



World Library and Information Congress: 71th IFLA General Conference and Council

"Libraries - A voyage of discovery"

August 14th - 18th 2005, Oslo, Norway

Conference Programme:

<http://www.ifla.org/IV/ifla71/Programme.htm>

July 13, 2005

Numéro de code :

007-F

Session :

101 Bibliothèques en sciences et technologies

CRIS + Libre accès = vers une diffusion du savoir scientifique sur le GRID

Keith G Jeffery

Director IT,

CCLRC^{NdT} Rutherford Appleton Laboratory, Chilton, Royaume-Uni

E-mail: kgj@rl.ac.uk

Traduction par Sandrine AVRIL,
INIST-CNRS, sandrine.avril@inist.fr

Résumé

En associant les CRIS (Current Research Information Systems ou systèmes d'information sur la recherche en cours) aux systèmes en libre accès, on réunit des systèmes de gestion de la R&D et des systèmes de libre accès aux publications scientifiques – la partie la plus visible de la production en R&D – au sein d'une infrastructure GRID européenne émergente. Le débat autour du libre accès est très vif, avec deux tendances qui s'opposent et se complètent : la voie « verte » (auto-archivage dans des archives institutionnelles) et la voie « dorée » (paiement de la publication par l'auteur/l'établissement). Les grands éditeurs expérimentent des services de type « dorés » et les archives institutionnelles « vertes » connaissent une croissance rapide. Les GRID sont aujourd'hui envisagés, notamment dans les NGG Reports (Next Generation GRID ou GRID de nouvelle génération), comme un espace informatique européen, actuellement mis en œuvre progressivement, sous la direction, notamment, de la DG INFSO F2 de la Commission

^{NdT} Council for the Central Laboratory of the Research Councils

européenne. Ils sont destinés à fournir un accès facile à l'information et au calcul. Les systèmes CRIS proposent un cadre pour l'évaluation des publications scientifiques (y compris la compréhension de leur environnement), un cadre de gestion de la R&D dans les organismes de recherche (des laboratoires nationaux aux universités) par les organismes de financement, ainsi qu'un mécanisme assurant l'interopérabilité de l'information sur la recherche et le développement.

1 Introduction

Nous vivons dans un monde de concurrence et d'évaluation et le milieu de la recherche scientifique n'est pas épargné par ce phénomène : la plupart des pays évaluent la qualité de leurs universités et organismes de recherche sur la base de résultats mesurables et prennent des décisions budgétaires en conséquence. Les publications scientifiques sont l'un de ces résultats mesurables.

Les systèmes CRIS fournissent l'information nécessaire à la gestion de la recherche et développement, aux transferts de technologie et à la création de richesse. Dans le monde de la R&D actuel, les CRIS sont souvent utilisés par les organismes de financement de la recherche, les organismes de recherche et les universités. Néanmoins ils portent davantage sur l'environnement de la recherche (projets, chercheurs, organismes, équipement, financement) que sur les résultats de la recherche à proprement parler (brevets, produits, publications).

La publication des résultats de la recherche (ou la transmission de l'information ou du savoir) est la pierre angulaire du développement de la culture technologique de l'humanité depuis des millénaires : depuis les peintures rupestres d'animaux (destinées peut-être à la formation des futurs chasseurs ou à honorer un projet de chasse) jusqu'aux dessins et écrits de Léonard de Vinci (les premiers rapports techniques). L'apparition de techniques d'impression bon marché a induit une croissance explosive des publications, et une certaine évaluation de la qualité était devenue nécessaire. Les sociétés savantes ont commencé à critiquer les publications, qui étaient souvent sous forme manuscrite et lues en public, la version ancestrale de l'évaluation par les pairs actuelle. Aujourd'hui, une publication électronique multimédia accessible par un hyperlien, comportant un code exécutable et des données associées, peut être évaluée par toute personne au moyen d'un commentaire en ligne.

Le fait d'extérioriser les concepts de l'esprit des chercheurs, de les consigner par écrit (avec les résultats et observations qui y sont associés), permet une conservation posthume des travaux des chercheurs et rend leurs résultats reproductibles et diffusables. Certains estiment que cette « conservation externe de la mémoire » est le signe distinctif de l'humanité.

C'est précisément pour parvenir à cette vision holistique de la recherche (et non seulement de ses résultats, dont les plus évidents sont les publications, mais aussi de son contexte), que nous souhaitons allier les systèmes CRIS et les bases de données de publications. Et plus spécifiquement, nous souhaitons utiliser la technologie des GRID pour associer les CRIS et les archives ouvertes de publications et créer ainsi les conditions pour une e-recherche ou recherche en ligne. L'utilisateur final peut ainsi avoir un accès électronique à des travaux antérieurs, au contexte de la recherche, aux expériences et aux données/logiciels. Le cas échéant, l'utilisateur peut suivre ou reproduire à distance des expériences et publier des articles nouveaux grâce à une

plate-forme performante et facilement accessible d'édition de documents électroniques. Tout cela constitue le corpus de la connaissance scientifique.

Parallèlement, un gestionnaire de la recherche peut analyser en contexte les résultats de la recherche et n'est plus obligé de se fier uniquement par exemple au seul nombre de publications. De plus, l'industriel peut utiliser les données du CRIS pour trouver de nouvelles idées créatrices de richesse.

