

# *Data processing in the service of water - or how to transform data into knowledge?*

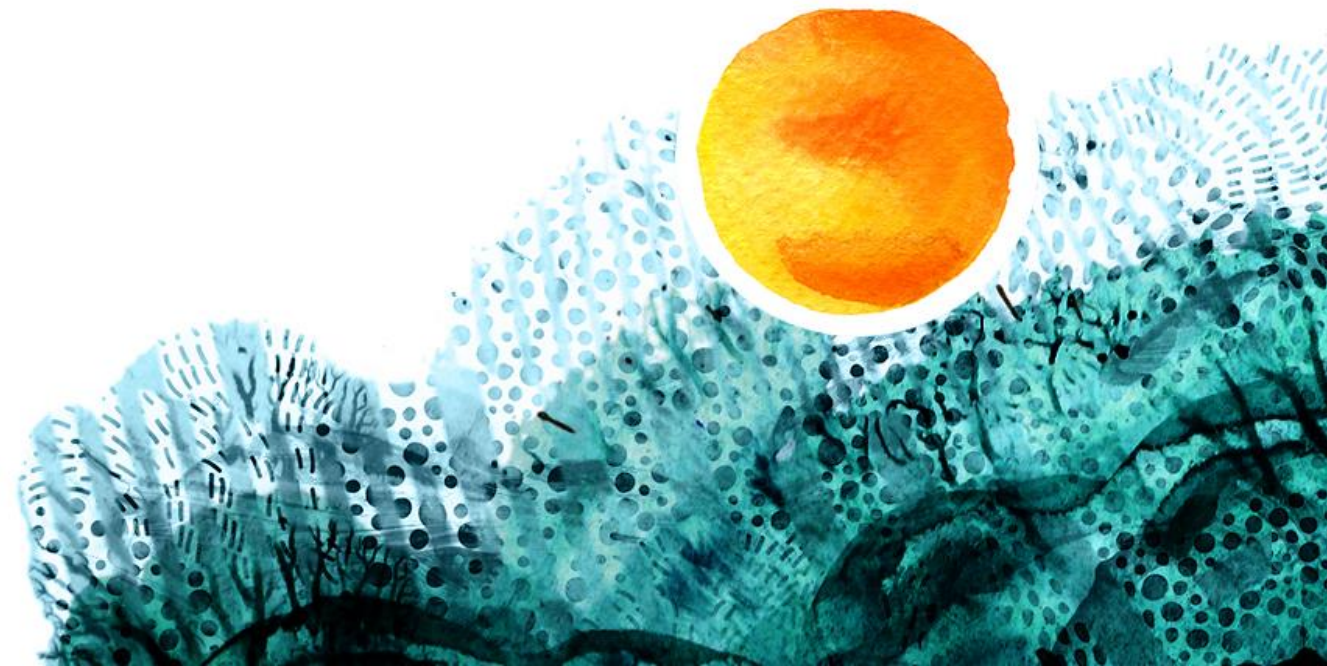


**Anna SCHEILI, Ph.D., Chemist**

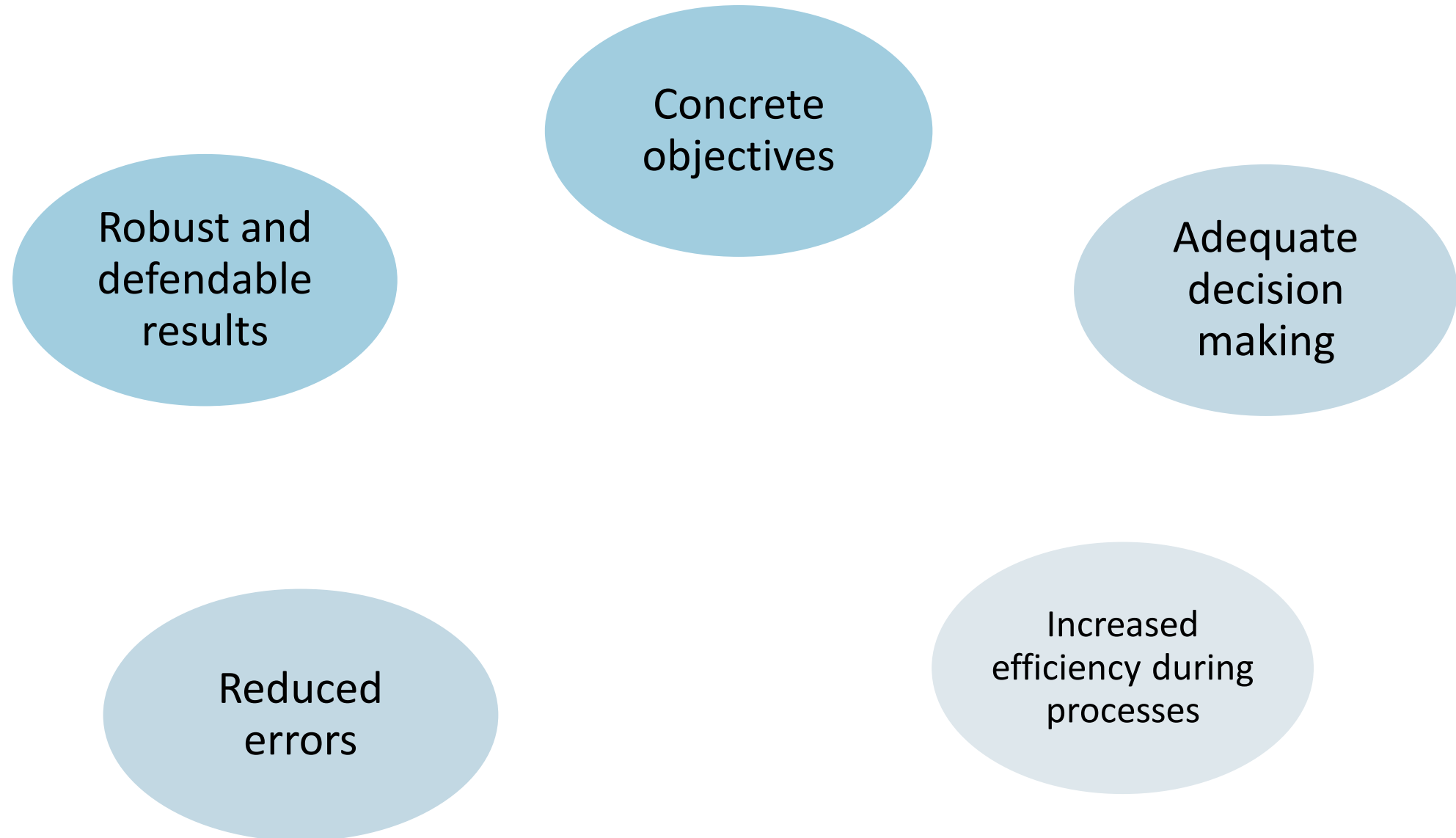
Project manager and expert consultant

WaterShed Monitoring

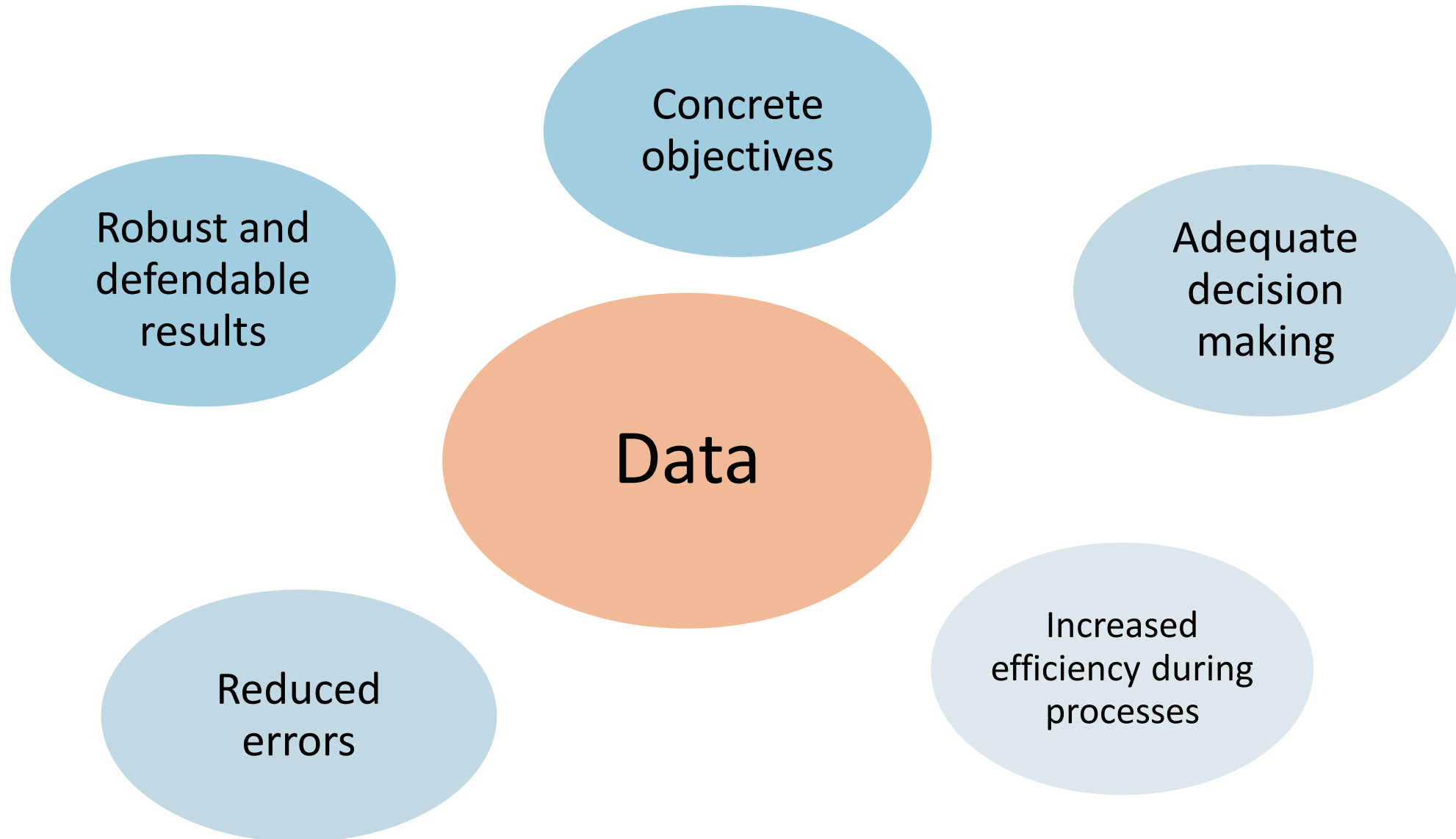
[ascheili@watershedmonitoring.com](mailto:ascheili@watershedmonitoring.com)



