

SiQaCo

Contrôle qualité de données sismiques

Olivier Geber

<https://github.com/IPGP/siqaco>

22 novembre 2018



1. Introduction
2. Cahier des charges
3. Description du produit
4. Questions

1. Introduction

Qu'est ce que SiQaCo ?



SiQaCo est un logiciel en cours de développement

Les principales fonctionnalités du logiciel sont :

1. De créer et mettre à jour une structure finale¹ de données et de méta-données validées à partir de différentes sources de données et de méta-données
2. D'effectuer le contrôle qualité sur les données et méta-données

1. Archive de données découpés de minuit à minuit, compression homogène, avec un label de qualité défini par l'utilisateur ("M" ou "Q")

1. Introduction

Contexte



- ▶ Besoin formulé par les observatoires volcanologiques de l'IPGP
- ▶ Récupération des données : trous + recouvrements dus à des pannes (station, transmission)
- ▶ Validation des données : effectuée par un ensemble d'étudiants
- ▶ Suppression des trous et recouvrements : AEQC + autres outils

→ Dans ce contexte, un cahier des charges a été élaboré à partir de l'expression des besoins des différents observatoires

2. Cahier des charges

Résumé



- ▶ Publié le 9 Avril 2018, diffusé à la communauté RESIF
- ▶ Disponible sur Github :
[github.com/IPGP/siqaco/blob/master/Cahier des charges Siqaco.pdf](https://github.com/IPGP/siqaco/blob/master/Cahier%20des%20charges%20Siqaco.pdf)
- ▶ Retours principalement de l'IPGP, Géoscope

2. Cahier des charges

Expression des besoins : administration



L'outil doit :

- ▶ Être synthétique
- ▶ Être simple à utiliser
- ▶ Avoir des rapports d'erreur facile et rapide à analyser
- ▶ Être collaboratif & évolutif

L'outil comportera également une interface de visualisation :

- ▶ De l'état du traitement des discontinuités
- ▶ Des rapports d'erreurs
- ▶ De l'analyse de métriques sur l'archive finale

2. Cahier des charges

Expression des besoins : fonctionnalités



Récupération de données

- ▶ Manuelle (clé USB, CD, SD...)
- ▶ Source distante
- ▶ Priorité des sources
- ▶ Disponibilité de la donnée

Bande passante

- ▶ Adaptation de l'outil en fonction de la puissance locale et distante
- ▶ Limitation en BP
- ▶ Optimisation des requêtes de données
- ▶ Système de file d'attente

Journaux

- ▶ Rapports traitement des données
- ▶ Logs du logiciel

Visualisation

- ▶ Visualiser l'archive et état du *complètement*
- ▶ Désactiver la surveillance de certaines stations
- ▶ Configurer le logiciel
- ▶ Afficher les *logs*

2. Cahier des charges



Expression des besoins : développement

Modularité

- ▶ Faciliter les développements futurs
- ▶ Pérenniser le logiciel
- ▶ Pas d'impact si changement de format dans les données ou changement de protocole

Logiciel orienté BDD

- ▶ Optimiser l'espace utilisé
- ▶ Optimiser le temps de calcul

Qualité des données

- ▶ Éviter de trop *toucher* à la donnée
- ▶ Blocage en 4K

2. Cahier des charges

Expression des besoins : contrôle qualité



Vérification des méta données

- ▶ Conformité aux standards
- ▶ Pas de recouvrement dans les époques des méta-données
- ▶ Cohérence des date et canaux dans méta-données

Vérification croisée méta-données et données

- ▶ Cohérence des données entre méta-données et séries temporelles
- ▶ Unicité des méta-données par échantillon

Validation des séries temporelles

- ▶ Vérification des NSLC
- ▶ Vérification du label de qualité
- ▶ Pas d'erreur à la décompression
- ▶ Fréquence d'échantillonnage et encodage homogène
- ▶ Taille des blocs, ordre des octets

3. Description du produit

Résumé



- ▶ Dossier de conception publié le 9 Août 2018, diffusé à la communauté RESIF
- ▶ Disponible sur GitHub :
github.com/IPGP/siqaco/blob/master/Dossier-de-conception.pdf

