

Distr.
GÉNÉRALE

CES/AC.71/2004/18 (Summary)
1^{er} mars 2004

FRANÇAIS
Original: RUSSE

COMMISSION DE STATISTIQUE et
COMMISSION ÉCONOMIQUE
POUR L'EUROPE

COMMISSION EUROPÉENNE

CONFÉRENCE DES STATISTICIENS
EUROPÉENS

OFFICE STATISTIQUE DES
COMMUNAUTÉS EUROPÉENNES
(EUROSTAT)

**ORGANISATION DE COOPÉRATION ET DE
DÉVELOPPEMENT ÉCONOMIQUES (OCDE)
DIRECTION DES STATISTIQUES**

Réunion mixte CEE/Eurostat/OCDE sur la gestion des systèmes d'information statistique
(Genève, 17-19 mai 2004)

Thème iii): Les produits à source ouverte et les consortiums de logiciels dans le domaine statistique

**Les logiciels libres et l'utilisation des technologies
de l'information dans les services de statistique**

Document d'appui

Présenté par le Comité d'État de statistique d'Azerbaïdjan¹

Résumé

Introduction

1. Les logiciels libres résultent de l'élaboration collective de programmes mis librement à la disposition de toutes les personnes intéressées, avec leurs codes source. La méthode des logiciels libres se différencie de la méthode d'élaboration classique «fermée» de la façon suivante:

- La création d'un logiciel ne passe pas par l'élaboration détaillée et l'approbation d'un projet unique;

¹ Auteur: Faiğ Jaliöv (ssc@azstat.org).

- Le logiciel est créé par un collectif virtuel de programmeurs indépendants, ouvert à tous ceux qui souhaitent participer;
- Les documents de travail et les textes des programmes sont entièrement ouverts à tous, et non seulement à ceux qui prennent part à l'élaboration.

2. L'objet du présent document n'est pas d'analyser le mouvement des logiciels libres, ni ses aspects psychologiques, moraux, sociaux et commerciaux. Il a été beaucoup écrit sur ce sujet, en particulier par des chefs de file du mouvement comme Richard Stallman, Eric Raymond, Tim O'Reilly et d'autres.

3. Nous nous contenterons d'aborder la question de l'utilisation de logiciels libres dans les services de statistique et d'étudier les perspectives d'application de cette méthode à l'élaboration de progiciels de traitement de données statistiques.

L'expérience de l'utilisation de logiciels libres

4. Il s'agit en premier lieu de l'utilisation de systèmes pour la construction du réseau corporatif du Comité d'État de statistique (Goskomstat), à savoir:

- **Le système d'exploitation Linux.** Depuis 2001, nous sommes passés entièrement de Novell et de Windows NT à Linux. À l'heure actuelle, tous les serveurs de notre réseau corporatif utilisent ce système d'exploitation;
- **Composants Internet-Intranet.** Service du domaine DNS sur la base BIND, serveur de fichiers Samba, serveur Web Apache intégré avec Perl, PHP et MySQL (notre site www.azstat.org utilise ce serveur), serveur de messagerie Sendmail, serveur de procuration Squid, passerelle et pare-feu iptable, etc.

5. En second lieu, il s'agit de l'utilisation des langues de programmation diffusées librement et de MySQL pour le traitement d'applications statistiques client-serveur. Dans ces applications, on utilise également une grande quantité de programmes et de modules diffusés librement. Il serait bien sûr dénué de sens de refuser systématiquement les produits commerciaux. Pour mettre au point des applications, nous utilisons en général les outils et systèmes d'élaboration de programmes des sociétés Microsoft, Borland, Oracle et autres.

Les perspectives d'élaboration de logiciels de statistique par la méthode des logiciels libres

6. Peut-on procéder au traitement de données statistiques au moyen de la méthode des logiciels libres? Dans l'affirmative, quel type de tâche peut-on effectuer et comment peut-on organiser le travail? Les services de statistique emploient habituellement un nombre limité de programmeurs et ceux-ci ne peuvent se charger de l'ensemble des tâches. Il faut donc soit augmenter les effectifs, soit confier une partie des travaux à des sociétés extérieures. Dans un cas comme dans l'autre, la dépense est importante et les inconvénients majeurs. La méthode ouverte, qui permet d'effectuer des tâches en réunissant un collectif virtuel de partisans des mêmes idées, peut être une solution partielle, mais aussi une solution à part entière. Examinons quelques aspects de la question:

- **Type de tâche.** Les tâches que des services de statistique peuvent être appelés à accomplir de manière ouverte doivent être suffisamment universelles et se prêter à des solutions originales et intéressantes. C'est une condition nécessaire pour réunir un collectif important de programmeurs indépendants. Il faut être conscient que n'est certes pas l'altruisme qui pousse les programmeurs à s'associer à un projet. Ceux-ci sont des utilisateurs potentiels qui, comme nous, sont intéressés par le résultat final. Il est inutile de compter sur la réunion d'un large collectif virtuel si la tâche proposée correspond à un domaine d'application très restreint.
- **Gestion du projet.** Le projet peut être publié sur un site de travail. Cependant, il ne doit pas être lancé à partir d'une proposition vide. Il importe de proposer dès le début un programme de travail avec des textes de départ. Ce programme peut être préliminaire, contenir des erreurs et fonctionner imparfaitement. Ce n'est pas important. Ce qui prime, c'est l'idée. Le projet doit s'adresser aux spécialistes hautement qualifiés susceptibles de se joindre au collectif.

