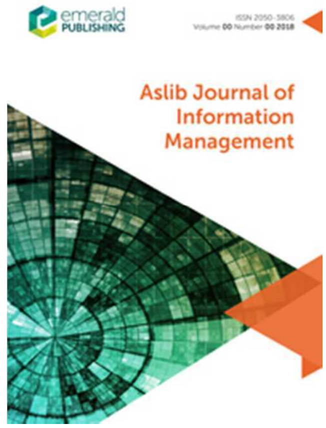
Article:

Cox, A.M. orcid.org/0000-0002-2587-245X and Tam, W. orcid.org/0000-0001-9978-3428 (2018) A critical analysis of lifecycle models of the research process and research data management. Aslib Journal of Information Management, 70 (2). pp. 142-157. ISSN 2050-3806

https://doi.org/10.1108/AJIM-11-2017-0251



A critical analysis of lifecycle models of the research process and research data management

Journal:	*Aslib Journal of Information Management*
Manuscript ID	AJIM-11-2017-0251.R2
Manuscript Type:	Research Paper
Keywords:	research, research data management, lifecycle, visualisation, research process, metaphor

A critical analysis of lifecycle models of the research process and research data management

Abstract

Purpose

Visualisations of research and research related activities including research data management as a lifecycle have proliferated in the last decade. This study offers a systematic analysis and critique of such models.

Design/methodology/approach

A framework for analysis synthesised from the literature presented and applied to nine examples.

Findings

The strengths of the lifecycle representation are to clarify stages in research and to capture key features of project based research. Nevertheless, their weakness is that they typically mask various aspects of the complexity of research, constructing it as highly purposive, serial, unidirectional and occurring in a somewhat closed system. Other types of models such as spiral of knowledge creation or the data journey reveal other stories about research. It is suggested that we need to develop other metaphors and visualisations around research.

Research implications

The paper explores the strengths and weaknesses of the popular lifecycle model for research and research data management, and also considers alternative ways of representing them.

Practical implications

Librarians use lifecycle models to explain service offerings to users so the analysis will help them identify clearly the best type of representation for particular cases. The critique offered by the paper also reveals that because researchers do not necessarily identify with a lifecycle representation, alternative ways of representing research need to be developed.

Originality/value

The paper offers a systematic analysis of visualisations of research and research data management current in the Library and Information Studies (LIS) literature revealing the strengths and weaknesses of the lifecycle metaphor.

Introduction

In the last decade Library and Information Studies (LIS) has shown a growing interest in the detail of the research process, in part due to the creation of a new depth of research support services in libraries (Corrall, 2014). The academic library's role has been turned "inside out" from mainly providing the user community with access to literature, to stewarding the knowledge being created within the institution and making it discoverable to the wider world (Dempsey, Malpas & Lavoie, 2014). This includes an increasing investment in helping local researchers manage data within the research process, stewarding different versions and types of outputs, including data, and also supporting and measuring all sorts of dissemination and impact beyond academia. Thus, libraries are seeking to offer services across the course of research as a whole, and as a consequence the black box of the research process has been opened. In this context a number of commentators and practitioners have commented on the proliferation of lifecycle models/visualisations in the research data management and research support area (Wilson, 2014; L'Hours, 2014; Carlson, 2014). Ball (2012) reviews nine such data lifecycle models; the Committee on Earth Observation Satellites (CEOS) Working Group on Information Systems & Services (2011) reviewed 44 models. There is a significant amount of variation in their purpose and assumptions, but we can learn a lot about how research is conceptualised from examining them.

The lifecycle has long been a favourite model in LIS (Ma and Wang, 2010a,b); in particular it is a core concept in records and archive management (Williams, 2006). The appeal is the temporal dimension the metaphor adds to our understanding of the differing activities in view. The metaphor seems to be particularly appealing in the Research area because it fits into thinking about designing systems workflows, be those administrative or IT-based. Yet the term lifecycle is also ambiguous implying the model of a birth to death journey or, somewhat in contrast, the pattern of birth and reproduction where the cycle is endlessly repeated or, indeed, enters a progressive upward cycle. A number of authors, notably Carlson (2014); Waddington, Green & Awre (2012) and Wissik & Durco (2015) have begun to reflect on these models. But there remain questions about what different types of model there are; what uses they have; and what assumptions are made in them. Given the way that the lifecycle is becoming virtually the default way of representing the research process, it is important at a theoretical level to question whether this adequately conceptualises research. At a practical level, given their increasing use by practitioners to conceptualise and explain services, it is important to weigh up whether they capture how researchers themselves view the research process.

In this context the aim of this paper is to explore the strengths and weaknesses of representing research and research data related activities in a lifecycle model, through a systematic comparison of some that have been published recently in the research data space. It does this by examining the literature to produce an analytic framework; undertaking a systematic analysis of a selection of nine models using the framework; and on this basis reaching some conclusions about the strengths and limits of this type of model/visualisation. Having probed the typical limits of the lifecycle approach, thought can be given to the benefits of radically different representations such as the knowledge spiral or data journey which offer more critical perspectives.

The lifecycle

A number of authors have already attempted to differentiate different types of lifecycle. According to Ma and Wang (2010a, b), the life cycle of information can be understood from two perspectives. The perspective of value focuses on how the worth of information changes, usually deteriorates, over time. The other more common perspective is of management. The authors identify six types of management based life cycle models. The commonest is what they call the chain model. It is a simple chain of steps. The second type, the matrix, expands the chain model by giving more detail on each of the stages in the chain. The third type is the circular model, with the crucial difference being that the end of the chain links back to the beginning thus restarting the cycle. The spiral extends the circular model by illustrating how each cycle is not just a repetition of previous cycles but builds and strengthens, as things such as audiences or the nature and structure of information change. An integrated model combines two of the other types as a way to indicate the complexity of processes or to bring in external factors. The final model is the wave, which is able to express an on-going activity but an overall decline in the value of data. This set of categories itself reveals the complexity of the concept.

