



Bases de données relationnelles

Intérêt, stockage, structuration, accès

Généralités & la base 'data LEESU'



- 1) Organisation de données 'classique' historique.
- 2) Organisation avec une Base de Données ?
- 3) Principe
- 4) La base 'data' au LEESU 
- 5) Les moyens d'accès
- 6) Types et flux de données
- 7) Gestion des données & Ouverture vers l'exterieur
- 8) le lien base 'data' <-> base 'matériels' 

1) organisation 'historique' des données

classement hiérarchique de fichiers hétérogènes

The image displays a file explorer window on the left, showing a hierarchical structure of data files. The main window shows three open files:

- 150610-PointB-M.txt - Bloc-notes**: A text file containing metadata for a data file, including the path, filename, cadence d'enregistrement (300 s), and various parameters (VOIE: 0 Prof, VOIE: 1 Temp, VOIE: 2 Condu, VOIE: 3 Oxyg, VOIE: 4 Fluor, VOIE: 5 Vbatt).
- NutrientsData_20181224_003200.dat - Bloc-notes**: A data file containing a list of parameters: "TOA5", "CR800Series", "CR800", "28032", "CR800.Std.26", "CPU:SolSCompansQuali", "TIMESTAMP", "RECORD", "Battv", "Temp", "Conduct", "pH", "Turb", "NoDePluie", "Nt", "TS", "RN", "Volts", "degre", "us/cm", "Units", "FNU", "Smp", "Smp", "Smp", "Smp", "Smp", "Smp", "Smp", "Smp".
- Microsoft Excel - 2016-08-16 stationLacESP.csv**: An Excel spreadsheet showing a table of data with columns B through H. The table contains numerical values for various parameters over time.

	B	C	D	E	F	G	H
214	datcnv_ox_ta	=	yes				
215	file_type	=	ascii				
216							
217	0	-0.15	-0.148	21.7352	9.41591	107.13045	-0.472
218	1	-0.15	-0.148	21.6822	9.69014	110.13743	-2.279
219	2	-0.15	-0.148	21.6591	9.70328	110.24429	-2.28
220	3	-0.15	-0.148	21.576	9.62979	109.22668	3.462
221	4	-0.15	-0.148	21.4997	9.64154	109.19831	-1.424
222	5	-0.15	-0.148	21.4517	9.68707	109.61847	4.465

1) organisation actuelle des données

Organisation hiérarchique de données fichiers

Inconvénients:

- hétérogénéité des formats: texte, csv, Excel, typé etc.
- format des données spécifiques de la source: entête, lignes/colonnes.
- Il faut connaitre tous les chemins, nommer le fichiers.
- pas de relation entre les données
- séries décomposées, de taille variable.
- gestion des droits d'accès lourd . gestion écriture/lecture simultanée...

2) organisation avec une BD

Organisation et stockage de données sur des Bases de données relationnelles



- les données sont centralisées (Bases) et sur un même niveau hiérarchique
- séparation des données et de leurs attributs.
- enrichissement avec des informations descriptives: Métadonnées
- structuration des données => 'Tables' stockées dans une base (ou des bases)
- relations entre les groupes (coté relationnelle)

id_data	mesure	date_heure	b	c	i	commentaire
1179695	1	2016-01-26 02:16:09	1.597	NULL	NULL	NULL
1179696	1	2016-01-26 02:21:09	1.596	NULL	NULL	NULL
1179697	1	2016-01-26 02:26:09	1.597	NULL	NULL	NULL
1179698	1	2016-01-26 02:31:09	1.597	NULL	NULL	NULL
1179699	1	2016-01-26 02:36:09	1.596	NULL	NULL	NULL
1179700	1	2016-01-26 02:41:09	1.595	NULL	NULL	NULL
1179701	1	2016-01-26 02:46:09	1.594	NULL	NULL	NULL
1179702	1	2016-01-26 02:51:09	1.593	NULL	NULL	NULL
.....	.	-----	NULL	NULL	NULL

Table: t_data

id_mesure	nom	parametre_instrument_FK	projet_FK	zone_FK
1	Profondeur pt B M	11	0	2
2	Temperature pt B M	2	0	2
3	Conductivité pt B M	3	0	2
4	O2 pt B M	4	0	2
5	Fluorescence pt B M	5	0	2
6	Phycocyanine pt B M	6	0	2
7	pH pt B M	7	0	2
8	Tension Bat pt B M	8	0	2
9	Profondeur pt A M	9	0	1
10	Temperature pt A M	10	0	1

