# MIDAS

## Multifaceted Intelligent Data Access and Sharing

**Author:** Sînică Alboaie, PhD, Axiologic Research

**Purpose:** A report summarising MIDAS initial public proposal
**Visibility**: Public
**Date:** February 2024
**Version:** 0.1

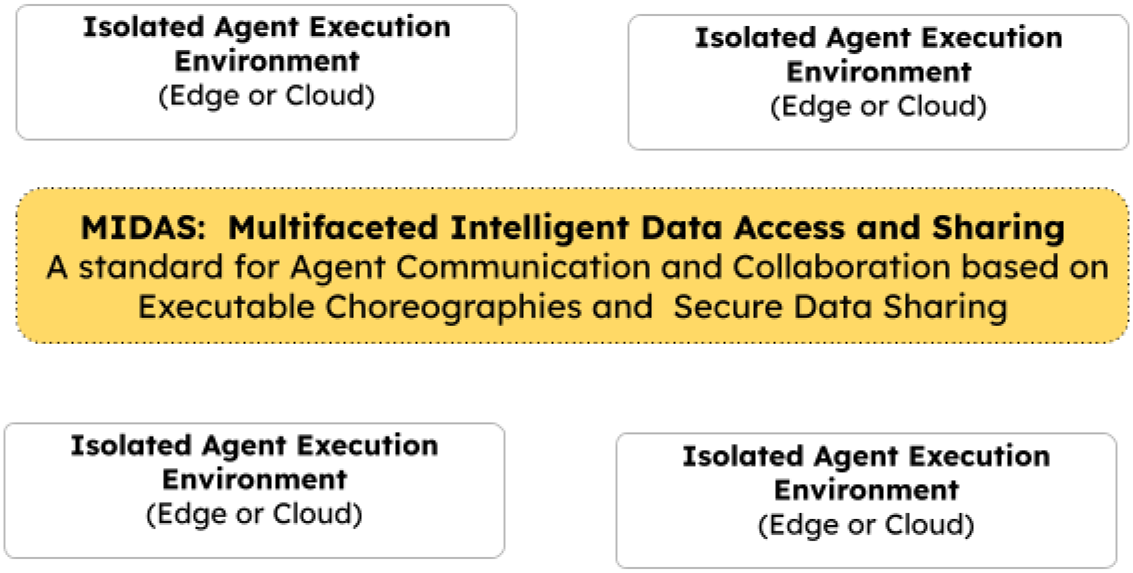
## Abstract

This document outlines Axiologic Research's vision for the evolution of artificial intelligence (AI), emphasizing the emergence of intelligent agent swarms as crucial for human safety. It argues that Artificial General Intelligence (AGI) will stem from the synergy of expert Large Language Models (LLMs), various AI technologies, and symbolic reasoning, forming a "swarm of intelligent agents." This concept enables complex interactions humans can still understand, laying the groundwork for advanced intelligence development. The paper stresses the importance of ensuring these agents' security, privacy, and alignment, suggesting they operate in isolated environments to mitigate threats from Advanced Persistent Threats (APTs) or unaligned superintelligences. MIDAS is introduced as a technology facilitating secure, verifiable communication among agents using executable choreographies, leveraging OpenDSU technologies for decentralised data sharing. Furthermore, the document previews the development of AssistOS, an open-source AI platform aiming to standardise communication in multi-agent systems. It also explores "executable choreography" as a means for decentralised, secure agent interaction, underscoring the necessity of these approaches for secure, privacy-preserving systems in multi-agent environments. The conclusion highlights the significance of establishing a secure communication and information-sharing standard among intelligent agents, positioning executable choreographies, decentralised sharing, and self-validating data as essential to obtain secure and aligned AI systems.

## What is MIDAS?

The vision set forth by Axiologic Research on the evolution of AI, as elaborated in our prior reports [IR1], [IR2], [IR3], posits that the next decade will see the emergence of swarms of intelligent agents as a pivotal development for human safety. These agents must be safeguarded through technical security measures, such as operating system security, cryptography, and strategies for integrating intelligence. This is to ensure that any defects or malicious intents within a group of agents do not compromise the collective alignment of the swarms.

