

FG-SOL : un nouveau service pour la préservation et la réutilisation des logiciels de la recherche

Teresa Gomez-Diaz

LIGM
Cité Descartes, Bâtiment Copernic - 5, bd Descartes
Champs sur Marne 77454 Marne-la-Vallée Cedex 2

Vincent Legoll

IdGC c/o IPHC
23 rue du loess - BP28
67037 Strasbourg cedex 2

Jérôme Pansanel

IPHC
23 rue du loess - BP28
67037 Strasbourg cedex 2

Sorina Pop

CREATIS
7 Avenue Jean Capelle
69100 Villeurbanne

Geneviève Romier

CC-IN2P3
21 avenue Pierre de Coubertin CS70202
69627 Villeurbanne

Pascal Wassong

IPHC
23 rue du loess - BP28
67037 Strasbourg cedex 2

Résumé

France Grilles - Software On Line (FG-SOL) est proposé par l'infrastructure de recherche France Grilles (<http://www.france-grilles.fr/accueil/>). Les codes développés dans nos laboratoires, appelés ici logiciels de la recherche, jouent un rôle fondamental dans le fonctionnement de la recherche puisque la plupart des résultats obtenus dans l'ensemble des domaines scientifiques s'appuient sur des analyses de données, simulations, ou calculs obtenus grâce à ces logiciels. Mais les algorithmes implémentés et les savoirs sous-jacents sont souvent peu décrits dans les publications associées. Il est alors difficile voire impossible de reproduire les résultats ou de faire évoluer cette recherche sans l'accès à ces logiciels.

FG-SOL a pour premier objectif de permettre aux scientifiques qui produisent des logiciels et souhaitent les faire connaître, de pouvoir les mettre à disposition facilement, accompagnés si besoin d'un jeu de données et des liens vers les articles scientifiques associés. FG-SOL fournit donc à la communauté

scientifique un environnement qui permet d'une part la reproduction des résultats, et d'autre part l'utilisation de ces codes avec de nouveaux jeux de données, l'accès au code source et son étude.

Le second objectif de FG-SOL est de permettre aux scientifiques à la recherche d'un logiciel pour leurs propres travaux de pouvoir découvrir, tester, réutiliser les logiciels mis ainsi à disposition sur une infrastructure de production sans avoir à installer eux-mêmes ces logiciels.

Après une introduction consacrée au contexte et aux besoins, nous présenterons le service du point de vue des utilisateurs puis l'architecture technique basée sur VIP (<https://www.creatis.insa-lyon.fr/vip/>) et nous conclurons sur les perspectives d'évolution.

Mots-clefs

logiciels de la recherche, code scientifique, reproductibilité, validation, réutilisation, test, accès, service, France Grilles, cloud, Free Open Source software, research software, reproducibility

1 Introduction

France Grilles offre depuis plusieurs années des services de cloud (voir la présentation FG-Cloud) aux chercheurs et ingénieurs de la communauté scientifique. Ces services permettent à la fois un déploiement facile de machines virtuelles (VMs), de containers embarquant un ou plusieurs logiciels ainsi que des données associées sur les infrastructures et l'accueil de nouveaux utilisateurs. Il est donc devenu plus simple de donner accès à ces logiciels à de nouveaux scientifiques.

Par ailleurs, les codes développés dans nos laboratoires, appelés ici logiciels de la recherche (LR), jouent un rôle fondamental dans le fonctionnement de la recherche puisque la plupart des résultats obtenus dans l'ensemble des domaines scientifiques s'appuient sur des analyses de données, simulations, ou calculs obtenus grâce à ces logiciels. Mais les algorithmes implémentés et les savoirs sous-jacents sont souvent peu décrits dans les publications associées. Il est alors difficile voire impossible de reproduire les résultats ou de faire évoluer cette recherche sans l'accès à ces logiciels.

