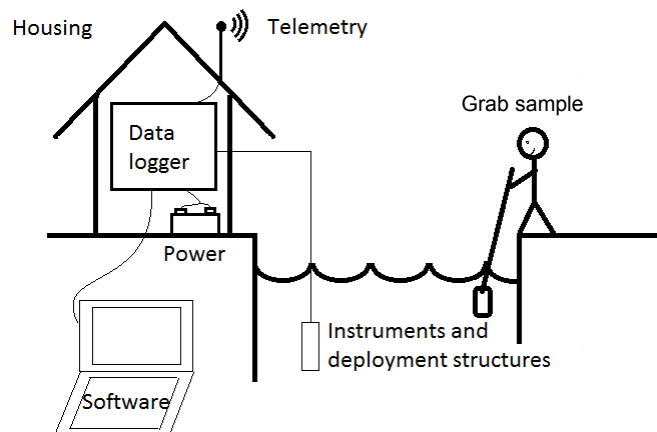


**datEAUbase: A powerful  
large capacity database  
for raw and validated  
water quality data  
with emphasis on their metadata**

Q. Plana, T. Kraft, J. Alferes, T. Maruéjols,  
P.A. Vanrolleghem  
*National Monitoring Conference, Tampa Bay  
May 5, 2016*



**Automated monitoring stations + Sampling/lab**



# Automated monitoring stations + Sampling/lab



© Plana et al., 2016

3

# Automated monitoring stations

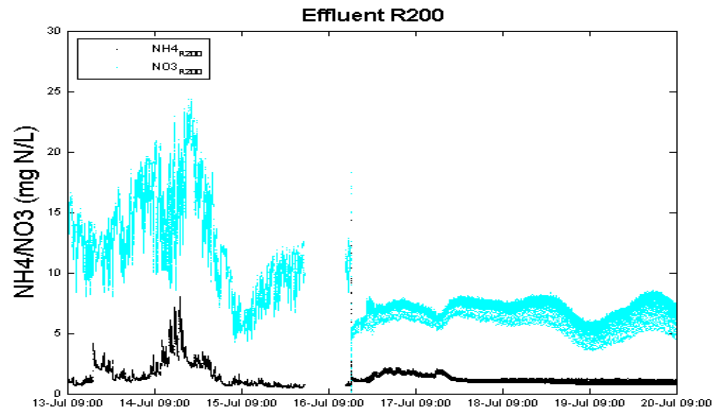
High frequency monitoring

Large amounts of relevant data

© Plana et al., 2016

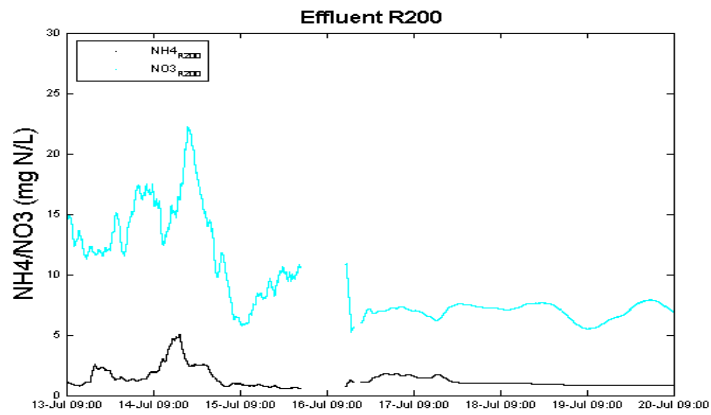
4

## Raw and filtered data



5

## Raw and filtered data



6

## Importance of the metadata (data about data)

- What does this single value mean?

