

# LES ONTOLOGIES DE RÉFÉRENCE AU SERVICE DE L'INTEROPÉRABILITÉ EN SANTÉ

L'informatique est devenue un outil incontournable dans tous les secteurs et la santé ne fait pas exception. Toutefois, du fait de la multiplicité de terminologies, du manque de standards pour coder des signaux, de différences entre les plateformes informatiques, etc., on rencontre d'importantes difficultés lorsque l'on veut partager des ressources informationnelles (tant des données que des modèles). Ce problème d'interopérabilité est lié directement à l'informatisation : si les humains n'éprouvent pas de réelles difficultés à résoudre les ambiguïtés de langage, il n'en est pas de même pour les ordinateurs, qui ont besoin de contextes sémantiquement explicites pour pouvoir explorer les différents cyberspaces qui touchent, notamment, aux domaines de la santé. Nous présentons ici quelques efforts réalisés dans ce sens pour tenter de favoriser l'interopérabilité grâce à l'adoption d'un ensemble commun d'ontologies de référence. La mise en place de cette stratégie et le succès de sa pénétration dépendront de la mise en place d'outils et d'interfaces qui permettront aux utilisateurs de s'affranchir de codes informatiques par trop rébarbatifs, nous présenterons aussi certains de ces outils en cours de mise en place.

Par **Stephen Randall THOMAS\***

---

\* Ph.D., DR CNRS - IR4M UMR8081, CNRS Université Paris-Sud, Orsay.

## INTRODUCTION

L'informatique est devenue un outil incontournable dans tous les secteurs, et la santé ne fait pas exception. Si la vision du robot médecin appartient encore au domaine de la science-fiction, il n'en reste pas moins que l'informatique rend bien des services aux médecins et aux services hospitaliers. Pour nous en convaincre, nous citerons quelques exemples faisant l'actualité : le Dossier Médical Personnel (DMP) (1), un dispositif qui se met progressivement en place et qui mettra à la disposition de tout citoyen son historique médical complet ; la télémédecine (ou la e-Santé) devient opérationnelle pour la surveillance à domicile de patients âgés et/ou à mobilité réduite ; et la chirurgie assistée par ordinateur est devenue une réalité pour de nombreux types d'intervention. Dans les domaines de l'industrie biomédicale, la modélisation fait partie intégrante du développement de nouveaux médicaments et constitue même le cœur de certains appareils implantés, tels les stimulateurs cardiaques (*pacemakers*) adaptatifs, qui tiennent compte du niveau d'activité du patient pour régler en temps réel la fréquence cardiaque.

Toutefois, dans ces domaines et dans leurs domaines annexes, du fait de la multiplicité des terminologies, du manque de standards en matière d'unités de mesure, des différences de plateformes informatiques (notamment), on rencontre d'importantes difficultés à partager des ressources informationnelles tant des données que des modèles (les DMR, pour *Data and Model Resources*), par exemple des informations épidémiologiques, des données anonymisées, des outils d'analyse d'images, des codes de modèles prédictifs, et d'autres encore. Ce problème d'interopérabilité est lié directement à l'informatisation. En effet, si les humains n'ont pas (ou n'ont que peu) de difficulté à résoudre les ambiguïtés de langage, il n'en est pas de même pour les ordinateurs, qui ont besoin de contextes sémantiquement explicites pour explorer les différents cyberspaces touchant aux domaines de la santé. Nous présenterons ici certains efforts allant dans ce sens qui tentent de favoriser l'interopérabilité grâce à l'adoption d'un ensemble commun d'ontologies de référence. La mise en place de cette stratégie et le succès de sa pénétration dépendront d'outils et d'interfaces qui permettront aux utilisateurs de s'affranchir de codes informatiques par trop rébarbatifs. Nous présenterons aussi des outils de cette nature actuellement en cours de mise en œuvre.

(1) <http://www.dmp.gouv.fr>

INTEROPÉRABILITÉ OU INTÉGRATION ?  
LE CŒUR DE LA VISION DU PROJET  
PHYSIOME/VPH

Plusieurs communautés d'acteurs ont besoin de disposer d'intégration ou d'outils d'interopérabilité. C'est le cas pour les acteurs du VPH (2) (*Virtual Physiological Human*, la partie européenne du Projet Physiome), les chercheurs en biologie systémique, en physiologie, en informatique biomédicale, en pharmacologie, en systèmes de santé individuels, en « e-infrastructure » des sciences de la vie...

