

Du bon usage d'Adonis

RunMyCode

La recherche académique en économie et gestion à portée de clic

RunMyCode est un projet de création et de déploiement à l'échelle internationale d'une plateforme de mise à disposition de codes de calculs scientifiques dans le domaine de l'économie et de la gestion. Ce projet repose sur le concept novateur de « site compagnon » de publication scientifique. Un site compagnon est un site Internet à partir duquel les utilisateurs peuvent exécuter en ligne et de façon interactive des programmes informatiques associés à une publication à partir de leurs propres données et/ou de leurs propres choix de paramètres. Ils utiliseront pour ce faire un simple navigateur internet. Les calculs nécessaires à la restitution des résultats seront effectués sur des serveurs de calcul de la grille de services du TGE Adonis ([Centre de calcul de l'IN2P3](#)). L'utilisation du site compagnon ne requiert donc aucune compétence particulière en informatique. Le site compagnon apporte ainsi une dimension nouvelle aux actions de recherche en permettant, au-delà de la publication scientifique – objet fixé dans le temps – de donner accès aux éléments ayant permis la recherche : les données et les programmes informatiques. Le projet part du constat suivant. La production scientifique dans les domaines de l'économie et de la gestion est énorme et en plein essor. En parallèle, le nombre d'utilisateurs potentiels (étudiants, chercheurs, entreprises, administrations) de nouveaux traitements et de nouvelles méthodes issues de la recherche croît également de façon exponentielle. Or, dans la plupart des contributions scientifiques, une implémentation informatique est rendue nécessaire, particulièrement pour les méthodes mathématiques, statistiques ou économétriques. Dès lors, de nombreux blocages viennent entraver la diffusion, la réplication et l'appropriation par la communauté de ces nouvelles méthodes. Ces problèmes sont souvent liés à la difficulté, pour les utilisateurs, de programmer ou de modifier le programme mis à disposition par les auteurs, afin de l'adapter à leurs propres données et à leurs propres choix de paramètres. Cet état de fait ralentit la diffusion des idées scientifiques notamment en dehors de la sphère académique. Il limite en outre la vérifiabilité et la réplication des résultats numériques décrits dans les publications scientifiques. Dans un article récent, Victoria Stodden, professeur de statistique à l'Université de Columbia à New York, insiste sur l'importance de la notion de réplicabilité des résultats des publications scientifiques dans l'ensemble des sciences dites computationnelles.

"[Reproducibility is] essential to addressing the growing credibility crisis in science. It is impossible to believe most of the computational results presented at conferences and in published papers today. Even mature branches of science, despite all their efforts, suffer severely from the problem of errors in final published conclusions. Traditional scientific publication is incapable of finding and rooting out errors in scientific computation, and standards of verifiability must be developed".

Victoria Stodden,
"Trust Your Science? Open Your Data and Code", *Amstat News*, Août 2011

Le concept de site compagnon constitue ici une réponse particulièrement adaptée tant en termes de diffusion que de réplicabilité des résultats scientifiques. Il peut être perçu comme un premier pas vers le concept générique de « publication exécutable » (*executable paper*) dans le domaine de l'économie-gestion.

Le service *RunMyCode* sera proposé dans un premier temps dans le domaine de l'économie et des sciences de gestion. Toutefois, le concept de site compagnon pourra être ensuite décliné dans d'autres disciplines où la programmation informatique tient une place centrale dans la démarche scientifique. On peut citer les domaines de l'ingénierie linguistiques, des statistiques, des mathématiques appliquées, de la bio-statistique, etc.

Objectifs

Le premier objectif du projet *RunMyCode* est de permettre aux chercheurs français et étrangers de diffuser de façon rapide à l'échelle internationale les résultats de leurs recherches à partir d'un site centralisant des ressources informatiques de calcul et de traitement, disponibles en ligne. Ce service devrait notamment permettre d'augmenter le potentiel de citations de certains articles scientifiques.

Le second objectif est d'offrir à une très large communauté d'utilisateurs, qui dépasse largement la seule communauté académique, la possibilité d'utiliser, en ligne et de façon interactive, des méthodes de calcul et de traitement scientifiques parmi les plus récentes pour leurs propres données et avec leurs propres choix de paramètres. Cette utilisation n'était jusque là pas possible pour l'ensemble des utilisateurs qui ne possédaient pas les compétences informatiques pour implémenter ces méthodes sur des progiciels dédiés.

Le troisième objectif est de fournir à la communauté académique (chercheurs, éditeurs, arbitres, etc.) un moyen de pouvoir répliquer les résultats scientifiques et d'en démontrer ainsi l'exactitude, la fiabilité et la portée générale.