Whereas Ma and Wang (2010a,b) emphasise the different visual structure of the models, Carlson (2014) working more directly in the research service area emphasises differing underlying audiences. He suggests that life cycle models in the data service context can be divided into three types:

1. Individual-based, which are for a particular project giving detail on how it unfolds.
2. Organisation-based, which are for showing how services fit to different stages of research or how researchers should access different services at different stages in their work.
3. Community-based, which are for a particular academic community or discipline (including professionals that support them) to define existing or good practice.

This seems to be an insightful categorisation, although, these are not mutually exclusive categories, for example a community based model could include elements of organisational support. He also mixes descriptive and prescriptive elements.

Another way of analysing such models is available in the work of Möller (2013). In the context of seeking to develop a data lifecycle model for the semantic web, he analyses lifecycle models of "data centric systems" such as for multimedia, elearning, digital libraries, knowledge and content management, ontologies and databases. He focuses on data in the context of technical systems, with less sense that they could be socio-technical systems including different social actors, activities etc. However, he does usefully suggest that such systems can vary across the following dimensions:

- Data/metadata – whether data and metadata are differentiated.
- Prescriptive/descriptive – whether the model seeks to define what would be good practice or is describing actual practice.
- Homogenous/heterogeneous – whether the data in the system is similar or of different types. This would seem to be more of a dimension of difference, than a yes no distinction.
- Closed/open – whether the system follows a given set of data, or whether new data can join the system after the beginning of the cycle.
- Centralised/distributed – whether the system is based on a single unitary infrastructure or is it spread over multiple systems.
- Lifecycle type – sequential/incremental/evolving. In a sequential model each step has to be worked through before the next can be started. In an incremental model a new lifecycle can begin before all the expected steps are worked through. In an evolving model several different iterations of the lifecycle can be occurring simultaneously.
- Granularity – whether it is a high level or fine grained model.

Some of these dimensions, such as the first may be specific to Möller's particular study, but most of the others appear to be productive for thinking about research based lifecycle models too. Thus we have the beginning of an analytic framework combining the visual dimension (Ma and Wang, 2010a,b); main purpose (Carlson, 2014); and characteristics of the elements or processes making up the model from Möller (2013).

Method

In order to carry out the study we collected a large number of models published in the last ten years in both the peer reviewed and practitioner LIS literature relating to research and research data. Inevitably this collection was not comprehensive. To represent the range of models being developed the search was not restricted to peer reviewed material, rather it was highly relevant to include examples being published by practitioners. We also made the decision to include models dealing specifically with data and as well as broader models of the research process as a whole, because they appeared to be published to the same audience and are often strongly inter-related. We then selected nine for more in-depth analysis. They were:

1. Research lifecycle (RIN/NESTA, 2010)
2. The research lifecycle at UCF (University of Central Florida Libraries' Research Lifecycle Committee, 2012)
3. The scholarly knowledge cycle (Lyon, 2003)
4. Research Lifecycle enhanced by an "Open Science by Default" Workflow (Grigorov, et al. 2014)
5. The Idealized Scientific Research Activity Lifecycle Model (Patel, 2011)
6. The integrated scientific life cycle of embedded networked sensor research (Pepe et al., 2010)
7. E-science and the lifecycle of research (Humphrey, 2006)
8. Create and Manage Data (Corti et al., 2014)
9. Digital Curation Centre (DCC) curation lifecycle model (Higgins, 2008)

Some such as “E-science and the lifecycle of research” and the “DCC digital curation lifecycle” have been very widely cited in the literature, and we know has been very influential at the practice level too. Some, including the DCC model, have also been well documented in the public domain, thus supporting more in-depth analysis. The “integrated scientific life cycle of embedded networked sensor research” was developed and used in a number of publications by one of the leading researchers in the field, Christine Borgman, so again is considered particularly worthy of consideration. The Open Science by default model reflects current thinking around open science. Thus nine models of research data and research, which appeared to be influential or interesting, were identified for analysis.

Using Ma and Wang (2010a,b), Carlson (2014) and Möller (2013) as starting points we defined a set of key dimensions by which the models could be compared, and through applying them to the sample models refined the criteria. This framework is in three parts: scope and point of view; elements and processes; and visualisation.

Scope and point of view consists of three elements:

1. The subject matter of the lifecycle, which could be anything from the whole research process at one end of the scale to a model just concerned with data, or some aspect of it, at the other.
2. Whether it is project-, organisation- or community-based, after Carlson (2014).
3. Prescriptive/descriptive – whether the intention is to describe a real world state of affairs or define how things should be ideally (Möller, 2013).

Elements and processes consists of:

4. Level of abstraction, i.e., granularity, after Möller (2013).
5. Homogenous/heterogeneous – whether constituent elements are similar or diverse (Möller, 2013).
6. Closed/open – whether it is presented as a bounded, self-contained system or is open to outside elements or processes (Möller, 2013).
7. Centralised/distributed – whether it rests on a single infrastructure or not (Möller, 2013).
8. Uni/multi-directionality – whether the flow of activities goes in a single or multiple directions. This seemed to be an aspect of Möller's (2013) "lifecycle type", but as this was rather complex we thought it more simply articulated through 8 and 9.
9. Seriality/simultaneity – whether activities occur as a set of stages or whether more than one process can occur at once.

Visualisation consists of

10. What type of lifecycle it is, after Ma and Wang (2010a,b).
11. Use of colour, this item and the ones that follow are more granular features of the visual representation that emerged from the data as useful.
12. Visuality/Textuality – the balance of image and text.
13. Recti-/curvilinearity – whether the forms were primarily organic curved forms or more rectilinear ones.

Findings

Description of table

[Insert Table 1 around here]

Table 1 offers a summary of the nine models selected, starting with the broadest, which deals with the whole of the research process, through to the most specific, purely concerned with data curation.

