

COST ACTION CA23104

MAINSTREAMING WATER REUSE INTO THE CIRCULAR ECONOMY PARADIGM (Water4Reuse)

WG5: Data and Databases for water REUSE



WG5 : Data and Databases for water REUSE



Vicky Ruano

Associate Professor

Department of Chemical Engineering -
School of Engineering
Universitat de València , Valencia, Spain
e-mail: m.victoria.ruano@uv.es

WG5-Leader



Dr. Eng. Anacleto Rizzo

IRIDRA Srl

Via A. La Marmora 51 - 50121 FIRENZE

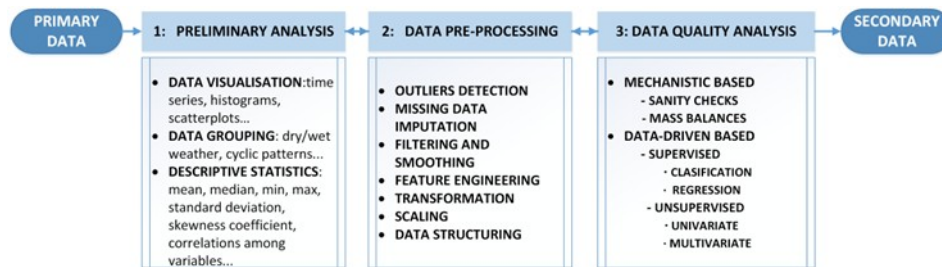
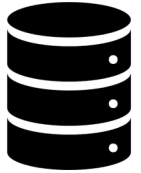
rizzo@iridra.com

WG5-Co-Leader

WG5 : Data and Databases for water REUSE

□ Objectives

- Assemble a panel of experts to propose reliable and usable methods for water reuse data management
- Generate databases for the water reuse value chain from collected data
 - Collaborate with WG1, WG2, WG3 and WG4 to define the databases to be generated
 - Common framework for data collection, organization and storage.
 - Applied data quality pipeline techniques to integrate sensor data quality meta-data



- Ensure **reliability, availability, accessibility, and usability** of databases for the water reuse value chain



WG5 : Data and Databases for water REUSE

☐ Members presentation:

- ☐ Background, current work and ideas to contribute to WG5

- ☐ Judit Babcsanne Kiss

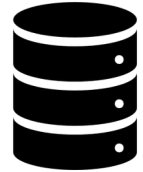
- ☐ Michael (Iggy) Litaor – geochemistry and geophysics

- ☐ Michal Snehota (CVUT) – subsurface hydrology (data mining and management), water flow in urban area, NbS for water recycling in building

- ☐ Janelcy Alferes (Vito) – data chain (data generation and management) in wastewater

- ☐ İbrahim Öztürk – data visualization (experience air pollutant data)

- ☐ Lorenzo Tombolini (UNIVPM) – wastewater engineer



WG5 : Data and Databases for water REUSE

Year		1				2				3				4			
Quartile		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
WP5. Data and databases for water REUSE																	
Tasks	Task 5.1: Collection data																
	Task 5.2: Categorized collected data																
	Task 5.3: Generate Meta-Data																
	Task 5.4: Generate Quality Meta-Data																
	Task 5.5: Create Databases																
Activities	Activity A5.1: Joint seminars for common framework for data acquisition and recognition																
	Activity A5.2: Promoting collaboration with expert panels (METACo, IWA Digital Water Programme																
	Activity A5.3: Promoting grants for YRIs from ITC to partipate in conferences, organize invited session or seminars to train students in Big-Data Management																
	Activity A5.4: Organization of international research schools by YRIs from ITCs on Data Science and Digitalization																
Deliverables	D5.1: Protocol for data processing																
	D5.2: Databases available in open repositories																
	D5.3:Database evaluation and validation																

Year	1				2				3				4			
Quartile	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
WP5. Data and databases for water REUSE																
Task 5.1: Collection data																

WG5 : Data and Databases for water REUSE

Task 5.1, 5.2, 5.3: Data Collection, Data Categorization, Meta-Data Generation

- **Actions:**

- **Data Governance Framework:**

- Reference Data sheets” for standardize data collection, structure and storage”.