# *What is optimized water management?*



# *What is optimized water management?*



# Where is data in water management?

Legislative and policy framework • Public involvement and awareness  
Research, science and technology • Guidelines, standards and objectives

Monitoring • Management • Operations • Communication



Sources



Treatment



Distribution system



Drinking water

Adapted from CCME multibarrier approach

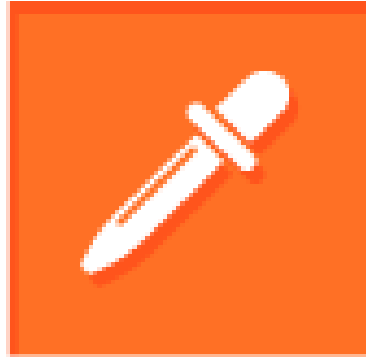
# Who are the data users?



# *Main challenges*



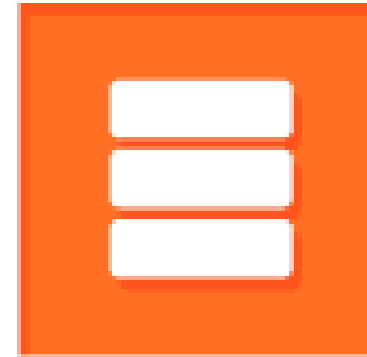
Planification,  
management and  
optimisation of  
water quality  
monitoring  
programs



Reliable sampling  
and analytical  
protocols and  
materials during  
field work and  
laboratory  
analyses



Data validation,  
interpretation and  
exploitation, data  
sharing

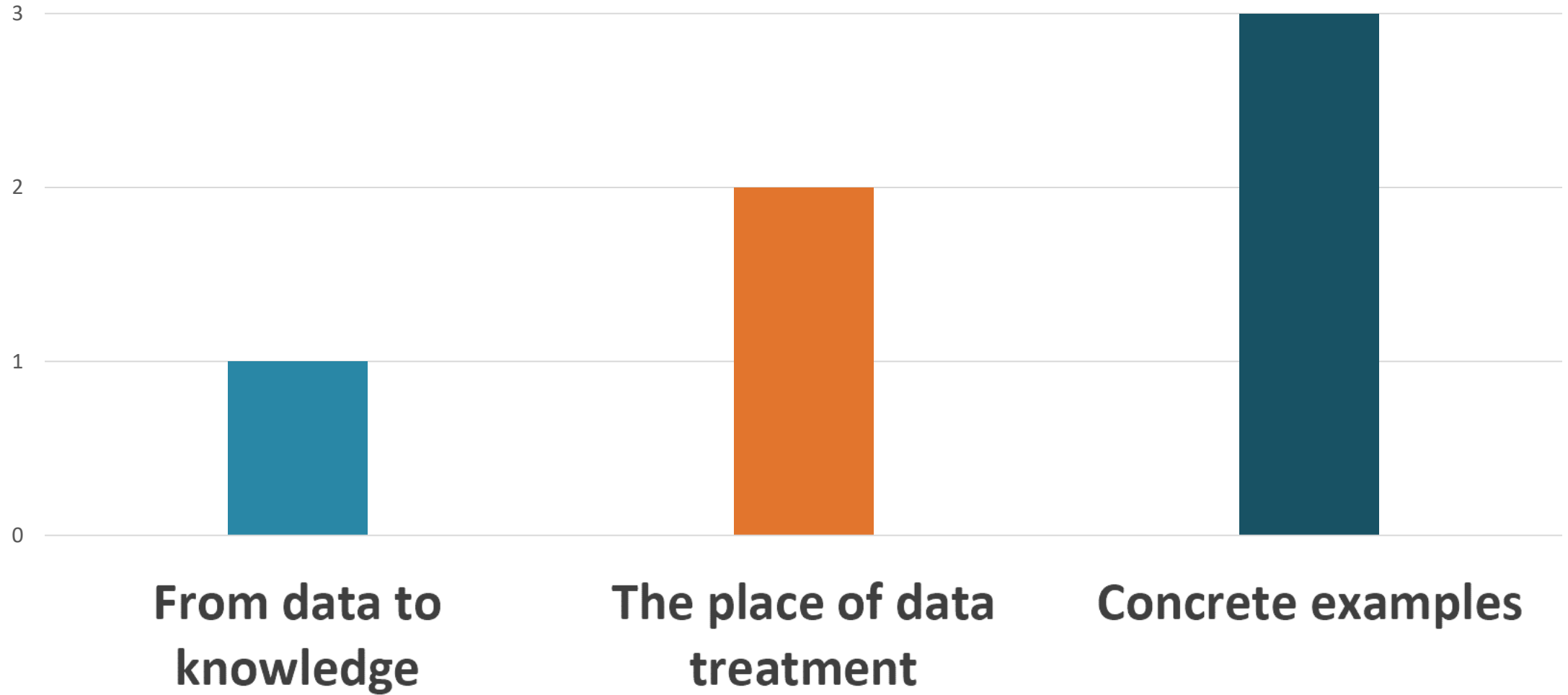


Integrated data  
management and  
inclusion of  
metadata

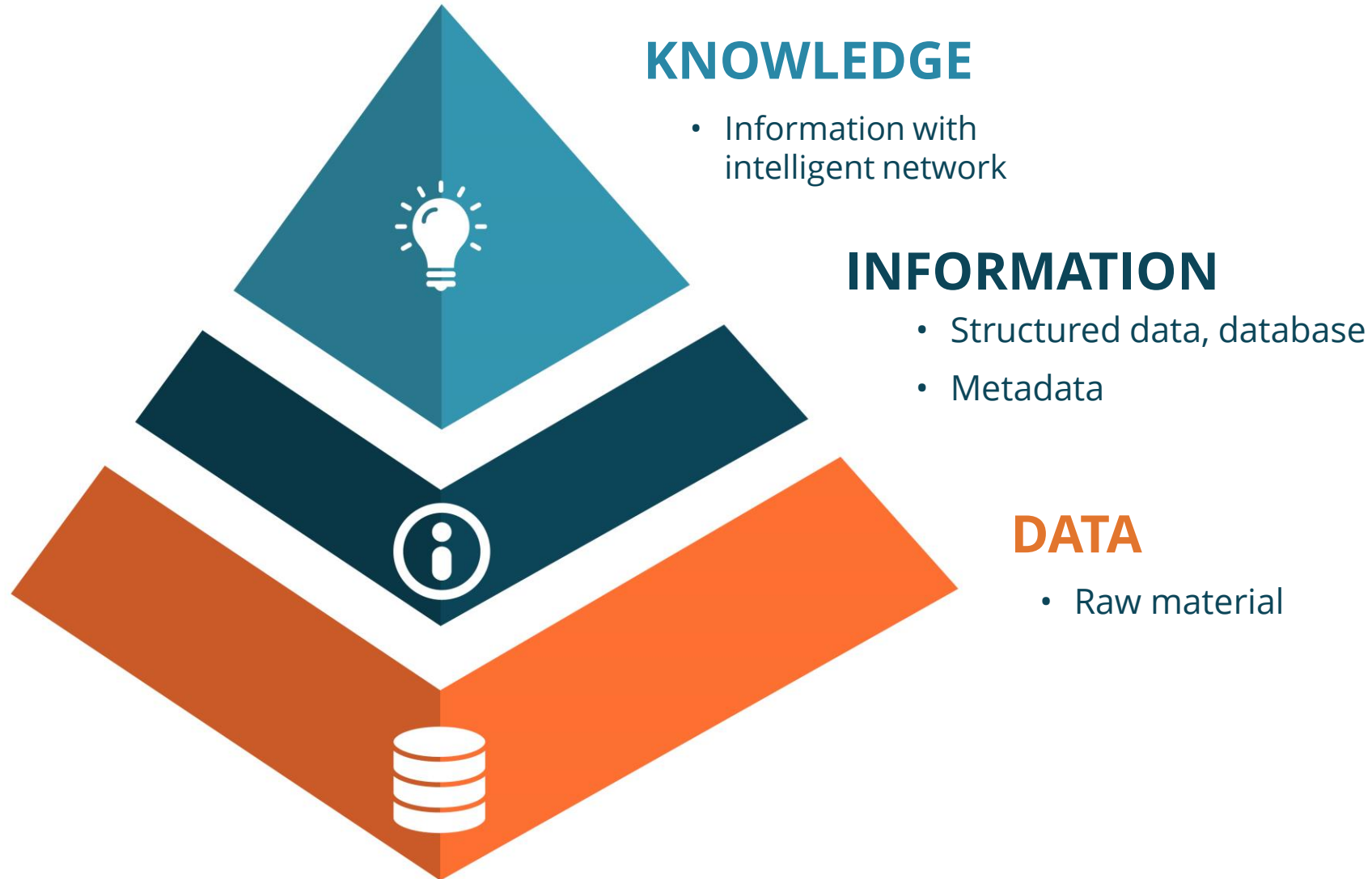


Data query and  
exportation

# Presentation plan



# *The pyramid of knowledge*





# *What is data?*

*« Things known or assumed as facts, making the basis of reasoning or calculation»*