Row	Conduct	Temp	DO	pH	Turbidity	DO	Temp	Conduct	Temp	DO	pH	Turbidity	DO
2	TimeUTC	Status	Quality	Value	Status	Quality	Value	Status	Quality	Value	Status	Quality	Value
3	42142.81901	78	0	3.000007227	78	0	9	1	0	7.28615197	1	0	11.96138902
4	42142.81927	78	0	3.000007227	78	0	9	1	0	7.28615197	1	0	11.96138902
5	42142.81953	78	0	3.000007227	78	0	9	1	0	7.279264039	1	0	11.96138902
6	42142.81979	78	0	3.000007227	78	0	9	1	0	7.279264039	1	0	11.96138902
7	42142.81944	78	0	3.000007227	78	0	9	132	0	11.80200904	1	0	7.223818213
8	42142.81950	78	0	3.000007227	78	0	9	1	0	7.279264039	1	0	11.96138902
9	42142.81956	78	0	3.000007227	78	0	9	1	0	7.279264039	1	0	11.96138902
10	42142.81962	78	0	3.000007227	78	0	9	1	0	7.279264039	1	0	11.96138902
11	42142.81968	78	0	3.000007227	78	0	9	1	0	7.280518148	1	0	11.96138902
12	42142.81974	78	0	3.000007227	78	0	9	1	0	7.280518148	1	0	11.96138902
13	42142.81979	78	0	3.000007227	78	0	9	1	0	7.280518148	1	0	11.96138902
14	42142.81986	78	0	3.000007227	78	0	9	1	0	7.280518148	1	0	11.96138902
15	42142.81991	78	0	3.000007227	78	0	9	1	0	7.281792329	1	0	11.96138902
16	42142.81997	78	0	3.000007227	78	0	9	1	0	7.281792329	1	0	11.96138902
17	42142.82002	78	0	3.000007227	78	0	9	1	0	7.282251068	1	0	11.96138902
18	42142.82008	78	0	3.000007227	78	0	9	1	0	7.282251068	1	0	11.96138902
19	42142.82014	78	0	3.000007227	78	0	9	132	0	11.80200904	1	0	7.223818213
20	42142.82020	78	0	3.000007227	78	0	9	1	0	7.282251068	1	0	11.96138902
21	42142.82026	78	0	3.000007227	78	0	9	1	0	7.282251068	1	0	11.96138902
22	42142.82031	78	0	3.000007227	78	0	9	1	0	7.282251068	1	0	11.96138902
23	42142.82037	78	0	3.000007227	78	0	9	1	0	7.282251068	1	0	11.96138902
24	42142.82043	78	0	3.000007227	78	0	9	1	0	7.282251068	1	0	11.96138902
25	42142.82049	78	0	3.000007227	78	0	9	1	0	7.282251068	1	0	11.96138902
26	42142.82054	78	0	3.000007227	78	0	9	1	0	7.282251068	1	0	11.96138902
27	42142.82060	78	0	3.000007227	78	0	9	1	0	7.282251068	1	0	11.96138902
28	42142.82066	78	0	3.000007227	78	0	9	1	0	7.282251068	1	0	11.96138902
29	42142.82072	78	0	3.000007227	78	0	9	1	0	7.282251068	1	0	11.96138902
30	42142.82078	78	0	3.000007227	78	0	9	1	0	7.282251068	1	0	11.96138902
31	42142.82083	78	0	3.000007227	78	0	9	132	0	11.80200904	1	0	7.223818213
32	42142.82089	78	0	3.000007227	78	0	9	1	0	7.295082569	1	0	11.96138902
33	42142.82095	78	0	3.000007227	78	0	9	1	0	7.295082569	1	0	11.96138902
34	42142.82101	78	0	3.000007227	78	0	9	1	0	7.295082569	1	0	11.96138902
35	42142.82106	78	0	3.000007227	78	0	9	1	0	7.295082569	1	0	11.96138902
36	42142.82112	78	0	3.000007227	78	0	9	1	0	7.295082569	1	0	11.96138902
37	42142.82118	78	0	3.000007227	78	0	9	1	0	7.295082569	1	0	11.96138902
38	42142.82124	78	0	3.000007227	78	0	9	1	0	7.295082569	1	0	11.96138902
39	42142.82130	78	0	3.000007227	78	0	9	1	0	7.295082569	1	0	11.96138902
40	42142.82136	78	0	3.000007227	78	0	9	1	0	7.295082569	1	0	11.96138902

7.2811

## Importance of the metadata

- What does this single value mean?

What has been measured?

When has the value been measured?

