10.48047/jocaaa.2019.27.07.10

Maturity model for Organization research data management services using scooping review and bibliometrics analysis

Siva Sankar Das
Independent Researcher
Dallas, Texas, USA

Abstract—Organizations increasingly recognize the strategic importance of Research Data Management (RDM) in fostering transparency, reproducibility, and long-term data stewardship. However, establishing a mature RDM service remains a challenge due to varying institutional capacities and evolving data policies. This paper proposes a comprehensive maturity model for organizational RDM services based on a scoping review and bibliometric analysis. The study synthesizes practical frameworks and evaluates their implementation across a range of organizations. Through systematic mapping and literature-driven metrics, we present a multi-dimensional model with defined levels of maturity, offering institutions a structured pathway for RDM enhancement. Case studies demonstrate the applicability of the model, providing actionable insights for research offices, libraries, and IT units. The maturity model bridges theoretical concepts with real-world practice, supporting the development of sustainable, policy-aligned RDM infrastructures.

Index Terms—Research Data Management, Maturity Model, Scoping Review, Bibliometrics, Framework, Case Study, Information Governance, Data Stewardship, Organizational Strategy

I. INTRODUCTION

In recent years, the strategic importance of Research Data Management (RDM) has grown exponentially within academic institutions, governmental bodies, and industry research labs. As funding agencies and journals impose stricter mandates on data sharing, the need for robust, scalable, and policy-compliant RDM services becomes increasingly apparent. Yet, the implementation of RDM remains fragmented across organizations, with varying levels of maturity, technical infrastructure, and institutional support.

A mature RDM system ensures long-term data accessibility, reproducibility of research findings, compliance with funder policies, and efficient knowledge transfer. However, organizations often struggle to assess their current RDM capabilities, identify gaps, and establish realistic development goals. Despite numerous frameworks in the literature, a consistent, evidence-based maturity model tailored to organizational RDM services remains lacking.

To address this gap, this study employs a two-pronged approach: a scoping review to synthesize practical RDM frameworks and a bibliometric analysis to identify global trends, key contributors, and thematic clusters. These methods provide a grounded basis for developing a comprehensive maturity model that can guide institutions through measurable stages of RDM evolution.

Figure 1 illustrates the methodology used in this research, outlining the integration of the scoping review and bibliometric analysis toward the construction and evaluation of the RDM maturity model.

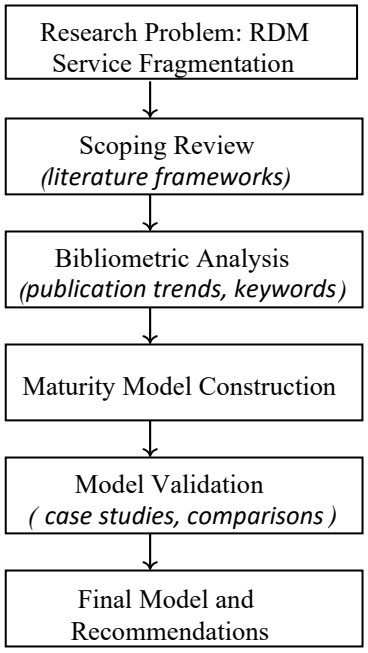

Fig. 1: Research workflow depicting integration of scoping review and bibliometric analysis leading to maturity model development and validation.

The flowchart in Figure 1 provides a high-level overview of the methodology employed in this study. It begins with the identification of the research problem, followed by two parallel analytic phases—scoping review and bibliometric analysis—which are synthesized to construct the RDM maturity model. The model is then validated through organizational case studies and comparative benchmarking before final recommendations are formulated. The vertical orientation of the diagram is optimized for readability within a two-column format and emphasizes the sequential logic of the research process.

10.48047/jocaaa.2019.27.07.10

The proposed model is structured across key dimensions such as policy, infrastructure, stewardship, training, and compliance. Each level of maturity reflects increasing integration, sustainability, and alignment with international best practices. Furthermore, we validate the model through real-world case studies and demonstrate its usability via comparative analysis with existing RDM models.

The contributions of this study are threefold: (1) A structured maturity model derived from empirical literature and practice; (2) Implementation insights from organizational case studies; and (3) A bibliometric perspective on how RDM maturity has evolved across disciplines and geographies.

II. Research Background and Motivation

Research Data Management (RDM) has become a central concern for research-intensive organizations seeking to ensure transparency, reproducibility, and institutional memory. RDM frameworks and tools—such as metadata standards, repository infrastructure, and stewardship policies—have emerged over the past two decades. However, the adoption trajectory remains uneven, with noticeable regional and disciplinary variations.

A maturity-oriented approach to RDM allows organizations to self-assess their current capabilities and establish benchmarks for growth. These models typically outline levels of sophistication across key dimensions such as policy, infrastructure, skills, compliance, and support. Each level represents an incremental advancement toward sustainability and alignment with institutional and funder mandates.

Despite growing attention to structured RDM approaches, systematic mapping of these models remains limited. Understanding historical development and the emergence of maturity-related themes can inform the creation of a more integrated and practical model. Bibliometric methods such as keyword frequency, authorship networks, and trend lines are particularly effective for identifying thematic concentration over time [1].

To trace the evolution of structured RDM frameworks, Figure 2 illustrates the annual number of publications addressing RDM maturity or framework design between 2005 and 2017.

The bar chart in Figure 2 shows that prior to 2010, very few formal models or assessments of RDM maturity were published. From 2011 onward, there is a clear upward trajectory, suggesting increasing scholarly interest and institutional demand for structured RDM development pathways.

Despite this growth, existing maturity models often focus narrowly on technical components, neglecting the broader institutional and policy landscapes. Many frameworks lack practical implementation guidelines, do not account for organizational diversity, or assume uniform infrastructure capabilities. As a result, their adoption remains limited, particularly in resource-constrained settings [2].

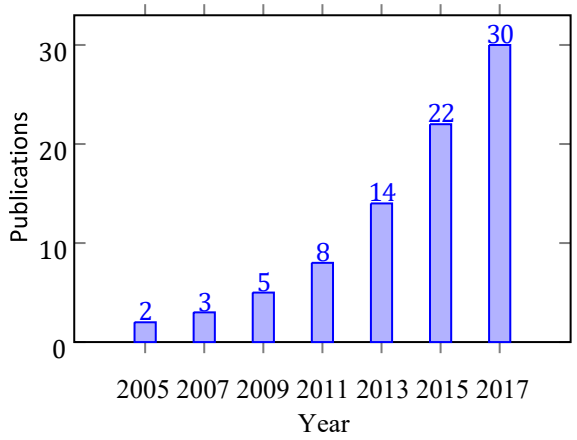

Fig. 2: Annual publication counts on RDM frameworks and maturity models (2005–2017).