» Propose a common framework for data collection:

- Signal data/Lab data
 - Data records: Timestamp + Data point + Qualifier
- Metadata (Basics/Advanced): Standardize nomenclature
 - Describing signal generation
 - Sensor data quality (measurement errors, diagnosis)
 - Annotations (Contextual information)

» Guidelines to structure and store data in Reference Data Sheets

Year		1				2				3				4			
Quartile		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
WP5. Data and databases for water REUSE																	
Tasks	Task 5.1: Collection data																
	Task 5.2: Categorized collected data																
	Task 5.3: Generate Meta-Data																
	Task 5.4: Generate Quality Meta-Data																

WG5 : Data and Databases for water REUSE

Task 5.1, 5.2, 5.3: Data Collection, Data Categorization, Meta-Data Generation

- **Actions:**

- **Data Governance Framework:**

- Database technology to be used:

- » Flat-file databases (csv, xls)
- » Relational databases (SQL Server, MySQL...)
- » Framework of the existing repositories

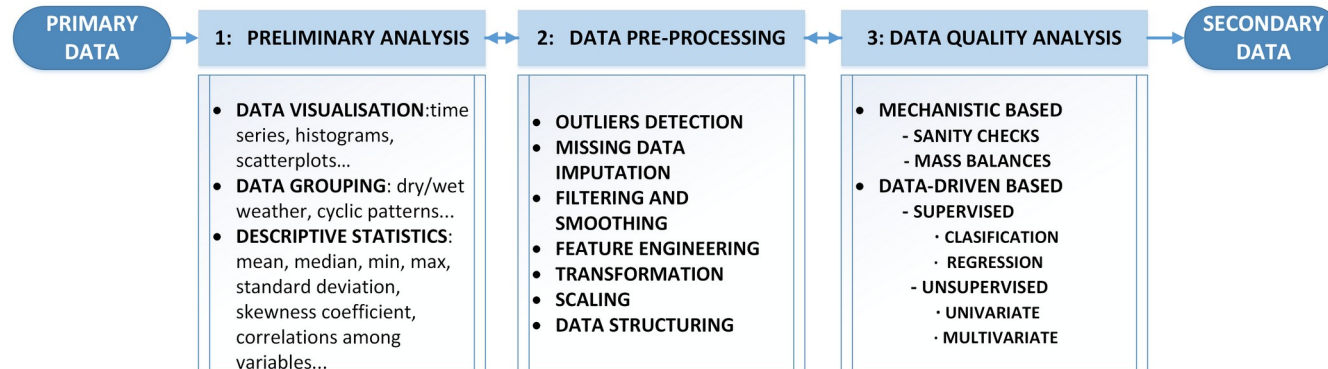
Year		1				2				3				4			
Quartile		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
WP5. Data and databases for water REUSE																	
Tasks	Task 5.1: Collection data																
	Task 5.2: Categorized collected data																
	Task 5.3: Generate Meta-Data																
	Task 5.4: Generate Quality Meta-Data																

WG5 : Data and Databases for water REUSE

□ Task 5.4: Smart-Data Generation.

– Actions:

- Implement **data quality pipeline techniques** to generate sensor signal quality meta-data
 - Assess the available reference data
 - Define the techniques to be applied within the data processing pipeline



- Review available software tools that facilitate the implementation of data quality techniques

WG5 : Data and Databases for water REUSE

Expected deliverable for 2025:

Year		2024		2025								2026								2027								2028																								
Year		1												2												3												4														
Quartile		1			2			3			4			1			2			3			4			1			2			3			4																	
Month		N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O			
WP5. Data and databases for water REUSE		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	
Tasks	Task 5.1: Collection data	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	
	Task 5.2: Categorized collected data							■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	Task 5.3: Generate Meta-Data							■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	Task 5.4: Generate Quality Meta-Data							■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	Task 5.5: Create Databases																										■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Activities	Activity A5.1: Joint seminars for common framework for data acquisition and recognition																																																			
	Activity A5.2: Promoting collaboration with expert panels (METACo, IWA Digital Water PRogramme																																																			
	Activity A5.3: Promoting grants for YRIs from ITC to partipate in conferences, organize invited sesion or seminars to train students in Big-Data Management																																																			
	Activity A5.4: Organization of international research schools by YRIs from ITCs on Data Science and Digitalization																																																			
Deliverables	D5.1: Protocol for data processing																	■																																		
	D5.2: Databases available in open repositories																																																			
	D5.3:Database evaluation and validation																																																			

D5.1. Protocol for data processing: (Feedback from Task 5.1, 5.2, 5.3 and 5.4)

- Should include:
 - “Reference Data sheets”; Databases types, data governance framework (definition, collection, structure, storage), guidelines for data quality techniques.

WG5 : Data and Databases for water REUSE

☐ Leaders of Actions:

- ☐ Databases to be created (ALL)
- ☐ Each database depending on whether its objective is specific or more general (Leader?)
- ☐ Common framework to define the minimum data to be collected:
 - ☐ Data records: Timestamp + Data point + Qualifier
 - ☐ Basic Metadata (Nomenclature)
- ☐ Database technology
 - ☐ Structure and storage the collected data
- ☐ Generate additional metadata for sensor quality data:
 - ☐ Data quality pipeline techniques

☐ Expected dissemination results:

- ☐ Journal paper:
 - ☐ "Towards a unified framework for specific Databases in the water sector"
 - ☐ "The importance of metadata to improve the usability of DataBases in Water Reuse Value Chain"
 - ☐ "Databases"
 - ☐ "..."

