IJDC | *General Article*

Research Data Lifecycle (RDLC):

An Investigation into the Disciplinary Focus, Intended Use Cases, Creator Backgrounds, Stages and Shapes of RDLC Models

Jie Jiang Danielle Maurici-Pollock

Rong Tang

Simmons University

Abstract

We report the results of a study examining 78 Research and Data Lifecycle (RDLC) models located in a review of the literature. Through synthesis-analysis and the nominal group technique, we investigated the RDLC models from the point of view of their disciplinary focus, intended use cases, and model creators, as well as the specific stages and shapes. Our study revealed that the majority of the disciplinary focus for the models was generic, science, or multi-disciplinary. Models originating in the social sciences and humanities are less common. The intended use cases varied in a wide spectrum, with a total of 34 different scenarios. The creators and authors of the RDLC models came from more than 20 countries with the majority of the models created as a result of collaboration within or across different organizations. Our stage and shape analysis also outlined key characteristics of the RDLC models by showing the commonalities and variations of named stages and varying structures of the models. As one of the first empirical investigations examining the deep substance of the RDLC models, our study provides significant insights into the context and setting where the models were developed, as well as the details with regard to the stages and shapes, and thereby identified gaps that may impact the use and value of the models. As such, our study establishes a foundation for further studies on the practical utilization of the RDLC models in LIS practice and education.

Submitted 23 November 2022 ~ *Revision submitted* 9 April 2024 ~ *Accepted* 27 January 2025

Correspondence should be addressed to Jie Jang. Email: jie.jiang@simmons.edu

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

Introduction

With increasing requirements for data management planning and the growing demand for research data management services, a variety of key stakeholders, including researchers, libraries, sponsored programs, data centers, and research organizations, have joined the efforts to develop Research Data Lifecycle (RDLC) models to model the process and the workflow associated with research data generation and management (Fan, 2019). Over the years, there have been research investigations into published RDLC models. As an example, Ball (2012) reviewed eight RDLC models, and highlighted the importance of RDLC models in guiding data management activities. Meanwhile, the Committee on Earth Observation Satellites (CEOS) Working Group on Information Systems and Services (2012), in an effort to compile a list of data lifecycle models and concepts, documented 55 published models. Pouchard (2015) evaluated two lifecycle models under the context of big data projects, and then proposed a Big Data Lifecycle model specifically targeting big data management activities. In the same vein, Cox and Tam (2018) conducted a systematic comparison of nine popular lifecycle models for research data management from the aspects of Scope and Points of View; Elements and Processes, and Visualisation. The authors found that while lifecycle models were effective in capturing key stages and features of research, they often obscured the complexity of the research process and frequently represented it as uni-directional and occurring in a closed system.

The current studies on existing RDLC models typically involve small sample sizes and have varying levels of depth on the specifics of individual models. To address these gaps, the present study began with a comprehensive literature review to collect/extract a large sample of RDLC models. Furthermore, our study explores the key characteristics of these models by studying the model's disciplinary focus, intended use cases, and model creators, as well as the specific stages and shapes. For our investigation, we decided to include both research lifecycle models that contained stages of data lifecycle as well as pure data lifecycle models for the purpose of understanding data practice both within and independently from research activities.

Literature Review

Research and Data Lifecycle Models (RDLC)

According to Lyon et al. (2020), a Research Lifecycle model (RLC) describes and identifies the different phases that a research project goes through from the beginning to the end. With their graphical representation, RDLC models have been applied as "a grounding framework" for researchers of diverse disciplines to plan for and organize a series of activities associated with the successful completion of their research projects. Carlson (2014) further explains the value of using Data Lifecycle models (DLC) to define and represent the flow of research data: "Life cycle models help to define and illustrate these complex processes visually, making it easier to identify the component parts or distinct stages of the research data. By breaking down the process of the data coming into being, growing, and evolving as it is applied toward fulfilling its purpose into interrelated stages, the specific needs of the researcher are more readily identified" (p. 63).

With the progress of research projects, research data are generated and collected as a necessary component of many types of research endeavours. It is critical for the research data to be retrievable and reusable for the sake of validating or reproducing the research output, and for scholarly communications in the long run (Faundeen et al., 2014). Carlson (2014) claims that, "The premise behind the application of life cycle models to research data is that data also progresses through a life cycle of sorts. From its inception to its use and completion, research data will likely undergo multiple transformations in its format, application, use, and perhaps even its purpose" (p. 63). In this sense, a RDLC model "offers a high-level overview of the individual actions, operations, or processes that must be undertaken at different stages"

(Faundeen et al., 2014, p. 2). Faundeen et al. (2014) further indicate that as a visual tool, RDLC models can "assist scientists in anticipating and planning for specific actions that need to be taken at each stage to manage the data, and thus help to ensure timely, comprehensive, and secure approaches to data curation" (p. 2).

The value of a RDLC model is further specified by Carlson (2014) from the perspective of data services: "Through identifying and naming the transformations that data will undergo as stages in a larger life cycle, organizations can better target their services toward addressing real-world situations and needs of the communities they seek to serve. The utilization of life cycle models can provide a useful framework to articulate these stages and to contextualize and communicate what kinds of data services could be provided to whom and when" (p. 63).

Visual Characteristics of RDLC Models

In their investigation into the shape and direction of the visual forms of nine RDLC models, Cox and Tam (2018) use "uni-/multi-directionality" to describe the flow of activities in a single or multiple directions. According to Cox and Tam (2018), researchers tend to be driven to complete a research project within a fixed time frame, in which a linear workflow might be more representative of the real-world progression of a research project, phase by phase. Nevertheless, from the standpoint of open science practices, a research project should not be closed without ensuring its outputs to be shareable and reusable. Consequently, a circular model emphasizing on data reuse at the end might better represent a research process that is iterative. The same thing goes for the visualization of the directions of the RDLC models in that the sequence of research activities tends to be "based on repeating steps a number of times or going back and forwards between different stages" (p.152). From that perspective, models taken in linear and unidirectional shapes "cuts against what we know of the typical character of research" (Cox & Tam, 2018, p.152). Further, Cox and Tam (2018) point out the problems in the design and documentation of the RDLC models: "A critical point is where the lifecycle model turns full circle. In a few cases, we have seen that the notion that there is a restart of the cycle has not really been explained. The real nature of how a process might be iterated is not developed" (p.153). Consequently, despite the proliferation of RDLC models, the models have limits in their explanatory power.

Research Questions

Our study aims to pursue answers to the following research questions:

RQ1. Among the RDLC Models, what specific disciplinary focuses are there?
RQ2. What are the possible use cases of the RDLC models intended by the creators?
RQ3. What are RDLC creators' organizational affiliations, countries, and partnerships?
RQ4. What specific stages do the RDLC models include?
RQ5. What specific shapes or visual design do RDLC models take on?

Methods

Literature Search and Synthesis Analysis

Our study employed a synthesis analysis of published RDLC models that relate to research and data activities. Three rounds of literature search were performed between August and November of 2020 on multiple databases, including EBSCO Academic Search Ultimate, ACM Digital Library, SCOPUS, LISA, and Google Scholar to retrieve relevant literature. We located more than 100 published works including at least one RDLC model. An inventory was created to record the relevant literature and RDLC model information. A total of 78 RDLC models in our inventory constituted the sample of our study. The comprehensive list of RDLC models sampled in this study along with their source information are provided in Appendix I.

Data Coding and Analysis

Nominal Group Technique (NGT)

We applied *Nominal Group Technique* (NGT) (American Society for Quality, 2022) while coding the disciplinary focus and the intended use case for each model. NGT is defined as a structured method for group brainstorming that encourages contributions from each member of the group "to achieve consensus" (Harb et al., 2021, p. 140). All three co-authors of this paper were involved in the group coding process. This process began with individual research team members (i.e. the three co-authors) proposing a coding label for a model, followed by open discussions of the proposed coding. The team then collectively finalized the coding based on group agreement. This approach ensured all members' perspectives were considered and informed the process. Occasional difficulties occurred in reaching consensus, which we addressed through additional discussions and, when necessary, majority voting.

