Additional File 2 (Addfile2) – Description of existing data lifecycles

Data Lifecycle for HPC Scientific data perspective: The research teams of Oak Ridge National Laboratory, USA and Jet Propulsion Laboratory, California, USA described a data lifecycle that focuses on the support of big data lifecycle from a High-Performance Computing (HPC) Data perspective that is important for end-to-end scientific discovery process [114]. This data lifecycle consists of the following phases, data collection, data sharing, data storage, consume.

Data lifecycle for cloud automation tools: In 2016, Y. Demchenko introduced a data lifecycle for cloud automation tools [40]. This data lifecycle contains the following phases, data collection, data _filtering, data analysis, and data delivery.

Data Lifecycle for Telco networks data management: In 2010, U. Olsson, a representative of the Ericsson company, presented a data lifecycle intending to manage Telco networks data [115]. This data lifecycle consists of the following phases, data storage (Vault layer), retrieve (Engine layer), aggregate (Grid layer), analysis (analysis layer), share (exposure layer), use, and publish (presentation layer).

Energy big data lifecycle: In 2019, T. Ku et al. introduced an energy big data lifecycle for effective renewable energy information management [116]. This lifecycle consists of the following phases, creation, acquisition understanding, _filter, enrichment, modeling, visualization, and destroy.

Data lifecycle for the Tobacco industry: In 2019, Y. Jiang et al. described a data lifecycle for the Tobacco industry. This data lifecycle focuses on the data security protection aspects of data in the Tobacco industry [117]. This data lifecycle consists of the following phases, data acquisition, storage, access, protection, and destruction.

Data lifecycle for cloud computing: In 2012, D. Chen and H. Zhao introduced data lifecycle for cloud computing and detailed analysis about data security and privacy challenges related to cloud computing across all phases of data lifecycle [118]. This data lifecycle includes the following phases, generation, transfer, use, share, storage, archival, and destructor.

Data lifecycle for cloud data: In 2017, A. M. El-Zoghby, and A. M. Azer explained a data lifecycle of cloud data that focuses on the seven phases of the data lifecycle and data security and privacy aspects [119]. The data lifecycle consists of the following phases, generate data, transfer data, data usage, share data, store data, archive data, dispose of data.

Data lifecycle for IoTs: H. CAO et al. presented a data lifecycle for the IoTs in 2019, intending to integrate a variety of computational resources and analytical capabilities corresponding to this lifecycle [120]. They described the following phases of this data lifecycle, raw data generation, data aggregation, data filtering (data extraction), data transformation, data storage, data share, and data feedback.

Personal data lifecycle: In 2019, J. Peng et al. described a personal data lifecycle and proposed a method known as a differential attribute desensitization system (DADS, to protect personal information during the phases of the lifecycle [121]. The personal data lifecycle consists of data release, data filtering, data enrichment, data storage, data analysis, data protection, and data use.

Data lifecycle about the coal mine industry: In 2017, L. Xianglan mentioned a data lifecycle about the coal mine industry [122]. This data lifecycle consists of the following phases, data collection, data filtering, enrichment, data storage, data analysis, data protection, data visualization, and data sharing.

Data lifecycle for smart healthcare: In 2017, A. Solanas et al. presented a data lifecycle for smart healthcare [123]. This data lifecycle consists of the following phases, data gathering, data storage, share, analysis (knowledge acquisition), and data usage (knowledge application).

Data Lifecycle Model for NSF: The National Science Foundation (NSF), USA proposed a data lifecycle in 2020. The data lifecycle focuses on NFS large facilities regarding physical and natural science research [94]. This data lifecycle consists of the following phases data capture, data filtering and processing, data storage, data archive, and dissemination.

Data lifecycle cycle for smart cities: In 2017, M. K. Sutherland and M. E. Cook described a data lifecycle that is specifically centered on smart cities [13]. This data lifecycle consists of the following phases, data acquisition, data enrichment, data analysis, data quality, data preservation data use, re(use), share, and deletion.

Storage data lifecycle: R. Hasan et al. in 2005 proposed a data lifecycle as a process to create a threat model for the data storage systems [124]. This data lifecycle consists of the following phases, data creation, data reception, output preparation, data retrieval, data backup, and data deletion.

Research data lifecycle: H. khan et al. mentioned a research data lifecycle in 2018 [125]. This data lifecycle consists of the following phases, data planning, data acquisition, data curation, data access, and data preprocessing, analysis, visualisation, and data access.

A lifecycle for CENS Data: In 2007, a research team of Center for Embedded Networked Sensing (CENS) proposed a lifecycle for digital research libraries [126]. This data lifecycle consists of the following phases, experiment design, calibration, data filtering, data integration, share, data analysis, publication, use, and storage.