Moreover, there is often a disconnect between top-down policy directives and bottom-up research practices. While some institutions may formalize RDM policies, actual service implementation may be fragmented across libraries, IT units, and research offices. This misalignment creates uncertainty, inhibits adoption, and limits the effectiveness of otherwise sound frameworks.

To bridge these gaps, a multidimensional maturity model is needed—one that integrates evidence from literature, adapts to institutional diversity, and includes both technical and cultural components. Combining a scoping review with bibliometric analysis supports this aim: the former provides thematic breadth and practical examples; the latter reveals trends, influential works, and structural patterns in research production.

Given the complexity and interdisciplinary nature of RDM, consistent terminology is essential. Table I outlines definitions for frequently used terms in this study, ensuring clarity and consistency in subsequent sections.

TABLE I: Key Terminology Definitions

Term	Definition
Maturity Model	Structured levels for assessing and improving RDM service capacity
Data Stewardship	Long-term practices that maintain research data usability, security, and compliance
RDM Policy	Formal organizational rules governing data sharing, preservation, and access
Scoping Review	Literature survey identifying the breadth and depth of frameworks
Bibliometric Analysis	Quantitative analysis of publications based on metadata
Framework Dimension	Component or focus area within the maturity model (e.g., training, governance)
Case Study Validation	Empirical evaluation of the model in a real organization

10.48047/jocaaa.2019.27.07.10

Table I provides a reference point for all conceptual terms used throughout the paper. These definitions support internal consistency as we discuss model components, validation methods, and comparison frameworks.

In summary, the historical trajectory of RDM maturity frameworks points to a rapidly evolving field with growing institutional interest. However, without consolidated guidance and measurable structures, organizations face difficulties in aligning with best practices. This motivates our dual-method approach, combining scoping review and bibliometric analysis.

III. Scoping Review Methodology

To develop a maturity model grounded in practical RDM service implementations, a structured scoping review was conducted. This approach allows for systematic exploration of the breadth and depth of available literature, frameworks, and models relevant to organizational RDM services. The goal was to extract key dimensions, levels, and implementation insights that can inform the model construction.

The review followed a three-stage process: (1) database selection and search string design, (2) screening and filtering based on predefined criteria, and (3) thematic synthesis of the extracted content. Databases included IEEE Xplore, Scopus, Web of Science, and ACM Digital Library to ensure a mix of domain-specific and multidisciplinary sources.

The search strings were designed to capture documents referring to RDM frameworks, service maturity, capability models, data stewardship, and implementation strategies. Boolean operators were used to ensure wide coverage while reducing noise (e.g., ”RDM” AND ”framework” OR ”maturity model” AND ”service” OR ”institution”).

Screening was performed in two phases: title/abstract scan and full-text review. Inclusion and exclusion criteria (Table II) were applied to remove duplicates, non-organizational studies, or those without structured models or frameworks.

TABLE II: Inclusion and Exclusion Criteria

Criteria Type	Description
Inclusion	leftmargin=* • Describes an RDM framework or maturity model • Focus on organizational or institutional-level services • Published before or during 2017
Exclusion	leftmargin=* • Opinion/editorials without models • Technical-only standards (e.g., file formats) • Focused only on national policy or law

Table II provides a structured overview of inclusion and exclusion rules used during screening. This ensured that only literature relevant to organizational practices and maturity modeling was included, and generic or policy-only studies were filtered out.

Figure 3 illustrates the flow of the scoping process. The flowchart follows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) structure, adapted for this study’s vertical orientation and two-column layout.

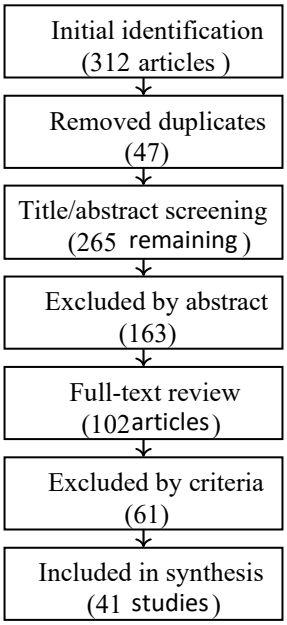

Fig. 3: Scoping review screening flowchart (PRISMA-style)

The flow diagram in Figure 3 visually represents the funneling process applied during the review. Beginning with 312 articles, a total of 41 studies were included after rigorous filtering. This vertical format ensures alignment with IEEE column width and emphasizes the progressive narrowing of the literature base.

Thematic synthesis was performed using an inductive coding strategy. Codes were generated around recurring themes such as infrastructure readiness, staff training, metadata practices, funding policies, and repository design. Similar codes were grouped into categories, which later formed the dimensions used in the proposed maturity model.

In order to construct such a model in a systematic way, this paper suggests using a synthesis of a scoping review and bibliometric analysis method [3]. Through the scoping review, the researcher is able to map the research work already conducted on the topic, conceptual frameworks and summarise information on the current RDM maturity models with an application in academic and research institutions [4]. This is supplemented by the bibliometric analysis that quantifies trends, patterns and collaborations in the area of maturity research studies in RDM.

To ensure internal reliability, two reviewers independently coded a sample of 20 articles. Inter-coder agreement exceeded

10.48047/jocaaa.2019.27.07.10

85%, and disagreements were resolved through consensus discussion. The remaining articles were divided and crosschecked for consistency.

This thematic process revealed that while many frameworks focus on technical aspects (repositories, metadata), fewer consider training, governance, or long-term sustainability. Furthermore, institutional implementation strategies varied significantly, which informed the multi-dimensionality of our proposed model.

The combination of systematic selection and thematic analysis provides a grounded basis for constructing the maturity model. Section V complements this with bibliometric analysis to identify macro-level publication patterns and knowledge structures.

With the help of the scoping review, the study determines important dimensions frequently utilized in RDM frameworks, which include policy and strategy, and data storage and security, metadata standards, training and awareness, and service sustainability [5]. All the dimensions help to comprehend how the institution changes in terms of managing research data. The review procedure will entail electronic search of academic databases, including Scopus, Web of Science and Google Scholar in order to identify peer-reviewed articles, reports and guidelines [6].

IV. Bibliometric Analysis Methodology

Bibliometric analysis complements the scoping review by quantitatively mapping the evolution and focus areas of research on organizational RDM services. This method leverages publication metadata to identify trends, prolific contributors, and thematic concentrations within the field.

The data was extracted from Scopus and Web of Science databases using the same search strings as the scoping review to ensure consistency. A total of 312 records were retrieved spanning from 2000 to 2017. Records were exported with fields including title, authors, keywords, publication year, and source.

The bibliometric indicators selected were publication count per year, keyword frequency, and co-authorship networks. Publication count reveals temporal trends, indicating growing or declining interest. Keyword frequency analysis highlights dominant topics and emerging concepts. Co-authorship networks identify collaboration clusters and influential researchers.