Table: t_mesures

2) organisation avec une BD

Organisation et stockage de données sur des Bases de données relationnelles

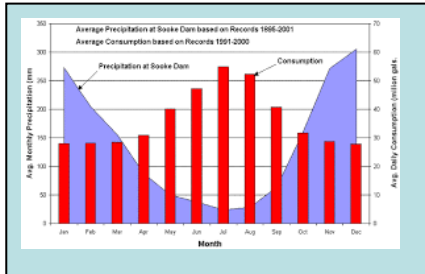
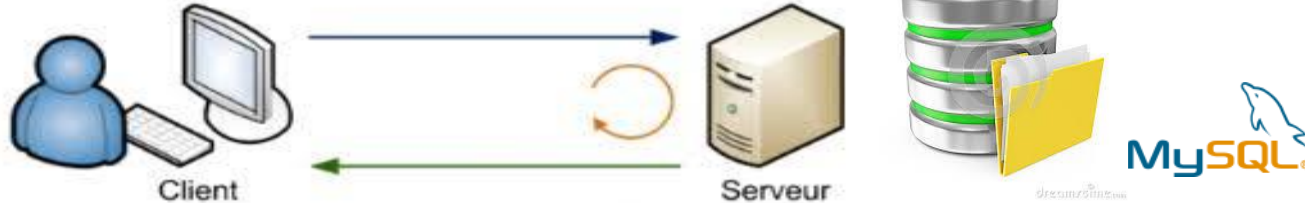
Avantages:

- uniformisation du format des données
- requêtes relationnelles entre plusieurs tables de données: exemple
 - à quelle période le pH > 8 sur 4 capteurs(4 tables) ?
 - quelle était la conductivité à ce moment la ?
 - on délocalise le trie des données sur le serveur
- stockage de données > 8 millions de Tera Octets par table avec MySql: (théorie)
 - séries chronologiques entières...
- un Système de Gestion de la Base de Données: (SGBDR) :
 - L'accès simultané entre utilisateurs, Les droits.
 - L'emplacement de mémorisation: Plus de gestion des chemins de stockage
 - Accès à des données centralisées distantes

3) principe de fonctionnement

Producteur /
Consommateur

Système gestionnaire
SGBRD



Application:

- lecture
- écriture

Base de données MySQL

Table

	Champ 1	Champ 2	...
Entrée	xxxxxx	xxxxxx	
Entrée	xxxxxx	xxxxxx	
Entrée	xxxxxx	xxxxxx	
Entrée	xxxxxx	xxxxxx	
.....	


