used to build systems that make it easier for businesses and modern farms to operate efficiently. Examples will be based on different problems that occur within the daily operations of a small business and how to creatively solve these problems with programming.

				Introduction to data analysis with Pandas – Series and data frames – Importing and exporting dataset, data cleaning and pre-processing – Exploratory data analysis – Computing descriptive statistics – Combining and merging datasets Introduction to data visualisation – Data visualisation with Seaborn – Plotting continuous data	It will also cover examples of how programming can be applied within an agricultural context and give a big picture overview of how technology powered by Python has been able to improve agricultural systems. At the end, students should understand how to frame business/process questions and how to solve these problems using Python.

				and categorical data with Seaborn	
	CS2 3	Introduction to Business Intelligence	Week 8	– Overview of business intelligence – Data base management system – Data modelling with Entity Relationship diagram – Relational model – SQL – OLAP – Descriptive and predictive analytics – Visualisation – Dashboard creation – Power business intelligence	This module introduces trainees to business intelligence and basic SQL concepts. The module is aimed at equipping learners with the skills to mine data from a relational database, extract valuable information and create meaningful dashboards that can be used by business owners to make day-to-day decisions. In addition, the module will give an introduction to some of the open-source business intelligence software and how to quickly set up and use it.

	CS2 4	Tech Skills: Option Based Option 1: Digital Marketing Option 2: React (Web)	Weeks 9, 10, 11	**Digital marketing** – Search engine optimisation – Search engine marketing – Content marketing – Email marketing – Social media marketing – Web analytics **React (Web)** – React UI – Routing – Form helpers – Type checkers – State management – API clients – Testing static generators	The **digital marketing** option exposes learners to the inherent possibilities of a digital economy through digital marketing. The study sessions are designed to expose learners to the various aspects of digital marketing. These include search engine optimisation (SEO), keyword research, social media marketing, email marketing, content marketing and web analytics. At the end of the course learners are expected to be able to create effective integrated digital marketing strategies for businesses. The **React (Web)** option teaches learners how to use React JavaScript library to develop an interactive user interface on the website. Different components of React are explored, such as the React User Interface (UI), routing, form helpers, type checkers, state management, application programming interface (API) clients, and testing static generators.

		Option 3: Angular Option 4: Docker (needed for CS3)		**Angular** – Angular components – Angular templates – Angular directives – Angular depending injection – Angular routing; Angular forms; Angular HTTP – Angular animation – Angular best practices – Angular project and review **Docker** – Docker overview and environment setup – Docker Images – Docker Networks and	The **Angular** option introduces learners to Angular, a platform and framework for building single-page client applications using HTML and TypeScript. Learners are expected to use Angular features to create dynamic web applications. The **Docker** option teaches learners how to use, create, deploy, and run applications by using containers. They will be taught how to use Docker Images, Docker Networks, Docker Containers, Docker Compose and how to troubleshoot Docker.

		Option 5: React Native (mobile)		Containers – Docker Compose and – Troubleshooting Docker **React Native (Mobile)** – Introduction and environment setup – Knowledge in React – Android and IoS components	The **React Native** option teaches learners how to build mobile apps using JavaScript. Learners should be able to deploy simple mobile apps on Android and IoS platforms.

Table 4. Course III – FAIR Data Management (CS3)

Course	Module	Module title	Week	Topics	Module description
Course 3	CS3.1	Introduction to Data Science II	Weeks 1, 2, 3	Introduction to statistical thinking	This module builds on the previous introduction to data science. Learners will be taught how to make statistical inferences to draw clear conclusions from data. It also

Computer Science III – FAIR Data (CS 3)				– Population data vs sample data – Parameter vs statistics – Descriptive statistics – Scale of measurement – Inferential statistics Introduction to machine learning – Supervised learning methods – regression, classification – Unsupervised learning methods – clustering,	introduces machine learning, supervised and unsupervised learning. Learners should be able to create machine learning models and discover underlying clusters in a dataset.

				factor analysis, principal component analysis – Training and evaluating models – Regression metrics – Classification metrics	
	CS3.2	Regulatory Framework	Week 2	– Key concepts of data regulatory framework – General Data Protection Regulation (GDPR) and its principles – Data regulation specific to Sudan, South Sudan, Somalia and Ethiopia	The emergence of the Internet as a global telecommunications network has had a huge impact on how we view and apply data protection and regulations. Before the massive expansion of the Internet, data was of minor interest and did not generate significant global interest. This module provides participants with an understanding of what a regulatory framework is and what it is used for. Learners will understand general data protection principles,