L'objectif de FG-SOL est de proposer l'environnement nécessaire où les logiciels de la recherche peuvent être cherchés, trouvés, étudiés, testés et réutilisés par d'autres chercheurs en vue de valider et de reproduire les résultats publiés, ou bien pour leur utilisation dans la production de nouveaux résultats et de nouvelles connaissances.

Ce service s'adresse aux concepteurs et développeurs de logiciels pour faire connaître, rendre visibles, mettre à disposition facilement, accompagnés si besoin d'un jeu de données et des liens vers les articles scientifiques associés. Il leur permet de proposer leurs logiciels, sans avoir à se préoccuper de contraintes techniques, dans un environnement adapté et intégré à une infrastructure de calcul et traitement de données.

FG-SOL s'adresse également aux scientifiques à la recherche d'un logiciel pour leurs propres travaux et leur permet de pouvoir découvrir, tester, réutiliser les logiciels mis ainsi à disposition sur une infrastructure de production sans avoir à installer eux-mêmes ces logiciels.

Il va de soi que les conditions légales d'utilisation et d'étude des logiciels sont prises en compte, les logiciels déployés étant sous licence libre.

FG-SOL fournit donc à la communauté scientifique un environnement qui permet d'une part la reproduction des résultats, et d'autre part l'utilisation de ces codes avec de nouveaux jeux de données, l'accès au code source et son étude.

2 FG-SOL, les services proposés

FG-SOL est un nouveau service au catalogue de France Grilles. Il se présente comme un ensemble cohérent de services concernant les logiciels de la recherche et destinés en premier lieu à la communauté

scientifique de France. La conception de ces services est le résultat d'une collaboration entre des membres du CNRS (LIGM¹, IdGC², IPHC³), de EGI.eu⁴, de IASA⁵ (Grèce) et de l'Université d'Édimbourg qui a été présentée en tant que proposition de projet H2020 en mars 2016. France Grilles propose cette année, sous le nom de FG-SOL, ces services pour la communauté de recherche en France. L'approche est pragmatique, centrée sur les utilisateurs et la structuration des services se veut souple, flexible et modulaire, elle est conçue pour s'intégrer aux services existants des e-infrastructures internationales telles que EGI⁶, EUDAT⁷, OpenAIRE⁸. La figure 1 ci-dessous représente schématiquement l'ensemble de ces services.

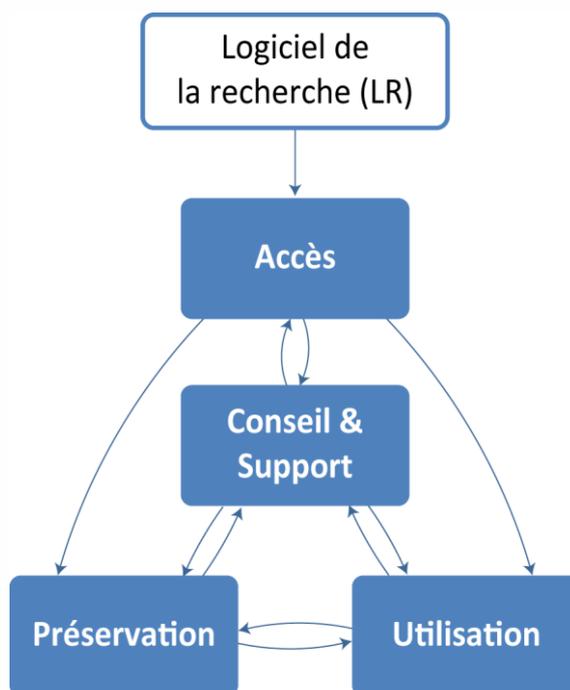


Figure 1

Bien que l'ensemble de ces services soit un tout cohérent présenté comme le service FG-SOL, chacun des services peut être considéré isolément.

2.1 Le service « Accès »

Le service Accès est composé principalement des descriptifs des logiciels et de mots clés et métadonnées pour décrire chaque logiciel et faciliter la recherche.

