The Data Berg Model

Introducing the Data Berg Model

A White Paper

Version 1.0 July 2026

Alan Green CITP FBCS

Contents

Document Information

Author

Alan Green CITP FBCS

Document

The Data Berg Model – White Paper

Version

1.0

Publication Date

July 2026

Document History

Version	Date	Notes
1.0	July 2026	First public release

1. Executive Summary

Organisations create, store and manage ever-growing volumes of long-lived data. Existing lifecycle models, records management practices and governance frameworks offer valuable guidance for managing that information, but they do not generally provide a common vocabulary for distinguishing between the different stages and events that occur throughout its lifetime. The Data Berg Model addresses this need.

The Data Berg Model is a conceptual model and reasoning framework for understanding how long-lived data forms, persists and is subsequently acted upon. It offers an implementation-agnostic behavioural vocabulary, applied consistently across both digital and physical information, that allows organisations to describe, reason about and communicate these dynamics.

The Data Berg Model describes the following phases, state and behaviours for long-lived data:

- Cooling
- Freezing
- Frozen
- Melt
- Thaw
- Calving
- Evaporation

These classifications rest on objectively identifiable phases, states and behavioural events, rather than assumptions about organisational intent, business process or system architecture. This distinction enables consistent reasoning across diverse organisational and technical environments while preserving a stable vocabulary.

By establishing a shared language for long-lived data, the model allows for more precise communication between organisations and the people responsible for preserving, managing, governing and using the data. This shared understanding reduces ambiguity and provides governance, systems design and implementation with a consistent conceptual foundation, helping organisations make more informed and consistent decisions throughout the data's lifetime.

2. Introduction

Data typically outlives the activities that created it. Business transactions complete, research projects conclude, systems are replaced and organisations evolve, yet the information produced by those activities frequently remains for years or decades.

Throughout that time, the data itself does not remain static. Its operational value changes, its context evolves and new obligations arise. All or part of the data may be preserved, reactivated, divided, copied or permanently removed. These changes are observable regardless of the technology used to store the data or the governance framework used to manage it.

Existing lifecycle models, records management practices and governance frameworks provide valuable guidance for controlling information, but they do not generally offer a common vocabulary for describing what long-lived data is doing as it transitions through long-term preservation. The Data Berg Model establishes that vocabulary, providing a conceptual lens through which those stages can be considered.

Viewed through that lens, long-lived data can be observed, reasoned about and discussed consistently by everyone involved in its stewardship. This shared understanding provides the foundation upon which the Behavioural layer and subsequent Governance, Systems and Implementation layers are built.

3. Principles of the Data Berg Model

Principle 1 – Behaviour-Centric

The Data Berg Model is behaviour-centric. It approaches long-lived data by first observing what the data is doing before considering how that data should be governed, managed or implemented. This is fundamental to the entire model.

Most approaches to managing data begin by deciding what rules, policies or technologies should be applied. The Data Berg Model takes a different starting point. Before deciding how long-lived data should be managed, it first seeks to identify what the data is doing and the conditions it is exhibiting.

Placing behaviour at the centre of the model provides a stable foundation upon which governance decisions can be made. Different organisations may apply different retention policies, security controls, technologies or operational processes, but the observable classifications remain the same.

Understanding these before determining how data should be managed provides a more consistent basis for governance, systems design and implementation. Rather than allowing governance decisions to define how data is interpreted, the model encourages governance to be informed by what can first be objectively observed. Organisational intent, ownership and technical implementation provide important context, but none of them alter the underlying classification being observed.

Principle 2 – Technology Independence

The Data Berg Model is independent of the technologies used to create, store, preserve or manage long-lived data. This independence allows the model to be applied consistently across different organisations and environments.

Long-lived data exists in many forms. It may be stored in modern cloud platforms, traditional databases, document management systems, physical archives or technologies yet to emerge. Although the systems used to manage that data may change over time, the underlying phases, states and behaviours described by the model remain the same.

By separating the model from any particular technology or implementation, it provides a stable conceptual foundation unaffected by changes in software, infrastructure or organisational practice. This allows organisations to adopt the model without changing their existing systems, and enables the same behavioural vocabulary to be applied across diverse technical environments.

Principle 3 – Implementation Independence

The Data Berg Model is implementation independent, too. It draws a clear line between understanding the behaviour of long-lived data and deciding how that behaviour gets implemented in practice — which is what lets it stay clear of the many different ways organisations choose to preserve, manage and use their data.

There is rarely a single correct implementation for managing long-lived data. Different organisations have different objectives, regulatory obligations, budgets, technologies and operational practices, so the same observed behaviour may be implemented in many different ways.

Separating behavioural understanding from implementation decisions provides a stable conceptual foundation that can be adopted without prescribing specific processes, products or technical solutions. Organisations remain free to implement the model in whatever way best meets their own requirements, while continuing to share a common behavioural understanding.

4. The Challenge of Long-Lived Information

Every organisation generates information in support of its activities. Business transactions complete, research projects conclude, systems are replaced and organisations evolve. While these activities may themselves be relatively short-lived, the information they produce often remains for years or decades.

