What is dARK?

dARK is the first implementation of a technology that could be the base for a **low-cost and decentralized services to assign/resolve persistent identifiers** (ARK PID for this first version).

dARK is an ARK implementation based on **institutional blockchain nodes**, the **data is owned, stored and controlled no by one single organization** but by **all participants in a "public good" network**.

We see this as the foundation for an **open and community driven project** fostered by **Instituto Brasileiro de Informação em Ciência e Tecnologia** (IBICT-Brazil) and **LA Referencia/RedCLARA** (supported by SCOSS) with the long-term objective of provide an **open non-centralized deduplicated persistent identifier factory and resolution service model for the global Open Science ecosystem** based on permissioned blockchain technology.

Why are we developing dARK?

- *The **need** not only of **persistent but unique/deduplicated** identifiers to build better research graphs, indicators and research assessment tools.*

- ***Lack of PID coverage** on **Global South** repositories/journals mainly because of the **costs** that those **services/fees** represents for the institutions.*

- *Most persistent identifier systems are **based on centralized models**, depending on **few agencies** that support the service infrastructure.*

What is ARK PID?

Archival Resource Keys (ARKs) serve as *persistent identifiers*, or stable, trusted references for information objects. Among other things, they aim to be web addresses (URLs) that don't return 404 Page Not Found errors. The ARK Alliance is an open global community supporting the ARK infrastructure on behalf of research and scholarship.

End users, especially researchers, rely on ARKs for long term access to the global scientific and cultural record. **Since 2001 some 8.2 billion ARKs** have been created by over [1000 organizations](#) — libraries, data centers, archives, museums, publishers, government agencies, and vendors.

Source: https://arks.org

Why decentralized?

Increased autonomy and flexibility, reduces costs: Decentralized management of PIDs allows for greater autonomy and flexibility among different stakeholders.

Reducing single points of failure: By distributing the responsibility of managing PIDs across multiple entities, a decentralized approach helps mitigate the risk of a single point of failure.

Promoting innovation and diversity: Decentralization fosters an environment that encourages innovation and diversity in the development and implementation of PID systems.

Scalability and adaptability: Decentralized systems can be more scalable and adaptable to changing circumstances

Trust and transparency: A decentralized approach promotes trust and transparency by distributing authority and decision-making power among multiple entities.

Why Blockchain?

Blockchain can be seen as a "public" database where all authorized participants (institutions/organizations) can store and query data (securely and reliably)
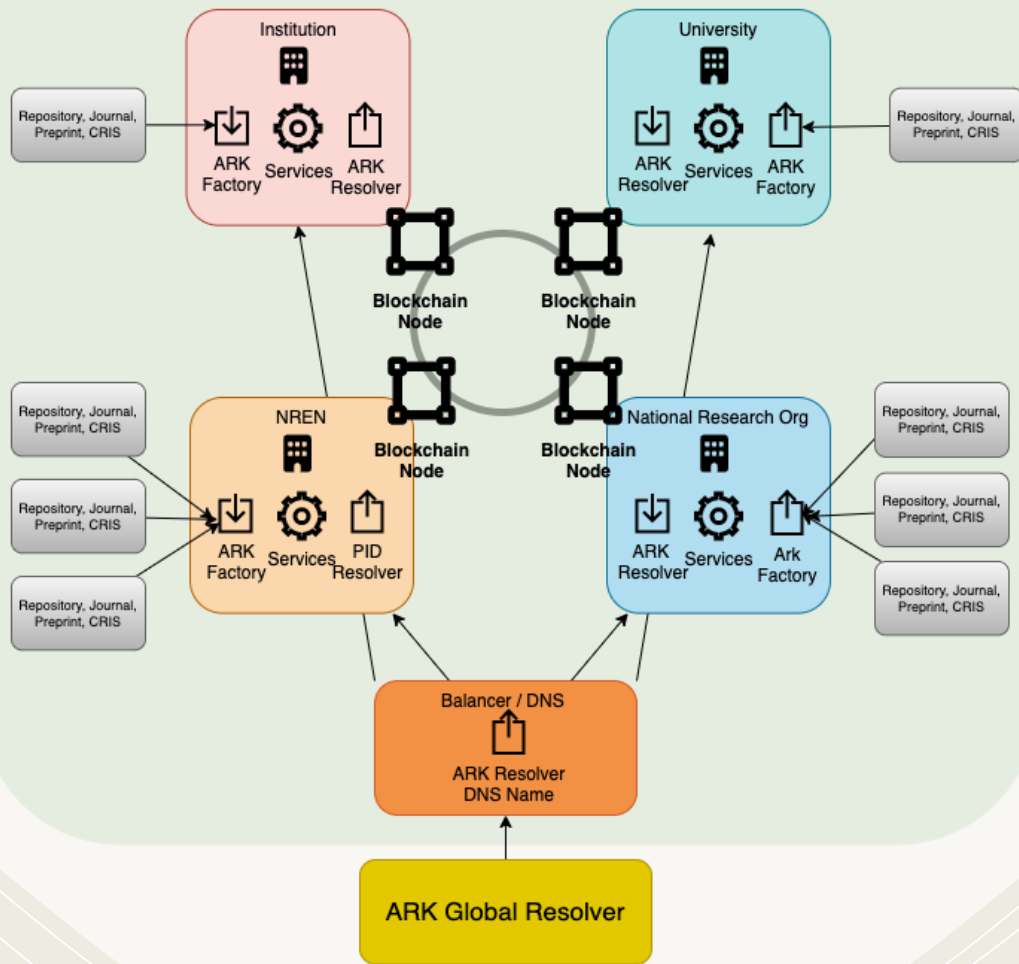
Blockchain technology provides **security, transparency, and interoperability,** the native decentralized nature mitigates security risks by **eliminating single points of failure**, ensures **data preservation, data immutability safeguards the integrity and authenticity**.