Where has the value been measured?

7.2811

Why has the value been measured?

Who has collected the value?

How has the value been measured?

## Importance of the metadata

- Collected data are **only useful** when:
  - They are well-documented
  - Their quality is assured

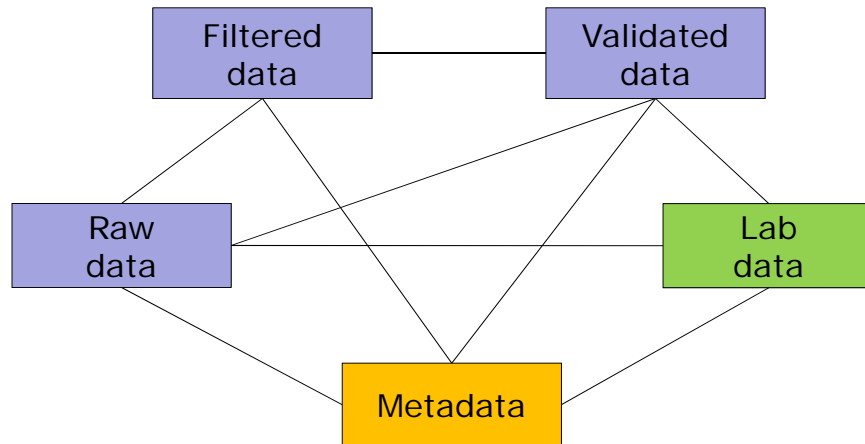


- Data storage task
- Data validation task

## Contents

- Introduction
- **Types of data**
- Challenges
- datEAUbase scheme
- datEAUbase design
- Examples of data
- Conclusions

## Types of data



## Contents

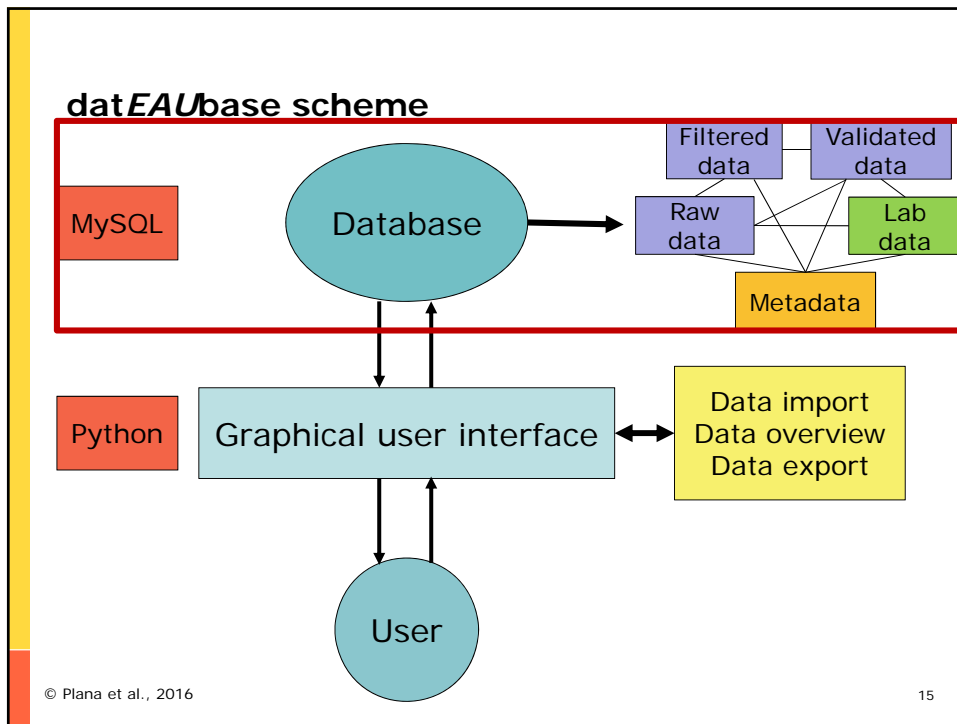
- Introduction
- Types of data
- **Challenges**
- *datEAUbase* scheme
- *datEAUbase* design
- Examples of data
- Conclusions

