

InriaGforge : leçons de 2 ans d'exploitation de GForge à l'INRIA

David Margery
INRIA/IRISA
Campus de Beaulieu, 35042 Rennes Cedex
David.Margery@inria.fr

Janet Bertot
INRIA Sophia Antipolis - Méditerranée
2004 route des lucioles - BP 93 FR-06902 Sophia Antipolis Cedex
Janet.Bertot@inria.fr

Christophe Demarey
INRIA Futurs
Parc Scientifique de la Haute Borne - 40, avenue Halley - Bât.A, Park Plaza - 59650 Villeneuve d'Ascq
Christophe.Demarey@inria.fr

Claude Inglebert
INRIA Futurs
Parc Orsay Université - 4, rue Jacques Monod - 91893 ORSAY Cedex
Claude.Inglebert@inria.fr

Résumé

Dans cet article, nous présentons un retour d'expérience sur l'exploitation du logiciel GForge à l'INRIA. Nous y décrivons l'infrastructure mise en place, quelques actions de sécurisation du logiciel, la volumétrie du service, ainsi quelques problèmes rencontrés lors de la mise en oeuvre.

Nous nous attardons aussi sur les usages constatés de ce type d'outil : comment les projets collaborent-ils ; quelle est leur taille ; quelles sont les fonctionnalités réellement utilisées ; qui sont les personnes collaborant à l'aide de ce type d'outils et quels sont les problèmes remontés aux administrateurs.

Nous verrons ainsi pourquoi un tel outil a trouvé toute sa place dans l'environnement de travail du personnel de l'INRIA.

Mots clefs

Forge, retour d'expérience, travail collaboratif, outils de développement, projet public vs privés, sécurisation...

1 Introduction

Une forge de type GForge est en théorie un environnement de travail collaboratif pour faciliter le développement de logiciels et la création de communautés autour de ces logiciels. Deux ans après son ouverture aux personnels de l'INRIA et à leurs collaborateurs dans le monde, le serveur InriaGforge continue d'être en phase de montée en puissance si l'on regarde les courbes d'usage, alors qu'à ce jour plus de 873 projets et 3128 utilisateurs sont enregistrés à ce service¹.

¹ Les données de cet article correspondent à la situation du service entre les 13 et 14 septembre 2007

Dans cet article de retour d'expérience, nous nous proposons d'analyser la réalité de l'usage de ce type d'outil afin de connaître les différences entre la théorie de l'usage d'une forge et sa pratique. Nous espérons ainsi pouvoir répondre à la question de savoir si la popularité du service est le signe de nouvelles pratiques en travail collaboratif ou tout simplement le fruit d'une agrégation réussie à assez grande échelle de services bien connus.

L'équipe chargée d'administrer InriaGforge est amenée à fournir un service utilisé dans le travail quotidien de centaines de personnes, qui s'attendent à une certaine confidentialité de leur travail et surtout de leurs données, et ce sur un environnement ouvert. Dans cet article, nous décrirons donc aussi la manière dont est administré le serveur pour assurer la meilleure continuité de service possible, et les mesures prises pour assurer au mieux la confidentialité des données sans remettre en cause la facilité d'utilisation du service : être suffisamment flexible et complet pour héberger les services utiles au développement collaboratif de logiciels. Nous décrirons aussi les principaux problèmes de passage à l'échelle rencontrés.

Pour finir, nous donnerons quelques indications chiffrées sur la volumétrie d'un tel service.

2 Présentation générale du service

2.1 Définition d'un service de forge

Une forge est un ensemble d'outils ou de composants plus ou moins intégrés qui favorisent la collaboration autour d'un projet. En général, une forge fournit des outils de gestion de versions, de suivi de bogues et de tâches à accomplir, et des outils de communication (forums, pages web, wiki, listes de diffusion, etc.). Une forge facilite la collaboration ; les outils et leurs utilisations s'adaptent à des besoins variés et évolutifs.

En plus de ces éléments, GForge fournit un service de gestion de documents et d'annonces, de l'hébergement de

pages PHP pour les espaces web ainsi qu'une zone de téléchargement des versions du projet mises à disposition.

2.2 Motivations pour un service de forge

Pour l'INRIA, mettre en place un service de forge répond à plusieurs motivations.

La première est de fournir à toutes les personnes travaillant à l'institut une infrastructure pour leurs collaborations scientifiques avec les partenaires internes ou externes à l'institut.

