

## Le système d'information Nutriflux

J. Vernet<sup>1</sup>, M. Reichstadt<sup>1</sup>, I. Ortigues-Marty<sup>1</sup>

	INRAE, UMRH1213, 63122 SAINT-GENÈS-CHAMPANELLE
	<p>Je suis actuellement en retraite depuis le 01/02/2020. A l'INRAE, j'étais ingénieur en traitement des données. De 1980 à 1999, je faisais partie de l'équipe des chambres respiratoires de Theix sous la direction de Michel Vermorel. L'objectif était essentiellement de déterminer les besoins énergétiques des animaux de ferme et des humains. Mon rôle était d'effectuer les mesures sur les chambres respiratoires et d'exploiter les données correspondantes (informatique, statistiques). De 1999 à l'heure actuelle, j'étais essentiellement associé aux recherches sur le métabolisme des tissus et organes sous la direction de Isabelle Ortigues-Marty avec comme objectif principal de trouver par modélisation de nouvelles lois de réponse du métabolisme des nutriments en relation avec l'alimentation. Mon rôle consistait essentiellement à produire des données issues de la bibliographie internationale (bases de données) et de former les étudiants aux méthodes et outils de méta-analyse. A partir de 2010, j'ai aussi été associé au projet ATOL (Animal trait ontology of livestock, département PHASE) dans lequel j'ai créé une partie importante de la base nutrition et alimentation. C'était de beaux challenges. J'espère que tout cela va continuer après mon départ.</p> <p><a href="mailto:jean.vernet@inrae.fr">jean.vernet@inrae.fr</a></p>
	<p>Je m'appelle Matthieu Reichstadt, Ingénieur d'Etude en bioinformatique et responsable des systèmes d'information à l'UMR1213 Herbivores au Centre ARA de Theix.</p> <p>Après mon master en bioinformatique j'ai été recruté au CNRS afin de travailler sur le développement d'une plate-forme bioinformatique de calcul parallèle, notamment dans le cadre de recherches sur le paludisme, basée sur le cloud computing et la grille de calcul.</p> <p>J'ai été embauché en 2009 à l'INRA au sein de l'Unité Mixte de Recherche sur les Herbivores, au sein de laquelle je suis responsable de l'équipe informatique. Spécialisé dans le web et les bases de données, je travaille pour mon unité, mais participe ou porte également des projets nationaux ou internationaux. J'ai collaboré sur le projet de bien-être animal européen welfur jusqu'en 2017 sur lequel je gérais la totalité de la partie informatique, et travaille actuellement principalement sur les développements des bases bioinformatiques, les ontologies, et la gestion des parcelles de l'INRA, en partenariat avec d'autres Centres ou d'autres instituts de recherche.</p> <p><a href="mailto:matthieu.reichstadt@inrae.fr">matthieu.reichstadt@inrae.fr</a></p>
	<p><a href="mailto:isabelle.ortigues@inrae.fr">isabelle.ortigues@inrae.fr</a></p>

**Résumé.** Nutriflux est un système d'information pour la modélisation de la nutrition des ruminants, à accès limités et contenant des données issues de publications internationales. Il permet de répondre à des questions de recherche bien spécifiques. Il comprend actuellement trois bases de données différentes. Les acteurs sont les chercheurs, les étudiants, un ingénieur en traitement des données et un informaticien. Le modèle relationnel permet de recevoir tout type d'expérience et tout type de données mesurées. Les données sont homogénéisées, identifiées de façon fine, certaines peuvent être estimées grâce à un système extérieur. Les processus sont l'accès et l'authentification des utilisateurs, la saisie des données, la validation des données, la caractérisation des aliments et l'extraction. La valorisation est faite hors Nutriflux par méta-analyse. Nutriflux permet une structuration forte des données ainsi qu'une description fine des animaux, des variables mesurées et des méthodes de mesure. Nutriflux a permis au moins 8 publications internationales et 5 thèses de doctorat.

