On the analysis of quality for data lifecycle models

Mauro Iacono[1,*,†], Michele Mastroianni[2,*,†], Christian Riccio[1,†] and Bruna Viscardi[3,†]

[1]Università degli Studi della Campania "L. Vanvitelli", viale Lincoln 5, 81100 Caserta, Italy
[2]Università degli Studi di Foggia, Via A. Gramsci 89/91, 71122 Foggia, Italy
[3]Independent expert, Italy

Abstract

Information systems are backed up by data and rely on them. The quality of information systems depend on a proper and accurate data management: the design of data management cycles should consequently be supported by sound quality analyses which may identify risks and qualify the phases in terms of quality dimensions. In this paper we present an analysis framework for the quality of data management cycles.

Keywords

Data management, Performance evaluation, Availability, Security, Privacy, Information systems, Information management, Quality

1. Introduction

Quality and effectiveness of information systems depend on a sound, consistent, and well posed management of the data sources on which they rely. Being those data lakes, big data repositories, relational or non-relational databases or other sources, a proper design, operation, and evolutionary maintenance of the data processing system is the fundamental basis to support the applications composing the information system. Complexity derives not only from data volumes, source and nature heterogeneity and technological articulation, but from non-functional specifications either of the data processing system or of the applications.

A structured approach to the choice of the right data management cycle should include a correct articulation in phases and a proper evaluation of the characteristics of each phase which can support the compliance with the desired non-functional requirements in each epoch of its lifetime, from the bootstrap of the system to its decommissioning. The adoption of well-known data management cycle may provide a good reference for designer and administrators to dominate complexity.

Literature offers several different Data Lifecycle Models (DLM) for data management [1, 2], but does not specifically focus on a structured approach to the choice. We proposed in [3] an approach for the selection of the most appropriate DLM given an overall qualitative evaluation of the relevant desired characteristics by leveraging a panel of experts: in this paper we focus on a framework for the qualification of each DLM phase to provide a finer grain support for DLM design.

2. Background: Data Lifecycle Models

A data lifecycle model (DLM) is a conceptual framework that describes the stages through which data passes during its existence, from initial creation or capture to its eventual deletion or archival [4]. The model typically includes phases such as data generation, collection, storage, processing, analysis, sharing, and disposal. It helps organizations manage data systematically to ensure data quality, compliance, security, and effective usage. A typical DLM is shown in Figure 1.

QualITA 2025: The Fourth Conference on System and Service Quality, June 25 and 27, 2025, Catania, Italy
*Corresponding author.
†These authors contributed equally.
mauro.iacono@unicampania.it (M. Iacono); michele.mastroianni@unifg.it (M. Mastroianni); christian.riccio@unicampania.it (C. Riccio); bruna_viscardi@hotmail.it (B. Viscardi)

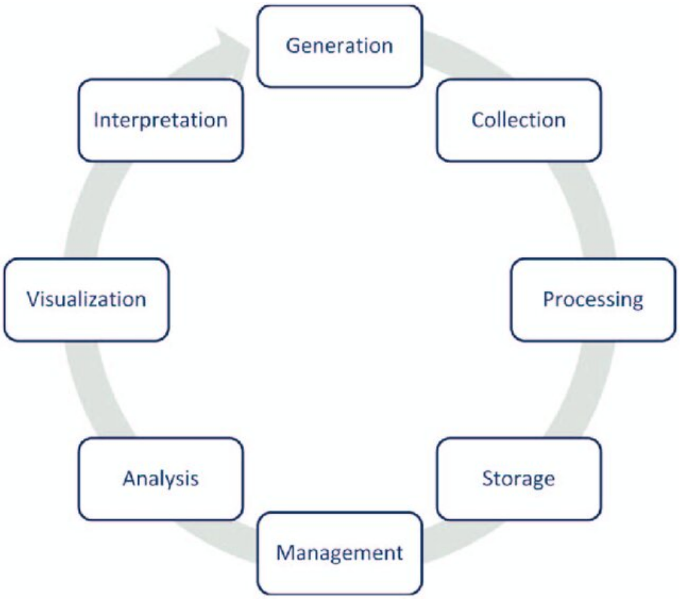

Figure 1: A typical Data Lifecycle Model (adapted from [5]).

In scientific and technical literature there are several data lifecycle models designed to deal with different data management issues or particular case studies. Some DLMs, such as USGS [6], DATAOne [7] or IBM DLM [8], have been widely used by a number of research group in their activity of data management in various different disciplines.

The availability of many best practices that could be used as *Standard* for the data lifecycle model could improve significantly the aspects related to data quality. Using that DLMs could have the following positive aspects:

- **Clarity**: literature models provide a clear, structured, and often visualized framework that helps clarify the stages of data management, making it easier for teams to understand and communicate processes [9];
- **Best Practices and Compliance**: these models are typically built on established best practices and are designed to support regulatory compliance, data quality, and security standards [10];
- **Efficiency and Consistency**: standard models reduce inefficiencies and risks associated with ad hoc approaches, supporting optimized data retention, cost savings, and improved decision-making;
- **Broad Applicability**: literature models are often designed to be widely applicable, providing a high-level framework that can be adapted across different domains and organizations [11];
- **Support and Documentation**: here is often extensive documentation, community support, and existing tools aligned with standard models, making implementation and troubleshooting easier.