La seconde est de mettre à la disposition des scientifiques de l'INRIA des outils qui favorisent de bonnes pratiques de travail collaboratif (que ce soit au sein d'un projet interne à l'INRIA ou ouvert à l'extérieur) et des outils qui aident à la diffusion du résultat de ce travail. Une forge permet en effet une collaboration évolutive autour d'un projet, grâce à des espaces privés et publics associés au projet ; l'environnement de travail évolue et s'adapte en fonction du projet. Alors que beaucoup de forges servent à l'hébergement de projets libres ou seulement interne à une entreprise, la notion de projets privés est une contrainte importante pour ce service. Dans ce contexte, un projet privé est un projet accessible seulement aux membres de celui-ci, que ces membres soient personnels INRIA ou pas. La possibilité d'avoir à la fois des projets publics et des projets privés et d'évoluer entre les deux, si nécessaire, est importante afin de favoriser les bonnes pratiques dès le début d'un projet y compris s'il n'est pas encore prévu de le rendre public.

2.3 Objectifs du service

Le service de forge à l'INRIA a donc été démarré avec trois objectifs principaux :

- offrir une grande facilité de collaboration,
- assurer la confidentialité des données des projets hébergés sur le serveur, en particulier pour les projets privés
- assurer la disponibilité du service, parce qu'il doit être utilisé au quotidien, dans plusieurs fuseaux horaires.

3 Infrastructure mise en place

3.1 Logiciel

Pour répondre à ces motivations et objectifs, un service a été mis en place, basé sur le logiciel GForge [1] et trois serveurs gérés par une équipe d'administrateurs.

Le logiciel GForge a été choisi parce qu'il

- permettait un démarrage rapide aux projets ayant déjà adopté CVS [2] comme outil de gestion de versions,
- était un logiciel ayant déjà fait ses preuves
- est un logiciel libre, ce qui permet de faire des adaptations du code au contexte d'utilisation.

Un certain nombre d'adaptations du logiciel ont été nécessaires pour nos objectifs et les besoins exprimés de nos utilisateurs, en particulier pour

- le support de projets privés,

- le support d'une personnalisation des pages décrivant l'accès au dépôt pour permettre aux projets de donner quelques informations supplémentaires ; 14 projets utilisent cette fonctionnalité (par exemple cgal [3]).
- la mise en place d'accès webdav aux dépôts subversion [4] (cette fonctionnalité a été utilisée par 292 utilisateurs entre le 7/08/2007 et le 9/09/2007).
- le support de la navigation des sources par les membres authentifiés des projets privés,
- prendre en compte les modifications faites à l'installation des services sur les serveurs pour une meilleure sécurisation des données.

Certaines de ces adaptations sont relativement faciles à mettre en œuvre. Néanmoins, comment partager ces différentes modifications avec la communauté ? Bien qu'une mise à disposition de ces adaptations soit possible, presque tous nos ajouts sont reliés à une utilisation très atypique (autour des projets privés...). Nous avons proposés les patches les plus simples et généraux à la communauté des développeurs de GForge mais faute d'interlocuteurs directs l'effort n'a pas été poursuivi.

3.2 Équipe d'administration

L'équipe chargée de l'administration du serveur a été constituée de six ingénieurs des différents centres de recherche mis à disposition pour une partie de leur temps de travail (10% en phase d'exploitation), d'un responsable opérationnel (à 25 %) chargé de l'animation de l'équipe et de deux personnes supervisant le service pour faire le lien avec les services des moyens informatiques et les services de développement et d'expérimentation dans les différents centres de recherche de l'INRIA.

A l'usage il s'avère que l'administration du service occupe 0,5 équivalent temps plein auquel il faut ajouter 0,1 pour les deux superviseurs. Pour le déploiement du service, un noyau de trois ingénieurs a investi 7HM pour le choix du système, du logiciel forge et pour la fiabilisation et la sécurisation du service, cela inclut toutes les procédures et la documentation de celles-ci. En phase de déploiement le temps investi par les superviseurs et par un expert sécurité est égal à 3,5 HM.

Pour la gestion des incidents et des requêtes des utilisateurs, l'outil RT [5] (gestionnaire de tickets déployé sur un serveur distinct) est utilisé. Au 3 août 2007, soit 26 mois après la mise en service, 2548 tickets ont été traités. Parmi ces tickets, 805 (32%) correspondent à des demandes utilisateurs dont :

- 73 (9%) pour des problèmes liés aux listes de diffusion, essentiellement à cause de leur mauvaise intégration au reste du mécanisme d'authentification
- 167 (~21%) pour des difficultés liées à l'utilisation de ssh pour se connecter au serveur pour accéder au dépôt ou à l'espace d'hébergement web
- 29+18 (~6%) pour des demandes liées à l'utilisation de subversion et CVS respectivement.

Les autres tickets sont générés par notre configuration de GForge (demandes d'ouverture de projet et d'annonces sur la Une du site InriaGForge) et par les différents mécanismes de surveillance du service que nous avons mis en place.