				– FAIR Guidelines	national data regulations, and the basics of FAIR Guidelines, as well as be able to explain why we need FAIR Guidelines and the benefits for their country.
	CS3.3	FAIR Data Management	Week 3	– FAIR Guidelines – Data management plan – FAIR data management – Platforms for creating FAIR DMPs	This module exposes learners to the FAIR Guidelines and FAIR data management plans (DMPs). What kind of questions make a good DMP and which tools should be used to create a DMP? In addition, learners will be able to practise creating a FAIR DMP.
	CS3.4	FAIR Data Point Installation	Weeks 4, 5	– Metadata – Catalogues – Dataset – Distribution – Docker installation	This module describes FAIR Data Points (FDPs), their objectives and elements. The main purpose of this module is to illustrate how an FDP can be deployed on a local machine and provide detailed steps for a successful installation. It also aims to explain how to publish machine-actionable (meta)data to an FDP.

				– Deployment of FDPs – Open Refine – Open Refine installation – A semantic data model – FAIRification process for a dataset	Another objective of this module is to illustrate how non-FAIR data can be assigned machine-readable metadata to enable them to be discoverable by individuals and machines. In addition, leaners will be taught how to work with Open Refine and how to create RDF triplets. Learners will be presented with a simulated cancer dataset shown how to FAIRify it by building a semantic data model from the dataset.
	CS3.5	Semantic Data	Weeks 6, 7	– Semantic web and linked data – Semantics modelling – Ontologies – eCRF as a FAIR tool – CEDAR as a FAIR tool	The module introduces learners to semantic web and linked data, and shows them how to use eCRF and CEDAR to create and explore metadata and as a FAIR tool.

	CS3.6	FAIR Data for Health	Weeks 8, 9	– FDPs and their role in research and medicine – FAIR Guidelines in research and healthcare – FAIR Data Trains	In this module students will learn the importance of FAIR Guidelines in healthcare research including how FAIR Guidelines can facilitate knowledge discovery from health data and how linked health data drives research, better use and learning from data, and contributions to patient care.
	CS3.7	Internship	Weeks 10, 11	– On-the-job training for learners	This internship focuses on the knowledge gained in the previous modules and provides on-the-job training where students can gain experience and knowledge and learn how to apply their skills.

3.2 Mode and duration

This programme will span 36 weeks (12 weeks per course) with a total of 3 courses: Computer Science I (CS1), Computer Science II (CS2) and FAIR Management Principles (CS3). The core topics that pertain to data stewardship will be Introduction to Data Science I and II, Regulatory Framework, FAIR: Data Management, Data Point Installation, and Data for Health and Semantic Data.

The weekly activities summary for each course is as follows:

- Week 1 – Registration and Orientation
- Weeks 2 to 11 – Learning Activities and Interaction
- Week 12 – Examination

Considering the possible locality of the target participants and the limited infrastructure available in such places, the distance education model will take a blended learning approach, in which online learning is combined with face-to-face interaction at partner universities. Each participant is expected to devote a minimum of 12 hours a week, of which 4 hours is for self-study of the provided learning resources, 4 hours for online activities and interactions, and 4 hours for assessments and assignments.

3.3 Expected learning outcomes, activities and assessments

In addition to registration in week 1, learners are mandated to participate in two short modules: Peace Building and Conflict Resolution (to expose them to the skills needed to coexist and resolve conflicts in order to maintain peace in their communities) and Trauma and Mental Health (to help them to cope with the violence and trauma that they might have experienced in times past). To enable them to access the enormous opportunities inherent in the IT world, the modules on Digital Technologies, Computer Networks, IT Service Management, and Project Management will be designed to teach a wide range of skills on digital technology, contents creation, software installation, basic cyber security, IT productivity

tools, hardware coupling and troubleshooting, maintenance of local area networks, and other relevant topics. The learners will be facilitated via a learning management system using activities such as video conferencing, chats, online forums and so forth for interaction between teachers and learners and also for peer-to-peer communication. Practical sessions will be organised for students to demonstrate the skills acquired. Quizzes and assignments will also be given to gauge outcomes and these will be graded. It is expected that the course will not only qualify learners for IT-related jobs, but that they will also be capacitated to perform exceedingly well in other areas using the skills acquired.

