

Streaming knowledge graph-based digital twin for biodiversity restoration

Piotr Sowiński

Second AIOTI Workshop on Semantic Interoperability for Digital Twins
24 March 2026, Frankfurt (Oder), Germany

About us

- **NeverBlink** – deep-tech startup from Poland
- Building a breakthrough platform for **real-time data integration and decision making** based on **neurosymbolic AI**
- This project: **OBEREK**, open call under HE RURBANIVE
 - NeverBlink – coordinator
 - Universitat Politècnica de València
 - Fundació Assut



Problem context

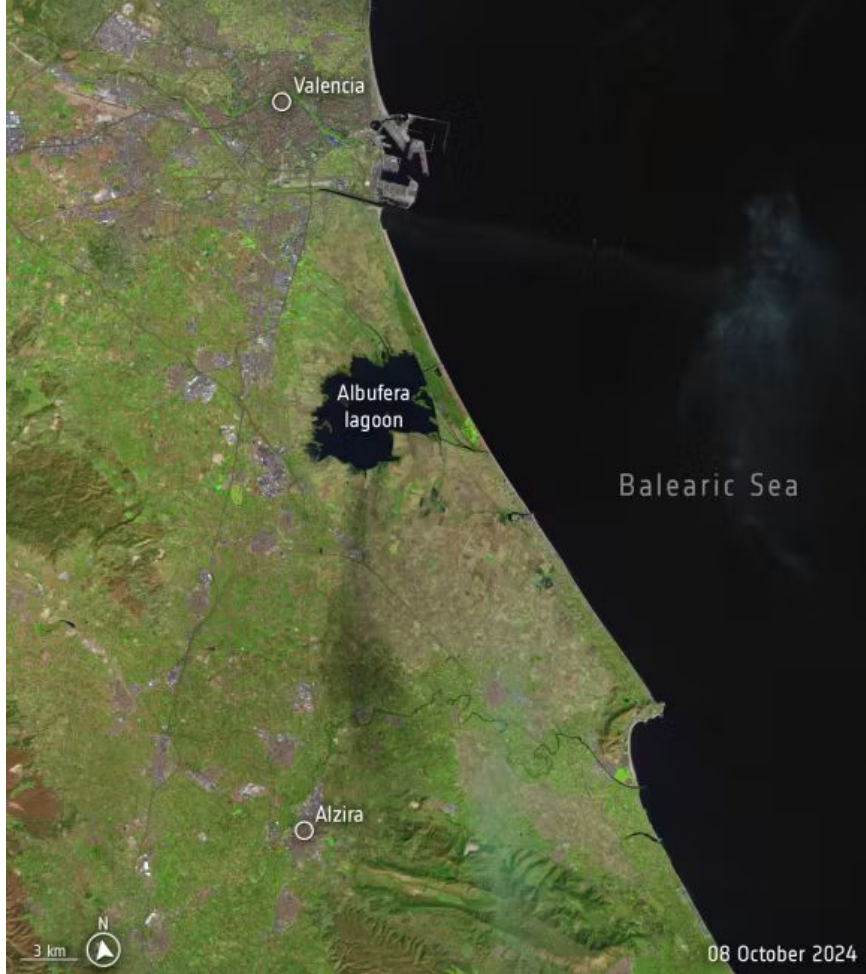
- **Albufera de València** – coastal lagoon, vital to region's biodiversity, food supply, and culture
- Albufera was **heavily damaged** in 2024 floods
- We need a **digital twin of Albufera** to track and guide biodiversity restoration efforts
- To build the digital twin, we need to merge **data from many sources**