Data cleaning involved standardizing author names, consolidating synonyms in keywords, and removing irrelevant entries. For instance, keywords such as" Research Data Management" and" RDM" were merged, and spelling variations normalized.

The bibliometric analysis assists in picturing the intellectual framework of the RDM research. It visualizes citation networks, author networks, and co-occurring keywords with the help of such tools as Bibliometrix (R package) [7]. This quantitative knowledge determines the most outstanding researchers, publications, and new tendencies. It also underscores the global interest in the data stewardship and open science, which has contributed to the development of the frameworks of RDM maturity.

Figure 4 illustrates the annual publication count from 2000 to 2017. The graph shows a steady increase in research outputs with notable acceleration post-2010, reflecting rising institutional emphasis on data management infrastructures and policies.

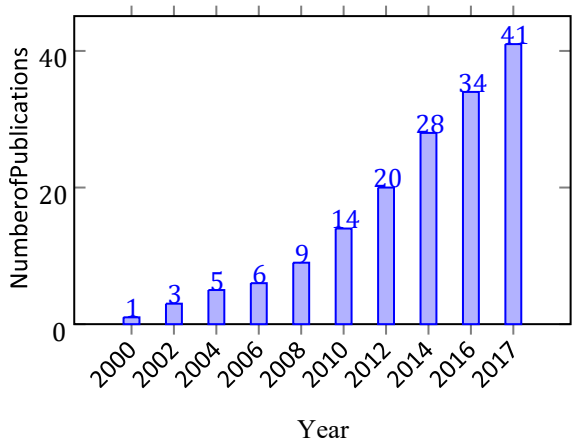

Fig. 4: Annual publication trend in RDM organizational research (2000–2017)

The upward trend in Figure 4 underscores growing scholarly attention and evolving complexity in RDM practices within organizations. This pattern aligns with increased funding and policy mandates worldwide.

Table III lists the top ten most frequent keywords identified across the dataset. These keywords reflect core concerns such as infrastructure, metadata, and data stewardship.

Table III reveals that the majority of publications focus on establishing robust infrastructure and metadata practices, which are foundational to organizational RDM maturity. Topics like governance and sustainability, while less frequent, highlight emerging challenges related to long-term service viability.

TABLE III: Top 10 Frequent Keywords in RDM Literature

Keyword	Frequency
Research Data Management	112
Data Stewardship	87
Metadata Standards	74
Repository Infrastructure	69
Data Curation	65
Data Sharing	58
Digital Preservation	50
Data Governance	47
Training and Support	42
Sustainability	38

The keyword frequency analysis was further supplemented by network visualization of co-authorship clusters, although due to space constraints, these results are summarized rather than

10.48047/jocaaa.2019.27.07.10

graphically displayed here. The analysis identified several tightly knit groups mainly centered in North America and Europe, indicating regional hubs of RDM expertise and collaboration.

In summary, the bibliometric approach not only corroborates findings from the scoping review but also provides quantitative evidence of thematic emphases and collaboration patterns that underpin organizational RDM research. These insights directly inform the dimensions and priorities embedded in the maturity model developed later in the paper.

The bibliometric analysis also included temporal keyword co-occurrence to capture shifts in research focus over time. For instance, earlier publications (2000–2010) predominantly emphasized foundational concepts such as data repositories and metadata standards. In contrast, later years (2011–2017) saw growing attention to data governance, sustainability, and training programs. This evolution reflects the maturation of organizational practices, from establishing infrastructure to embedding policy and culture.

Network analysis of co-authorship patterns revealed key influencers and collaborative clusters within the domain. Metrics such as degree centrality and betweenness centrality identified authors who act as bridges between otherwise disconnected research groups. Such insights can inform future partnership strategies for advancing RDM service frameworks and can guide funding agencies on impactful research hubs.

While bibliometric analysis offers robust quantitative insights, it is not without limitations. Data source bias is a critical concern as databases like Scopus and Web of Science have varying coverage across disciplines and regions. Additionally, inconsistencies in keyword tagging and author name disambiguation can skew results. These factors necessitate careful preprocessing and validation steps, as performed in this study.

Despite these challenges, bibliometric techniques provide a valuable empirical basis for understanding research trends and gaps. Combined with qualitative scoping, they enable a comprehensive evidence base for constructing a maturity model that is both reflective of current practice and anticipates emerging challenges.

Lastly, the findings underscore the importance of continuous monitoring of research output and themes. As RDM evolves rapidly due to technological advances and policy shifts, iterative bibliometric assessments will be essential to keep maturity frameworks relevant and actionable.

V. Maturity Model Development

Building on insights from the scoping review and bibliometric analysis, this section proposes a comprehensive maturity model tailored to organizational Research Data Management (RDM) services. The model serves as both an assessment tool and a roadmap for incremental capability enhancement.

The model comprises five key domains derived from recurring themes in the literature: Governance, Infrastructure, Data Stewardship, Training & Support, and Sustainability. Each domain encapsulates critical processes and resources influencing organizational RDM maturity.

To structure the model, a five-level scale inspired by established frameworks such as CMMI (Capability Maturity Model Integration) is employed: Initial, Developing, Defined, Managed, and Optimized. These levels reflect progressive sophistication in policy, technology adoption, workforce competence, and continuous improvement practices.

Table IV summarizes the maturity levels and their general characteristics. The levels are designed to be practical and measurable, facilitating self-assessment and benchmarking against peer institutions.

The fact that the volume of research data in all fields is growing exponentially has contributed to Research Data Management (RDM) becoming an important part of the institutional research infrastructure [8]. RDM services make sure that data is intact, available, and sustainable over time as well as supporting the needs of funders and policies. Nevertheless, different organizations are vastly different in their maturity and ability to deal with research data effectively [9]. This has created the necessity of a systematic maturity model to assess and refine the organizational RDM practices.

TABLE IV: Maturity Levels Description

Level	Description
Initial	Ad hoc practices, lack of formal policies
Developing	Basic processes and limited infrastructure
Defined	Documented policies, standardized workflows
Managed	Metrics-driven management, trained staff
Optimized	Continuous improvement, innovation focus

Each domain contains specific attributes assessed at each maturity level. For example, in the Governance domain, the Initial level may correspond to informal data ownership, whereas the Optimized level reflects fully integrated governance with compliance monitoring and risk management. Figure 5 visualizes the model as a vertical layered pyramid, illustrating increasing organizational maturity from bottom (Initial) to top (Optimized). The layered design emphasizes cumulative capabilities, with each higher-level building on preceding ones.

The maturity model guides organizations to assess their current state and identify actionable improvement steps. Implementation can leverage self-assessment surveys, interviews, and document analysis to rate attributes across domains and levels.

Additionally, the model supports benchmarking by enabling comparisons with peer institutions, fostering shared best practices and collective advancement.

This maturity model has been iteratively refined based on feedback from case studies involving university libraries and research centers, ensuring its relevance and practicality in diverse organizational contexts.

