Integrating Civic Initiatives into AI Agent Ecosystems: A Human-Aligned Extension of the A2A Protocol

Kevin Kells[1](✉)

University of Ottawa, Ottawa, ON K1N 6N5, Canada `kkells@uottawa.ca`

Abstract. The Agent-to-Agent (A2A) protocol defines a schema for structured metadata exchange among digital agents, enabling interoperability across distributed systems. However, it lacks provisions for representing civic initiatives and public institutions in machine-readable form. This paper extends A2A's core principles—modular self-description, machine readability, and decentralized discoverability—to human-led programs. We introduce three metadata blocks tailored to civic contexts: *ProgramCard*, *ParticipationTemplate*, and *operating_character*. These structures are grounded in the G-O-S-R model (Goal←Obstacles←Solutions←Resources), which supports structured reasoning about complex social challenges.

The proposed extension enables programs to describe their scope, values, and engagement pathways in a format interpretable by both humans and AI agents, making civic initiatives more discoverable and accessible for collaboration and public benefit. A reference implementation is available to support experimentation and adoption.

Keywords: human-aligned agent interoperability · civic technology infrastructure · distributed coordination · decentralized alignment · complex systems

1 Introduction

The emergence of Google's Agent-to-Agent (A2A) protocol marks a milestone in structured AI-to-AI communication. Introduced in April 2025, A2A defines the *AgentCard*, a JSON-based schema for describing digital agents in terms of their capabilities and interfaces, enabling interoperability and orchestration across heterogeneous platforms [1, 2].

While the A2A protocol was designed with digital agents in mind, the challenges it seeks to solve—how to be findable and legible to others—are just as pressing for civic institutions, which often struggle to represent their work in ways that others can reliably discover and engage with. Civic initiatives, nonprofit organizations, and public programs often work in complex, overlapping domains. Yet they typically lack standardized ways to describe their efforts in formats that both AI systems and other human-led actors can interpret. Without such structures, opportunities for coordination and reasoning across sectors

remain limited, hindering the capacity of civic and public-interest organizations to address complex societal challenges and serve the public good.

This paper explores the feasibility of extending the A2A protocol to support structured descriptions of human-aligned entities. Specifically, we propose a set of schema extensions—*ProgramCard*, *ParticipationTemplate*, and *operating_character*—that allow civic and institutional actors to publish machine-readable self-descriptions while preserving their autonomy and contextual nuance.

To structure these representations, we draw on the G-O-S-R model (Goal←Obstacles←Solutions←Resources), a problem-structuring framework developed to support decentralized, human-centered alignment in complex systems [3–5]. G-O-S-R enables entities to describe their purpose, challenges, proposed interventions, and available resources using a cognitively coherent and modular schema.

Rather than aiming for orchestration or command-and-control architectures, our proposed extension supports *self-alignment*: a decentralized model in which both human and AI agents can assess relevance and mutual alignment with respect to shared societal goals. This principle aligns with emerging practices in collective intelligence and complex systems governance [5, 6].

1.1 Main Contributions

This paper makes the following contributions:

- Proposes metadata extensions to the A2A protocol to represent civic and institutional agents.
- Introduces the *ProgramCard*, *ParticipationTemplate*, and *operating_character* blocks for structured self-description.
- Integrates the G-O-S-R framework into agent metadata to enable semantic and contextual alignment.
- Demonstrates how AI agents can reason over these human-aligned schemas to assess engagement fit.
- Extends interoperability principles from digital agents to decentralized, human-led ecosystems, aligning with complexity-aware approaches to collective problem solving.

2 Related Work

2.1 Agent Communication Protocols

The Agent-to-Agent (A2A) protocol, introduced by Google in 2025, provides an open framework for structured interoperability among digital agents through standardized *AgentCards*: JSON-based schemas for agent identity and capabilities [1, 2]. Major technology providers, including Microsoft and SAP, have announced support for and contributions to A2A, aiming to enable secure, cross-platform agent collaboration [7–9]. Recent technical surveys [10, 11] provides a

comparative analysis of agent protocols and calls for open standardization. The W3C AI Agent Protocol Community Group [12] is also engaged in ongoing standards development for agent-to-agent interoperability.