*(Oxford Languages, 2022)*

PARAMETER	VALUE	UNIT
Fluorescent Dissolved Organic Matter (fDOM)	45.82	QSU
Fluorescent Dissolved Organic Matter (fDOM)	15.28	RFU
Chlorophyll a [479-61-8]	2.29	µg/l
Chlorophyll a [479-61-8]	0.55	RFU
	0.36	µg/l
Cyanobacteria (by fluorescence method)	0.3	RFU
Turbidity	7.7	NTU
pH	7.71	pH
	0.262945	g/l
Specific conductance	404.5	µS/cm
Dissolved oxygen	9.68	mg/l
Oxygen saturation	96.7	%

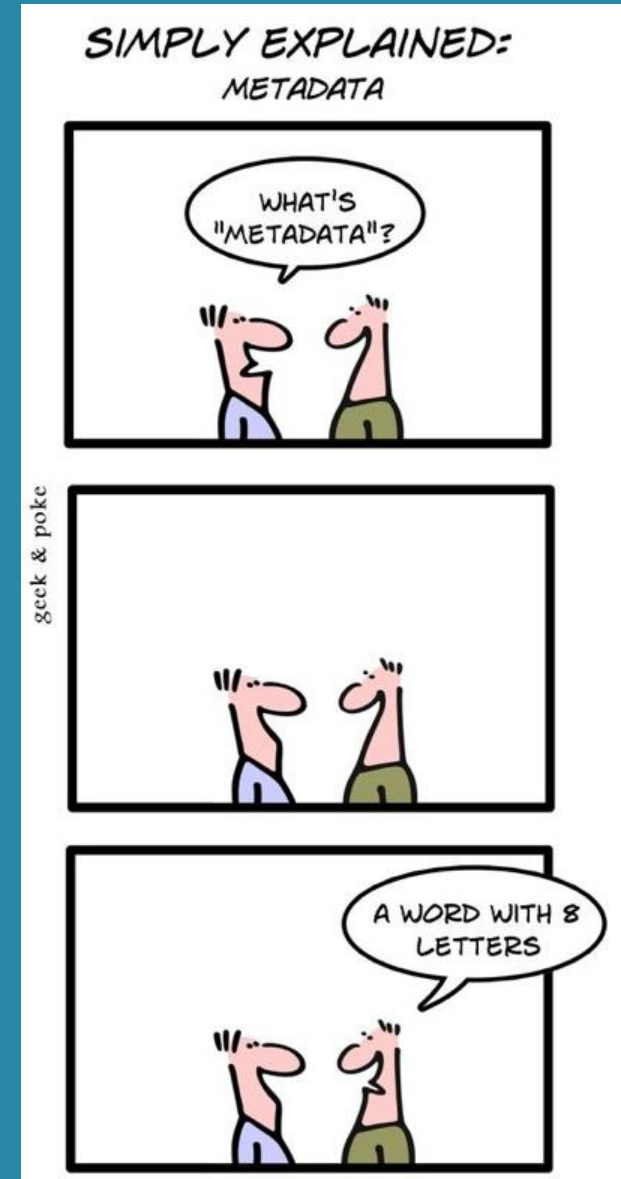
# What is metadata?

« Data that provides information about other data »

Techtarget, 2022

«Structured information that helps describe, explain, locate and use data »

Adproxima.fr, 2022



# Context #308776

Updated on : September 19, 2022 09:52

Edit

## OBSERVED BY

Laura and Thomas

## SAMPLING START

September 13, 2022 12:37

## SAMPLING END

September 13, 2022 12:41

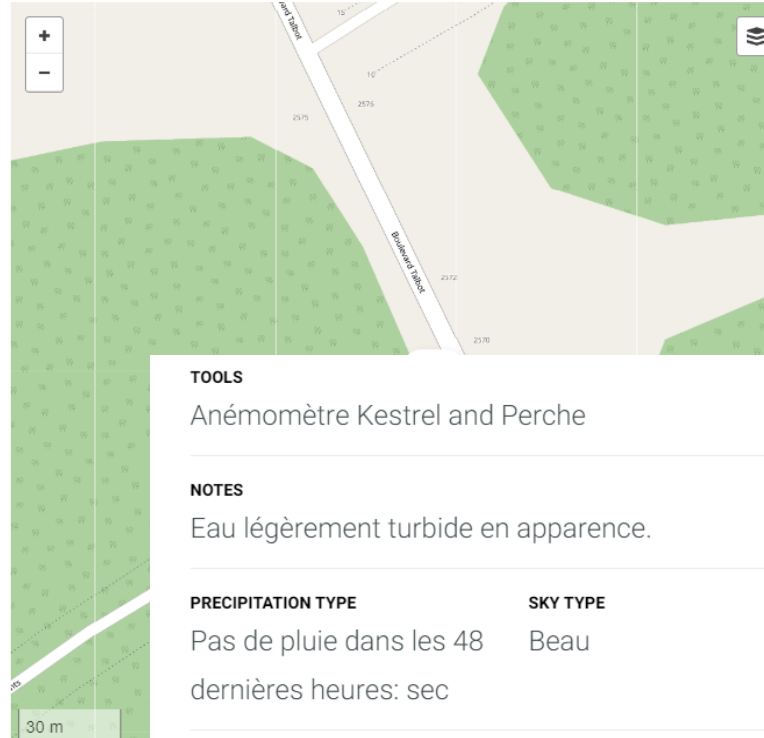
## SAMPLING PROTOCOLS

Proposition d'un programme de suivi de la qualité d'eau à l'échelle du bassin versant de la rivière Saint-Charles, Québec (Chapitre 4: Protocoles)

- [Chapitre\\_4.pdf](#) (20 Mar 14:15) **1.07 MB**
- [Behmel2010\\_Programme\\_suivi\\_BV\\_saint\\_charles.pdf](#) (20 Mar 14:14) **6.27 MB**

## SAMPLING OBJECTIVE

- Enquêtes approfondies sur zones de contaminations potentielles des cours d'eau



### TOOLS

Anémomètre Kestrel and Perche

### NOTES

Eau légèrement turbide en apparence.

