



The LifeWatch ERIC Biodiversity
& Ecosystem eScience Conference

PLOVDIV, BULGARIA

7-10 JULY 2026

Scientific Knowledge Graphs for Biodiversity Research

*From machine-actionable DMPs to connected, interoperable knowledge — the LifeWatch
ERIC × OpenAIRE case*

Presenter: Joaquín López Lerida - ICTCore

July 2026



1

What is an SKG?

Graphs, meaning, and why science needs them

2

How SKGs are built

Nodes, edges, identifiers, and shared vocabularies

3

maDMP as a source of nodes

Turning a data management plan into graph material

4

A worked example

The real LifeWatch ERIC maDMP, rendered as a graph

5

Interoperability

Connecting LifeBlock to the OpenAIRE Graph



PART 1

What is a Scientific Knowledge Graph?

Starting from zero — graphs, meaning, and why modern science needs them.

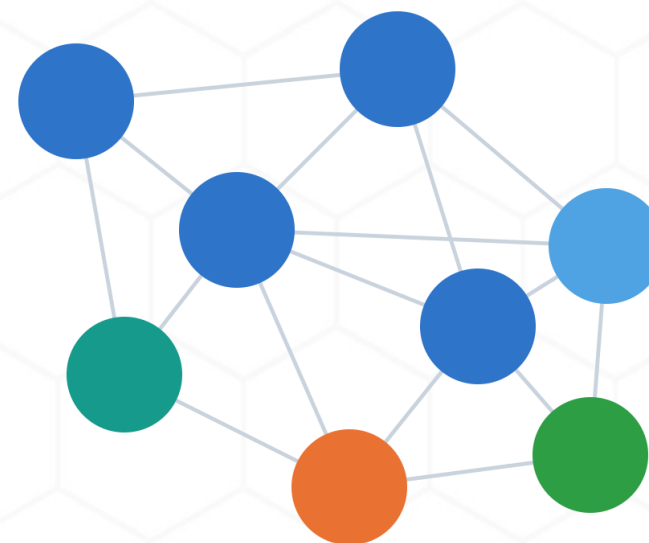
The problem: research is a pile of disconnected outputs

Today: scattered outputs



no links — discovery & reuse are manual

With an SKG: connected

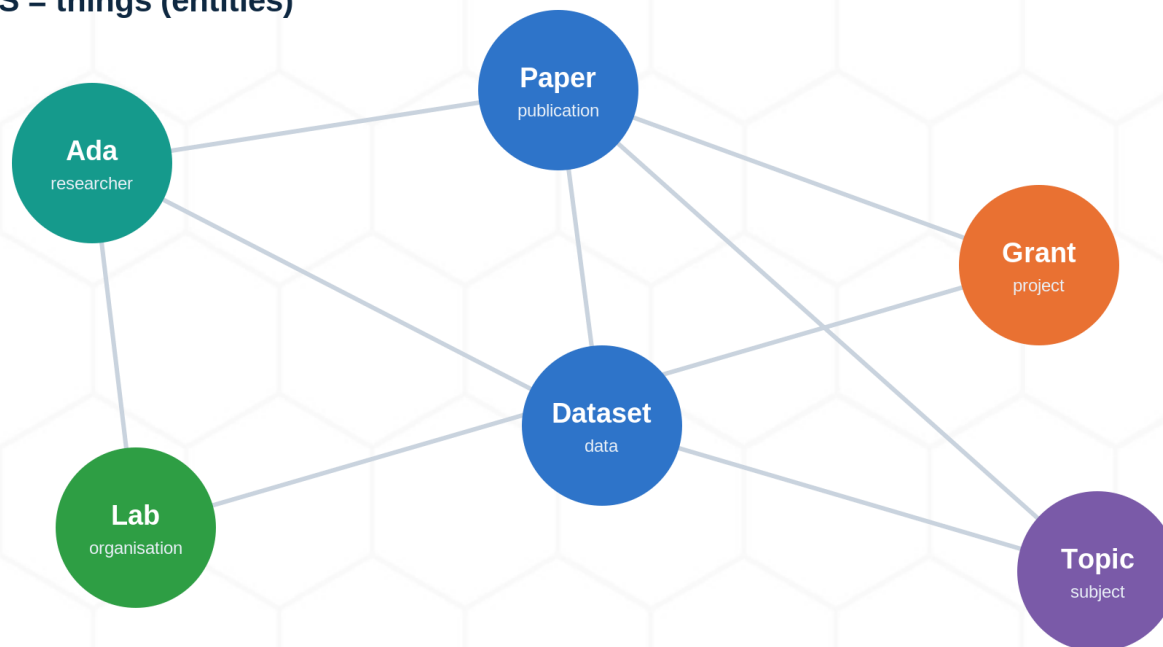


traversable — automatic discovery, provenance & impact

A paper here, a dataset there, the software somewhere else, the people and the funding recorded in yet another system.
Nothing knows about anything else.