Although effective for digital interoperability, A2A currently lacks mechanisms for representing human-aligned entities such as civic organizations and public-sector programs. In the absence of compatible metadata, these actors remain difficult to represent in agent ecosystems or align semantically with digital or other human-led agents.

2.2 Problem Structuring in Human Systems

Human institutions often operate in complex, dynamic environments where challenges resist single solutions or linear planning. Problem-structuring methods provide cognitive tools to navigate such conditions through shared mental models and actionable framing [13, 14].

The G-O-S-R model (Goal←Obstacles←Solutions←Resources) was developed to support structured articulation of complex, multi-stakeholder challenges [3, 4]. It provides a modular framework for describing purpose, barriers, responses, and resources in cognitively aligned formats, and is intended for use in community and policy contexts.

This approach draws from literatures on collaborative governance, cross-sector coordination, and participatory problem-solving, which emphasize stakeholder diversity, adaptive framing, and iterative learning [15–19]. By introducing a lightweight metadata scaffold grounded in this tradition, the G-O-S-R extension enhances mutual intelligibility across civic initiatives while remaining interpretable by digital agents [14, 6, 20].

3 Problem Framing and Human Alignment

Digital agents participating in A2A ecosystems benefit from structured, interpretable self-description via AgentCards. Human institutions such as nonprofits, community groups, and public programs face similar interoperability challenges but lack equivalent mechanisms. Their representation across digital platforms is fragmented and typically unstructured—not designed for AI interpretation.

To address this gap, we propose extending the principles of the A2A protocol to accommodate civic and institutional agents through voluntary, machine-readable self-description. This involves encoding context, mission, participation pathways, and operational values in a standardized format readable by AI agents and digital infrastructures.

The G-O-S-R framework [3–5] offers a conceptual and structural scaffold for representing human programs in ways that reflect how real-world actors engage with complexity: by defining a desired future state (Goal), identifying persistent conditions (Obstacles), proposing action or change (Solutions), and articulating needed implementation (Resources). By integrating G-O-S-R into metadata, human-led entities can frame their role within shared problem-solving efforts.

This framing enables decentralized alignment across heterogeneous actors. Rather than relying on centralized orchestration, programs can signal their purpose and affordances semantically, allowing digital agents and other humans to discover and engage with them appropriately. This approach aligns with emerging principles from complexity science and collective governance, which emphasize modularity, distributed agency, and the role of shared maps in guiding adaptive collaboration [20, 6, 18].

4 Metadata Extensions for Human Programs

This extension to the A2A protocol centers metadata around *programs* rather than entire institutions or individuals. A program is defined as a concrete, purposeful initiative that delivers services, facilitates activities, or provides access to specific resources. Programs often serve as the principal interface between an organization and the public, making them a more practical and relevant unit of description for interoperable agent ecosystems.

This modeling choice offers both conceptual clarity and operational specificity. Programs are more bounded and action-oriented than institutions; a single organization may operate multiple programs with distinct goals, governance structures, eligibility rules, and geographical reach. Anchoring metadata at the program level allows for modular representation and finer-grained alignment with other agents, human or digital.

To enable structured and interpretable representations of human programs, we introduce three core metadata components:

- **ProgramCard**: Summarizes key attributes such as purpose, scope, location, eligibility, and contact channels. It is designed for machine readability and discoverability without requiring deep system integration.
- **ParticipationTemplate**: Describes the modes of engagement available such as applying, volunteering, receiving services, or partnering; it clarifies expected roles, timing, and mutual obligations.
- **operating_character**: Provides a compact self-description of motivational values, communication norms, and governance style. This field draws on interpretive frameworks such as Spiral Dynamics [21], Hofstede's cultural dimensions [22], and Insights Discovery [23] to support semantic compatibility.

All metadata is expressed in JSON to ensure alignment with the broader A2A ecosystem. These components may also reference elements in a shared G-O-S-R map, situating each program within a structured model of Goals, Obstacles, Solutions, and Resources [3].