The Computer Science Level 2 (CS2) modules were developed to teach the learners intermediate skills such as computer programming with Python, Introduction to Data Science, Business Intelligence, Digital Marketing, Front End Web Development with Angular, Docker and React (Web), and React (Native). Learners are expected to be able to write Python programs, as this is essential for data science. The Data Science and Business Intelligence modules will groom learners in the world of machine learning, data analytics and business analytics. Tech skills, which can provide a career path, are also taught. Marketable skills will be taught, such as skills in using digital marketing concepts to manage the digital platforms of business organisations and create digital advertising campaigns for small and medium scale businesses; skills in JavaScript to teach front end web development; and skills in Angular, Docker, and React to expose learners to software engineering. Similar activities of facilitation, engagement, practical and assessment as in CS1 will be introduced to teach, assess and encourage learners.

Computer Science Level 3 (CS3) modules are an extension of Computer Science Level 2 (CS2). Learners will be exposed to statistical thinking, supervised and unsupervised learning, and regression. Another interesting topic in FAIR Data is called FAIR Data Trains. The students will be exposed to FAIR Data for Health, which explores how linked health data drives research, better use and learning from data, and further contributions to patient care. In addition, learners will be taught the FAIR Guidelines for data

management as well as FAIR Data Point installation, Docker installation, the creation of machine-readable metadata, catalogues, datasets, and distribution. Students will also be shown how to FAIRify existing datasets using linked data and semantics modelling. The main objective of the course at this level is to understand the role of a data scientist in the industry and become acquainted with different data presentation formats, understand basic statistical thinking, understand machine learning techniques (such as supervised and unsupervised learning), understand basic concepts such as (sensitive) personal data and FAIR Guidelines, apply the FAIR Guidelines, know what data management and a data management plan (DMP) are, know the content elements that make up a DMP, be able to develop a FAIR DMP, and learn tools and techniques for the FAIRification of data.

4. Conclusion and further developments

This article reviewed existing curriculum, such as the EDISON framework, for Data Science Professionals. The presented profiles are defined based on the ESCO taxonomy and include the following groups: managers (DSP01–DSP03), professionals (DSP04–DSP09), professional data management/handling (DSP10–DSP13), professional (database) technical (DSP14–DSP16), professional technicians (DSP17–DSP19), and support and clerical workers (DSP20–DSP22). This framework defines data steward relevance and profile as DSP10. It is anticipated that all educational requirements of a data steward were met in the curriculum provided, which blends the skills involved in data stewardship and the FAIR Guidelines. A student that has satisfied all requirements will be certified as a FAIR data scientist and will be able to serve as both a FAIR data steward and analyst. In-depth and focused curricula development with diverse participants will build a core of FAIR data scientists. This will encourage the rapid adoption of FAIR Guidelines for data for research and development.

Authors' contributions

All of the authors contributed to the writing and provided critical feedback to help shape this article. **Francisca Oladipo** contributed to the conception and design of the work, as well as approval of the final the version to be published. **Sakinat Folorunso** contributed to the conception and design of the work, data collection, drafting of the article, and critical revision of the article. **Ezekiel Ogundepo** contributed to the drafting of the article, data collection, data analysis and interpretation, and critical revision of the article. **Obinna Osigwe** contributed to the drafting of the article and critical revision of the article. **Akindele Akinyinka** contributed to the conception or design of the work, drafting of the article and critical revision of the article.

Acknowledgements

We would also like to thank Misha Stocker for managing and coordinating this Special Issue (Volume 4) and Susan Sellars for copyediting and proofreading. We also acknowledge VODAN-Africa, the Philips Foundation, the Dutch Development Bank FMO, CORDAID, and the GO FAIR Foundation for supporting this research.

Conflict of interest

All of the authors declare that they have no competing interests.