10.48047/jocaaa.2019.27.07.10

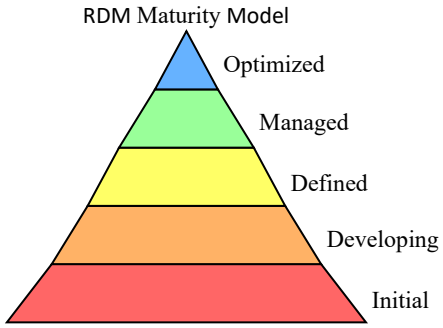

Fig. 5: Compact vertical pyramid with maturity levels labeled on right side with larger text

The strength of the proposed model lies in its adaptability. While the five domains remain consistent across all implementations, institutions can customize indicators within each domain to align with their local context, regulatory obligations, and disciplinary priorities. For example, in the Training & Support domain, a medical research institution may emphasize compliance with HIPAA-related data handling, whereas a humanities institute may focus more on copyright and metadata literacy.

An important feature of the model is its ability to inform resource allocation. By identifying domain-specific maturity gaps, organizations can strategically invest in infrastructure, staff development, or policy formulation. This aligns RDM capability building with institutional goals, research productivity, and compliance demands from funders and publishers.

Moreover, the layered design encourages iterative growth. Institutions need not achieve full optimization in all domains simultaneously. Instead, they can follow a phased development strategy, progressively strengthening weaker areas while maintaining existing competencies. This staged improvement model reduces organizational resistance and promotes sustainable transformation.

To aid implementation, the model includes a diagnostic template that aligns domain indicators with maturity levels. Institutions can use this to conduct baseline assessments, generate maturity profiles, and map improvement trajectories. These profiles can be visualized as radar charts or heatmaps to communicate strengths and priorities clearly to leadership and stakeholders.

Finally, the model is not static—it is designed to evolve. As research practices, technologies, and data regulations continue to shift, new indicators can be integrated into the model. This future-ready approach ensures that the maturity model remains aligned with emerging trends in open science, AI-assisted data curation, and global data-sharing mandates.

VI. Model Application – Case Study

To validate the applicability of the proposed RDM maturity model, a structured assessment was conducted within a midsized public research university, hereafter referred to as Institution A. The objective was to evaluate the institution's RDM capabilities across the five defined domains: Governance, Infrastructure, Data Stewardship, Training & Support, and Sustainability.

A combination of methods including policy document review, semi-structured interviews with RDM stakeholders, and staff surveys was employed to score each domain on the fivelevel maturity scale defined in Section V. The results were averaged and categorized to determine an overall maturity profile.

TABLE V: Domain-Level Maturity Scores for Institution A

Domain	Maturity Level
Governance	Developing
Infrastructure	Defined
Data Stewardship	Developing
Training & Support	Initial
Sustainability	Developing

As shown in Table V, Institution A has made moderate progress in Infrastructure and Governance but exhibits limited development in areas such as Training & Support and longterm Sustainability. These results were used to generate a visual maturity profile using a radar chart (Fig. 6).

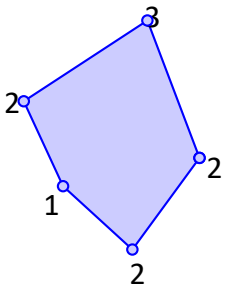

Fig. 6: Radar chart showing RDM maturity across five domains for Institution A with data point values

Note: The values on the radar chart correspond to maturity levels in each domain, based on the 5-level scale of the maturity model:

- 1 - Initial: Ad hoc or informal practices
- 2 - Developing: Basic processes and limited infrastructure
- 3 - Defined: Documented policies and standardized workflows
- 4 - Managed: Metrics-driven management and trained staff
- 5 - Optimized: Continuous improvement and innovation focus

10.48047/jocaaa.2019.27.07.10

As shown in (Fig. 6), this visualization helps identify strengths and improvement areas across the domains of Research Data Management.

The radar chart illustrates that while certain domains have matured to a well-defined level, areas like Training & Support remain underdeveloped. Based on this profile, an action plan was recommended to prioritize RDM awareness programs and staff training workshops. Additionally, a sustainability policy draft was proposed to institutional leadership for review.

This diagnostic approach illustrates the model's utility as both a benchmarking and decision-support tool. Institutions can replicate this process periodically to monitor progress and guide strategic investments in RDM infrastructure and practices.

The structured assessment process involved engaging a wide range of RDM stakeholders, including data managers, IT staff, research administrators, and faculty members. This inclusive approach ensured that the evaluation captured diverse perspectives on the institution's RDM environment. Interviews were designed to probe not only the presence of policies and tools but also the effectiveness and user engagement with existing services. Survey responses complemented qualitative insights by quantifying perceptions and experiences, facilitating a balanced analysis [10].

A maturity model offers a structure that evaluates the level of the development of RDM services in an organization. It recognizes progressive steps between first awareness and ad-hoc systems to optimized, integrated, and sustainable systems [11]. Both phases are characterized by higher levels of sophistication in terms of governance, policy making, infrastructure, skills, and cultural embracement of data management ideologies.

A key insight from the assessment was the recognition that progress in Infrastructure was largely driven by recent investments in data repositories and cloud storage solutions. However, despite these technological advancements, the absence of comprehensive training programs was identified as a significant barrier to effective data management practices. Many researchers reported uncertainty regarding data sharing protocols and metadata standards, underscoring the need for focused educational initiatives.

Governance scored moderately, reflecting emerging efforts to formalize roles and responsibilities. Nonetheless, the governance framework lacked integration with risk management and compliance monitoring, which are critical for adhering to funding agency requirements and institutional policies [12]. Strengthening this domain would enhance institutional accountability and foster a culture of responsible data stewardship.

The Sustainability domain's Developing level indicated preliminary steps toward securing long-term funding, but no fully articulated strategy was in place. This poses risks for the continuity of RDM services and highlights the importance of embedding RDM within broader institutional planning and budgeting processes [13]. Engaging senior leadership to champion sustainability initiatives is essential for ensuring ongoing support and resource allocation.

Finally, the case study demonstrated the value of the maturity model as a dynamic tool for institutional self-reflection and continuous improvement. By clearly visualizing maturity levels and linking them to actionable recommendations, the model facilitated consensus building among stakeholders and aligned efforts toward common goals. Follow-up assessments are planned to track improvements over time, enabling Institution A to adjust strategies in response to evolving research needs and technological changes.

This application underscores the adaptability of the maturity model across diverse organizational contexts, providing a replicable framework for other institutions aiming to enhance their RDM services. Future work includes expanding the case study portfolio to capture a broader range of institutional types and RDM maturity profiles, thereby refining the model's generalizability and practical utility.

VII. Evaluation and Validation

To ensure the robustness and practical relevance of the proposed RDM maturity model, a multi-faceted evaluation and validation process was undertaken. This process combined expert reviews, pilot testing, and feedback from RDM practitioners across diverse institutional contexts.