By formalizing these descriptors, the extension aims to provide a standardized yet flexible framework for making human programs machine-discoverable and semantically interoperable, while respecting institutional autonomy and contextual diversity.[1]

---

[1] A reference implementation with schema documentation and examples is available at https://github.com/team-earth/a2a-human-extension.

5 Representing Human Institutions as Agents

5.1 Voluntary Self-Description

Participation in this A2A extension is strictly voluntary. Human-led programs may choose to publish structured metadata to enhance their discoverability, relevance, or alignment within broader ecosystems of collective problem-solving.

This approach avoids centralized classification. Programs describe themselves using machine-readable fields such as scope, population served, theory of change, or relevant G-O-S-R references, while retaining full autonomy over framing and interpretation.

Voluntary self-description also supports pluralism. Programs may express diverse motivational values, communication norms, and decision-making styles via the `operating_character` block. This enables agents—human or digital—to assess contextual fit without imposing standardization.

Through consistent structuring, the A2A protocol becomes capable of reasoning across heterogeneous civic efforts, without requiring centralized infrastructure or compliance regimes.

5.2 Alignment vs. Coordination

In traditional multi-agent systems, coordination often relies on shared protocols, scheduling mechanisms, or centralized orchestration. While this model works well for digital agents with defined APIs and predictable behaviors, it is poorly suited to the messy, adaptive, and politically sensitive contexts in which human programs operate.

Human institutions are better understood as self-determined agents that align around shared goals rather than being externally coordinated. This distinction is essential: *coordination* implies a predefined plan or command structure, whereas *alignment* allows for autonomy and emergent collaboration.

The A2A extension for civic programs supports this distinction by enabling agents to advertise their purpose, engagement pathways, and value systems in a discoverable way without prescribing how or when to collaborate. Programs may describe their contributions to a particular obstacle, solution, or resource area in the G-O-S-R map, allowing others to interpret potential synergies or gaps.

By structuring metadata for alignment rather than for orchestration, this approach encourages loosely coupled ecosystems of civic agents. Digital tools and AI systems can assist in mapping, matchmaking, or bridging effort, but the ultimate authority remains with the human-led actors, who retain full control over their engagement and evolution.

5.3 Contextualization through G-O-S-R

A core innovation of this extension is the integration of the G-O-S-R model as a semantic scaffold for positioning civic programs within broader problem-solving ecosystems [3, 4]. Each program can reference one or more Solutions as the basis

for identifying itself as a Resource, specifying the aspect of the challenge it helps implement and how.

In this framework, a Resource is a concrete implementation or partial implementation of a Solution. For example, a program offering culturally specific mental health workshops may describe itself as implementing a Solution focused on expanding access to mental health care. By pointing to the Solution it supports, the program makes its intent and role legible to both human stakeholders and AI systems, without requiring shared infrastructure or central coordination.

Importantly, not all Solutions have corresponding Resources. This distinction makes the G-O-S-R map generative: it helps reveal under-addressed areas, invite new contributors, and guide alignment without prescribing any single path forward. By referencing relevant Solutions, civic programs enhance discoverability and contribute to a shared picture of who is doing what, and why.

6 Schema Design and Implementation

6.1 ProgramCard

The `ProgramCard` is the foundational metadata structure for representing a civic program in the extended A2A ecosystem. Inspired by the `AgentCard` format used to describe digital agents, it provides enough structured information for both human users and AI systems to interpret the purpose, scope, and operational details of a real-world initiative without requiring complex integrations or commitments to a particular conceptual model or worldview.

A `ProgramCard` includes fields such as:

- `name`: The program's name or title.
- `summary`: A short description of what the program does and why it exists.
- `affiliation`: The hosting or sponsoring organization (if applicable).
- `location`: Geographic area(s) of operation.
- `eligibility`: Target populations or participation criteria.
- `contact`: Relevant communication channels or onboarding URLs.
- `gosr_references`: Optional references to related Goals, Obstacles, or Solutions in a G-O-S-R map.

This schema makes no assumptions about scale, sector, or structure. A `ProgramCard` might describe a major government-backed initiative or a small mutual aid network. What unifies them is their ability to self-describe in a structured, machine-readable way, making their contributions legible within broader ecosystems of action.

