Proof of Personhood, Dual Agent Systems & Standards Path

privacymage — IKP Co-Chair Reconstruction

Session summary

This is the core PoP session — the one that directly feeds our working group item on Proof of Personhood: Biometrics, Systems + Agent Duality as a Personhood Preservation Mechanism. The session brought together biometric identity providers, decentralized identity researchers, standards body practitioners, and the IKP working group to workshop the three-graph model, hear from contributing protocols, explore the dual agent architecture, and map the ISO standardization path for the paper.

What landed

Three graphs, one overlap

privacymage presented the foundational model: three graphs that, when overlapped, produce proof of personhood. (1) **Knowledge substrate** — a knowledge graph representing domain understanding within a context (the “spell web” visualization was demonstrated live). (2) **Promise/intent graph** — the actions, commitments, and traversals a person makes through that knowledge. (3) **Trust graph** — the emergent property of knowledge meeting promise over time. The key insight: these three systems can sit separately, meaning the coordinate where knowledge meets promise to produce trust is extremely unique and difficult to replicate or spoof.

Biometric proof of personhood — contributing protocol presentation

A major biometric PoP protocol presented their approach: iris-based biometric capture using a hardware orb, processed into an iris code via secure multi-party computation (sMPC) to prove uniqueness without revealing identity. The protocol emphasizes: never saying *who* the human is, only that it’s a *unique* human. Use cases presented included age verification (ZKP-based proof of being over 18 without revealing date of birth) and human verification badges on social platforms.

Key technical detail: the orb captures biometric data (iris + face images) signed by the hardware device, kept by the user alone. The iris code is sent to sMPC to confirm uniqueness. Government ID can be linked locally (face matching between orb capture and passport) to build an evidence chain — all held privately on-device, with selective ZK disclosure.

The discussion between privacymage and the protocol team produced a critical distinction: proof of *a* human (uniqueness) vs. proof of *a specific* human (identity). The protocol deliberately stays far from specific identification, targeting anonymity sets no smaller than ~1 million people. For narrower disclosure (e.g., proving a bachelor’s degree), better ZK identity tooling is needed where the individual holds all evidence locally and discloses only the minimum required.

Dual credential system — First Person Project

privacymage presented the two-credential model from the First Person Project: (1) **Personhood credentials** — top-down, ecosystem-issued credentials recognizing someone as a member (World ID as an example; BGIN could issue its own based on its governance requirements). (2) **Verifiable relationship credentials** — bilateral, between two nodes in the trust graph, created through key ceremonies. The personhood credential makes you a node; the relationship credentials are the edges.

First customer: the Linux core kernel, using the dual credential system to manage maintainer access and code merge permissions. Results expected within 6-12 months.

Knowledge substrate as proof mechanism

privacymage demonstrated the knowledge graph tool live, having the room collectively navigate a path through the “agent privacy” knowledge substrate. Each participant chose nodes to visit, creating a unique constellation — a group-traversed path through shared knowledge. This constellation can be: downloaded as a machine-readable artifact, wrapped in a ZK proof, attached to a credential (DID/VC), and used as proof of having observed and understood a body of knowledge.

The concept extends to bilateral overlap: if two people each have their own knowledge substrate, they can create a controlled intersection — “discovery with zero knowledge” — finding where their interests overlap without disclosing all information. This is framed as the key to agent delegation: the knowledge substrate defines the AI agent’s path, while the cryptographic keys remain separate.

Dual agent architecture — the human stack

privacymage presented the “human stack” model, mirroring the IP stack: UI and trust tasks at the top, credentials and cryptography at the bottom. The dual agent architecture separates conflicting constraints: one agent handles boundary-making and information protection, the other handles composition and knowledge access. These two agents must never share their full information surface — the security comes from the gap between them.

Emergent properties of the dual system: **connect** (top agent — willing to share knowledge after trust tasks) and **reflect** (bottom agent — adding time dimension through blockchain measurements). Over time, the agents develop inference capability based on the gap between them, creating what privacymage described as “agent communication with polarity.”