On the other side, there is also to consider some drawbacks:

- **Lack of Customization**: literature models may not address specific organizational needs or unique workflows, potentially leading to gaps in coverage or relevance for specialized projects [11], [9];
- **Oversimplification**: standard models may mask the real-world complexity of data processes, presenting them as linear or closed systems, which can be misleading in dynamic or multi-directional research environments [9] [12];
- **Rigidity**: adhering strictly to a standard model can limit flexibility and innovation, especially when dealing with novel data types or emerging business requirements.

3. The Analytic Hierarchy Process (AHP) and its variant AHP-Express

AHP is a structured technique for organizing and analyzing complex decisions proposed by Saaty in 1987 [13] [14] [15]. It works by breaking down a problem into a hierarchy of criteria and alternatives, then making pairwise comparisons to establish priorities.

Firstly, it is necessary to structure the decision problem as a hierarchy, so it is decomposed into three primary levels:

1. Goal: the overall objective;
2. Criteria: the factors contributing to the goal, $\{C_1, C_2, \ldots, C_m\}$;
3. Alternatives: the choices to be evaluated, $\{A_1, A_2, \ldots, A_n\}$

$$\text{Goal} \longrightarrow \{\ C_1, C_2, ..., C_m\ \} \longrightarrow \{\ A_1, A_2, ..., A_n\ \} \tag{1}$$

For each element in a level, a pairwise comparison matrix A is constructed. The entries a_{ij} are based on the Saaty's 1-9 scale [14] of relative importance.

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \tag{2}$$

The matrix has two key properties: i) Reciprocity, $a_{ji} = \frac{1}{a_{ij}}$ and ii) Diagonal elements are $a_{ii} = 1$. So, the matrix A can be represented as:

$$A = \begin{bmatrix} 1 & a_{12} & \cdots & a_{1n} \\ 1/a_{12} & 1 & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 1/a_{1n} & 1/a_{2n} & \cdots & 1 \end{bmatrix} \tag{3}$$

The priority vector $\mathbf{w}$, which represents the relative weights of the elements being compared, is found by solving the eigenvalue problem:

$$A\mathbf{w} = \lambda_{\text{max}}\mathbf{w} \tag{4}$$

where λ_{max} is the *principal (largest) eigenvalue* of matrix A. and $\mathbf{w}$ is its corresponding *right eigenvector*, normalized so that its elements sum to 1.

$$\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} \quad \text{with} \quad \sum_{i=1}^{n} w_i = 1 \tag{5}$$

In practice, the eigenvector is often approximated using the geometric mean method, which is simpler and yields very similar results. The steps are:

- Calculate the geometric mean $\tilde{w}_i$ for each row i:

$$\tilde{w}_i = \left(\prod_{j=1}^{n} a_{ij} \right)^{1/n} \tag{6}$$

- Normalize the geometric means to get the priority vector $\mathbf{w}$:

$$w_i = \frac{\tilde{w}_i}{\sum_{j=1}^{n} \tilde{w}_j} \tag{7}$$

- This gives the approximated priority vector $\mathbf{w} = (w_1, w_2, \ldots, w_n)^T$.

AHP measures the consistency of the decision-maker's judgments using the *Consistency Ratio (CR)*. To do that, firstly Consistency Index have to been computed using the following formula:

$$\text{CI} = \frac{\lambda_{\text{max}} - n}{n - 1} \tag{8}$$

It is possible to calculate the Consistency Ratio (CR)

$$\text{CR} = \frac{\text{CI}}{\text{RI}} \tag{9}$$

In which the Random Index (RI) is extracted from the known table for the matrix size n as described in [14]. A CR value of 0.10 or less is considered acceptable, while if CR > 0.10, the judgments may be too random and should be revised.

The final step is to aggregate the weights from all levels of the hierarchy. Obtain the priority vector for alternatives under each criterion ($\mathbf{w}_{\text{alt}}^j$ for criterion C_j) and the priority vector for the criteria themselves ($\mathbf{w}_{\text{crit}}$). For each alternative A_i, its final global priority G_i is calculated by summing the products of its local priority under each criterion and the weight of that criterion.

$$G_i = \sum_{j=1}^{m} (\text{local weight of } A_i \text{ under } C_j) \times (\text{weight of } C_j) \tag{10}$$

Or, in vector form for all alternatives:

$$\mathbf{G} = W_{\text{local}} \cdot \mathbf{w}_{\text{crit}} \tag{11}$$

where the columns of W_{local} are the local priority vectors for each criterion.

The alternative with the highest global priority G_i is the preferred choice.

3.1. The AHP-Express variant

AHP-Express is a variant of the well-known AHP method proposed in [16]. The core difference between the two methods lies in how they handle the evaluation of alternatives, which leads to significant differences in the number of required judgments and the underlying mathematical procedure.