### PRECIPITATION TYPE

Pas de pluie dans les 48 dernières heures: sec

### SKY TYPE

Beau

### WAVE TYPE

-

### WIND TYPE

- / -

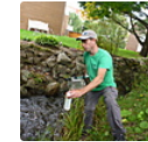
### HARMFUL ALGAL BLOOM

Unspecified

### HARMFUL ALGAL BLOOM TYPE

-

## IMAGES



## Probes and laboratories

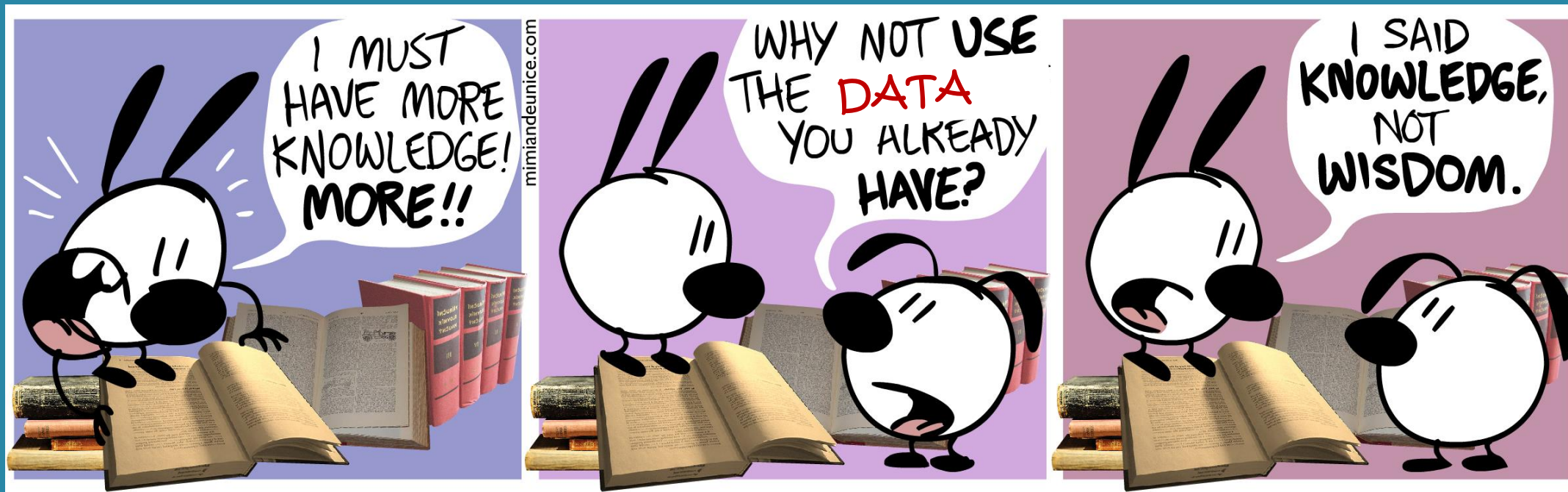
## Observations in context

Edit

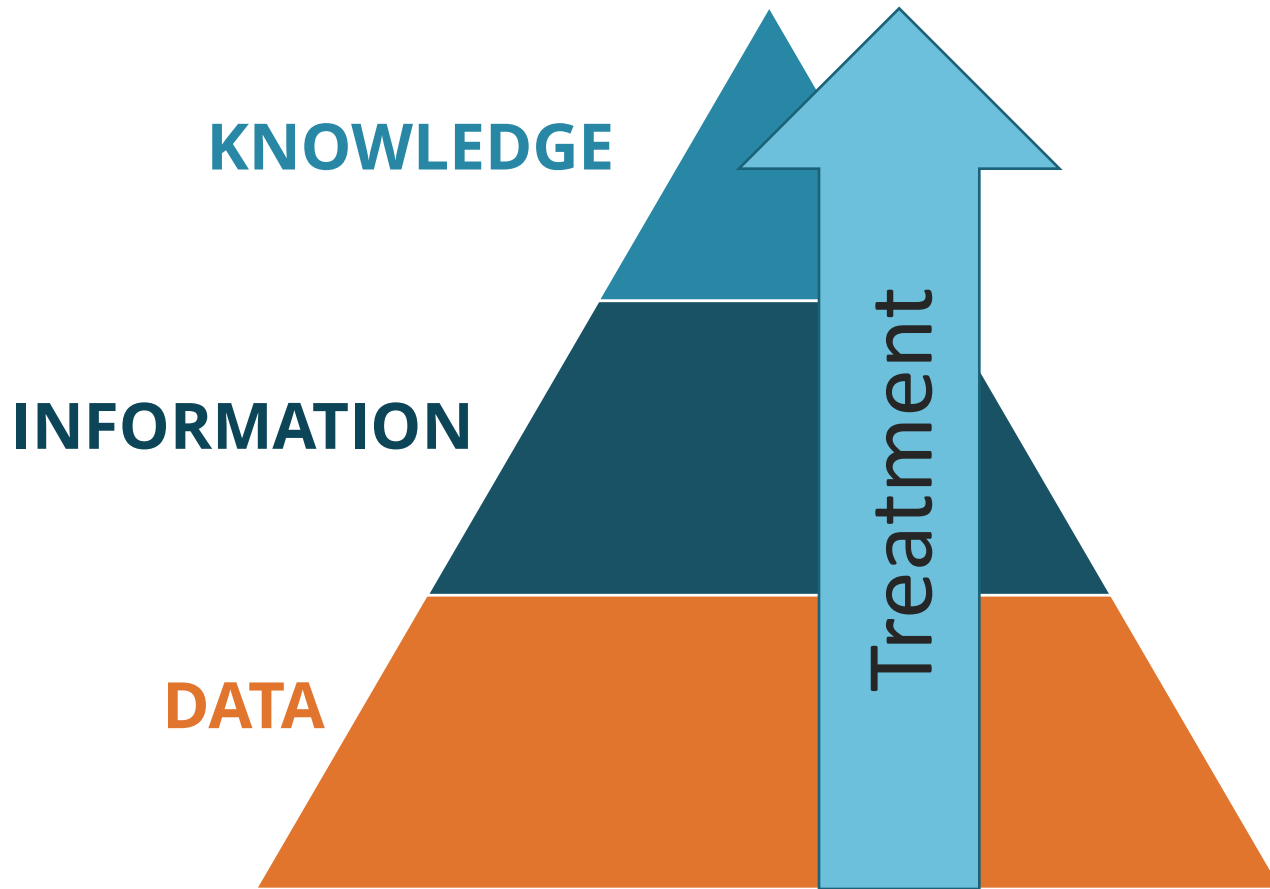
# *What is information?*

*« Data that have been recorded, classified, organized, related and interpreted within a framework so that meaning emerges »*

*Statistics Canada, 2017*

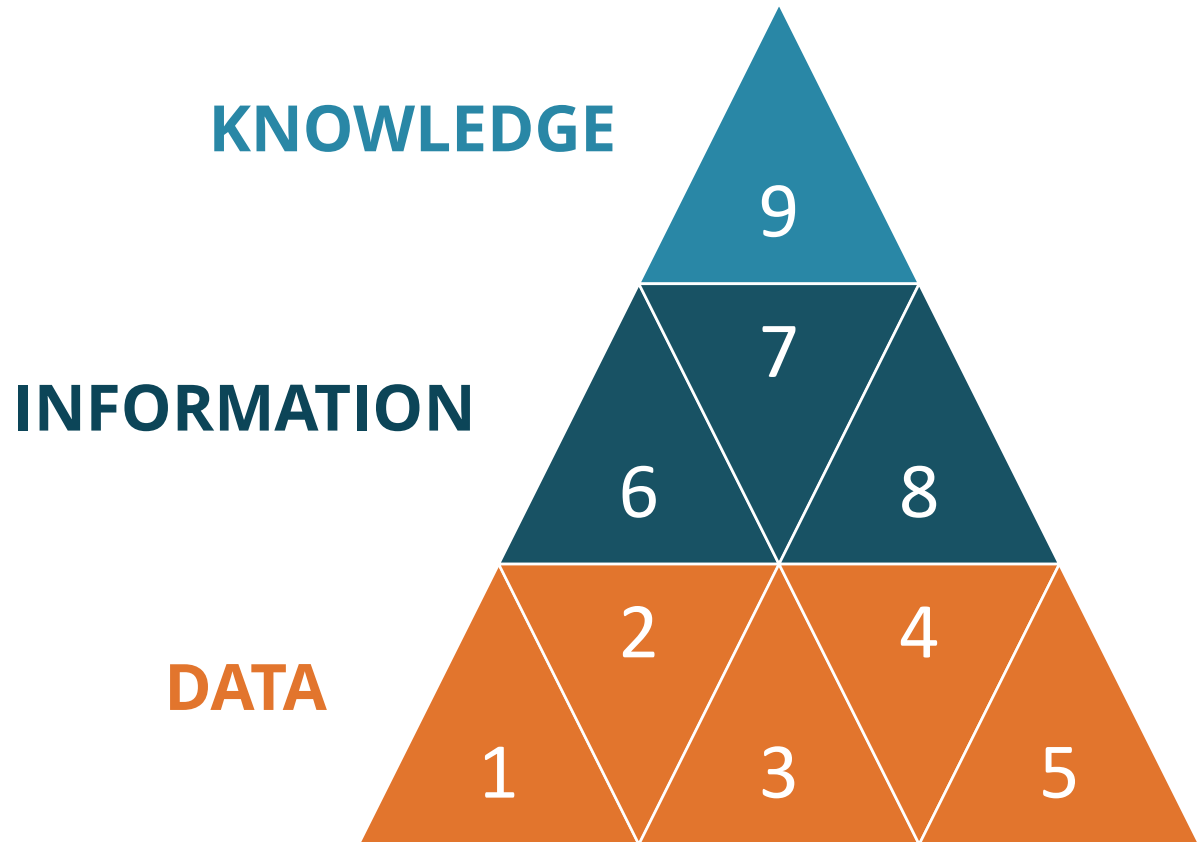


# Objectives



- Well defined objectives
  - ex. conformity, vulnerability or risk analysis, study of extremes, etc.
- Steps of data treatment and management are defined by the objectives
- Applied treatments define obtained knowledge
- A solid basis and structure
  - Robust data and relevant treatment