**Mots clés :** système d'information, base de données, Nutriflux

**Abstract.** Nutriflux is an information system used for the modelling of ruminant nutrition, with limited access and containing data from international publications. It allows to answer very specific research questions. It currently includes three different databases. The actors are researchers, students, a data processing engineer and a computer scientist. The relational model allows to receive any type of experience and any type of measured data. The data are homogenized, finely identified, some can be estimated using an external system. The processes are user access and authentication, data input, data validation, food characterization and extraction. The valorization is done outside Nutriflux by meta-analysis. Nutriflux allows a strong data structuration and a detailed description of animals, measured variables and measurement methods. Nutriflux has enabled at least 8 international publications and 5 doctoral theses.

**Keywords :** information system, database, Nutriflux

## **Introduction**

L'objet de cet article est de présenter le système d'information Nutriflux du point de vue de l'utilisateur. La mise en œuvre de cet outil répond à deux exigences : le besoin de données pour la modélisation chez le ruminant (Sauvant *et al.* 2008) et le besoin de structurer des données très disparates provenant des publications. L'initiative en revient à un groupe de chercheurs d'INRAE de Theix travaillant au départ sur les flux sanguins de nutriments. Aujourd'hui, le système d'information Nutriflux est utilisé par les différents chercheurs, doctorants et stagiaires impliqués dans différents projets liés à ce système.

# Le Cahier des Techniques de l'Inra 2020 (99)

## Développement

### Généralités sur Nutriflux

Nutriflux est un système d'information conçu pour recevoir les données de la bibliographie internationale. Il sert à préparer les données qui seront utilisées pour la modélisation (Sauvant *et al.* 2008). Il permet de répondre à des questions de recherche bien spécifiées. Par exemple, il peut fournir les données correspondant aux flux sanguins de nutriments énergétiques mesurés/déterminés au niveau de certains organes chez des ruminants de différents stades physiologiques et recevant plusieurs types d'aliments (Loncke *et al.* 2009). Nutriflux est un système à accès limités : les utilisateurs doivent être impliqués dans les projets de recherches sous tendus par Nutriflux.

Nutriflux est actuellement composé de trois bases de données différentes. La base Flora permet d'étudier l'influence des variations de flux sanguins de nutriments à travers les tissus et organes en fonction des variations des conditions d'alimentation (Loncke *et al.* 2009). La base Alicar permet d'étudier la composition des produits animaux (carcasse, muscle) en fonction des variations des conditions d'alimentation (Al Jammal *et al.* 2016). La base  $\omega$ 3meat permet de prédire les concentrations en certains acides gras du muscle à partir des concentrations en d'autres acides gras des mêmes muscles (Mourot *et al.* 2013).

Les animaux sont ceux rapportés dans les publications internationales. Ils sont étudiés en groupes de plusieurs animaux. Les données mesurées sur les animaux sont moyennées pour le groupe. Les données d'alimentation sont celles rapportées dans les publications internationales et ce, de façon plus ou moins complète par rapport aux besoins des chercheurs. Afin de compléter ces données, Nutriflux est relié aux systèmes d'alimentation pour les ruminants dont il utilise les données (p. ex. INRA 2018).

### Les acteurs

Les acteurs sont les chercheurs, les étudiants, l'ingénieur d'études en traitement des données, et l'informaticien. Les chercheurs, Isabelle Ortigues-Marty, Pierre Nozière et Christelle Loncke pour l'INRA, ainsi que d'autres chercheurs de pays étrangers, associés aux différents projets sous-tendus par les 3 bases définissent les questions de recherche, la façon de les traiter et le suivi du travail des étudiants. Les étudiants (thésards, masters statistiques) participent à la saisie des données, font des extractions et étudient les questions de recherche par méta-analyse (Sauvant *et al.* 2008), ce dernier point se faisant hors Nutriflux. L'ingénieur d'études en traitement des données, Jean Vernet, transmet les méthodes et outils aux étudiants, assure le suivi au quotidien, saisit et valide les données et vérifie les extractions. L'informaticien, Matthieu Reichstadt, construit, met au point et fait évoluer le système d'information. Il interagit avec l'ingénieur d'études en traitement des données pour les demandes particulières et les modifications à apporter.