Photo: Fundació Assut

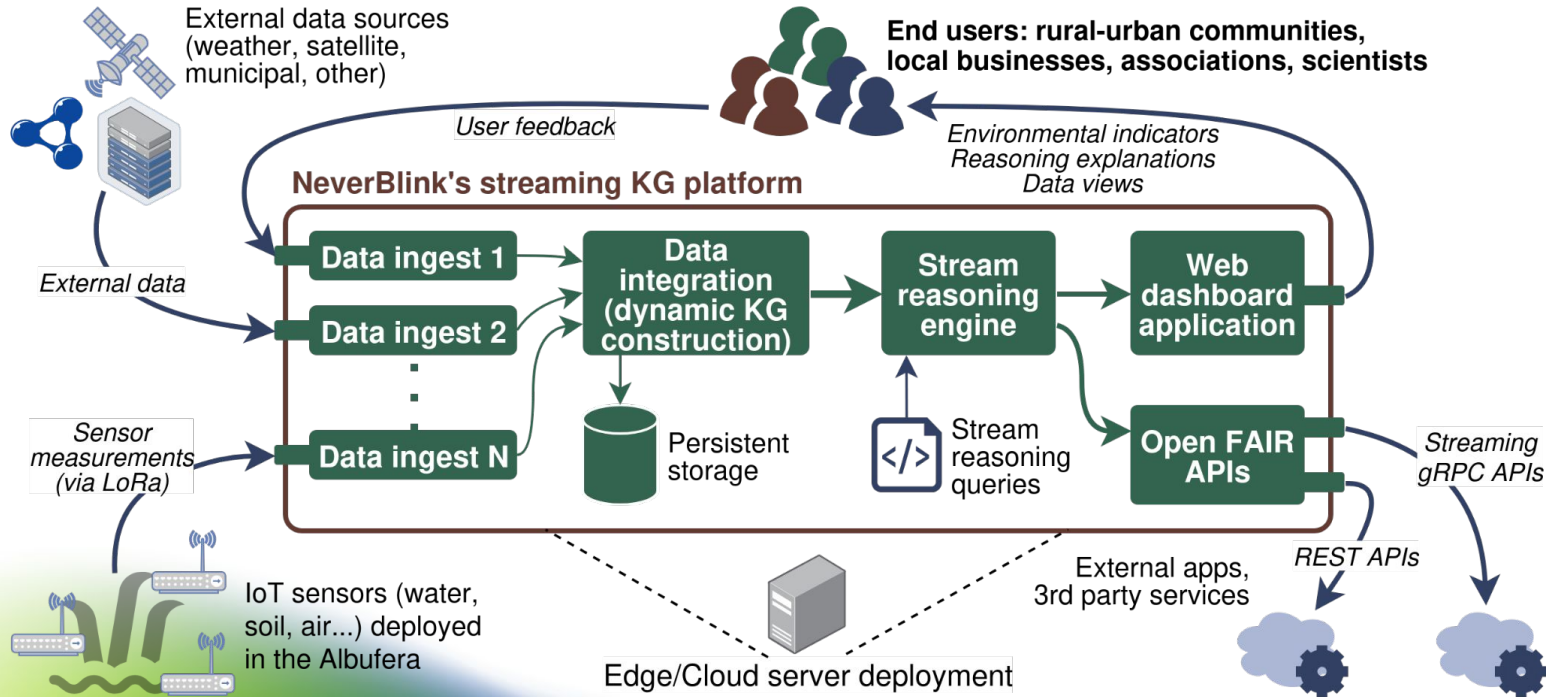


Photo: Le Monde



Source: [USGS](#), processed by ESA, via CNN Weather

Solution architecture



Data sources



IoT: water quality (3 sites)



Multispectral satellite imagery (Sentinel-2)

datos.gob.es
potenciando nuestra economía digital

EN

Data Catalogue > Datasets > Redes de control...

Redes de control: Estaciones de la Albufera.

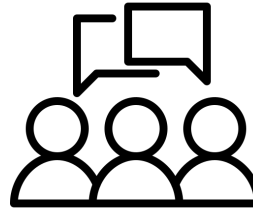
Environment

Publisher: Confederación Histórica del Júcar
Administration level: State Administration
Entity: Public
License: <http://datos.gob.es/?causale=legal>

Description: Información sobre la red de seguimiento para el control de la evolución del nivel de la lámina de agua y caudales en l' Albufera.

Data Information Technical sheet Quality

Municipal APIs (e.g., water flow data)



Community feedback

Territorio: **TODO EL TERRITORIO**

Estación	Temperatura				Punto de Humedad Relativa				Riesgo AVAMET		Zonas Infiltrables	
	actual	max	min	prom	actual	max	min	prom	0-1	2-3	4-5	6-7
Albufera	°C	°C	°C	°C	%	%	%	%	mm	mm	mm	mm
Valencia (Albufera)Pisc. de l'Albufera (Infiltración)	9,9	7,2	10,8	7,6	9,3	8,6	0,2	6	50	18		
Valencia (Albufera)Tancat de la Pica	9,1	7,0	11,4	7,0	8,7	0,0	0	0	0	0	16	
Sueca (Albufera)Má. de l'Estret	9,6	5,9	9,7	7,9	8,2	8,9	0,2	10	0	0	16	
Sueca (Albufera)Tancat de l'Estret	9,6	5,7	9,8	7,8	8,1	8,9	0,0	10	0	0	16	

858 estaciones operativas actualizadas: 22-03-2026 09:35

Autoría: **El Júcar**
Ejecución, auditoría y validada por AVAMET favorablemente para todos los sensores.
Ejecución, auditoría y validada por AVAMET favorablemente.