Quantitative Coding

Quantitative data coding and synthesis analysis were performed later on the inventory for the information about each RDLC model, including title, developer, model origin, disciplinary focus, intended use cases, stages (e.g., stage count and name), and visualization features including the shape, directions, and the starting point. To standardize our coding process, the three co-authors of this paper collaboratively coded a subset of the sampled lifecycle models. This initial joint coding session was crucial for setting clear standards and ensuring consistency in the individual coding phases that followed. Subsequently, each team member independently coded a portion of the sample, applying these agreed-upon standards. Upon completion of this coding, the team reconvened to review and discuss any questions or uncertainties that arose during the process. The combination of independent coding followed by collective discussions allowed the authors of this paper to leverage the diverse perspectives within the team while ensuring a cohesive and reliable analysis. See Appendix II for our quantitative coding schema.

Furthermore, it should be noted that our analysis was not singularly based on the components or visual structure of the RDLC models' images, but more importantly, it was based on a combination of model authors' own description, explanation, and documentation of the model, together with the visual characteristics of the models themselves. If a model lacked the authors' own explanation or visual representation, we excluded the model from our study sample. For all the model samples in our study, we had both authors' own documentation and the visual of the model.

Inclusion Criteria

As we selected published RDLC models our criteria of inclusion were as follows:

1. The model needed to be described or introduced in the literature as a lifecycle model relevant to research and/or data process.

2. The model had to have a visual representation of some sort, illustrating the processes or key concepts/activities in the lifecycle.

3. The labels/texts within the model needed to be in English.

As such, our study sample included lifecycle models which exclusively attend to a granular aspect of research or data, such as the UK Data Archive Model, which focuses on digital data archiving and discovering processes; and the Grant Lifecycle Model, developed by Grants.gov which explains in detail the series of actions, from the application and management, to the close of a research grant. [1] We encountered notable challenges during the sampling process, such as the obsolescence of model links, and the absence of visual representations for some models. These issues prevented us from accessing certain resources through their original online sources. To address the issue of link obsolescence, we utilized the Wayback Machine digital web archive, to retrieve the necessary information. Our research included six models recovered through the Wayback Machine: University of Oxford Research Data Management Chart, The JISC Research Model, IWGDD's Digital Data Lifecycle Model, Loughborough University Library Research Lifecycle Model, the Records Model, and the Generic Science Model. Those models that have no visual representation were excluded from our research sample.

Results

Overview

Of the 78 models that we examined, the majority focus on "data" (n=57, 73.08%), followed by "research" (n=17, 21.79%). Two models (n=2, 2.56%) cover both research and data, and another two (n=2, 2.56%) center around grants.

Disciplinary Focus

The research team coded the disciplinary focus of each model by closely examining the authors/creators' own description in their publications about the context where the RDLC was used or could be applied to. As such, the granularity may vary by models. We acknowledge this is a limitation of our study. Of 78 RDLC models, we identified 19 individual disciplines of origin, together with categories of "generic" and "multi-discipline." The majority of the RDLC models can be characterized as "generic" (n=55, 70.51%), not originating from or intended to be used within a specific field, whereas six models (7.69%) appear to be "multi-discipline." Among individual disciplines, the most frequent discipline was Science (n=19, 24.36%), followed by Business, Environmental Science, Geospatial, and Nanotechnology and Material Sciences; all had four RDLC models (5.13%) each. Table 1 displays those disciplines that originated more than one RDLC model.

Table 1. Distribution of RDLC models' disciplinary focus.

Discipline	Number of RDLCs	Percentage
Generic	55	70.51%
Science	19	24.36%
Multidiscipline	6	7.69%
Business	4	5.13%
Environmental Science	4	5.13%
Geospatial	4	5.13%
NanoTechnology & Material Science	4	5.13%
Chemistry	3	3.85%

[1] https://www.grants.gov/

Discipline	Number of RDLCs	Percentage
Engineering (Civic Eng., Industry Eng.)	3	3.85%
BioTechnology	3	3.85%
AstroPhysics	2	2.56%
Humanities	2	2.56%
Health & Medicine	2	2.56%
Public administration, Policy, & Gov.	2	2.56%
Social Sciences	2	2.56

To enable a macro-level perspective of domain patterns across the RDLC models we grouped them into areas of STEM (science, technology, engineering, medicine, and math), Social Sciences, and Humanities. For details about the disciplines that were classified into each domain, please refer to the Discipline Focus section of Appendix II. Figure 1 illustrates the domain distribution of the RDLC models. In addition to the "generic" (n=55, 70.51%) and "multi-discipline" (n=6, 7.69%) categories, the predominate domain origin was STEM (n=49, 62.82%), followed by Social Sciences (n=9, 11.54%), while Humanities originated relatively few RDLC models (n=2, 2.56%).

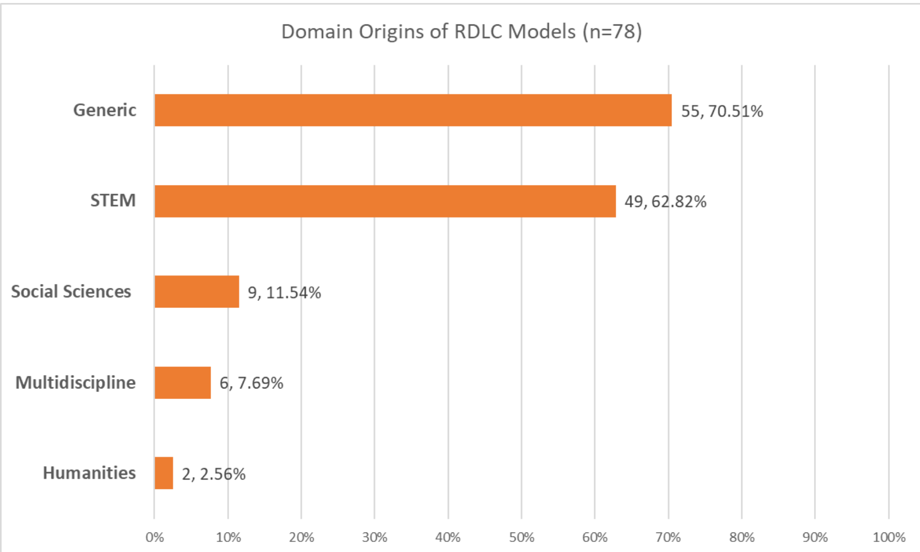

Figure 1. Domain origins of RDLC models (n=78).

Intended Use Cases

Of 78 RDLC models, we identified 33 different use cases. By intended use cases, we meant the specific activities, settings, or types of data for which the RDLC models were designed and intended to inform data practices by the original author or from the source. Among 33 use cases, 19 cases were covered by more than 1 model. Table 2 includes the 19 repeated use cases. The most common use case was "Big Data" (n= 13, 16.67%), typically in the case of large facilities. "Data Preservation and Archival Storage" was the use case for 12 models (15.38%). Other common use cases were "Digital Data" (n=11, 14.10%), "Government, Public

Administration & Policy" (n=9, 11.54%), and "LIS Practice" (n=8, 10.26%). The 14 unique use cases included "Digital Humanitarianism Practice," "Smart City Data," "Disaster Relief & Aid," "National Population Data," "Product Development & Utilization," and "Linked Data," among others.

Table 2. Distribution of RDLC models' intended use cases.

Use cases	Number of RDLCs	Percentage
Big Data, Large facility	13	16.67%
Data Preservation & Archival Storage	12	15.38%
Digital Data	11	14.10%
Gov, Public administration & Policy	9	11.54%
LIS Practice	8	10.26%
Open Data/Science	6	7.69%
Data Security, back-up/Privacy	5	6.41%
All-encompassing	5	6.41%
DM Practices & Assessment	4	5.13%
Data Value Chain	4	5.13%
Data Quality Assurance & Control	4	5.13%
Grant-funded research	4	5.13%
Institutional Practice	3	3.85%
e-Science	3	3.85%
Cloud computing/service	2	2.56%
Commercial & Business Data	2	2.56%
Technology Research	2	2.56%
Metadata	2	2.56%
Large-scale Ecological research	2	2.56%

RDLC Model Creator Affiliations

Types of Organizations

The creators of the RDLC models were affiliated primarily with four types of organizations. The most common was academic institutions (n=45, 57.69%), followed by government agencies (n=24, 30.77%), and industry (n=9, 11.54%). The fewest authors were from data-related organizations (n=4, 5.13%).