Afin d'éviter de toujours répondre aux mêmes requêtes, une Foire Aux Questions ([FAQ](#)) est maintenue par l'équipe d'administration et accessible depuis la page d'accueil de la forge. De même, pour aider les utilisateurs novices, un petit tutoriel a été élaboré : « [A gentle introduction to InriaGforge](#) ». Celui-ci explique comment se créer un compte, demander l'ouverture d'un projet, configurer ses accès, son environnement, etc.

La FAQ et le tutoriel ont été rédigés en anglais de manière à pouvoir être lus et compris par le plus grand nombre (nombreuses collaborations internationales). L'ensemble des pages InriaGforge est disponible en 2 langues : anglais et français (selon les préférences du compte utilisateur ou de son navigateur), toujours dans le but de faciliter l'accès de tous à la forge.

Afin de favoriser la communication et l'échange d'informations au sein de l'équipe d'administration de la forge, équipe qui est géographiquement dispersée, une liste de diffusion et un wiki ont été mis en place. Le wiki qui est externe au serveur GForge est la « mémoire » de l'équipe d'administration et permet à chaque membre de venir étoffer et corriger la documentation. Le wiki comprend :

- les comptes-rendus des réunions d'équipe
- une description de la mise en place des serveurs et de leur configuration
- les procédures à suivre en cas de défaillance du serveur, d'alertes de surveillance, d'alertes de sécurité (mise à jour du noyau ou installation de paquets nécessaires), ou de restauration de données.
- une liste non-exhaustive d'aides mémoire concernant les manipulations régulièrement effectuées par les administrateurs
- un document relatif à la sécurité des serveurs. Celui-ci décrit les services utilisés, les actions entreprises pour les sécuriser ainsi qu'une analyse des risques potentiels et quelques pistes pour améliorer la sécurité.

3.3 Serveurs

Nous utilisons trois serveurs pour l'administration du service : le serveur de *production*, un serveur de qualification et un serveur pour le développement. Le serveur de qualification est identique au serveur principal au niveau matériel et toutes les mises à jour y sont qualifiées avant de passer sur le serveur de production. Le serveur pour le développement est utilisé par l'équipe pour adapter notre installation, ajouter des fonctionnalités, déboguer nos scripts, etc. (étant donnée son utilisation, ce serveur est une machine banale). Ceci a l'avantage de laisser libre le serveur de qualification pour toute mise à jour de sécurité à traiter en urgence.

Sur chaque serveur, nous distinguons trois entités distinctes pour l'administration :

- le système de base, maintenu en appliquant les patches de sécurité de notre distribution sur chaque serveur,

- les fichiers de configuration propres à chaque serveur, mis à jour en utilisant un petit utilitaire (edadm²) permettant de conserver l'historique des modifications,
- l'ensemble des logiciels/composants spécifiques à GForge, maintenus cohérent entre les serveurs à l'aide de scripts basés sur rsync.

Sur le serveur de développement, les versions de tous les logiciels et composants de notre version personnalisée de GForge correspondent aux dernières versions présentes dans le dépôt subversion de notre code. Ainsi, si un bogue est corrigé sur la version en ligne sur ce serveur par un des administrateurs, il peut par une simple commande de notre logiciel de gestion de versions enregistrer cette nouvelle version dans ce dépôt. L'historique des modifications est donc aisément conservé. Mieux, il est possible, en utilisant les fonctionnalités de subversion, d'obtenir le patch des modifications locales depuis que le système est en place. Celui-ci pourrait être utilisé pour adapter des nouvelles versions de GForge à nos besoins.

Sur le serveur de qualification, des scripts, basés sur le logiciel rsync, permettent de pousser les modifications faites sur le serveur de développement vers ce serveur. Nous pouvons alors tester les fonctionnalités de base sur ce serveur à l'aide d'un script de test pour valider les services essentiels tels que l'accès aux dépôts et la confidentialité de ceux-ci. Pour des modifications plus complexes (e.g. changement de version de GForge), nous avons une suite complète de tests de non régression à dérouler manuellement. De plus, le passage par un serveur de qualification permet également de vérifier notre documentation décrivant des différences entre la GForge originale et notre InriaGforge.

Pour le passage en production, les mêmes scripts à base de rsync sont utilisés pour mettre à jour le serveur principal. De plus, tous nos fichiers de configuration du logiciel GForge (configuration web, fichiers include pour les scripts...) sont générés à partir d'un fichier de configuration maître, géré par edadm, et d'un script synchronisé par rsync. Il y a donc peu de chances de faire des erreurs de configuration.