Tipo de estación: **Estación propia de AVAMET**
Estación propia de AVAMET

Actualización: **0**
Hora de la última actualización (cada 5 minutos)

Local weather stations (AVAMET)

Data sources – challenges

- **Diverse formats** (JSON, custom text, CSV, TIFF, Excel, ZIP...)
- **Diverse access modes** (MQTT, HTTP...)
- **Non-uniform units** of measure
- Data source **context must be preserved**
- **Dense numerical data** from the satellite

Solution: semantic interoperability!

But how?

KG lifting – YARRRML

- Used for CSV & JSON sources
- Easier to write than RML
- Still requires deep understanding of RDF!
-> **we are working on easier alternatives**

```
1 INDI_CHJ, Fecha, Nivel (m.s.n.m), Caudal (m³/s), Volumen (hm³), Estimado
2 08A01, 05/02/2026, 0.381,,, FALSE
3 08A02, 05/02/2026, 0.426,,, FALSE
4 08A03, 05/02/2026, 0.393, 1.143, 0.099, FALSE
5 08A04, 05/02/2026, 0.340, 2.462, 0.213, TRUE
6 08A05, 05/02/2026, 1.004, 4.155, 0.359, TRUE
7 08A06, 05/02/2026, 0.536, 1.657, 0.143, FALSE
```

```
10 sources:
11   defaultSource:
12     access: data.csv
13     referenceFormulation: csv
14     delimiter: ",",
15
16 mappings:
17
18   FlowVolume:
19     sources: defaultSource
20     s:
21       value: blank_$(INDI_CHJ)_$(Fecha)_FlowVolume
22       type: blank
23     po:
24       - [ a, saref:Observation ]
25       - [ saref:madeBy, "$(INDI_CHJ)" ]
26       - [ saref:hasTimestamp, "$(Fecha)" ]
27       - [ saref:observes, s4watr:FlowVolume ]
28       - p: saref:hasResult
29         o:
30           value: blank_$(INDI_CHJ)_$(Fecha)_FlowVolumeResult
31           type: blank
32
33   FlowRate:
34     sources: defaultSource
35     s:
36       value: blank_$(INDI_CHJ)_$(Fecha)_FlowRate
37       type: blank
```

KG lifting – data transforms

- Sometimes you just get a completely custom data format 🙄
- **Solution:** data transforms with NeverBlink's RDF extensions

```
1 local lines = std.split(input, "\n");
2 local chunks = std.flattenArrays(
3   [std.split(line, "|") for line in lines]
4 );
5
6 [
7   rdf.triple(
8     "nb:" + std.split(chunk, "=")[0],
9     "saref:hasValue",
10    rdf.literal(std.split(chunk, "=")[1])
11  )
12  for chunk in chunks
13  if std.length(std.split(chunk, "=")) ≥ 2
14 ]
```

```
1 |esta=tancatestell|data=22/03/26 11:39|temp=15.1|temp_max=15.1|temp_mxt=11:39|
temp_min=5.7|temp_mnt= 7:07|hrel=72|hrel_min=70|hrel_max=95|pres=1011.9|pres_min=1011.
6|pres_max=1013.0|vent=6.4|vent_dir=177|vent_max=17.7|vent_mxt=11:11|prec=0.0|
prec_mes=47.8|prec_any=91.0|prec_epi=0.0|prec_mid=0.0|prec_mih=----|eto=0.61|
eto_men=45.57|eto_any=193.14|rad=686|rad_max=686|rad_tmx=11:38|fulla=0|vent_max1=9.7|
vent_max2=9.7|vent_max10=11.3|vent_max15=11.3|vent_max60=11.3|prec_in10=0.0|
prec_in1h=0.0|prec_in12h=0.0|prec_10=0.0|prec_1h=0.0|prec_12h=0.0|
```

What about satellite data?