Model 1: Research lifecycle (RIN/NESTA 2010)

The scope of the model is the whole research lifecycle for any discipline. Published in the RIN/NESTA study "Open to all?", the lifecycle was developed to illustrate openness in research, dealing with issues such as what material outputs are made open and to whom, at each stage of the research lifecycle. In terms of elements and processes, it is at a high level of abstraction with the whole research process represented in just 7 stages: Conceptualising and networking; Proposal writing and design; collecting and analysing; Infrastructuring; Documenting and sharing; Publishing and reporting; Translating and engaging. The process is seen as largely closed (with a number of set elements in the system), unidirectional (traveling forward, without any sense of iteration or return to earlier steps) and serial (with a single sequence of steps taken one after the other) – though "infrastructuring" seems to happen in parallel with collecting and analysing. There is no indication that the processes would necessarily be occurring within one particular system, so it can be seen as distributed. Visually the lifecycle type is circular. It is a true circular lifecycle in the sense it is clear that the stage of "translating and engaging" to disseminate research to different communities naturally merges into the "conceptualising and networking" around the next research idea, so one can understand a logic of the cycle repeating itself. A related table lists types of documentary or data outputs created at each stage in the lifecycle; thus it could be seen as a circular matrix within Ma and Wang's (2010a,b) terminology, because it is laid out as a circle, but has a table of explanation for each step in the circle. The low use of colour and text is consistent with a high level of abstraction.

Model 2: The research lifecycle at UCF (University of Central Florida Libraries' Research Lifecycle Committee, 2012)

As with the first model, the scope is the whole research lifecycle for any discipline. In this case it is a lifecycle for a particular organisation, because it is intended to show how different support services fit into the research workflow. This is what Carlson (2014) calls an organisation based model. In terms of elements and processes it is a lot more detailed than the previous model, but remains unidirectional and serial. But it does include more heterogeneous elements of actors and their roles at different stages and appears to be distributed across multiple systems. Using Ma and Wang's terminology (2010a, b) it is an integrated model, incorporating four different circular cycles (planning, project, publication and "21st Century digital scholarship" – which includes dissemination and preservation) as well as some indication of other flows. It is clear how the cycle restarts with ideas springing from one piece of research naturally restarting a new planning cycle. It is one of the strongest designs with high use of colour to retain clarity despite its inclusion of lots of low level detail. The curvi-linear style is strongly suggestive of an organic process.

Model 3: The scholarly knowledge cycle (Lyon, 2003)

The scholarly knowledge cycle is a high level abstraction, which, unusually, encompasses both research and teaching – reflecting the reuse of research data and outputs in the teaching context. The text describing the model suggests that it is generic, but elements are clearly relevant to a specific, individual project. Although it is labelled as a knowledge cycle, this early model has a strong focus on data and information management, e.g., key elements are databases. In terms of elements and processes, the order of research stages is less clearly identified than in some of the other models, particularly because many of the linkages are bi-directional. This captures complexity but reduces readability. Similarly, its inclusion of heterogeneous elements (institutional repositories, databases, metadata, research and teaching processes) and differing infrastructures, means it captures complex reality, but has reduced legibility. In terms of visualisation two cycles (research and teaching) are happening simultaneously.

Model 4: Research Lifecycle enhanced by an "Open Science by Default" Workflow (Grigorov, et al. 2014)

This lifecycle is again for the whole of the research process, with a prescriptive objective to explain how open science elements could potentially fit in at each stage of research. An inner circle sets out nine basic research stages, with a strong emphasis on dissemination and public engagement. An outer circle mirrors each step with open science innovations, e.g., research data management (RDM) around data, open access for outputs and citizen science applied to engagement activities. An additional element is linking to the author's unique ID meandering through the whole process, implying the way it ties together disparate activities. In terms of elements and processes it is at a high level of abstraction and there is a strong sense of a serial, uni-directional flow. Visually it is presented as two concentric circles, but there is a clear logic to how the process is repeated from engaging publics to having a new idea.

Model 5: Idealized Scientific Research Activity Lifecycle Model (Patel, 2011)

The scope is again of the whole research lifecycle, specifically for science. The use of the word idealised in the title implies a prescriptive intent. The model is intended to show processes of a typical physical science experiment project and some idealised stages for supporting long-term research data management (Patel, 2011). In terms of elements and processes there is quite a bit more complexity than the previous models. Further, there is a definite sense of processes going on simultaneously in different areas, and some of the arrows suggest multi-directional flows. Activity is going on across a distributed set of systems not in a single one. It is unusual compared to most of the models in being presented in in the form of a mass of text in rectangular boxes linked by straight lines, suggesting that it has been developed out of linking together a number of chains of steps. However, in fact there is a strong sense of circularity. Four key heterogeneous elements (types of activities) are included: research, publication, administration and archiving – and linked by information flows. Each type of activity is colour coded to help the reader pick out related processes; but the heavy textual orientation makes the whole diagram feel very complex. In seeking to present significant detail the effect is of considerable complexity, and on first viewing the visualisation is hard to read. This may be more related to the quality of graphic design than conceptual. Legibility might also have been enhanced by organising some of the elements into clearer stages, e.g., there is no link between 'publications databases' and 'publish research.'

Model 6: The integrated scientific life cycle of embedded networked sensor research (Pepe et al. 2010)

Again, the model's scope is the whole research process, though in this case it relates to one specific research community. Like most of the other models the cycle is closed, uni-directional and serial. It is a closed circular model of the research lifecycle at a medium level of abstraction and involving three sequential processes: 1. experiment design and device calibration; 2. data capture, cleaning, and analysis; 3. publication and preservation. What seems to be indicated by this visualisation is the importance of seeing design, analysis and publication as three relatively discrete processes. The low use of text and colour reflects little attempt to lend complexity to the representation. Although the model is presented visually as circular, the potential link between last and first stage is not clear – since there is no focus on new idea generation or data reuse which might restart the process. Thus in reality the three sub-processes seem to be linked together in a chain process, ending with publication.