First — what is a “graph”?

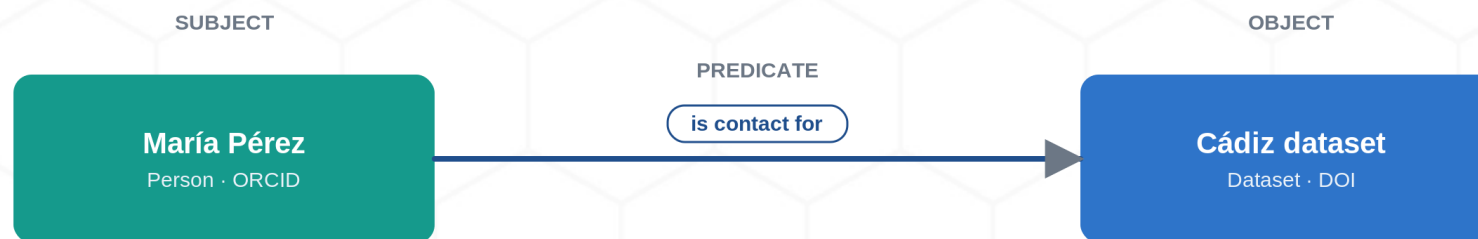
● NODES = things (entities)



— EDGES = relationships between them

A graph is just things (nodes) joined by relationships (edges). Nothing more mysterious than that.

Adding meaning: from graph to knowledge graph



Every fact is a triple: subject – relationship → object

Millions of triples that share the same identifiers weave into one graph.

A knowledge graph gives every node and every edge an explicit, machine-readable meaning.

So, what is a Scientific Knowledge Graph?

A Scientific Knowledge Graph (SKG) is a database of research information — products, processes and the actors behind them — expressed as connected, machine-readable entities and relationships.



Research products

Publications, datasets, software, workflows, protocols, and — yes — DMPs themselves.



Actors & agents

The people and organisations who create, fund, host and steward those products.



At scale

The graphs links relationships across hundreds or millions of research objects.



Findability & discovery

Ask questions no single database can — “show every dataset from this grant, and who reused it.”



Provenance & trust

Follow the chain from result back to raw data, instrument, method and person. Reproducibility becomes traceable.



Automation

Machines act on the links: auto-fill reports, trigger FAIR assessments, validate against funder rules.



Impact & reporting

Roll up outputs, citations and reuse to people, projects and infrastructures — evidence for policy and funders.

The FAIR principles — Findable, Accessible, Interoperable, Reusable — are what an SKG operationalises at scale.



PART 2

How SKGs are built

The building blocks: nodes, edges, identifiers, and shared vocabularies.



Nodes (entities)

The “things”:

- a dataset
- a person
- an organisation
- a project, a topic

Each has a type and an identifier.



Edges (relationships)

The typed links between them:

- authored by
- funded by
- stored in
- is about

Direction and meaning matter.



Properties

Facts attached to a node or edge:

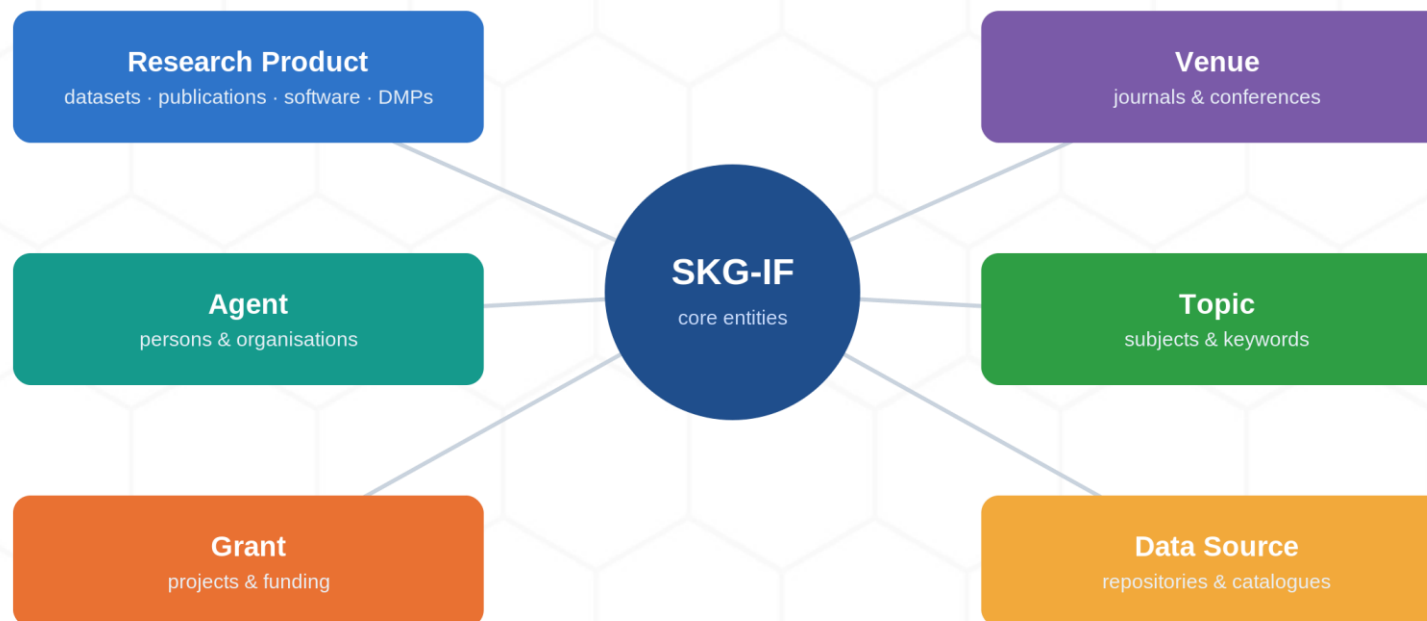
- title, date, language
- identifier (DOI/ORCID)
- licence, size, format

The metadata payload.

Get these three right and consistent, and a graph practically assembles itself.

A shared model: the SKG-IF core entities

A small shared vocabulary of entity types — the common language of scholarly graphs



SKG-IF reduces the whole scholarly world to a handful of entity types — a common language every graph can speak.

The glue: Persistent Identifiers (PIDs)



DOI

Digital Object Identifier — datasets, publications, software.



ORCID

A permanent ID for a researcher. Distinguishes people who share a name.



ROR

Research Organization Registry — a stable ID for institutions.



RAiD

Research Activity ID — a persistent ID for a project or activity.

Why PIDs are everything: a PID is a globally unique, permanent name. When two systems mention the same DOI or ORCID, they are provably talking about the same thing — so the graph can merge them into one node instead of two look-alikes.



Darwin Core

The biodiversity metadata standard — species occurrence, taxonomy, location. Speaks the language of our domain.



ENVO

The Environment Ontology — controlled terms for habitats and environmental features, so “coastal zone” means one thing.



schema.org / DataCite

General-purpose vocabularies for describing resources and citation metadata across the whole web of science.

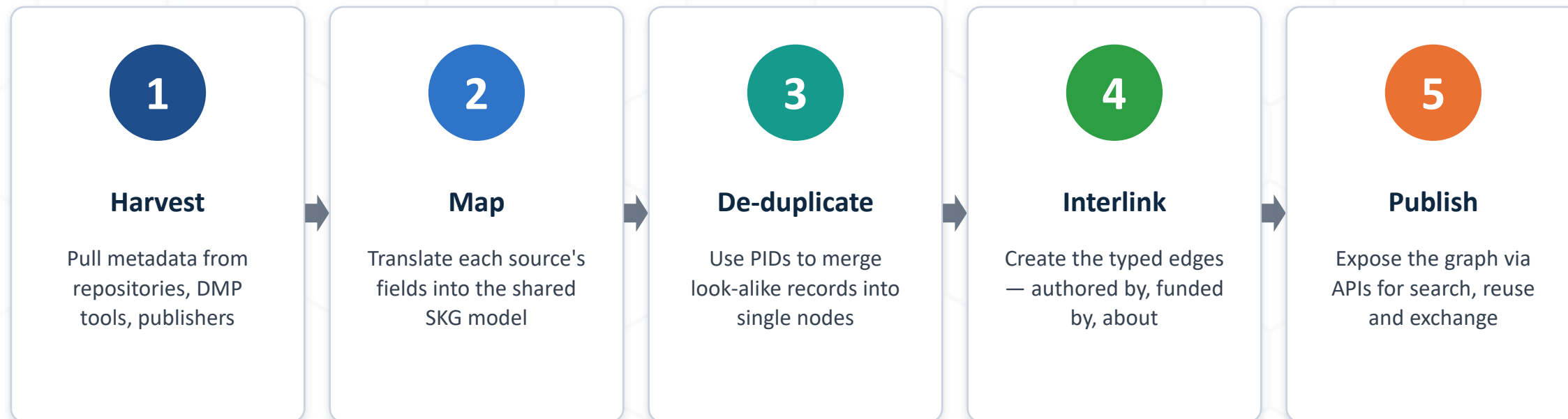


Why it matters

A shared vocabulary means “subject” or “habitat” is understood identically by every tool — no translation guesswork.