## Data management challenges

- Variability of raw data formats
- Databases grow continuously and monitoring programs change
  - Database must be adapted
- Need for high-performance storage and data access
- Personnel that is collecting and managing data is changing over time
  - Inconsistencies

## Contents

- Introduction
- Types of data
- Challenges
- **datEAUbase scheme**
- datEAUbase design
- Examples of data
- Conclusions



**Contents**

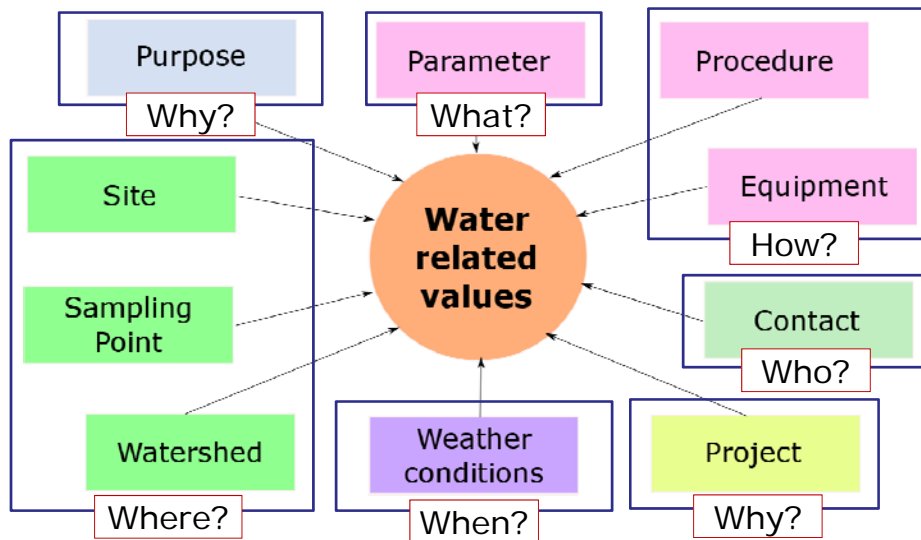
- Introduction
- Types of data
- Challenges
- datEAUbase scheme
- **datEAUbase design**
- Examples of data
- Conclusions