Pour finir, sur le serveur de qualification et sur le serveur principal, toute modification sur le système de base est précédée d'une sauvegarde du système de base pour permettre un retour arrière rapide en cas de problèmes lors d'une mise à jour.

3.4 Sécurité et confidentialité des données

Nous avons mis en place plusieurs mesures pour s'assurer de l'intégrité et de la confidentialité des données.

3.4.1 Environnements restreints

Afin de limiter les désagréments liés à des attaques éventuelles, les serveurs hébergeant la forge sont constitués de l'arborescence classique d'un système Unix ainsi que de deux environnements restreints (chroot) imbriqués. Le premier d'entre eux héberge le serveur web alors que le deuxième héberge les données des utilisateurs.

² Script basé sur RCS permettant de garder un historique de tous les changements effectués.

Ces environnements restreints sont maintenus en cohérence avec le système de base par synchronisation locale grâce à des scripts basés sur rsync qui permettent, entre autres, de propager presque automatiquement les mises à jour de sécurité.

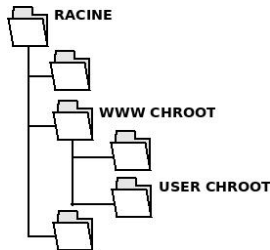


Figure 1 : Encapsulation des environnements restreints

3.4.2 suPHP

La mise en place de suPHP [6], outil permettant d'exécuter un script PHP avec les droits du propriétaire du script, s'est avérée nécessaire. En effet, la forge permet aux utilisateurs d'écrire leurs propres scripts PHP et de les interpréter grâce au serveur web de la forge. Sans suPHP, les scripts s'exécutent avec les droits de l'utilisateur apache qui dispose d'accès privilégiés aux dépôts des projets, notamment pour parcourir le code source des projets en ligne. Un utilisateur malicieux aurait alors pu parcourir le code source d'un projet auquel il n'aurait pas dû avoir accès en écrivant le script PHP adéquat.

Cette mesure limite aussi l'impact des attaques via des scripts PHP mal ficelés. Ce problème montre les limites entre la protection offerte par notre installation et l'utilisation souhaitée par nos utilisateurs. Clairement, il est impensable de prohiber du PHP sur la forge, nous pouvons seulement informer et éduquer nos utilisateurs sur la meilleure façon de s'assurer que quelqu'un ne peut pas détourner leur code PHP pour des actions malveillantes. L'équipe a un rôle pédagogique à jouer.

3.4.3 Intégrité du serveur

L'intégrité des binaires et de la configuration des différents services utilisés par InriaGforge est assurée par l'outil integrity [7] et des scripts permettant d'alerter les administrateurs en cas de changement sur le système de fichiers.

Cette vérification ne pouvant s'appliquer aux données des utilisateurs qui évoluent régulièrement, une surveillance de l'espace disque utilisé par les projets et les utilisateurs a été mise en place. Cette surveillance se traduit par l'envoi d'un mail aux personnes concernées leur demandant de bien vouloir s'assurer que toute augmentation significative de l'utilisation de l'espace disque est normale. Ceci permet par exemple de déceler des fichiers qui auraient été déposés à l'insu du propriétaire du compte. En pratique, plusieurs seuils ont été définis pour les comptes utilisateurs et les comptes des projets. Par exemple, pour les comptes des projets qui contiennent entre autres le répertoire *htdocs* servi par le serveur web, le premier palier correspond à « aucune donnée hébergée », le suivant à 10Mo et les paliers suivants augmentent progressivement.

3.5 Résultats

3.5.1 Évolutivité du serveur

Nous avons déployé ce service en nous basant sur le très jeune (à l'époque) GForge 4.0. Nous décidons d'explorer la migration à une version 4.5 au bout d'un an, car cette version semble apporter des améliorations intéressantes, en particulier le support de deux sous-groupes par projet pour permettre de faire la distinction entre les membres ayant des droits d'écriture sur le dépôt et les autres et le plugin wiki. La version 4.5 avait une restructuration majeure de la base de données de la GForge, imposant une migration complexe. Le service offert par la forge ne devait pas subir une longue interruption. Une pré-étude de deux semaines nous a permis de comparer les deux versions sur des données et de définir la procédure de migration. Une fois qualifiée, le passage sur le serveur de production a nécessité une heure d'indisponibilité.

3.5.2 Volumétrie du service

Depuis sa mise en service, la bande passante utilisée par InriaGforge connaît une croissance régulière comme le montre le diagramme ci-dessus. L'utilisation de la bande passante est corrélée aux nombres de projets, d'utilisateurs et l'activité sur les dépôts. La moyenne mensuelle en 2007 est de 281 Go en sortant et 24,5 Go en entrant.

