LifeBlock for FAIR data citation

Joaquín López Lérida
Data e-Science Management Plans and Blockchain Officer,
LifeWatch ERIC

https://orcid.org/0000-0001-9697-7710

Mastering Data Citation: Insights from the BioDT Research Infrastructures
Online, 20, Nov 2023

Funded by
the European Union

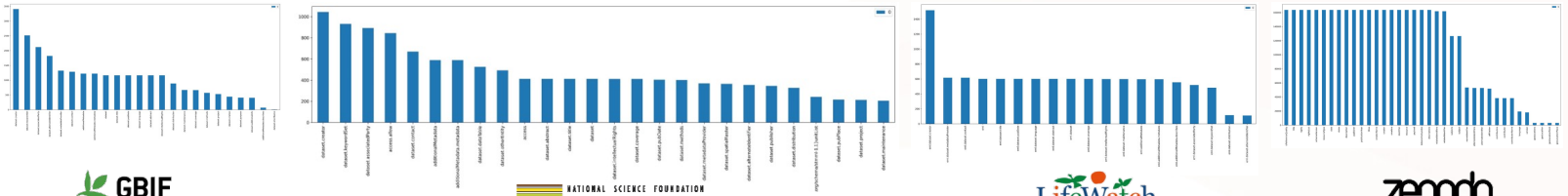
Starting point for LifeWatch ERIC for FAIR data citation

- We needed to be able to support the most common biology metadata formats. These must include EML and DataCite in order to cater for the needs of non-biological metadata if needed and in the case a provider doesn't support EML.

Metadata Format	GBIF	LTER	ZENODO	METADATA CATALOG
EML	X	X		X
DataCite	X		X	X

- A statistical analysis of the prevalence of the different fields across some RI files was done for metadata to have a clear idea about design needs.

Frequency of EML terms one level deep into the Metadata XML for some RIs.



GBIF

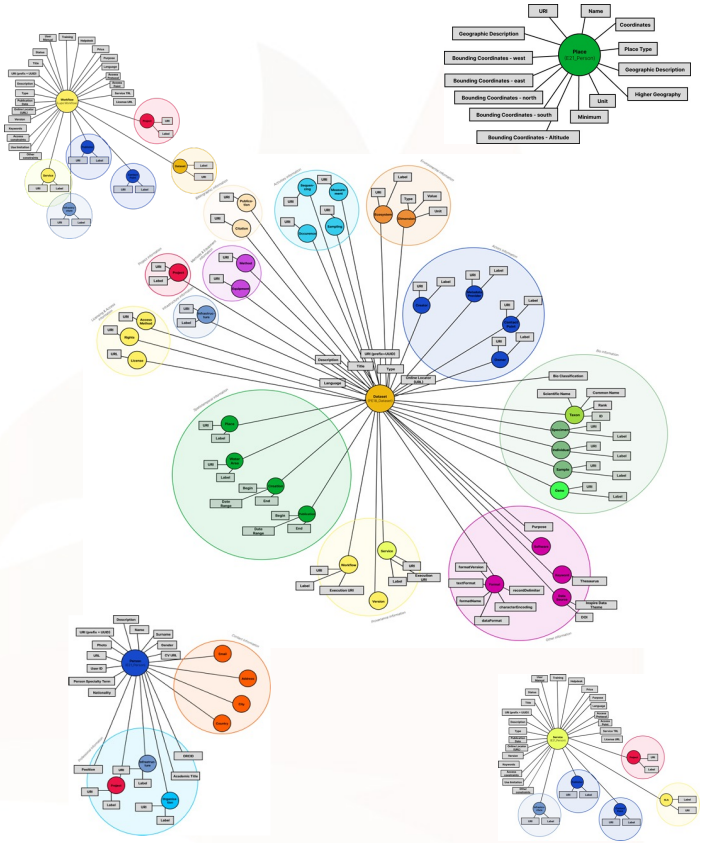
NATIONAL SCIENCE FOUNDATION LTER NETWORK LONG TERM ECOLOGICAL RESEARCH

LifeWatch ERIC

zenodo

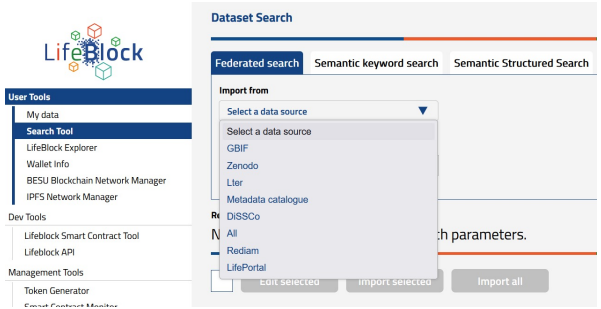
Our challenges for LifeBlock metadata schema

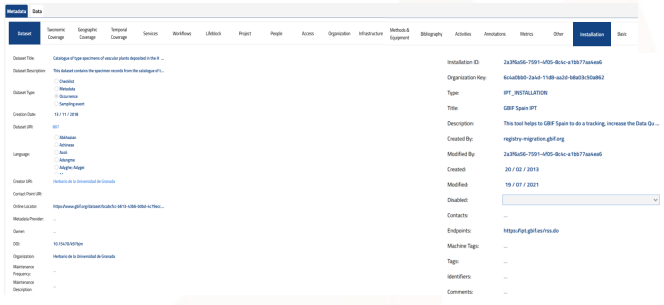
- **RIs Diversity:** Each organization has its own language and structure; we had to translate and unifies these languages.
- **Multilingualism and Formats:** Multiple languages and formats, with linguistic and technical barriers.
- **Text to Knowledge:** Our goal is to convert unstructured information into valuable knowledge, ready to be explored and applied, particularly to our VREs.
- **Interoperability:** We need that VREs and RIs from different sources and formats can "talk" to each other seamlessly. Definition of a user data space where we can unify the different sources of information.
- **Queries:** Need of intuitive interface for querying and exploring RIs, both in federated syntactic search notation and in semantic and artificial intelligence search notation.