The AHP-Express method simplifies the original AHP by replacing the m matrices for alternatives with direct rating. For each criterion C_j, the decision-maker directly assigns a local priority r_{ij} to each alternative A_i; the scale may be the same Saaty 1-9.

These ratings are then normalized per criterion to create the local priority vector for each criterion. The most common normalization is the additive normalization:

$$w_{ij}^{\text{local}} = \frac{r_{ij}}{\sum_{i=1}^{n} r_{ij}} \quad \text{for } j = 1, 2, ..., m \tag{12}$$

where r_{ij} is the rating of alternative i under criterion j. The other steps are identical to the original AHP method.

4. A qualification framework for DLM phases

Pointed out that using standard DLMs could significantly boost data quality in a research project, it is crucial to determine how to choose the best suited Data Lifecycle Model. first of all, it is necessary to detect the various issues related to data quality and to refer in which phase of the DLM are they related. From an accurate analysis of scientific literature, has been populated the following Table 1 [17], [18], [19], [20].

Lifecycle Stage	Quality risks	Mitigation Strategies
Data Collection	- Incomplete data - Biased sampling - Format inconsistencies	- Define clear data requirements - Use validation rules - Standardize collection methods
Data Storage	- Data corruption - Unauthorized access - Redundancy	- Implement encryption/backups - Role-based access control - Deduplication checks
Data Processing	- Transformation errors - Loss of data fidelity - Inconsistent aggregations	- Automated data validation - Logging/auditing - Version control
Data Analysis	- Misinterpretation - Outdated data - Sampling bias	- Data lineage tracking - Freshness checks - Peer review of models
Data Archival/Deletion	- Data degradation - Compliance violations - Unintended retention	- Retention policies - Secure deletion methods - Regular audits

Table 1
Quality Challenges Across Lifecycle Stages

In the Table below several data quality risks - and the related mitigation strategies - have been highlighted. A possible strategy devoted to select the best alternative standard DLM for the given problem is to compare the various DLM to find if there are the right mitigation strategies in every examined phase.

The comparison between DLMs could be evaluated using widely used Multi-criteria decision-making (MCDM) methods, such as AHP or TOPSIS.

There is also to underline that qualitative issues in DLM apply on a vertical and on a horizontal direction: the vertical direction concerns a single phase of the DLM and affects all applications which involve that phase; the horizontal direction concerns all phases involved by a single application. Moreover, depending on how the data lifecycle works in the real case (e.g. sporadic queries; continuous data flow; data burst, etc.) there are different paths in DLM phases.

In other words, there are two different series of parameters to be applied in the comparison:

1. set of parameters based on how the specific phase of the specific DLM performs in the requirements listed in Table 1;
2. set of parameters based on how the specific phase of the specific DLM is crucial in the contest of the whole lifecycle (relative importance of the single phase).

A well-known MCDM method that could be used in this case is the Analytic Hierarchy Process (AHP) by Saaty [14]. In this method, are to be populated the Decision Matrix A[i,j] and the Weight Vector w(k); the Decision Matrix has to be populated using the first set of parameters while the second set of weights will converge in Weight Vector.

4.1. A Specialized Support Tool

To validate what discussed so far, we applied it to two case studies (Section 5) and the evaluated alternatives using the AHP process. The decision framework is based on the AHP-Express variant, which reduces elicitation by fixing a reference criterion and comparing each remaining criterion against it according to Saaty's 1–9 scale. The tool requires as input a CSV/Excel file with DLMs as rows and criteria (factors) as columns, computes criterion priorities and DLM scores, and returns the final ranking together with bar and radar charts for visual and quick highlights. For a detailed discussion of the tool, please refer to the following paper [3].

5. Case studies

In order to practically illustrate the approach, two appropriate case studies has been analyzed. Firtly, it is necessary to select a series of DLMs to be used as alternatives in AHP method as described in Section 3.

The DLMs included in out assessment are:

- *DCC Curation Lifecycle*: DCC is a data life cycle model that allows data to be managed and preserved effectively, from its creation or receipt to its final preservation. It involves a planning phase in which strategies are implemented to preserve digital material throughout its life cycle, for example through the use of standards and technologies. In fact, it includes plans for the management and administration of all actions in the preservation life cycle.[21]
- *DDI Lifecycle*: this DLM consists of eight phases (concept, collection, processing, archiving, distribution, discovery, repurposing) in which attention is paid to reducing conceptual inconsistency in data through the use of ontologies in collection and the importance of archiving them in order to enable data reuse by the user. This dlm was created to address the lack of work on ontological dynamics in a manner consistent with the data lifecycle. Ontologies enable the promotion of data management, standardisation and integration.[22]
- *CIGREF Lifecycle*: It is a data management-oriented DLM. It includes phases such as collection, consolidation and structuring, followed by distribution and exploitation of data. In our opinion, it pays particular attention to the governance, compliance and security/privacy requirements typical of corporate contexts. [23].
- *CRUD Lifecycle*: This is a DLM that involves the adoption, for each of its five phases, of strategies aimed at reducing data security risks. The phases of creation, Storaging and Destruction are mandatory, while Use and Share and Archive phases are optional. [24]
- *DataONE Lifecycle*:Although this DLM comprises numerous stages, it cannot be defined as complex. In each stage, different people can interact directly with the data, and it is unlikely that a single person will interact with the data in all stages. The data life cycle is useful because it can be used to identify data flows and work processes for scientists, librarians, or others associated with the scientific data process. It is not suitable in the context of Big Data. [7]

The criteria has been chosen as in [3], and represents the different phases of DLMs.

