Elettra Sincrotrone Trieste

Elettra
Sincrotrone
Trieste

Decentralised Persistent Identification: Design Problems and Generalised Specification Proposal

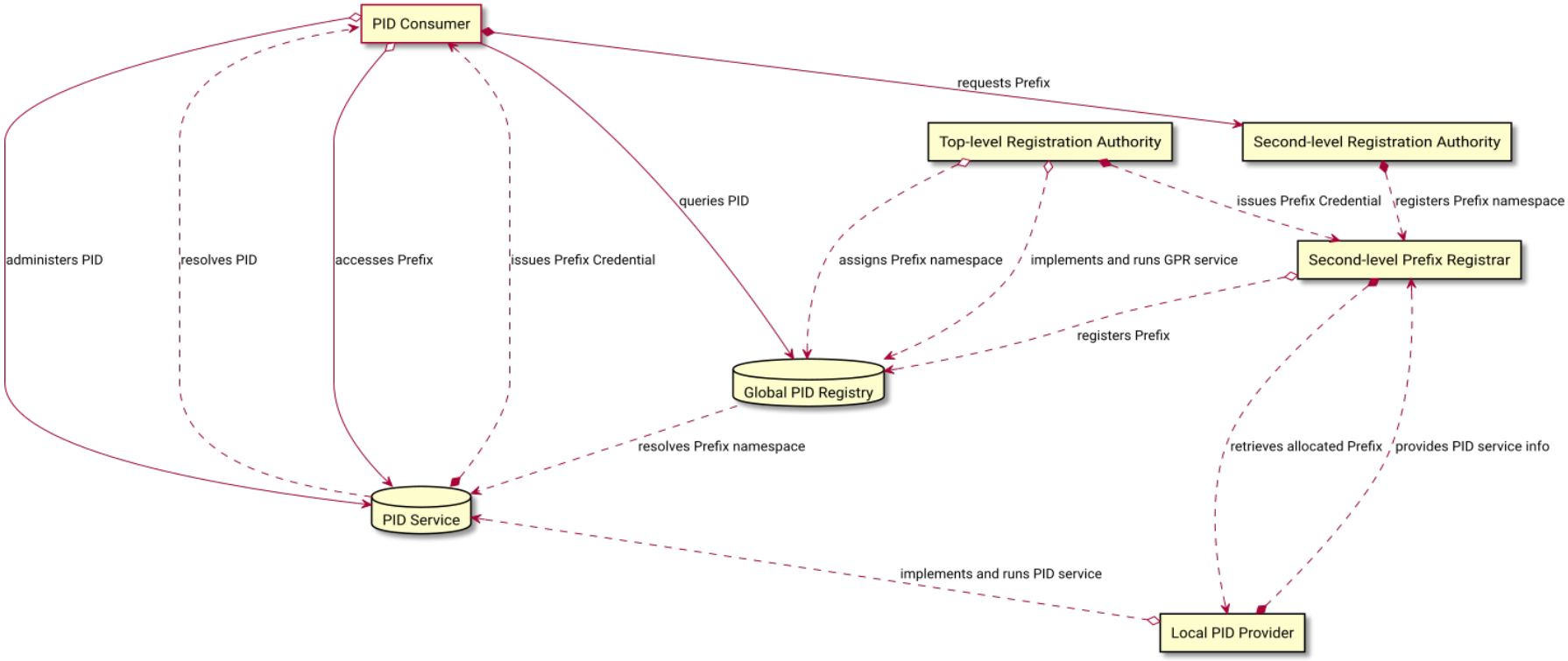
Andrey Vukolov, Elettra Sincrotrone Trieste