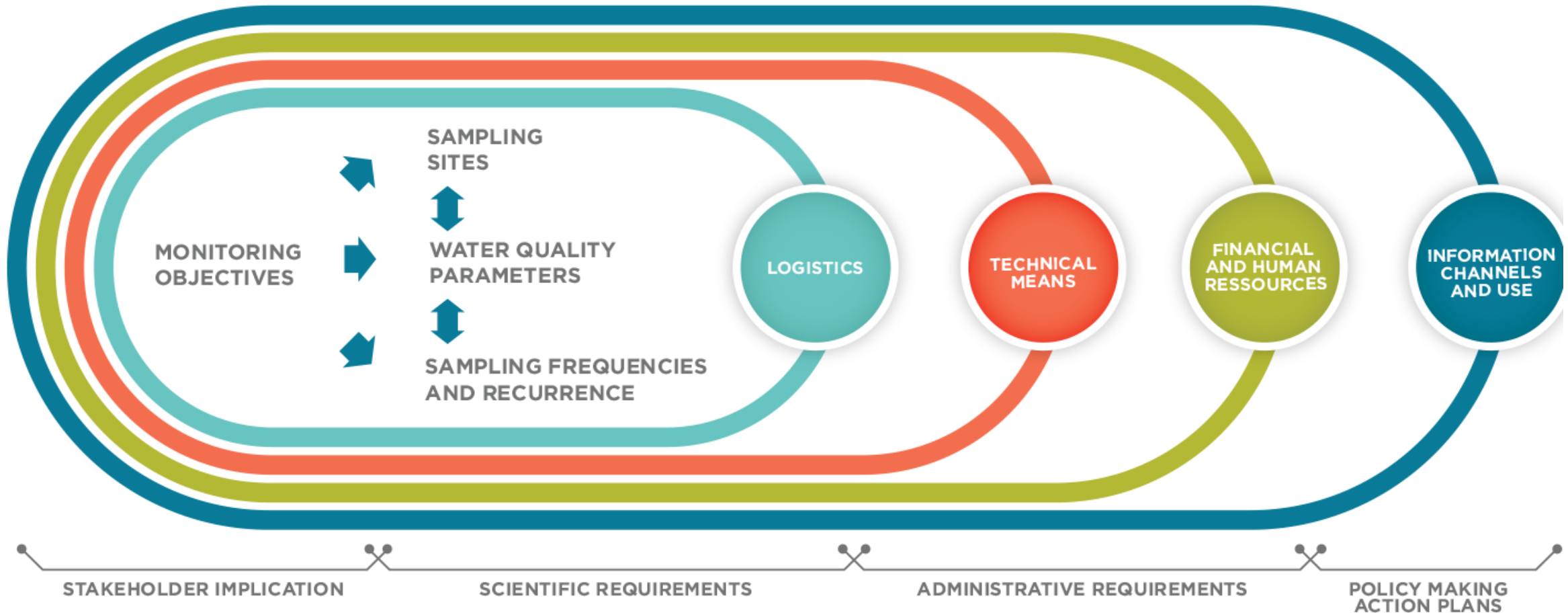
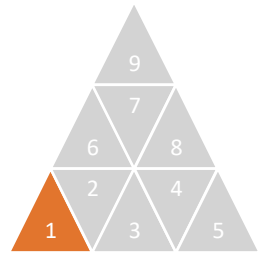
# Applied data treatment steps



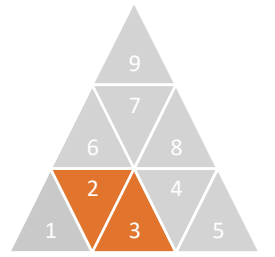
- 1 – Acquisition
- 2 – Cleaning
- 3 – Formatting
- 4 – Validation
- 5 – Qualification and correction
- 6 – Structuration
- 7 – Storing
- 8 – Linking data
- 9 – Analysis and exploitation



# Data acquisition



# Cleaning and formatting



- Cleaning
  - Identification of relevant data sets
  - Duplicates: keep or eliminate?
- Formatting
  - Crucial step for the rest of the process
  - Standardization of intrants
  - Consistency and exploitability
  - Need efficient collaboration and agreement between parties

## Probes and laboratories parameter

English \*

NAME \* Salicylic acid

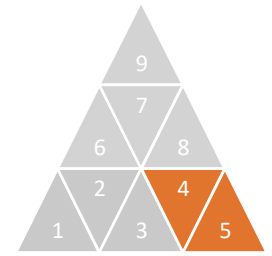
DESCRIPTION Salicylic acid is an organic compound with the formula  $\text{HOC}_6\text{H}_4\text{CO}_2\text{H}$ . A colorless, bitter-tasting solid, it is a precursor to and a metabolite of

French

NAME Acide Salicylique

DESCRIPTION

# Validation, qualification and correction



## ■ Validation

- Establishing validation rules according to the objectives (ex: units, parameter names, etc.)
- Possible automatization of strict rules based on
- Semi-automatized validation: initial filter, than individual treatment, final approval by a human

## ■ Qualification and correction

- Identification of incorrect data
  - Definition of rules (ex: conformity according to values)
  - Searching for contextual causes (ex: calibration, protocols, staff, etc.)
  - Definition of specific treatment: conservation, exclusion of calculations, elimination, correction
- Data distorsion (ex: extremes)
- Artificial intelligence

### ERROR

Unknown parameters : Lab\_ID

Unknown measuring units : 22H1479-01

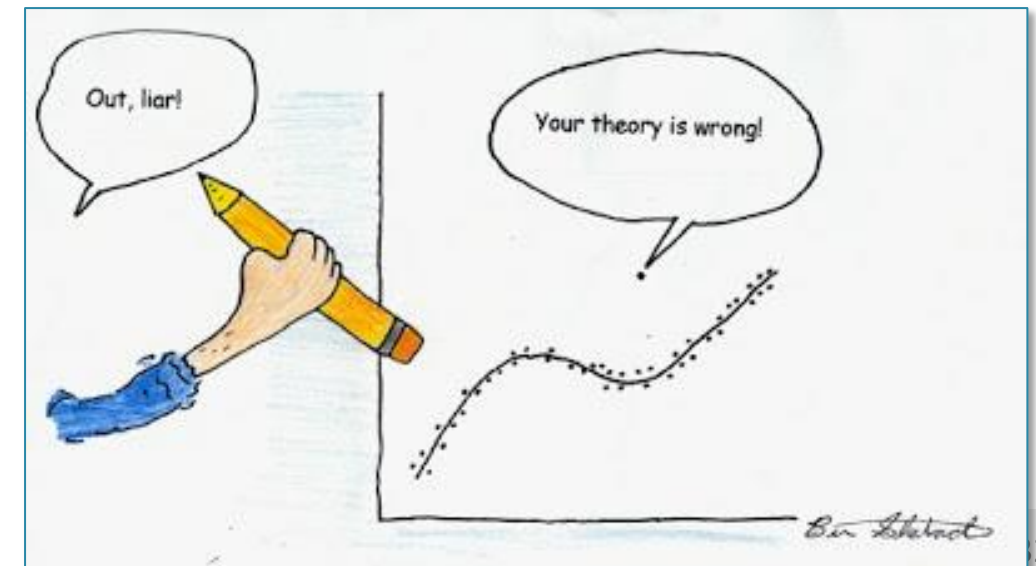
### Import logs

---

Sampling point name not found at line 2

Sampling point name not found at line 3

Context not found - Pumphouse Raw Water Line / 2022-05-23 at line 4

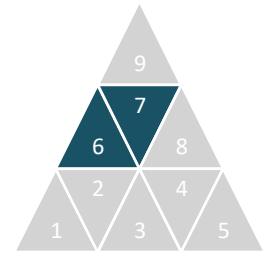


# *What is a database?*

*« A database is a data structure that stores organized information »*

*<http://techterms.com/definition/database>, 2022*

# Storage and structuration

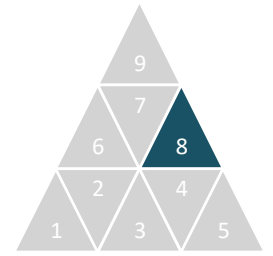


## ■ Relational databases

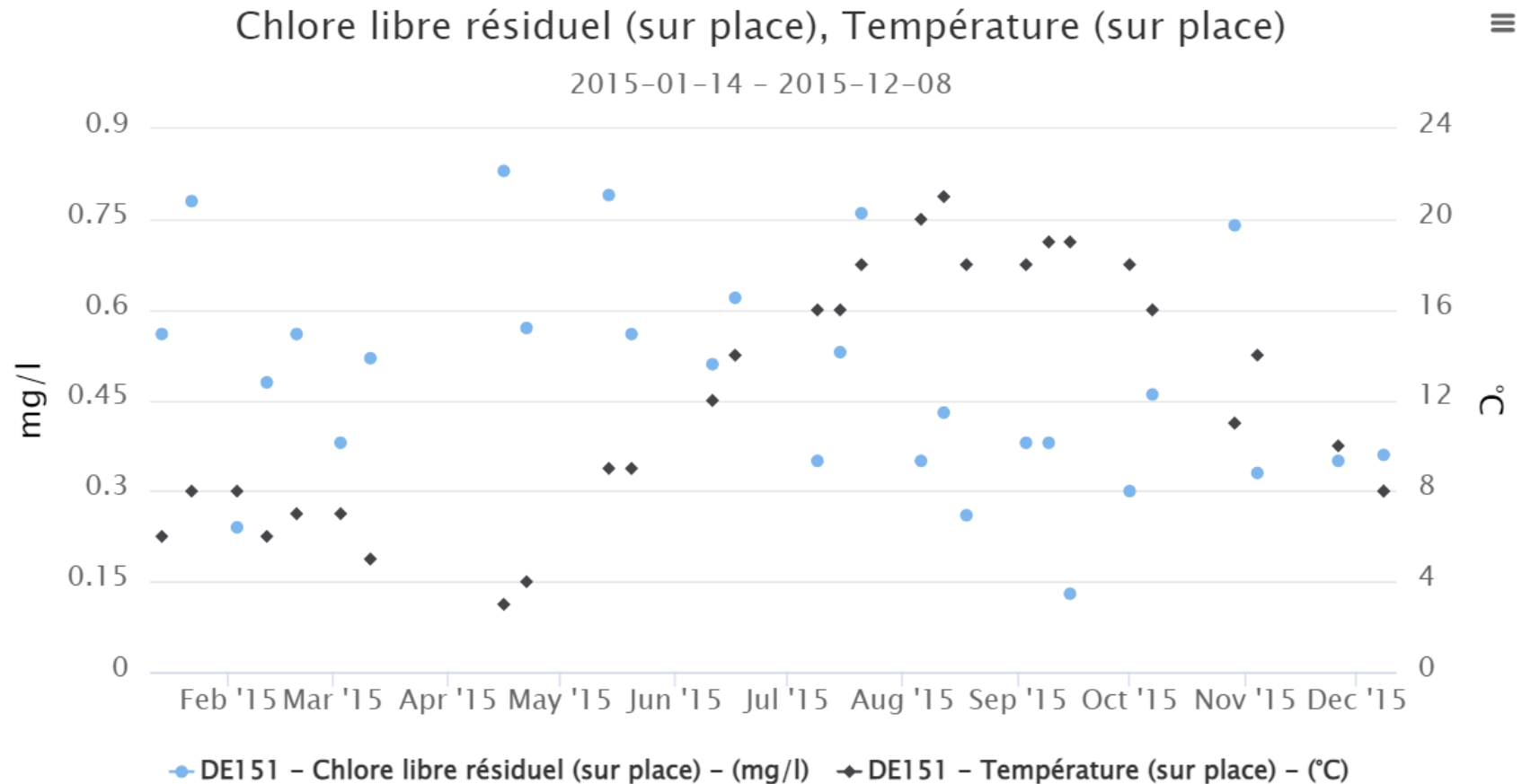
- Access, update, search information based on the relationships of data stored in different tables
- Run queries that involve multiple databases
- Store text or numeric data, but also other data types: geographical information, sound clips, pictures, videos, etc.