Category	**Criteria**	**Meta-Phases**
A	Starting	Planning Collection
	Administration	Use/Reuse/Feedback Share Governance
	End-of-Life	Archival Disposal
B	Data Assessment	Preparation Quality
	Computation	Analysis Visualization Storage
	Security	Access Protection

Figure 2: The structured set of criteria; the meta-phases are detailed in [2].

Table 2
Values for each phase and each DLM. For each DLM the assigned phase score is given in terms of quality importance for that specific phase.

DLM	Starting	Assessment	Computation	Administration	Security	End-of-life
DataOne	4	8	10	0	0	0
DCC	8	7	9	6	7	6
CRUD	5	7	4	7	10	8
CIGREF	8	5	10	7	0	0
DDI	9	10	6	5	0	7

Table 3
Case study 1. Criterion priorities by AHP-Express: category-A and -B views and the final normalized vector used for scoring.

Phase	A	B	Final
Starting	0.3323	0.3323	0.3323
Assessment	0.0665	0.0665	0.0665
Computation	0.1108	0.1108	0.1108
Administration	0.3323	0.3323	0.3323
Security	0.1108	0.1108	0.1108
End-of-life	0.0475	0.0475	0.0475

Subsequently, the values for all criteria have been assigned by a team of experts using Saaty scale (1 to 9) according to the the comprehensiveness of the quality requirement as described in Table 1. The assigned scores are shown in Table 5.

We designed two case studies that stands as complementary viewpoints in data governance and quality, and are related to the management of water data. The only difference across cases is the reference criterion used during pairwise elicitation, which encodes the perspective of the decision maker.

5.1. Producer of certified data

In this first case study we took the perspective of an authority responsible of certifying the quality of water resources. Because early lifecycle failures spread and are expensive to fix, upstream governance—procedures, sampling plans, legal compliance, and stakeholder coordination—is essential in such a situation. In order to align the comparisons with a pre-analytical quality emphasis, we thus choose **Starting** as the AHP-Express reference criterion. What we expected from this first trial was that models that delay quality controls to later phases are expected to receive lower relative priority in this perspective, which tends to favor DLMs that explicitly address planning, provenance capture, protocol standardization, and intake controls when using Starting as the reference. The obtained results are presented in Sec. 5.1.1

5.1.1. Case 1 Results

In Table 3 are reported the criterion weights where is possibile to notice that most of the importance is concentrated on *Starting* and *Administration* (each ≈ 0.332), followed by *Computation* and *Security* (≈ 0.111), while *Assessment* and *End-of-life* are comparatively less emphasized. The resulting ranking (Table 4) places **DCC** first (7.174), followed by **CIGREF** (6.424), **CRUD** (6.383), **DDI** (6.313), and **DataOne** (2.968). A visual of the final rank, if presented in Fig. 3.

The radar plot in Fig. 4 compares the normalized profiles of DLMs over phases. Here, DLMs with robust profiles over phases tends to prevails over the others. Specifically, **DCC** has good benefits in *Starting* and *Administration*, while **CIGREF** is very strong in *Computation/Administration* and presents

Table 4
Case study 1. Final ranking of DLMs and scores.

Rank	DLM	Score
1	DCC	7.174
2	CIGREF	6.424
3	CRUD	6.383
4	DDI	6.313
5	DataOne	2.968