Ethics statement

Tilburg University, Research Ethics and Data Management Committee of Tilburg School of Humanities and Digital Sciences REDC#2020/013, June 1, 2020-May 31, 2024 on Social Dynamics of Digital Innovation in remote non-western communities

Uganda National Council for Science and Technology, Reference IS18ES, July 23, 2019-July 23, 2023

References

[1] Desjardins, J.: How much data is generated each day? [Online]. World Economic Forum (2019). Available at: https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/. Accessed 8 March 2021

[2] NEJM Catalyst: Healthcare big data and the promise of value-based care. *NEJM Catalyst,* Massachusetts Medical Society (2018)

[3] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al.: The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3(1), 1–9, (2016)

[4] Mohanasundaram, K.: Curriculum design and development. *Journal of Applied and Advanced Research* 3(Suppl. 1), S4–S6 (2018)

[5] Diamond, C. C., Mostashari, F., Shirky, C.: Collecting and sharing data for population health: A new paradigm. *Health Affairs (Millwood)* 28(2), 454–466 (2009)

[6] Rosenbaum, S.: Data governance and stewardship: Designing data stewardship entities and advancing data access. *Health Service Research* 45(5)(2), 1442–1455 (2010)

[7] Wildgaard, L.: Reframing data stewardship educations in Denmark and abroad [Online]. Data Stewardship in Education Project, National Coordination of Data Steward Education in Denmark (2020)

[8] Dutch Techcentre for Life Sciences: Data stewards interest group [Online]. Available at: https://www.dtls.nl/community/interest-groups/data-stewards-interest-group/. Accessed 8 March 2021

[9] Whyte, A., de Vries, J., Thorat, R., Kuehn, E., Sipos, G., Cavalli, V., Kalaitzi, V., Ashley, K.: EOSCpilot deliverable D7.3: Skills and capability framework, EOSCpilot project deliverable [Online]. The European Open Science Cloud for Research Pilot Project (2018). Available at: https://eoscpilot.eu/sites/default/files/eoscpilot-d7.3.pdf. Accessed 8 March 2021

[10] EDISON Community: Data science competence framework [Online]. Available at: https://github.com/EDISONcommunity/EDSF/tree/master/data-science-competence-framework. Accessed 8 March 2021

[11] DigCurV: A curriculum framework for digital curation [Online]. Available at: https://www.digcurv.gla.ac.uk. Accessed 1 January 2021

[12] Wildgaard, L., Vlachos, E., Nondal, L., Larsen, A. V., Svendsen, M.: *National coordination of data steward education in Denmark: Final report to the National Forum for Research Data Management (DM Forum).* DM Forum (2020). DOI: 10.5282/zenodo.3609516

[13] CODATA: Call for applications – CODATA RDA Data Steward School 2019 – Deadline 1 June 2019 [Online]. Committee on Data, International Science Council. Available at: https://codata.org/initiatives/strategic-programme/research-data-science-summer-schools/data-steward-school-2019/. Accessed 8 March 2021

[14] EDISON Community: Data science body of knowledge [Online]. Available at: https://github.com/EDISONcommunity/EDSF/tree/master/data-science-body-of-knowledge. Accessed 8 March 2021

[15] EDISON Community: Data science model curriculum [Online]. Available at: https://github.com/EDISONcommunity/EDSF/tree/master/data-science-model-curriculum. Accessed 8 March 2021

[16] EDISON Community: Data science professional profiles [Online]. Available at: https://github.com/EDISONcommunity/EDSF/tree/master/data-science-professional-profile. Accessed 8 March 2021

[17] European Commission: European skills, competences, qualifications and occupations (ESCO) [Online]. Available at: https://ec.europa.eu/esco/portal/home. Accessed 8 March 2021

[18] European Commission: *Realising the European Open Science Cloud. First report and recommendations of the Commission High Level Expert Group on the European Open Science Cloud.* European Commission, Brussels (2016). Available at: https://ec.europa.eu/research/openscience/pdf/realising_the_european_open_scien_cloud_2016.pdf. Accessed 8 March 2021

[19] Dutch Techcentre for Life Sciences: GO FAIR initiative [Online]. Available at: https://www.dtls.nl/fair-data/go-fair/. Accessed 8 March 2021

[20] Dutch Techcentre for Life Sciences: FAIR Data Principles [Online]. Available at: https://www.dtls.nl/fair-data/fair-principles-explained/. Accessed 8 March 2021

[21] European Open Science Cloud for Research Pilot Project: Stewardship skills for open science and scholarship, EOSCpilot workpackage WP7 [Online]. Available at: https://www.eoscpilot.eu/themes/wp7-skills. Accessed 8 March 2021