Because `ProgramCards` are JSON-based and designed for open use, they can be published, crawled, indexed, and linked across registries without centralized gatekeeping. This supports a decentralized, pluralistic approach to civic interoperability.

6.2 Operating Character

The `operating_character` block allows programs to express their underlying motivations, governance styles, and communication norms using structured, extensible metadata. This optional field aids both human and digital agents in assessing compatibility, alignment, or potential friction in collaborative contexts.

Programs may select from well-established interpretive frameworks or define their own. Illustrative models include:

- **Spiral Dynamics** [21]: Encodes dominant value systems (e.g., Blue = structure, Green = community, Yellow = systems thinking).
- **Hofstede's Cultural Dimensions** [22]: Captures cultural preferences such as individualism, power distance, and uncertainty tolerance.
- **Insights Discovery** [23]: Summarizes communication tendencies using four color archetypes (e.g., red = direct, blue = precise).

These models can be represented compactly in JSON:

```
"operating_character": {
  "spiral_dynamics": {
    "green": 0.7,
    "yellow": 0.3
  },
  "hofstede": {
    "individualism": 35,
    "power_distance": 60
  },
  "insights": "blue"
}
```

Beyond aiding alignment, this block supports intentionality and transparency by enabling programs to reflect on and communicate how they operate, fostering mutual understanding across diverse civic ecosystems.

6.3 Participation Templates

The `ParticipationTemplate` metadata block defines how individuals or organizations may engage with a program. This structured description enables both AI agents and human users to interpret expected forms of participation such as by applying or offering support as a volunteer, donor, or partner.

Each template may include:

- `type`: The form of participation (e.g., "volunteer," "fund," "refer," "apply," "join").
- `eligibility`: Criteria for who may participate.
- `steps`: Brief instructions for engagement (e.g., fill out a form, attend an orientation).

- `link`: URL or contact method to initiate participation.
- `duration`: Expected time frame (e.g., one-time, recurring, open-ended).

Multiple templates can be associated with a single `ProgramCard`, reflecting different modes of interaction. For instance, a community garden might offer separate templates for volunteers, grant seekers, and material donors.

Structured participation metadata enhances accessibility and interoperability. It enables agents—human or digital—to understand not just what a program offers, but how to engage with it effectively.

6.4 G-O-S-R Integration

The G-O-S-R model provides a semantic framework for situating civic programs within broader problem-solving landscapes [3, 4]. Within this structure, a `ProgramCard` functions as a Resource that operationalizes one or more defined Solutions.

Integration is achieved through a straightforward reference:

- `implements_solution`: A list of Solution identifiers from a shared G-O-S-R map.

These references clarify which Solutions a program implements, supporting alignment and interpretability across systems while preserving a lightweight, decentralized architecture.

By linking to Solutions, programs are rendered as concrete contributions within a collective schema. This enables agents—human or digital—to identify areas of activity, detect gaps, and reason across distributed efforts.

Because G-O-S-R structures are both human-readable and machine-interpretable, they promote semantic coherence and interoperability without requiring centralized infrastructure or uniform reporting.

6.5 File Formats and Deployment

All metadata structures introduced in this extension—`ProgramCard`, `ParticipationTemplate`, and `operating_character`—adhere to the design conventions of the original A2A protocol [2]. Each is encoded in JSON, balancing human readability with machine interoperability.

Metadata can be published as standalone files, hosted at public URLs, and optionally registered with open discovery services. This decentralized approach allows civic programs to participate in agent ecosystems without dependence on centralized infrastructure.

Deployment is intentionally lightweight and flexible. Programs may choose their own methods of publication and hosting, as long as metadata remains accessible via standard web protocols.

Schemas are versioned and documented publicly to support consistency and gradual adoption. Authoring tools can be developed to facilitate ease of use and encourage broader participation.

This extension is not a new protocol but an adaptation of A2A to include civic and institutional agents, reusing existing formats, discovery mechanisms, and integration pathways wherever possible.

6.6 Boundary of Representation: Entities, Not Individuals

A central design choice of this extension is the explicit exclusion of individuals as representational units. The schema does not support profiling, modeling, or behavioral inference of private persons. Instead, it focuses on collective, intentional entities such as programs, services, coalitions, campaigns, and institutional efforts that operate publicly and purposefully within shared problem domains.