Model 7: E-science and the lifecycle of research (Humphrey, 2006)

Despite the title of this model implying it is about the whole of research, its main focus is data, its use and re-use, not all aspects of research. The main chain of activity is from study concept and design, through data collection, processing, access and dissemination to analysis. This is at a high level of abstraction. Above the data access chevron, a data repurposing loop starts. Thus while the lifecycle is presented visually as a chain in that the main processes are indeed set out in order, the looping arrows point to data repurposing as a circular process, as well as also suggesting a parallel process of research outcomes stimulating new study design. According to the author, the model aims to provide a better understanding of the relations between stages in research and increase the awareness of potential information losses between stages (Humphrey, 2006). The focus on data loss is interesting, however, in reality the concept of data loss is not very well visualised – it is supposedly indicated by gaps between each main activity in the chain, but this is not obvious to understand to the viewer. As with most of the other models there is a strong sense of uni-directionality and seriality. Because it is presented in rectinlinear forms with reliance on text it does not convey much sense of an organic process.

Model 8: Create and Manage Data (Corti et al. 2014)

In terms of scope unlike most of the other models it is lifecycle purely concerned with research data. It is specifically for social science research. It is a data-centric view of the research process, thus it does not show the wider context of how a project is conceived or planned (and in this sense is not optimised for the researcher to link to their own process). In terms of elements and processes it has a low level of abstraction, with much detail provided in accompanying text. It is largely closed, unidirectional, serial. Although presented as a circle it could be considered a chain matrix, because it consists of a series of stages, with some (textual) detail offered of activities that occur at each stage. Although presented as a circle there is no real basis for a "rebirth": There is a discontinuous jump between making one's own data accessible (an end of project activity) to then searching for that of others (implying a new project).

Model 9: DCC curation lifecycle model (Higgins, 2008)

In terms of scope, unlike the other models the DCC curation lifecycle is specifically concerned with data curation only. It focuses not on the life of data from the researcher's point of view, but on its preservation: indeed, it seeks to instantiate the OAIS reference model, a widely cited conceptual framework for digital archives (Higgins, 2008). The intention is to define how things should be done, so it is a prescriptive model. It does not encompass the whole research process, such as idea generation. "Conceptualise" as the starting point for data creation is symbolically set outside the main circular cycle. Indeed, the model does not examine what the researcher might do in terms of collecting, manipulating etc. data. Data creation is mentioned because the advice is that data has to be managed from birth, but the focus is on processes after its original purposes of collection. This makes it poor for explaining research services to end-users, but it is a strong representation of how data can be curated with an emphasis on reuse. In terms of elements and processes, it gives considerable low level detail of processes. It is closed and unidirectional, but has a strong sense of multiple processes happening simultaneously. If only because it concerned with one specific activity so activities seem to be happening within a single centralised system, unlike most of the other models. it is hard to understand in what sense "transform" the last step in the cycle can then lead on to "create/receive" so it is arguably, like the OAIS reference model, essentially a chain or number of chains, even if presented visually as circular and sequential. Sub processes are nested within the circle thus producing an onion or target model, a type of visualisation not identified by Ma and Wang (2010 a,b).

Discussion

The nine models analysed here represent a wide range of approaches and purposes in representing research and/or data: from abstract representations of the whole research lifecycle through to much more detailed models of data creation and reuse, and even just of data curation. Each has its strengths relevant to its particular purpose. Some are specific to a particular academic field or institution, others seek to generalise to any form of research. Although not an exhaustive picture of the range of types of model that have been developed in the last few years in the research/research data arena the analysis seems to support the value of the descriptive framework synthesised from previous authors.

A few patterns are immediately apparent. Sometimes the title of the model misrepresents the true scope, e.g., where a data lifecycle model is described as a research lifecycle (model 7). Similarly, it seems quite frequently to be the case that the visualisation misrepresents the process, e.g., is circular when in fact there is little explanation of how the loop is recontinued (Models 6,8,9) or is set out visually as a chain when there are strong suggestions of circularity (Model 5). More complex models seem to be more successful when they clearly identify sub-cycles or distinct stages. The DCC's onion-type model, as a variant on the circular model, is a type of visualisation not mentioned by Ma and Wang (2010a,b). The UKDA model shows how a circular model can also have a matrix element, with more detailed descriptions of each stage.

The strengths of the lifecycle model become clear through the analysis. It is a simple and understandable visualisation. Its typically curvilinear shapes imply organic rather than highly managed or rationalised processes. As a means to make sense of research it offers more insights than simple lists of scholarly primitives (Unsworth, 2000) or activities (Palmer and Cragin, 2008) because we need to have some notion of how such elements fit together. A temporal approach, often identifying discrete stages and their order, is the obvious way to link them. Lifecycles also improve on the complex "information flow maps" presented in RIN's (2009) descriptions of life science research processes. These may reflect reality more fully, but are not as memorable or understandable, because they seem more like snapshots in time and do not give us a sense of research as a movement of stages towards an end-goal.

Furthermore, the strength of the lifecycle is that to some extent it reflects how researchers themselves perceive research. For example, in showing different stages of research they echo many research methods books which also represent research in flow chart form, moving across a number of stages (e.g., Pickard, 2013) and visualised in what Ma and Wang (2010a,b) would categorise as a chain. At a fundamental level much research is driven by the need to produce an output within the scope of a time limited project and so is progressive and linear. Jeng et al. (2016) found that many researchers when asked to draw their research process represented it in a chain form.