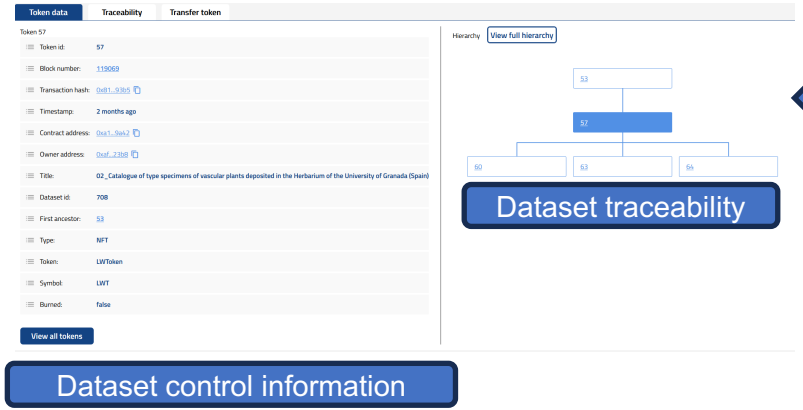


Federated and semantic Search over main RI´s

Common metadata schema for RDF transformation





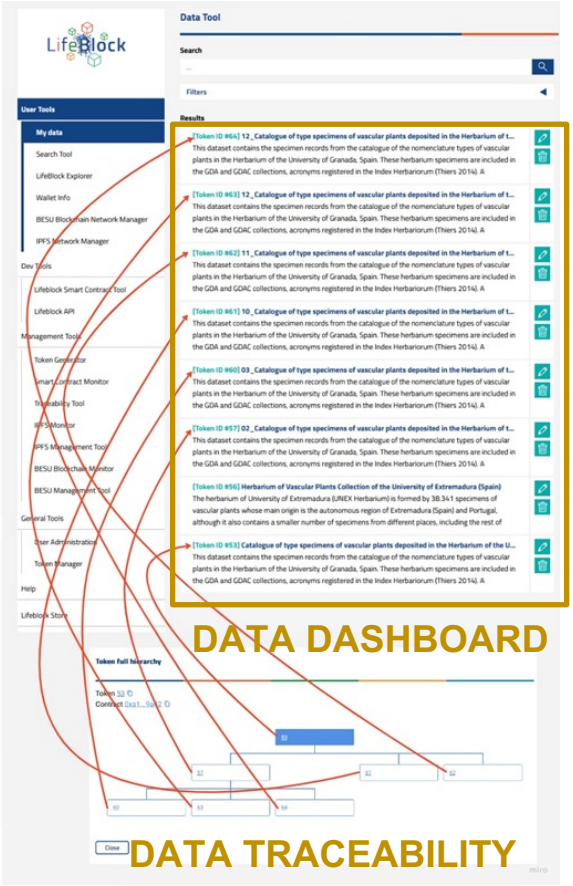


Dataset traceability

Dataset control information

Usage of NFT (**ERC-721**) to storage metadata Information as unique and trazable information container with **IPFS** as storage resource for datasets. Each token is a dataset that has an unique number and an unique owner.

- **F**indability - Provide **user-friendly** search functions, intuitive for any user and above all, easy to use.
- **A**ccessibility - Provide accessible environments for users and **easy to understand** and use, even for semantic search.
- **I**nteroperability - Provide data beyond the resources generated by LifeWatch ERIC by **extending the databases** accessible from the tool to other RIs, DB, etc.
- **R**eusability – Provide data from LifeBlock that can be used in any other environment, workflow, SKG, VRE, ..., always preserving their **traceability** in order to be guaranteed with successive reproductions.



Advantages of using LifeBlock for FAIR data citation

- **Recognition and Attribution of Data**: LifeBlock´s transparency and ERC-721's unique identification capabilities ensure precise tracking and attribution of each dataset, enhancing recognition of data creators and promoting open data sharing.
- **Standard Practices in Data Citation**: ERC-721 tokens enable clear tagging and citation of datasets, creating a standardized and respectful data citation practice, ensuring proper acknowledgment of creators and sources.
- **Improvement in Data Discovery and Reusability**: Lifeblock's immutable record, combined with ERC-721's detailed metadata, improves the findability and reusability of datasets, accelerating biodiversity research.
- **Credit to Data Managers**: LifeBlock highlights the contributions of data curators and managers, transparently demonstrating their vital role in maintaining and updating datasets.
- **Collaboration with Research Infrastructures**: Blockchain integration with research infrastructures like GBIF, Zenodo, LTER or Rediam and LifeWatch fosters seamless data sharing and enhances data citation practices through collaborative expertise.
- **Essential Techniques for Data Citation**: The usage of ERC-721 in data citation introduces key techniques for precise and efficient data referencing, offering a solid framework for recognizing contributions in biodiversity research.