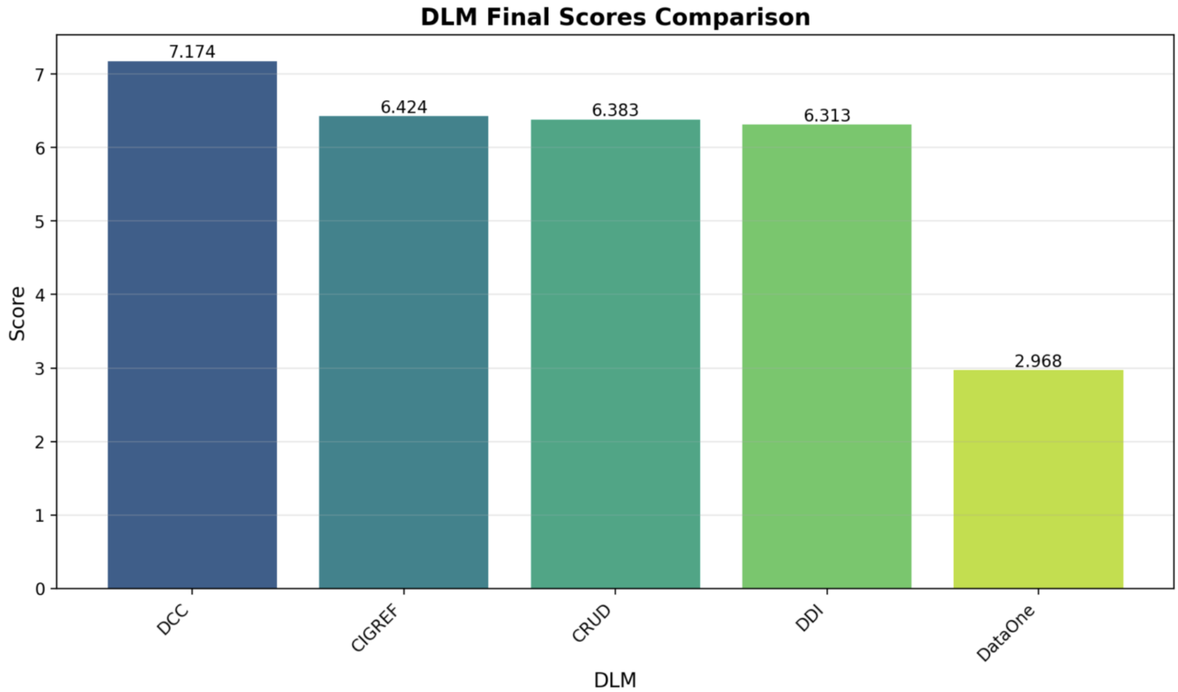

Figure 3: Case study 1. DLMs final scores.

null or lowers scores in *Security* and *End-of-life*, which penalizes the overall evaluation over such phase. **CRUD** excels in *Security* and *End-of-life*, but these two have lower weights than the two main ones; **DataOne** is ranked as last.

The radar aligns between the models profile and the weight structure visible in the final ranking bar plot.

5.2. Generic data user

Here, we assume a generic data user/company who uses data that has already been certified by third parties. Transformation quality, computational reproducibility, and processing pipeline performance/-traceability become the primary risk factors. Thus, we chose **Computation** as the reference criterion for AHP-Express.

DLMs that specify processing standards, validation during transformations, versioning, reproducibility, and auditability of computational steps should be given preference in this viewpoint. Strong intake models with less information about processing controls might be ranked lower overall. The obtained results are discussed in Section 5.2.1