WG5 : Data and Databases for water REUSE

ACTIVITIES for 2025:

Year		2024				2025					2026					2027					2028																														
Year		1										2										3										4																			
Quartile		1		2		3		4		1		2		3		4		1		2		3		4		1		2		3		4																			
Month		N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O		
WP5. Data and databases for water REUSE		□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	
Tasks	Task 5.1: Collection data	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	
	Task 5.2: Categorized collected data							■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	Task 5.3: Generate Meta-Data							■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	Task 5.4: Generate Quality Meta-Data							■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	Task 5.5: Create Databases																										■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	
Activities	Activity A5.1: Joint seminars for common framework for data acquisition and recognition																																																		
	Activity A5.2: Promoting collaboration with expert panels (METACo, IWA Digital Water PProgramme																																																		
	Activity A5.3: Promoting grants for YRIs from ITC to partipate in conferences, organize invited session or seminars to train students in Big-Data Management																																																		
	Activity A5.4: Organization of international research schools by YRIs from ITCs on Data Science and Digitalization																																																		
Deliverables	D5.1: Protocol for data processing																	■																																	
	D5.2: Databases available in open repositories																																																		
	D5.3:Database evaluation and validation																																																		

A5.3

- 14th IWA International Conference on Instrumentation, Control and Automation (ICA) in Oslo (30th June -2nd July 2025)
- Possible training schools for YRIs to acquired skills on Big-Data Management

WG5 : Data and Databases for water REUSE

ACTIVITIES for 2025:

Year		2024					2025					2026					2027					2028																													
Year		1										2										3										4																			
Quartile		1		2		3		4		1		2		3		4		1		2		3		4		1		2		3		4																			
Month		N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O		
WP5. Data and databases for water REUSE		□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	
Tasks	Task 5.1: Collection data	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	
	Task 5.2: Categorized collected data							■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	Task 5.3: Generate Meta-Data							■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	Task 5.4: Generate Quality Meta-Data							■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	Task 5.5: Create Databases																										■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	
Activities	Activity A5.1: Joint seminars for common framework for data acquisition and recognition																																																		
	Activity A5.2: Promoting collaboration with expert panels (METACo, IWA Digital Water PProgramme																																																		
	Activity A5.3: Promoting grants for YRIs from ITC to partipate in conferences, organize invited session or seminars to train students in Big-Data Management																																																		
	Activity A5.4: Organization of international research schools by YRIs from ITCs on Data Science and Digitalization																																																		
Deliverables	D5.1: Protocol for data processing																	■																																	
	D5.2: Databases available in open repositories																																																		
	D5.3:Database evaluation and validation																																																		

A5.3

- 14th IWA International Conference on Instrumentation, Control and Automation (ICA) in Oslo (30th June -2nd July 2025)
- Possible training schools for YRIs to acquired skills on Big-Data Management

Notes

- **We need participants to WG5!!!!!!**
- Database from case studies? Not clear yet
- Open source existing database? No funds for buying software
- **Zenodo** a good option? Already experienced by Michal
 - No problem of space
 - Data in Brief or other data journals
 - Usually csv, suitable for us?
 - It seems not interactive data storage
- Importance framework for the reuse, not the dataset per se
- Starting with dataset and moving in a second phase with database?
- **Impossible to choose the technology before knowing the kind of data we have to manage**
- **Pro and cons of existing tools for database?** To be ready to choose the most suited once the dataset will be more clear
- **Close database instead of live database**
- Start to decide how to store dataset in readable format
- Clarify available data packages
- Initial steps to develop dataset (data, column, separation)
- Ask amount and quantity of samples data to WGs and sample
- **Data paper accompanying Review paper from other WGs**
- Which data are we looking for?
- Do not reinvent the wheel....common way of doing review risk to loose a lot (e.g. SCOPUS), a lot of research not findable, future dataset/database suitable for AI with meta data..... Zenodo can fit the purpose?
- Prioritize long term time series



Suggested first steps?

- State of the art on available data frameworks/data bases (open sources) that could accommodate generic water related time series. Look for pros and contras in preparation of decision later (in function of specific datasets from the case studies/WGs) - **Janelcy**
- The structure of data packets from available IoT devices can be requested from other groups to ensure compatibility and integration. For example, details such as the specific IoT device being used, the communication protocol it follows, and sample data formats can be shared to facilitate seamless data exchange and interoperability - **Ibrahim**
- To prepare the questions/questionnaire on case studies for the meetings with other WG's. - **Vicky/Anacleto/Michal**
- Anacleto, Vicky, for the case studies, consider checking <https://www.hotspotreuse.com/> (from Water Reuse Europe association) and Water360 de for direct potable reuse, nice web-based mapping for visualisation water reuse case studies - **Quim**



COST ACTION CA23104

SUGGESTIONS....

WG5: Data and Databases for water REUSE