Our primary research hypothesis contends that Artificial General Intelligence (AGI) will arise from the collaborative efforts of expert Large Language Models (LLMs), various AI technologies, and symbolic reasoning rather than from an isolated LLM. This concept, referred to as a "swarm of intelligent agents," highlights the ability of specialised, intelligent components to perform complex interactions in a manner that humans can understand, termed choreographies. This lays the foundation for the development of advanced intelligence.


Isolated Agent Execution Environment (Edge or Cloud)
Isolated Agent Execution Environment (Edge or Cloud)
MIDAS: Multifaceted Intelligent Data Access and Sharing
A standard for Agent Communication and Collaboration based on Executable Choreographies and Secure Data Sharing
Isolated Agent Execution Environment (Edge or Cloud)
Isolated Agent Execution Environment (Edge or Cloud)

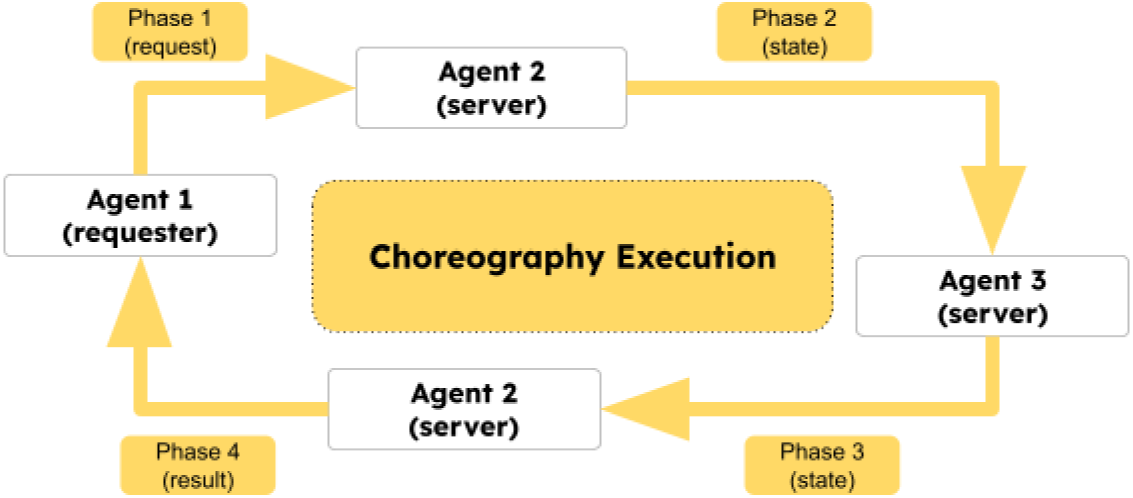
The most crucial insight we aim to convey at this level is that the AI systems of the future will be multi-agent. However, to ensure the agents' security, privacy, and alignment from the perspective of Advanced Persistent Threats (APTs) or superintelligences that may go out of control, agents must operate in as isolated environments as possible. Additionally, other ideas for restricting agents are required [IR5]. It suffices to say that achieving AGI or Superintelligent General Intelligence (SGI) will likely necessitate hundreds, if not thousands, of isolated agents working together. This collaboration entails real-time communication as well as working on shared documents. MIDAS is envisioned as a technology that enables agents to communicate securely and verifiably using executable choreographies and share documents in a granular and decentralised manner among themselves. Here, we refer to the intention to use technologies from OpenDSU, such as DSUs (Data-Sharing Units)[7] and SVD (Self-Verifying Data)[8].

To bring these ideas to fruition, Axiologic Research is embarking on an ambitious project to develop an AI platform for open-source applications that leverage generative AIs and multi-agent systems. This project is expected to evolve into a fully-fledged open-source operating system called AssistOS. For an overview of AssistOS, refer to our report [IR4]. Additionally, for a deeper understanding of our perspectives on AI alignment, we recommend reviewing our report [IR5].