Ils ont besoin d'intégration, certes, mais pour gérer quoi, au juste ? Ils utilisent (et produisent) des données scientifiques/expérimentales pour paramétrer/calibrer/valider des modèles, mais aussi des données cliniques destinées à être stockées dans différents systèmes, dans différents hôpitaux, dans différents États membres de l'Union européenne et dans différentes bases de données destinées à la recherche...

Ils utilisent aussi des connaissances spécifiques à tel ou tel patient et en lien avec les connaissances spécifiques à leurs domaines. Ils utilisent des informations sur les différentes parties et les différents processus physiologiques du corps humain, et ce, dans une compréhension systémique de la pathophysiologie. Ils ont besoin de connaissances acquises sous forme numérique au travers de métadonnées, d'ontologies et de modèles. Ils ont besoin d'avis (bonnes pratiques, etc.) émis par les laboratoires de recherche et par la pratique clinique : ceux-ci seront formalisés dans des guides, des standards et des protocoles qui seront utilisés pour promouvoir une traduction des connaissances fondamentales et des modèles intégrés en bénéfices cliniques.

Pour rendre toutes ces ressources accessibles aux différentes communautés, il existe de nombreux vecteurs de diverses natures : des langages *Markup* pour le partage de modèles (comme SBML (3), CellML(4) et JSim's MML (5)), des standards communs pour les formats de données (pas seulement des images, mais aussi des signaux et des *workflows*), des annuaires de données et de modèles tous ayant été produits par des groupes de travail (par exemple, autour du « patient numérique »). Les ontologies sont le « liant » de cet ensemble : mises en cohérence, elles constituent le langage commun qui permet que ces espaces puissent former une unité explorable informatiquement.

(2) HUNTER (P) & al., "A vision and strategy for the virtual physiological human: 2012 update", *Interface Focus* 3(2), 2013 (Doi: 10.1098/rsfs.2013.0004).

(3) <http://sbml.org>

(4) <http://www.cellml.org>

(5) [http://www.physiome.org/jsim/docs/MML\\_Intro.html](http://www.physiome.org/jsim/docs/MML_Intro.html)

Ici, nous nous concentrerons sur un sujet particulièrement pertinent pour les domaines de la biologie computationnelle que nous avons mentionnés plus haut, celui d'établir des liens entre les paramètres et les variables de modélisations mathématiques de structures et de fonctions biologiques : les données cliniques, les données biologiques et les données expérimentales (ce que l'on appelle des DMR, pour *Data and Model Resources*).

Quelles sont les ontologies nécessaires pour mener à bien cette tâche ?

Il nous en faut pour pouvoir représenter des éléments tels que l'anatomie, les espèces et les sous-espèces d'animaux, les types de cellules, les fonctions (flux, concentrations, indicateurs cliniques...), la terminologie (définitions de paramètres de mesure, d'espèces chimiques...), les techniques et les conditions expérimentales, les pathologies liées, les fonctions moléculaires, etc.

Certaines de ces ontologies existent déjà (6), et d'autres sont en cours de développement.

#### Le projet RICORDO et ses outils

Le projet RICORDO (7) réunit plusieurs acteurs industriels (pharmacie, dispositifs médicaux...) et académiques pour développer des outils d'interopérabilité en anatomie et en physiologie. RICORDO est repris dans le projet DDMoRe (*Drug Disease Modeling Resource* (<http://www.ddmore.eu/>) qui s'inscrit dans le cadre de l'IMI (*Innovative Medicines Initiative* – <http://www.imi.europa.eu/content/ddmore>).

RICORDO pose cette question (à laquelle il s'efforce de répondre) : « Comment la documentation pour des projets et des bases de données est-elle réalisée actuellement, et comment peut-on faire mieux ? ».

Chaque projet individuel désigne/nomme/décrit les choses à sa façon, en utilisant généralement du texte libre de façon non uniforme entre les différents groupes ou entre les différentes communautés de recherche.

Ces libellés ne sont donc pas utilisables dans des recherches automatisées (recherche documentaire sur papier, recherches sur le *Web*...). Souvent, les terminologies sont « à plat », elles restent non structurées. Quelquefois des vocabulaires et des ontologies contrôlés (des CVO, pour *Controlled Vocabulary and*

*Ontology*) sont utilisés, mais tout le monde n'utilise pas les mêmes, et différentes ontologies se chevauchent. En même temps, la couverture des domaines est incomplète : il faut remplir les « trous ».