Circular lifecycles can also be seen as having a strength in improving on the visualisation of research as a chain, by expressing the desire for data reuse, stressing that in some sense the process is to be repeated. Hence Briney (2015) writes about the "old data lifecycle" as picturing the purpose of data as achieved when a paper is written. The "new data lifecycle adds data sharing, preservation, and data reuse as steps in the research process." (Briney, 2015: 11). Yet on close inspection many of the data oriented models examined here do not really capture a circular process. At a pragmatic level the organisational form of a lifecycle is a good way of talking about delivering services at the point of need/within the user/researcher workflow (Carlson, 2014). It reflects a general recognition of the need to reflect patterns in the life of the user, rather than expecting users to fit into the life of the library (Connaway, 2015).

Some models go further in capturing complexity, e.g., the "idealised scientific research activity lifecycle model". This arises from seeking to represent a number of processes occurring simultaneously, multi-directionality, and relying on text to convey detailed activities. A counter trend is to boil things down to a minimum of basic elements, ones that resonate with all researchers regardless of discipline or method. This is exhibited, for example, in many university web sites where the simplest categories are used to organise an explanation of RDM such as "create, organise, keep, find and share" (University of Leicester, 2017).

Nevertheless, there are weaknesses with the lifecycle model for representing research and research data. In particular, most of the models are closed, serial and unidirectional. This cuts against what we know of the typical character of research. In the real world research is often accepted to be:

Unique: each piece of research has its own pattern, shaped by the context of the research, particular choices in research design and contingent events.

Shaped by discipline/sub-discipline and by methodology: there is a strong sense that the process of research differs between different academic "tribes", particularly because of the use of different research methodologies. Research methods text books suggest that at a high level quantitative and qualitative research may have fundamentally different structures. But there are also many mixed methods research designs (Cresswell and Plano Clark, 2011).

Iterative and non-linear: research tends to be based on repeating steps a number of times or going back and forwards between different stages, and most significantly, it is non-linear. It may simply not progress by clear stages. Stages may run in parallel or could be skipped entirely (Pepe et al., 2010). A researcher may hit a problem and have to repeat steps early in the cycle (Mattern et al., 2015) or plan the research as highly iterative. Furthermore, at some point repeating the same method that has been used before has to break down and a new paradigm be constructed. So a linear narrative of research neglects the creativity and intuition that is also an important part of research. By definition there cannot be any recipes or prescriptions for the exploration of the unknown.

Heterogeneous: research involves multiple types of information and data. The RIN/Nesta (2010) information flow maps express well the sheer variety of types and sources of data used in a single field of research; lifecycles typically do this less well.

Open: new data and material can be drawn into research at almost any time (e.g., collecting new data or finding new literature). Particularly in qualitative research with an exploratory or emergent design, there is a sense that new material can be brought into the research even at a late stage.

> **Distributed**: the infrastructure is typically not a unified one. This is true whether that be technical infrastructure, such as tools or storage spaces or socio-technical infrastructure, e.g., the people, communities, places that make research possible.
>
> **Concatenated**: researchers often pursue an interest across multiple contexts, in multiple projects. The nature of the intellectual journey here is much richer and more complex to summarise than that encompassed by one lifecycle.
>
> **Messy**: this characteristic captures a sense of real world disorganisation, unplanned and unexpected outcomes (Brannen, 1992). Research does not play out in a predictable way. Research is a bricolage: making the best of the material to hand to create a new understanding of a phenomena (Berry, 2006).

Lifecycle models are explicitly abstractions and simplifications; that is their value. They clarify by making relevant simplifications (Carlson, 2014). In terms of the elements and processes dimension they tend to represent research as homogeneous, closed, unidirectional and serial. In terms of visualisation they are simple visually, and use low colour and text. In doing so they mask much of the complexity captured in the adjectives above. Hence it is a weakness of many of the models, that though they represent research as containing heterogeneous elements and existing on distributed architectures they see it as a rather closed system. They did not represent wider processes and influences. Typically, they are unidirectional and based on a series of pre-set, serial stages. None of them were evolving spirals, which would have reflected deepening understanding and learning. This begins to open up a more critical viewpoint on the lifecycle model, which despite its proliferation, has limits of explanatory power.

Lifecycles may make most sense where research is seen as a time limited project. And indeed this may be an increasing reality, one reinforced by funding, and academic promotion and tenure structures. It is consistent with some views of research that Brew (2001) found in her study of the experience of research: such as of it as a series of steps (dominoes conception) or as in a social market place (trading conception).

However, Brew (2001) also found that some researchers see the experience of research as a journey of personal transformation. A lifecycle is weak in capturing this. Rather, we have resonances of the much reproduced visualisation “the island of research”, that imagines research as an exciting but hazardous journey through an exotic island landscape (Harburg, 1966). Here the process is complex, uncertain and messy; and research is understood as an intellectual but also an emotional, even physical challenge. A lifelong career of concatenated research where rather intangible underlying themes are explored in an open ended way, has a very different quality to the individual project with its clear time horizon and outputs. Researchers’ career narratives may often be presented as an open ended quest for knowledge. In a study at University of Pittsburgh libraries one participant represented research as an “unending, repeating cycle” (Mattern, 2015:415). In the course of this resources are drawn on in an open way, are heterogeneous, and the infrastructure is distributed. This is part of what is lost in many lifecycle models: another weakness. In the Pittsburgh study this participant drew their research as a daisy shape, with activities around an enduring core of the scholar’s own interests (Jeng et al., 2016). She is quoted as saying “In order to unlock a phenomenon, I’ve got to understand who I am as a researcher” (Mattern et al., 2015: 415). From a theoretical viewpoint it is important that the concept of research in LIS reflects this complexity. It may be that researchers respond more strongly to these types of representations, than visualisations that are more simplistic and reductive.