The **transparent and auditable** nature of Blockchain enables the tracking and verification of all activities, **enhancing transparency and accountability,** adds a new **layer of trust to the PID ecosystem** by allowing the community to check all the transaction history.

**We use "Hyperleder besu" blockchain** https://www.hyperledger.org/use/besu

Energy consumption and environmental impact are important aspects to us. In this context, we design our implementation considering **Proof of authority (PoA) consensus mechanism**, which is both environmentally friendly and **operates without the high CPU usage** associated with other consensus algorithms.

The dARK Persistent ID Network (first goal)

The dARK Ecosystem (long term goal)
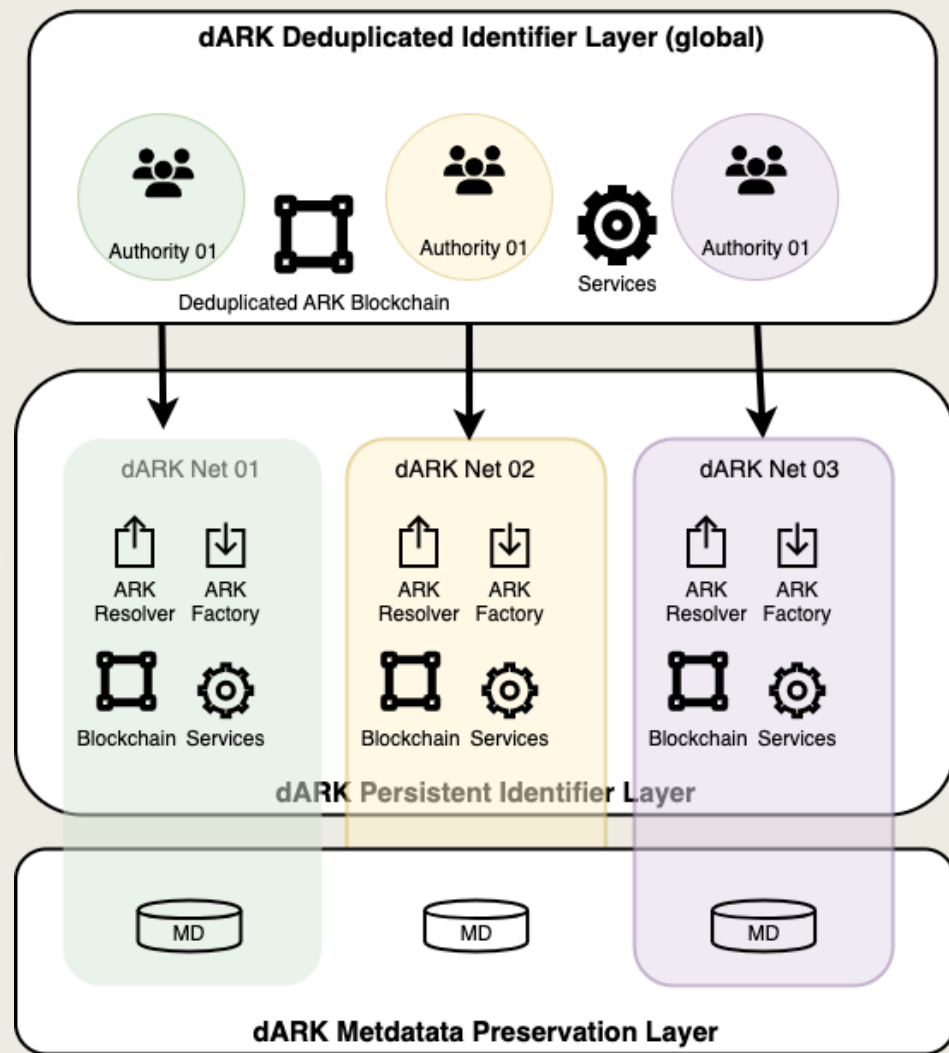
2022–2023 ROADMAP – Work in progress

