

THE COMPATIBLE DATA GENERATION

FINALLY, EVOLVING TO DATASETS
THAT WORK TOGETHER

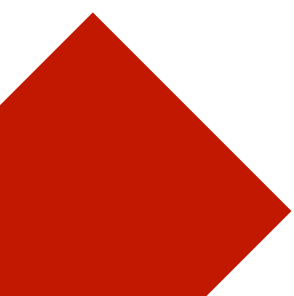
AUTHOR: ROBERT MACK, PH.D.

PUBLICATION: MAY 2025

Abstract

Compatible datasets help enterprises deliver value faster, better, and cheaper than traditional siloed datasets.

Research sponsored by:



COGINITI

EXECUTIVE SUMMARY

Since the dawn of computing, IT has relied on siloed datasets. Since then, businesses have rapidly evolved, and siloed datasets and their associated data administration methods have been updated in an attempt to meet those evolving business needs. However, we have reached an impasse where siloed datasets can no longer meet modern business needs. The complexity of the data architecture combined with the volume and variety of datasets is inundating IT. **There is no reasonable path forward using current data administration methods, and IT is falling far behind.**

A new generation of compatible data assets appears on the horizon. This groundbreaking compatible data generation represents a major step forward in business data strategy. Compatible data assets mean the end of siloed datasets and complex ETL-based data pipelines. This new generation of compatible data assets is built on a foundation of compatible interoperable datasets that seamlessly blend to form a distributed data fabric. This data asset modernization is designed to meet evolving business requirements now and for years to come. We are talking about a giant step forward for IT where datasets actually work together.

It is time to modernize your datasets and architecture if you want to use modern technologies effectively, like AI/ML. These modern technologies need a smooth, seamless data fabric foundation filled with high-quality data instead of the fragmented, disjointed, and distrusted siloed data foundation common to most businesses today. Has the time passed for your business to modernize its data foundation?

Working with compatible datasets is far faster, better, and cheaper than with siloed datasets. We will highlight these three fundamental drivers of business value throughout this whitepaper, but the impact on business opportunity is the most impressive. **Compatible data promotes the utility of seamless, high-quality data even beyond the enterprise and across industries.** Our independent research indicates that all datasets can be made compatible to provide unobstructed access to trusted information.

CONTENTS

Executive Summary.....	2
Introduction	3
Data Foundation Choices	4
Data Foundation modernization.....	4
The Data Compatibility Standards	5
The Data Architecture Modernization.....	6
Our Distributed Golden Data Fabric	6
The Efficiency of Data Compatibility	8
In Summary.....	8
About Maxxphase	9
About coginiti	9

About the Author:

Robert Mack, Ph.D.
Maxxphase, CEO

With over 35 years of IT experience, Bob has personally designed and implemented data systems for many clients in a variety of industries, including 8 Fortune 100 companies. Much of his most recent work has been implementing business intelligence solutions.

Bob has applied for and been issued 10 US patents on dataset and architecture design and implementation. He is the inventor of the Data Compatibility Methods that will reform business data administration and governance for many years to come.

INTRODUCTION

In the Compatible Data Generation, we at Maxxphase have discovered several flaws in current data administration methods that are detrimental to all IT organizations. We have engineered several innovative solutions to correct these data management flaws that significantly improve the efficiency and effectiveness of any IT organization. We combined our data management solutions to form Data Compatibility Methods.

In this age of AI/ML, outdated data strategies and data management methods impact your business's ability to utilize your data assets effectively. IT has always lagged as businesses evolve, and many business leaders express frustration with how IT manages their data assets.

Common reasons for business leader's frustrations include: ([Why Data Strategy is Key to Business Success.](#))

1. **Siloed Data:** All datasets, as designed, are disparate siloed datasets as they lack direct dataset interoperability.
2. **Poor Data Quality:** Today's datasets are known for their inconsistent, incomplete, or outdated data, often leading to inaccurate analyses and reduced trust in data-driven decisions.
3. **Inefficient Use of Resources:** Modern data management methods are very labor-intensive. Organizations invest heavily in data tools but fail to keep up with the existing workload.
4. **Poor Data Accessibility:** Decision-makers and teams often lack the means to fully leverage data, leading to underutilization of the data assets.
5. **Lack of Clear Data Strategy and Governance:** Data management can become chaotic and non-compliant with regulations without well-defined policies and accountability.
6. **Evolving Expectations:** As technologies like AI and machine learning grow in importance, some organizations struggle to adapt their data strategies fast enough to keep up.