Fonctionnement

Un site compagnon est un site Internet à partir duquel les utilisateurs peuvent exécuter en ligne des programmes informatiques associés à une publication scientifique. Le site permettra de télécharger le programme informatique et un jeu de données de démonstration proposé par l'auteur. En ce sens, le site compagnon peut être conçu comme une vitrine ou un écrin de la publication scientifique, offrant à l'utilisateur la possibilité de télécharger les ressources liées (codes et données, si ces dernières ne sont pas soumises à des droits particuliers) et/ou d'utiliser directement ces ressources afin d'en évaluer la robustesse, la performance et la validité.

Une des innovations du projet consiste en la réalisation d'une application informatique qui permet à l'auteur de créer en ligne son propre site compagnon de façon complètement personnalisée. Dans le cadre d'un processus en six étapes, l'auteur fournit différentes informations concernant sa publication, ses co-auteurs etc. ainsi qu'une description formalisée des variables et des paramètres

qui entrent dans son programme informatique (*inputs*). Aucune description des sorties n'est requise : le site compagnon reproduit intégralement le format et la mise en page des résultats que l'auteur aurait obtenu sur son ordinateur personnel. À l'issue du processus, qui ne prend au total que quelques minutes, l'auteur obtient une prévisualisation de son futur site. Cette application développée sous la technologie J2EE a fait l'objet d'un dépôt auprès de l'APP (Agence de Protection des Programmes) et d'un dépôt de brevet aux États-Unis sous l'égide du FIST-CNRS et de la cellule de valorisation de l'InSHS.

Aucun effort particulier de programmation n'est demandé aux auteurs : le procédé *RunMyCode* permet tout simplement d'encapsuler et de transformer en un site exécutable les programmes développés dans le cadre de l'activité habituelle de recherche en vue de la publication.

Le futur site compagnon est alors soumis à différentes phases de validation. La première validation sera réalisée par l'auteur et l'ensemble de ses co-auteurs. Une seconde validation scientifique portera sur la publication et veillera à ce que celle-ci soit en adéquation avec la ligne éditoriale de *RunMyCode*. La dernière validation portera alors sur le code informatique lui-même, notamment sur les performances et la robustesse selon des paramètres différents. Une fois validé, le site est alors déployé en ligne sur la grille de services du TGE Adonis.

Les sites compagnons sont accessibles soit directement sur le web (via un moteur de recherche), soit à partir du portail *RunMyCode*. Ce dernier permet d'effectuer différentes recherches et d'obtenir des informations sur le service. Chaque auteur dispose par ailleurs d'un espace personnalisé recensant l'ensemble de ses productions. Au travers de cet espace, il peut obtenir différentes statistiques relatives à la visibilité de son site compagnon sur Internet : nombre de visites, nombre d'exécutions de son code, nombre de téléchargements, etc.

Structuration du projet

Le projet *RunMyCode* est porté par un ingénieur informaticien CNRS, Yvan Stroppa, et trois professeurs (Gilbert Colletaz, Christophe Hurlin et Christophe Pérignon), appartenant au GREGHEC (HEC Paris, UMR2959) et au LEO (Université d'Orléans, UMR6221). Avec le soutien du TGE Adonis, d'HEC Paris et de l'Université d'Orléans, l'équipe a été renforcée par deux informaticiens de la société ATOS Origin et des stagiaires issus de la licence et du master MIAGE d'Orléans.

Sur les conseils de Philippe Baptiste (directeur de l'Institut des sciences informatiques et leurs interactions, INS2I), une collaboration interdisciplinaire a été initiée avec le LIFO (Laboratoire d'informatique fondamentale d'Orléans) et le LI (Laboratoire d'informatique de Tours). Elle porte notamment sur deux axes : l'ordonnancement et l'optimisation des calculs distribués sur la grille ; l'extension du concept de site compagnon à la recherche opérationnelle.

La phase de développement a duré un peu plus d'une année, mobilisant deux personnes sous la direction technique d'Yvan Stroppa. Durant cette phase, le projet a bénéficié de l'expertise des équipes du TGE Adonis et du Centre de calcul de l'IN2P3, notamment en vue de réaliser le couplage des sites avec la grille de services. Ce partenariat, sous l'égide de Jean-Luc Pinol et de Richard Walter, s'est avéré déterminant dans la phase de développement de ce projet ambitieux. Aujourd'hui, le projet est prêt à passer dans une phase opérationnelle.

Perspectives

Le service *RunMyCode* sera ouvert à la fin du mois de janvier 2012 et accessible directement à l'adresse www.runmycode.org. Après une première période d'évaluation, une campagne de promotion et de communication sera lancée auprès de la communauté scientifique internationale en collaboration avec le TGE Adonis. Des liens seront mis en place avec différents systèmes d'archivage des publications, et notamment Hal-SHS. En outre, des partenariats devraient être établis avec des éditeurs scientifiques et des revues scientifiques afin de diffuser le procédé. Durant cette phase de lancement le service sera également proposé à différentes associations savantes en économie et gestion ainsi qu'à des organisateurs de congrès ou de conférences.