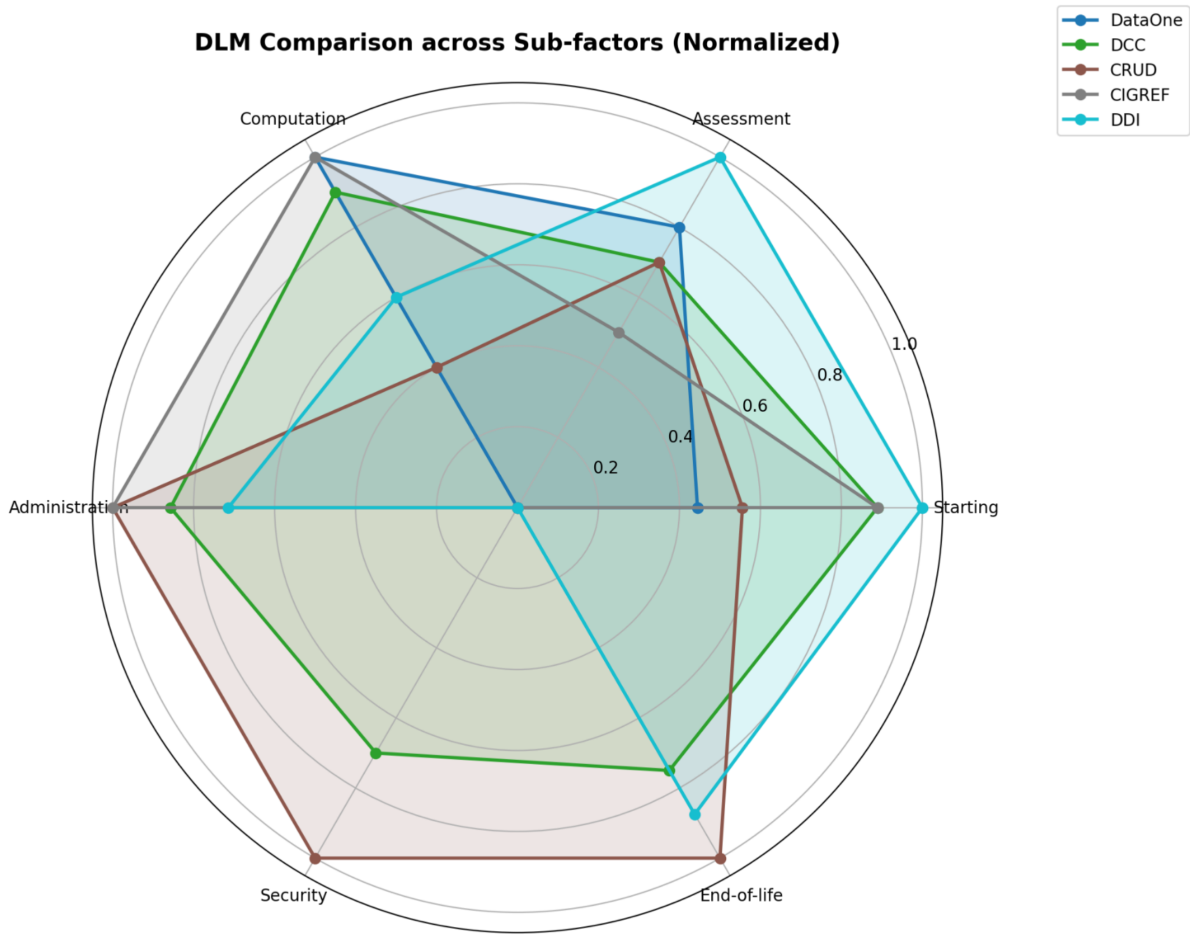

Figure 4: Case study 1. Normalized criteria profiles by DLM (radar).

Table 5
Case study 2. Criterion priorities by AHP-Express: category-A and -B views and the final normalized vector used for scoring.

Phase	A	B	Final
Starting	0.0641	0.4484	0.2560
Assessment	0.1495	0.0641	0.1070
Computation	0.4484	0.1495	0.2990
Administration	0.0641	0.0641	0.0640
Security	0.2242	0.2242	0.2240
End-of-life	0.0498	0.0498	0.0500

5.2.1. Case 2 Results

The criterion weights are reported in Table 5. here, most of the importance is concentrated on *Computation* (≈ 0.299), *Starting*(≈ 0.256), and *Security* (≈ 0.224), while *Assessment* (≈ 0.107), *Administration* (≈ 0.064), and *Endof-life* (≈ 0.050) are comparatively less emphasized. The resulting ranking (Table 6) places **DCC** first (7.740), followed by **CRUD** (6.313), **CIGREF** (6.021), **DDI** (5.836), and **DataOne** (4.868). A visual of the final rank is presented in Figure 5.

From a visual inspection of the radar plot in Figure 6 we can see that (paying focus to the reference criteria *Computation*), models that are strong on *Computation* and *Security* tend to prevail, provided they do not show structural gaps on the remaining axes. **DCC** exhibits a balanced and wide polygon, combining high *Computation* with strong *Starting*/*Administration*. **CRUD** excels on *Security*

Table 6
Case study 2. Final ranking of DLMs and scores.

Rank	DLM	Score
1	DCC	7.740
2	CRUD	6.313
3	CIGREF	6.021
4	DDI	5.836
5	DataOne	4.868

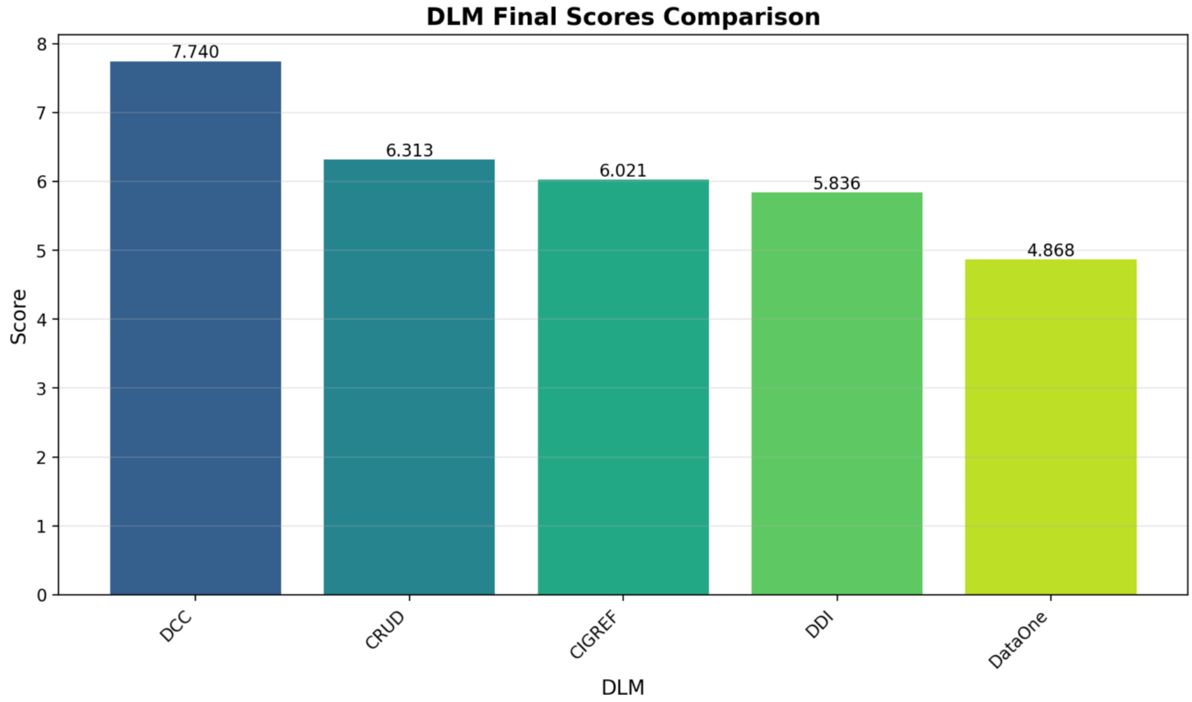

Figure 5: Case study 2. DLMs final scores.