Persistent Identifiers: current mainstream implementation and design problems

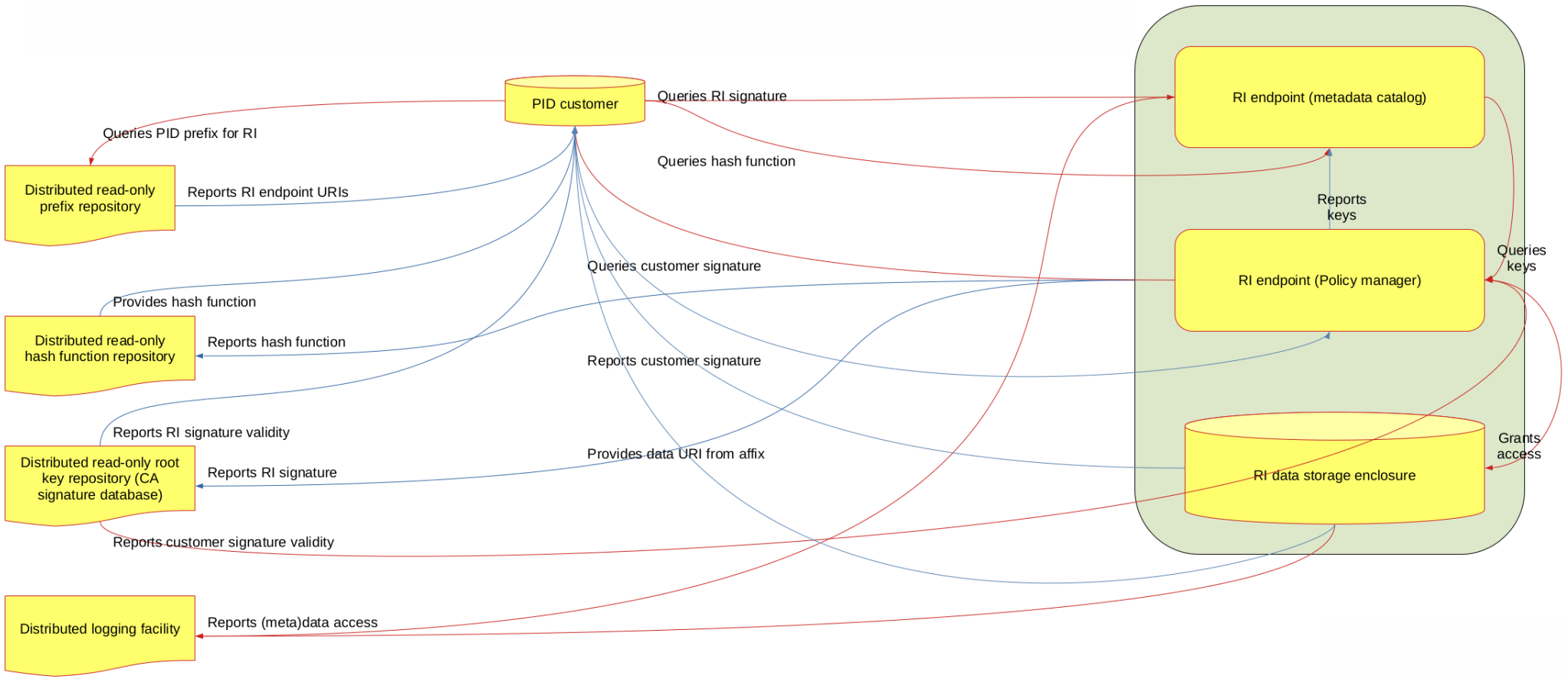
- Persistent identification and binding of the metadata to material and digital entities are now implemented through large-scale centralised providers.
- Data, metadata and PID entities are existence-agnostic about each other within the infrastructure, data linking is not implemented.
- The resolving operation of the given PID is not transparent and the client obtains provenance only from the root trusted authority that is actually self-signed by a provider.
- Resolution is declared as an unidirectional process, so it is not possible for the end user to do a content-based reverse lookup for the given PID on the client's side.

Elettra Sincrotrone Trieste

Persistent Identifiers: current mainstream implementation and design problems



Decentralised PID: a model with integrated provenance



Decentralised PID Infrastructure: expected features and social contracts

- **Decentralization**. The infrastructure should not depend on any kind of central authority or organisation that can mangle, censor or intercept the resolution.
- **Reliability**. The infrastructure must be resilient to failures, attacks, and changes in underlying networks or infrastructures.
- **Transparency**. The infrastructure must be open and public in all aspects, including specifications, working policies and source code so that any node on the network can check the history and current state of PID registration, conversion and resolution.
- **Safety**. The infrastructure should prevent unauthorized or abusive addition, modification, or deletion of PIDs, and provide mechanisms to prove these operations in a convenient and transparent way, if needed.
- **Flexibility**. The infrastructure must allow PIDs within the node-owned subsets to be added, changed or deleted by the agreement of all decentralized network nodes that have the right to publish them, without the need to change code or system parameters.
- **FAIR-compliance**. The infrastructure should implement all the described features with a maximal level of compliance with FAIR Data Management Principles, including all kinds of underlying access control and provenance