Gilbert Colletaz, LEO UMR6221, Université d'Orléans / CNRS,

Christophe Hurlin, LEO,

Christophe Pérignon, GREGHEC UMR2959, HEC Paris / CNRS,

& Yvan Stroppa, LEO

contact&info

▶ Gilbert Colletaz, LEO

gilbert.colletaz@univ-orleans.fr

▶ Christophe Hurlin

christophe.hurlin@univ-orleans.fr

▶ Christophe Pérignon

perignon@hec.fr

▶ Yvan Stroppa

yvan.stroppa@univ-orleans.fr

▶ Pour en savoir plus

www.runmycode.org

UNE QUESTION / UNE RÉPONSE

Qu'est-ce qu'un site compagnon ?

de Richard Walter, avec la collaboration des responsables du projet RunMyCode

Le principe dit du « site compagnon » est de mettre à disposition des procédures effectives de calculs scientifiques. À partir d'un site internet, les utilisateurs peuvent exécuter en ligne et de façon interactive des programmes informatiques associés à une publication à partir de leurs propres données avec la possibilité d'intervenir sur les paramètres de la modélisation. Ce concept de « site compagnon » rejoint l'idée d'*Executable Paper*^{*}, en mettant à disposition l'ensemble des ressources (codes comme données) associées aux articles, ce qui en change radicalement la lecture. Cette dimension nécessite une centralisation des ressources informatiques.

À titre d'exemple, citons une publication récente^{**} : Robert Engle (prix Nobel d'économie 2001) et Christian Brownless proposent une nouvelle mesure de risque systémique pour les institutions financières, appelée MES (*Marginal Expected Shortfall*). Cette mesure vise à quantifier l'impact d'une faillite potentielle d'une banque sur le système financier dans son ensemble. Rendue nécessaire par la crise de 2008 et la faillite de Lehman Brothers, une telle mesure s'inscrit dans le contexte des nouvelles réglementations bancaires de Bâle III.

Pour publier cet article, les auteurs ont développé des codes sous un logiciel d'économétrie spécialisé (comme Matlab par exemple) afin de rendre opérationnelle la mesure et d'en illustrer différentes propriétés. Ce code est relativement complexe puisqu'il fait notamment appel à une estimation non paramétrique d'une espérance tronquée (*Expected Shortfall*).

Un site compagnon de l'article d'Engle et Brownless, crée à partir du code des auteurs, comporte un formulaire permettant à l'utilisateur d'introduire ses propres données portant sur n'importe quelle banque et sur n'importe quelle période. Le formulaire permet en outre de choisir différentes options concernant la méthode statistique retenue pour calculer la MES. À l'issue des calculs effectués sur le site compagnon, l'utilisateur obtient sur son navigateur internet la prévision (à un horizon fixé) de la mesure de risque systémique de cette institution. Il peut, par ailleurs, vérifier la robustesse des résultats publiés par les auteurs, et ceci sans avoir à modifier le code. Enfin s'il le souhaite, l'utilisateur peut télécharger le code.

Ce concept de « site compagnon » peut être décliné dans d'autres problématiques où l'outil informatique tient une place centrale dans la démarche scientifique. Il est évident dans le domaine des données à référence spatiales, mais aussi des bases d'information, du multimédia et du traitement du texte.

* Voir *Executable Paper Grand Challenge* : <http://www.executablepapers.com>

** Engle, R.F. & Brownless, C.T., (2010) *Volatility, correlations and tails for systemic risk measurement*, New York Stern University

contact&info

► Nadine Dardenne
Chargée de la communication
et de la structuration des réseaux
nadine.dardenne@tge-adonis.fr
► Pour en savoir plus
www.tge-adonis.fr

la lettre de l'INSHS

- ▶ **Directeur de la publication** Patrice Bourdelais
- ▶ **Directeur de la rédaction** François-Joseph Ruggiu
- ▶ **Responsable éditoriale** Sandrine Clérisse com-shs@cncrs-dir.fr
- ▶ **Conception graphique** Sandrine Clérisse & Bruno Roulet, Secteur de l'imprimé PMA
- ▶ **Graphisme Bandeau** Valérie Pierre, direction de la Communication CNRS
- ▶ **Crédits images Bandeau**
© Photothèque du CNRS / Hervé Théry, Émilie Maj, Caroline Rose, Kaksonen
- ▶ **Pour consulter la lettre en ligne**
www.cncrs.fr/inshs/Lettres-information-INSHS/lettres-informationINSHS.htm
- ▶ **Pour s'abonner / se désabonner**
com-shs@cncrs-dir.fr
- ▶ **Pour accéder aux autres actualités de l'INSHS**
www.cncrs.fr/inshs

Institut des sciences humaines et sociales CNRS

• 3 rue Michel-Ange 75794 Paris cedex 16 •