Countries

The 78 RDLC models were developed by people who work at organizations from a variety of countries or geographical areas across the globe. The model authors were from a total of 21 countries, with 10 countries being the home of more than one model, and 11 having one model each. The top country where more than half of the model authors originated was the USA (n=40, 51.28%), followed by the UK (n=15, 19.23%), and Germany (n=7, 8.97%). South Korea and Spain have authored three models (3.85%) each. Two organizations that were responsible for a RDLC model were internationally based. Two RDLC models (2.56%) each originated from authors in Canada, China, Netherlands, and New Zealand. Table 3 lists the countries affiliated with more than one RDLC model. Examples of single model originating countries include Malaysia, Pakistan, Saudi Arabia, and France, among others.

Table 3. Country affiliation of RDLC model creators

Country Affiliation	Number of RDLCs	Percentage
USA	40	51.28%
UK	15	19.23%
Germany	7	8.97%
South Korea	3	3.85%
Spain	3	3.85%
Canada	2	2.56%
China	2	2.56%
International	2	2.56%
Netherlands	2	2.56%
New Zealand	2	2.56%

Collaborations

Of 78 RDLC models, 10 (12.82%) were created by a single named individual, 33 (42.31%) were created by more than one named individual. Thirty-four models (43.59%) were credited to an organization and two (2.56%) were credited to more than one organization. Table 4 shows the count and proportions of the single or multiple authorship or by single or multiple organizations.

Table 4. Types of authorship and affiliated organizations.

Types of authorships	Number of RDLCs	Percentage
Single Author	10	12.82%
Multi-Author	33	42.31%
Single Org	35	43.59%
Multi-Org	2	2.56%

We also examined the institutional affiliations of individual creators for evidence of cross-organizational collaborations. Among 78 RDLC models, 49 (62.84%) were created by people from a single institution, whereas 29 (37.18%) were developed collaboratively by authors from multiple institutions. Nine (11.54%) of these collaborations were by creators from multiple countries.

Stage Analysis

For this part of the analysis, we examined the specific stage labels along with the total number of stages or entities within a given model. Across the 78 models analyzed, we identified a total of 130 distinct stages/entities. Of these, 75 were common to multiple models, while 55 were unique to individual models. The most prevalent stages found include "Collect/Acquire" (n=31, 39.74%), "Analyze/Analysis" (n=30, 38.46%), "Plan/Conceptualize/Design" (n=29, 37.18%), "Preserve/Preservation" (n=22, 28.21%), "Store/Databank" (n=22, 28.21%), "Archive" (n=18, 23.08%), "Discover/Discovery" (n=18, 23.08%), and "Process" (n=17, 21.79%). Table 5 lists stages that occurred repeatedly among more than 10% of the models. Examples of stages occurring in less than 10% of the models are "Maintain" (n=7, 8.97%), "Transform" (n=5, 6.41%), "Data Output/Result" (n=4, 5.13%), "Documenting" (n=3, 3.85%), "Deliver" (n=2, 2.56%), and "Secure Funding" (n=2, 2.56%). Examples of the 55 stages appearing in one model are "Clean/Cleaning", "Data Harmonisation", "Inventory", "Organizational Structure", "Repackage", and "Scholarly Communication."

We then counted the number of stages. All 78 models had multiple stages, ranging from as few as 3 to as many as 13. Both the mode and median number of stages were 6, and the average number of stages was 6.51. Twenty models (25.64%) have six stages, while six models have more than 9 stages, with four models having 13 stages.

Table 5. Distribution of RDLC models' stage labels.

Stage labels	Number of RDLCs	Percentage
Collect/Acquire	31	39.74%
Analyze/Analysis	30	38.46%
Plan/Conceptualize/Design	29	37.18%
Store/Databank	22	28.21%
Preserve/Preservation	22	28.21%
Archive	18	23.08%
Discover/Discovery	18	23.08%
Process	17	21.79%
Create	15	19.23%
Publish/Publication	15	19.23%
Use	14	17.95%
Dispose/Destroy/Delete/Discard	14	17.95%
Access	13	16.67%

Exploit/Re-use	12	15.38%
Assure (data quality)	11	14.10%
Share	10	12.82%
Disseminate	10	12.82%
Idea	10	12.82%
Appraise/Evaluate/Assess	9	11.54%
Describe (metadata)	9	11.54%
Experiment/Test/ Research Project	9	11.54%

Shape Analysis

We then analyzed the shapes of the models. We documented whether the model has a clear starting point, whether there is a consistent QA/QC (quality assurance or quality control) element across all stages or multiple stages, and whether there is an indication of the direction(s) of the flow. In coding the QA/QC element, if a model incorporates the element only in several but not all stages, it was coded as QA/QC being present in focused stages. For example, the DCC model does not have an explicit starting point, does not have a QA/QC element, and is circular and multi-directional. A mere three models (3.85%) consistently have a QA/QC factor across all stages.

In terms of flow direction, some models are either multidirectional or unidirectional, while others lack a clear directional indication. Out of the 78 models, only 38 (48.72%) have an explicit starting point.

In our analysis of the models' shapes, we differentiated them based on their layout in the following three categories: *circular*, *linear*, or *special shapes.* Adopting this categorization, the research team came to a consensus on the criteria that circular models organize their stages in a circular format, often suggesting a recurring or cyclical process. Linear models have their stages presented in a sequential manner, indicating a start-to-finish flow, while special-shaped models possess some unique configurations, such as models that represent stages using a staircase, tables, or other non-linear, non-circular layouts. Notably, most circular shaped models lack a clear starting stage, unless it is marked on the graphics of specific stage numbers.

Model Directions

More than half (n=48, 61.54%) of the models are unidirectional, which means the progression of the stages is illustrated in one direction. Twenty-one (26.92%) models have multi-directional movements among the stages. Finally, nine models (11.54%) have no indication of directions.

Model Shapes

More than half (n=41, 52.56%) of the models are in the circular shape. Twenty-six (33.33%) models take the linear shape, whereas the remaining 11 (14.10%) have a special shape. Among those with a special shape, there are four different types, including matrix (n=5) (e.g. "Support Your Data" rubric), staircase (n=3) (e.g. Ku and Gil-Garcia Data Lifecycle model), flowchart (n=2) (Core Scientific Metadata 'CSMD' model, the Research Lifecycle at UCF), and non-sequential hub (n=1) (Data Value Network model).

Some circular and linear shaped models have additional shapes embedded, including circular with a loop (n=1) (see Figure 2. RIN/NESTA Research Lifecycle model), linear with a loop (n=5) (see Figure 3, University of Virginia data life cycle model), and circular with time series (n=1) (see Figure 4, the Evolution of Scientific Information Model, or Subramanyam's Model). Table 6 provides the summary of the models' direction and shapes.

Table 6. Model directions and shapes.

Directions	
Uni-Directional	48 (61.54%)
Multi-Directional	21 (26.92%)
No Direction	9 (11.54%)
Shapes	
Circular	41 (52.56%)
Linear	26 (33.33%)
Special Shapes	11 (14.10%)

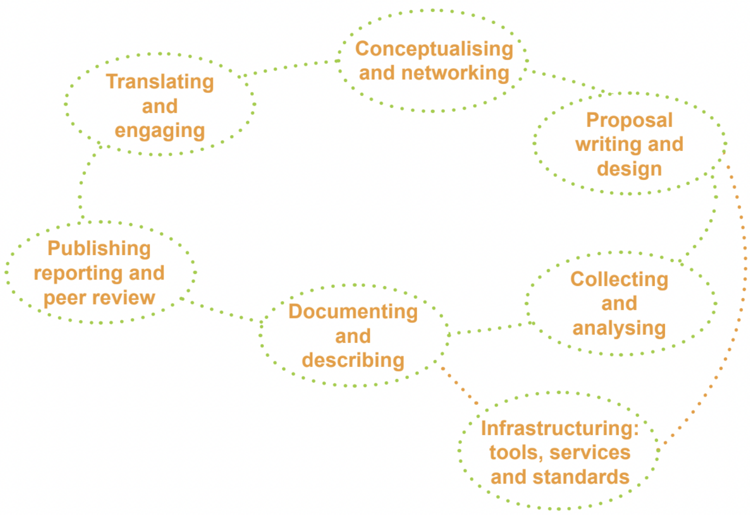

Figure 2. The RIN/NESTA Research Lifecycle model in the shape of circular with a loop. From Sandy et al. (2020, p. 202).