### **Les difficultés techniques et contraintes à surmonter**

Pour la conception du modèle des données, il existe plusieurs types d'expériences, avec des plans d'expérience différents que le système doit être capable de recevoir. Ensuite, il faut pouvoir recevoir tout type de données, appartenant à des domaines de recherche différents et pouvoir faire des extractions fines sur ces données. Enfin, les publications sont hétérogènes : le vocabulaire est hétérogène, p. ex. les noms des variables désignant une seule et même chose sont différents d'une publication à l'autre. Aussi, les données rapportées ne sont pas forcément les mêmes d'une publication à l'autre, p. ex. pour la composition chimique des rations, certaines publications rapportent la teneur en matières azotées totales, d'autres rapportent la teneur en énergie métabolisable. De plus, il existe de nombreuses données manquantes par rapport aux besoins des chercheurs, p. ex. les caractéristiques des animaux et l'alimentation (quantités ingérées, composition chimique des rations).

Pour l'administration du système, les accès des utilisateurs étant limités, il est possible de donner un accès à un utilisateur impliqué dans un des projets de recherches sous tendus par Nutriflux et de mettre en place une procédure d'authentification pour cet utilisateur. Pour la saisie, afin d'accélérer la saisie des tableaux de données des publications, il est possible de les saisir sous Excel (Microsoft-Corporation 2016) et de les importer ensuite dans Nutriflux. Pour la caractérisation des aliments des publications, qui sert à estimer les données manquantes de composition chimique des aliments et des rations nécessaires aux chercheurs, nous utilisons un système externe, le système INRAE d'alimentation pour les ruminants, p. ex. INRA 2018. L'extraction des données est modulaire, p. ex. on doit pouvoir extraire I) seulement les mesures sur les animaux ou II) les mesures sur les animaux et l'alimentation. On doit aussi pouvoir faire des choix sur les animaux, p. ex. sur le type d'animal et/ou le stade physiologique et aussi sur les variables mesurées sur les animaux.

### **Le modèle relationnel et l'homogénéisation des données**

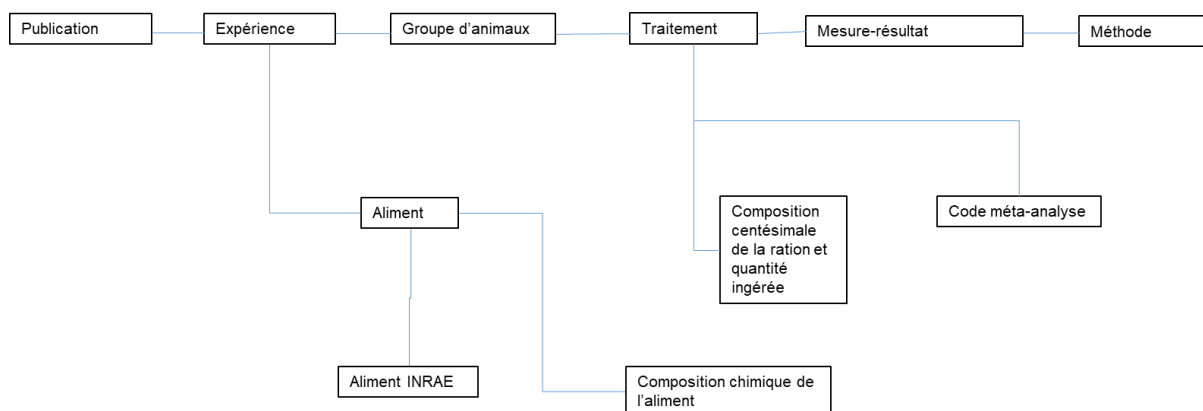
Afin de modéliser le système d'information et surmonter les différentes difficultés techniques et contraintes, nous avons, tout d'abord, étudié des publications représentatives des trois bases. Puis, forts de l'expérience antérieure sur une première version de Flora (Vernet & Ortigues-Marty 2006), un modèle relationnel a été conçu pour Nutriflux en utilisant la méthode Merise ([http://fr.wikipedia.org/wiki/Merise\\_\(informatique\)](http://fr.wikipedia.org/wiki/Merise_(informatique)), consultation le 24/03/2020).