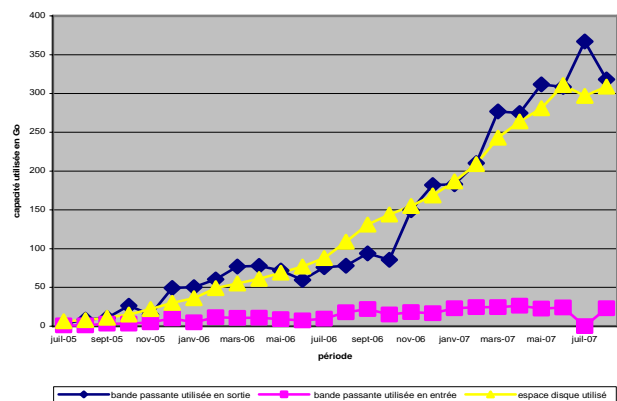


Figure 2 : Volumétrie du service INRIAGforge

Concernant l'espace disque utilisé, sa progression est constante (+200% chaque année). Actuellement 116 Go sont nécessaires au stockage des données des utilisateurs. Notre serveur est dimensionné de façon à réserver 450 Go pour les données utilisateurs. Ce service étant encore jeune, il est difficile de prévoir l'espace disque nécessaire pour les 2-3 ans à venir.

3.5.3 Problèmes rencontrés

Deux problèmes liés à la distribution linux ont été rencontrés. Tout au début, l'outil de surveillance de nos disques durs nous a contraints d'utiliser une version non stable du noyau linux. Deux semaines après l'ouverture, nous avons atteint une limite système sur le maximum de groupes Unix auxquels un utilisateur peut appartenir. Cette limite est facilement contournée, mais nécessite néanmoins

une recompilation de la libc avec les en-têtes du noyau à chaque mise à jour majeure de celui-ci.

Finalement comme tout système ouvert, nous avons vu un certain nombre de tentatives d'attaques, de toute évidence faites par des robots. Ces tentatives, souvent par le biais d'attaque de codes PHP, ont pour but d'utiliser l'espace disque de la forge pour distribuer des fichiers qui ne sont pas associés aux projets sur la forge.

4 Usages d'InriaGforge

4.1 Profil des projets

Sur les 873 projets actifs sur InriaGforge, il y a 228 projets publics et 645 projets privés.

Voici la répartition des projets selon leur description :

- développement de logiciels scientifiques : 633 projets
- rédaction collaborative scientifique : 164 projets
- développements non-scientifiques : 76 projets

Le service a été ouvert avec l'objectif d'offrir une plateforme pour les collaborations scientifiques de l'institut. Le développement de logiciels est évidemment l'utilisation habituelle d'une forge. La forge accueille également des projets pour la rédaction d'articles, livres, etc., activité typique des collaborations scientifiques de nos utilisateurs.

Cette activité collaborative ne nécessite que l'utilisation d'un gestionnaire de versions. De plus, elle a une durée de vie très courte. Bien qu'une forge paraisse un outil trop puissant pour héberger ces projets, nous avons décidé de fermer un serveur CVS dans le but de favoriser l'adoption de la forge et d'avoir un outil unique pour ces deux activités collaboratives (développement logiciel et rédaction d'articles).

La seule interrogation autour des projets de rédaction est la gestion de leur terminaison. Actuellement nous n'avons reçus que moins d'une dizaine de demandes de fermeture.

La dernière catégorie de projets sur la forge correspond aux développements hors de la production scientifique, elle inclut les développements effectués par les équipes techniques de l'INRIA et les développements d'utilitaires faits par des scientifiques. Clairement, ces développements doivent profiter de cette plate-forme qui facilite l'adoption de bonnes pratiques de développement.

4.2 Profil de collaboration des projets

L'étude du nombre de membres d'un projet, du nombre de membres ayant un accès en lecture et du nombre de membres ayant un accès en écriture à des dépôts de codes donne quelques enseignements intéressants.

On constate que peu de projets utilisent les possibilités offertes par la forge de restriction d'accès aux dépôts. Ce résultat est une surprise, sachant que les possibilités d'un contrôle fin des accès au dépôt étaient un besoin exprimé par nos utilisateurs. Ce comportement est-il dû à la nature d'une forge qui suppose une confiance mutuelle entre les développeurs ? Nous devrions interroger nos utilisateurs pour tenter de comprendre si le niveau de contrôle offert est inadéquat ou si les utilisateurs sont mal informés ou si à

l'utilisation, ce genre de restriction s'avère inutile parce que les utilisateurs adoptent une approche plus open source.