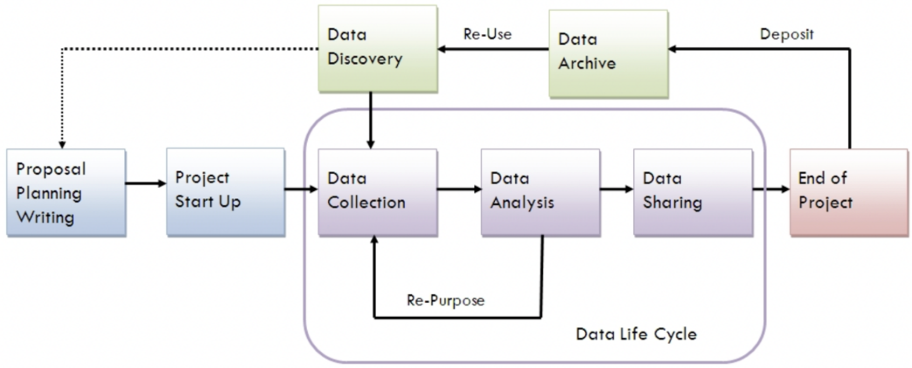

Figure 3. The University of Virginia Data Lifecycle model in the shape of linear with a loop. From University of Virginia (n.d.).

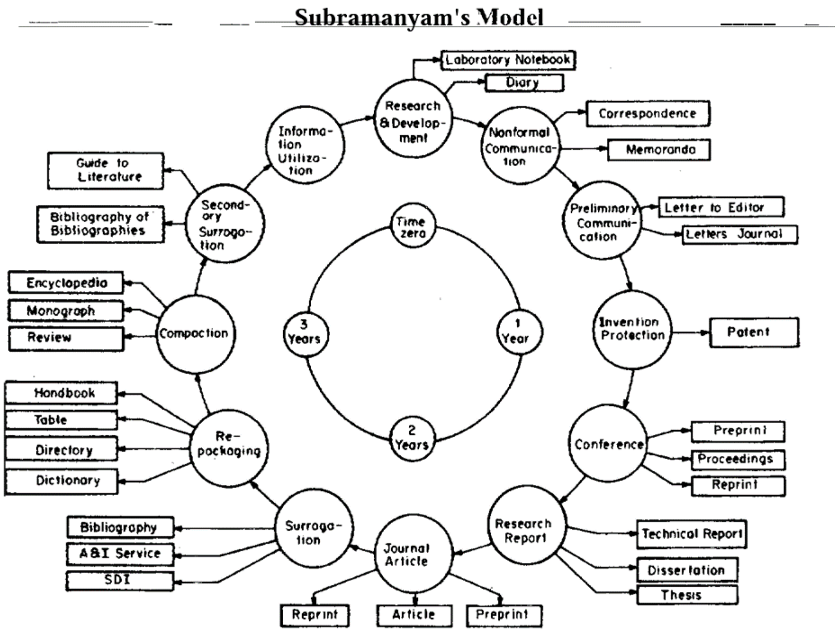

Figure 4. Screenshot of Evolution of Scientific Information model or Subramanyam's model in the shape of circular with time series. From Curl (2001, p. 458).

Discussion

RDLC Models Key Characteristics

As one of the first studies examining research and data lifecycle models from the point of view of their disciplinary focus, use cases, and model creators, as well as the specific stages and shapes, our results provided significant insights into the contexts and settings where the models were developed. Our study also identified gaps that may impact the use and value of the models.

Some notable gaps include (1) insufficient documentation and archiving of the model and (2) lack of detailed explanation of the features of the model. Throughout our study process, we discovered that several models, including some well-known or popular RDLC models when they were initially incepted, somehow vanished without a clear indication from relevant agencies/institutions concerning the whereabouts of the model or whether the model has been phased out or replaced by a newer version. We had to frequently rely on the Internet Archive's Wayback Machine to retrieve some of the disappearing models. For several models that we could find, the literature and explanation of the model's specific features were lacking.

As to specific visual displays of the models, the majority (51.28%) did not have a clear/explicit starting point. The model creators might have assumed some level of visual literacy and cultural orientation of the audience/readers of the model, and/or may have intended to indicate that the RDLC has no one single starting point, but without clear visual indicators, this could be confusing to users with diverse backgrounds. In addition, the effectiveness of the 10% of the models that lacked directional indicators might negatively impact on the effectiveness of their utilization. This insight is consistent with Cox and Tam's (2018) criticism of the RDLC models and should inform the creation of more robust, accessible and intuitively-understandable models.

Our study revealed that the majority of models with a disciplinary focus originated in STEM. Models originating in the social sciences and humanities are less common. The use cases varied across a wide spectrum, with a total of 34 different scenarios. The creators and authors of the RDLC models were primarily from academic institutions (close to 60%), while very few came from data-related organizations. We speculate that this may because there are fewer data-related organizations. These have become established relatively recently to respond to stakeholder needs in this area. In terms of country of origin, the model creators came from more than 20 countries with the majority of the models created as a result of collaboration within or across different organizations. We also note that the variations in the distribution of authors' countries might be due to each country's own initiatives, legislative effort, and funding focus on the value of data lifecycle.

Our stage and shape analysis also outlined key characteristics of the RDLC models by showing the commonalities and variations of named stages and varying structure of the models. With 78 models in total containing 130 different stages (75 common stages and 55 unique stages), the RDLC models certainly encompass an overwhelmingly wide variation in their labeling of specific stages. This vast variation makes it difficult for the consumer/user of RDLC models to comprehend and employ the model. Our recommendation to the creators of RDLC models is to examine other published models and provide labels that mirror earlier models, and also provide definitions of each stage and the rationale behind the label for a given stage. It would be helpful for users to be able to crosswalk various models that they have encountered and apply them in a flexible and evolving manner. In terms of the model shapes, while we found that the majority of models take a circular or linear shape, the shape often is not singular, and it often embeds another sub-shape, adding complexity. Nevertheless, both circular and linear are easier to understand for users, offering clearer visual pathways for understanding progression as opposed to the intricate and abstract configurations that potentially obscure stage flow. Therefore, from the usability and accessibility perspective, creating a RDLC model utilizing circular and linear shape is most acceptable. In terms of directionality, with the majority unidirectional, the models are straightforward and again, accessible. However, there were also more than 10% of the models that had no direction, which could be difficult to comprehend any movements among stages.

Study Limitations

Even though we examined 78 RDLC models, we recognize that our search might not have captured all the published RDLC models in existence, including some that were referenced in the literature, but for which we were unable to locate a visualization. Further, this study focused on models published in English; non-English models were excluded.

Moreover, we recognize that there is no agreed upon definition for RDLC models used consistently across studies. In some cases, it is unclear which definition of RDLC models that the authors of a given study referred to as they described a model as being an RDLC model. This inevitably presented difficulty in collecting our study sample. One or two RDLC models in our study sample might not be precisely a lifecycle model but we included them because they were described by other literature as an RDLC model. In performing shape analysis, we also noted the presence/absence of the QA/QC element, however, it was obvious certain RDLC models were not focused on the quality assurance/quality control component of the lifecycle, and therefore were missing this particular piece.

Furthermore, contextual information provided by authors/creators varied between models. As our coding was based on this information, granularity, consistency, and depth of coding did differ between models. This impacted our coding of both disciplinary focus and use cases. With certain disciplines and use cases presented originally by authors/creators as rather specific, others might be rather broad in scope. While we acknowledge the issue of consistency, granularity, and depth in the coding, our reliance on authors/creators' own description of their models gives us confidence as this is how they viewed the model's disciplinary focus and use cases.

