



*Joint COOPEUS, ENVRI and EUDAT workshop
on persistent digital identifiers
(PID) for open time series data
WP6/Biodiversity/Lifewatch*

Francisco Manuel Sanchez Cano fco.manuel.sanchez83@gmail.com [CSIC & CWI]

Fernando Aguilar aguilarf@ifca.unican.es

Jesús Marco marco@ifca.unican.es

Instituto de Física de Cantabria (IFCA), CSIC, SANTANDER, SPAIN

- Highlights from the Motivation:

A major prerequisite for the proper use of persistent identifiers (PID) e.g. within data citations is the persistence of both, identifiers as well as the **integrity of the associated data set**. This poses questions when **PIDs are to be used for unfinished data sets** or open time series data. Such data is typically generated within research infrastructures (RI) during long lasting experiments such as satellite missions, **environmental monitoring campaigns**, or in permanent installations such as natural hazard detection and **early warning systems**.

Open time series data are often used in research during ongoing experiments and potentially published earlier than the underlying data set has been closed and is publicly released. It is therefore important to enable the scientific community to properly cite these data in their publications and the proper use of PIDs is of key importance to reach this goal

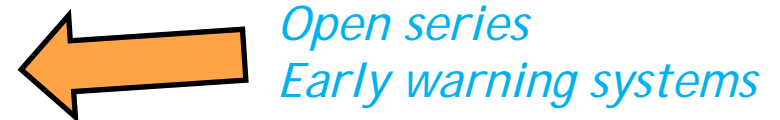
- DATA LEVEL (in increasing complexity/volume)

- Level 1: paper / published results
- Level 2: data in simplified format (outreach)
- Level 3: data used by researchers
- Level 4: RAW data



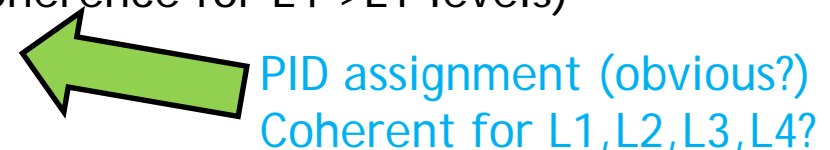
- ACTIONS (reverse time order)

- Data Quality Management on RAW data
 - include ID assignment [Period, RUN, ...] and conditions DB links
- Data Certification & Software Validation
- Data Preservation (guaranteeing coherence for L4->L1 levels)