A critical point is where the lifecycle model turns full circle. In a few cases we have seen that the notion that there is a restart of the cycle has not really been explained. The real nature of how a process might be iterated is not developed. If we are following data, it may be reused but it is likely to be in a different context; often by another researcher. The simple circle of reproduction pattern does not capture this complexity. Ma and Wang (2010a,b) emphasise that a commonality of writing about lifecycles is that the value of data is typically seen as declining from the moment it is created – a characteristic best captured in the wave model. On the other hand, if we are following a research lifecycle, it will never be simply asking the same question again. If we are building knowledge a better way to represent growing understanding would be some sort of upwards spiral, since the whole point is that overall understanding is being accumulated. One thinks immediately of the spiral models in Knowledge Management such as the SECI model proposed by Nonaka, Toyama and Komo (2000). This after all is a model of knowledge creation: should not models of research also capture this upward spiral? It is noticeable that this is one of Ma and Wang's (2010a,b) types of lifecycle visualisation that is not found in this collection of lifecycles.

From the information professional perspective, the (organisational and community) lifecycle is appealing as a way of bringing the library into the scholar's "workflow". But that currently popular term seems rather reductive, and suggests standardisation, an administrative or IT perspective on research. For some researchers the most engaging and highly meaningful aspects of research are not captured by the idea of workflow. So while this approach goes further in seeing research from the researcher's perspective, the lifecycle still often lacks the inner grasp of research as a personal transformation or intellectual journey.

A further weakness of the lifecycle model, is that within such models data seems to be a neatly bounded "thing". Data is being represented as a spreadsheet or output that appears to exist unproblematically as a distinct entity and is worked on through a series of stages. Yet what we know about research data is the way it changes itself through the lifecycle: Data is collected, quality checked, selected, combined, transformed (e.g., through simulations), it may be embargoed or released in edited form. The UKDA model comes nearest to capturing some of this. But the other models we have looked at here do little to capture this sense of data as changing and relational (Haider and Kjellberg, 2016). Even the DCC model tracks data as if it were a single entity. Yet this is to blackbox one of the key information management challenges.

We might also want to recognise the way that research data is managed within a wider Personal Information Collection. Thus data could be just one element within a scholar's PICs related to research and wider PICs for other tasks such as teaching and administration (Al Omar and Cox, 2016).

This thought directs our attention to another alternative viewpoint and weakness in the lifecycle approach. This arises where we focus on the data itself as mobile across different social contexts. This is a perspective captured in the metaphor of the data journey (Bates et al., 2016). Here we follow data, from its creation, often outside a research context as it is moved, transformed and repurposed in the different domains within which it is used. To exemplify this journey, Bates et al. (2016) refer to the way weather data can be created in different contexts such as in weather stations or through extracting historic data from ship logs, and later used by the national meteorological office, but also move to be exploited in future markets in the City of London. Movement across contexts is accompanied by mutability in the content and meaning of the data as it is combined and managed differently. Data is created by managers of a local weather station in a spirit of openness, but is much more closely protected by others in the chain. Bates et al. (2016) visualise the journeys of weather data within and outside the research context with an underground map type diagram with different coloured lines and different stops. This is actually a chain representation, with sharp disjunctions as data is fundamentally transformed in different contexts. It articulates scepticism about the meaning of open data. It is a sharp reminder that many cyclic models are driven by policy prescriptions about reuse. Most of the models we have looked at are project oriented reflecting the context of the use of data (and often its first creation) in one research project. The journeys model prompts us to consider the wider life of the data; just as the personal journey model prompts us to think about the wider narrative of the data in the researcher's career.

Conclusion

This paper presents a systematic investigation comparing the proliferating number of lifecycle models that have been generated in the area of research support and research data management, both in the peer reviewed and practitioner literatures. The analysis of the models revealed the very differing perspectives on research that are current and the different value of these, as well as reminding us that no model can capture a comprehensive viewpoint. On a practical level, the analysis will help practitioners select or design models appropriate to the task at hand. A number of radically alternative visualisations/metaphors were also considered, such as the perspective held by many researchers of research as a transformational journey (and that might be best represented as a spiral) and the rather discontinuous journey of data itself, which could be visualised in subway map form. Given the complexity of both research and data our discussion has revealed both the flexibility and some of the perils in over-reliance on this one metaphor. Burgi and Roos (2003) suggest that when dealing with complex concepts, multiple metaphors are needed to avoid being trapped into simplistic assumptions. The lifecycle idea is an extremely useful metaphor, but it tends to encourage thinking that research processes are highly purposive, unidirectional, serial and occurring in a closed system. Research is often not like this, and the analysis has exposed the limited thinking created by such an assumption. The lifecycle model also often implies a repeated cycle when there is no real basis for this. The conclusion must to be suggest a need to add other visualisations for research to our repertoire of conceptual models. The knowledge spiral and the data journey map are just two such examples that reveal that viewing research through different metaphors enriches our understanding. This is important theoretically because LIS needs to develop a convincing understanding of the research process. It is important practically too, because while lifecycles are increasingly used to explain service offerings, the analysis shows that this may not always reflect researchers' own understanding of the research process. In failing to do so they can alienate potential users.

The paper developed and tested a framework for systematic comparison that can be reused or provide the basis for further study. This highlights three sets of interdependent features, around scope and point of view; elements and processes; and visualisation. These features seem to capture the degree of relevant complexity retained in each model.

As LIS continues to unravel the complexities of research it seems likely that more models will emerge, helping us to build up an even richer account of research. In this context there is great value in asking researchers themselves about the metaphors through which they understand research. The drawing based method used at Pittsburgh by Jeng et al. (2016) and Mattern et al. (2015) is a useful approach to elicit such representations. Clearly what the researcher is being asked to represent is key: it could usefully try to get a representation of research career, not just an individual project. Developed with academics from a wider range of subjects and levels of experience would enable us to gain more insight into the nature of research. New powerful visualisations to explain research, will also help all stakeholders understand their role and to explain the place of services in the research process.