Phase 1: Concept Paper Development: Develop a comprehensive concept paper outlining the objectives, scope, and technical details of the dArk project. **DONE!** https://doi.org/10.5281/zenodo.7686101

Phase 2: Proof of Concept (PoC) Development: Validate the technical feasibility of the dARK project by implementing a simplified version of smart contracts and demonstrating key features in a small-scale network. **DONE!**

Phase 3: Integration with External Systems: Enable cross-referencing and resolution of identifiers between dARK and external systems to ensure interoperability. **WORKING**

Phase 4: Pilot Implementation in Brazil: Set up and configure the infrastructure by installing Hyperledger Besu nodes compatible with the target network configuration for a pilot implementation in Brazil. **WORKING**

2024-2025 ROADMAP – Next steps

Phase 5: Testing, improvement, and open-source release: Conduct comprehensive testing, including unit, integration, and stress testing, to improve the system's stability and release a first stable open-source version.

Phase 6: Documentation and Training: Create comprehensive documentation, including installation guides and smart contract APIs, and conduct training workshops to educate stakeholders on dARK identifier usage and benefits.

Phase 7: Governance and Upgrades: Establish a governance model for decision-making, consensus upgrades, and dispute resolution, and plan for regular enhancements based on feedback and evolving requirements.

Phase 8: Community Building and Adoption: Foster a community around the dARK project by engaging with researchers, institutions, and developers through community events, forums, and communication channels for collaboration and knowledge sharing.

TEAM

Lautaro Matas, LA Referencia lautaro.matas@lareferencia.redclara.net

Washington Segundo, IBICT, BRAZIL washingtonsegundo@ibict.br

Thiago Nóbrega, IBICT, BRAZIL thiagonobrega@ibict.br

J. Edilson S. Filho, IBICT, BRAZIL edilsonfilho@ibict.br

Jesús Mena-Chalco, UFABC, BRAZIL jesus.mena@ufabc.edu.br

More LA Referencias's dARK team members are coming in second half of 2023 thanks to SCOSS pledges

The Global Sustainability Coalition for Open Science Services (SCOSS)

SCOSS

LA Referencia
Red de repositorios de acceso abierto a la ciencia

LA Referencia
THE FEDERATED NETWORK OF LATIN AMERICAN OPEN SCIENCE REPOSITORIES

LA Referencia provides some essential elements for the development of the international Open Science ecosystem.

It was founded 10 years ago and brings together twelve countries along with RedCLARA, our regional research and education network. It provides a tool that facilitates the international harvesting, validation, and enrichment of metadata from repositories. It offers value-added services and responds to the interoperability challenges of the new OpenAIRE guidelines. This tool is increasingly being adopted outside of Latin America (Spain, Portugal, Africa). It addresses the growing need for integration with research management systems (CRIS) and aims to produce statistics to serve national decision-makers when evaluating research.

By supporting LA Referencia, you will help us develop new services such as a decentralised, sustainable and global persistent identifier service based on blockchain and extend the current usage statistics service to other infrastructures/regions contributing to a more equitable international publishing system.

WHY HAS IT BEEN DEEMED AN ESSENTIAL INFRASTRUCTURE?

LA Referencia enables open science as a public good through its federated network of open access institutional repositories in Latin America. It is governed and funded by Latin American governments and supports the implementation of Open Science policies. SCOSS funding will help develop two innovative services around sustainable persistent identifiers and usage statistics, promoting bibliodiversity, and the development of new indicators based on open science. LA Referencia is aligned with the SCOSS theme FAIR Open Access repository services.

"It's not **dARK** yet, but it's gettin' there"

Bob Dylan

THANK YOU – GRACIAS – OBRIGADO

Lautaro Matas

lautaro.matas@lareferencia.redclara.net

Washington Segundo

washingtonsegundo@ibict.br

Frequent Asked Questions

Why we use ARK instead other PID?

- ARKs are cheaper, more flexible, and less centralized, letting you **create unlimited identifiers without paying for the right to do so**,

- add **any kind of metadata**, including no metadata,

- append extensions and **query strings during resolution**,

- **link directly** to an article, image, or spreadsheet that is **immediately usable by people and software without making them first stop at a landing page**,

- make millions of ARKs resolvable by managing just one ARK, via a mechanism called **suffix passthrough**.

Source: https://arks.org

Could this very same technology be used for DOI/Other?

Yes! We want to work on a "PID agnostic" distribution of this solution soon for supporting other PIDs.

For the moment other PIDs can be appended to the dARK record (as alternatives) and will be supported by the resolution service.