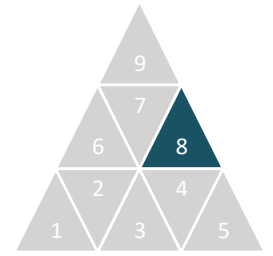
# Linking data and creating knowledge



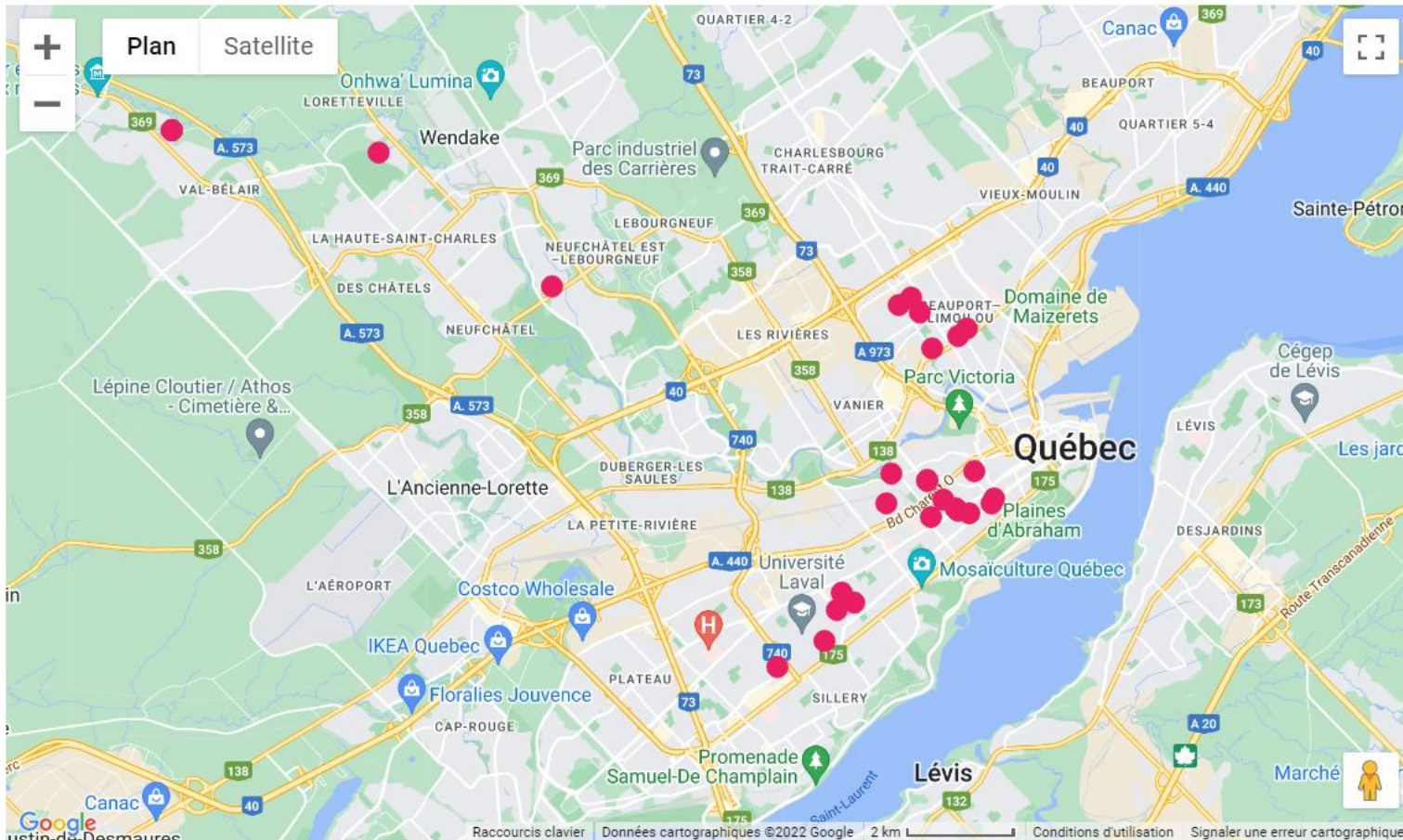
- Linking same type of data
  - Ex: water quality in the distribution system: Free chlorine vs Temperature



# Linking data and creating knowledge



- Linking different type of data, from different sources
  - Ex: type of consumer complaints VS localisation VS drinking water quality



BY DATE

FROM

TO

PROBLEM TYPE

▼

PROBLEM FREQUENCY

▼

LEAD INLET?

▼

CONNECTION CROSSED?

▼

SIGNS OF CORROSION?

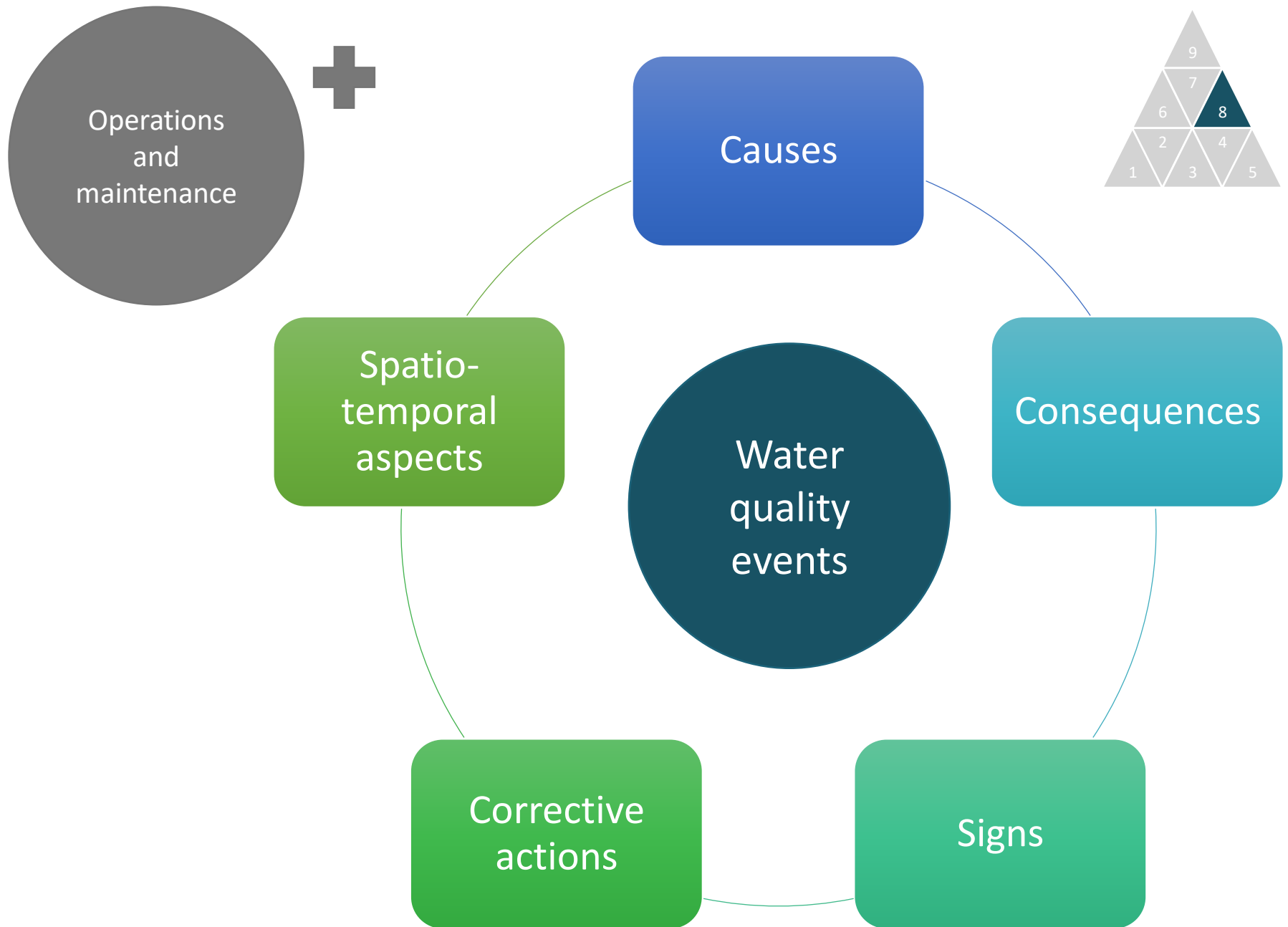
▼

NEIGHBOR IMPACTED?