Le catalogue de logiciels donne accès aux descriptifs. Ces logiciels ont été proposés par leurs concepteurs et développeurs au service FG-SOL dans l'objectif de les mettre à disposition pour étude, test ou utilisation. Les descriptifs permettent de prendre connaissance des caractéristiques des différents logiciels.

¹ Laboratoire d'informatique Gaspard-Monge

² Institut des Grilles et du Cloud

³ Institut pluridisciplinaire Hubert Curien

⁴ La fondation EGI (EGI.eu) est une association de droit néerlandais établie à Amsterdam en 2010. Elle coordonne l'e-infrastructure EGI pour le compte des participants au « EGI Council »: e-infrastructures nationales, telles que France Grilles et European Intergovernmental Research Organisations (EIROs).

⁵ Institute of Accelerating Systems and Applications, Athènes

⁶ <https://www.egi.eu/>

⁷ <https://www.eudat.eu/>

⁸ <https://www.openaire.eu/>

Ce catalogue bénéficie de l'expérience de **PLUME – RELIER**⁹ et est conçu pour être interopérable avec l'EGI AppDB¹⁰ (<https://appdb.egi.eu/>). La description de chaque logiciel comprend, outre la description des fonctionnalités et des caractéristiques essentielles du logiciel, différents types de méta données dont par exemple les disciplines scientifiques concernées, des liens vers des publications ou données associées.

Les descriptions sont accessibles à travers une interface de recherche.

Dans la version actuelle des services, le catalogue est publié sous une forme préliminaire. Il est prévu dans le futur que la présentation du catalogue, en particulier la recherche, soit enrichie.

2.2 Le service « Conseil & Support »

Le service « Conseil & Support » est, comme son nom l'indique, destiné à apporter conseil et support autour des logiciels de la recherche aux utilisateurs du service (ceux-ci seront décrits plus en détail dans la section 3) dans deux dimensions : la bonne utilisation du service FG-SOL (pour les producteurs de logiciels et pour les utilisateurs de ces logiciels) et la bonne diffusion des logiciels. Ce service bénéficie de l'expérience de **PLUME – RELIER**, en particulier le thème « Patrimoine logiciel d'un laboratoire »¹¹.

Il s'agit de guider les utilisateurs et de leur apporter informations, connaissance des bonnes pratiques sous forme de références à des documents publiés et de documentations. Un support technique pour l'intégration des logiciels de recherche et de leur contexte d'utilisation sur les infrastructures France Grilles est également fourni. On notera que ce support reste léger et qu'il n'est pas prévu de support au développement des logiciels. En pratique ce service est assuré par les experts France Grilles.

Le projet PRESOFT¹² complète ce service en étudiant l'utilisation de plans de gestion de logiciels (Software Management Plans) dans quelques laboratoires pilotes.

2.3 Le service « Préservation »

L'objectif du service est de permettre d'accéder à une copie figée du logiciel pendant un temps défini à l'avance. Ce peut être par exemple la version du logiciel et l'ensemble des données ayant servi aux calculs liés à une publication.

2.4 Le service « Utilisation »

Ce service permet aux utilisateurs d'exécuter un des logiciels proposés sur l'infrastructure France Grilles. L'architecture technique est décrite dans la section 4 de ce document.

En pratique, ce service permet à un utilisateur désireux de tester un logiciel de recherche de l'exécuter avec les données proposées par le concepteur du logiciel dans un contexte prédéfini sans avoir à se préoccuper de l'installer, de le configurer.

L'utilisateur peut également utiliser ce logiciel pour sa propre production scientifique avec ses propres jeux de données sur l'infrastructure France Grilles. Cette utilisation se fait dans les mêmes conditions que l'utilisation des ressources France Grilles.