Conclusion

Our study is one of the first empirical investigations examining the substance of the RDLC models from their key characteristics including disciplinary contexts, use cases, stages and shapes, and more. Through our analysis, we have noted that while RDLC models have proliferated, serious gaps and inconsistencies occur across the models. Our recommendations to future model creators/authors include: (1) provide a clear definition of the RDLC model, (2) provide a detailed documentation of the model development history, (3) make the model visually explicit in terms of the starting point and ending point (if any), the directions, the stages, and the relationship between stages and entities in the model, (4) the labeling of the model stage/entities need to be clear and may either crosswalk or be consistent with stages/entities of previous models if appropriate, and (5) make sure the visual of the model is comprehensible rather than challenging to readers of varying visual literacy levels.

Moreover, in our study, we found that the majority of RDLC models come from STEM, rather than social sciences or humanities. This inevitably impacts the shapes models take and the stages that are considered part of the lifecycle. When such a model is deployed in another context, such as the design and development of research data services (RDS), this may lead to certain stages and activities being emphasized, while others are rendered less visible, or invisible, resulting in services that do not meet the needs of researchers from all disciplines.

Even though we have provided descriptive analysis of various characteristics of the RDLC models, further analysis is planned to explore whether there are significant differences by discipline, country, and organizational affiliation in the way research and data lifecycles are visually represented and the stages that are explicitly defined in the life cycle.

As RDLC models can serve as the basis for conceptualizing the design of research data services (RDS), understanding which models are in use and how they are employed can help us explore the assumptions underlying the design of RDS, including if there are potential stages or activities that receive less attention. Future studies examining the use and practical usefulness of RDLC models in the context of library and information science education and practice might yield fruitful and practically valuable results.

References

American Society for Quality. (2022). *What is nominal group technique?* https://asq.org/quality-resources/nominal-group-technique

Ball, A. (2012). *Review of data management lifecycle models*. IDMRC, University of Bath, Bath, UK. https://researchportal.bath.ac.uk/en/publications/review-of-data-management-lifecycle-models

Carlson, J. (2014). The use of life cycle models in developing and supporting data services. In Ray, J. M. (Ed.), *Research data management: Practical strategies for information professionals* (pp. 63-86). Purdue University Press. https://doi.org/10.2307/j.ctt6wq34t.6

Cox, A. M., & Tam, W. W. T. (2018). A critical analysis of lifecycle models of the research process and research data management. *Aslib Journal of Information Management*, *70*(2), 142-157. https://doi.org/10.1108/AJIM-11-2017-0251

Curl, S. R. (2001). Subramanyam revisited: Creating a new model for information literacy instruction. *College & Research Libraries*, *62*(5), 455-464 (2001). https://doi.org/10.5860/crl.62.5.455

Fan, Z. (2019). Context-based roles and competencies of data curators in supporting research data lifecycle management: Multi-case study in China. *Libri*, *69*(2), 127-137. https://doi.org/10.1515/libri-2018-0065

Faundeen, J., Burley, T. E., Carlino, J. A., Govoni, D. L., Henkel, H. S., Holl, S. L., ... & Zolly, L. S. (2014). The United States geological survey science data lifecycle model (No. 2013-1265). US Geological Survey. https://doi.org/10.3133/ofr20131265

Harb, S. I., Tao, L., Peláez, S., Boruff, J., Rice, D. B., & Shrier, I. (2021). Methodological options of the nominal group technique for survey item elicitation in health research: A scoping review. *Journal of Clinical Epidemiology*, *139*, 140-148. https://doi.org/10.1016/j.jclinepi.2021.08.008

Lyon, L., Jeng, W., & Mattern, E. (2020). Developing the tasks-toward-transparency (T3) model for research transparency in open science using the lifecycle as a grounding framework. *Library & Information Science Research*, *42*(1), 100999. https://doi.org/10.1016/j.lisr.2019.100999

Pouchard, L. (2015). Revisiting the data lifecycle with big data curation. *International Journal of Digital Curation*, *10*(2), 176-192. https://doi.org/10.2218/ijdc.v10i2.342

Sandy, H. M., Million, A. J., & Hudson-Vitale, C. (2020). Innovating support for research: the coalescence of scholarly communication? *College & Research Libraries*, *81*(2), 193-214. https://doi.org/10.5860/crl.81.2

University of Virginia Library. (n.d.). *Research data management.* https://library.virginia.edu/data/data-management

Working Group on Information Systems and Services (WGISS). (2012). *CEOS data life cycle models and concepts version 1.2*, Committee on Earth Observation Satellites (CEOS). Retrieved from https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/WGISS_Data-Lifecycle-Models-And-Concepts.pdf, last accessed 2024/03/25.

Appendix I

Table of the Sampled Research & Data Lifecycle Models and Source Information

RDLC Model	Source
DataONE Data Lifecycle	Allard, S. (2012). DataONE: Facilitating eScience through collaboration. *Journal of eScience Librarianship*, *1*(1): 3. https://doi.org/10.7191/jeslib.2012.1004
Digital Curation Centre (DCC) Lifecycle	Digital Curation Center (DCC). (n.d.). *Curation lifecycle model.* https://www.dcc.ac.uk/guidance/curation-lifecycle-model
BLM Data Management Handbook	Sinaeepourfard, A., Masip-Bruin, X., Garcia, J., & Marín-Tordera, E. (2015). A survey on data lifecycle models: Discussions toward the 6Vs challenges. *Technical Report (UPC-DAC-RR-2015–18).*
Data Security Lifecycle 2.0	Rich. (2011, September 6). *Data Security Lifecycle 2.0*. Securosis. https://securosis.com/blog/data-security-lifecycle-2-0/
The DigitalNZ Content model	LeFurgy, B. (2012, February 21). *Life cycle models for digital stewardship*. The Signal [Webpage]. The Library of Congress. https://blogs.loc.gov/thesignal/2012/02/life-cycle-models-for-digital-stewardship
Data Documentation Initiative (DDI)	DDI 3.3 (2020) documentation. (n.d.). *Introduction*. Retrieved March 25, 2024, from https://ddi-lifecycle-documentation.readthedocs.io/en/latest/User%20Guide/Introduction.html
United States Geological Survey (USGS) Science Data Lifecycle	Faundeen, J., Burley, T. E., Carlino, J. A., Govoni, D. L., Henkel, H. S., Holl, S. L., ... & Zolly, L. S. (2014). The United States geological survey science data lifecycle model. US Geological Survey Open File Report No. 2013-1265. http://pubs.usgs.gov/of/2013/1265
The Ecoinformatics model	Michener, W. K., & Jones, M. B. (2012). Ecoinformatics: Supporting ecology as a data-intensive science. *Trends in ecology & evolution*, *27*(2), 85-93.
The Generic Science model	Scientific data management (SDM) for government agencies: Report from the