- Surely you can't put a 15-channel hi-res multispectral image into RDF and process it?
- **Yes you can – with RDF-tensor!**
<https://w3id.org/rdf-tensor>
- TIFF -> tensor literal
- Native tensor representation in memory
(dense numerical array)
-> zero overhead!



```
1 PREFIX saref: <https://saref.etsi.org/core/>
2 PREFIX tensor: <https://w3id.org/rdf-tensor/vocab#>
3
4 [] saref:hasValue
5   | "{\\"type\\": \\"float32\\", (...data)}"^^tensor:DataTensor
6 .
```

```
1 {
2   "type": "float32",
3   "shape": [
4     512,
5     512,
6     15
7   ],
8   "data": [
9     0.2621999979019165,
10    0.31869998574256897,
11    0.35339999198913574,
12    0.3711000084877014,
13    0.36890000104904175,
```

Unique capability – tensors & ML in SPARQL

- “Numpy in SPARQL”
- Hardware acceleration for tensor ops (CPU, GPU, NPU...)
- Everything done in a unified inference engine

<https://w3id.org/rdf-tensor>

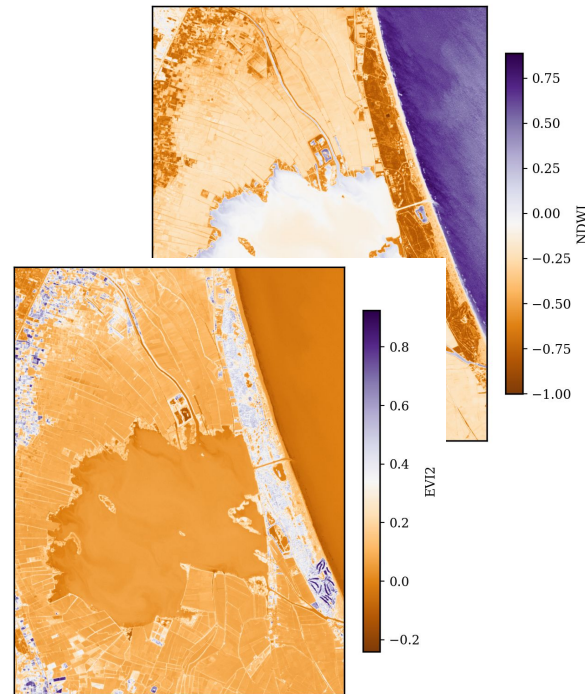
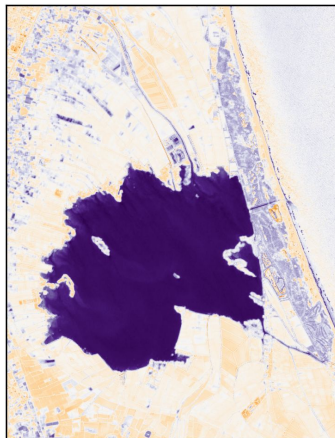
```
107 BIND (t:divide(t:subtract(?b03, ?b08), t:add(t:add(?b03, ?b08), t:create("float32", 1E-06))
108 BIND (t:multiply(?ndwi, ?mask_cloud) AS ?ndwi_masked)
109
110 BIND (t:divide(t:subtract(?b08, ?b11), t:add(t:add(?b08, ?b11), t:create("float32", 1E-06))
111 BIND (t:multiply(?ndmi, ?mask_cloud) AS ?ndmi_masked)
112
113 BIND (t:subtract(?b08, ?b04) AS ?evi_numer)
114 BIND (t:add(t:subtract(t:add(?b08, t:multiply(t:create("float32", 6.0), ?b04))), t:multiply
115 BIND (t:multiply(t:create("float32", 2.5), t:divide(?evi_numer, t:add(?evi_denum, t:create
116 BIND (t:multiply(?evi, ?mask_cloud) AS ?evi_masked)
117
118 BIND(
119   rusaLka:infer(
120     "fcover",
121     "fcover_map",
122     "input_image", ?fcover_input
123   ) AS ?fcover
124 )
125
```

← ML inference directly in SPARQL!