2.4.1 La reproductibilité

Le service « Utilisation » permet de reproduire l'exécution d'un traitement et de l'étudier dans la mesure où le producteur met à disposition l'ensemble du logiciel (code source, exécutable, documentation...), les

⁹ <https://www.projet-plume.org/relier>

¹⁰ Ce composant du système d'information de l'infrastructure pan-européenne EGI est développé et opéré par IASA avec le soutien de différents projets Européens et de la communauté EGI. Le « software market place » de cette base de données est un catalogue de logiciels utilisés dans le cadre de l'e-infrastructures EGI. Une partie de ces logiciels sont des logiciels de la recherche utilisés dans différentes disciplines scientifiques.

¹¹ L'ensemble des documents produits dans ce cadre est accessible à la page <https://www.projet-plume.org/patrimoine-logiciel-laboratoire>

¹² http://devlog.cnrs.fr/_media/jdev2017/poster_jdev2017_presoft_teresa_gomezdiaz.pdf?id=jdev2017%3Aposters&cache=cache

données qu'il a utilisées pour sa recherche et des liens vers les publications associées et assure un support (léger) lié à cette utilisation. Il est clair que FG-SOL ne répond pas à toutes les questions liées à la reproductibilité, néanmoins il fournit à la communauté scientifique un environnement qui facilite la reproductibilité (voir (1)).

3 Les publics ciblés

L'interaction des différents types d'utilisateurs avec les services FG-SOL est schématisée par la figure 2. FG-SOL s'adresse en premier lieu aux producteurs de logiciels sans lesquels bien sûr, il n'a pas de sens, mais aussi aux utilisateurs de ces logiciels scientifiques.