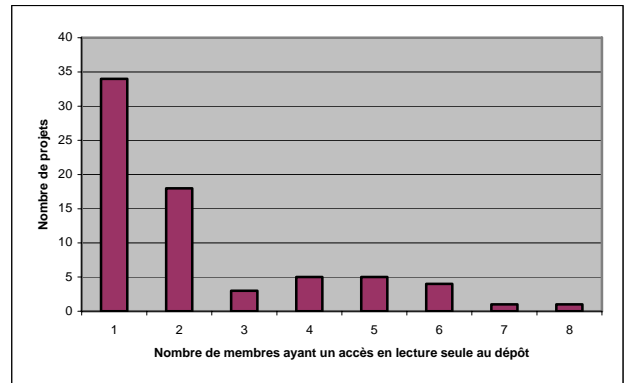


Figure 3 : Membres avec un accès en lecture seule au dépôt

On constate également que la plus grande part des projets sont de petite taille puisqu'un peu plus de 48% des projets ont au maximum 3 membres. Pour les autres projets, 40% ont entre 4 et 10 membres et 12% (112 projets) ont plus de 10 membres.

Il n'est pas très surprenant de voir beaucoup de petits projets (de 1 à 3 développeurs) ; les développements faits dans un institut de recherche en informatique commencent souvent par la validation d'idées scientifiques par un chercheur et ses étudiants. Par contre le chiffre de 40% des projets ayant entre 4 et 10 membres laisse supposer l'existence de développements nécessitant une force de travail conséquente.

La forge a été adoptée à la fois par des projets mûrs qui auparavant administraient leurs propres environnements de développement tel que Coq [8], et par des projets jeunes qui cherchent à construire une communauté open source, tel que OpenAlea [9].

L'étude de l'origine des membres de chaque projet basée sur l'adresse mail que ceux-ci ont associée à leur compte ne nous donne qu'une vision partielle sur la diversité des collaborations. On trouve 287 projets (33%) dans lesquels tous les membres ont des adresses des domaines INRIA, parmi lesquels 103 projets ont un seul utilisateur. On identifie aussi 247 projets (28,5%) comme des collaborations franco-françaises (des membres de l'INRIA et des membres avec des adresses dans le domaine « fr »). Enfin, 241 projets (27,5%), sont identifiés comme des collaborations internationales : au moins un membre possède une adresse mail avec un suffixe d'un autre pays ou « edu ». Un dernier groupe de 98 projets (11%) ont des adresses mail chez un fournisseur de boîtes aux lettres gratuites : yahoo.com, gmail.com, hotmail.com, et laposte.net. Sont-ils des projets internationaux ou franco-français ?

La lecture de ces chiffres met en évidence la nécessité pour un organisme comme l'INRIA de mettre à la disposition de

ses membres des espaces numériques de travail pour des groupes ouverts à l'extérieur.

4.3 Outils de collaboration utilisés

Une forge est une agrégation de services. Afin de comprendre quels sont les services utilisés, nous avons fait une analyse de leur utilisation depuis le 1^{er} janvier 2007 :

- La zone de téléchargement est considérée comme utilisée si au moins un fichier a été ajouté durant la période.
- L'outil de suivi de bogues est considéré comme utilisé si au moins 4 bogues ont été créés et un fermé au cours de la période examinée.
- Le gestionnaire de tâches est considéré comme utilisé si au moins 5 tâches ont été créées au cours de cette période.
- Les forums sont pris en compte si plus de 10 messages ont été créés par au moins 2 auteurs différents.
- Les listes de diffusions sont considérées comme utilisées par un projet si celui-ci possède au moins 2 listes de diffusion (en sachant qu'une liste pour les *commits* au dépôt de gestionnaire de version est créée automatiquement pour chaque dépôt). Leur activité n'a pas été observée.
- L'hébergement web est considéré comme utilisé si les fichiers constituant le site ont une taille totale de plus de 10ko.
- Un projet est considéré comme utilisant un wiki s'il a activé la fonctionnalité wiki ou si on trouve des fichiers dont le nom comporte la chaîne wiki dans l'espace web hébergé (15 projets ont installés leur propre wiki).
- Un projet est considéré comme ayant utilisé le service d'annonces si au moins une annonce a été postée.
- Un projet est considéré comme utilisateur du gestionnaire de documents s'il possède au moins 2 documents gérés.
- Un projet est considéré utilisateur de son dépôt de code s'il y a eu au moins 5 contributions à ce dépôt au cours de la période examinée.

La figure 4 montre que le dépôt de versions est de loin la fonctionnalité la plus utilisée. Viennent ensuite les outils de communication tels que les sites web, les wiki ou les listes de diffusion, puis les outils permettant de mettre à disposition des fichiers comme le gestionnaire de documents. Ces fonctionnalités semblent former la base des services utiles auquel il faut ajouter une zone de téléchargement pour les releases.