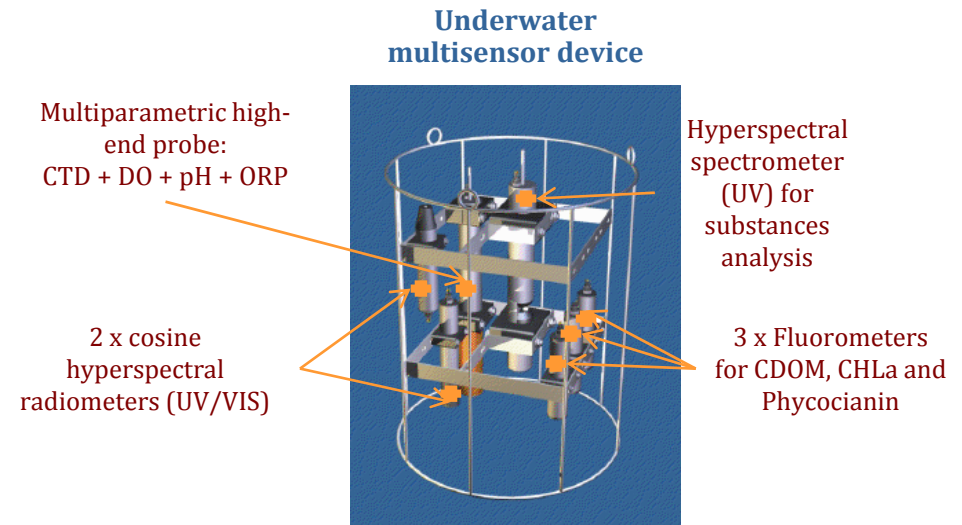


- ROLES

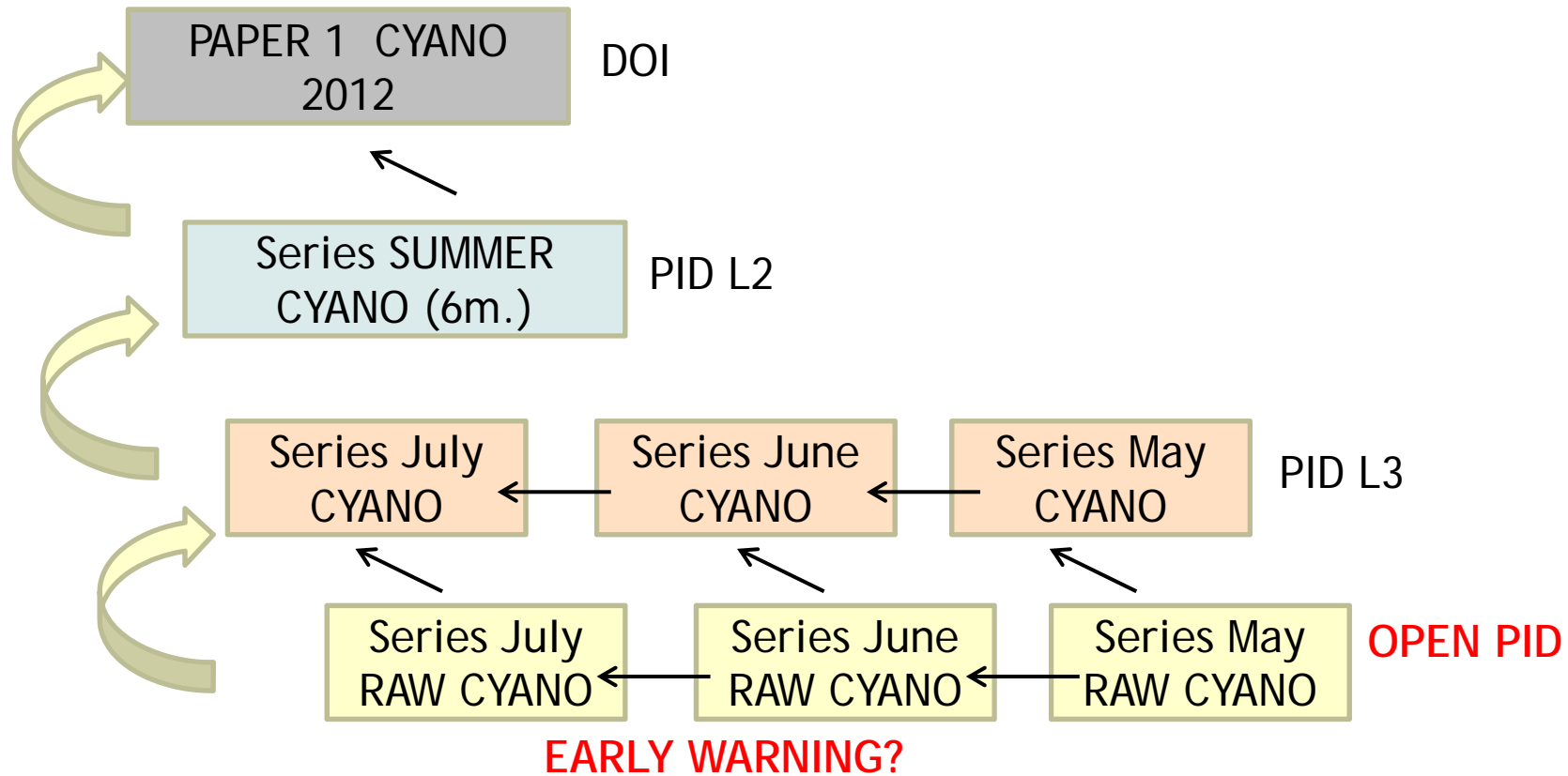
- Instrument operators
- Data/Software process experts
- Team Researchers doing "initial" analysis (basis of instrument?)
- General Researchers (accessing to open data)
- Public