© Plana et al., 2016

16



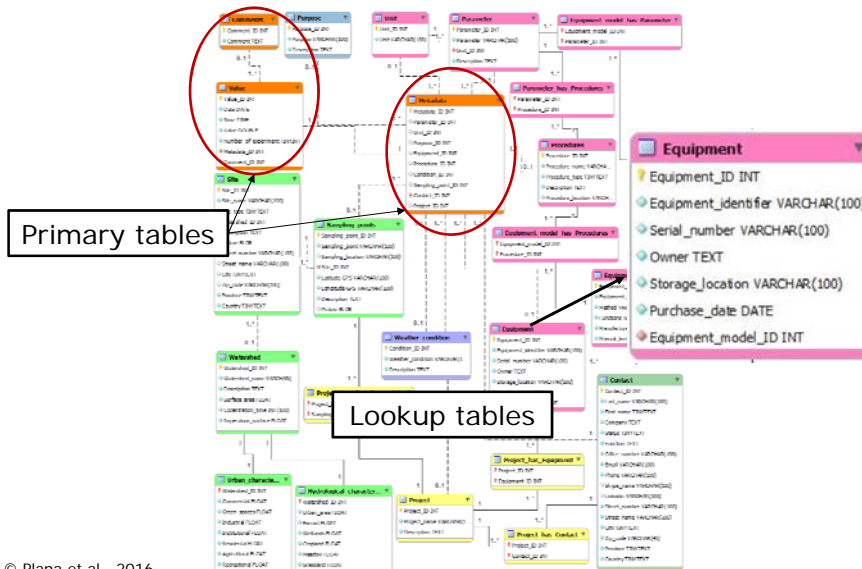
## datEAUbase design



© Plana et al., 2016

17

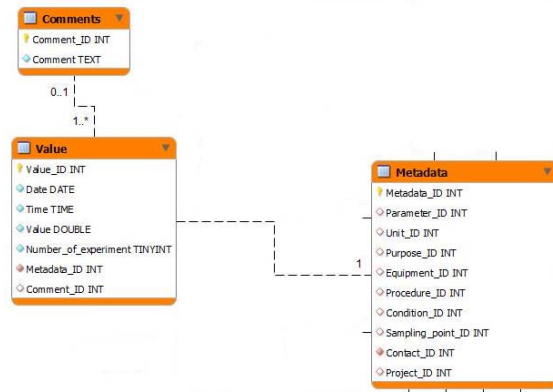
## datEAUbase design



© Plana et al., 2016

18

## datEAUbase design – Primary tables



© Plana et al., 2016

19

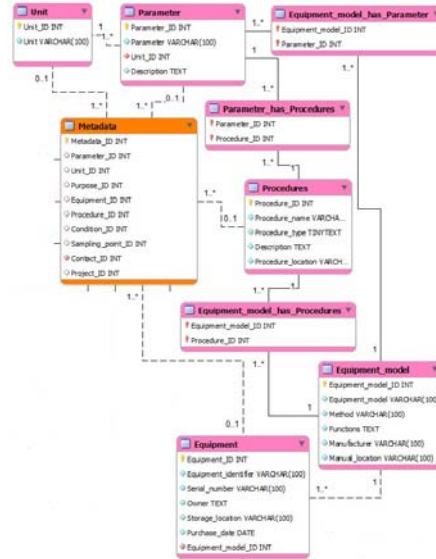
## datEAUbase design – Primary tables

- **Date:** May 29, 2015
- **Time:** 15:36:02 GMT
- **Value:** 7.32
- **Parameter:** pH
- **Units:** [-]
- **Sensor:** pH\_003
- **Conditions:** dry weather
- **Purpose:** calibration
- **Procedure:** ISO-15839
- **Site:** Grandes-Piles F/AL
- **Sampling point:** inlet
- **Responsible:** Plana
- **Project:** monEAU
- **Comment:** Unsuccessful calibration

© Plana et al., 2016

20

## datEAUbase design – Equipment & procedures



© Plana et al., 2016

21

## datEAUbase design – Equipment & procedures

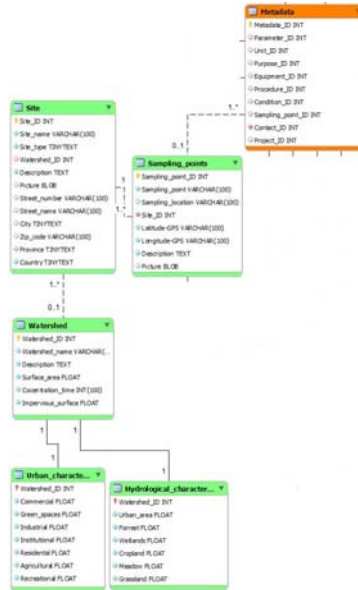
- **Parameter:** pH
- **Units:** [-]
- **Sensor:** pH\_003
- **Brand:** Hach
- **Model:** DPD1P1
- **Serial number:** 2659777
- **Principle:** Differential of electrical potential
- **Current location:** Grandes-Piles F/AL
- **Manual:** pHD sc Digital Differential Sensor. User manual
- **Manual location:** <http://modeleAU...>