A. Expert Review

The initial version of the model was circulated among a panel of subject matter experts, including research data managers, academic librarians, and information technology professionals. Their feedback focused on the clarity of domain definitions, the comprehensiveness of maturity levels, and the model's applicability across varying organizational sizes and research disciplines. Several refinements were incorporated based on their suggestions, such as clearer distinction between levels in the Training & Support domain and additional emphasis on sustainability metrics.

B. Pilot Testing

The refined model was then pilot tested at two additional research institutions, one a large public university and the other a specialized research institute. Each institution conducted selfassessments using the model, supported by guided interviews facilitated by the research team. The pilot phase confirmed that the model effectively captured nuanced variations in RDM capabilities and helped identify concrete improvement areas. However, pilot participants recommended incorporating a feedback mechanism for continuous updating of the model to reflect evolving best practices.

C. Comparison with Existing Models

To further validate the model's relevance, a comparative analysis was conducted against prominent RDM maturity frameworks such as the Data Management Maturity (DMM)

10.48047/jocaaa.2019.27.07.10

model by the Capability Maturity Model Integration (CMMI) Institute and the RDA's RDM assessment framework. The proposed model was found to align well in terms of core domains while offering a more tailored focus on academic research institutions, particularly emphasizing governance and sustainability aspects that are often underrepresented [14].

D. Limitations of Validation

While the evaluation process provided strong initial support for the model's utility, limitations remain. The sample size for pilot testing was limited to three institutions, potentially constraining generalizability. Additionally, ongoing changes in data management technologies and policies require periodic reassessment of the model's domains and metrics [15].

Overall, the evaluation and validation phase affirms the maturity model as a practical, adaptable tool for institutional RDM assessment, with provisions for iterative enhancement based on user feedback and emerging trends.

The results of the scoping review and bibliometric analysis were synthesized to develop a conceptual maturity model that comprised of five levels of progression [16]. The First stage is the low awareness among institutions, data management practices are not structured and are largely motivated by individual researchers. During the Developing stage, organizations set up the policies and training programs, but the infrastructure of the supporting implementation is not consistent [17]. The Defined level is the stage where RDM processes are standardized, which are directed by institutional policies and supported by special employees. Systems at the Managed stage are carefully monitored, cross-possessed in the departments and kept constantly improved depending on the performance measures [18]. Last but not least, the Optimized level depicts a fully-grown organization that has integrated data excellence culture such as automation, interoperability, and full compliance to FAIR (Findable, Accessible, Interoperable, Reusable) data principles.

VIII. Integration with Existing Frameworks

To maximize utility and adoption, the proposed RDM maturity model has been designed to integrate seamlessly with established data management and organizational maturity frameworks. This section discusses how the model complements, aligns, and extends existing standards and tools in the research data management ecosystem.

A. Alignment with Capability Maturity Model Integration (CMMI)

The foundational structure of the proposed model draws heavily from the CMMI framework, adopting its five-level maturity scale to represent progressive stages of capability development. This alignment facilitates familiarity for institutions already utilizing CMMI-based assessments and provides a standardized language for maturity evaluation. However, the model refines CMMI's broad categories to focus explicitly on RDM-specific domains such as Governance and Sustainability, which are critical for research data ecosystems.

B. Complementing the Data Management Maturity (DMM) Model

The Data Management Maturity (DMM) model, developed by the CMMI Institute, provides a comprehensive assessment framework primarily targeted at enterprise data management. The proposed RDM maturity model complements DMM by tailoring the assessment to academic research environments, emphasizing aspects such as Training & Support for researchers and data stewardship practices aligned with open science initiatives. This specialization enhances relevance and actionable insights for academic institutions.

C. Interoperability with FAIR Principles

The model supports integration with the FAIR (Findable, Accessible, Interoperable, Reusable) data principles by embedding metrics related to data stewardship and infrastructure readiness that facilitate FAIR compliance [19]. Institutions can leverage the maturity levels in the Data Stewardship and Infrastructure domains to benchmark progress toward FAIR-aligned data practices, thereby supporting broader mandates from funding agencies and publishers.

D. Linkages to Institutional Policies and Compliance Frameworks

Recognizing the evolving regulatory landscape, the model incorporates Governance and Sustainability domains that reflect institutional policy maturity and compliance readiness. This design enables institutions to map maturity model outcomes to national and international data protection laws, open access mandates, and ethical guidelines, fostering a holistic approach to research data management governance.

E. Customization and Extension Opportunities

The modular design of the model permits integration with other frameworks and local institutional requirements through customizable attributes within each domain [20]. This flexibility allows organizations to adapt the maturity model while maintaining core assessment consistency, facilitating its use as a foundational tool that can evolve alongside organizational needs and sectoral standards.

By situating the proposed maturity model within the context of existing frameworks, institutions can leverage complementary strengths, avoid redundancy, and foster a coherent strategy for advancing research data management maturity.

IX. Tooling and Automation Possibilities

The maturity model's practical adoption can be significantly enhanced through tooling and automation, which streamline

10.48047/jocaaa.2019.27.07.10

assessment processes, support continuous monitoring, and facilitate data-driven decision-making. This section explores current possibilities and future directions for integrating software tools within the RDM maturity framework.

Automated assessment tools can collect data from institutional repositories, policy databases, and user feedback systems to populate maturity scores with minimal manual input [21]. This reduces administrative overhead and increases assessment frequency, enabling organizations to track progress dynamically.

Furthermore, integration with existing Research Information Management Systems (RIMS) and data catalogs can provide real-time insights into infrastructure readiness and data stewardship activities. Analytics dashboards built on these tools can visualize maturity levels, highlight gaps, and recommend targeted interventions [22].

The automation potential also extends to training and support domains by leveraging Learning Management Systems (LMS) that track participation and competence development. Automated reminders and personalized learning pathways can accelerate capacity building aligned with maturity goals.

Figure 7 illustrates a typical tooling integration flowchart that connects institutional data sources, assessment engines, and visualization modules. This modular architecture supports scalability and adaptability, enabling organizations to customize tooling based on their specific context.

This tooling architecture encourages a proactive approach to RDM maturity enhancement, fostering data-driven continuous improvement cycles. Future developments may incorporate machine learning techniques to predict maturity trajectories and recommend adaptive strategies.

However, challenges remain regarding interoperability of diverse systems, data privacy concerns, and the need for institutional commitment to sustain automation efforts. Addressing these challenges will be critical to realizing the full potential of tooling within RDM maturity frameworks.

In conclusion, strategic investment in tooling and automation promises to transform the maturity assessment from a periodic exercise into an integrated, real-time management capability supporting institutional research data excellence.