	workshop to improve SDM. Workshop held June 29 - July 1, 2010, Washington D.C. March 2011. Report No. CENDI/2011-1. Co-sponsored by the Environmental Protection Agency (EPA), CENDI (The Federal STI Managers Group), and the Interagency Working Group on Digital Data (IWGDD). https://web.archive.org/web/20170718022038/http:/semanticommunity.info/Other/Scientific_Data_Management_for_Government_Agencies/Report_from_the_Workshop_to_Improve_SDM#Figure_C2.
Geospatial Data Lifecycle	Federal Geographic Data Committee. (March 31, 2010). Stages of geospatial data lifecycle a16 supplemental reference. https://www.fgdc.gov/policyandplanning/a-16/stages-of-geospatial-data-lifecycle-a16.pdf/view
The LOD2 (Linked Open Data) Stack model	Auer, S., Bühmann, L., Dirschl, C., Erling, O., Hausenblas, M., Isele, R., ... & Williams, H. (2012). Managing the life-cycle of linked data with the LOD2 stack. In, *The Semantic Web–ISWC 2012: 11th International Semantic Web Conference, Boston, MA, USA, November 11-15, 2012, Proceedings, Part II 11* (pp. 1-16). Springer Berlin Heidelberg.
The University of Deusto model	Sinaeepourfard, A., Masip-Bruin, X., Garcia, J., & Marín-Tordera, E. (2015). A survey on data lifecycle models: Discussions toward the 6Vs challenges. *Technical Report (UPC-DAC-RR-2015–18).*
The Records model	University Archives | Michigan State University. (2018, May 19). *Records Management.* https://web.archive.org/web/20180519074507/http:/archives.msu.edu/records/
The JISC Research model	Jisc. (2021, April 27). *Research Data Management Toolkit.* https://web.archive.org/web/20210427075604/https:/rdmtoolkit.jisc.ac.uk/research-data-lifecycle/
The UK Data Archive Model [UK Data Service Data Lifecycle Model]	Sinaeepourfard, A., Masip-Bruin, X., Garcia, J., & Marín-Tordera, E. (2015). A survey on data lifecycle models: Discussions toward the 6Vs challenges. *Technical Report (UPC-DAC-RR-2015–18).*
Beijing University Model	Yu, X., & Wen, Q. (2010, December). A view about cloud data security from data life cycle. In *2010 international conference on computational intelligence and software engineering* (pp. 1-4). IEEE.
IWGDD's Digital Data Lifecycle Model	National Science and Technology Council. (2009). Harnessing the power of digital data for science and society.

	https://web.archive.org/web/20190306105757/https://www.nitrd.gov/about/harnessing_power_web.pdf [Wayback Machine, Exhibit B]
Astroparticle Data Lifecycle	Tokareva, V., Bychkov, I., Demichev, A., Dubenskaya, J., Fedorov, O., Haungs, A., ... & Zhurov, D. (2021, July). German-Russian astroparticle data life cycle initiative to foster big data infrastructure for multi-messenger astronomy. In 37th International Cosmic Ray Conference (ICRC 2021), Online, 12.07. 2021–23.07. https://arxiv.org/pdf/1907.13303.pdf
McDowall's Data Lifecycle Model	McDowall, R. D. (2019). Data integrity focus, Part VII: A data life cycle for chromatography. *LC-GC North America*, *37*(8), 532-537. https://www.chromatographyonline.com/view/data-integrity-focus-part-vii-data-life-cycle-chromatography
Digital Data Lifecycle	Roth, S., & Luczak-Roesch, M. (2020). Deconstructing the data life-cycle in digital humanitarianism. *Information, Communication & Society*, *23*(4), 555-571. https://doi.org/10.1080/1369118X.2018.1521457
"Support Your Data" rubric	Borghi, J., Abrams, S., Lowenberg, D., Simms, S., & Chodacki, J. (2018). Support your data: A research data management guide for researchers. *Research Ideas and Outcomes*, *4*, e26439. https://riojournal.com/article/26439/
DaMaRO Research Data Lifecycle Model	Carlson, J. (2014). The use of life cycle models in developing and supporting data services. In Ray, J. M. (Ed.), *Research data management: Practical strategies for information professionals* (pp. 81-82). Purdue University Press. https://doi.org/10.2307/j.ctt6wq34t.6
Open Archival Information System (OAIS) Reference Framework	Consultative Committee for Space Data Systems (CCSDS) (2012). Reference model for an open archival information system (OAIS). https://public.ccsds.org/pubs/650x0m2.pdf
Data Lifecycle for Large Facilities, by NSF	Christopherson, L., Mandal, A., Scott, E., & Baldin, I. (2020). Toward a data lifecycle model for NSF large facilities. In *Practice and Experience in Advanced Research Computing* (pp. 168-175).
Capability Maturity Model for Scientific Data Management	Crowston, K., & Qin, J. (2011). A capability maturity model for scientific data management: Evidence from the literature. *Proceedings of the American Society for Information Science and Technology*, *48*(1), 1-9. https://doi.org/10.1002/meet.2011.14504801036
Big Data Lifecycle	Demchenko, Y., De Laat, C., & Membrey, P. (2014, May). Defining architecture

	components of the Big Data Ecosystem. In *2014 International conference on collaboration technologies and systems (CTS)* (pp. 104-112). IEEE. DOI: 10.1109/CTS.2014.6867550
IBM Lifecycle	IBM Software. (n.d.). Wrangling big data: Fundamentals of data lifecycle management. Retrieved March 25, 2024, from https://silo.tips/download/ibm-software-wrangling-big-data-fundamentals-of-data-lifecycle-management
Capability of Privacy Protection (CoPP) of Data Lifecycle	Lin, L., Liu, T., Hu, J., & Zhang, J. (2014, December). A privacy-aware cloud service selection method toward data life-cycle. In *2014 20th IEEE international conference on parallel and distributed systems (ICPADS)* (pp. 752-759). IEEE.
PII (Personal Identifiable Information) Lifecycle	Michota, A., & Katsikas, S. (2015, October). Designing a seamless privacy policy for social networks. In *Proceedings of the 19th panhellenic conference on informatics* (pp. 139-143). https://doi.org/10.1145/2801948.2801998
Enterprise Data Lifecycle	Chaki, S. (2015). Enterprise Information management in practice. Nueva York: Apress. ihttps://link.springer.com/book/10.1007/978-1-4842-1218-9
Hindawi Data Lifecycle	Khan, N., Yaqoob, I., Hashem, I. A. T., Inayat, Z., Mahmoud Ali, W. K., Alam, M., ... & Gani, A. (2014). Big data: Survey, technologies, opportunities, and challenges. *The scientific world journal*, 2014. https://doi.org/10.1155/2014/712826
Data Life Cycle Labs (DLCLs) model	Pouchard, L. (2015). Revisiting the data lifecycle with big data curation. *International Journal of Digital Curation*, *10*(2), 176-192. https://doi.org/10.2218/ijdc.v10i2.342
Big Data Lifecycle model	Demchenko, Y., De Laat, C., & Membrey, P. (2014, May). Defining architecture components of the big data ecosystem. In *2014 International conference on collaboration technologies and systems (CTS)* (pp. 104-112). IEEE.
Data Value Network-DVN Model	Attard, J., Orlandi, F., & Auer, S. (2016, October). Data value networks: Enabling a new data ecosystem. In *2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI)* (pp. 453-456). IEEE. 10.1109/WI.2016.0073
Comprehensive Scenario Agnostic Data Life Cycle (COSA-DLC)	Sinaeepourfard, A., Garcia, J., Masip-Bruin, X., & Marin-Tordera, E. (2016). A comprehensive scenario agnostic data lifecycle model for an efficient data complexity

	management. *2016 IEEE 12th International Conference on E-Science (e-Science)*, Baltimore. https://escience-2016.idies.jhu.edu/wp-content/uploads/2016/11/Garcia-Jordi-slides.pdf, see p.10
Ku and Gil-Garcia Data Lifecycle	Sutherland, M. K., & Cook, M. E. (2017, June). Data-driven s mart cities: A closer look at organizational, technical and data complexities. In *Proceedings of the 18th annual international conference on digital government research* (pp. 471-476). https://doi.org/10.1145/3085228.3085239
Data Value Spectrum	Lim, C., Kim, K. H., Kim, M. J., Heo, J. Y., Kim, K. J., & Maglio, P. P. (2018). From data to value: A nine-factor framework for data-based value creation in information-intensive services. *International journal of information management*, *39*, 121-135. https://www.sciencedirect.com/science/article/pii/S0268401217300816, see Fig.2
Core Scientific Metadata (CSMD) model	Matthews, B., Sufi, S., Flannery, D., Lerusse, L., Griffin, T., Gleaves, M., & Kleese, K. (2010). Using a core scientific metadata model in large-scale facilities. *The International Journal of Digital Curation*, *1*(5), 106-118.
Michigan State University (MSU) Records Lifecycle	Faundeen, J. L., & Hutchison, V. B. (2017). The evolution, approval and implementation of the US geological survey science data lifecycle model. *Journal of eScience Librarianship, 6*(2).
University of Virginia Data Lifecycle model	UVA Library. (n.d.). *Research data management*. Retrieved March 25, 2024, from https://library.virginia.edu/data/data-management
Data Lifecycle Model for Macrosystems Ecology (MSE) Research Data	Rüegg, J., Gries, C., Bond-Lamberty, B., Bowen, G. J., Felzer, B. S., McIntyre, N. E., ... & Weathers, K. C. (2014). Completing the data life cycle: Using information management in macrosystems ecology research. *Frontiers in Ecology and the Environment*, *12*(1), 24-30.
University of Oxford Research Data Management Chart	Research Data Management UAS. (Last updated November 10, 2010)., *Research Data Management*. https://web.archive.org/web/20120104141426/http:/www.admin.ox.ac.uk/rdm/
NOAA Environmental Data Life Cycle Functions	National Oceanic and Atmospheric Administration (NOAA) Observing Systems Council. (n.d.). *NOAA Environmental Data Management Framework.*