© Plana et al., 2016

22

## datEAUbase design – Sampling points



© Plana et al., 2016

23

## datEAUbase design – Sampling points

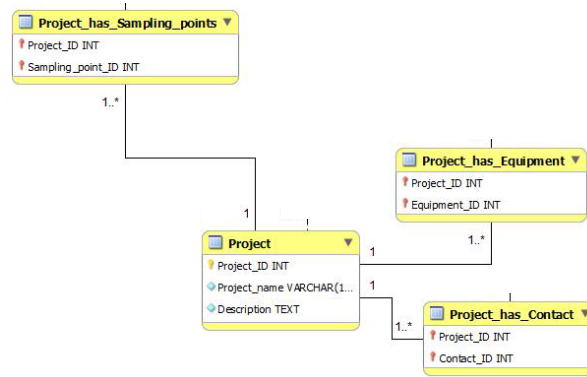
- **Site:** Grandes-Piles F/AL
- **Sampling point:** Inlet
- **Address:** 267-303 5e Av.,  
Grandes-Piles, GOX 1H0, QC,  
Canada
- **Coordinates:** 46° 41' 04" N  
72° 42' 59" W
- **Watershed:** Saint-Maurice  
river
- **Surface area:** 43 300 km<sup>2</sup>
- **Concentration time:** 2 days
- **Impervious surface:** 4 %
- **Urban characteristics:**
  - 54.25 % of green spaces
  - 2.25 % of industrial area
  - 13.5 % of residential area
  - 22 % of agricultural area
  - 8 % of recreational area



© Plana et al., 2016

24

## datEAUbase design - Projects



© Plana et al., 2016

25

## datEAUbase design - Projects

- **Project name:** monEAU
- **Description:** AMS to study the water quality
- **Sampling point:** inlet in Grandes-Piles F/AL
- **Equipment:**
  - conductivity\_001
  - pH\_003
  - ammolyser\_001
- **Personnel involved:**
  - Alferes
  - Plana
  - Vanrollegem

© Plana et al., 2016

26

## datEAUbase design – Other lookup tables

- **Contact information:**

- **First name:** Queralt
- **Last name:** Plana
- **Company:** Université Laval
- **Status:** PhD student
- **Address:** 1065, avenue de la Médecine, room PLT-2954, Québec, G1V 0A6, QC, Canada
- **E-mail:** [queralt.plana.1@ulaval.ca](mailto:queralt.plana.1@ulaval.ca)
- **Phone:** +1 418 656 2131, ext. 8730

## datEAUbase design – Other lookup tables

- **Weather conditions:**

- **Condition:** wet day
- **Description:** rainfall of more than 3 mm/d

- **Purpose:**

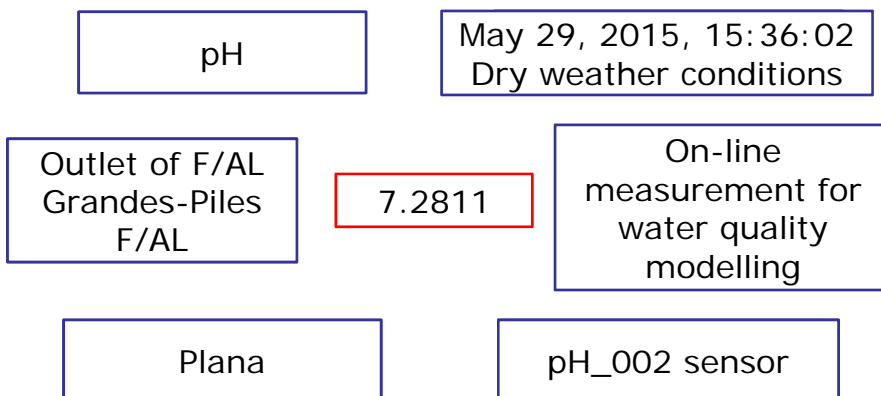
- **Purpose:** sensor validation
- **Description:** routine sensor validation activity for verification of proper operation

## Contents

- Introduction
- Types of data
- Challenges
- datEAUbase scheme
- datEAUbase design
- **Examples of data**
- Conclusions

## Example

- Understandable value



## Contents

- Introduction
- Types of data
- Challenges
- datEAUbase scheme
- datEAUbase design
- Examples of data
- **Conclusions**

## Conclusions

- datEAUbase is a powerful tool to store data using a single consistent format, giving quality to the content
- Any environmental parameter can be stored into it
- It offers flexibility and it can be modified and adapted for future studies
- Its design is explicitly providing relevant metadata information to the measured values
  - The metadata is fundamental to understand the measured values for use in further studies
- It combines raw, filtered, validated, lab and **metadata**



## Acknowledgments



*Canada Research Chair  
in Water Quality Modeling*



Fondation canadienne pour l'innovation  
Canada Foundation for Innovation



UNIVERSITÉ  
LAVAL