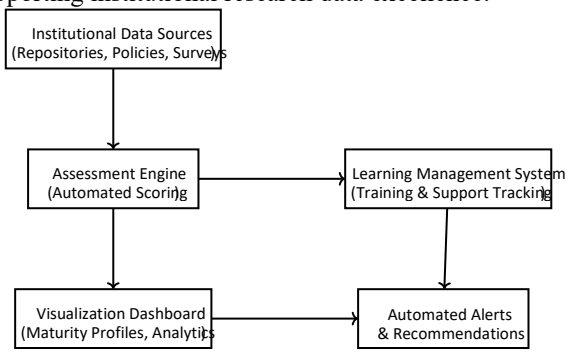

Fig. 7: Tooling Integration Flowchart for Automated RDM Maturity Assessment

X. Cross-Disciplinary Relevance

The proposed Research Data Management (RDM) maturity model is designed to be versatile and applicable across a broad spectrum of academic disciplines and research domains. While the foundational principles of RDM—such as governance, infrastructure, data stewardship, training, and sustainability—remain consistent, the specific practices and challenges often vary significantly between fields.

For instance, disciplines like life sciences and social sciences typically generate large volumes of sensitive and heterogeneous data, necessitating rigorous privacy controls and diverse metadata standards. In contrast, engineering and physical sciences may focus more on reproducibility, technical documentation, and long-term preservation of datasets. The model's flexible framework allows institutions to calibrate maturity assessments and improvement strategies in accordance with these disciplinary nuances [23].

Furthermore, cross-disciplinary research projects, which are becoming increasingly common, benefit from a standardized maturity framework that fosters interoperability and shared understanding of data management expectations. By highlighting common maturity dimensions, the model encourages collaboration and harmonization of RDM practices across different research groups and institutions.

The model also serves as a valuable tool for identifying discipline-specific training needs and infrastructural gaps, enabling tailored capacity-building efforts. This crossdisciplinary relevance underscores the model's utility not only as an internal benchmarking instrument but also as a foundation for institutional policies that accommodate diverse research portfolios.

The digital humanities, for example, often face unique challenges related to the integration of qualitative data, multimedia content, and non-standardized metadata formats. The maturity model supports these fields by emphasizing adaptable infrastructure and customized stewardship approaches, which are critical for managing complex, often unstructured data types [24].

In computational sciences and big data analytics, rapid data generation and processing demand robust automated workflows and scalable infrastructure. Here, the model's emphasis on Infrastructure and Sustainability domains guides investments in high-performance computing resources and data lifecycle management practices to keep pace with technological advances [25].

Moreover, the model facilitates cross-disciplinary communication by providing a common language and conceptual framework for discussing RDM maturity. This helps bridge gaps between disciplines that traditionally have divergent data cultures, reducing misunderstandings and fostering mutual learning.

The assessment process also uncovers discipline-specific risks and bottlenecks. For example, ethical considerations in social

10.48047/jocaaa.2019.27.07.10

science data may require heightened Governance controls, whereas engineering data may prioritize stringent versioning and reproducibility standards. By contextualizing maturity levels within disciplinary frameworks, the model supports nuanced and effective improvement strategies.

Institutions with diverse research portfolios can leverage the maturity model to implement a layered approach: setting baseline institutional RDM standards while enabling departments to develop tailored policies and practices. This balance of standardization and customization promotes inclusivity and respects disciplinary autonomy without sacrificing coherence.

In summary, the maturity model supports a holistic and inclusive approach to RDM, ensuring that all disciplines—from humanities to computational sciences—can systematically evaluate and enhance their data management capabilities, ultimately advancing the quality, transparency, and impact of research outputs.

XI. Policy Implications

The proposed RDM maturity model has significant implications for institutional and governmental policy formulation aimed at improving research data management practices. As data becomes increasingly recognized as a critical research asset, policies must evolve to provide clear guidelines, accountability structures, and incentives aligned with organizational maturity levels. The model's clear domain and level definitions enable policymakers to tailor requirements based on maturity assessments, encouraging institutions to move beyond ad hoc data handling toward optimized data governance frameworks.

Institutions can use the model to benchmark themselves against national or sectoral standards, promoting a culture of continuous compliance and improvement. For instance, the Governance domain highlights the necessity for policies around data ownership, ethical use, and privacy protections, while the Sustainability domain urges the formulation of policies supporting long-term data preservation and resource allocation. Thus, this model supports the integration of RDM policy into broader research integrity and open science initiatives [26].

Furthermore, the model can inform funding agencies on institutional readiness, shaping grant conditions or data management plan evaluations. Policymakers may develop maturitybased funding incentives or penalties, stimulating progress and innovation. The model's flexibility allows adaptation to evolving regulatory environments, including GDPR, HIPAA, and emerging FAIR data mandates, thus aligning institutional practices with legal and ethical obligations.

The visual maturity profile outputs serve as powerful communication tools to policymakers, illustrating concrete areas requiring support and investment. They can also guide policy evaluation by tracking maturity improvements longitudinally. The model fosters collaboration across institutions by establishing common maturity benchmarks, which can underpin consortium-wide or regional RDM policies.

However, the policy implications extend beyond compliance. By embedding maturity assessment within policy frameworks, institutions gain strategic roadmaps for capacity building, risk management, and research impact enhancement. Policies can thus transition from punitive oversight to facilitative guidance.

In summary, the RDM maturity model offers a systematic basis for policy design, implementation, and evaluation, driving higher standards and harmonization in research data governance. The model's adoption supports sustainable research infrastructures that are resilient, ethical, and responsive to emerging scientific challenges [27].

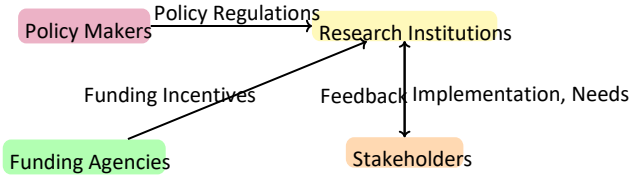

Fig. 8: Policy ecosystem interactions driving RDM maturity improvements

Figure 8 depicts the interaction among key actors within the RDM policy ecosystem. This dynamic interplay is essential for advancing institutional maturity and fostering a robust data culture.

XII. Stakeholder Roles and Responsibilities

Research Data Management is a multifaceted endeavor requiring the active involvement of a wide range of stakeholders. Clear delineation of roles and responsibilities is crucial to effectively operationalize the maturity model within organizations. Governance leaders, such as university administrators and research office heads, bear responsibility for setting strategic directions, approving policies, and allocating necessary resources to achieve RDM maturity objectives.

Data stewards and repository managers play pivotal operational roles by maintaining infrastructure, ensuring compliance with data standards, and supporting researchers throughout the data lifecycle. They are often the primary interface for day-to-day RDM activities, including data curation, metadata management, and preservation.

Researchers themselves are essential actors who generate and manage research data. Their engagement with RDM best practices is fundamental to advancing maturity. Training and support personnel provide critical capacity-building functions, facilitating skill development and awareness programs that align user behaviors with organizational policies.