Elettra Sincrotrone Trieste

Decentralised PID Infrastructure: Conceptual Specification Proposal

- The proposed decentralised PID infrastructure is built from the set of **peers** initially connected over **IPFS** network.
- Every peer may work as an **untrusted resolver node** performing global resolution and reverse lookup for every valid PID.
- The PID consists of **superset** (prefix) ID, **subset** ID and **record** ID (ideally hashed) with optional version affix:
  `<superset>/<subset>~<record>[_<affix>]`
- The **supersets** are managed by the finite set of **trusted peers** allowing the essential operations via all-agree **ratification** mechanism.
- The trusted peers yield the **subset** management to dedicated peers which take responsibility for publishing and verification of the subset record resolution tables over IPFS network.

Decentralised PID Infrastructure: Conceptual Specification - Peers

- **Trusted Peer**. The trusted peers have a full right to assign, manage, revoke and delete discrete PIDs and their sub- and supersets available within the infrastructure. Every trusted peer signs any operation request within the infrastructure with his unique private key.
- **Publisher Peer**. The untrusted peers in the decentralized network that hold the right to publish PIDs under the dedicated subset. The responsible Trusted peer has a right to mark individually the published PIDs as Verified, or Suspicious, or to leave them in a Normal state by default.
- **Untrusted Peer**. The untrusted peers in the decentralized network that have no rights to publish PIDs, subsets or supersets, but hold a right to **resolve** and **reproduce** the existing PIDs and **retrieve** the underlying data/metadata, also to **provide** a public resolution gateway.

Decentralised PID Infrastructure: Integrity Maintenance

- The ratification mechanism requires **identical resolution table IDs** published over the peers' IPFS **entry points**.
- Trusted peers can ratify declaration of the given PID as **verified** or **suspicious** (also the weighted mechanism is possible).
- Every peer inside the infrastructure shares the **same software**, and the software is also identified with a dedicated PID available for the content-dependent reverse lookup.
- Every ratification operation within the infrastructure should be **signed** by the cryptographic key owned by every trusted peer.

Decentralised PID Infrastructure: IPFS Entry Points per Peer

Name	Published by			Description
	Trusted	Publisher	Untrusted	
TRUSTED_PEERS	+	-	-	Trusted Peers List
PENDING_OPERATIONS	+	-	-	Pending Operations List
SUPERSET	+	-	-	Superset List
SUBSET	-	+	-	List of PIDs issued under the given Subset
VERIFIED	+	+	-	List of Verified PIDs declared by the given Trusted Peer in the given Subset
SUSPICIOUS	+	+	-	List of PIDs declared Suspicious by the given Trusted Peer in the given Subset
BROADCAST	+	+	+	Plain-text mutable broadcast message
PEER_INFO	+	+	+	Public peer information and ID

Decentralised PID Infrastructure: Network Status Consensus Maintenance

- Trusted peers are obliged to maintain the ratification status and identicity of the ID of the **Trusted Peer List**, **Superset List**, and **Subset Publishers List** published on their entry points.
- Publisher peers are obliged to check the integrity of the **Trusted Peer List** and **Subset Publishers List** for its owned Subset.
- Subsets are built over the IPNS public keys owned by the given **Publisher Peer**, and every published state of the PID subset is signed by his Private key.
- The Untrusted peer must verify the ratification status of **Trusted Peer List**, **Superset List**, and **Subset Publishers List** every time when initialized.
- Every peer has a right to provide a public resolution gateway for any user outside the network. During the resolution he is obliged to check both digital signatures from the **Publisher Peer** who published the **Subset**, and all **Trusted Peers** ratified this Publisher Peer to enter the list.

Decentralised PID Infrastructure: Roadmap and Discussion Points

- Possibilities to integrate with dPID.
- Discussion about minimizing social contracts within the specifications, with a possibility to provide subsets for the societies requiring web-of-trust principle.
- Disseminate the specification around the interested research institutions summoning the core maintainers to start the decentralised project.
- Systematize and group the operations and corresponding tasks for the Peers.
- Explore the possibilities of reverse compatibility and cross-resolution for the traditional PIDs like DOI.

Elettra
Sincrotrone
Trieste

Thank you!

Contact email: andrey.vukolov@elettra.eu

Elettra
Sincrotrone
Trieste

www.elettra.eu