▼

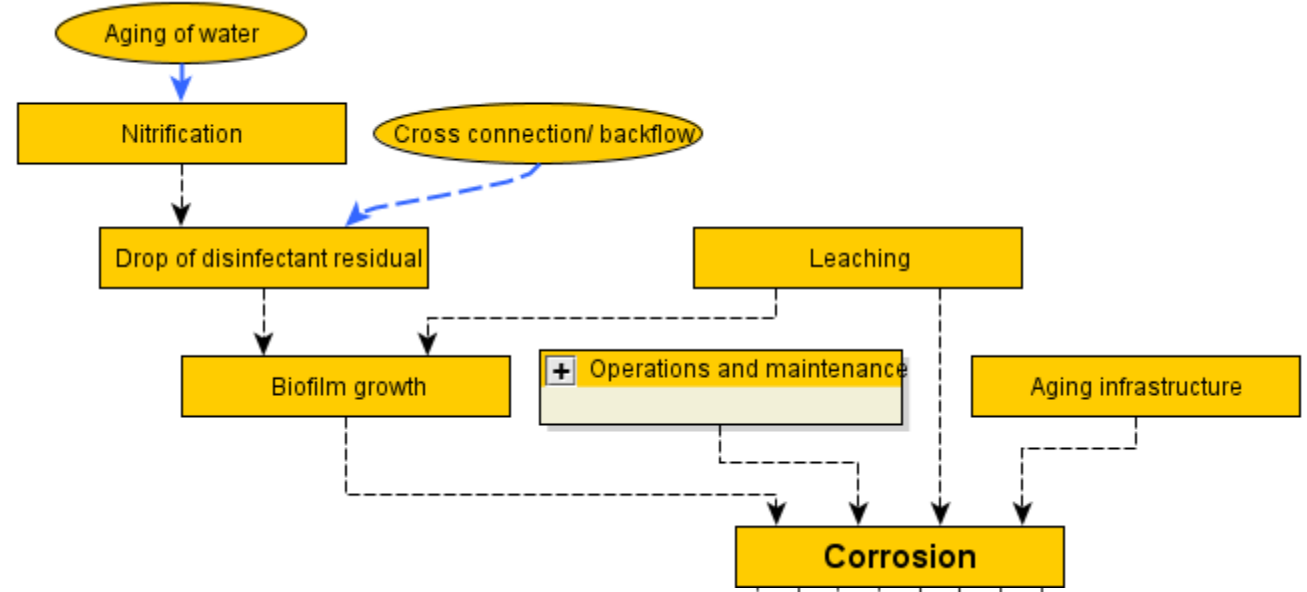
■ **Complex and multi-level linking**

- Ex: Understanding water quality events in the distribution system

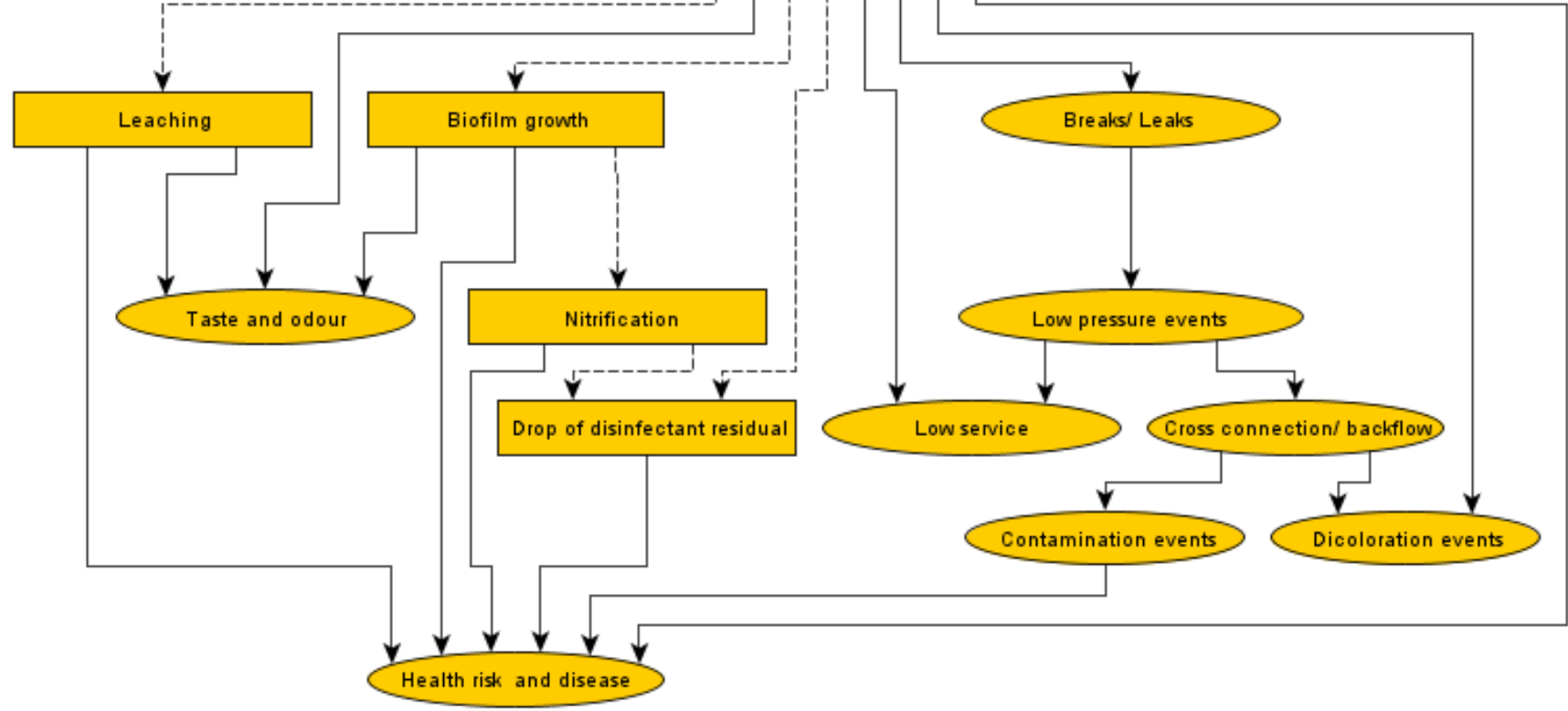




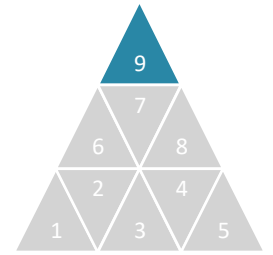
## Causes



## Consequences



# Analysis and exploitation



- Communication, visualisation and analysis
  - Graphs, maps, indices (ex: water quality index)



## Water Quality Index

6341 Rue Ducolombier, Québec, QC G2K 2M2, Canada

*Based on water quality data available on 3 novembre 2022.*

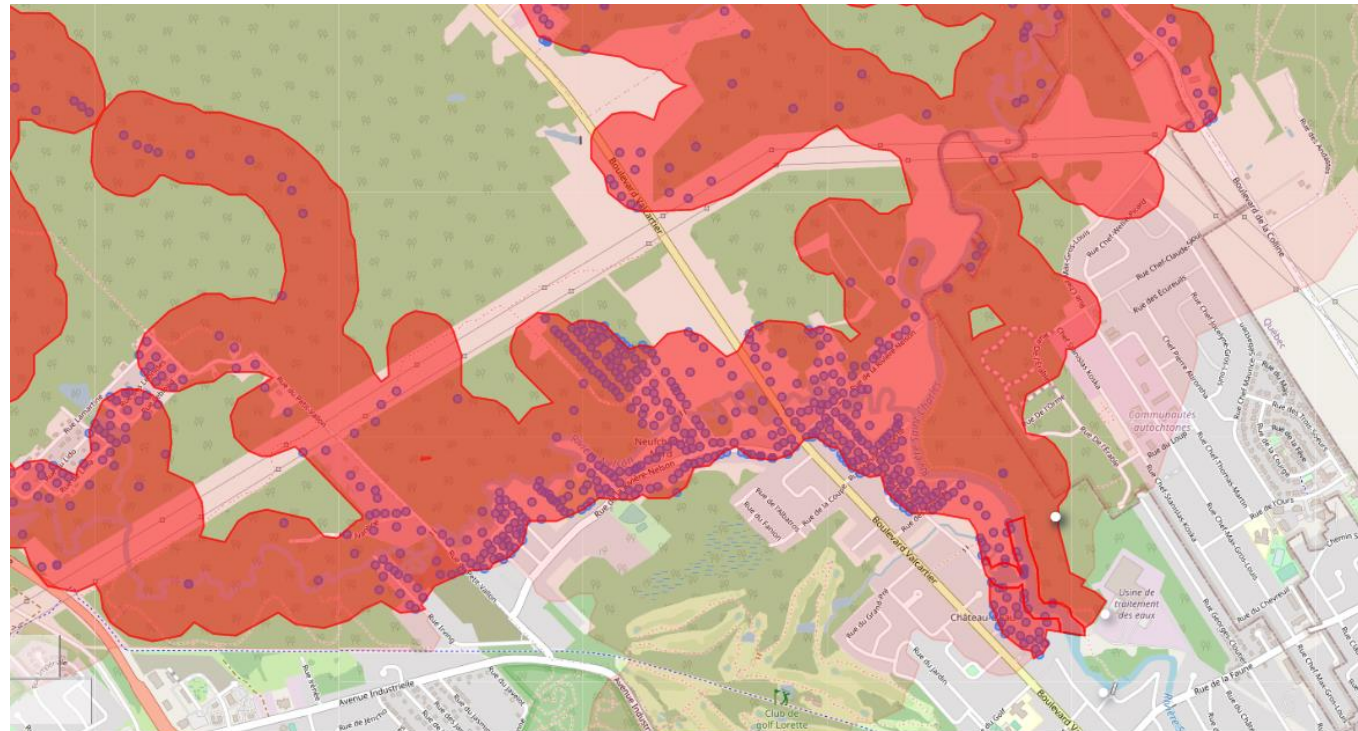
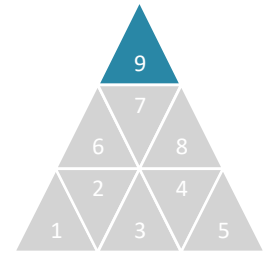
99 / 100