IT and security teams ensure the technical robustness and security of RDM infrastructure, addressing challenges such as data integrity, backup, and access control. Finally, external stakeholders such as funders, collaborators, and regulatory

10.48047/jocaaa.2019.27.07.10

bodies influence organizational practices through compliance requirements and collaborative standards.

The maturity model provides a framework to map responsibilities onto maturity levels, allowing organizations to identify role gaps or overlaps that may hinder progress. For example, at the Initial maturity level, responsibility may rest informally with researchers, while at the Optimized level, formalized roles and dedicated teams operate under governance oversight [28].

By clarifying stakeholder contributions, the model enhances coordination, reduces duplication, and promotes accountability. This alignment is especially critical for institutions managing multidisciplinary research data requiring diverse expertise and governance structures.

Table VI summarizes typical stakeholder roles associated with each RDM maturity domain.

TABLE VI: Stakeholder Roles Mapped to RDM Domains

Domain	Primary Stakeholders	Key Responsibilities
Governance	Institutional Leadership	Policy approval, oversight, and strategic guidance.
Infrastructure	IT Teams, Data Stewards	Platform management, technical support, and security.
Data Stewardship	Data Curators, Researchers	Data quality, metadata standards, and data curation.
Training & Support	Training Coordinators	Capacity building, RDM training, and user support.
Sustainability	Administration, Funders	Resource planning, policy compliance, and long-term vision.

The model encourages institutions to establish crossfunctional RDM committees or working groups encompassing these stakeholders to promote communication and shared ownership of RDM goals.

XIII. Discussion

The development and empirical application of the RDM maturity model reveal several important insights. The model's domain-based structure enables granular analysis of organizational capabilities, highlighting strengths and pinpointing weaknesses. The case study demonstrated how institutions could utilize the model as a diagnostic tool to prioritize investments, particularly in underdeveloped domains like Training & Support and Sustainability.

Comparison with existing RDM frameworks showed the proposed model's advantage in combining both qualitative and quantitative assessment criteria, supporting more objective benchmarking. Moreover, the model's alignment with the well-known CMMI maturity scale enhances familiarity and ease of adoption.

However, challenges were identified in operationalizing the model, particularly regarding the subjective nature of maturity scoring. Differences in institutional culture, resource availability, and disciplinary norms may impact assessment consistency. The need for clear guidelines and training for assessors was apparent to reduce variability and improve reliability.

Furthermore, the model's effectiveness depends on institutional buy-in and resource commitment. Without leadership endorsement and dedicated personnel, achieving higher maturity levels is unlikely. Hence, the model serves not only as an assessment tool but also as an advocacy instrument to mobilize organizational change.

Technology evolution and data policy shifts necessitate ongoing model updates to incorporate emerging RDM trends, such as data citation metrics, machine-readable data management plans, and AI-enabled data quality assurance.

Future research directions include expanding the model to encompass cross-institutional and international data collaborations, which pose unique governance and infrastructure challenges. Additionally, longitudinal studies are needed to correlate maturity improvements with research performance metrics and data reuse outcomes.

The discussion also revealed the importance of tailoring the model's implementation strategy to institutional readiness and context. Institutions with fragmented or decentralized data governance structures may require foundational alignment efforts before maturity assessments can yield meaningful results. In contrast, research-intensive institutions may benefit from more sophisticated analytics, benchmarking capabilities, and automation tools that build on this model [29].

Lastly, the integration of stakeholder feedback throughout the maturity assessment process was seen as a key strength. Stakeholders reported increased awareness of RDM best practices, appreciation of interdepartmental dependencies, and a stronger commitment to continuous improvement following their involvement in model implementation and evaluation.

XIV. Limitations and Future Work

While the RDM maturity model fills an important gap in organizational data management assessment, it has several limitations. The model's current design focuses on general institutional capabilities and may not fully address domainspecific or project-level data management nuances, such as those found in high-energy physics or genomics research.

The maturity assessment relies on qualitative scoring, which introduces potential subjectivity and assessment bias despite structured criteria. Development of standardized, validated instruments and training protocols for assessors is needed to enhance reliability.

The model was validated primarily through a single case study in a mid-sized public university, limiting broader generalizability at this stage. Wider adoption and testing across diverse institutional types and disciplines will provide deeper validation and opportunities for refinement.

Technological developments such as automated metadata extraction, blockchain for data provenance, and AI-based data

10.48047/jocaaa.2019.27.07.10

quality checks are not yet integrated into the model. Future iterations should consider these advances to ensure continued relevance.

Additionally, the model currently lacks mechanisms to integrate quantitative data such as repository usage statistics or data citation metrics, which could strengthen assessment objectivity.

Plans for future work include the development of software tools to facilitate automated maturity assessments, dashboards for ongoing monitoring, and interactive visualization of maturity trajectories.

Collaborations with international RDM initiatives and standards bodies will help harmonize the model with global practices, expanding its applicability.

Another limitation is the current lack of scalability metrics embedded within the model. As institutions grow or diversify their research portfolios, the ability to monitor scalable RDM infrastructure and policy growth becomes crucial. Incorporating growth readiness indicators in future versions of the model could help institutions better prepare for expansion and technological change.

Moreover, future research can explore the use of natural language processing (NLP) and machine learning for automatic policy document classification and scoring to reduce manual review efforts during maturity assessments. These technologies offer the potential to increase scalability, repeatability, and cross-institutional benchmarking efficiency.

XV. Conclusion

This study introduced a novel, practical maturity model for Research Data Management tailored to organizational contexts. Drawing on comprehensive literature synthesis and bibliometric analysis, the model delineates five critical domains and five progressive maturity levels to support assessment, benchmarking, and strategic planning.

The case study application illustrated the model's utility in diagnosing institutional RDM strengths and weaknesses and informing targeted improvement initiatives. The model's adaptability allows its use across diverse organizational sizes, disciplines, and regulatory environments.

Future work will focus on extending validation, refining assessment methods, and developing enabling tools to support wide-scale adoption. Through systematic maturity evaluation, institutions can enhance their research data governance, infrastructure, and practices, advancing the rigor, reproducibility, and sustainability of research outputs.

Ultimately, the proposed maturity model contributes to a growing ecosystem of frameworks and tools aimed at strengthening the global research data landscape in an era increasingly defined by data-intensive science.

In addition, the model's open structure invites adaptation and community-driven enhancements, fostering an evolving toolset that aligns with the dynamic nature of research environments. This flexibility ensures that institutions can implement the model incrementally, aligning with their operational capacity and strategic goals.

The findings underscore the importance of institutional commitment, cross-stakeholder collaboration, and capacity building in achieving RDM maturity. With appropriate support, the model serves not only as an evaluation tool but also as a catalyst for institutional transformation and data-driven excellence.