## Swarm Communication and Choreographies

The concept of "choreography" involves having a formalised description of a communication protocol between independent entities. For instance, detailing the inputs and outputs for a web API constitutes a choreography. The API call facilitates data flow between a client and a server and then back to the client. This is like a “choreography execution”. The “choreography” concept becomes more useful when multiple entities are engaged in a singular communication "flow", allowing for discussions on choreographies or orchestrations. Orchestration refers to scenarios where a central conductor manages all calls and aggregates results. In contrast, "choreography" applies when no central conductor exists, and the entities coordinate their interactions independently.

An "executable choreography" represents a choreography  (a flow or communication protocol description )as a dynamic script that can be executed across various network nodes. This approach enables complex interactions across a distributed network, emphasising the autonomous and coordinated execution of processes without needing a centralised controller.



The concept of "swarm communication" [2] is just a method for implementing "executable choreographies" [3]. In swarm communication, each phase of the communication flow is represented as code associated with locations of execution.  The swarm communication comes with the intuition that the model of the  computational flows is a  jump between  independent nodes, each gathering new insights about each node. It is like an RPC call from a client, but instead of hitting only one server, there is a series of executions in multiple nodes, and eventually, the results are returned to the client. Even more complex “communication patterns” could be described; the linear one is the simplest.

Each jump will also carry some data obtained in each node; therefore, the execution of a choreography has some state that is transferred with each jump.

From the privacy and security point of view, the best practice of swarm communication is to put the results of computations within the flow while deliberately avoiding transferring confidential data from one node to another.  This could involve trust issues about the execution, but digital signatures could help.  It is simple but not a very common concept because most use cases are straightforward, and RPC,  with eventually some orchestration, is enough. However, with the advent of Large Language Models (LLMs), executable choreographies could become essential to future AI systems architectures. This is because they facilitate proper isolation of each agent's execution environment and enable the verification of communication between agents while maintaining the decentralised nature of the agent swarm.

An example that clearly illustrates this point is that with the increasing intelligence of LLMs (Large Language Models), applications will begin incorporating "strings" that represent instructions for LLMs. The most appropriate place for these strings is within choreographies rather than scattered throughout the application's code. From a security standpoint, it is sometimes necessary for agents to be highly restricted in their responses, or they might have a role-based system that allows different types of agents to respond differently, or not at all, to unauthorised questions or unauthorised agents.

An important concept is that of 'verifiable choreographies. Static or dynamic checking of the choreographies becomes easier, as it does not involve checking the whole code but only the communication protocol itself. The idea is that each agent offers generic APIs and capabilities. Still, the business logic stays at the level of the choreography, offering a decoupling from the long-term code and the ease of changing the business code. Although choreographies inherently promote decentralisation and peer-to-peer (P2P) communication among agents, various verification techniques can be implemented transparently, primarily in the communication layer. This layer typically involves message queues for each agent. Still, it can also be more formally established when the choreography is approved by each agent or by the system owners where the swarm of agents operates. Other verification techniques could involve cryptographic methods and more advanced technical approaches, such as Zero Knowledge Proof or self-validating data [8], where each agent digitally signs its contributions or ensures specific characteristics related to computational integrity.

## Applying OpenDSU Research in MIDAS

Our team is a significant contributor and the leader of the OpenDSU open-source project [6].