Table

Table

3) principe de fonctionnement

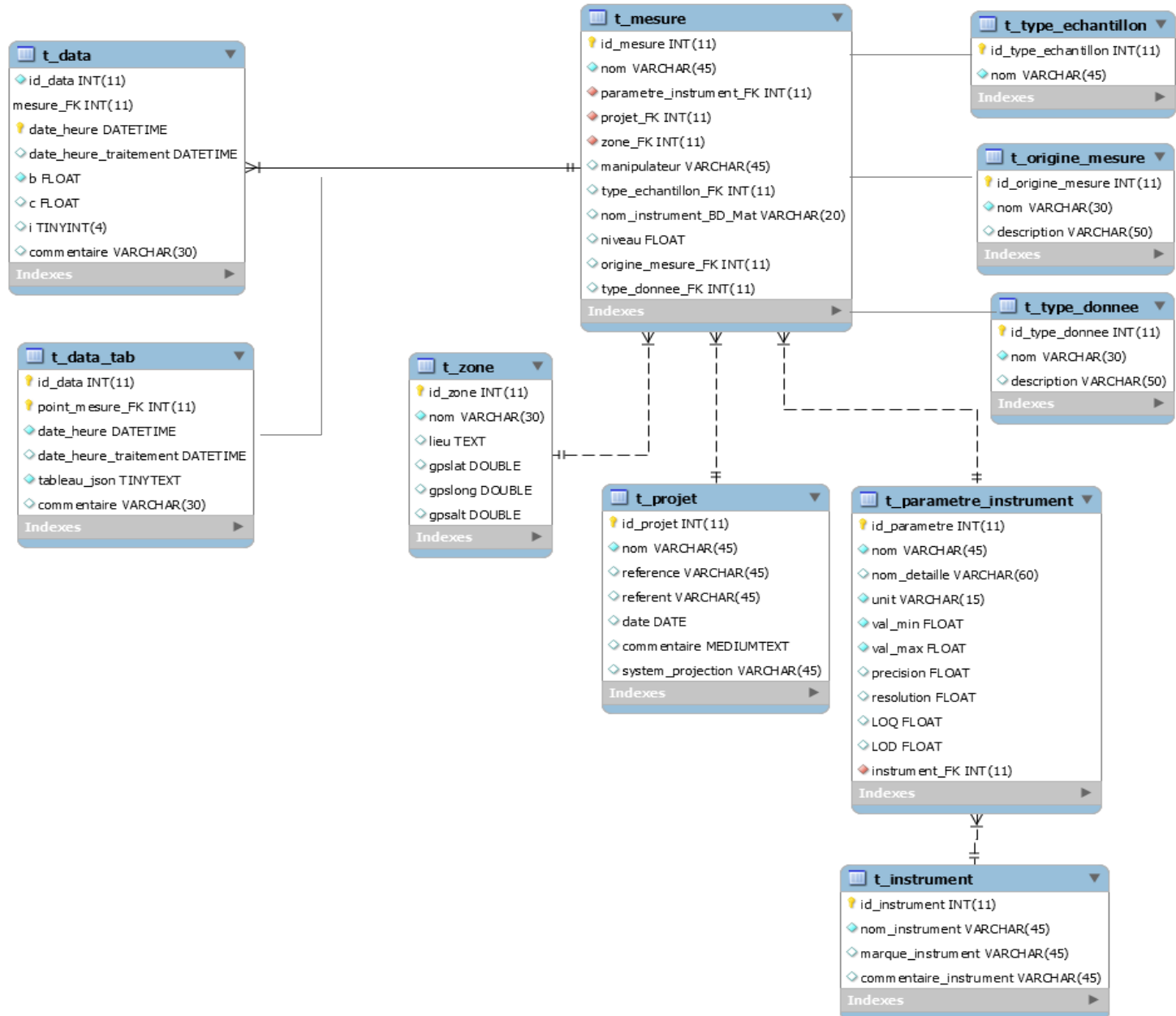


À propos des bases de données relationnelles

- Utilisation depuis : 1980
- Incontournables : industrie, e-Commerce, banque, recherche scientifique etc.
- Bases sur le marché: nombreuses, spécialisées ou non, payante ou non.
- Un choix de SGBDR:  gratuite, la plus utilisée, bien documentée.
- Spécialité: concepts, terminologie, administration, sécurité
- Définir une structure adaptée et ouverte. Etablir des liaisons
- Langage d'accès : SQL *Structured_Query_Language* (optionnel)

4) une base 'data' au LEESU

organisation de données: Modèle de conception



4) une base 'data' au LEESU

: Tables et relations

id_data	mesure	date_heure	b	c	i
1403774	1	2015-12-22 12:59:00	0.081	NULL	NULL
1403775	1	2015-12-22 13:04:07	0.911	NULL	NULL
1403776	1	2015-12-22 13:09:07	1.266	NULL	NULL
1403777	1	2015-12-22 13:14:07	1.271	NULL	NULL
1403778	1	2015-12-22 13:19:07	1.272	NULL	NULL
1403779	1	2015-12-22 13:24:07	1.271	NULL	NULL

	id_mesure	nom	parametre_instrument_FK	projet_FK	zone_FK	type	nom_instrument
▶	1	Profondeur pt B M	11	0	2	0	198
	2	Temperature pt B M	2	0	2	0	198
	3	Conductivité pt B M	3	0	2	0	198
	4	O2 pt B M	4	0	2	0	198
	5	Fluorescence pt B M	5	0	2	0	198
	6	Phycocyanine pt B M	6	0	2	0	198
	7	pH pt B M	7	0	2	0	198

id_parametre	nom	nom_detaille	unit	val_min	val_max	precision	resolution	LOQ	LOD
1	Profondeur	Profondeur Sambat	m	-1	10	NULL	NULL	NULL	NULL
2	Temperature	Temperature Sambat	°C	-5	35	NULL	NULL	NULL	NULL
3	Conductivité	Conductivité Sambat	mS	0	2	NULL	NULL	NULL	NULL
4	O2	O2 Sambat	%	0	200	NULL	NULL	NULL	NULL
5	Fluorescence	Fluorescence Sambat	µg/l	0	500	NULL	NULL	NULL	NULL
6	Phycocyanine	Phycocyanine Sam...	Cel/ml	0	10000	NULL	NULL	NULL	NULL
7	pH	pH sambat	u	0	14	NULL	NULL	NULL	NULL

Inst	id_instrument	nom_instrument	marque_i
	0	Refresh data re-executing the origin	
	1	ctd_nke_mpx1	NKE
	2	ctd_nke_mpx2	NKE
	3	temp_NKE_SP2T	NKE
	4	Station nutriment	LEESU
	5	Meteo ciel	Meteo ciel
	6	wheather API	NULL
	7	Station Proliphyc Météo	LEESU

id_projet	nom	reference	referent	date	commentaire	system_projection
0	Roulepur Compans	NULL	MC Gromaire	NULL	NULL	lambert 93
1	SenseCity	NULL	MGC&MS	NULL	NULL	lambert 93
2	Bassin Villette	NULL	Brigitte Vinçon	NULL	NULL	lambert 93

id_zone	nom	lieu	gpslat	gpslong
0	Eau brute	Compans	48.995225	2.649119
1	Noue filtrante	Compans	48.995546	2.648298
2	Accotement filtr...	Compans	48.9962	2.64666
3	Arbres	Champs ...	48.8418	2.59096
4	Amont	la Vilette	48.8915	2.385471
5	Baignade	la Vilette	48.8858	2.373795

id_origine_mesure	nom	description
0	capteur	La mesure est issue d'un capteur dans le milieu
1	analyse	La mesure est issue d'une analyse en laboratoire