This long-lived information continues making demands on the organisation as long at it continues to exist. It has to be retained, protected, understood and made available when needed. Regulatory enquiries, subject access requests, audits, litigation and historical investigations may all require information to be located, interpreted and trusted long after the people, systems and processes that created it are gone.

Managing long-lived information therefore extends beyond simply storing it; it requires the continued investment of organisational energy, including the people, expertise, time and financial resources needed to preserve, understand and act upon that information throughout its lifetime. Whether provided by employees, specialist contractors or external consultants, these resources represent an ongoing organisational commitment that persists long after operational activity has ceased.

Long-lived information should not, however, be viewed solely as an organisational obligation. The same information that carries continuing cost, risk and responsibility may also retain significant future value. Historical information may provide evidence, support business intelligence, enable new analytical insight, inform strategic decision-making or provide valuable source material for artificial intelligence.

The challenge for organisations is therefore twofold: they must manage the continuing obligations associated with long-lived information while remaining able to recognise and realise its potential future value. Achieving both requires more than effective governance or reliable storage; it requires a consistent way of understanding what long-lived information is, how it behaves throughout its lifetime, and how those behaviours may be observed and discussed across different organisational disciplines.

Existing lifecycle models, records management practices and governance frameworks provide valuable guidance here, explaining how information should be retained, protected, classified and ultimately disposed of. These are essential disciplines; however, they do not generally provide a common behavioural vocabulary for describing the observable classifications exhibited by long-lived information, independently of the technologies, governance frameworks or implementation approaches used to manage it.

The Data Berg Model addresses this challenge by offering a behaviour-centric conceptual lens through which long-lived information can be consistently observed, understood and discussed. Rather than replacing existing governance frameworks or lifecycle models, it complements them, providing a common behavioural understanding that allows organisations to reason more consistently about the continuing obligations, accumulated risk and potential future value of long-lived information.

5. The Data Berg

A Data Berg is the stable, long-lived state that results once data has completed its operational purpose and entered long-term preservation. Although it may remain unchanged for many years, it should not be regarded as static or inactive: throughout its lifetime, a Data Berg continues to exist within an organisational context, subject to ongoing obligations while retaining the potential for future value.

The Data Berg name is drawn from the natural behaviour of icebergs, and the metaphor is intended as a conceptual representation rather than a technical one: it describes the observable behaviour of preserved data over time, not physical storage or system architecture. Like an iceberg, a Data Berg may remain stable for long periods before being acted upon by internal or external influences that change its state.

The Data Berg Model traces the formation and subsequent behaviour of a Data Berg throughout its lifetime. It begins with data as it cools and, where appropriate, progresses through Freezing to form a stable, long-lived Data Berg. Once formed, a Data Berg may subsequently be subject to a series of observable behaviours, including Melt, Thaw, Calving and Evaporation.

Together, these phases, states and behaviours form a consistent conceptual framework for understanding how long-lived data forms, persists and gets acted upon throughout its life.

6. The Behavioural Journey

Not every Data Berg will pass through every phase or behaviour. Some data may never progress beyond Cooling, while preserved Data Bergs may be subject to different behaviours depending upon organisational needs and external events.

The sections below introduce each phase, state and behaviour of the Model.

Cooling

Cooling is the first phase of the Data Berg Model. It begins when data has stabilised and is no longer being actively changed — for example, a completed business transaction, a finalised document, a customer record, an email address, or any other information that has reached a stable condition.

Cooling is an observable phase rather than an action: it reflects a change in the condition of the data itself, not a decision made by the organisation. Cooling does not imply that the data will become a Data Berg; it simply identifies the point at which the data becomes a candidate for future decisions. It may remain within an operational system, be deleted, or progress towards long-term preservation.

Recognising Cooling provides organisations with a consistent way of identifying data that has become stable, marking the beginning of its potential long-lived journey.

Freezing

Freezing is the second phase of the Data Berg Model. Within the model, Freezing represents the concept of freeze-framing rather than the physical phenomenon of freezing a fluid. In this context, Freezing preserves not only the data itself, but also the surrounding context needed to understand its meaning long after the systems, processes and people that created it may no longer exist.

Freezing is the deliberate process of preparing stable data for long-term preservation by capturing sufficient context to ensure it can be understood, trusted and used in the future.

The extent of Freezing is determined by the needs of the organisation. The more effectively context is preserved during this phase, the more likely the resulting Data Berg will remain understandable, usable and valuable throughout its lifetime.

Frozen

Frozen is the stable, long-lived state created once the Freezing phase has been completed. At this point, the data has entered long-term preservation as a Data Berg, together with the context needed to support its future understanding.

Although a Data Berg may remain unchanged for many years, it should not be regarded as inactive. Throughout its lifetime it continues to accumulate organisational obligations and risk, while also retaining the potential for future value. Regulatory change, evolving business needs, technological change and new analytical opportunities may all alter the significance of a Data Berg without changing the data itself.