Data lifecycle for industry: In 2017, M.Y. Santos et al. mentioned a data lifecycle for the industry [127,128]. This data lifecycle consists of the following data collection, data preparation, data storage, data access, data analysis, and data visualization.

A lifecycle for big scholarly data: S. Khan et al. in 2017 described a lifecycle for big scholarly data [129]. This data lifecycle consists of data collection, data integration, data enrichment, data analysis, data storage, use, share, and data visualization.

A lifecycle for social and economic data: In 2017, D. Blazquez and J. Domenech described a data lifecycle with the aim to manage the processing and integration of social and economic data [46]. They proposed the following phases of the abovementioned lifecycle, studying and planning, data collection, data quality, data integration, data enrichment, analysis, publish, share, data storage, and data reuse.

Data lifecycle for manufacturing: In 2018, F. Tao et al. described a lifecycle for manufacturing data [130]. This data lifecycle consists of the following phases, data collection, transmission, storage, pre-processing, filtering, analysis, mining, visualization, and use.

Research data lifecycle: L. Lyon et al. mentioned a data lifecycle for research data in open science that focuses on research transparency [131]. This data lifecycle consists of the following phases, design & plan, collection, process, analysis, visualization, store, publish, and archive.

A lifecycle for big healthcare data: In 2018, H. Khalou_ et al. presented a big healthcare data lifecycle with the aim to provide recommendations to resolve data security threats and attacks in every phase of healthcare data lifecycle [132]. This data lifecycle consists of the following phases, data collection, data transformation, data modeling, and knowledge creation.

Data lifecycle: In 1993, A. V. Levitin and T. C. Redman mentioned a data lifecycle that consists of other sub-cycles like acquisition cycle and usage cycle [133]. They proposed the following phases of the data lifecycle data creation, storage, usage, analysis, quality, share, feedback, and destruction.

A lifecycle for environmental research data: In 2012, W. K. Michener et al. presented a data lifecycle for environmental research data [134]. This data lifecycle consists of the following phases, collect, integrate, assure, describe, store, analysis, access, archive, publish, and use.

A lifecycle for big data value creation: In 2020, Faroukhi et al. presented a data lifecycle for big data value creation. They focus on the big data value creation that facilitates data monetization during data value chains [26]. This lifecycle consists of the following phases, data generation, data acquisition, data pre-processing, storage, data analysis, data visualization, and data exposition.

A lifecycle for big data analytics for psychologists: In 2019, S. Muthy et al. presented a data lifecycle for big data analytics solutions that help psychologists in order to analyse social media data [135]. This data lifecycle consists of the following phases, domain study, planning, exploration, data preparation (integration, enrichment), modeling (analysis), implementation, visualization, archive, result synthesis, and operationalization.

The information pyramid of Reynolds and Busby lifecycle: - J. G. Creuseveau, Les mentioned this lifecycle in 2014, and the same is referred to in [17]. This lifecycle focuses on transforming a company raw data into information in order to constitute knowledge of the company. The lifecycle consists of the following phases: collection, integration, Analysis, and Publication.

Yuri Demchenko data lifecycle: Yuri Demchenko introduced this data lifecycle in 2014. The Demchenko data lifecycle mainly focuses on the big data [10]. This model comprises complicated and iterative processes to gather and mine (big) data and transform it into valuable information. This lifecycle consists of the following phases: data collection, data storage, data filtering/Enrichment/classification, data analysis, data visualization, data archive, data usage.

Data lifecycle: This data lifecycle is defined in 2018. This data lifecycle considers the destroy phase based on respective PAs practices [58]. The destroy phase can be used in terms of an end of life or end of the usefulness of data. This data lifecycle consists of the following phases: plan, collect, store, use, access, process, publish, and destroy.

SCC-data lifecycle: A Smart City Comprehensive (SCC) data lifecycle model, published in 2016, is an extension of the above-mentioned COSA-DLC. This lifecycle is primarily focused on the management of smart city data with Fog to Cloud resources management [44, 54]. SCC-DLC consists of three blocks and their respective phases: Data Acquisition (Data Collection, Data Filtering, Data Quality and Data Description), Data Processing (Data Process, and Data Analysis) and Data Preservation (Data Classification, and Data Archive and Data Dissemination).

Data value cycle: In 2014, OECD defined a data value cycle. This lifecycle focuses on data value creation through Data-driven innovation [136]. This lifecycle consists of the following phases, datafication and data collection, store, data analytics, use, knowledge base, achieve, share, and decision making.

Lifecycle in databases: H. Kilov introduced an elementary model in 1990. This lifecycle focuses on database operations such as read, and update. This lifecycle consists of these traditional phases, which are not used in the recently proposed data lifecycle models [137]. This lifecycle consists of the following phases: create, read, update, delete, store, and use.