An autonomous multisensorial device to monitor and forecast the dynamics of toxic cyanobacteria in a remote water reservoir

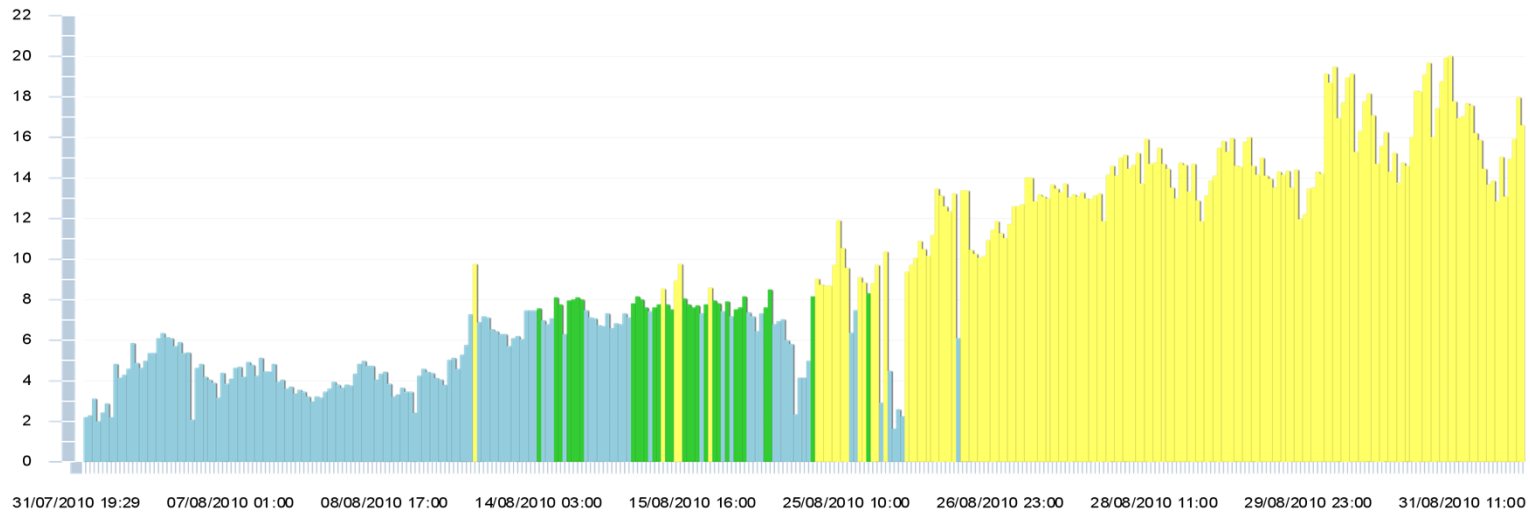


- ⊕ Open Time Series issues :
 - ⊠ Fluorometers measurements (hourly profiles, 2 min. measurements)
 - ⊠ EARLY WARNING SYSTEM
 - ⊠ DQM applied later
- ⊕ Typical RUN period: 1 month (time lapse between revisions)
- ⊕ Use of two different DBMS: On-line & Off-line (DQM applied)
- ⊕ Starting to work now on PID...



Phycocyanin concentrations could be toxic

- ⊕ Different warning levels
- ⊕ based on data taken every 2 minutes while profiling



Series: August 2010



Status of initial tests of PID for Lifewatch

- PIDs are being tested using SARASurf test EPIC's PID service on wireless sensor network deployments
 - Identify sensor nodes
 - URL
 - SERIAL
 - LOCATION
 - DATASHEET
 - Identify phenomena + metadata

The SURF SARA logo consists of the words "SURF" and "SARA" in white, bold, uppercase letters. "SURF" is inside a black rounded rectangle, and "SARA" is to its right. A black speech bubble tail points from the "SURF" box towards the text.