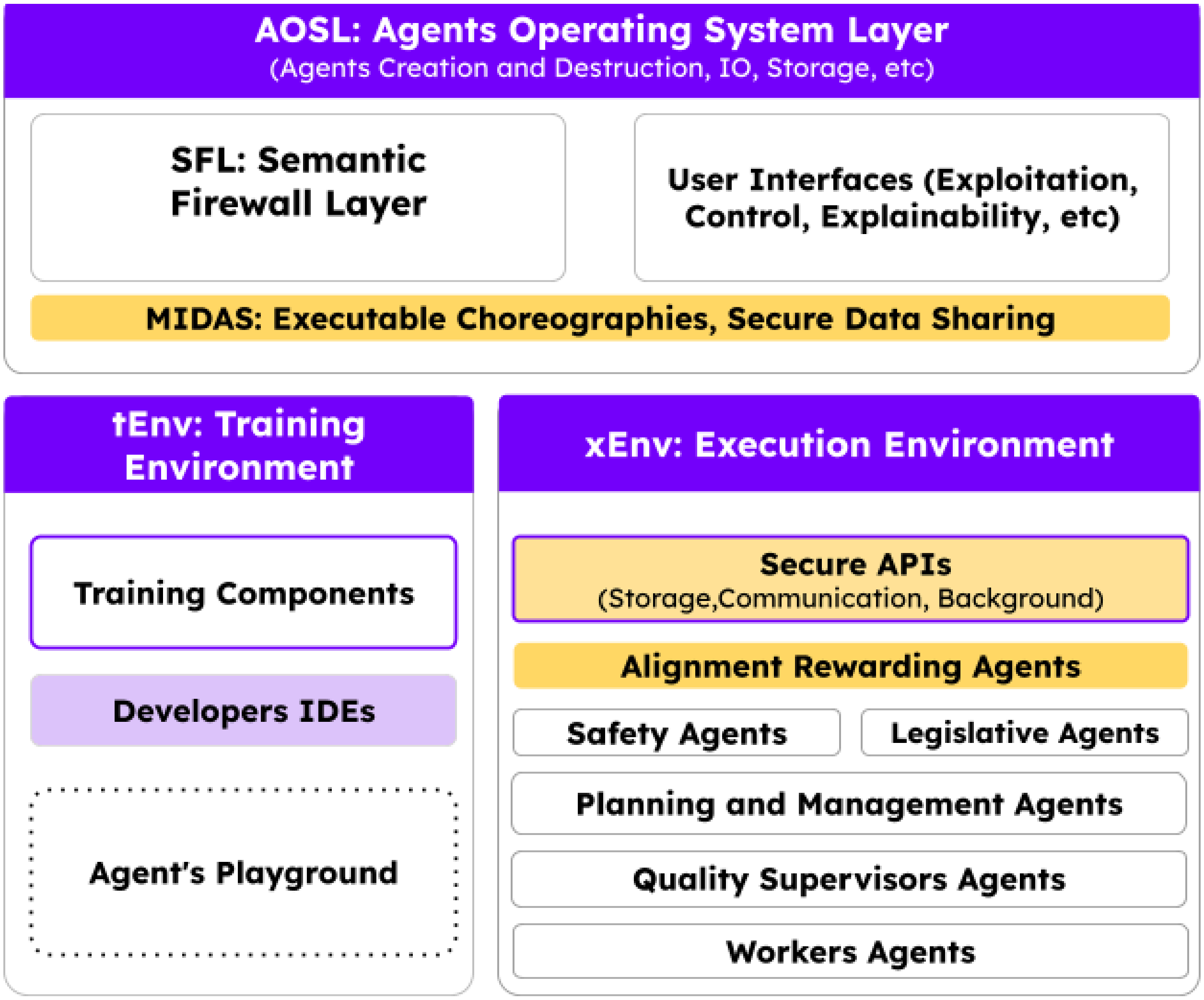
OpenDSU presents an innovative solution to decentralised data-sharing challenges by leveraging distributed ledgers, such as blockchains. In scenarios that demand secure data sharing, each entity manages a "domain" for data segregation and shares a common "domain" for "digital twins" or "business processes" involving multiple stakeholders.

OpenDSU provides a standard data-sharing protocol for decentralised applications operating in multiple access domains. Most importantly, it offers an elegant way of working with cryptography, where each DSU (Data Sharing Unit) can be seen as a document with a unique identity controlled by a cryptographic key. It allows for sharing in read-only or write mode by simply sharing derived keys or adhering to cryptographic key derivation protocols and pre-shared keys. The fundamental intuition is that in complex multi-agent environments, each agent will have access only to specific DSUs[7] they work on, allowing choreographies to transmit complex data without actually transporting big messages, thereby providing security by design and privacy by design. A centralised system used for sharing is an obvious target for attacks, but if a cryptographic key controls sharing, the attack surface is significantly reduced.

The concepts of "executable choreography" based on "swarm communication," DSU as a decentralised data-sharing method, and SVD as a method to implement microledgers are fundamental outcomes of our research over the past decades. They have the potential to become elementary concepts taught in university courses. The complexity introduced by multi-agent systems makes us confident that these approaches are extremely necessary for MIDAS to create proper AI systems in terms of security and privacy.