Nous avons conçu un système qui permet de recevoir les données de tout type d'expérience et tout type de données mesurées par rapport à des domaines de recherche différents. Grâce aux référentiels mis en place, il est possible d'homogénéiser toutes ces données. Par exemple, dans Alicar, nous avons retenu pour une variable le nom tissu adipeux sous cutané, %, équivalent à subcutaneous fat, % en anglais comme nom standard valide pour toutes les appellations désignant ce même tissu. Pour permettre des extractions bien ciblées, les mesures sont identifiées de façon fine, par le type de mesure, le compartiment anatomique, la matrice de mesure

## Le Cahier des Techniques de l'Inra 2020 (99)

et l'unité. Il est aussi possible d'estimer les données nécessaires et manquantes de composition chimique des aliments. Pour cela, nous faisons appel à un système extérieur à Nutriflux : le système INRA d'alimentation des ruminants (p. ex. INRA 2018) qui rapporte les compositions chimiques de nombreux aliments obtenues chacune sur un grand nombre d'échantillons. Dans ce système, il existe des aliments équivalents en nature à ceux rencontrés dans les publications internationales de Nutriflux.

Le modèle physique de données (MPD) respecte les trois premières formes normales (Codd 1972). La figure 1 présente le modèle relationnel simplifié de Nutriflux. Cinquante règles de gestion ont été rédigées pour la saisie des données dans les différentes tables et la caractérisation des aliments. Le développement de Nutriflux est réalisé par Matthieu Reichstadt, ingénieur en informatique, pour la base de données en MySQL (<https://fr.wikipedia.org/wiki/MySQL>, consultation le 18/03/20) et pour les interfaces utilisateurs en PHP (<https://fr.wikipedia.org/wiki/PHP>, consultation le 18/03/20).



**Figure 1. Le modèle relationnel simplifié de Nutriflux**

*Publication : table rassemblant toutes les publications, Expérience : table rassemblant toutes les expériences, Groupe d'animaux : table rassemblant tous les groupes d'animaux, Traitement : table rassemblant tous les traitements expérimentaux, Mesure-résultat : table rassemblant toutes les mesures-résultats, Méthode : table rassemblant toutes les méthodes de mesure, Aliment : table rassemblant tous les aliments, Composition centésimale de la ration et quantité ingérée : table rassemblant la composition centésimale des rations en aliments et les quantités ingérées par l'animal, Code méta-analyse : table rassemblant tous les codes de méta-analyse déjà créés, Aliment INRAE: table rassemblant les aliments du système INRAE pour ruminants (p.ex. INRA 2018) correspondant en nature aux aliments des publications, Composition chimique des aliments : table rassemblant toutes les valeurs de composition chimique des aliments des publications.*

### Les processus

Les processus comprennent l'accès et l'authentification des utilisateurs, la saisie des données, la validation des données, la caractérisation des aliments et l'extraction (Figure 2).

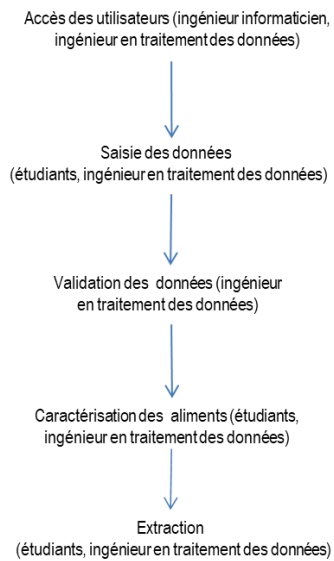


Figure 2 Les différents processus et les acteurs impliqués

L'accès des utilisateurs à Nutriflux se fait par une interface Web. Leur authentification se fait par login/mot de passe LDAP. Les droits d'accès sont délivrés pour chaque utilisateur à une base de données et enregistrés dans la table Administrateur.