SURF SARA



Status of initial tests



idx	type	parsed data	data	timestamp	ttl type	ttl	refs	privs
1	URL	http://sensenets.project.cwi.nl/lifewatchsensors/scripts/utis/scaninfo.php?uid=S68	IGH0dHA6Ly9zZW5zZW5ldHMucHJvamVjdC5jd2kubmwwbGlmZXdhdGNoc2VuY3R5cy9zY3JpcHRzL3V0aWxzL3NjYW5pbmZvLnBocD91aWQ9UzY4	2012-12-03T17:58:33Z	0	86400		rwr-
2	SERIAL	OKLPIJU79	T0tMUEIKVTc5	2012-12-03T17:58:33Z	0	86400		rwr-
3	OWNER_EMAIL	sanchez@cwi.nl	c2FuY2hlekBjd2kubmw=	2012-12-03T17:58:33Z	0	86400		rwr-
4	LOCATION	336465	MzM2NDY1	2012-12-03T17:58:33Z	0	86400		rwr-
5	BACKGROUND	SMART CITY	U01BUIQgQ0IUWQ==	2012-12-03T17:58:33Z	0	86400		rwr-
6	DATASHEET	http://www.libelium.com/documentation/waspmote/waspmote-datasheet_eng.pdf	aHR0cDovL3d3dy5saWJlIG11bS5jb20vZG9jdW11bnRhdGlvbi93YXNwbW90ZS93YXNwbW90ZS1kYXRhc2hlZXRfZW5nLnBkZg==	2012-12-03T17:58:33Z	0	86400		rwr-

PID Resolution



Monitoring Gas Sensor S02

Location Printer corridor

Last update received at 2013-05-23 18:53:43

Sensor	Current Value	Average Value
Battery Level	89.00%	67.99%
Temperature	26.12	25.505
Humidity	22.40%	36.804%
CO	0.51%	0.509%
CO2	324.89	327.771
Air Pressure	196.35mm	192.630mm



Status of initial tests



- PIDs are used not only to identify unique specific devices (sensors)
- PIDs contain relevant information associated with the device referenced
- Which information should be stored within PIDs in order to optimize their usage and make the most of their potential?





PIDs are destined to become...



- Tools for creating semantically correct input
 - Register sensor platform
 - Standard form offering a wide sensor description range (measurement + characteristics)
 - EPIC PID generator
 - LifeWatch PID Database





PIDs are destined to become...



Sensor Data Stream Form

PID Generator:

Owner:

Display Sensor Description:

IPv6 Endpoint for data collection:

GPS Location

Latitude:

Longitude:

Kind of Board Sensor


▾

Timestamp Unit

▾

Phenomenon

▾

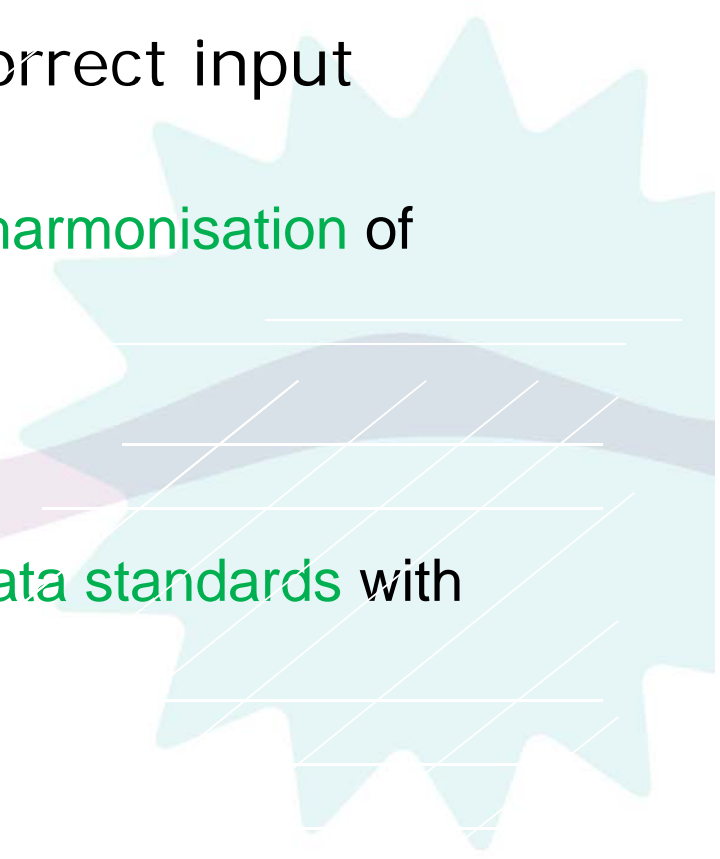
A close-up photograph of a lynx's head in profile, looking towards the left. The lynx has thick, brown fur with darker spots and stripes, and its ears are pointed upwards.



