



**HAL**  
open science

# La circulation de l'information, de son acquisition, à son exploitation....

Charlotte Allain, François Laperruque

## ► To cite this version:

Charlotte Allain, François Laperruque. La circulation de l'information, de son acquisition, à son exploitation..... Cahier des Techniques de l'INRA, 2018, N° Spécial: Phénotypage animal: de la donnée brute à sa valorisation, N° Spécial: Phénotypage animal: de la donnée brute à sa valorisation, pp.49-52. hal-02617660

**HAL Id: hal-02617660**

**<https://hal.inrae.fr/hal-02617660v1>**

Submitted on 10 Sep 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License



## Unités Expérimentales, acquisition de l'information

Sur les Unités Expérimentales, l'information peut être collectée de différentes façons (**Figure 1**). L'information peut être saisie directement sur le terrain, grâce à l'utilisation de boîtier de saisie durci de type PDA (personal digital assistant). Le Cati Sicpa a développé plusieurs applications sur PDA afin de faciliter la saisie d'informations. Cela évite les erreurs dues aux doubles saisies, papier puis encodage informatique. Pour la plupart des applications, le PDA doit être connecté en Wi-Fi à une base de données locale (base de données relationnelles MySQL) stockée sur un serveur de stockage intermédiaire. Cela donne l'avantage de pouvoir utiliser plusieurs PDA en même temps sur le même chantier. Cela est possible sous condition d'avoir une bonne connexion Wi-Fi sur le lieu d'élevage. Cependant, il existe aussi des cas dans lesquels un fonctionnement déconnecté est requis (mauvaise couverture Wi-Fi, choix dans l'architecture de l'application), dans ce cas, le chantier de saisie est embarqué sur le terminal.

Avant d'aller sur le terrain, il faut initialiser le chantier avec la liste des animaux. L'ensemble des SI (espèces et multi espèces) permettent de le faire facilement depuis leur logiciel PC. Après le chantier, il faut télécharger les données collectées, cette opération se réalise depuis le PC dans l'élevage. Dès cet instant, les données sont envoyées au CTIG (Centre de Traitement de l'Information Génétique) pour stockage, et rendues accessibles pour les scientifiques.

L'information peut être également collectée par des automates (de pesée ou de contrôle laitier). Dans ce cas également, le SI espèce permet de préparer facilement les fichiers à charger sur les automates et de télécharger facilement les données collectées pour envoi au CTIG. Il reste aussi la possibilité d'enregistrer des données via une saisie clavier sur PC ou même par intégration manuelle d'un fichier. Ces deux derniers cas ne sont fiables que si la quantité et la fréquence de ces actions sont limitées.

Le Cati Sicpa a développé des systèmes de distributions d'aliments capables de collecter, en masse, des données sur le comportement des animaux, lors de leur alimentation. Ces systèmes, qui n'existent pas dans le commerce avec les fonctionnalités souhaitées, sont complètement automatiques, de l'acquisition de l'information jusqu'à sa remontée vers le stockage centralisé au CTIG. Ils s'appuient sur le réseau local des UE (fibre, Wi-Fi ou Ethernet) et sur un réseau spécifique utilisant des adresses non routées (cf RFC1918). Ce 2<sup>e</sup> réseau fonctionne de façon autonome et interconnecte des distributeurs, des caméras et un serveur de stockage intermédiaire, situé dans l'UE. L'ensemble des systèmes d'acquisition de phénotypes, sont donc « isolés », ce qui constitue un niveau intéressant de sécurité informatique contre les intrusions. Les données collectées par les distributeurs sont dupliquées en temps réel, vers le serveur de stockage local. Ce dernier, ayant accès au réseau spécifique et au réseau de l'UE, envoie automatiquement ses données, la nuit par Internet, vers le stockage centralisé de Sicpa Expérimentations. Ses grandes capacités de stockage et de traitement, font qu'il remplace les mini serveurs de type « Linutop » lorsque des besoins Big Data apparaissent dans les UE.

De ce fait, l'ensemble des outils de collecte de données sur les animaux, ayant besoin d'un stockage local, convergent vers l'utilisation d'un même serveur, dimensionné en fonction des volumes de stockages attendus.

## Interconnexions

La connexion à Internet des sites expérimentaux, est gérée et financée par les équipes de la DSI (Direction du Système d'Information) de l'Inra, qui s'appuient sur le GIP (groupement d'intérêt public) RENATER et ses solutions de connectivité. Ces sites sont raccordés via des technologies différentes (fibre, Radio, CN2, SDSL, ADSL) en fonction des possibilités locales. Ces interconnexions sont stratégiques car il est bien évident qu'il va être impossible de faire transiter d'importants volumes de données via un lien sous dimensionné. Pour

ses propres besoins, le Cati Sicpa assure une supervision de 1<sup>er</sup> niveau de ces raccordements, grâce au logiciel SmokePing (<https://oss.oetiker.ch/smokeping/>), qui permet de tester la réponse, (temps et qualité) d'un matériel (routeur ou serveur) situé sur un site expérimental (**Figure 2**). Pour le support des interconnexions réseaux, nous avons la chance de pouvoir profiter de l'appui des EIC (Equipes Informatiques de Centre) qui connaissent bien nos demandes, en matière de raccordements.