The critical principle: the knowledge substrate (mage) should never touch the wallet/keys (ward). The only way to ensure this is mathematical separation into two systems, potentially using trusted execution environments.

Standardization path

An ISO standards expert provided the practical path: BGIN as Class A liaison to ISO TC307 can submit documents for publication. This paper belongs in the TR (Technical Report) category — a survey of the state of the world, requiring only one voting round (~4 months).

The IS (International Standard) path takes 2-4 years through multiple drafts and voting stages.

Critical reality check from a veteran standards practitioner: “If Apple and Google decide to do something, it doesn’t matter what we say — that’s the de facto standard.” The key to successful standards: bring something half-baked from real-world implementation to the standards body, don’t start from scratch in committee. Standards developed by “professional standards personnel who don’t have product responsibility” tend not to get adopted. IEEE 7012 (machine-readable personal privacy terms) was cited as a case where 8 years of standards work is only now reaching adoption — and may gain traction as AI data consumption creates market demand.

Process recommendation from the room: distribute the document to contributors with a deadline for comments, set an agenda for review, and target Block 15 (October, DC) for a fast release version open to public comment.

Overlap with active IKP work

1. Proof of Personhood & Agent Duality — this IS the session

Everything here directly feeds the working group item. The three-graph model (knowledge × promise × trust), the dual credential system (personhood + relationship), the dual agent architecture (mage + ward), and the human stack are all novel contributions that need to be formalized in the document. The biometric protocol’s distinction between proof-of-unique-human and proof-of-specific-human maps to our existing tension between biometric and social proof approaches.

2. Key Management — direct connection

The dual agent architecture’s principle that “the wallet should never touch the knowledge substrate” is the same principle established in the Key Management session: one system holds keys, another composes transactions. The human stack model (UI/trust tasks at top, credentials/crypto at bottom) provides the theoretical framework for the three-tiered key management model (security → signing → power) from the previous session.

3. RPP / Understanding as Key — implementation layer

The knowledge substrate tool and constellation mechanism demonstrated live is the technical implementation of the Relationship Proverb Protocol discussed in the Accountable Wallet session. The constellation (a unique path through a knowledge graph) is the machine-readable version of a proverb compression — both encode proof of understanding.

4. Taxonomy of Harms — identity failure dimension

The discussion on AI replicating observable human behavior raises a harm category: identity spoofing at the personhood level. If an AI can replicate any observable digital behavior, then proof of personhood must rely on what can’t be digitally observed — the gap between the two agents, the biometric substrate, or the bilateral ceremony.

5. PQC connection

The entire credential stack (personhood credentials, relationship credentials, ZK proofs over biometric data) depends on cryptographic primitives vulnerable to quantum attack. The PoP paper must address PQC migration of the credential infrastructure as a requirement, not an afterthought.

Pending actions triggered

- **Document timeline:** Distribute the PoP paper to contributors with comment deadline. Target Block 15 (October, DC) for fast release version open to public comment. Prepare for potential ISO TC307 TR submission after that.
- **Three-graph model formalization:** Write up the knowledge × promise × trust framework as a core theoretical contribution in the paper.
- **Dual agent architecture section:** Expand the human stack / mage-ward separation into a formal section with the connect/reflect emergent properties.
- **Contributing organizations:** Confirm and expand the list. The biometric protocol team confirmed as contributor. First Person Project as infrastructure partner. Seek additional protocols for the comparative evaluation.
- **PQC integration:** Add a section on quantum vulnerability of the credential stack and PQC migration requirements for personhood infrastructure.
- **Knowledge substrate tool:** Continue developing for BGIN use — potential to host BGIN's research corpus as a navigable knowledge graph that contributors traverse and annotate.

Compression seal: 🧬📋🔮 — *The person is proven at the overlap: what you know, what you promise, and the trust that emerges between the two agents that can never fully see each other.*