In addition to the above list, we would add that the disparate data architecture currently used is a significant source of frustration. The disparate data architecture is fragmented and disjointed because of the siloed datasets. The poor data quality is an unintended artifact of the disparate data architecture. Within this whitepaper, we will address our approach to remedy these listed sources of frustration for business leaders.



According to Gartner, "Every year, poor data quality costs organizations an average \$12.9 million. Apart from the immediate impact on revenue, over the long term, poor quality data increases the complexity of data ecosystems and leads to poor decision making".

[How to Improve Your Data Quality - Gartner](#)

DATA FOUNDATION CHOICES

There are only two choices regarding the data foundation for your business. Those choices are either the disparate data foundation or the compatible data foundation.

The main distinction between data foundation choices is elementary. **The disparate data foundation is comprised of disparate siloed datasets that do not work together. This disparate data foundation is fragmented, disjointed, and filled with redundant data of poor quality.** Businesses that want to advance using modern technologies, such as AI/ML, cannot, as the disparate data architecture does not provide a data foundation conducive to such modern endeavors.

In contrast, at Maxxphase, we explicitly designed the compatible data foundation to be seamless, flexible, and highly scalable. We use our patented Data Compatibility Standards to enforce data integrity between compatible datasets for the highest data quality. Compatible datasets are directly interoperable and so compatible that a distributed data fabric spontaneously forms. Now, businesses that want to advance using modern technologies have the proper data foundation for now and for many years to come. These interoperable datasets herald the start of Compatible Data Generation. Please refer to [AI and Data Compatibility are Perfect Together](#) for more information.



DATA FOUNDATION MODERNIZATION

Our Data Compatibility Methods provide an innovative modernization methodology for enriching your siloed datasets to become compatible and directly interoperable. **The enrichment process is very straightforward and, most importantly, noninvasive to the existing datasets or data architecture.**

Our independent research has found that each dataset's data context is critical to its interoperability with other datasets. This data context is the shell of master data that encapsulates each dataset. This data context shell establishes the significance of the entire dataset relative to other datasets. **We have determined that siloed datasets have a disparate data context. As such, the disparate datasets are isolated because they lack data context consistency and master data commonality.** This lack of a standardized data context in siloed datasets forced the use of expensive and time-consuming data integration methods with complex ETL-based data pipelines.

First Law of Direct Dataset Interoperability

Siloed datasets have a disparate data context, while Directly Interoperable Datasets have a standardized data context.

Using our Data Compatibility Methods, we encapsulate any siloed dataset within our standardized data context to form a compatible dataset. Most importantly, the dataset encapsulation process is a noninvasive enrichment, so we don't change your existing data or architecture. We merely add to the data context of each dataset.

Working with compatible, directly interoperable datasets is a pleasure compared to the tedium of working with siloed data.

The data centralization bottleneck created by complex data pipelines with handcrafted data transformations for integrating siloed datasets is no longer a problem. Instead, you can modernize each dataset independently using our Data Compatibility Methods. Since the standardized data context is reusable, the more datasets you modernize, the less work required per dataset. The data foundation modernization can be achieved quickly and at minimal cost due to the reduced work effort and rapid time to market.

Second Law of Direct Dataset Interoperability

Any siloed dataset can be noninvasively enriched to become a directly interoperable dataset.

THE DATA COMPATIBILITY STANDARDS

Data Compatibility Standards are master data domain-specific reusable components of a Standardized Data Context. Data Compatibility Standards used with a dataset depend on the master data domains within the dataset. A Standardized Data Context can contain several master data domain-specific Data Compatibility Standards. When copies of a Data Compatibility Standard exist in multiple compatible datasets, the copies can be joined, thus joining the compatible datasets and supporting direct dataset interoperability.