L'idée adoptée par RICORDO est de développer une annotation commune des métadonnées standards, pour utiliser les CVOs avec un minimum de recouvrements, après avoir choisi un ensemble d'ontologies de référence. Ensuite, il faut rendre les métadonnées interopérables sémantiquement, ce qui est difficile pour des concepts aussi complexes que ceux des processus physiologiques : il faut donc les définir à partir des termes figurant dans les ontologies de référence. Il est nécessaire également de construire un format commun pour les annotations des DMR (*Data & Model Resources*) (par exemple (8) : MIRIAM, RDF, OWL...). Il faut aussi fournir, sur le *Web*, une boîte à outils permettant la construction de termes complexes, l'annotation des données et le stockage de métadonnées dans un répertoire/catalogue.

Un point important est la possibilité d'une dé-corrélation entre, d'une part, les métadonnées d'annotation et, d'autre part, les ressources correspondantes. Le cas échéant, les métadonnées d'annotation se limitent à l'indication de ce qui existe dans les ressources, sans que les ressources elles-mêmes soient forcément directement accessibles dans la DMR elle-même. Les communautés cliniques se sont mises d'accord pour adopter cette position.

La Figure 1 (de la page suivante) présente cette structuration sous l'intitulé de « triplet d'annotation ». Il y a en effet trois niveaux : les données et les modèles (DMRs), les annotations et, enfin, les ontologies.

Les DMRs comprennent les ressources scientifiques et constituent les « observations » vues du niveau annotations/métadonnées. De l'autre côté, les ontologies fournissent à ces métadonnées le sens de ce qu'elles annotent. En résumé, les annotations fournissent un lien entre les DMR et leur signification médicale, qui est basée sur les ontologies. Un exemple est présenté ensuite pour illustrer la question de l'annotation de termes complexes. Le schéma (voir la Figure 1) en est reproduit ci-après (d'après DE BONO & *al.*, 2011, BMC Research Notes 4.313).

Deux ontologies sont mobilisées dans cet exemple, celle de l'anatomie et celle des qualités biologiques (taille, vitesse, etc.). La construction des termes complexes à partir de ces termes de base est une particularité de RICORDO. Par ailleurs, on voit que les relations entre termes de base permettant de représenter le terme complexe sont elles aussi de différentes natures correspondant à une variété de sens. Les anno-

(6) Voir le "OBO Foundry" of Open Biological and Biomedical Ontologies  
<http://www.obofoundry.org>

(7) DE BONO & *al.*, *The RICORDO approach to semantic interoperability for biomedical data and models: strategy, standards and solutions*, BMC Res Notes 4: 313, 2011 (DOI: 10.1186/1756-0500-4-313) et <http://www.ricordo.eu>

(8) DE BONO (B.), HOEHNDORF (R.), WIMALARATNE (S.), GKOUTOS (G.) & GRENON (P.), *The RICORDO approach to semantic interoperability for biomedical data and models: strategy, standards and solutions*, BMC Res Notes, 4, 313, 2011 (Doi: 10.1186/1756-0500-4-313).

tations sont en accès public : toute personne peut faire des requêtes à ce niveau. En revanche, tous les DMR ne sont pas en libre accès. Ces ressources sont de différentes natures. Dans l'exemple présenté, on trouve ainsi : un modèle de l'infarctus du myocarde, des données concernant l'expression du gène cardiaque, une liste de références des termes de la pathologie, etc.

Les principales ontologies de référence sont les suivantes : FMA (*Foundational Model of Anatomy*), GO (*Gene Ontology*), ChEBI (*Chemical Entities of Biological Interest*), PATO (*Ontology of Phenotypic Qualities*), *OBO Relationship Ontology* et *OBO Cell Type Ontology*.