Knowledge process-lifecycle: S. Staab et al. introduced knowledge process-lifecycle in 2001. This lifecycle focuses on knowledge processes [138]. This lifecycle consists of the following phases, Create, Capture, Retrieve/Access, and use. After the implementation of a Knowledge management system, knowledge processes principally revolve around the aforesaid phases.

Capability Maturity Model (CMM) for Scientific Data Management process life-cycle: In 2012, K. Crowston and J. Qin introduced the Capability Maturity Model (CMM) for Scientific Data Management (SDM) process lifecycle. It covers all aspects of effective scientific data management practices [63]. This data lifecycle consists of the following phases: Data acquisition, processing and quality assurance (capture, acquire, process, prepare, validate, and audit data), Data description and representation, Data dissemination (distribute date, provide access), and Repository services/preservation (archive, curate, migration, deliver).

Data lifecycle: Qingqing et al. presented a data lifecycle centered on data privacy in a cloud computing environment [139]. They explained the following three phases of their proposed lifecycle, data collection, processing, and privacy management.

PII lifecycle: In 2015, M. Michota and S. Katsikas defined the Personal Identifiable Information (PII) lifecycle. This lifecycle intends to keep a balance between sharing and the protection of personal data [17]. This lifecycle contains the "deletion/destroy" phase. This lifecycle consists of the following phases, collection, access, use, modification, storage (user can manage or delete his/her information), transfer (internal sharing, external distribution/publication), maintenance (user can manage or delete information).

Data lifecycle: In 2013, X. Cheng et al. described a data lifecycle that centers on data management in Virtual DataSpace [140]. The lifecycle consists of the following phases, data produce, data integration, data evaluation, use, feedback, and data demise.

Data Lifecycle Process Model: The research team of Nanjing Telecommunication Technology Institute, China, presented a data lifecycle process model in 2010 [141]. This data lifecycle consists of mainly three main process

phases that include design, manufacturing, and application. This lifecycle contains the following key phases, data collection, integration, filtering, analysis, quality, storage, and data archive.

Data lifecycle for cloud computing security: In 2010, X. Yu and Q. Wen described the data lifecycle for cloud data security and suggested a design process of cloud data security [142]. This data lifecycle consists of the following phases, create, store, use, share, archive, and destruct.

IHRB Data Lifecycle: This data lifecycle is proposed by the Institute for Human Rights and Business - IHRB, the UK in 2016. This lifecycle is centered on individual personal data, particularly its usage for commercial purposes and present (big) data analytical model [143]. This model consists of the following phases, Collection, Datafication, Storage, Processing, and Usage.

Security data lifecycle: The security data lifecycle is presented by Huaglory Tian-eld, Glasgow Caledonian University, the UK, in 2016 [144]. This data lifecycle is primarily concerned with data security and consists of the following phases, data collection and filtering (data preprocessing), data storage, analysis, and visualization and these phases are aligned with the process of Cyber Security Situational Awareness (CSSA).

Knowledge lifecycle for e-learning: The Knowledge lifecycle for e-learning is pioneered by the University of Southampton's research team, the UK, in 2006. This lifecycle is a semantically driven knowledge management lifecycle [145]. These model phases include knowledge acquisition, knowledge modeling, knowledge reuse, and knowledge maintenance.

IRIS Data Lifecycle: The Systems Integration and Re-engineering (IRIS), a Spanish research group, founded this lifecycle in 2019 [62]. This data lifecycle concentrates on the actions concerning big data. This data lifecycle consists of the following phases, content (Data collection, Data Generation, Data Obtain), Acquisition (Data Integration, Data Ingestion), Enhancement (analysis, modeling, training, and validation of the models, data storage), Inquiry (access to data and find data), sharing, and Visualization (presentation and visualization).

Scientific Data Lifecycle Management (SDLM) Model: Y. Demchenko and P. Memrey presented a Scienti_c Data Lifecycle Management (SDLM) Model in 2013 [146]. It consists of the following phases, planning, data collection, analysis (processing), share/publishing, use, feedback archive, and destruction.

Big data (privacy) lifecycle: In 2016, A. Mehmood et al. introduced big data (privacy) lifecycle [95]. This study explains state-of-the-art privacy-preserving mechanisms in each phase of the big data life cycle. This data lifecycle is reused in [96] and named the life cycle of Big data analytics. This lifecycle consists of the following phases, data generation, storage, processing, and protection.

DataONE lifecycle model: The DataONE lifecycle model is used for scientific research data. This lifecycle is introduced in 2012 by Michener, W.K et al. [15]. The DataONE lifecycle model phases include plan, collect, assure (format, coding, quality assurance, quality control), describe, preserve, discover, integrate (transform), and analyze.

DDI (Data Documentation Initiative) Lifecycle: This lifecycle is initially introduced by Structural Reform Group in 2004 and subsequently refined by X. Ma et al. in 2014. The Data Documentation Initiative (DDI) lifecycle is

used for research data, including social science data [14, 150]. The DDI lifecycle consists of the following phases, concept study, data collection, analysis (data Processing), use, data archive, store, data distribution (share), and repurposing.