Unique capability – tensors & ML in SPARQL



**SPARQL +
RDF-tensor**



Streaming pipelines

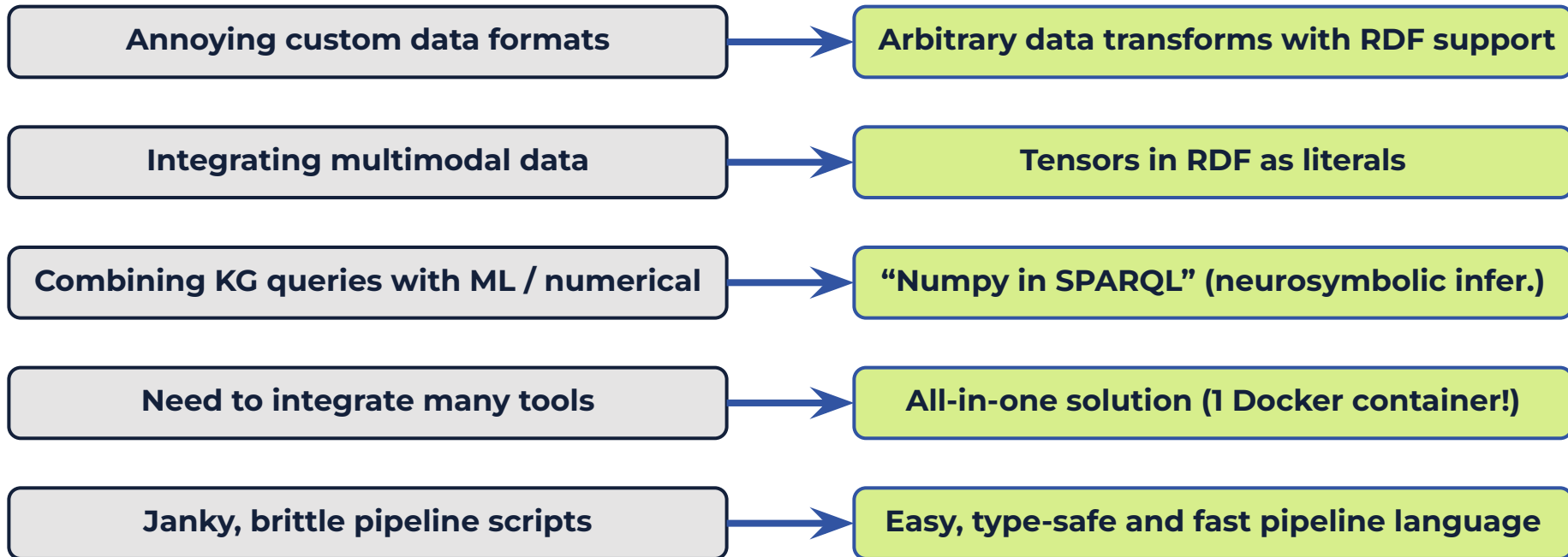
```
1 type: composite-runnable
2 components:
3   # Subscribe to IoT JSONs from MQTT
4   - source: mqtt-subscribe
5     host: mosquitto
6     port: 1883
7     topics:
8     | - iot/data/north
9
10  # Lift JSON to RDF using RML
11  - flow: stream-from-single
12  - flow: lift-json-to-rdf-rml
13  | model: iot-water-yarrrml
14
15  # Send for further processing
16  - sink: broadcast
17
```

```
1 type: composite-runnable
2 components:
3   - source: from-broadcast-zip
4     from:
5       - ingest-iot-N
6       - ingest-iot-S
7       - ingest-iot-SW
8       - csv-ingest-mapped-yarrrml
9       - copernicus-temp
10    in-type: rdf-dataset
11    merge-strategy: take-latest
12
13  - flow: merge-rdf-datasets
14
15  - flow: sparql-stream-graph-query
16    query: |
17      CONSTRUCT {
18        nb:OBEREK a nb:RUE ;
19        | saref:madeExecution ?obs_iri .
20
21        ?obs_iri a saref:Observation ;
22        | saref:hasTimestamp ?now_timestamp ;
23        | saref:observes nb:KEI-1 ;
24        | saref:hasFeatureOfInterest ?foi_iri ;
25        | saref:hasResult ?prop_val_iri .
26
27        ?foi_iri a saref:FeatureOfInterest ;
28        | geo:hasBoundingBox ?geom_iri .
29
```

Presentation – live dashboard



Summary: challenge -> solution



Do you have a challenge for us?

 piotr@neverblink.eu

 <https://neverblink.eu/>

 [company/neverblink](https://www.linkedin.com/company/neverblink)

NeverBlink sp. z o.o.
ul. Wspólna 56, 00-684,
Warszawa Poland
VAT: PL7011239754

This project has indirectly received funding from the European Union's Horizon Europe research and innovation programme under project RURBANIVE (Grant Agreement number: 101136597). Total project budget: €100,000.00.

OBEREK's deliverables and public materials reflect only views of the OBEREK consortium. The European Commission or the RURBANIVE project are not liable for any use that may be made of the information contained therein.



Co-funded by
the European Union



UNIVERSITAT
POLITÀCNICA
DE VALÈNCIA