## D'AUTRES INITIATIVES D'INTEROPÉRABILITÉ

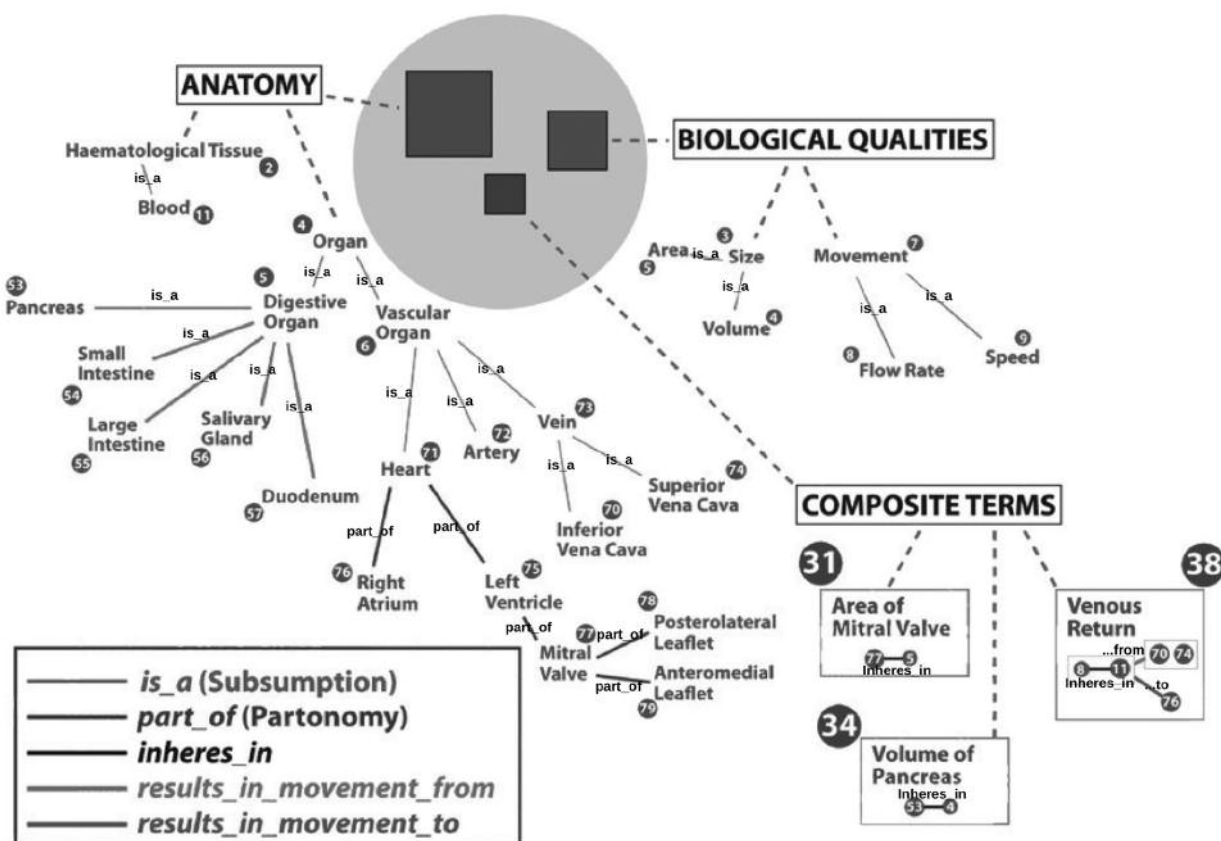
Il existe, bien sûr, de nombreuses autres initiatives pour améliorer l'interopérabilité. J'en citerai brièvement trois, qui sont complémentaires à l'initiative RICORDO, à savoir : l'interface ApiNATOMY, l'ontologie de simulation en physiologie humaine

HuPSON et les *mappings* génotype-phénotype cGP ("*causally cohesive Genotype-Phenotype*" mappings)

ApiNATOMY (9) fournit un outil de cartographie interactive liant l'anatomie et la physiologie multi-échelles. Sa méthodologie s'appuie sur la topologie des graphes d'ontologies d'anatomie, surtout la FMA (*Foundational Model of Anatomy*). Concrètement, elle crée des cartes arborescentes navigables en 3D (voir la Figure 2 de la page suivante) représentant les différentes parties du corps humain, avec les méta-datas sémantiques liées aux ontologies (pour plus de détails, le lecteur est invité à se reporter au site *Web* concernant cet outil et à sa vidéo) (10). Le but, à terme, est de fournir des interfaces (*Graphical User Interfaces*) et des tableaux de bord pour la gestion de ressources biomédicales et l'analyse de données.