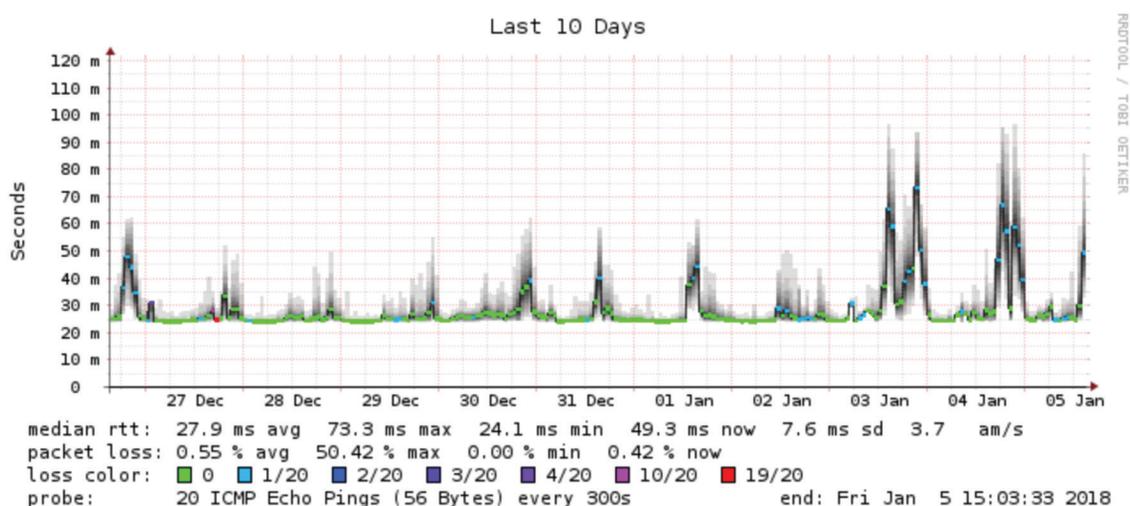


Figure 2. Exemple de graphique généré par l'application SmokePing.

## CTIG, stockage de l'information

Le Centre de Traitement de l'Information Génétique (Jouy-en-Josas), Unité de Service du Département de Génétique Animale, héberge l'ensemble des SI en production du Cati Sicpa. Pour cela, selon le type de service, des machines physiques ou virtuelles sont mises à notre disposition. Ces machines sont administrées par l'équipe système du CTIG (montée des versions de logiciels, correction des failles de sécurité, etc.), qui assure aussi leur sauvegarde. Ce dernier point important, donne lieu à des sauvegardes complètes et incrémentales des systèmes de fichiers et des bases de données, avec copies sur supports externalisés. D'autre part, les machines virtuelles sont aussi sauvegardées « à chaud » sur une baie de disques, installée dans la salle Inra du Data Center Ile-de-France. L'ensemble de ces sécurités s'inscrit dans le cadre d'un PRA (plan de reprise d'activités) qui permet au CTIG de repartir de 0, en cas d'incident majeur. Un logiciel d'archivage permet de déplacer les données sur des supports de stockage moins coûteux, dans un format pérenne. Les métadonnées associées aux données originelles, permettent à leur propriétaire de gérer finement leur cycle de vie (rétention, alertes, etc.).

## Unités de Recherche, exploitation de l'information

Sur les Unités de Recherche, les données sont mises à disposition des scientifiques, facilement. Depuis les applications client/serveur, il est possible d'extraire les données depuis les SI, en format texte, grâce à un module d'extraction de données, de sélectionner les variables à garder, de choisir les critères de sélection et de définir des clés de tri des résultats.

L'extraction des données peut également se réaliser directement sur le logiciel SAS (statistical analysis system) grâce à une interrogation des SI en langage SQL, avant une analyse statistique.





L'application SIDExTRACT (<https://forge-dga.jouy.inra.fr/projects/sidextract/wiki>), disponible au CTIG, permet aux scientifiques d'extraire les données d'une expérimentation gérée par Sicpa Expérimentations. Actuellement, ce logiciel interroge le serveur MariaDB, mais dans un avenir proche, il sera interfacé avec un stockage NoSql, vraisemblablement sous « Cassandra » pour les données acquises en masse.

Des solutions Web, interfacées aux SI espèces, sont en cours de développement (ex : Sicpa Porcins). Elles permettront d'accéder aux données d'un élevage pour chaque personne ayant les autorisations nécessaires, et sans avoir d'application à installer sur son PC. Ces interfaces Web permettront de construire des requêtes personnalisées en tenant compte des structures de la base de données hôte.

## Conclusion

La bonne circulation de l'information, entre tous les acteurs de la recherche, nécessite des infrastructures techniques modernes, des outils performants mais aussi et surtout, des personnels passionnés et formés pour les exploiter. Même si nous n'avons pas pu tout évoquer dans cet article, nous savons que sans ces trois conditions, rien n'est possible. Les agents sont et resteront à la base du bon fonctionnement du système global, chacun à leur niveau d'intervention.