L'élaboration de logiciels de statistique par la méthode corporative

7. Il s'agit d'une méthode de mise au point collective d'applications par quelques organisations intéressées. Ce travail peut être organisé en vertu d'un accord entre ces organisations. Les tâches mêmes peuvent être effectuées à l'aide de la méthode classique ou de la méthode des textes ouverts. La méthode corporative peut s'appliquer aux tâches qui ne sont pas suffisamment universelles et qui correspondent à un domaine d'application restreint. Il ne faut pas s'attendre, dans ce cas, à réunir un large collectif virtuel de programmeurs indépendants. Ces tâches n'intéressent que quelques services de statistique, qui peuvent aussi se charger conjointement de leur réalisation.

8. Compte tenu des restrictions concernant le volume des documents, nous nous sommes contentés de mentionner certaines questions et d'indiquer certaines directions. Nous espérons présenter ultérieurement, dans la version intégrale du document, une analyse plus approfondie du problème.

9. Les logiciels libres sont un mode d'élaboration collective de programmes selon lequel ceux-ci sont diffusés, avec les codes source, à tous ceux qui le souhaitent. Ce mouvement n'est pas parti de rien. Dès les premiers temps d'Unix et d'Internet, les programmeurs ont échangé leurs textes source. Lorsque la société AT&T s'est vu interdire toute activité commerciale dans le domaine de la technologie informatique, Ken Thompson et Dennis Ritchie ont commencé à distribuer le code source d'Unix à tous ceux qui le souhaitaient. Grâce à cela, des milliers de programmeurs ont eu la possibilité de développer librement ce système d'exploitation, y intégrant leurs idées et leurs programmes. Unix s'est ensuite développé sur une base commerciale, ce qui n'a pas influé de manière très positive sur son sort ultérieur. C'est vers la fin des années 70 qu'a commencé à se créer spontanément parmi les développeurs indépendants et les utilisateurs d'Unix et de quelques autres programmes un réseau d'échange de messages, de fichiers et d'informations qui, par la suite, est devenu Internet.

10. Aujourd'hui, le mouvement des logiciels libres a pris une ampleur considérable. Il représente un mode d'élaboration selon lequel:

- La création d'un logiciel ne passe pas par l'élaboration détaillée et l'approbation d'un projet unique;
- Le logiciel est créé par un collectif virtuel ouvert à tous ceux qui souhaitent participer;
- Les documents de travail et les textes des programmes sont entièrement ouverts à tous, et non seulement à ceux qui participent.

11. Linux. Les produits Linux sont fondés sur un texte de base écrit par Linus Torvalds et contiennent en général des centaines d'autres ensembles de textes ouverts, dont la plupart des plus récents sont également exploitables dans d'autres environnements, y compris Unix, Windows, Windows NT, MacOS, etc.

- FreeBSD, OpenBSD et les autres produits de l'Unix de Berkeley. À l'heure actuelle, on accorde surtout de l'attention à Linux, alors que le développement de BSD Unix a également une grande importance.
- Les moyens de programmation. Dans le cadre du projet GNU, financé par la *Free Software Foundation*, un ensemble d'outils de programmation perfectionnés a été créé, dont le compilateur GCC pour la traduction du langage C, les compilateurs g++ et C++, l'éditeur Emacs et le débogueur GDB. Par ailleurs, deux autres éléments absolument inestimables de la culture des textes ouverts ont été créés en dehors du projet GNU, à savoir le programme patch de Larry Wall, qui permet aux développeurs d'échanger des petites corrections dans des programmes plutôt que de se transmettre l'ensemble des textes, et le système CVS, qui permet de travailler à partir de différentes versions des textes source.
- Les langages. La première place parmi les langages élaborés selon le principe des textes ouverts revient au Perl de Larry Wall; les langages Tcl de John Ousterhout et Python de Guido van Rossum ont aussi de nombreux partisans.
- Apache. Comme on l'a mentionné précédemment, le groupe Apache domine le marché des serveurs Web. D'après l'enquête la plus récente de la société Netcraft (www.netcraft.co.uk/survey), 54 % de l'ensemble des serveurs Web présents sur le réseau fonctionnent à l'aide d'Apache, 23,5 % ont été conçus à l'aide d'outils Microsoft et 6,5 % à l'aide d'outils Netscape. Un groupe d'environ 12 développeurs principaux, secondés par une communauté d'utilisateurs nombreuse et active, continuent de perfectionner Apache et de lui apporter leur soutien.
- Samba. Ce produit a été élaboré par un groupe international dirigé par l'Australien Andrew Tridgell. Il permet aux systèmes Unix et Linux de servir de serveurs de fichiers et de serveurs d'impression dans les réseaux tournant sous Windows 95/98/NT. Il utilise la technologie de «l'homme invisible», qui permet aux administrateurs d'incorporer Linux au réseau et d'administrer celui-ci de façon invisible.
- Sendmail. Ce système, mis au point initialement en tant que composant de BSD Unix, est aujourd'hui un élément essentiel de l'infrastructure du système de messagerie sur Internet.