La saisie des données se fait dans un ordre bien déterminé (Figure 3). Elle se fait grâce à des formulaires mais aussi par l'intermédiaire de fichiers Excel (Microsoft-Corporation 2016). Ces derniers doivent être au format csv et sont importés grâce à une procédure spécifique d'importation. Pour la mise en œuvre, l'utilisateur appelle la procédure à partir du menu utilisateur. Puis, il choisit le fichier Excel (Microsoft-Corporation 2016) et la table Nutriflux de destination ainsi que le format du fichier Excel (csv, Microsoft-Corporation 2016). Puis, la procédure propose sous la forme d'un tableau les colonnes du fichier Excel (Microsoft-Corporation 2016) et les champs de la table Nutriflux de destination. L'utilisateur met en relations au moyen de listes déroulantes les colonnes du fichier Excel (Microsoft-Corporation 2016) et les champs de la table de destination. Puis, l'utilisateur valide ses choix. L'importation a lieu. Enfin, la procédure retourne le nombre de lignes importées.

## Le Cahier des Techniques de l'Inra 2020 (99)

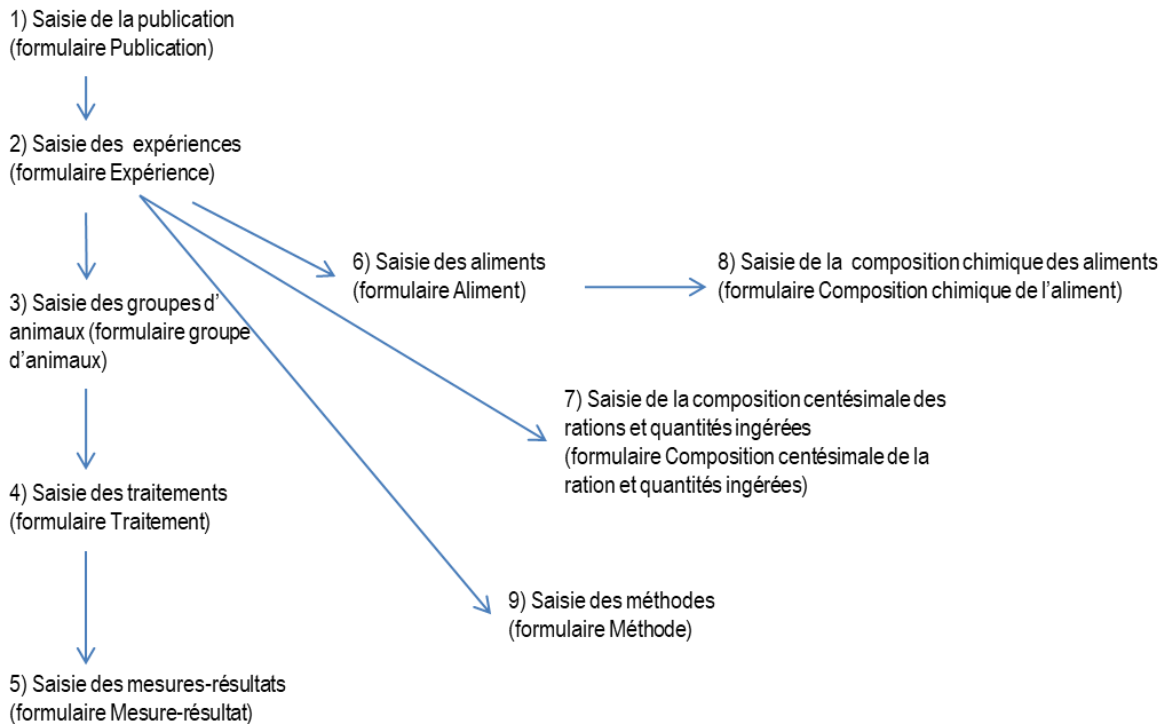


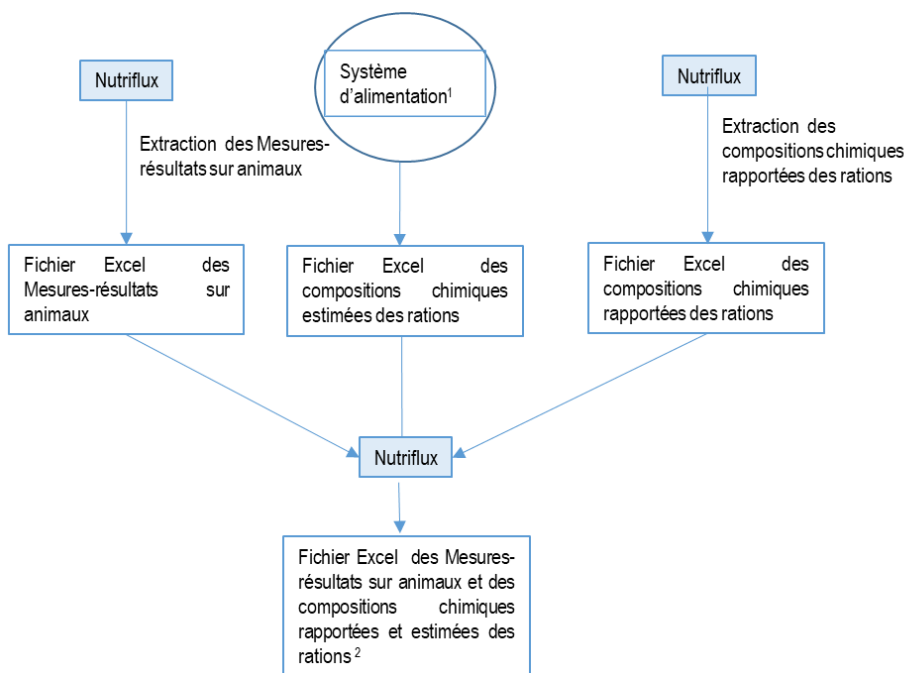
Figure 3. Déroulement d'une saisie typique dans Nutriflux

Par souci de simplification, certaines tables ont été conçues pour recevoir des informations de natures différentes. C'est le cas de la table Composition centésimale de la ration et quantités ingérées qui peut recevoir i/ des compositions centésimales de la ration en aliments et ii/ des quantités de ration ingérées par l'animal. Des données de référence ont été créées afin d'harmoniser la saisie et éviter les incohérences. Elles concernent en particulier les animaux, les variables et les méthodes de mesure ; elles sont le résultat d'un travail de recherche important de la part d'un chercheur (Isabelle Ortigues-Marty), des étudiants en thèse et de l'ingénieur en traitement des données.