Identifiers say which thing; vocabularies say what kind of thing and what its fields mean.

How a graph gets populated



Steps 2 and 3 are where a good shared model and clean PIDs pay off — they make mapping and de-duplication almost mechanical.



PART 3

maDMP as a source of nodes

Your data management plan is not paperwork — it is graph material.

A DMP describes how a project's data will be:

- collected and generated
- organised, documented and stored
- shared, licensed and preserved
- who is responsible for each step

Required by funders. Horizon Europe, the ERC and most national agencies mandate a DMP for every funded project.



The catch

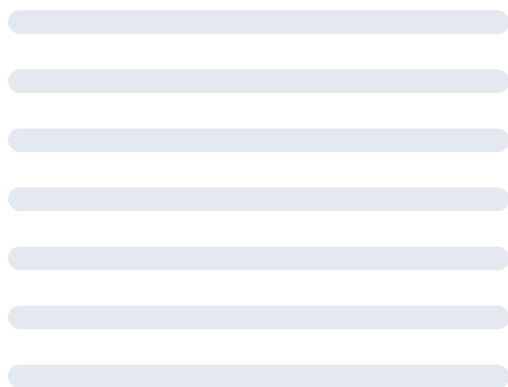
A traditional DMP is a Word or PDF document.

It is written for humans, goes stale the moment it is filed, and cannot be read by any machine.

All that structured knowledge is trapped in prose.

From DMP to machine-actionable DMP (maDMP)

Traditional DMP



Word / PDF · for humans · static



maDMP (JSON · RDA DCS)

```
{ "dmp": {  
  "title": "MARBEFES DMP",  
  "project": [ ... ],  
  "contact": { "orcid": ... },  
  "dataset": [ {  
    "dataset_id": "doi:...",  
    "distribution": [...]  
  } ]  
}}
```

structured · for machines · linkable by PID

Same information — but expressed as structured JSON that machines can read, validate, link and harvest.

The maDMP standard: RDA DCS + OSTRails profile



RDA DMP Common Standard

The international backbone. Defines the core objects: dmp, project, contact, contributor, dataset, distribution, host, licence, cost.



OSTRails Application Profile

Extends the standard via an “extension” block for what pilots need: spatiotemporal coverage, methods, quality, semantics, provenance.

Every one of those objects is a node-in-waiting:

project → Grant **contact / contributor** → Agent (Person) **institution** → Agent (Organisation)

dataset → Research Product **distribution.host** → Data Source **keyword** → Topic