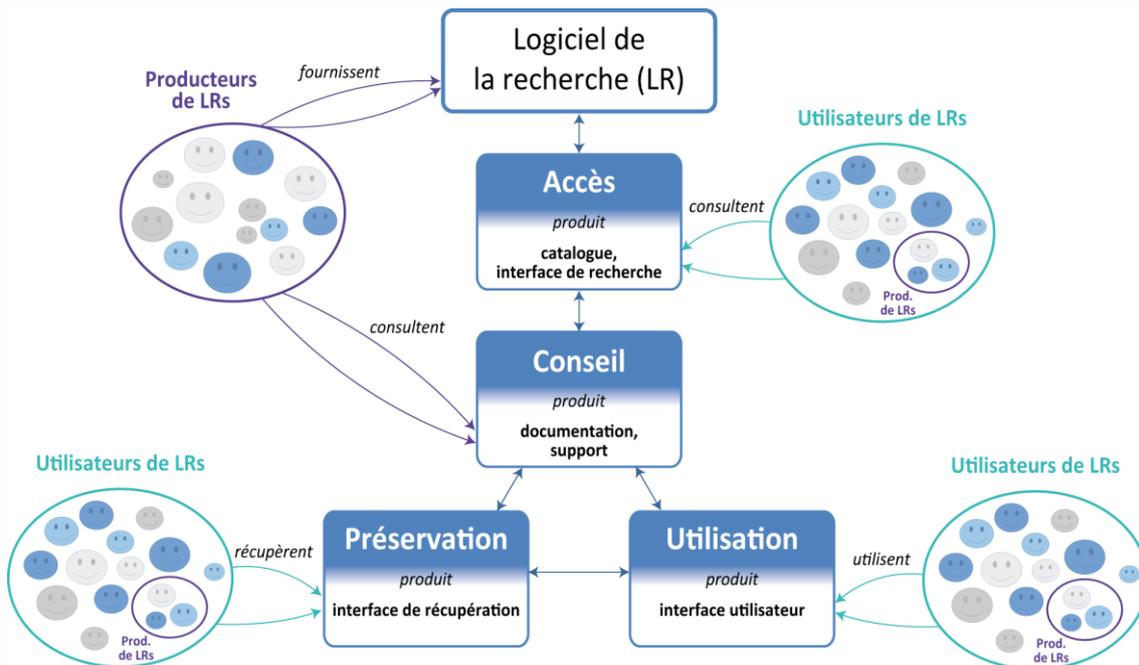


Figure 2

3.1 Les producteurs de logiciels scientifiques

Les concepteurs et développeurs de logiciels de la recherche sont la première cible de FG-SOL qui leur propose un environnement adapté pour faire connaître et proposer à l'utilisation leurs logiciels. Le service « Conseil & Support » s'adresse essentiellement à eux pour les aider dans cette démarche pas toujours évidente, souvent par manque de temps pour en prendre en compte tous les aspects (licences, conditions techniques...).

FG-SOL leur permet de bénéficier d'un ou plusieurs des différents services : seulement « Accès », ou aussi « Préservation » et/ou « Utilisation ».

3.2 Les utilisateurs de logiciels scientifiques

Pour les utilisateurs, FG-SOL apporte outre un catalogue, un environnement technique fiable où tester et utiliser un logiciel mis à disposition par des collègues. Ils n'ont plus besoin d'installer le logiciel pour pouvoir le tester, et s'il répond aux besoins, l'utilisation est possible sur une infrastructure solide.

3.3 Les laboratoires, tutelles, financeurs

Une troisième catégorie d'utilisateurs est l'ensemble des entités de la recherche. En effet, les laboratoires ont besoin de faire connaître leur production logicielle. Le catalogue est pour eux un bon outil qui peut être utilisé dans le cadre de leur évaluation par exemple.

4 FG-SOL, l'architecture technique

L'architecture technique à proprement parler concerne les services « Préservation » et « Utilisation ». Comme indiqué dans la présentation des services, le service « Conseil & Support » repose principalement sur des ressources humaines. Il n'en est donc pas question dans cette partie.

L'architecture adoptée est celle de VIP (<https://www.creatis.insa-lyon.fr/vip/>) (voir (2)). Elle est schématisée par la figure 3 ci-dessous.