id_type_echantillon	nom
3	air
4	eau
1	sédiment
2	sol
5	systeme

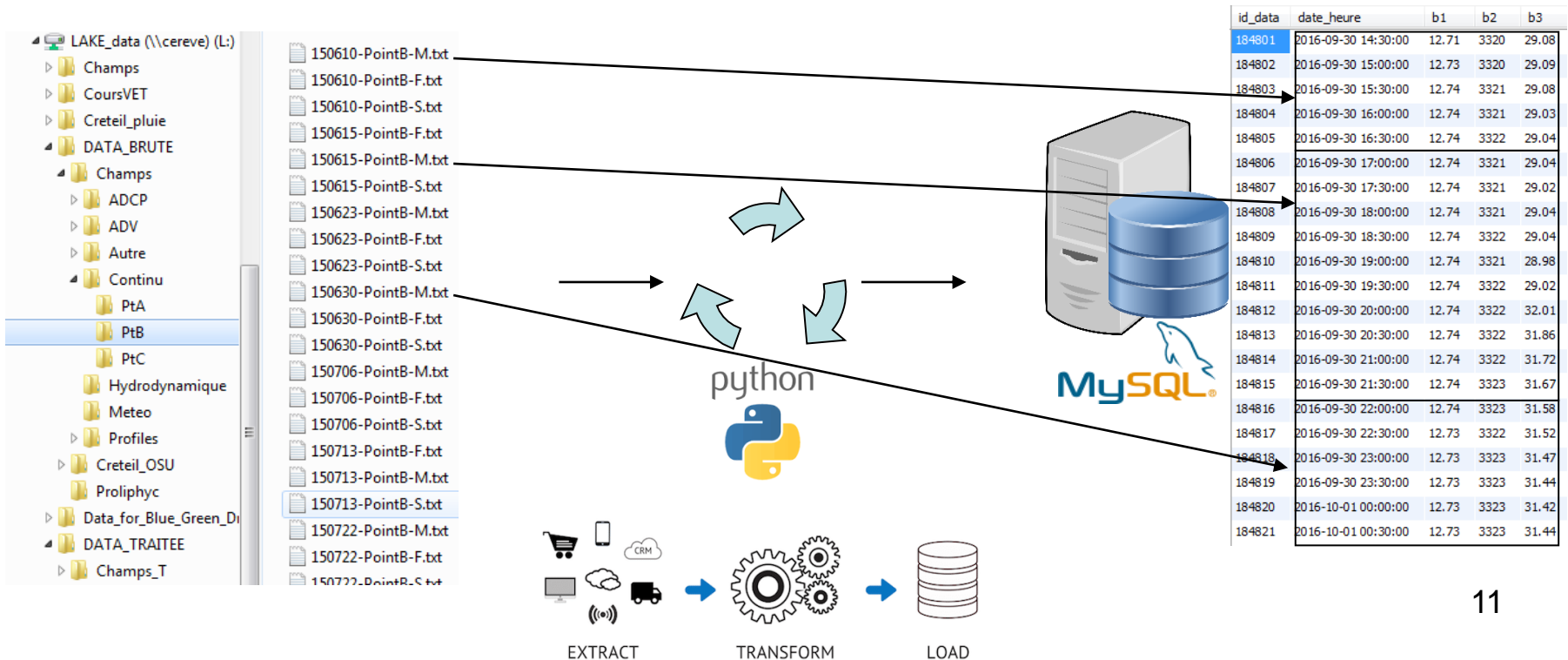
id_type_donnee	nom	description
1	Tableau de valeurs numériques	Une tableau de valeurs: Chaîne
2	valeur numérique	Une valeur de type réel

4) une base 'data' au LEESU

Ecriture des données sur la base

Automatiser:

- Conserver le dépôt manuel ou automatique(télétransmission) des fichiers brutes
- Une routine automatique d'écriture des nouvelles données.
- Il vérifie régulièrement si des nouvelles données arrivent (5 fois/jour)
- L'utilisateur final des données n'a rien à gérer:
 - Le format d'origine, la mise 'bout à bout' des séries



5) les moyens d'accès

Mode d'accès aux données: DIRECT via SQL

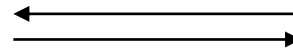
- À partir votre langage de programmation préféré : nécessite de s'intéresser au langage SQL

```
SELECT b FROM t_data Where id_mesure = 'Debit_entrant' where date > 17/06/2015
```

Sélectionner les valeurs depuis la table t_data ou id_mesure = Debit_entrant et ou date est > 17/06/2015

C'est la méthode de bas niveau offrant le plus de possibilité dans les requêtes!