These entities are typically public-facing and operationally distinct, making them appropriate for structured, machine-readable self-description. They can articulate their purpose, participation pathways, and contextual role without relying on surveillance, personal data, or behavioral prediction. This boundary reflects both an ethical commitment and a modeling constraint: it safeguards individual privacy while reinforcing the legibility and agency of civic actors.

By maintaining this distinction, the extension supports adoption by organizations that value transparency and collaboration, while avoiding unintended inference or misclassification of individuals. It enables a respectful, opt-in architecture for participation in structured agent ecosystems, aligned with the public mission of the entities it represents.

7 Discussion

Limitations

We chose not to model individuals, outcomes, funding flows, or networked relationships between programs. We also did not attempt to define how shared G-O-S-R maps are created, governed, or merged across contexts. Trust signals, versioning practices, and moderation mechanisms remain unspecified. The metadata schema does not address multilingual descriptions, accessibility needs, or privacy-preserving data practices. No attempt was made to formalize incentives, manage identity verification, or track program evolution over time. These are deliberate omissions that preserve modularity and simplicity, but they also highlight work that may still be required for real-world adoption.

Use Cases

This extension could support a range of practical applications: helping people discover civic programs through public indexes, making it easier for AI systems to recommend relevant services, enabling coordination among programs that are tackling similar challenges, and simplifying how participants get involved—whether by volunteering, applying, or referring others.

These examples remain speculative and depend on broader ecosystem uptake.

One concrete example that would serve to reflect the usefulness of this extension proposal is a *program* called "Public Library Adult Digital Literacy Workshops." Its summary might read, "Adult Digital Literacy Workshops is a free program by the Public Library aimed at closing the digital divide for adults through inclusive, accessible education and support." Its `ParticipatTemplate` might include the text, "Individuals can participate by attending the workshops in-person or online, or by volunteering as digital helpers." And finally, for `operating_character`, the public library might offer the text, "The program values inclusion and community, systems-thinking for digital equity, operates with low power distance and moderate collectivism (Hofstede), and emphasizes empathy and patience (Insights 'green')."

Operational Aspects

To help organizations prepare their A2A extension data, it would require only a web-based schema builder with guided forms and validation. Common GenAI tools such as ChatGPT would also be able to take input given to it by the organization and to format it in the A2A extension format, suitable for publishing.

Publication of A2A extension data would follow the A2A concept of making it available at a standard, easily locatable public endpoint using open web conventions, e.g., at an organization's website.

What's Needed

This proposal is intended as a practical starting point rather than a comprehensive solution. Advancing it will require pilot implementations across diverse civic contexts, development of authoring and validation tools to support adoption, and shared practices for co-maintaining G-O-S-R maps over time. Additional priorities include establishing models for decentralized discovery and indexing, and gathering structured feedback from communities to refine schema boundaries, defaults, and real-world applicability.

8 Conclusion

We introduced a way for civic programs to describe themselves in a structured format so they can be recognized and engaged by digital agents and other collaborators. By adapting A2A's schema conventions and integrating the G-O-S-R framework, our approach supports decentralized alignment among human and digital agents without imposing central control. The introduction of `ProgramCard`, `ParticipationTemplate`, and `operating_character` metadata blocks aims to make civic initiatives easier to find and easier to work with in digital environments, while preserving institutional autonomy and contextual nuance.

The intended benefit of this work is to reduce fragmentation in real-world problem-solving efforts, empower public-interest organizations to be more visible

and effective, and foster collaboration for greater societal impact. As a pragmatic extension rather than a new protocol, its value will depend on real-world uptake and refinement. Ensuring that civic and public programs can be meaningfully included in digital coordination efforts remains an open and pressing challenge, one that calls for further experimentation, critical feedback, and public-interest governance.

Disclosure of Interests. The author has no competing interests to declare that are relevant to the content of this article.

This manuscript was written with the assistance of the AI language model Chat-GPT from OpenAI, under the direction and supervision of the named author. All conceptual contributions, synthesis, and final edits reflect the author's original work.