La validation des données se fait de façon manuelle i) à la fin de la saisie d'une publication et ii) hors Nutriflux, de façon automatique, au moment de l'analyse des données par méta-analyse et plus particulièrement du repérage des données éloignées et/ou aberrantes (Sauvant *et al.* 2008).

La caractérisation des aliments est faite en faisant appel à un système externe à Nutriflux : le système INRAE d'alimentation pour les ruminants, p. ex. INRA 2018. Elle est faite actuellement à l'aide d'une feuille de calcul sous Excel (Microsoft-Corporation 2016) externe à Nutriflux. Elle consiste à choisir des aliments INRAE proches en nature de ceux de la publication et à ramener leur composition chimique dans la feuille de calcul. Pour la validation, les valeurs estimées pour la ration à l'aide des aliments INRAE, (p. ex. la teneur en Matières azotées totales), sont comparées aux valeurs rapportées dans la publication. On cherche à minimiser l'écart entre les valeurs estimées et les valeurs rapportées. Le résultat de la caractérisation des aliments est alors saisi dans la table Aliment INRAE pour les calculs ultérieurs de composition chimique des rations.

Les extractions permettent de répondre à des questions de recherche bien précises, p. ex. Quelle est la valeur des flux sanguins de nutriments dans le tube digestif et le foie des ruminants et quelles sont les caractéristiques de leur alimentation ? (Loncke *et al.* 2009). Elles sont modulaires dans le sens où on peut sortir plusieurs types d'informations, p. ex. les informations sur les animaux et/ou les informations sur l'alimentation. Il est possible de faire des choix sur les caractéristiques des animaux et sur toutes les composantes des variables mesurées sur les animaux (type de mesure, compartiment anatomique, matrice, unité). La figure 4 donne un exemple de séquence d'extraction.



1 : p. ex. INRA 2018.

2 : ce fichier contient aussi les identifiants, la description des animaux et des rations et les quantités ingérées.