Our patented Data Compatibility Standards provide:

- **Master metadata and data content consistency**
- **Data quality improvements from data integrity enforcement**
- **Data warehouse functionality within and across compatible datasets**
- **Direct dataset interoperability for immediate data access from a seamless data fabric**



Data Compatibility Standards are reusable data model objects and reusable dataset objects. The same Data Compatibility Standard is typically used throughout the business and may be used by many organizations or industries. You can utilize a specific Data Compatibility Standard in hundreds or thousands of compatible datasets. Fortunately, when enriching datasets, we do not touch the existing data; instead, we enrich or add to the dataset. **Once we establish the Data Compatibility Standards, entire datasets can be modernized in a few minutes or hours and immediately incorporated into a distributed data fabric. Gone are the complex ETL-based data pipelines that have no place in a Compatible Data Architecture.**

THE DATA ARCHITECTURE MODERNIZATION

When datasets are directly interoperable, a fascinating thing happens. Directly interoperable datasets are so compatible that they spontaneously form a seamless distributed data fabric. There is no limit to the number of compatible datasets you can incorporate into the data fabric, and any dataset that complies with the Data Compatibility Standards can participate. Now, all your datasets work together in ways not possible with previous data architectures. Every compatible dataset enriches every other as we focus on the entire data architecture instead of individual siloed datasets.

Third Law of Direct Dataset Interoperability

Directly interoperable datasets collectively function as a seamless distributed data fabric.

Each compatible dataset is a modular plug-and-play component of a dynamic distributed data fabric. Since we enforce end-to-end data integrity within the data fabric, we eliminated the data silos that have been a recognized hindrance to businesses for over three decades. Your data fabric is seamless, and your data quality has vastly improved over disparate data architectures. You can now blend any compatible datasets on demand.



Some of our goals for the distributed data fabric architecture are:

- ✓ Enrich data silos to be compatible interoperable datasets
- ✓ Eliminate data integration and their ETL-based data pipelines
- ✓ Remove data architecture complexity
- ✓ Provide end-to-end data integrity to improve data quality
- ✓ Provide unlimited data fabric scalability
- ✓ Provide a seamless data architecture
- ✓ Reduce the cost of ownership
- ✓ Improve time to market and ROI

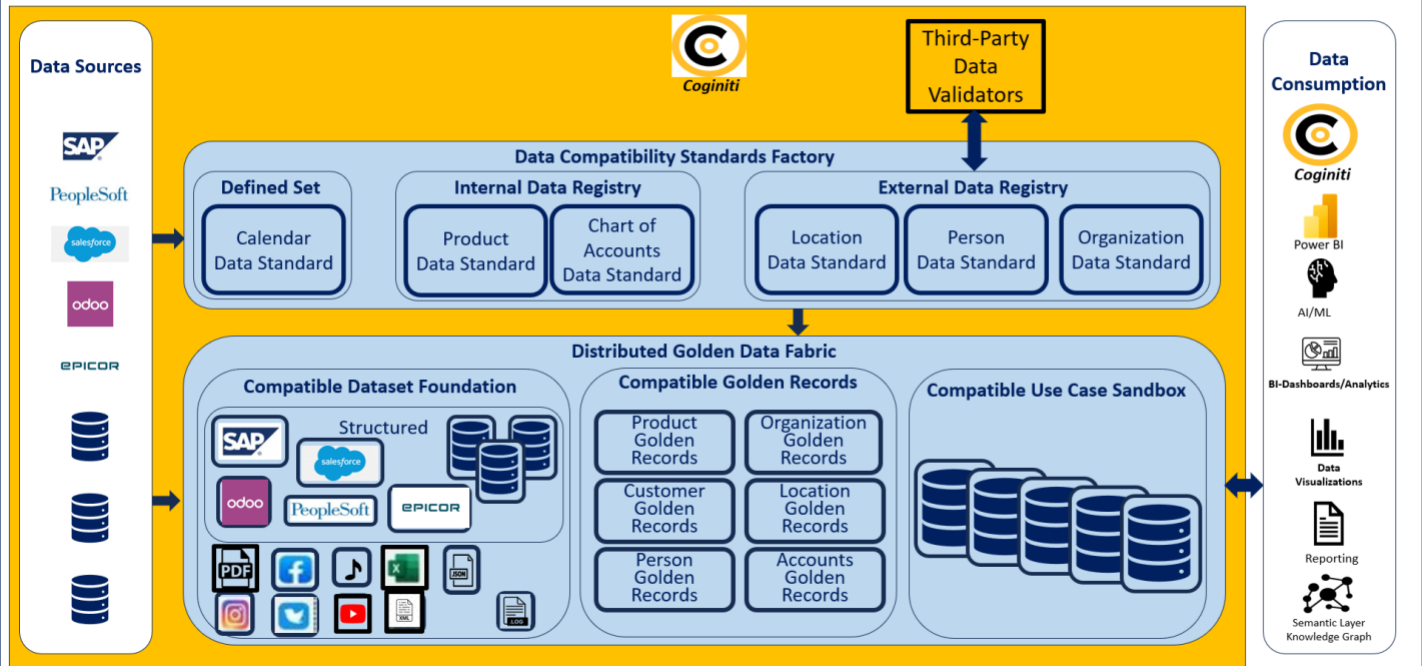
We meet all these goals with our Distributed Golden Data Fabric.