CIGREF data lifecycle: This lifecycle is proposed by CIGREF in 2014 and is referred in [17]. This data lifecycle focuses on data quality control. This data lifecycle consists of the following phases, data quality control, acquisition, transformation, analysis, and data value. Different data lifecycles are modeled from the CIGREF framework to enable their lifecycle implementation within businesses.

Australian National Data Service (ANDS) Data Sharing Verbs: This data lifecycle is introduced by the team of Australian National Data Services (ANDS) in 2009 [85] and focuses on research data and its sharing and reuse amongst the research community. This data lifecycle consists of the following phases, create (generation, aggregation, collation, transformation, curation), store (preservation, archive), describe (The generation or acquisition of metadata supporting the storage, discovery, access, and exploitation of data), identify, register, discover, access, share, and exploit (re-use).

Research Data Lifecycle (UKDA): In 2010, UK based entity "UK DATA Archive - UKDA" founded the research data lifecycle. This data lifecycle is specific to a scientific research field [48]. This lifecycle consists of the following phases, create, process, analysis, store, archive, access, and reuse.

Web Data Lifecycle: The web data lifecycle is described by Lóscio BF et al. in 2015. This lifecycle emphasizes on web data [93, 92]. This data lifecycle consists of the following phases, planning, creation, publication, access, consumption, share, feedback, refinement, process, publish and destroy.

Data Digital Curation Center (DCC) Lifecycle: Digital Curation Centre (DDC), UK research team founded the DDC lifecycle in 2008. This lifecycle model offers a graphical, high-level summary of the phases essential for effective curation and preservation of data [90, 147]. The DDC lifecycle consists of the following phases, create or receive (data generation and data collection from other sources), appraise, and select (evaluate, select, some cases sent data to other custodians), ingest (archive, sent to a repository, data center, share), preservation action (quality control, cataloging, classifying, generating fixed data), store, access, use & re-use, and transform.

OGD lifecycle: In 2015, J. Attard et al. mentioned an Open Government Data lifecycle [52]. This data lifecycle consists of the following phases, data creation, data selection, data analysis, data curation, data publishing, data discovery, data exploration, data storage, and data exploitation.

An IoT Data lifecycle: In 2016, S. Ahn et al. described a data lifecycle for IoT [148]. This data lifecycle consists of the following phases, data source profiling, data collecting, data tagging, context awareness, context sharing, data usage, and data post-processing.

Open Government Data Lifecycle: Zuiderwijk, A. and Janssen, M. et al., introduced the open government data lifecycle in 2015. This data lifecycle highlights a high-level representation of open data processes of the public sector [185]. This data lifecycle consists of the following phases, creation of data (producing, collecting, storing), a publication of data, finding data, analysis of data, data processing (enhancing, visualizing, linking, combing), use, share, and feedback on open data.

Data Lifecycle for IoTs data: In 2016, Kiourtis A. et al. mentioned a data lifecycle that focuses on data management and data exchange between IoTs [149]. This data lifecycle model consists of the following phases, conceptualize, create, appraise, and select, ingest, preserve, store, access, and transform.

Information Lifecycle: The Information lifecycle, introduced by L. LIN et al. in 2014, focuses on data security in the cloud computing environment throughout the lifecycle [88]. This lifecycle consists of the following phases: Data Generation, Data Storage, Data Access, Data Reuse, Data Archiving, Data Protection, and Data Disposal.

COSA-DLC: In 2016, the Advanced Network Architectures Lab (CRAAX), Universitat Politècnica de Catalunya (UPC, BarcelonaTech), Barcelona, Spain, described a comprehensive scenario agnostic data lifecycle (COSA-DLC) [42, 43] . This DLC consists of the following three blocks and their respective phases: Data Acquisition (Data Collection, Data Filtering, Data Quality and Data Description), Data Processing (Data Process, Data Quality, and Data Analysis), and Data Preservation (Data Classification, Data Quality, Data Archive and Data Dissemination(share).

USGS data lifecycle: U.S. Geological Survey (USGS) team developed a Science Data Lifecycle Model (SDLM) in 2013 [50]. This data lifecycle is specific to the scientific research field. This data lifecycle comprises of the following phases, Plan, Acquire, Process, Analyze, store, archive, publish (share), describe (data, metadata, handle quality, documentation), use, secure, and backup.

APDLM - Abstract Personal Data Lifecycle Model: In 2017, researchers from the University of Oxford, UK, reuse and extend the Abstract Data Lifecycle Model to support the management and traceability of the flow of personal data [48]. This model consists of the following phases, conceptual modeling, initiation, collection, retention, access, review, disclosure, share, usage, and destruction.

<><><><><><><><><>