Use Case 1: create a maDMP with ARGOS

1

Map the template

Map every field of your DMP template to the RDA DCS + OSTRails AP

2

Load into ARGOS

Deploy the machine-actionable template in ARGOS, OpenAIRE's DMP platform

3

Connect services

Wire in local systems — repository, PID providers, CRIS

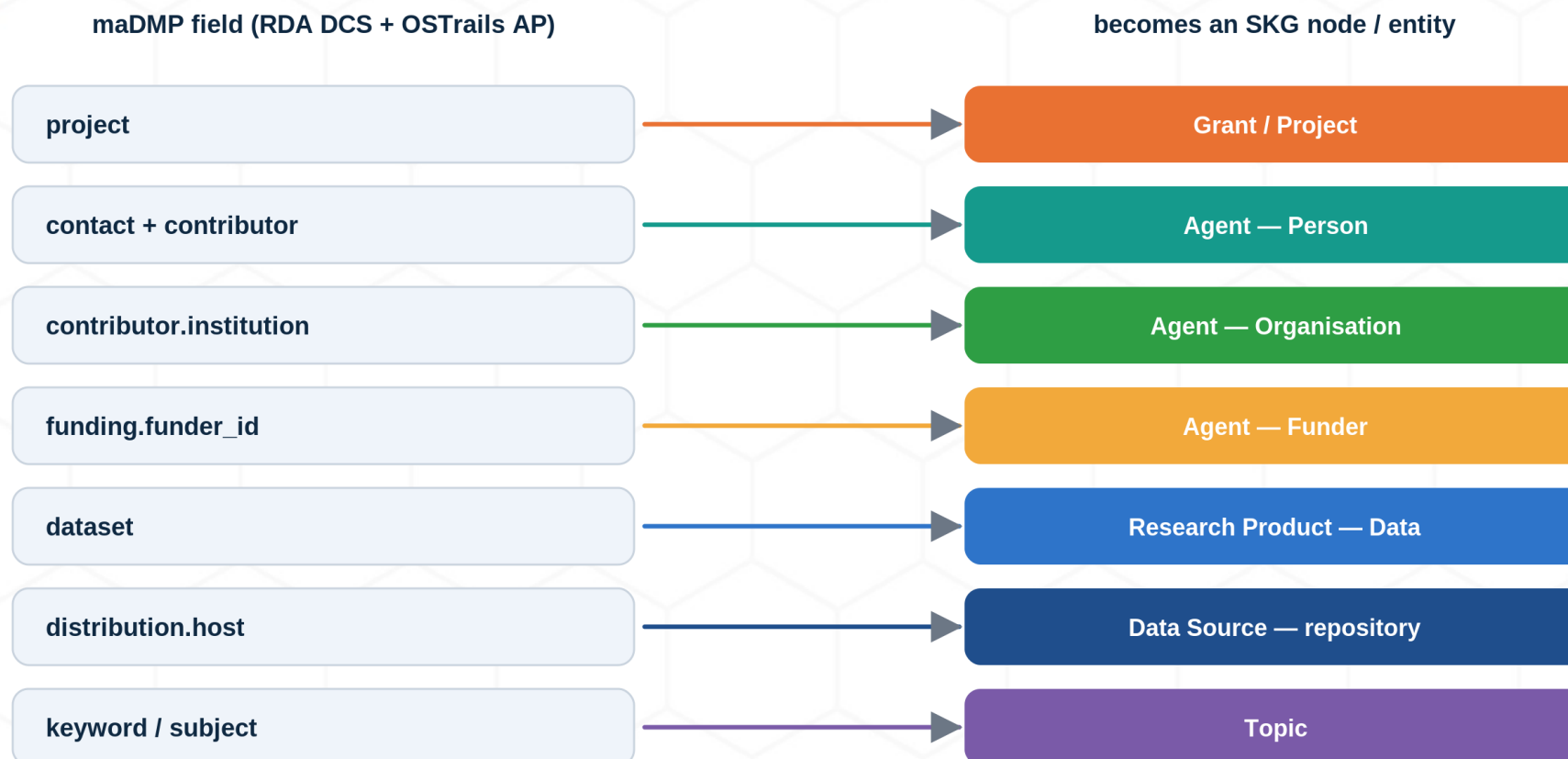
4

Export & harvest

ARGOS emits standard maDMP JSON, ready to feed the graph

This is exactly what LifeWatch ERIC did in OSTRails Thematic Pilot 7, using ARGOS by OpenAIRE.

Composing nodes: every field is a ready-made node





PART 4

A worked example

The real LifeWatch ERIC maDMP — turned into an actual sub-graph.

● Project	MARBEFES — coastal biodiversity monitoring (Horizon Europe)
● Dataset	Coastal imagery set — Cádiz 2025 (UAV RGB + segmentation)
● Contact	María Pérez · ORCID 0000-0002-1825-0097
● Data manager	Joaquín López · ORCID · unit: e-Biodiversity
● Organisation	LifeWatch ERIC · ROR https://ror.org/01bj3aw27
● Repository / host	LifeBlock — biodiversity SKG & repository
● Licence	Creative Commons Attribution 4.0 (CC-BY 4.0)
● Semantics	Darwin Core (subjects) · ENVO (habitats)

Step 1 — extract the nodes

 maDMP MARBEFES Funder (fundref) María Pérez Joaquín López LifeWatch ERIC Cádiz imagery LifeBlock My Lifewatch workflow Jupyter service Darwin Core / ENVO CC-BY 4.0

Twelve nodes, pulled straight from one JSON file — no manual data entry.