References

AL-Omar, M., & Cox, A. M. (2016). Scholars' research-related personal information collections: A study of education and health researchers in a Kuwaiti University. *Aslib Journal of Information Management*, *68*(2), 155–173. http://doi.org/10.1108/AJIM-04-2015-0069

Ball, A. (2012). *Review of Data Management Lifecycle Models*. Bath: University of Bath. http://opus.bath.ac.uk/28587/1/redm1rep120110ab10.pdf

Bates, J., Lin, Y.-W., & Goodale, P. (2016). Data journeys: Capturing the socio-material constitution of data objects and flows. *Big Data & Society*, *3*(2). http://doi.org/10.1177/2053951716654502

Berry, K. (2006). Research as bricolage: Embracing relationality, multiplicity and complexity. In K.Tobin and J. Kincheloe (Ed.s), *Doing educational research: A handbook*, Rotterdam: Sense publishers, 87-116.

Brannen, J. (Ed.). (1992). Introduction. In J. Brannen (Ed.) *Mixing methods: Qualitative and quantitative research*. Routledge, xi-xvi.

Brew, A. (2001). Conceptions of Research: A phenomenographic study. *Studies in Higher Education*, *26*(3), 271–285. http://doi.org/10.1080/03075070120076255

Briney, K. (2015). *Data Management for Researchers: organize, maintain and share your data for research success*. Exeter: Pelagic Publishing.

Carlson, J. (2014). The Use of Life Cycle Models in Developing and Supporting Data Services. In J. M. Ray (Ed.), *Research Data Management. Practical Strategies for Information Professionals.* West Lafayette: Purdue University Press., 63-86.

Committee on Earth Observation Satellites (CEOS) Working Group on Information Systems & Services (2011) http://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/WGISS_DSIG_Data-Lifecycle-Models-And-Concepts-v13-1_Apr2012.docx.

Connaway, L. S. (2015). Introduction. In L.S. Connaway (Ed), *The Library in the Life of the User: Engaging with People Where They Live and Learn*. Dub: OCLC, i-ix. Retrieved from http://www.oclc.org/content/dam/research/publications/2015/oclcresearch-library-in-life-of-user.pdf

Corrall, S. (2014). Designing libraries for research collaboration in the network world: An exploratory study. *LIBER Quarterly*, *24*(1), 17–48.

Corti, L., Van den Eynden, V., Bishop, L., & Woollard, M. (2014). *Managing and sharing research data: a guide to good practice*. Los Angeles: Sage.

Creswell, J. W., & Clark, V. L. P. (2011). *Designing and Conducting Mixed Methods Research* (2nd ed.). Thousand Oaks, United States: Sage Publications.

Dempsey, L., Malpas, C., & Lavoie, B. (2014). Collection Directions: The Evolution of Library Collections and Collecting. *Portal: Libraries and the Academy*, *14*(3), 393–423. http://doi.org/10.1353/pla.2014.0013

Dooley, J. (2015). *The Archival Advantage: Integrating Archival Expertise into Management of Born-digital Library Materials*. Dublin, Ohio: OCLC Research. Retrieved from http://www.oclc.org/content/dam/research/publications/2015/oclcresearch-archival-advantage-2015.pdf

Grigorov, I., Carvalho, J., Davidson, J., Donnelly, M., Elbaek, M., Franck, G., Jones, S., Melero, R., Knoth, P., Kuchma, I., Orth, A., Pontika, N. Rodrigues, E. and Schmidt, B. (2014). Research Lifecycle enhanced by an "Open Science by Default" Workflow. doi: 10.5281/zenodo.49960

Haider, J., & Kjellberg, S. (2016). Data in the making: Temporal aspects in the construction of research data. In K. Rekers, J.V. and Sandell (Ed.), *New big science in focus: Perspectives on ESS and MAX IV* (pp. 143–164). Lund University.

Harburg, E. (1966). The Island of Research. *American Scientist*, *54 (4)*, 470.

Higgins, S. (2008). The DCC Curation Lifecycle Model. *International Journal of Digital Curation*, *3*(1). http://doi.org/10.2218/ijdc.v3i1.48

Humphrey, C. (2006). e-Science and the Life Cycle of Research. Retrieved from https://era.library.ualberta.ca/files/bvq27zn54q/Lifecycle-science060308.pdf

Jeng, W., Mattern, E., He, D., & Lyon, L. (2016). Unpacking the “ Black Box ”: A Preliminary Study of Visualizing Humanists and Social Science Scholars ’ Data and Research Processes. In *iConference 2016 Proceedings*. http://doi.org/10.9776/16173

L’Hours, H. (2014). Workflows and lifecycles: the long-term view from a national data centre. RDMF11, DCC 20 June 2014. Retrieved from http://www.dcc.ac.uk/sites/default/files/documents/RDMF11/HERVE.pdf

Lenhardt, W. C., Ahalt, S., Blanton, B., & Christopherson, L. (2014). Data Management Lifecycle and Software Lifecycle Management in the Context of Conducting Science, *Journal of Open Research Software*, *2*(1), 1–4. http://doi.org/10.5334/jors.ax

Lyon, L. (2003). eBank UK: Building the links between research data, scholarly communication and learning. *Ariadne*. Retrieved from http://www.ariadne.ac.uk/issue36/lyon/

Ma, F., & Wang, J. (2010)a. A Literature Review of Studies on Information Lifecycle I: the Perspective of Value. *Journal of the China Society for Scientific and Technical Information*, *29*(5), 939–947.

Ma, F., & Wang, J. (2010)b. The Review of Studies on Information Lifecycle II: the Perspective of Management. *Journal of the China Society for Scientific and Technical Information*, *29*(6), 1080–1086.