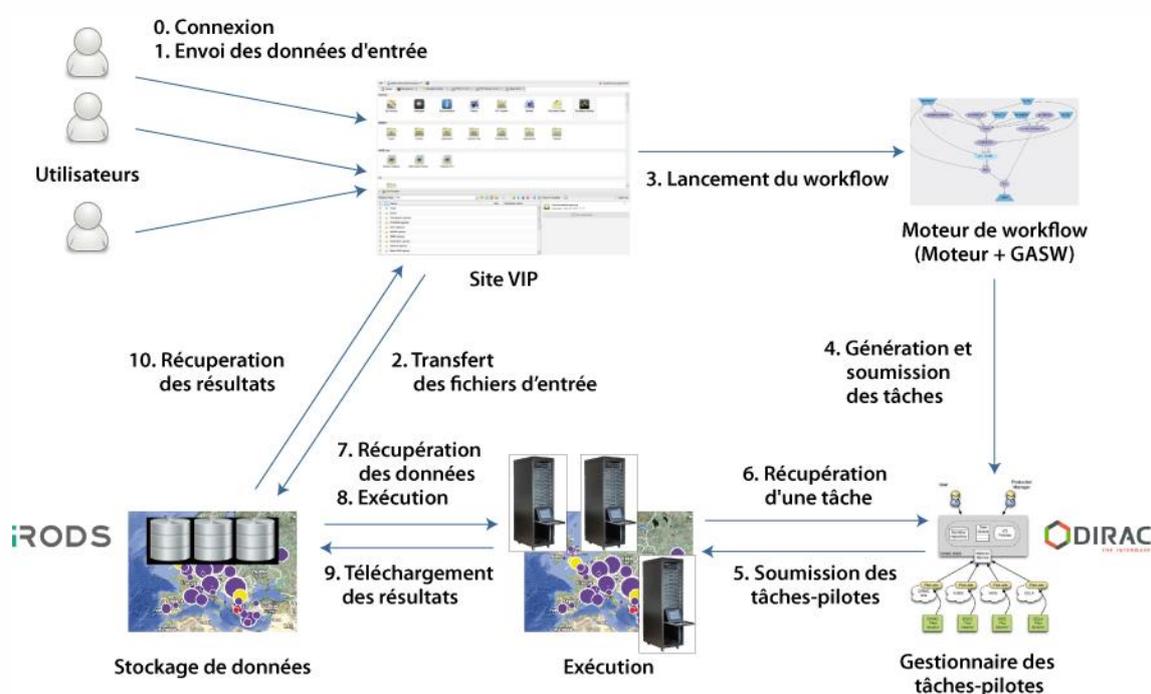


Figure 3

4.1 Le portail utilisateurs et les services sous-jacents

Le portail utilisateur de FG-SOL est une instance de VIP installée à l'IPHC et qui s'appuie sur ses propres instances de Moteur, GASW et DIRAC. Le portail permet à un utilisateur une fois identifié de choisir un logiciel, des données, de lancer l'exécution des traitements, de prendre connaissance des résultats, en bref, d'effectuer l'ensemble des tâches nécessaires à l'utilisation d'un logiciel de la recherche. La figure 4 est une copie d'écran du portail pour un utilisateur identifié :

- la page principale, en haut à gauche.
- la liste des applications lancées en haut à droite
- le panneau permettant des transferts de fichiers en bas

La figure 5 montre le suivi d'une application qui a été lancée.