OUR DISTRIBUTED GOLDEN DATA FABRIC

Our design strategy for the Distributed Golden Data Fabric was to provide a seamless data environment with end-to-end data integrity and built-in master data governance. We implement our data governance in both the Data Compatibility Standards and the golden data records for critical master data management within the business. **This data architecture was specifically designed to support modern technologies such as AI/ML, semantics, and knowledge graphs.**

Within the data fabric, all datasets are analytics-ready, modular plug-and-play components. You may combine any datasets on demand to collectively function like a single dataset. Each dataset individually supports the functionality of a data warehouse, as do multiple datasets collectively. Within the data fabric, we also provide compatible golden data records for your business's major master data domains. This area of the data fabric provides 360° master data content datasets in the form of compatible components to be shared throughout the fabric. Compatible datasets can also be easily combined and materialized to support specific use cases whenever desired. These materialized datasets remain compatible with others, including the governed golden records datasets.

Distributed Golden Data Fabric Architecture



1. **Source Datasets**— Incompatible structured data sources generated internally or externally. Structured datasets are replicated into the Golden Data Fabric to separate them from their operational data systems.
2. **Data Compatibility Standards Factory**—Data Compatibility Standards are created and maintained in the factory. Each standard is specifically designed for a master data domain. Each master data domain instance from each source dataset is uniquely identified and standardized in the factory for use in the data fabric.
3. **Distributed Golden Data Fabric** – The data fabric contains all compatible datasets. All compatible datasets are directly interoperable and characterized as analytics-ready, modular, and plug-and-play. We subdivided the Golden Data Fabric by types of compatible datasets, such as:
 - a. **Compatible Dataset Foundation**—This subdivision of the data fabric contains the most granular structured, semi-structured, and unstructured datasets. These datasets are the foundation of the data fabric. Source datasets that are associated with an operational data system are typically replicated into this subdivision. Other original non-operational datasets can be stored in the compatible dataset foundation as well. All datasets in the foundation are enriched with Data Compatibility Standards from the Data Compatibility Standards Factory.
 - b. **Compatible Golden Records**—This subdivision of the Golden Data Fabric contains compatible master datasets where the business's golden records are maintained. Of course, the compatible golden data records can be shared with any other compatible datasets as needed.
 - c. **Compatible Use Case Information** – Any subset of data combined, derived, and aggregated from the Compatible Dataset Foundation datasets can be materialized into a use case-focused compatible dataset for a more focused business use, such as KPI dashboards, periodic reporting, compatible data warehouse, and, of course, AI/ML.
4. **Data Consumption**—Data management and other dataset tools to support analytics, report generation, and business intelligence have access to the distributed golden data fabric and other areas of the architecture.

Beyond the obvious technical advantages of the Golden Data Fabric, we wish to list several other essential technical features:

- ✓ **Centralized Data Governance Implemented with Standards**
- ✓ **Golden Master Data Management and Data Stewardship**
- ✓ **The Chronological History of Golden Records**
- ✓ **Data Audit Trail Retained from Source Dataset to Insight and Back**
- ✓ **Autonomous Dataset Ownership Retention**

The improvements between the Disparate Data Generation and the Compatible Data Generation in efficiency, agility, scalability, reliability, and time to market are astonishing. Beyond improved features, the Compatible Data Generation removes the complexity of working with your data while slashing ownership costs. Perhaps the best news is that enriching siloed datasets to become compatible datasets is relatively easy while being noninvasive to the existing datasets or the implemented data administration methods.