Mattern, E., Jeng, W., He, D., Lyon, L., & Brenner, A. (2015). Using participatory design and visual narrative inquiry to investigate researchers' data challenges and recommendations for library research data services. *Program*, *49*(4), 408–423. http://doi.org/10.1108/PROG-01-2015-0012

Möller, K. (2013). Lifecycle models of data-centric systems and domains. *Semantic Web*, *4*(1), 67–88. http://doi.org/10.3233/SW.2012.0060

Nonaka, I., Toyama, R., & Konno, N. (2000). SECI , Ba and Leadership : a Unified Model of Dynamic Knowledge Creation. *Long Range Planning*, *33 (1)*, 5–34. https://doi.org/10.1016/S0024-6301(99)00115-6

Palmer, C. L., & Cragin, M. H. (2008). Scholarship and disciplinary practices. *Annual Review of Information Science and Technology*, *42*(1), 163–212. http://doi.org/10.1002/aris.2008.1440420112

Patel, M. (2011). Idealised Scientific Research Activity Lifecycle Model. University of Bath. Retrieved from http://opus.bath.ac.uk/35186/1/I2S2_ResearchActivityLifecycleModel_110407.pdf

Pepe, A., Mayernik, M., Borgman, C. L., & Sompel, H. Van De. (2010). From Artifacts to Aggregations : Modeling Scientific Life Cycles on the Semantic Web, *61*(3), 567–582. http://doi.org/10.1002/asi

Pickard, A. J. (2013). *Research Methods in Information* (2nd ed.). London: Facet Publishing.

Research Information Network (RIN). (2009). Patterns of information use and exchange: case studies of researchers in the life sciences. RIN/British Library http://www.rin.ac.uk/system/files/attachments/Patterns_information_use-REPORT_Nov09.pdf

Research Information Network (RIN)/ National Endowment for Science Technology and the Arts (NESTA). (2010). *Open to All? Case studies of openness in research*. Retrieved from http://www.rin.ac.uk/system/files/attachments/NESTA-RIN_Open_Science_V01_0.pdf

University of Central Florida Libraries Research Lifecycle Committee. (2012). The research lifecycle at UCF [Online Graphic]. Retrieved from https://library.ucf.edu/about/departments/scholarly-communication/overview-research-lifecycle/

University of Leicester (2017). Data management support for researchers. Retrieved from http://www2.le.ac.uk/services/research-data

Unsworth, J. (2000). Scholarly Primitives: what methods do humanities researchers have in common, and how might our tools reflect this? Retrieved from http://people.virginia.edu/~jmu2m/Kings.5-00/primitives.html

Williams, C. (2006). *Managing Archives: Foundations, Principles and Practice* (1st ed.). Oxford: Chandos Publishing.

Wilson, J. (2014). Mapping Life-cycle models to the University – in theory and in practice. RDMF Meeting, 20th June, 2014. Retrieved from http://www.dcc.ac.uk/sites/default/files/documents/RDMF11/James.pdf

Wissik, T., & Ďurčo, M. (2015). Research Data Workflows : From Research Data Lifecycle Models to Institutional Solutions. In *CLARIN 2015 Selected Papers, Linköping Electronic Conference Proceedings, Annual Conference 2015, October 14–16, 2015, Wroclaw, Poland* (pp. 94–107). Linköping University Electronic Press, Linköpings universitet. Retrieved from http://www.ep.liu.se/ecp/123/008/ecp15123008.pdf

Name of the model	1. Research lifecycle (RIN & NESTA 2010)	2. The research lifecycle at UCF (University of Central Florida Libraries' Research Lifecycle Committee, 2012)	3. The scholarly knowledge cycle (Lyon, 2003)	4. Research Lifecycle enhanced by an "Open Science by Default" Workflow (Grigorov, et al. 2014)	5. Idealized Scientific Research Activity Lifecycle Model (Patel, 2011)	6. The integrated scientific life cycle of embedded networked sensor research (Pepe et al. 2010)	7. E-science and the lifecycle of research (Humphrey, 2006)	78. Create and Manage Data (Corti et al. 2014)	9. DCC curation lifecycle model (Higgins, 2008)
	Scope and point of view								
Subject matter	Research lifecycle (with material outputs table)	Research lifecycle	Data and Information Management, despite the use of the term "knowledge" in the name	Research lifecycle	Research and data lifecycle	Research lifecycle	Although the title suggests it is a research lifecycle model, it has a strong focus on data use	Data lifecycle	Data curation lifecycle
Individual / organization / community based	Community based	Organization based	Individual / organization based	Community based	Community based	Community based	Community based	Community based	Community based
Prescriptive or descriptive	Descriptive	Mixed	Mixed	Prescriptive	Prescriptive	Descriptive	Descriptive	Mixed	Prescriptive
	Elements and processes								
Level of abstraction	High	Low	Medium	High	Medium	Medium	High	Low	Low
Homogenous/ heterogeneous	Homogenous	Heterogeneous	Heterogenous	Heterogenous	Heterogenous	Heterogeneous	Homogeneous	Homogenous	Homogenous
Closed/open	Closed	Closed	Open	Open	Closed	Closed	Closed	Closed	Closed
Centralized/distributed infrastructure	Distributed	Distributed	Distributed	Distributed	Distributed	Distributed	Centralized	Distributed	Centralized
Uni/ multi-directionality	Uni-directional	Uni-directional	Multi-directional	Uni-directional	Multi-directional	Uni-directional	Uni-directional	Uni-directional	Uni-directional
Seriality/ siimultaneity	Serial	Serial	High simultaneity	Serial	High simultaneity	Serial	Serial	Serial	High simultaneity
	Visualisation								
Lifecycle type	Circular or circular matrix	Integrated of four circular cycles	Integrated of two circular cycles	Circle	Circular/chain	Circular presentation but could be seen as chain	Chain with circular elements	Chain matrix	Chain Onion model
Use of colour	Low	High	Low	Low	Medium	Low	Medium	Low	Medium
Visuality/ textuality	Low text	Low text	Medium text	Medium text	High text	Low text	Low text	Low text (but matrix is highly textual)	Low
Recti- / curvilinearity	Curvi-linear	Curvi-linear	Mixed	Curvi-linear	Rectilinear	Curvilinear	Rectilinear	Curvilinear	Curvilinear

Table 1. A comparison of nine lifecycle models