Requêtes SQL direct depuis
son propre programme



python



Remplacer vos fonctions de lecture fichier:

-openfile

-readLines + vos fonctions de sélection

-read

-close



-Par l'équivalent SQL...en mieux



5) les moyens d'accès : direct

exemple de code



Appel de la bibliothèque	<pre>import mysql.connector</pre>
Se connecter à la base	<pre>conn = mysql.connector.connect(host=data_leesu,user=ld,password=PW, database=db_opur) cursor = conn.cursor()</pre>
Lancer la requête SQL	<pre>cursor.execute('SELECT * FROM t_data_compans_debits')</pre>
Récupérer le résultat	<pre>rows = cursor.fetchall()</pre>
Fermer la connexion	<pre>conn.close()</pre>

5) les moyens d'accès : direct

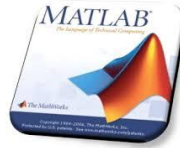
exemple de code



Appel de la bibliothèque	<code>library(RMySQL)</code>
Se connecter à la base	<code>mydb = dbConnect(MySQL(), user='user', password='PW', dbname='db_opur', host='data_leesu')</code>
Lancer la requête SQL	<code>rs = dbSendQuery(mydb, "SELECT * FROM t_data_compans_debits ")</code>
Récupérer le résultat	<code>data = fetch(rs, n=-1)</code>
Fermer la connexion	<code>dbDisconnect(mydb)</code>

5) les moyens d'accès : direct

exemple de code



Via DataBAsE ToolBox + driver ODBC

Pas d'appel de la bibliothèque	<i>Configurer une connexion via ODBC</i>
Se connecter à la base	<code>conn = database('MySQL','username','pwd');</code>
Lancer la requête SQL	<code>data = select(conn,'SELECT * FROM t_data_debits')</code>
Récupérer le résultat	<i>Données déjà dans data</i>
Fermer la connexion	<code>close(conn)</code>

5) les moyens d'accès

Via une interface 'généraliste' Laboratoire



Affichage graphique et exports de données, d'image, tableau etc.

Data Base Plot
Ver 08/02/2021 PhD

Liste des mesures à tracer

- Conductivité pt B M;b;1.0
- Fluorescence pt B M;b;1.0
- O2 pt B M;b;1.0
- pH pt B M;b;1.0
- Temperature pt B M;b;1.0

Début de période
janvier, 2021

	lun.	mar.	mer.	jeu.	ven.	sam.	dim.
53	28	29	30	31	1	2	3
1	4	5	6	7	8	9	10
2	11	12	13	14	15	16	17
3	18	19	20	21	22	23	24
4	25	26	27	28	29	30	31
5	1	2	3	4	5	6	7

Fin de période

Détail

Unité: °C, Niveau(m): -1.5
Minimum: -5,0, Maximum: 35,0
Compartment: eau
Origine: capteur
Type donnée: valeur numérique
Instrument Ref: 198

Figure 1

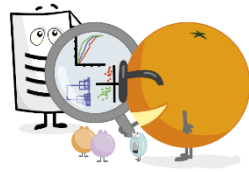
Graph showing temperature data (Temperature pt B M;b;1.0) over time (05/02/2018 to 12/02/2018). The graph displays multiple data series for different locations (TeneurEC5(1)_b (%) to TeneurEC5(9)_b (%)).

python
Ph Dubois
Exécutable évolutif

5) les moyens d'accès

Programmes spécialisés, interfaces et web

- à partir d'un soft d'affichage et /ou calculs ,SIG etc



Orange data mining



- depuis une interface d'administration ex: MySQL workbench



- à partir d'une interface web : l'exemple du projet BALNEAU



6) types et flux de données

Différents types de données

- Continu : mesures automatiques horodatées
- Ponctuel : données à fréquence indéterminée
- Séries : tableaux de mesures (ex profils, 3D)
 - Avec la fonction 'magique': serialise