and *End-of-life*—axes that carry non-negligible weight here, so it overtakes **CIGREF**, which is very strong on *Computation/Administration* but shows null values on *Security/End-of-life*. **DDI**, strong on *Starting/Assessment* but only moderate on *Computation*. **DataOne** improves, compared to the previous case study, thanks to *Computation*, yet it remains last because of gaps on *Security/End-of-life*. The radar, also here aligns with the profile of the models with the weight structure and explains the ordering observed in the final scores bar plot.

6. Conclusions and future work

In this paper a possible framework to be used to improve data quality has been presented, using AHP method to select the best alternative DLM to be used, having in mind the data quality improvement. Future work will involve the implementation of a series of test cases to assess the relevance of the results.

By comparing the results obtained in the two presented case studies, is possible to note that there is a shit in the final rank of the DLMs, where CRUD and CIGREF basically switches. This behavior is due to the fact that who produces the data gives more importance to phases weighted differently in the two models, i.e. Starting/Administration, whereas who uses the final data may give more importance to phases like Computation.

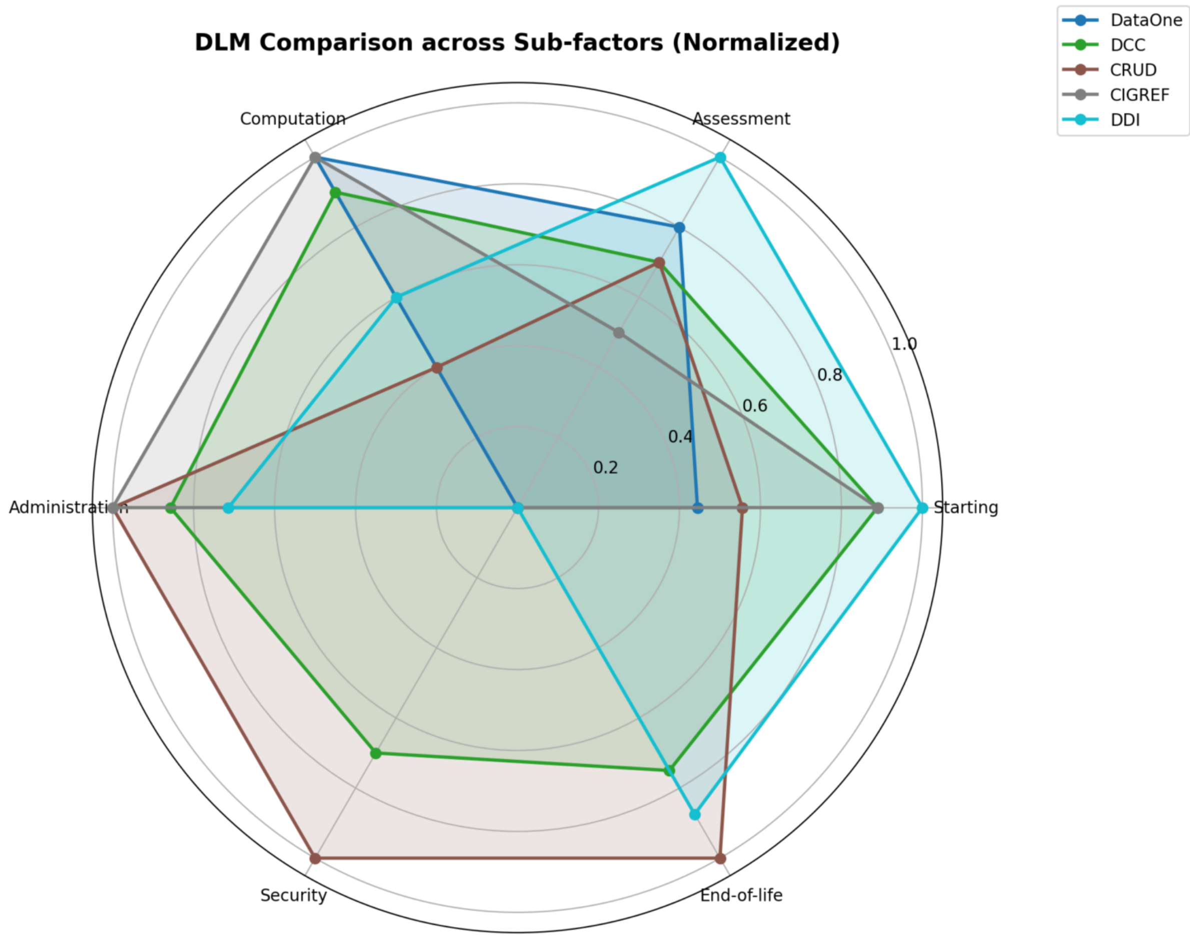

Figure 6: Case study 2. Normalized criteria profiles by DLM.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] M. El Arass, N. Souissi, Data lifecycle: from big data to smartdata, in: 2018 IEEE 5th international congress on information science and technology (CiSt), IEEE, 2018, pp. 80–87.