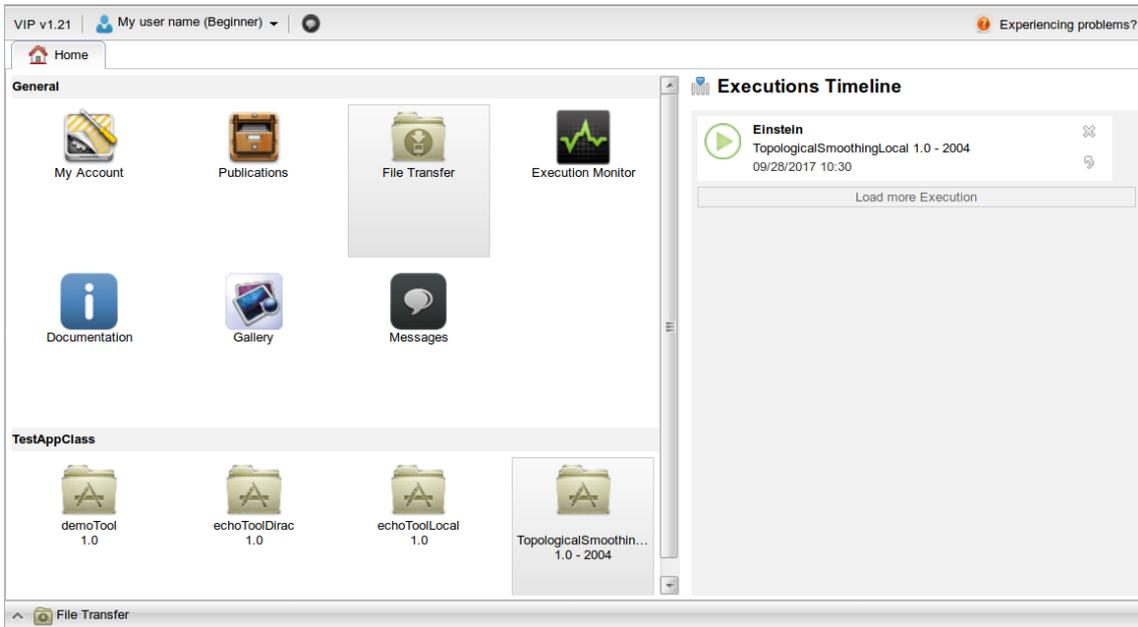


Figure 4

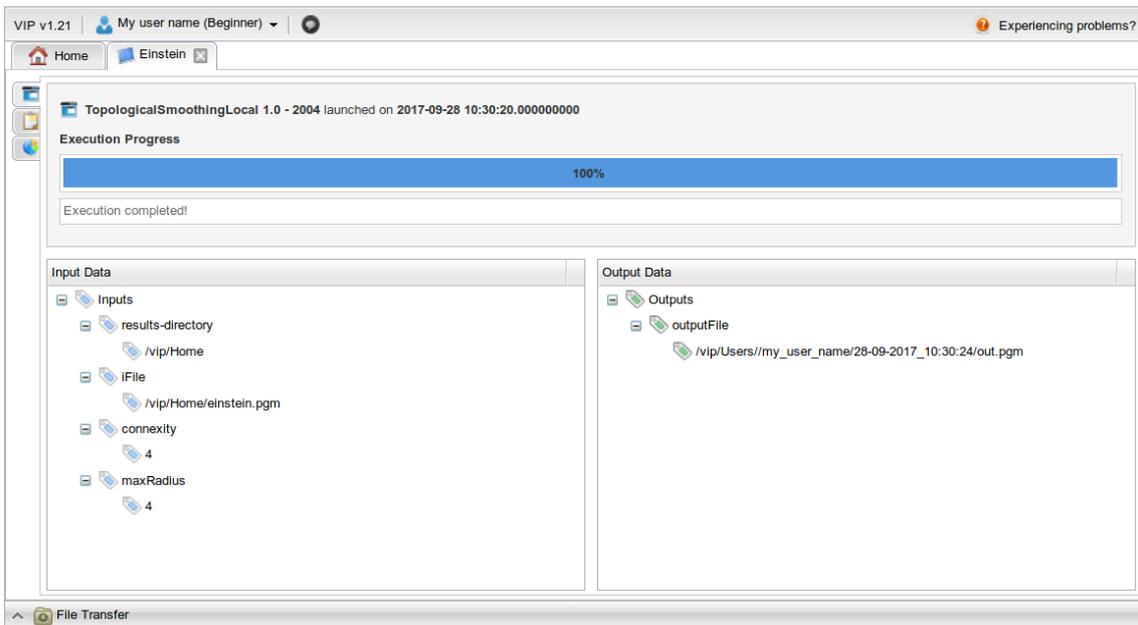


Figure 5

4.2 Les infrastructures

L'infrastructure utilisée est un sous-ensemble de l'infrastructure France Grilles. L'IPHC fournit les ressources nécessaires aux tests. Dans le cas où l'utilisation d'un logiciel par un utilisateur devient une utilisation de « production », l'utilisateur s'inscrit à la VO France Grilles et bénéficie alors des conditions d'utilisation « standard » des ressources France Grilles.

Les données sont stockées sur l'infrastructure, éventuellement à travers le service iRODS.

4.3 Le service « Accès »

Le point d'entrée aux services France Grilles est son site web www.france-grilles.fr sur lequel est publié le catalogue des services proposés. Il est donc naturel que le point d'entrée de FG-SOL soit également une des pages du catalogue sur le site web. En pratique, différents plugins ont été ajoutés au Wordpress du site pour permettre des formulaires d'interaction avec les utilisateurs, la présentation des descriptifs des logiciels, la recherche de logiciels et le lien vers les autres services. Ce service est encore à l'état de prototype.

5 Avancement du projet et perspectives d'évolutions

5.1 Etat du projet et perspectives

FG-SOL est un ensemble de services proposé par France Grilles.

Les services de « Préservation » et d' « Utilisation » sont construits de façon pragmatique sur la base de VIP et de services de calcul, traitement de données et stockage stables, en production depuis plusieurs années et répondant aux critères de qualité EGI¹³. Ces services évolueront avec les services sur lesquels ils sont basés pour rester au plus près des avancées technologiques du domaine.