	https://nosc.noaa.gov/EDMC/documents/NOAA_EDM_Framework_v1.0.pdf
IWGDD's Digital Data Life Cycle Model	Shepanek, R., B. C. Carroll, J. Candlish, J. Campbell, D. Duncan, K. Kirby, L. Petterson, J. Pratt, H. Ramapriyan, H. Surbaugh, AND B. Wilson. (2011). Harnessing the power of digital data: Taking the next step. Scientific Data Management (SDM) for Government Agencies: Workshop to Improve SDM, Washington, DC, June 29 - July 01, 2010. U.S. Environmental Protection Agency, Washington, DC, EPA/600/R-11/053. https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=OSIM&dirEntryId=235192
Linear Data Life Cycle	Shepanek, R., B. C. Carroll, J. Candlish, J. Campbell, D. Duncan, K. Kirby, L. Petterson, J. Pratt, H. Ramapriyan, H. Surbaugh, AND B. Wilson. (2011). Harnessing the power of digital data: Taking the next step. Scientific Data Management (SDM) for Government Agencies: Workshop to Improve SDM, Washington, DC, June 29 - July 01, 2010. U.S. Environmental Protection Agency, Washington, DC, EPA/600/R-11/053. https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=OSIM&dirEntryId=235192
Generic Science Data Lifecycle	Shepanek, R., B. C. Carroll, J. Candlish, J. Campbell, D. Duncan, K. Kirby, L. Petterson, J. Pratt, H. Ramapriyan, H. Surbaugh, AND B. Wilson. (2011). Harnessing the power of digital data: Taking the next step. Scientific Data Management (SDM) for Government Agencies: Workshop to Improve SDM, Washington, DC, June 29 - July 01, 2010. U.S. Environmental Protection Agency, Washington, DC, EPA/600/R-11/053. https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=OSIM&dirEntryId=235192
FGDC Stages of the Geospatial Data Lifecycle pursuant to OMB Circular	Federal Geographic Data Committee. (March 31, 2010). Stages of geospatial data lifecycle a16 supplemental reference. https://www.fgdc.gov/policyandplanning/a-16/stages-of-geospatial-data-lifecycle-a16.pdf/view
NDIIPP Preserving Our Digital Heritage	Digital Preservation. (2011). Preserving our digital heritage: The national digital information infrastructure and preservation program 2010 report. https://www.digitalpreservation.gov/documents/NDIIPP2010Report_Post.pdf
U.S. Department of Health and Human Services Key Components	Clinical Tools, Inc. (n.d.). *Guidelines for responsible data management in scientific research.* https://ori.hhs.gov/images/ddblock/data.pdf

GeoMAPP Geoarchiving Process Lifecycle	Committee on Earth Observation Satellites (CEOS). (April 19, 2012). Data Lifecycle models and concepts CEOS version 1.2. https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/WGISS_Data-Lifecycle-Models-And-Concepts.pdf
Joint Information Systems Committee (JISC) Information Environment	Beagrie, N. (2004). The Continuing Access and Digital Preservation Strategy for the UK Joint Information Systems Committee (JISC). *D-Lib Magazine*, *10*(7/8). https://doi.org/10.1045/july2004-beagrie
Open Government Data Lifecycle	Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. *Government information quarterly*, *32*(4), 399-418.
Big Data Activity/Value Chain	Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). Big data: The next frontier for innovation, competition, and productivity. https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_full_report.pdf
CIGREF Data Lifecycle	Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). Big data: The next frontier for innovation, competition, and productivity. https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_full_report.pdf. See Exhibt 37.
The Data Lifecycle Model	Redman, T.C., Fox, C., & Levitin, A. (2017). Data and data quality. In McDonald, J.D., & Levine-Clark, M. (Eds.). *Encyclopedia of Library and Information Sciences (4th ed.)* (1171-1182). CRC Press. https://doi.org/10.1081/E-ELIS4
Big Data Structure Lifecycle	Demchenko, Y., De Laat, C., & Membrey, P. (2014, May). Defining architecture components of the Big Data Ecosystem. In *2014 International conference on collaboration technologies and systems (CTS)* (pp. 104-112). IEEE. 10.1109/CTS.2014.6867550
Scientific Workflow	Demchenko, Y., De Laat, C., & Membrey, P. (2014, May). Defining architecture components of the big data ecosystem. In *2014 International conference on collaboration technologies and systems (CTS)* (pp. 104-112). IEEE. 10.1109/CTS.2014.6867550

Learning the Grant Life Cycle	Carlson, J. (2014). The use of life cycle models in developing and supporting data services. In Ray, J. M. (Ed.), *Research data management: Practical strategies for information professionals* (pp. 69). Purdue University Press. https://doi.org/10.2307/j.ctt6wq34t.6
Grant Lifecycle	Grants.gov. (n.d.). The grant lifecycle. Retrieved from https://www.grants.gov/learn-grants/grants-101/the-grant-lifecycle
Research 360 Institutional Research Lifecycle	McKen, K., Pink, C., Lyon, L., & Davidson, M. (2012). Research360: Data in the Research Lifecycle. https://researchportal.bath.ac.uk/en/publications/research360-data-in-the-research-lifecycle
RIN/NESTA Research Lifecycle	Sandy, H. M., Million, A. J., & Hudson-Vitale, C. (2020). Innovating support for research: the coalescence of scholarly communication? *College & Research Libraries*, *81*(2), 193-214. https://doi.org/10.5860/crl.81.2
The Research Lifecycle at UCF	UCF Libraries. (n.d.). *Overview: Research Lifecycle.* Retrieved March 25, 2024, from https://library.ucf.edu/about/departments/scholarly-communication/overview-research-lifecycle/
Research Lifecycle Enhanced by an "Open Science by Default" Workflow	Grigorov, I., Carvalho, J., Davidson, J., Donnelly, M., Elbaek, M., Franck, G., Jones, S., Melero, R., Knoth, P., Kuchma, I., Orth, A., Pontika, N., Rodrigues, E., & Schmidt, B. (2016, April 18). Research Lifecycle enhanced by an "Open Science by Default" Workflow. Zenodo. https://doi.org/10.5281/zenodo.49960
I2S2 Idealized Scientific Research Activity Lifecycle Model	Patel, M. (2011, Apr 7). I2S2 idealised scientific research activity lifecycle model. https://researchportal.bath.ac.uk/files/11246894/I2S2_ResearchActivityLifecycleModel_110407.pdf
the Integrated Scientific Lifecycle of Embedded Networked Sensor research	Pepe, A., Mayernik, M., Borgman, C. L., & Van de Sompel, H. (2010). From artifacts to aggregations: Modeling scientific life cycles on the semantic web. *Journal of the American Society for Information Science and Technology*, *61*(3), 567-582. https://arxiv.org/pdf/0906.2549.pdf
Humphrey's e-Science and RDLC model: e-science and the lifecycle of research	Humphrey, C. (2006). E-Science and the Life Cycle of Research. Education and Research Archive | University of Alberta. https://doi.org/10.7939/R3NR4V
ULS/T3 Model	Lyon, L., Jeng, W., & Mattern, E. (2020). Developing the tasks-toward-transparency (T3) model for research transparency in open science using the lifecycle as a grounding framework. *Library & Information Science*