[2] S. I. H. Shah, V. Peristeras, I. Magnisalis, Dalif: a data lifecycle framework for data-driven governments, Journal of Big Data 8 (2021) 89.

[3] M. Iacono, M. Mastroianni, C. Riccio, B. Viscardi, A framework for data lifecycle model selection, Future Internet 17 (2025). URL: https://www.mdpi.com/1999-5903/17/9/390. doi:10.3390/fi17090390.

[4] D. International, DAMA-DMBOK: Data Management Body of Knowledge (2nd Edition), Technics Publications, LLC, Denville, NJ, USA, 2017.

[5] B. Lund, T. Wang, What does information science offer for data science research?: A review of data and information ethics literature, Journal of Data and Information Science 7 (2022) 16–38. URL: https://doi.org/10.2478/jdis-2022-0018. doi:10.2478/jdis-2022-0018.

[6] J. Faundeen, T. E. Burley, J. A. Carlino, D. L. Govoni, H. S. Henkel, S. L. Holl, V. B. Hutchison, E. Martín, E. T. Montgomery, C. Ladino, et al., The United States geological survey science data lifecycle model, Technical Report, US Geological Survey, 2014.

[7] S. Allard, Dataone: Facilitating escience through collaboration, Journal of eScience Librarianship (2012).
[8] IBM Software, Wrangling big data: Fundamentals of data lifecycle management, Technical Report, IBM, 2013.
[9] A. M. Cox, W. W. T. Tam, A critical analysis of lifecycle models of the research process and research data management, Aslib Journal of Information Management 70 (2018) 142–157.
[10] J. Zhang, J. Symons, P. Agapow, J. T. Teo, C. A. Paxton, J. Abdi, H. Mattie, C. Davie, A. Z. Torres, A. Folarin, et al., Best practices in the real-world data life cycle, PLOS digital health 1 (2022) e0000003.
[11] A. Sinaeepourfard, X. Masip-Bruin, J. Garcia, E. Marín-Tordera, A survey on data lifecycle models: Discussions toward the 6vs challenges, Technical Report (UPC-DAC-RR-2015–18) (2015).
[12] T. Weber, D. Kranzlmuller, Methods to evaluate lifecycle models for research data management, Bibliothek Forschung und Praxis 43 (2019) 75–81.
[13] R. W. Saaty, The analytic hierarchy process—what it is and how it is used, Mathematical modelling 9 (1987) 161–176.
[14] T. L. Saaty, Decision making with the analytic hierarchy process, International journal of services sciences 1 (2008) 83–98.
[15] M. Tavana, M. Soltanifar, F. J. Santos-Arteaga, Analytical hierarchy process: Revolution and evolution, Annals of operations research 326 (2023) 879–907.
[16] J. E. Leal, AHP-express: A simplified version of the analytical hierarchy process method, MethodsX 7 (2020) 100748.
[17] C. Batini, M. Scannapieco, DATA AND INFORMATION QUALITY: Dimensions, Principles and Techniques, SPRINGER, 2018.
[18] M. Ghasemaghaei, G. Calic, Assessing the impact of big data on firm innovation performance: Big data is not always better data, Journal of Business Research 108 (2020) 147–162. URL: https://www.sciencedirect.com/science/article/pii/S0148296319305740. doi:https://doi.org/10.1016/j.jbusres.2019.09.062.
[19] K. Lee, N. Weiskopf, J. Pathak, A framework for data quality assessment in clinical research datasets, in: AMIA Annual Symposium Proceedings, volume 2017, 2018, p. 1080.
[20] L. L. Pipino, Y. W. Lee, R. Y. Wang, Data quality assessment, Communications of the ACM 45 (2002) 211–218.
[21] S. Higgins, The dcc curation lifecycle model, International Journal of Digital Curation (2008).
[22] X. Ma, P. Fox, E. Rozell, P. West, S. Zednik, Ontology dynamics in a data life cycle: challenges and recommendations from a geoscience perspective, Journal of Earth Science 25 (2014) 407–412.
[23] S.BOUTEILLER, Business data challenges. How to manage company data to create value?, Technical Report, CIGREF, 2014.
[24] X. Yu, Q. Wen, A view about cloud data security from data life cycle, in: 2010 International Conference on Computational Intelligence and Software Engineering, 2010, pp. 1–4. doi:10.1109/CISE.2010.5676895.

A. Online Resources

The software tool used in this paper is available as an interactive web application on Streamlit at https: //ahp3-python.streamlit.app/ (accessed on 02 Sep 2025). The source code is hosted on GitHub at https://github.com/christianriccio/ahp3-python.git (accessed on 02 Sep 2025).