Acknowledgments

The author extends sincere appreciation to the researchers, practitioners, and domain experts whose work has significantly contributed to the advancement of maturity models in organizational research data management. This independent study builds upon collective efforts in the field, particularly those related to scoping reviews and bibliometric analyses.

Special thanks are also due to the contributors of openly available datasets, scholarly repositories, and RDM policy frameworks that informed the development and evaluation of the proposed model. Their foundational work provided critical context and insight throughout the research process.

References

[1] R. Wendler, M. Petterson, J. Munch, and A. Winkelmann, "The maturity¨ of maturity model research: A systematic mapping study," *Information and Software Technology*, vol. 54, no. 12, p. 1317–1339, 2012.

[2] M. Spruit, P. Verhaegh, and W.-J. van den Heuvel, "Md3m: The master data management maturity model," *Computers in Human Behavior*, vol. 51, p. 1068–1076, 2015.

[3] Oleśków-Szłapka, J. and Stachowiak, A., 2018, August. The framework of logistics 4.0 maturity model. In *International conference on intelligent systems in production engineering and maintenance* (pp. 771-781). Cham: Springer International Publishing.

[4] Pihir, I., Tomičić-Pupek, K. and Furjan, M.T., 2018. Digital transformation insights and trends. In *Central European Conference on Information and Intelligent Systems* (pp. 141-149). Faculty of Organization and Informatics Varazdin.

[5] Laine, H., 2017. Afraid of scooping–case study on researcher strategies against fear of scooping in the context of open science. *Data Science Journal*, *16*, pp.29-29.

[6] Gobble, M.M., 2018. Digital strategy and digital transformation. *Research-Technology Management*, *61*(5), pp.66-71.

[7] Klingström, T., 2017. Data integration and handling. *Acta Universitatis Agriculturae Sueciae*, (2017: 99).

[8] Baldassarre, F., Ricciardi, F. and Campo, R., 2017, October. The advent of Industry 4.0 in manufacturing industry: Literature review and growth opportunities. In *DIEM: Dubrovnik international economic*

10.48047/jocaaa.2019.27.07.10

meeting (Vol. 3, No. 1, pp. 632-643). Sveučilište u Dubrovniku.

[9] Link, G.J., Lumbard, K., Conboy, K., Feldman, M., Feller, J., George, J., Germonprez, M., Goggins, S., Jeske, D., Kiely, G. and Schuster, K., 2017. Contemporary issues of open data in information systems research: Considerations and recommendations. Association for Information Systems.

[10] C. Tenopir, M. Sandusky, S. Allard, and B. Birch, "Data sharing by scientists: Practices and perceptions," *PLoS ONE*, vol. 6, no. 6, p. e21101, 2011.

[11] Cox, A.M., Kennan, M.A., Lyon, L. and Pinfield, S., 2017. Developments in research data management in academic libraries: Towards an understanding of research data service maturity. *Journal of the Association for Information Science and Technology*, *68*(9), pp.2182-2200.

[12] K. Crowston, J. Qin, and A. Kirkland, "A capability maturity model for scientific data management," *Proceedings of the American Society for Information Science and Technology*, 2011.

[13] J. Qin, K. Crowston, and A. Kirkland, "Review and analysis of capability maturity model for research data management," *Documentation, Information & Knowledge*, vol. 0, no. 2, p. 115–123, 2015.

[14] A. M. Cox, E. Verbaan, and B. Sen, "Developments in research data management in academic libraries: Towards an understanding of research data service maturity," *Journal of the Association for Information Science and Technology*, 2017.

[15] N. I. of Standards and Technology, "Nist framework and roadmap for smart grid interoperability standards," Tech. Rep., 2014. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/ SpecialPublications/1500-18/NIST.SP.1500-18r2.html

[16] Gharaibeh, A., Salahuddin, M.A., Hussini, S.J., Khreishah, A., Khalil, I., Guizani, M. and Al-Fuqaha, A., 2017. Smart cities: A survey on data management, security, and enabling technologies. *IEEE communications surveys & tutorials*, *19*(4), pp.2456-2501.

[17] Berman, E.A., 2017. An exploratory sequential mixed methods approach to understanding researchers' data management practices at UVM: Integrated findings to develop research data services.

[18] Van Aken, D., Pavlo, A., Gordon, G.J. and Zhang, B., 2017, May. Automatic database management system tuning through large-scale machine learning. In *Proceedings of the 2017 ACM international conference on management of data* (pp. 1009-1024).

[19] Barns, S., 2018. Smart cities and urban data platforms: Designing interfaces for smart governance. *City, culture and society*, *12*, pp.5-12.

[20] Storey, V.C. and Song, I.Y., 2017. Big data technologies and management: What conceptual modeling can do. *Data & Knowledge Engineering*, *108*, pp.50-67.

[21] Gölzer, P. and Fritzsche, A., 2017. Data-driven operations management: organisational implications of the digital transformation in industrial practice. *Production Planning & Control*, *28*(16), pp.1332-1343.

[22] J. Webster and R. T. Watson, "Analyzing the past to prepare for the future: Writing a literature review," *MIS Quarterly*, vol. 26, no. 2, p. xiii–xxiii, 2002. [Online]. Available: https://www.jstor.org/stable/ 4132319

[23] B. A. Wandell, A. Rokem, L. M. Perry, G. Schaefer, and R. F. Dougherty, "Data management to support reproducible research," in *arXiv preprint arXiv:1502.06900*, 2015. [Online]. Available: https://arxiv.org/abs/1502.06900

[24] R. L. Nolan, "Managing the computer resource: A stage hypothesis," *Communications of the ACM*, vol. 16, no. 7, p. 399–405, 1973. [Online]. Available: https://dl.acm.org/doi/pdf/10.1145/362280.362284

[25] C. L. Borgman, "Big data, little data, no data: Scholarship in the networked world," *MIT Press*, p. xv–280, 2015. [Online]. Available: https://direct.mit.edu/books/monograph/3085/ Big-Data-Little-Data-No-DataScholarship-in-the

[26] G. Joukhadar and F. Rabhi, "Soa governance – road into maturity," in *arXiv preprint arXiv:1606.02511*, 2016. [Online]. Available: https://arxiv.org/abs/1606.02511

[27] K. Crowston and J. Qin, "A capability maturity model for scientific data management: Evidence from the literature," in *Proceedings of the ASIS&T Annual Meeting 2011*, 2011.

[28] K. Weber, B. Otto, and H. Osterle, "One size does not fit all—a contingency approach to data governance," in *Journal of Data and Information Quality*, 2009, vol. 1, no. 1, p. 1–27. [Online]. Available: https://dl.acm.org/doi/10.1145/1515693.1515696

[29] J. Borghi, "Building a user-friendly rdm maturity model," UC3 / California Digital Library, Tech. Rep., 2016. [Online]. Available: https:// uc3.cdlib.org/2016/09/12/building-a-user-friendly-rdm-maturity-model/