id_data_type	data_type
2	continu
5	fichiers
1	ponctuel
3	series
4	textephotos

Des champs réservés à la validation

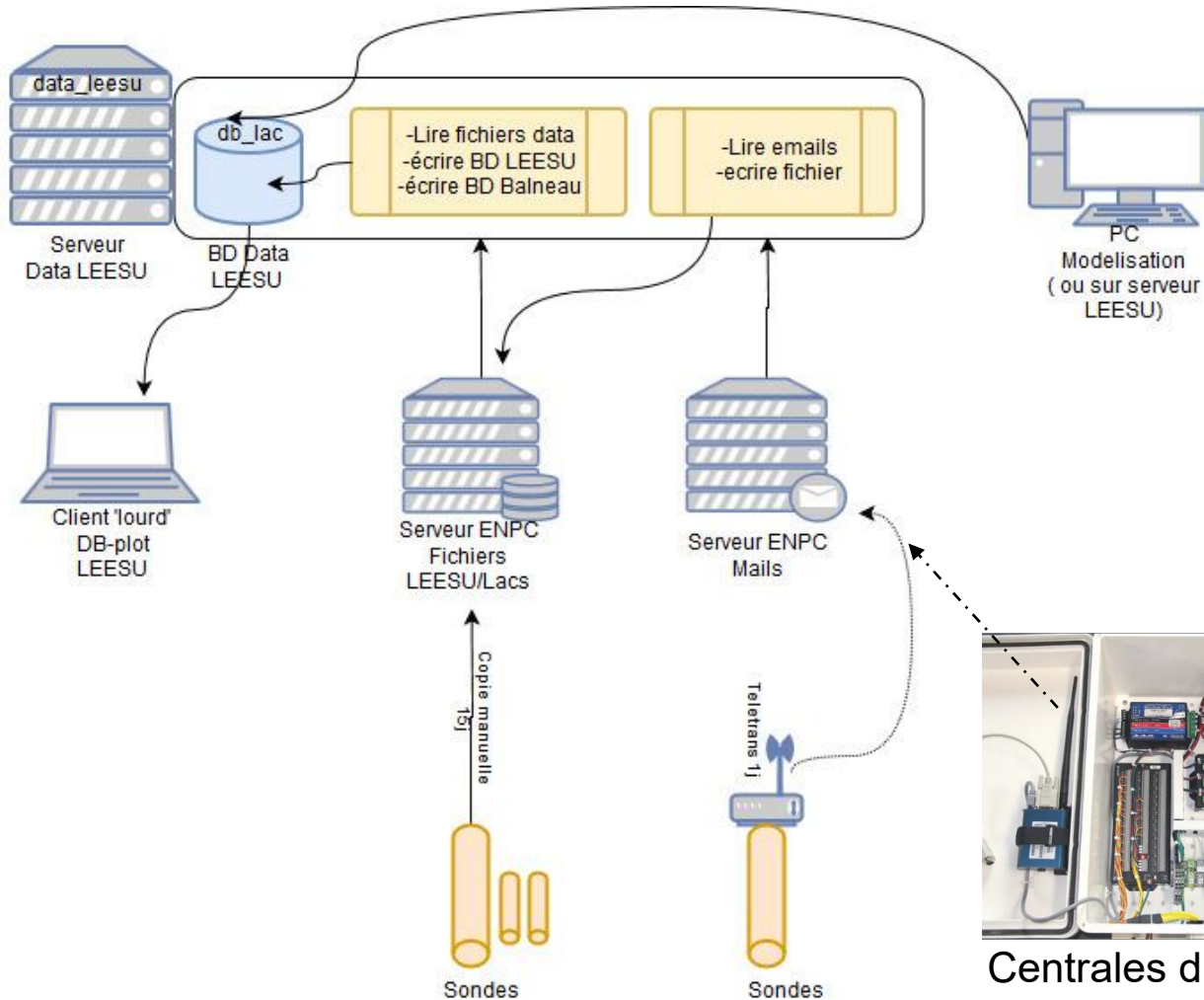
- bx : données brutes
- cx : données corrigées
- lx : indicateur sur correction

id_data	mesure	date_heure	b	c	i	commentaire
1179695	1	2016-01-26 02:16:09	1.597	NULL	NULL	NULL
1179696	1	2016-01-26 02:21:09	1.596	NULL	NULL	NULL
1179697	1	2016-01-26 02:26:09	1.597	NULL	NULL	NULL
1179698	1	2016-01-26 02:31:09	1.597	NULL	NULL	NULL
1179699	1	2016-01-26 02:36:09	1.596	NULL	NULL	NULL
1179700	1	2016-01-26 02:41:09	1.595	NULL	NULL	NULL
1179701	1	2016-01-26 02:46:09	1.594	NULL	NULL	NULL
1179702	1	2016-01-26 02:51:09	1.593	NULL	NULL	NULL

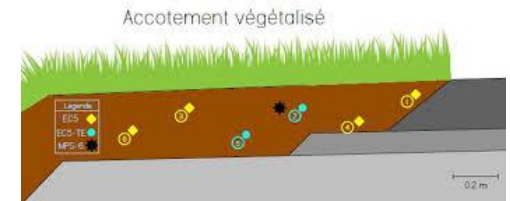
7) types et flux de données

Exemples de flux de donnée automatisé

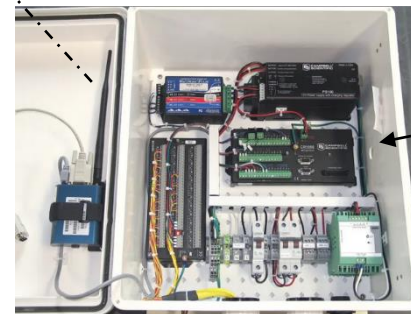
- Routine d'ajout automatique des fichiers de mesure en continu dans la base:
- ⇒ > 13 millions de mesures 'en continu' actuellement pour LACs / Champs.
- ⇒ > 5 millions Opur / Roulepur / Compans



Stations 'lacs'



Stations Roulepur 'sol & eau'