Step 2 — connect them with typed edges

maDMP


describes

MARBEFES

MARBEFES


is funded by

Funder

maDMP


has contact

María Pérez

MARBEFES


has data manager

José López

José López


is affiliated with

LifeWatch ERIC

maDMP


plans

Cádiz imagery

Cádiz imagery


is stored in

LifeBlock

Cádiz imagery


is about

Darwin Core / ENVO

Cádiz imagery


is licensed under

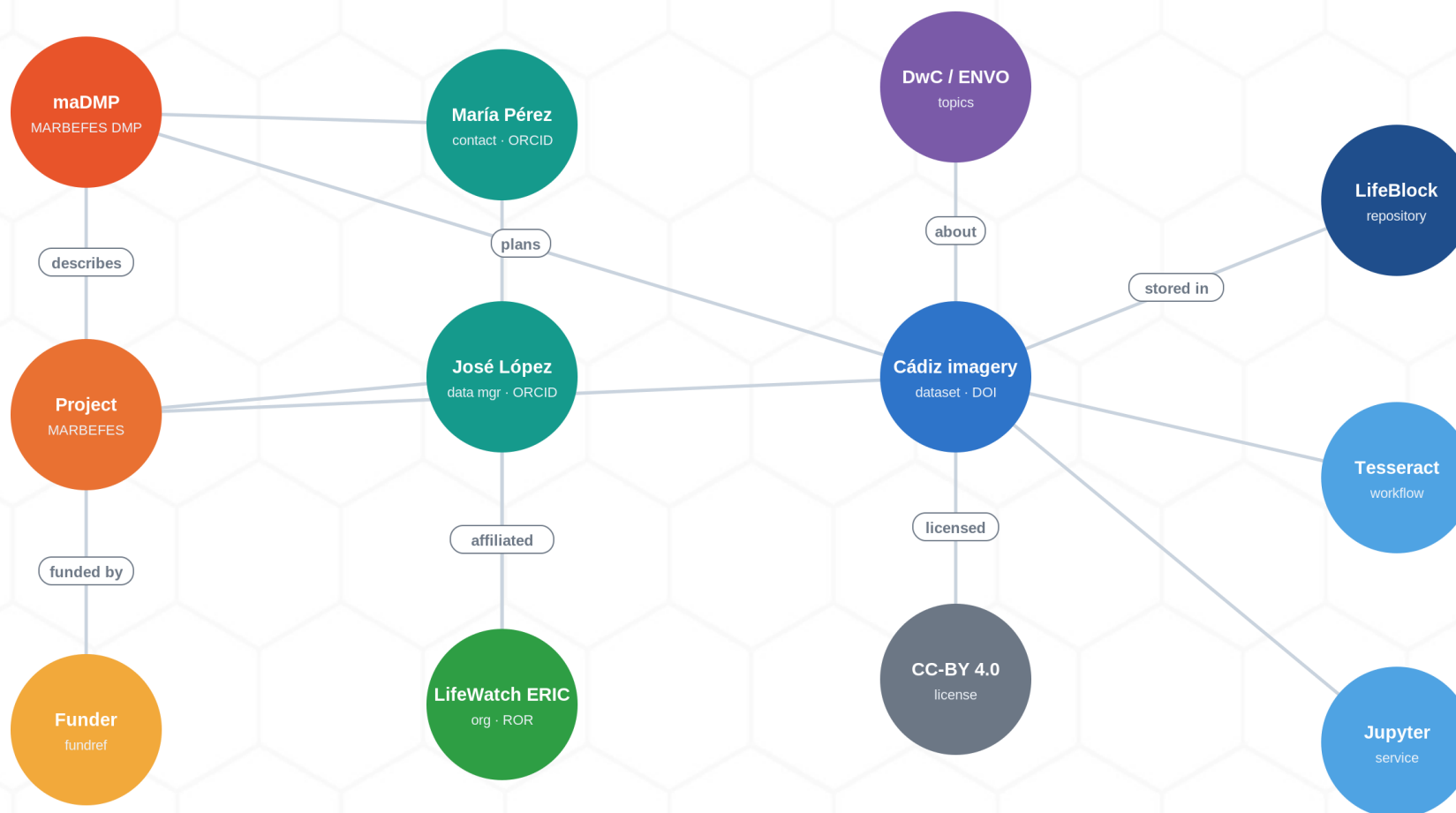
CC-BY 4.0

Cádiz imagery


was produced by

Tesseract

The result: a LifeBlock sub-graph from one maDMP





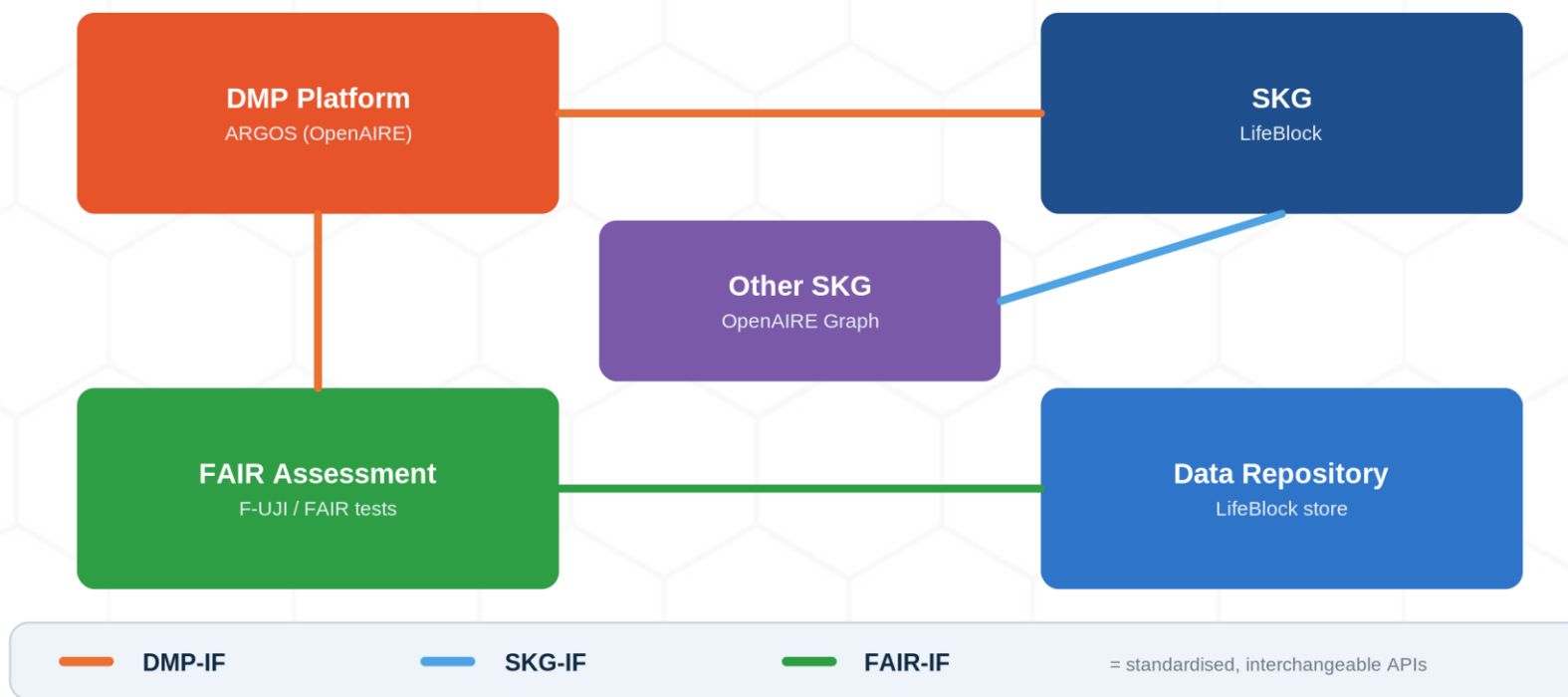
PART 5

Interoperability

Connecting our graph to the world — LifeBlock × OpenAIRE.

The OSTRails interoperability frameworks

One reference architecture · three interoperability frameworks



Three coordinated frameworks let independent tools plug together — and let one SKG talk to another.



A shared data model

The common entity types from part 2:

Research Product, Agent,
Grant, Venue, Topic,
Data Source.

One vocabulary for all.



An exchange format

A defined JSON / JSON-LD serialisation.

Any graph can export its content in it; any other graph can import it without custom glue.