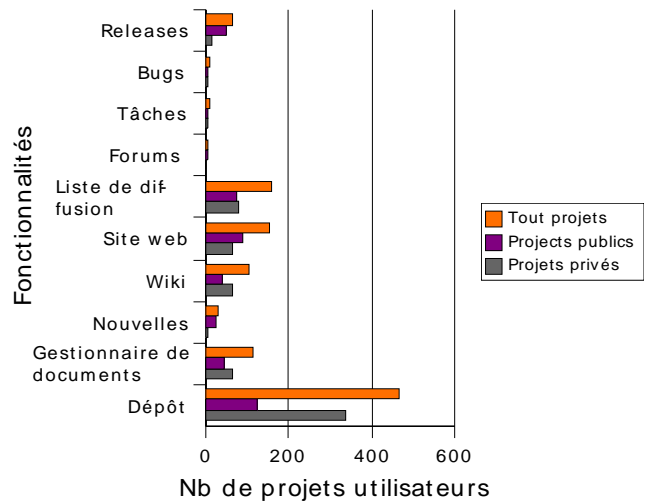


Figure 4 : Fonctionnalités utilisées

Elle montre aussi que les projets publics sont moins utilisateurs du dépôt de code par rapport aux projets privés. Ceci s'explique par le fait que certains groupes maintiennent deux dépôts (un public et un privé) et n'utilisent le projet public qu'à des fins de communications et de téléchargement.

Par contre, à l'exception de l'utilisation du dépôt de versions, aucune hiérarchie dans l'utilisation des autres fonctionnalités ne semble se dessiner.

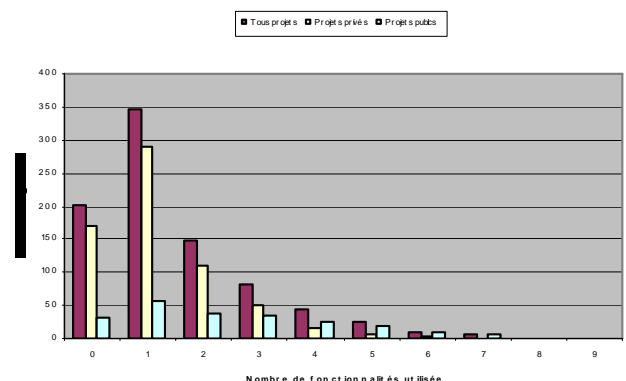


Figure 5 : Nombre total de fonctionnalités utilisées (total, projets privés, projets publics)

La figure 5 détaille le nombre de fonctionnalités utilisées par projet. On remarquera que les projets publics semblent utiliser plus de fonctionnalités, puisqu'ils sont plus de 41% à utiliser 3 fonctionnalités ou plus, alors que pour les projets privés, ils ne sont qu'à peine plus de 11%. On remarquera que près de 190 projets semblent plutôt inactifs. Parmi ceux-ci, la plupart sont des collaborations de rédactions scientifiques qui sont terminées (soumission d'articles). Néanmoins, on distingue quelques projets mettant à disposition des benchmarks ou d'autres documents et dont la seule activité enregistrée est le téléchargement de ces fichiers.

Une conclusion à la lecture de ces chiffres est que les projets privés et les projets publics ont des préoccupations différentes.

4.4 Services de plus haut niveau

Le logiciel GForge offre un certain nombre de possibilités de bâtir des services de plus haut niveau.

Le serveur exporte une API SOAP qui permet d'automatiser la plupart des tâches effectuées manuellement. En particulier, cette API permet la construction de release, la consultation et l'ajout de nouveaux bogues et tâches. Cependant, ces huit derniers mois, le serveur n'a enregistré qu'une seule utilisation de cette API. La non-utilisation est due probablement au manque de promotion de celle-ci. La seule promotion de la forge a été assurée par quelques présentations dans les centres de recherche lors du lancement d'InriaGforge et ces fonctionnalités avancées n'étaient pas leur objectif premier.

Une autre interface pour exploiter les données d'un projet InriaGforge est la mise à disposition de scripts. Ceux-ci permettent la consultation des dernières nouvelles du projet, de l'activité des forums de discussion et le suivi des bogues. Un exemple d'utilisation est donné par la page par défaut de l'espace d'hébergement web associé à chaque projet. Seuls 25 projets semblent utiliser cette possibilité.