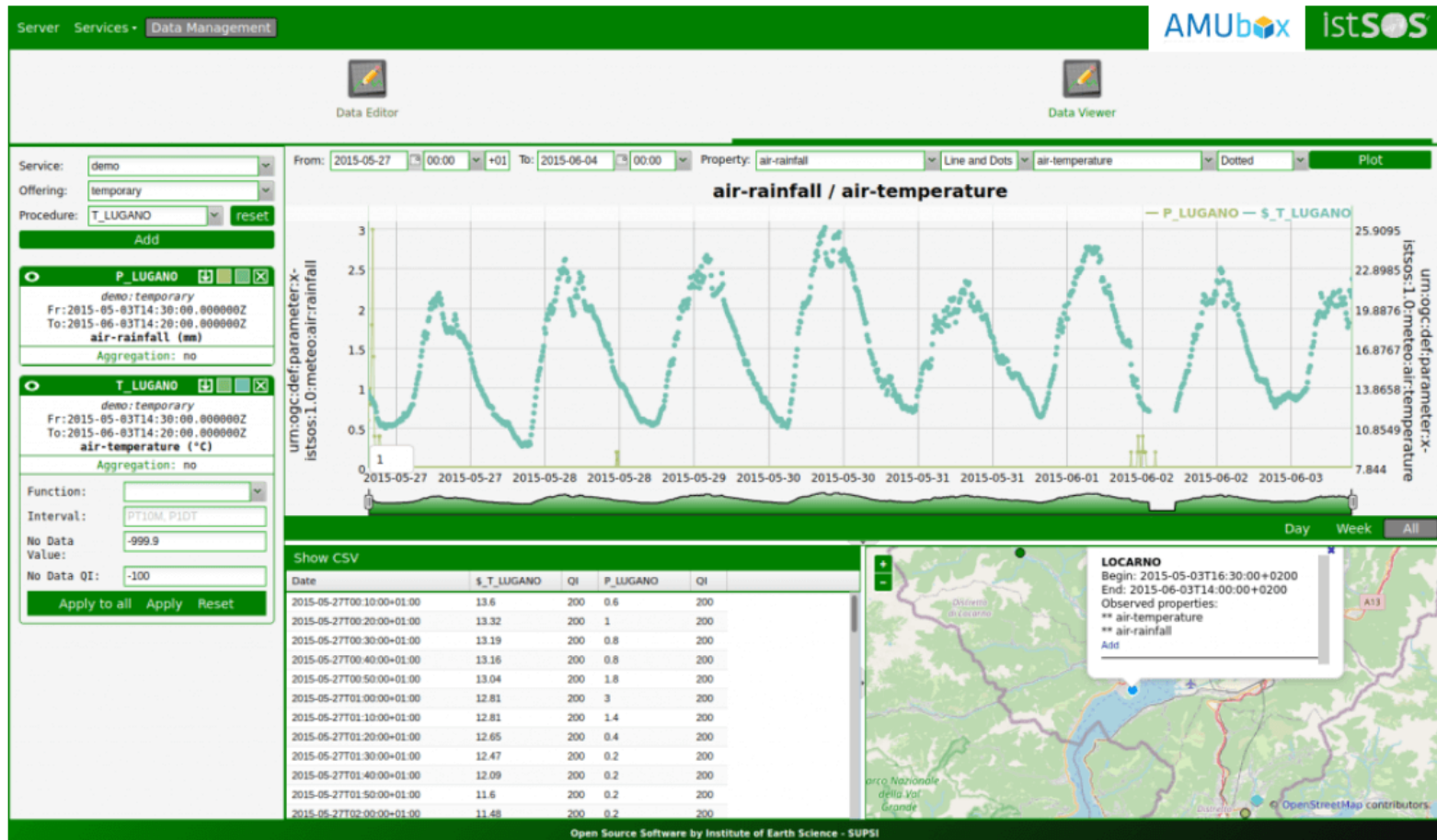
Standard
Indépendance
Modularité

Centrales d'acquisition + GPRS

7) Gestion des données & Ouverture vers l'extérieur

Plate forme web de visualisation

-istSOS Outils générique en ligne : hébergement locale

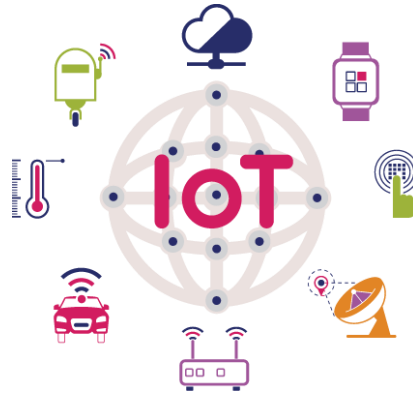


-Alimentation automatisée à partir de la base 'Data LEESU' selon protocole standard SOS