3. Description du produit

Langage & Framework



Environnement

- ▶ Linux
- ▶ Python 3.6
- ▶ Anaconda

Framework

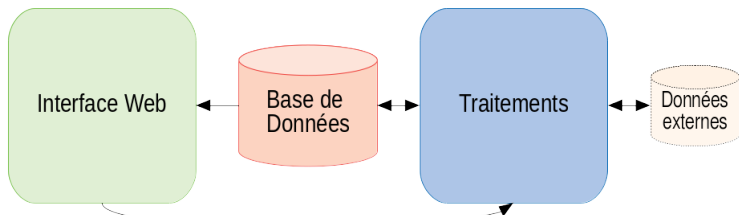
- ▶ Obspy 1.1.0
- ▶ Django 2.7

SGBDR

- ▶ MySQL 14 (pour le développement)
- ▶ ORM : Django

3. Description du produit

Principe général



3. Description du produit

Fonctionnalités



Traitements

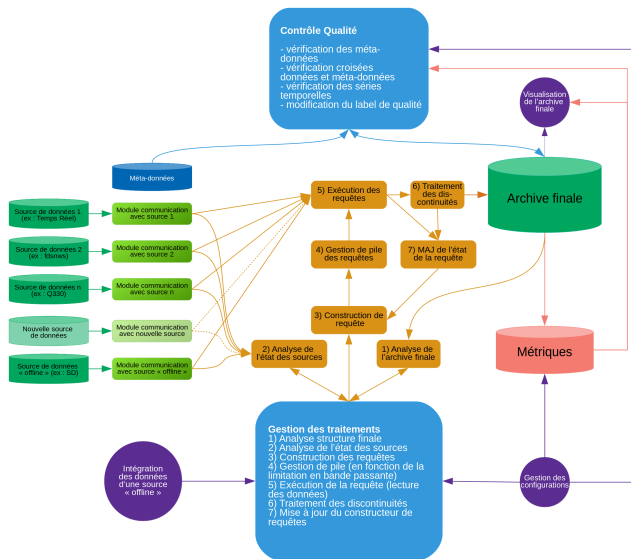
- ▶ Analyse de l'archive finale (métriques, liste des trous)
- ▶ Disponibilité des sources
- ▶ Inventaire de l'état des sources
- ▶ Construction des requêtes pour traiter les discontinuités
- ▶ Gestion des requêtes
- ▶ Exécution de la requête et mise à jour de la requête
- ▶ Traitement des discontinuités

Contrôle Qualité

- ▶ Vérification des métadonnées
- ▶ Vérification croisée des données avec les métadonnées
- ▶ Vérification des séries temporelles
- ▶ Modification du label de qualité

3. Description du produit

Schéma global



3. Dossier de conception

Interface



Visualisation

- ▶ Visualiser l'état de l'archive finale
- ▶ Visualiser les métriques
- ▶ Configurer la visualisation (désactiver certains canaux par ex.)
- ▶ Visualiser les tâches en cours
- ▶ Visualiser les logs, erreurs, historique des traitements

Surcharge

- ▶ Récupérer des données manuellement
- ▶ Annuler ou relancer une ou plusieurs requêtes
- ▶ Désactiver les alertes
- ▶ Modifier l'archive finale manuellement

Admin

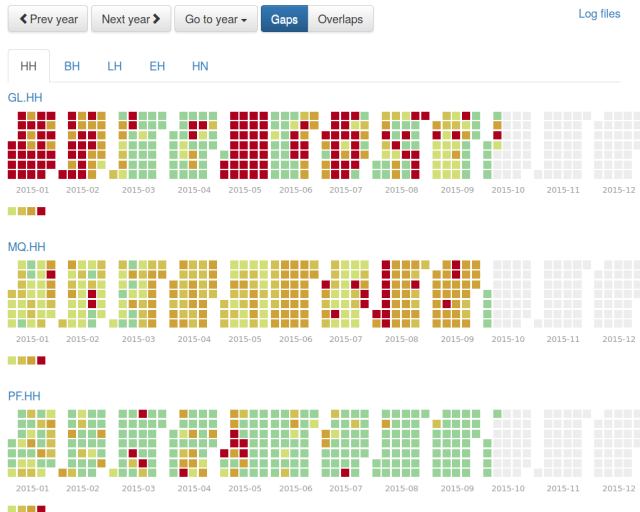
- ▶ Configurer le logiciel
- ▶ Gérer les utilisateurs

3. Description du produit

Interface : visualisation de l'archive



Network statistics



3. Description du produit

Interface : visualisation de l'archive



◀ Back to networks

HH BH LH EH HN other

GL.AMC.00.HH

GL.CAG.00.HH

GL.CDE.00.HH

GL.DHS.00.HH

GL.DSD.00.HH

GL.LKG.00.HH

GL.MML.00.HH

GL.MOD.00.HH

GL.MOG.00.HH

GL.SAV.00.HH

GL.SCG.00.HH

GL.TAG.00.HH

GL.TAG.90.HH

GL.TAGZ.00.HH

GL.TAGZ.90.HH

◀ Prev year

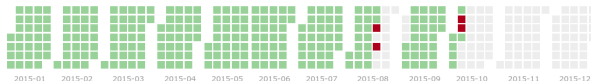
Next year ▶

Go to year ▼

Gaps

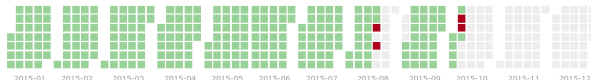
Overlaps

GL.LKG.00.HHN.D



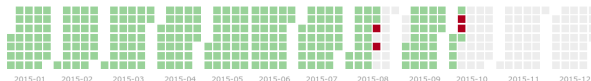
■ ■ ■ ■

GL.LKG.00.HHE.D



■ ■ ■ ■

GL.LKG.00.HHZ.D



■ ■ ■ ■

3. Description du produit

Calendrier prévisionnel



- ▶ Fin février 2019 : Version Beta
- ▶ Mars - Avril 2019 : Phase de test en observatoires
- ▶ Avril - Mai 2019 : Présentation du logiciel

Questions ?



Merci pour votre attention