Le service « Accès » peut être considéré actuellement comme un prototype. Il est construit sur le même CMS Wordpress que le site web de France Grilles, il est perfectible et sera l'objet d'évolutions en fonction des retours des utilisateurs.

Le service « Conseil et Support » est réalisé par des ingénieurs des laboratoires participants.

Les premiers logiciels de la recherche proposés dans ce cadre sont issus des laboratoires pilotes (IPHC, LIGM, LAAS) et proposés par des chercheurs volontaires. Pour amorcer un contenu, plusieurs logiciels libres utilisés assez largement dans la communauté scientifique sont également mis à disposition.

Les perspectives sont nombreuses :

Les retours des utilisateurs pilotes seront pris en compte et le service sera progressivement ouvert à d'autres laboratoires. Une ouverture aux utilisateurs non académiques de logiciels scientifiques est également envisageable pour permettre une meilleure diffusion des productions logicielles de nos laboratoires dans les entreprises.

Au niveau technique, pour le fonctionnement des services, il est nécessaire d'automatiser certaines tâches, d'apporter des améliorations au catalogue et à la recherche.

Des évolutions seraient envisageables pour proposer d'autres plateformes de calcul qui permettraient la reproductibilité de traitements dans différents environnements matériels ou avec des compilateurs différents par exemple.

Enfin, l'intégration du service au niveau européen reste un objectif à moyen terme. Cette intégration, en élargissant à la fois le champ des producteurs et celui des utilisateurs apporterait une masse critique indéniable et une ouverture internationale qui serait très certainement profitable à la diffusion des logiciels de nos laboratoires.

6 Conclusions

FG-SOL propose aux scientifiques l'environnement nécessaire où les logiciels de la recherche peuvent être cherchés, trouvés, étudiés, testés et réutilisés en vue de valider et de reproduire les résultats publiés, ou bien pour leur utilisation dans la production de nouveaux résultats et de nouvelles connaissances.

¹³ Les informations sur la qualité de service des différents sites « France Grilles » sont actuellement disponibles sur le site http://argo.egi.eu/lavoisier/site_reports?ngi=NGI_FRANCE&report=Critical&accept=html

Ce service s'adresse aux concepteurs et développeurs de logiciels. Il leur permet de faire connaître, rendre visibles, mettre à disposition facilement leurs logiciels, accompagnés si besoin d'un jeu de données et des liens vers les articles scientifiques associés. Il leur permet de proposer leurs logiciels, sans avoir à se préoccuper de contraintes techniques, dans un environnement adapté et intégré à une infrastructure de calcul et traitement de données.

FG-SOL s'adresse également aux scientifiques à la recherche d'un logiciel pour leurs propres travaux et leur permet de pouvoir découvrir, tester, réutiliser les logiciels mis ainsi à disposition sur une infrastructure de production sans avoir à installer eux-mêmes ces logiciels.

Le service FG-SOL est un nouveau service au catalogue de France Grilles en 2017 qui s'appuie sur les infrastructures et services existants ; il évoluera dans les mois prochains et s'étendra au-delà des laboratoires pilotes

Bibliographie

1. **D. L. Donoho, A. Maleki, I. U. Rahman, M. Shahram and V. Stodden.** Reproducible Research in Computational Harmonic Analysis. *Computing in.* 2009, Vol. 1.
2. **Glatard, T, Lartzien, C, Gibaud, B, Ferreira da Silva, R, Forestier, G, Cervenansky, F, Alessandrini, M, Benoit-Cattin, H, Bernard, O, Camarasu-Pop, S, Cerezo, N, Clarysse, P, Gaignard, A, Hugonnard, P, Liebgott, H, Marache, S, Marion, A, Montagnat, J.** A Virtual Imaging Platform for multi-modality medical image simulation. 2013, Vol. 32, 1.