THE EFFICIENCY OF DATA COMPATIBILITY

Our Data Compatibility Methods are faster, better, and less expensive because we enrich existing datasets without altering the existing data or metadata. Our methods are noninvasive to the existing datasets or the existing data architecture. Data Compatibility Methods are better because they provide a seamless data fabric of the highest data quality. No other existing data architectures come close to the capabilities of the Distributed Golden Data Fabric. It is cheaper because far less work is needed, and we do not utilize complex handcrafted data pipelines. Our Data Compatibility Standards are reused on every compatible dataset. As such, we implement universal dataset interactions on every compatible dataset, and each enriched dataset has the same added capabilities.

Almost everyone wants to use modern technologies like AI/ML. However, when your data architecture is based on siloed datasets, data fragmentation and the lack of data quality become intolerable. Your business has outgrown your disparate data foundation, and it is no longer suitable for your evolving business needs. When you make your datasets compatible, you save time, effort, and capital for your current business data initiatives and any future projects. Information technology has been waiting for a solution like data compatibility for over three decades.

IN SUMMARY

We believe that disparate siloed data foundations for IT are fast becoming obsolete. The volume and variety of data have increased dramatically, and IT has struggled to keep pace with evolving business needs. Compounded by the emergence of newer technologies such as AI/ML, a digital data foundation based on direct dataset interoperability is essential.

We explicitly designed our Data Compatibility Methods to handle an unlimited quantity of datasets while providing a Distributed Data Fabric of reliable, high-quality data. Data Compatibility provides a simplified path forward for information technology that will become the data foundation for many years to come.

Important Facts to Remember:

- ✓ Compatible Datasets are Analytics-Ready, Modular, and Plug-and-Play.
- ✓ Compatible Datasets are formed by adding Data Compatibility Standards.
- ✓ We don't touch your data. - Data Compatibility Standards are noninvasive.
- ✓ A Distributed Data Fabric spontaneously forms from Compatible Datasets.
- ✓ Data Compatibility is easy to implement and maintain.

ABOUT MAXXPHASE

With ten US-issued patents, we are the world's leading experts in modern dataset and data architecture design, implementation, and management. We are data strategists who enhance traditional datasets to be directly interoperable. For the first time, datasets blend seamlessly to provide on-demand accessibility to any needed information. These compatible datasets are so interoperable that they spontaneously form a distributed data fabric. The distributed data fabric provides seamless data of very high quality required by modern technologies like Artificial Intelligence, Machine Learning, Semantics, and Knowledge Graph.



ABOUT COGINITI

Coginiti is an AI-enabled collaborative data operations platform. Fortune 500 enterprises and government agencies—including the largest institutions in insurance, banking, healthcare, technology and retail—trust Coginiti to build, publish, and consume trusted data products.

Coginiti empowers data engineers, analysts, and business users to harness the full power of their data. Coginiti connects to any data source, and with its unique hybrid query capabilities, enables flexible, cost-efficient, and scalable data processing.



Get Started Today:

Most organizations request a presentation/demonstration to begin their Data Compatibility journey. Once it is time to engage more directly, we find use case discussions useful, often followed by a Proof-of-Concept project. We have also done Data Architecture Assessments to design a data model and an enriched Compatible Distributed Data Fabric. Please contact us at info@Maxxphase.com to request whatever you may need.

Intellectual Property Notice

Maxxphase Inc. holds US Patents 7,979,475; 8,554,801; 8,874,619; 9,552,380; 10,417,263; 10,545,937; 11,341,171; 11,893,046 B2; RE48,312 E; and RE50,273 E along with multiple pending patents. We claim trademarks on the terms Data Compatibility Standards™, Compatible Data Modeling™, Direct Dataset Interoperability™, and Distributed Data Fabric™. © 2025 Maxxphase Corporation. All rights reserved.