	Research, *42*(1), 100999. https://doi.org/10.1016/j.lisr.2019.100999
The Evolution of Scientific Information Model, or Subramanyam's Model	Curl, S. R. (2001). Subramanyam revisited: Creating a new model for information literacy instruction. *College & Research Libraries, 62*(5), 455-464 (2001). https://doi.org/10.5860/crl.62.5.455
Research Publication Model, based on the evolved Subramanyam's Model (i.e. "The Evolution of Scientific Information" Model)	Carlson, J. (2014). The use of life cycle models in developing and supporting data services. In Ray, J. M. (Ed.), *Research data management: Practical strategies for information professionals* (pp. 68). Purdue University Press. https://doi.org/10.2307/j.ctt6wq34t.6
Loughborough University Library Research Lifecycle Model	Loughborough University Library. (n.d.). *Research Support*. Retrieved from https://web.archive.org/web/20140103091749/https:/www.lboro.ac.uk/services/library/research/
Model of a Scholarly Communication Lifecycle	Carlson, J. (2014). The use of life cycle models in developing and supporting data services. In Ray, J. M. (Ed.), *Research data management: Practical strategies for information professionals* (pp. 67). Purdue University Press. https://doi.org/10.2307/j.ctt6wq34t.6
Survey Lifecycle model	Survey Research Center. (2016). Guidelines for Best Practice in Cross-Cultural Surveys. Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan. Retrieved March 25, 2024, from http://www.ccsg.isr.umich.edu/
Academic Research Lifecycle Model	Sheombar, A. (2019). Reflections on social media use along the academic research life cycle. In *12th IADIS International Conference on Information Systems (IS 2019)*, Utrecht, The Netherlands. https://doi.org/10.13140/RG.2.2.15423.76963
Research Lifecycle of Nanotechnology	Lee, S. H., & Kim, H. S. (2012, November). The research life cycle and innovation through grey literature in nanotechnology in Korea. In *Fourteenth International Conference on Grey Literature* (p. 31). https://www.greynet.org/images/GL14-S1S,_Lee.pdf
Scientific Research Lifecycle	Chung, E., Kwon, N., & Lee, J. (2016). Understanding scientific collaboration in the research life cycle: Bio-and nanoscientists' motivations, information-sharing and communication practices, and barriers to collaboration. *Journal of the association for information science and technology*, *67*(8), 1836-1848. https://doi.org/10.1002/asi.23520

Collaborative Dynamics (in the context of the scientific research lifecycle)	Chung, E., Kwon, N., & Lee, J. (2016). Understanding scientific collaboration in the research life cycle: Bio-and nanoscientists' motivations, information-sharing and communication practices, and barriers to collaboration. *Journal of the association for information science and technology*, *67*(8), 1836-1848. https://doi.org/10.1002/asi.23520
EPA (Project Lifecycle)	Shepanek, R., B. C. Carroll, J. Candlish, J. Campbell, D. Duncan, K. Kirby, L. Petterson, J. Pratt, H. Ramapriyan, H. Surbaugh, AND B. Wilson. (2011). Harnessing the power of digital data: Taking the next step. Scientific Data Management (SDM) for Government Agencies: Workshop to Improve SDM, Washington, DC, June 29 - July 01, 2010. U.S. Environmental Protection Agency, Washington, DC, EPA/600/R-11/053. https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=OSIM&dirEntryId=235192
The Lifecycle Model of Research Knowledge Creation	Humphrey, C. (2006). E-Science and the Life Cycle of Research. Education and Research Archive | University of Alberta. https://doi.org/10.7939/R3NR4V

Appendix II

Quantitative Coding Framework

Categories	Sub-category	Criteria	Approach
Model Title		According to the detail provided by the model creators or the source descriptions	Quantitative coding based on model source
RDLC Model Creator Affiliations	Country	USA UK Germany New Zealand Spain South Korea International China Netherlands ...	Quantitative coding based on model source, synthesis analysis.
	Type of Affiliated Organization	Industry (e.g. IBM, Microsoft Research) Academic Institution (e.g. University, University library) Government Agencies (e.g. Environmental Protection Agency (EPA), Grants.gov, European Commission under European Union, Library of Congress) Data Organization (e.g. JISC, Data Documentation Initiative)	Quantitative coding based on model source, synthesis analysis.
	Cross-institutional collaboration or not	Whether the original creators are associated with two or more institutions	Quantitative coding
	Cross-border collaborations or not	Whether the original creators are from two or more countries	Quantitative coding based on model source, synthesis analysis.
Model Type	Data Lifecycle (DLC) Model	According to the detail provided by the model creators or the source descriptions, the model is intended to describe the journey of or management activities around data	Quantitative coding based on model source, synthesis analysis.

	Research Lifecycle (RLC) Model	According to the detail provided by the model creators or the source descriptions, the model is intended to describe or inform specifics about research process	Quantitative coding based on model source, synthesis analysis.
	Research and Data Lifecycle (RDLC) Model	According to the detail provided by the model creators or the source descriptions, the model includes both descriptions about data journey and research process	Quantitative coding based on model source, synthesis analysis.
	Other (e.g. Grant Lifecycle)	According to the detail provided by the model creators or the source descriptions, the model is intended for specific area of use, such as research grant, survey study lifecycle	Quantitative coding based on model source, synthesis analysis.
Stage	Stage Labels	Original stage labels presented in the model	Quantitative coding based on model source, synthesis analysis.
	Number of stages in each model	Number of stages appeared in the model	Quantitative coding based on model source, synthesis analysis.
Shape	Circular Shape	The flow of stages involves a recurring or cyclical process	Quantitative coding based on model source, synthesis analysis.
	Linear Shape	Linear models have their stages presented in a sequential manner, indicating a start-to-finish flow,	Quantitative coding based on model source, synthesis analysis.
	Special Shapes	The model's shape possesses some unique configurations, such as models that represent stages using a staircase, tables, or other non-linear, non-circular layouts.	Quantitative coding based on model source, synthesis analysis.
Starting Point	There is a starting point.	There is a clear indication of the beginning stage in the model's visualization.	Quantitative coding based on model source, synthesis analysis.
	There is not a starting point.	No clear indication of the starting point in the model's visualization.	Quantitative coding based on model source, synthesis analysis.
Direction	Uni-directional	The progression of the stages is illustrated in one direction.	Quantitative coding based on model

			source, synthesis analysis.
	Multi-directional	there may be back and forth directions, or multiple directional angles indicated in the model	Quantitative coding based on model source, synthesis analysis.
	No-direction	No presentation or indication of the flow of stages in the model	Quantitative coding based on model source, synthesis analysis.
QA/QC Highlight	There is a highlight of the QA/QC element throughout the stages	There is a highlight of the QA/QC throughout the lifecycle stages in the model's visualization or from the source description.	Quantitative coding based on model source, synthesis analysis.
	There is not a highlight of the QA/QC throughout the lifecycle stages	There is NO highlight of the QA/QC throughout the lifecycle stages in the model's visualization or from the source description.	Quantitative coding based on model source, synthesis analysis.
Intended Use Case	According to the description made by the original author or from the source.	specific activities, settings, or types of data for which the RDLC models were designed and intended to inform data practices by the original author or from the source.	Quantitative coding based on model source.
Disciplinary Focus	Generic	According to the detail provided by the model creators or the source descriptions, the model is not originating from or intended to be used within a specific field.	NGT, Quantitative coding based on model source, synthesis analysis.
	STEM	According to the detail provided by the model creators or the source descriptions, the model is intended to be used within the following list of discipline areas: Science Environmental Science Geospatial NanoTechnology & Material Science Chemistry Engineering (Civic Engineering, Industry Engineering) BioTechnology Astro-Physics Health & Medicine Energy Biology	NGT, Quantitative coding based on model source, synthesis analysis.

	Physics Ecology Computer Science	
Social Sciences	According to the detail provided by the model creators or the source descriptions, the model is intended to be used within the following list of discipline areas: Business Public Administration Policy Government Social Sciences Law	NGT, Quantitative coding based on model source, synthesis analysis.
Multidiscipline	According to the detail provided by the model creators or the source descriptions	NGT, Quantitative coding based on model source, synthesis analysis.
Humanities	According to the detail provided by the model creators or the source descriptions	NGT, Quantitative coding based on model source, synthesis analysis.