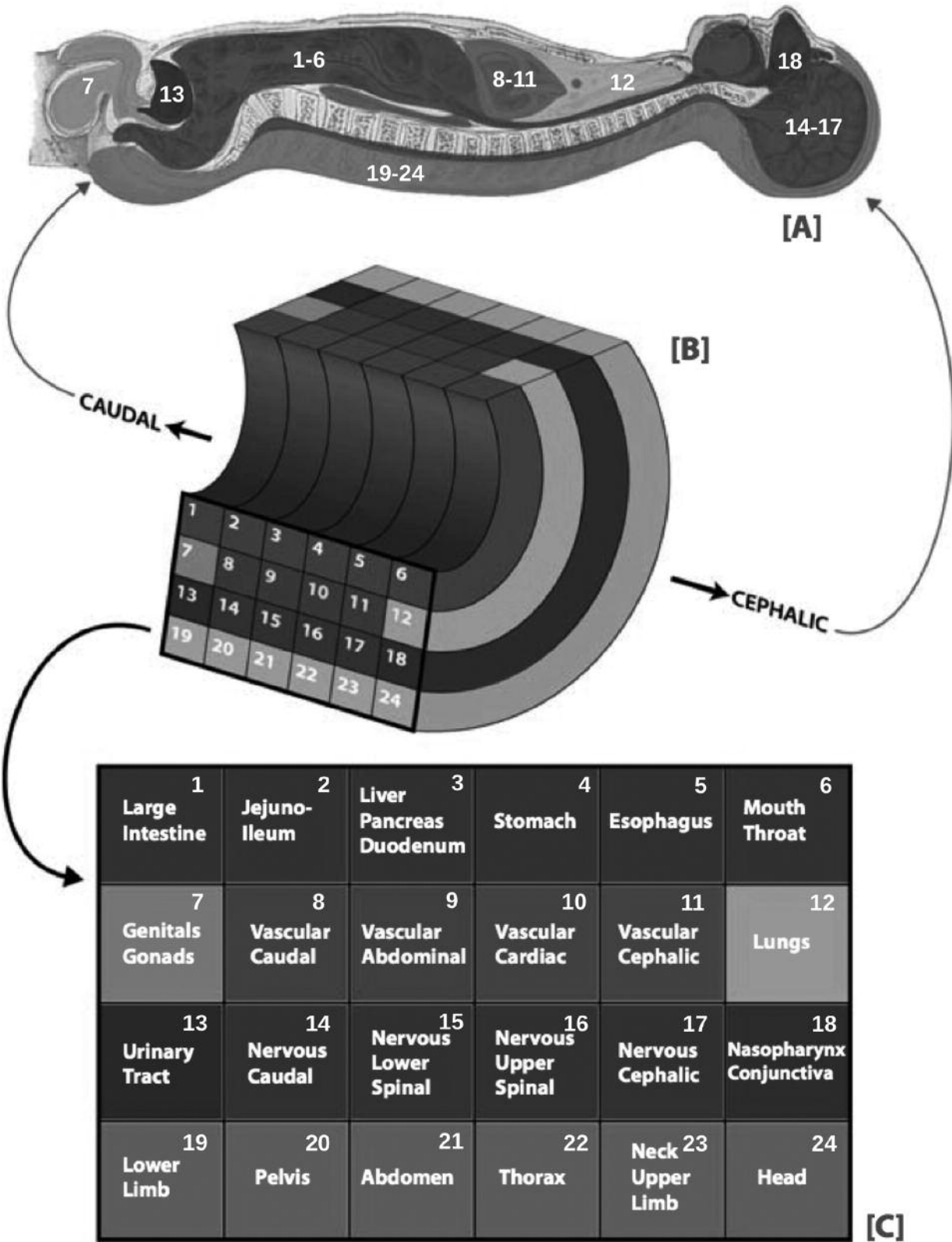
(9) DE BONO (B.), GRENON (P.) & SAMMUT (S.J.), "ApiNATOMY: a novel toolkit for visualizing multiscale anatomy schematics with phenotype-related information", *Hum Mutat*, 33(5), pp. 837-848, 2012.

(10) <https://sites.google.com/site/apinatomy/>



**Figure 1** : Une illustration de l'utilisation des ontologies de référence pour représenter des termes (ou expressions) décrivant des concepts complexes/composés. On voit que l'ontologie des particularités biologiques (*Biological Qualities*) n'utilise que la relation de subsumption.