Finalement, l'accès ssh donné au serveur pour mettre à jour les pages web du projet ainsi que les accès ssh et webdav nécessaires à l'accès au dépôt permettent à un projet de construire un service de version journalière (*nightly build*), de test d'un logiciel ou de mise à jour automatique de site web. Seulement 24 connexions régulières ont pu être cependant détectées (une connexion régulière étant définie comme plus de 10 connexions à la même heure à 10mn près durant le dernier mois). Si on descend le seuil à 4 connexions pour tenir compte des tâches programmées hebdomadaires, on trouve 48 connexions récurrentes potentielles.

On peut cependant noter qu'au mois de juin 2007, 1103 adresses IP différentes ont lu les flux rss proposés par le serveur.

Il semble donc que, dans l'ensemble peu de projets aient bâti des services complexes au dessus d'InriaGforge. Une présentation sur une utilisation plus poussée de la forge s'appuyant sur quelques projets réels, tel qu'OpenAlea [9], est à l'ordre de jour.

4.5 Mobilité des utilisateurs

Une analyse des logs de connexions aux dépôts indique que les deux possibilités de connexion (ssh, webdav) sont utilisées ; pour la période allant du 18 juillet au 9 septembre 2007, 1080 utilisateurs différents connectés, dont 884 par ssh et 292 par webdav (donc 96 en utilisant les 2 méthodes).

Afin d'avoir une idée de la mobilité de nos utilisateurs, nous avons classé les adresses IP de connexions par

groupement de 256x256 adresses IP (équivalent à un réseau de classe B). Lorsqu'on analyse ce classement, on trouve que 394 utilisateurs se sont connectés depuis plus d'un regroupement (36%). Parmi ceux-ci, 179 ont utilisé au moins 3 groupements différents, 88 au moins 4. Il y a donc une proportion assez importante des utilisateurs qui sont mobiles en ce qui concerne le réseau d'attache depuis lequel ils se connectent.

La mobilité des utilisateurs ne se résume donc pas à l'utilisation d'un espace numérique personnel partagé entre le domicile et le lieu de travail mais bien plutôt à une réelle mobilité des utilisateurs alors qu'ils collaborent avec d'autres. Dans ce contexte, il serait probablement utile d'ajouter un service de messagerie instantanée voir de voix sur IP à un serveur tel qu'InriaGforge pour faciliter la collaboration en mobilité.

5 Conclusions

L'engouement pour ce service montre qu'il répond à un vrai besoin pour les collaborations, qu'elles soient en interne ou avec des collaborateurs extérieurs. Bien que les différents niveaux d'intégration des composants a été source de quelques critiques, car certes améliorables, la croissance régulière témoigne de son appropriation par les scientifiques de l'INRIA.

Jusqu'à maintenant l'effort d'animation autour de l'InriaGforge s'est concentré sur l'adoption des fonctionnalités simples (les transferts de dépôts du code, connexion ssh, configuration de listes de diffusion...). La grande majorité des projets se satisfait des outils qui leur sont familiers. Les services plus avancés offrent des fonctionnalités intéressantes. La prochaine mission pour l'équipe est d'exposer ces fonctionnalités de haut niveau très peu utilisées afin d'automatiser certaines tâches effectuées manuellement. Elle pourra s'appuyer sur l'expérience des utilisateurs actuels plus avertis.

La forge a sa place pour les projets de petite taille comme pour ceux de grande taille. Mettre un tel outil à disposition des chercheurs visait à libérer les chercheurs de l'installation et la sécurisation de leurs propres serveurs de développement, permettant à ceux-ci de se concentrer sur leurs activités scientifiques. La contrepartie est l'investissement effectué par l'équipe d'administration pour assurer la confidentialité et l'intégrité des données des projets et pour maximiser la disponibilité du service. Ce coût est raisonnable comparé au succès du service rendu.

Les usages de la forge sont très variés et une forge peut s'adapter aux besoins de tous et tout au long de la vie d'un projet. Les composants utilisés dépendent fortement du type du projet (public vs privé) : les projets publics étant plutôt orientés outils de communications (annonces, wiki, site web...), à l'opposé les projets privés sont plus orientés vers le dépôt de code.

Finalement, nous nous interrogeons sur un éventuel remplaçant du logiciel GForge lui-même. Bien que des prétendants existent, il faut être capable de faire une telle transition en douceur et d'accompagner nos utilisateurs vers ce nouvel outil.

6 Références

- [1] <http://gforge.org>
- [2] <http://www.nongnu.org/cvs/>
- [3] https://gforge.inria.fr/scm/?group_id=52
- [4] <http://subversion.tigris.org/>
- [5] <http://bestpractical.com/rt>
- [6] <http://www.suPHP.org/Home.html>
- [7] <http://integrit.sourceforge.net/>
- [8] <http://coq.gforge.inria.fr>
- [9] <http://openalea.gforge.inria.fr>