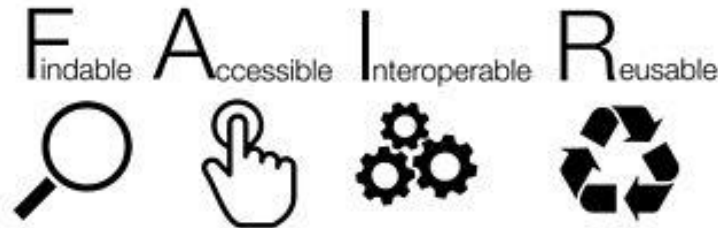
7) Gestion des données & Ouverture vers l'extérieur

Notions connexes

- IOT : Internet Of Things



- Bonnes pratiques



-Directive INSPIRE: INfrastructure for SPatial InfoRmation in Europe

Partage de données sur internet via infrastructure:
favoriser la protection de l'environnement

- Obligations: catalogue, accès gratuit, ...
- Données: 34 domaines
- Enjeux: améliorer l'information sur les données,
faciliter les échanges de données,...



7) Gestion des données & Ouverture vers l'extérieur

-Alimentation de plates formes d'obervatoires

Infrastructures de données spatiales (IDS)

Catalogue de Métadonnées



-SNO Observil Service National d'observation

- IDS basé sur [GeoSUNA](#)
- Regroupe les données de 11 laboratoires/observatoires en France
- Données de 4 sites de mesure en continu pour OPUR

-ZA Seine

- IDS basé sur [GeoNetwork](#) version 3.4.0,
- un serveur cartographique : [GeoServer](#) version 2.12.1



- Source de préparation des méta données / données vers

- DMP : Data Management Plan
- Data paper

8) le lien base data -> matériel

La base de donnée 'MATERIELS'

Rôle:

<https://basemateriel.enpc.fr>

- Inventaire de l'existant: => réutilisation, bilan matériel annuel.
- Traçabilité des opérations de maintenance: historique
- Associer des documents: procédures, notices, images etc (8Mo)
- Réservations

Base 'DATA' + Base 'Matériel' :

-Traçabilité

- Lien: le champs 'nom appareil'

109 seconde - 7.62 Mio

GLPI 9.2.4 Copyright (C) 2015-2018 Teclib' and contributors - Copyright (C) 2003-2015 INDEPNET Development Team



Data Base PLOT: traçage et export de mesures stockées sur la base de données

Leesu
laboratoire eau environnement systèmes urbains

Ver 08/02/2021 PhD

Liste des mesures à tracer

- Proliphyc eau CTD Température;b;1,0
- Proliphyc eau CTD Turbidité;b;1,0
- Proliphyc eau FP_Bluegreen;b;1,0
- Proliphyc eau FP_Cryptophyta;b;1,0
- Proliphyc eau FP_Diatoms;b;1,0

Tracer Exporter

Mode points Multi dates

Début de période: mars, 2021

	lun.	mar.	mer.	jeu.	ven.	sam.	dim.
8	22	23	24	25	26	27	28
9	1	2	3	4	5	6	7
10	8	9	10	11	12	13	14
11	15	16	17	18	19	20	21
12	22	23	24	25	26	27	28
13	29	30	31	1	2	3	4

Fin de période: mars, 2021

	lun.	mar.	mer.	jeu.	ven.	sam.	dim.
8	22	23	24	25	26	27	28
9	1	2	3	4	5	6	7
10	8	9	10	11	12	13	14
11	15	16	17	18	19	20	21
12	22	23	24	25	26	27	28
13	29	30	31	1	2	3	4

Détail

Unité: µg/L Niveau(m): 0,0

Minimum: 0,0 Maximum: 0,0

Compartment: eau

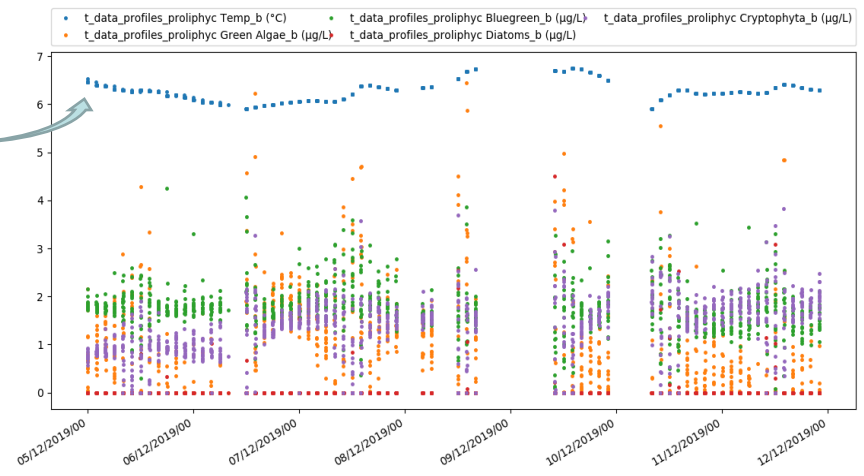
Origine: capteur

Type donnée: valeur numérique

Instrument Ref: 184

Base Materiel

Concentration Diatomés Fluoroprobe station Proliphyc



merci