L'ontologie d'Anatomie se sert aussi de la relation de partonomie. Notez que, si les termes composés ont chacun un identifiant unique, ils se réfèrent quand même explicitement aux URI (*Unique Resource Identifiers*) des ontologies standards de référence. Dans le projet RICORDO, les ontologies de référence et les termes composés sont les uns comme les autres formalisés en OWL (*Web Ontology Language*) (D'après de Bono & al., 2011, BMC Research Notes 4.313).



**Figure 2 :** (A) Une section longitudinale du corps humain montrant les surfaces externes et internes et les organes. Les couleurs sont reprises dans les panneaux (B) et (C).

(B) Un plan du corps humain idéalisé en 24 régions (les “wbAU” – *whole-body Asymmetric Unit*) radiales symétriques. Cette homunculus présente un axe de rotation central longitudinal localisé dans le tractus intestinal et orienté dans la direction céphalo-caudale. Par une opération de rotation autour de cet axe, on représente le plan du corps entier symétriquement, en 3D.

(C) L’unité asymétrique de 24 tuiles décrit des tissus colocalisés par région. Ceux-ci sont généralement impliqués dans des échanges convectifs avec les tissus des tuiles voisines (d’après de Bono et *al.*, 2012).

HuPSON (11) est une ontologie compréhensive qui couvre les besoins des domaines de modélisation et de simulation de la physiologie humaine et qui permet aussi d'incorporer des équations. Ce type d'ontologie manque cruellement pour faciliter le partage de simulations, de modèles, d'algorithmes et d'autres ressources du domaine. Basée sur la *Basic Formal Ontology* [BFO], elle permet d'annoter les expériences de simulation, de récupérer les informations utiles pour la modélisation, d'implémenter l'interopérabilité des approches algorithmiques utilisées en simulation biomédicale, de comparer les résultats de simulations et, enfin, de relier des approches basées sur la connaissance aux approches basées sur la simulation. Les cGP, ou cartes Génotype-Phénotype (12), partent de l'idée que, dans un modèle dynamique bien validé d'un trait physiologique complexe, une part importante de la variation génétique sous-jacente se manifeste à travers les paramètres du modèle. Ces paramètres sont eux-mêmes des traits phénotypiques (voir, par exemple, WANG et al., 2012) (13). La poursuite de cette approche pourrait faciliter une compréhens-

sion causale de la carte génotype-phénotype de traits complexes multi-factoriels, avec l'espoir de retombées importantes pour la technologie phénotypique.

---

## CONCLUSION

Toutes les ressources présentées ici s'appuient sur la communauté Physiome/VPH (14). Le VPH (*Virtual Physiological Human*) trouve son origine dans une succession d'initiatives : IUPS Physiome (*International Union of Physiological Sciences*) est issu de la physiologie et des collaborations de recherche dans ce domaine autour de modèles et de partage de données expérimentales, il vise à rendre cette recherche plus collaborative. Un projet STEP (*Seeding The EuroPhysiome*, FP6) a produit une *roadmap* européenne sur le sujet. Ensuite, ont été lancés les projets VPH (FP7) avec l'objectif d'amener ces réflexions vers la clinique : le VPH-NoE (réseau d'excellence) et les projets VPH-I (I comme Initiatives : une cinquantaine de projets appliqués à différentes maladies). Au travers de ces projets, l'Union européenne a fourni l'essentiel du financement au niveau mondial. Enfin, l'Institut VPH (15) a été créé. Il est autofinancé et son objectif est de prendre la relève de façon pérenne.

---

(11) <http://www.scai.fraunhofer.de/en/business-research-areas/bioinformatics/downloads.html> (fichiers OWL) et <http://bishop.scai.fraunhofer.de/scaiview/> (navigateur).

(12) GJUVSLAND (A. B.), VIK (J.O.), BEARD (D.A.), HUNTER (P.J.) & OMHOLT (S.W.), "Bridging the genotype-phenotype gap: what does it take?", *The Journal of Physiology*, 591(Pt 8), pp. 2055-2066, 2013.

(13) WANG (Y.), GJUVSLAND (A. B.), VIK (J.O.), SMITH (N.P.), HUNTER (P.J.) & OMHOLT (S.W.), "Parameters in dynamic models of complex traits are containers of missing heritability", *PLOS Computational Biology*, 8 (4): e1002459, 2012.

---

(14) HUNTER (P.J.) & al., "A vision and strategy for the virtual physiological human, 2012 update", *Interface Focus*, 3(2), 2013 (Doi: 10.1098/rsfs.2013.0004).

(15) <http://www.vph-institute.org/>