Figure 4. Exemple de séquence d'extraction de Nutriflux

La sortie rapporte p. ex. les identifiants, la description des animaux, les mesures résultats sur les animaux et l'alimentation. Il y a une colonne par variable et une ligne par traitement expérimental. Les mesures-résultats sont présentées par bloc (un bloc par variable). L'identification de la colonne est faite de façon complète dans l'en-tête de la colonne.

### La valorisation des données

Elle est effectuée hors Nutriflux. Elle se fait sur un jeu de données (ensemble de publications choisies et triées selon des critères bien précis) à partir des tableaux de données obtenus par extraction. Les études sont faites par méta-analyse (Sauvant *et al.* 2008). Elles aboutissent à des publications scientifiques comme p. ex. les



## Le Cahier des Techniques de l'Inra 2020 (99)

réponses des flux sanguins de nutriments énergétiques mesurées/déterminés au niveau de certains organes chez des ruminants à différents stades physiologiques aux variations de l'alimentation (Loncke *et al.* 2009).

### Conclusion

Nutriflux présente plusieurs intérêts. Il permet de recevoir tous les types de publications, quel que soit le plan d'expérience et tout type de données issues de domaines de recherche différents. Grace aux référentiels mis en place, il permet de résoudre l'hétérogénéité des publications. Il permet une description fine des animaux, des aliments, des variables et des méthodes de mesures. Il apporte une structuration forte des données. La saisie est facilitée par une interface utilisateur et la possibilité d'importer directement des tableaux de données Excel (Microsoft-Corporation 2016) au format csv. Il est possible d'estimer les données manquantes nécessaires aux chercheurs en faisant appel au système INRAE d'alimentation des ruminants. De plus, les extractions sont ciblées, modulaires et font appel à des filtres pour répondre à des questions de recherche précises. Nutriflux a déjà été utilisé dans le cadre d'au moins huit publications internationales et cinq thèses de doctorat.

Cet article est publié sous la licence Creative Commons (CC BY-SA).



<https://creativecommons.org/licenses/by-sa/4.0/>

Pour la citation et la reproduction de cet article, mentionner obligatoirement le titre de l'article, le nom de tous les auteurs, la mention de sa publication dans la revue « Le Cahier des Techniques de l'Inra », la date de sa publication et son URL).

### Bibliographie

Al Jammam M, Agabriel J, Vernet J & Ortigues-Marty I (2016) The chemical composition of carcasses can be predicted from proxy traits in finishing male beef cattle: A meta-analysis. *Meat Science* **119** : 174-84.

Codd EF (1972) Further normalization of the data base relational model. *Data base systems* : 33-64.

INRA (2018) *Alimentation des ruminants*. Edition QUAE, 78026 Versailles Cedex.

Loncke C, Ortigues-Marty I, Vernet J, Lapierre H, Sauvant D & Nozière P (2009) Empirical prediction of net portal appearance of volatile fatty acids, glucose, and their secondary metabolites ( $\beta$ -hydroxybutyrate, lactate) from dietary characteristics in ruminants: A meta-analysis approach. *Journal of animal science* **87** : 253-68.

Microsoft-Corporation (2016) *Microsoft Excel*.

Mourot BP, Gruffat D, Durand D, Chesneau G, Mairesse G & Lebert A (2013) *Prediction of polyunsaturated fatty acid content in bovine muscle [abstract]*. In: Proceedings of the 64th Annual Meeting of the European Federation of Animal Science, Nantes, France. Wageningen Academic Publishers.

Sauvant D, Schmidely P, Daudin JJ & St-Pierre NR (2008) Meta-analyses of experimental data in animal nutrition. *Animal* **2** : 1203-14.

Vernet J & Ortigues-Marty I (2006) Conception and development of a bibliographic database of blood nutrient fluxes across organs and tissues in ruminants: data gathering and management prior to meta-analysis. *Reproduction Nutrition Development* **46** : 527-46.

## **Remerciements**

Les auteurs remercient MM. Pierre NOZIERE et Michel VERGER pour leurs précieux conseils au moment de l'écriture de cet article ainsi que le relecteur, Alexandre JOURNAUX, pour ses remarques avisées.