Report Water Quality Issue

The water quality index is calculated with parameters measured regularly by the City of Québec in your area, which characterize the physicochemical and microbiological quality of the water. The value of the index can vary widely depending on the season and the City's areas, however, the water quality remains in line with current standards. If this is not the case, the City issues opinions on the situation in your area.

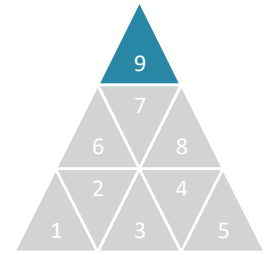
# Analysis and exploitation

- Decision making and actions
  - Ex: Vulnerability and risk analysis, protection plans



24 743 activités anthropiques répertoriées

# Analysis and exploitation



- Decision making and actions
  - Vulnerability and risk analysis, protection plans
  - Alerts



**ALERT TYPE \***  
Avis d'ébullition préventif

**ALERT ZONES \***  
Zone 31  
Select all None

**CAUSE OF THE ALERT \***  
construction

**EFFECTIVE FROM \***  
20 octobre 2022

**EFFECTIVE UNTIL**  
22 octobre 2022

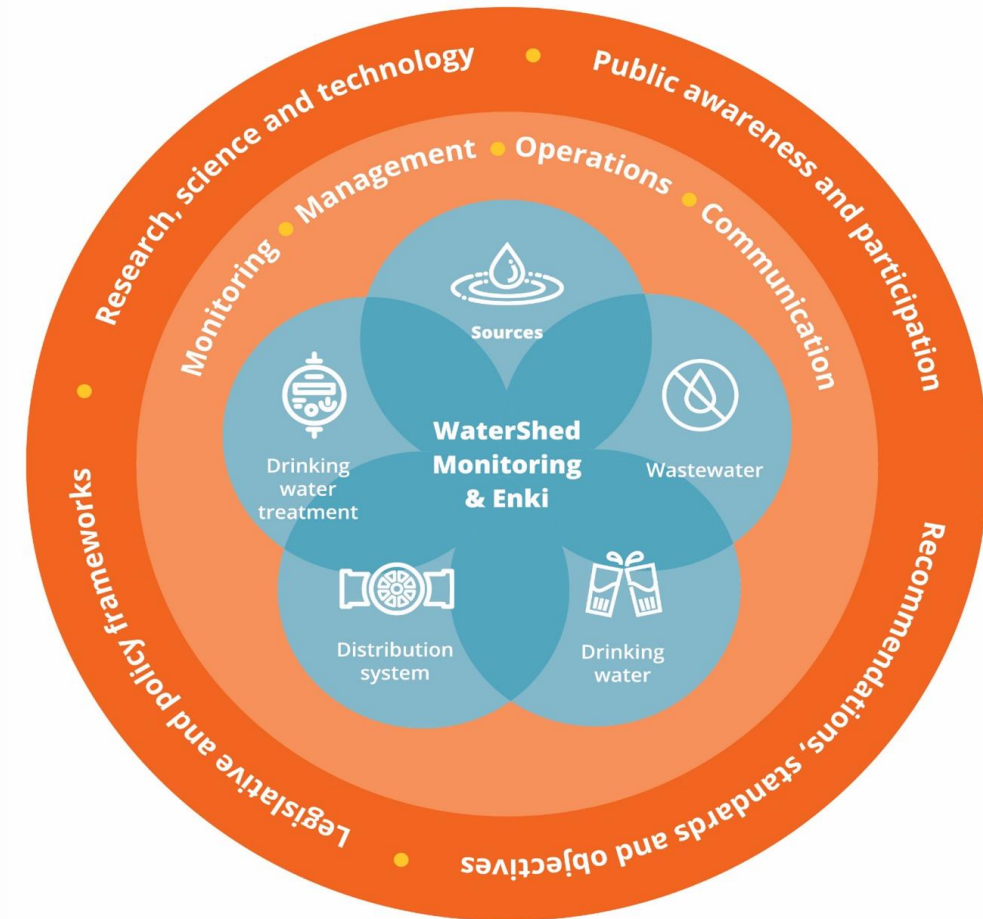
**RECOMMENDED ACTION**  
Lors d'un avis d'ébullition, il s'agit de faire bouillir l'eau destinée à la consommation pendant au moins une minute. L'utilisation de l'eau bouillie est recommandée pour la préparation des aliments et le brossage des dents. Les autres usages de l'eau potable

# *Conclusions*

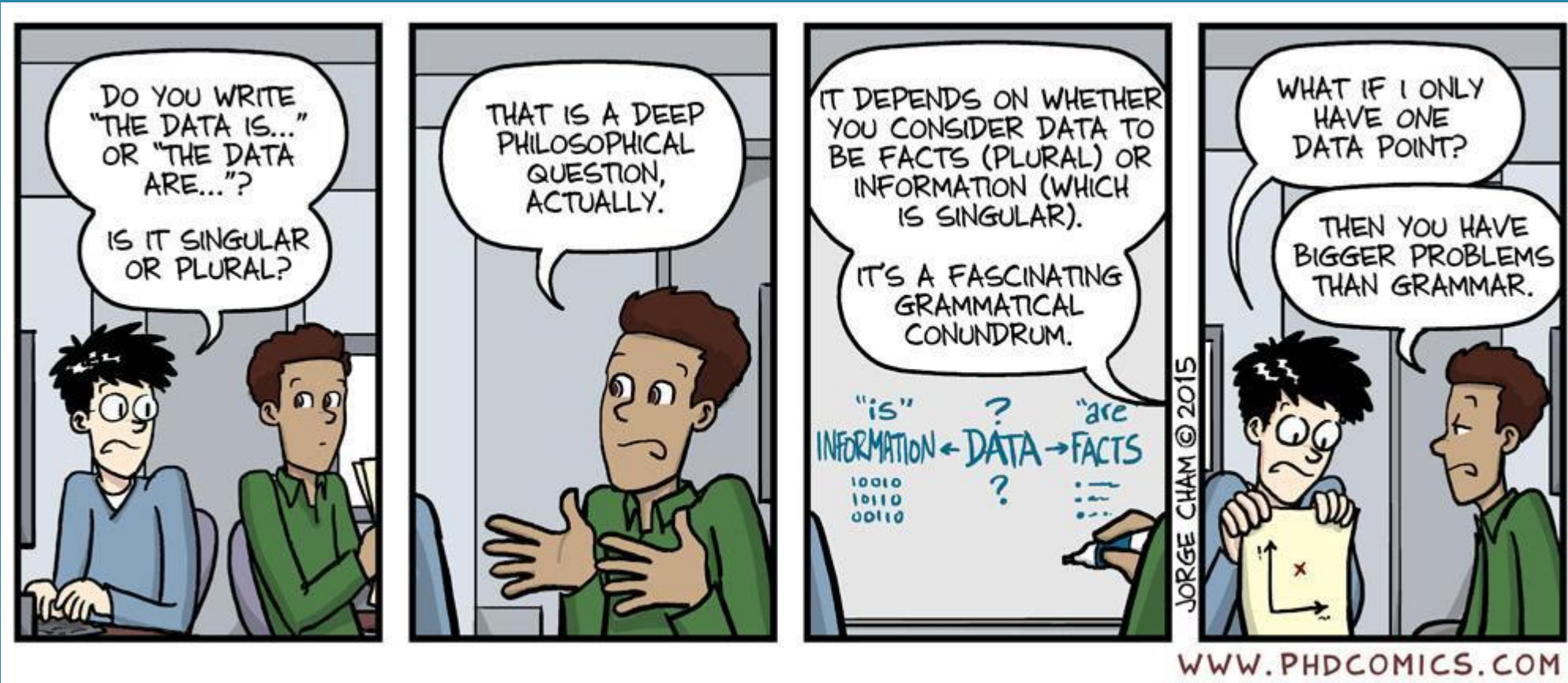
“Data are just summaries of thousands of stories – tell a few of those stories to help make the data meaningful.”

*Chip and Dan Heath, authors of “Made to Stick” and “Switch.”*

# Where can adequate data treatment take us?







Anna Scheili, PhD  
ascheili@watershedmonitoring.com