The Frozen state provides the foundation from which the subsequent behaviours of the Data Berg Model may occur. A Data Berg may remain Frozen indefinitely, or it may subsequently be subject to Melt, Thaw, Calving or Evaporation as organisational needs and external events arise.

Melt

Melt is the deliberate reactivation of all or part of a Data Berg to achieve a specific organisational objective, such as analysis, migration, business intelligence, research or business change.

Melt is a planned and controlled behaviour, initiated, resourced and owned by the organisation. During Melt, long-lived data is deliberately brought back into active use while the Data Berg itself remains preserved. Melt therefore represents the intentional reactivation of long-lived data, rather than a permanent change to the Data Berg itself.

Melt demonstrates that a Data Berg should not be regarded as permanently dormant. Long-lived data may remain preserved for many years before being intentionally reactivated in response to changing organisational priorities, new opportunities or evolving business needs.

Thaw

Thaw is the reactivation of all or part of a Data Berg in response to an external requirement or event, such as a regulatory enquiry, legal proceedings, an audit or other external obligation.

Thaw is an externally imposed behaviour. The organisation does not determine when it occurs, but must nevertheless respond to the external circumstances that require long-lived data to be brought back into active use. During Thaw, the Data Berg itself remains preserved while the required data is reactivated.

Thaw demonstrates that a Data Berg may be reactivated independently of organisational intent. Long-lived data may remain preserved for many years before external events require it to be located, understood and used, often with little opportunity for advance planning. The organisation cannot control when a Thaw occurs, only its readiness to respond.

Calving

Calving occurs when all or part of a Data Berg separates and continues as a new, independent Data Berg.

Once Calving has occurred, the new Data Berg exists independently of its parent. Although it may initially contain the same information and context, it begins its own behavioural journey. From that point onwards, it may evolve independently, accumulating its own context, obligations and risk.

Calving demonstrates that long-lived data does not always remain as a single, indivisible entity. As Data Bergs separate and evolve independently, they may gradually drift apart, becoming increasingly distinct despite sharing a common origin.

Evaporation

Evaporation is the deliberate and irreversible destruction of all or part of a Data Berg once its purpose for continued preservation has come to an end.

Unlike the other behaviours of the Data Berg Model, Evaporation permanently removes data from existence. Once Evaporation has occurred, the affected data no longer forms part of the Data Berg and cannot subsequently participate in any further behaviours.

Evaporation demonstrates that long-lived data should not be preserved indefinitely. As organisational obligations, risk and potential future value change over time, there comes a point where permanent destruction may become the most appropriate outcome for all or part of a Data Berg.

7. Why the Model Matters

The Data Berg Model provides organisations with a common behavioural vocabulary for understanding long-lived data, independently of technology, governance frameworks and implementation approaches. By distinguishing the observable phases, states and behaviours of a Data Berg, the model enables information to be discussed more consistently across business, governance and technical disciplines.

This shared understanding helps organisations reason more clearly about long-lived data before deciding how it should be governed, supported by systems or implemented in practice. Rather than replacing existing approaches, the Data Berg Model complements them by providing a stable behavioural foundation upon which those decisions can be based.

As organisations continue to create increasing volumes of long-lived information, understanding how that information behaves throughout its lifetime becomes increasingly important. The Data Berg Model provides a consistent conceptual framework through which organisations can better understand the continuing obligations, accumulated risk, potential future value and organisational energy associated with long-lived data.

8. Future Publications

This white paper introduces the Behavioural Layer of the Data Berg Model and provides the conceptual foundation for understanding the observable phases, states and behaviours of long-lived data.

Future publications will expand the Data Berg Model by describing additional layers that build upon this behavioural foundation. These are expected to include:

- Behavioural Layer Specification — the formal behavioural definitions, properties and classifications of the Data Berg Model.
- Governance Layer Specification — guidance for governing long-lived data using the behavioural foundation established by the model.
- Systems Layer Specification — architectural considerations for designing systems that support the Data Berg Model.
- Implementation Guidance — practical guidance and examples illustrating how the Data Berg Model may be implemented within different organisational environments.

Each publication will be capable of being adopted independently while sharing a common behavioural foundation, allowing organisations to adopt those elements of the Data Berg Model most appropriate to their own requirements.

9. Conclusion

Organisations have become increasingly skilled at creating, storing and governing information. Yet much of that information will spend the majority of its existence as long-lived data, where its value, obligations and risk continue to evolve long after the activities that created it have ended.

The Data Berg Model offers a different way of looking at that information. Rather than beginning with technology, governance or implementation, it begins by observing the behaviours of long-lived data itself. From that behavioural understanding comes a clearer basis for discussing how long-lived information should be preserved, reactivated, divided and ultimately removed.

The Data Berg Model is intended to encourage organisations to think differently about their long-lived information — not simply as data that has been stored away, but as an enduring organisational asset whose behaviours continue throughout its lifetime. By providing a common behavioural vocabulary, the model supports clearer communication, more consistent reasoning and better-informed decision-making wherever long-lived data exists.

Copyright

Author: Alan Green CITP FBCS

Website: https://www.alan-green.co.uk

LinkedIn: https://linkedin.com/in/alan~green/