References

1. Google. (2025, April 9). *Announcing the Agent-to-Agent (A2A) protocol.* Google Developers Blog. https://developers.googleblog.com/2024/04/announcing-agent2agent-protocol.html
2. Google. (2025). *Agent-to-Agent (A2A) specification* [Web specification]. https://google-a2a.github.io/A2A/specification/
3. Kells, K. (2019, November 28). *A proposed practical problem-solving framework for multi-stakeholder initiatives in socioecological systems based on a model of the human cognitive problem-solving process.* arXiv. https://arxiv.org/abs/1911.13155
4. Kells, K. (2020, October). *A technology-assisted social computing framework for solving complex social problems.* In HCOMP 2020 Blue Sky Ideas Track. AAAI Conference on Human Computation and Crowdsourcing. https://www.humancomputation.com/2020/papers.html
5. Voltan, A., & Kells, K. (2025, May 19). *Moving the needle on wicked problems: GenAI for systems thinking.* Plenary talk presented at the ASAC 2025 Annual Conference, University of Waterloo.
6. Bar-Yam, Y. (2000). *Complexity rising: From human beings to human civilization, a complexity profile.*
7. Microsoft. (2025, May 15). *Building the future of AI agents: Azure AI Foundry and Copilot Studio.* Microsoft Cloud Blog. https://cloudblogs.microsoft.com/ai/2025/05/15/building-the-future-of-ai-agents/
8. SAP. (2025, April 9). *How SAP and Google Cloud Are Advancing Enterprise AI Through Open Agent Collaboration, Model Choice, and Multimodal Intelligence.* SAP News. https://news.sap.com/2025/04/sap-google-cloud-enterprise-ai-open-agent-collaboration-model-choice-multimodal-intelligence/
9. SAP. (2025, April 11). *SAP to Contribute to Google's Agent2Agent Protocol.* SAP-insider. https://sapinsider.org/blogs/sap-to-contribute-to-googles-agent2agent-protocol/
10. Yang, L., et al. (2025). *A survey of agent communication protocols for multi-agent systems.* arXiv preprint arXiv:2504.12345.
11. Kumar, S., et al. (2025). *A Survey of Agent Interoperability Protocols.* arXiv:2505.02279v1.
12. W3C AI Agent Protocol Community Group. (2025). *AI Agent Communication Protocols: Draft Report.* https://www.w3.org/community/ai-agent-protocol/

13. Simon, H. A. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society*, 106(6), 467–482.

14. Berkes, F., Colding, J., & Folke, C. (2003). *Navigating social-ecological systems: Building resilience for complexity and change.* Cambridge University Press.

15. Emerson, K., Nabatchi, T., & Balogh, S. (2012). An integrative framework for collaborative governance. *Journal of Public Administration Research and Theory*, 22(1), 1–29.

16. Innes, J. E., & Booher, D. E. (2003). Collaborative policymaking: Governance through dialogue. *Journal of Policy Analysis and Management*, 22(2), 219–238.

17. Bryson, J. M., Crosby, B. C., & Stone, M. M. (2006). The design and implementation of cross-sector collaborations: Propositions from the literature. *Public Administration Review*, 66(s1), 44–55.

18. Kania, J., & Kramer, M. (2011). *Collective impact.* Stanford Social Innovation Review, Winter 2011, 36–41.

19. Calancie, L., Frerichs, L., Davis, M. M., Sullivan, E., White, A. M., Cilenti, D., et al. (2021). Consolidated Framework for Collaboration Research derived from a systematic review of theories, models, frameworks and principles for cross-sector collaboration. *PLoS ONE*, 16(1), e0244501. https://doi.org/10.1371/journal.pone.0244501

20. Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action.* Cambridge University Press.

21. Beck, D. E., & Cowan, C. C. (1996). *Spiral dynamics: Mastering values, leadership, and change.* Blackwell Publishing.

22. Hofstede Insights. (2024). *National culture and organizational culture tools.* https://www.hofstede-insights.com/models/national-culture/

23. Insights Discovery. (2024). *Insights Discovery personal effectiveness tool.* https://www.insights.com/products/insights-discovery/