## MIDAS in AssistOS Architecture

MIDAS is planned to play a significant role in the architecture of AssistOS, and possibly even beyond AssistOS, as it aims to establish a standard for communication and data sharing in multi-agent environments. Implementing MIDAS within AssistOS is intended to be an open-source realisation of this standard without precluding other implementations. MIDAS seeks to become the TCP/IP or HTTP for the future internet, where millions or billions of intelligent agents interact and share data.


AOSL: Agents Operating System Layer
(Agents Creation and Destruction, IO, Storage, etc)
SFL: Semantic Firewall Layer
User Interfaces (Exploitation, Control, Explainability, etc)
MIDAS: Executable Choreographies, Secure Data Sharing
tEnv: Training Environment
Training Components
Developers IDEs
Agent's Playground
xEnv: Execution Environment
Secure APIs
(Storage,Communication, Background)
Alignment Rewarding Agents
Safety Agents
Legislative Agents
Planning and Management Agents
Quality Supervisors Agents
Workers Agents

We recommend the reader to read our other research reports at www.axiologic.net/research for details on the multi-agent architecture envisioned for AssistOS and an overview of various methods for achieving alignment in multi-agent systems [IR5]. In brief, however, the idea is that MIDAS serves as a layer that allows for oversight by a Semantic Firewall component in all environments that offer agent virtualisation. This setup ensures the correct adherence to security rules and desired alignment constraints.

## Conclusions

One of the most critical uses of the Internet is the search for information with the hope of gaining the highest quality insights about the experiences and knowledge of other people. We believe that delegating this effort to intelligent agents is becoming plausible, and our research aids in creating solid and secure architectures for data sharing and searching on the next-generation Internet, which could be imagined as a vast and decentralised environment of AI-controlled agents. This report aims to explain the necessity of creating a standard for secure communication and information sharing among intelligent agents as simply as possible. It also intends to demonstrate why executable choreographies, decentralised sharing, and self-validating data —technologies offered by OpenDSU— are steps in the right direction and somewhat inevitable if we wish to keep pace with the evolution of artificial intelligence.

Inspired by the story of King Midas, whose touch converted all to gold, the MIDAS programming model is envisioned to provide AI agents with the potential to generate value.

## References

[IR1] AI Market Evolution report (2023) https://www.axiologic.net/downloads/report_ai_market.pdf
[IR2] Internal Report Super-intelligence Classification (2023)
https://www.axiologic.net/downloads/report_super_intelligences.pdf
[IR3] An internal report on the Social Impact of Controllable AIs (Aligned AIs)
https://www.axiologic.net/downloads/report_social_impact.pdf
[IR4] AssistOS Initial Vision https://www.axiologic.net/downloads/report_assistos.pdf
[IR5] AI Alignment Research Report (2024)
https://www.axiologic.net/downloads/report_ai_alignment.pdf

[1] https://en.wikipedia.org/wiki/Active_message
[2] "Swarm Communication – a Messaging Pattern Proposal for Dynamic Scalability in Cloud"
L. Alboaie, S. Alboaie, P. Andrei, At 15th IEEE International Conference on High-Performance Computing and Communications (HPCC 2013). Zhangjiajie, China, November 2013.
[3] Extending swarm communication to unify choreography and long-lived processes
L. Alboaie, S. Alboaie, T. Barbu, 3rd International Conference on Information Systems Development (ISD2014 Croatia) pp. 375–382, 2014.
[4] "Levels of Privacy for e-Health systems in the Cloud era" S. Alboaie, L. Alboaie, A. Panu, 24th International Conference on Information Systems Development. Harbin, China, August 25-27, 2015.
[5] Consider checking the www.axiologic.net/research page that provides additional reports and details on the AssistsOS components and concepts.
[6] OpenDSU Documentation ( www.opendsu.org )
[7] DSU Concept https://www.opendsu.org/pages/concepts/DSU%20Introduction%20(RFC-001).html
[8] SVD Concept https://www.opendsu.org/pages/contributors/Self-Validating%20Data%20(RFC-036).html