PIDs are destined to become...



- Tools for creating semantically correct input
 - Advantages:
 - Attributes standardized facilitate **harmonisation** of (meta)data standards
 - Facilitating WSN deployment
 - Supporting adherence to **metadata standards** with **zero-programming** effort

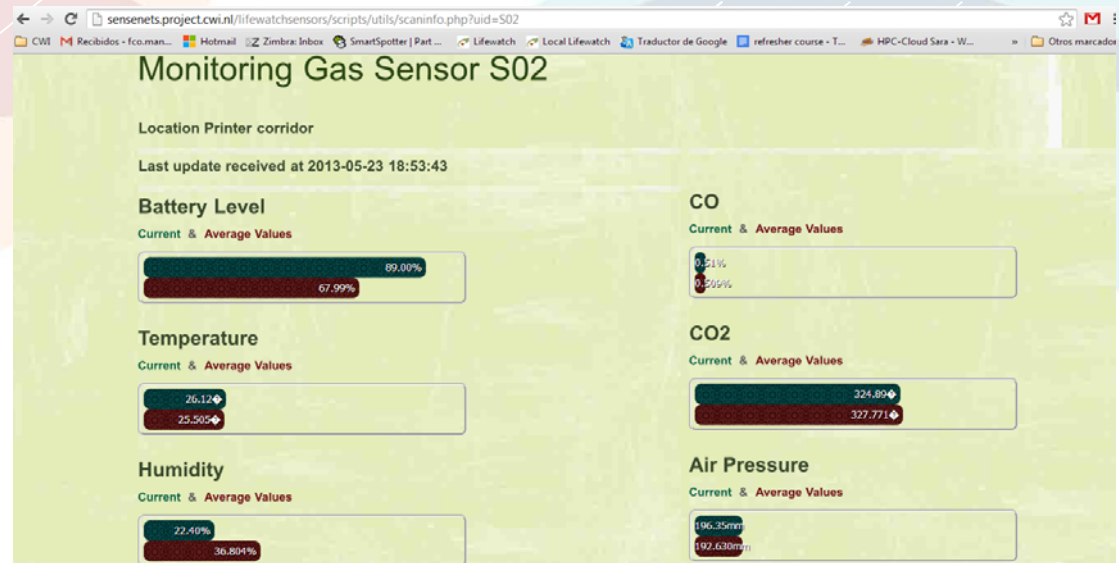
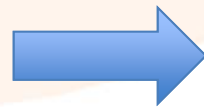




PIDs are destined to become...



- Tools for accessing data
 - QR App
 - QR code contains PID reference
 - Resolution facilitates access to sensor information





PIDs are destined to become...



- Tools for accessing data
 - Code Snippet App
 - Extracts information from sensor PID attributes in order to generate:
 - Human readable information
 - M2M communication feedback
 - RDF Triples
 - RestDesc + EYE reasoner N3 files