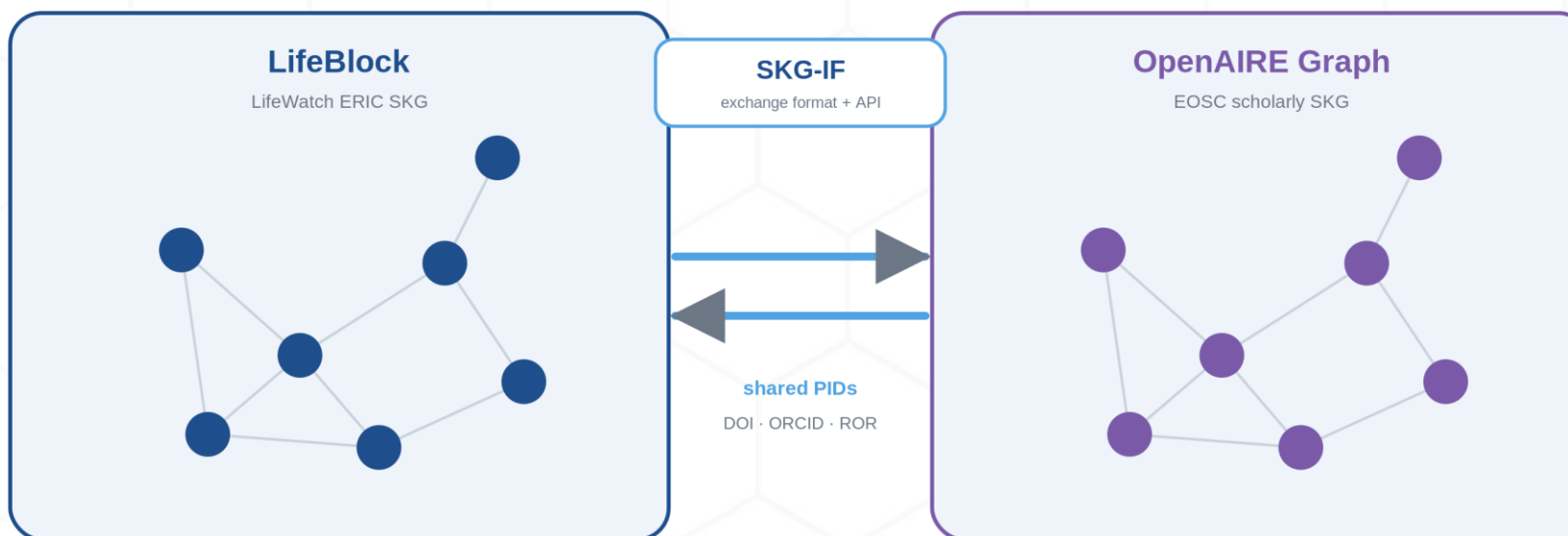
Agreed APIs

An SKG API so graphs can query each other:

search products, fetch a record, follow a link across graph boundaries.

RDA-endorsed, OpenAIRE-led. When two graphs speak SKG-IF, exchange stops being a bespoke integration project and becomes routine.

Connecting LifeBlock to the OpenAIRE Graph



Same identifiers + shared model → both graphs recognise the same entities and link across.

Shared PIDs identify the same entities; SKG-IF carries them; the two graphs link and enrich each other.

What you gain — and what's next



The chain we built

DMP → maDMP (ARGOS) → nodes & edges → LifeBlock SKG → SKG-IF → OpenAIRE. Each arrow is a standard, not a hack.



Write once, reuse everywhere

The plan you had to write anyway now fills the graph, feeds FAIR checks, and pre-populates your reporting.



FAIR by construction

PIDs and shared vocabularies mean your data are Findable, Accessible, Interoperable and Reusable from day one.



Next: we build one

In the hands-on session we'll assemble a small SKG together — from a maDMP, exactly as shown here.

The plan is no longer paperwork. It is the seed of a connected, interoperable knowledge graph.

Practical demonstration: LifeWatch SKG Studio

<https://joaquinlwe.github.io/skg-studio/>

The screenshot shows the LifeWatch SKG Studio web application. At the top left is the logo 'LifeWatch SKG Studio' with 'LIFEWATCH ERIC · OSTRAILS' below it. At the top right is a progress bar for 'New LifeWatch ERIC Data Management...' at 58%. The main content area has a light blue background with a grid of dots. In the center is a dark blue hexagonal icon. Below it is the title 'LifeWatch SKG Studio' and a paragraph: 'Author a machine-actionable Data Management Plan, then watch it become an interactive Scientific Knowledge Graph — connected to OpenAIRE. Built for LifeWatch ERIC · OSTRAILS Use Case 1.' Below this is a workflow diagram: '1. Build your maDMP' (with a document icon) followed by an arrow and '2. See it become a Knowledge Graph' (with a network icon). Below the diagram are two white cards. The first card is titled '1 maDMP Builder' and describes it as 'An ARGOS-style, OSTRAILS-compliant wizard to author your Data Management Plan — plan overview, project & funding, people, datasets, costs and the OSTRAILS organisation profile.' It has a button '→ Open the Builder'. The second card is titled '2 Knowledge Graph' and describes it as 'Explore your maDMP as a typed graph of research products, people, organisations and more — edit it freely and connect it to the OpenAIRE Graph via shared PIDs.' It has a button '→ Open the Graph'. At the bottom are two buttons: 'Load MARBEFES example' and 'Import maDMP JSON'.

LifeWatch SKG Studio
LIFEWATCH ERIC · OSTRAILS

New LifeWatch ERIC Data Management... 58%

LifeWatch SKG Studio

Author a machine-actionable Data Management Plan, then watch it become an interactive Scientific Knowledge Graph — connected to OpenAIRE. Built for LifeWatch ERIC · OSTRAILS Use Case 1.

1. Build your maDMP → 2. See it become a Knowledge Graph

1 maDMP Builder

An ARGOS-style, OSTRAILS-compliant wizard to author your Data Management Plan — plan overview, project & funding, people, datasets, costs and the OSTRAILS organisation profile.

→ Open the Builder

2 Knowledge Graph

Explore your maDMP as a typed graph of research products, people, organisations and more — edit it freely and connect it to the OpenAIRE Graph via shared PIDs.

→ Open the Graph

Load MARBEFES example Import maDMP JSON

Demo tool. Availability is not guaranteed, and functionality is limited.



The LifeWatch ERIC Biodiversity
& Ecosystem eScience Conference

PLOVDIV, BULGARIA

7-10 JULY 2026

THANK YOU!

Questions?

Joaquin.lopez@lifewatch.eu