Le World Wide Web et les moteurs de recherche qui y sont associés permettent un accès facile à l'information, même si sa pertinence (exactitude) et son obtention (exhaustivité) peuvent parfois être controversées. La technologie des GRID, issue du métacalcul (ou le fait de relier entre eux plusieurs supercalculateurs pour obtenir un ordinateur géant), s'est aujourd'hui bien développée et a intégré des services Web dans une architecture OGSA (Open GRID Services Architecture) : elle est donc devenue transparente grâce au Web et a vu ses capacités de calcul augmentées. Le principe du GRID en Europe (mais pas nécessairement en Amérique du Nord) est le suivant : l'utilisateur final lance une requête, l'ensemble du système GRID interprète la requête, fait une offre à l'utilisateur final (qui peut être financière et/ou porter sur un échange de droits), et, si l'offre est acceptée, exécute la requête en utilisant des agents ou courtiers qui utilisent des sources et ressources hétérogènes.

Le processus de la recherche peut être considéré comme un flux de production dans lequel les résultats, ainsi que les données produites, les logiciels et les références croisées à d'autres travaux, sont enregistrés à différents stades : idées de départ, propositions de projets, rapports intermédiaires et publication finale.

Ceci peut être résumé en une pseudo-équation : CRIS + Libre accès = vers une diffusion de la connaissance scientifique sur le GRID.

La suite de cet article aborde les thèmes de la connaissance scientifique, des CRIS, du libre accès, des GRID et fait une synthèse de tous ces thèmes.

2 La connaissance scientifique

Le contexte de la recherche, du développement et de l'innovation génère de la propriété intellectuelle sous la forme de produits, de brevets et de publications (au sens large, à savoir, de toute représentation tangible de l'intellect humain). Alors que les publications scientifiques traditionnelles (littérature blanche) produisent une propriété intellectuelle visible ; la littérature grise, qui constitue la « partie immergée de l'iceberg », demande encore à être organisée. Ceci représente généralement la base de son savoir ou de son savoir-faire [Je99], [JeAsRe00]. La question soulève certains points juridiques : de nombreuses entreprises ou organismes protègent leurs droits de propriété intellectuelle par des brevets ou des prépublications ; la législation américaine sur la liberté de l'information et protection des données fait contrepoids au droit d'auteur et au droit sur les bases de données. L'innovation, les transferts de technologie, la création de richesses, la qualité de la vie font partie des grandes priorités de la recherche et du développement, et attirent les investissements des gouvernements, des entreprises privées, des associations philanthropiques, voire des personnes physiques. La majorité de la technologie que nous utilisons couramment aujourd'hui est la résultante de la R&D des décennies antérieures, et

de même, nous devons la qualité de vie que nous connaissons aujourd'hui en grande partie à la R&D dans les domaines de la médecine, de l'éducation et de l'environnement. C'est cela la propriété intellectuelle.

Chaque entreprise ou organisme a besoin d'utiliser la propriété intellectuelle qu'elle produit pour ses activités commerciales (y compris les activités de R&D) et pour ses activités de marketing ou de relations publiques. Ceci suppose que tout organisme ou entreprise doit connaître la propriété intellectuelle qu'elle détient, la cataloguer, la préserver et comprendre les bénéfices commerciaux qu'elle peut en retirer, pour documenter les décisions à prendre à l'avenir en matière d'investissements en R&D, investissements qui généreront à leur tour de la propriété intellectuelle. La base de cette connaissance est constituée par la littérature blanche, mais aussi par la littérature grise, véritable iceberg, qui enfouit le savoir-faire de l'entreprise ou de l'organisme dans des rapports techniques, des manuels d'utilisation, de formation, entre autres. De plus, la propriété intellectuelle réside de plus en plus dans des données brutes (par exemple les résultats des essais cliniques sur les médicaments), dans des bases de données (par exemple, informations sur les relations clients) ou dans des logiciels (qui englobent les processus métiers de l'entreprise). Ce sont autant d'informations qui peuvent être soumises à l'examen attentif d'audits ou de demandes au titre de la liberté de l'information ; les entreprises ou organismes sont donc fortement incités à prendre des mesures pour bien gérer ces ressources en propriété intellectuelle.

Dans le monde de la R&D, la plupart des organismes publics de financement évaluent aujourd'hui la production (la propriété intellectuelle) des organismes qu'ils financent (par exemple, une université ou un organisme de recherche) et fondent leurs décisions de financement au moins partiellement sur de telles évaluations. C'est pourquoi l'archivage, la conservation et la gestion de la propriété intellectuelle sont d'une importance cruciale pour les organismes de recherche. De même, pour de nombreuses entreprises modernes, la qualité de la propriété intellectuelle est un facteur déterminant pour sa réussite et pour les investissements que sont prêts à consentir les actionnaires. Nous en déduisons donc qu'il est nécessaire de disposer d'un système et d'un environnement qui permettent à l'entreprise ou à l'organisme de gérer de manière efficace sa propriété intellectuelle en rapprochant la propriété intellectuelle (par exemple la littérature grise) des structures organisationnelles et des objectifs de l'établissement.

3 CRIS (*Current Research Information Systems*: systèmes d'information sur la recherche en cours)

D'un point de vue historique, les CRIS (ou systèmes d'information sur la recherche en cours) sont issus du monde de la recherche d'informations. Leur création a été motivée par le fait que l'information archivée se présentait souvent sous une forme similaire à celle des fiches catalographiques des bibliothèques {titre/ auteur / date / mots-clés / adresse de la source}. Utilisés de manière autonome, ces systèmes étaient appropriés jusqu'à ce que d'autres besoins se fassent jour, comme les analyses statistiques, l'intégration de données provenant d'autres SGBD, la souplesse de l'édition de rapports (y compris l'intégration de l'environnement client) et la gestion des données multimedia (ou hypermedia). Ces besoins sont apparus dès le milieu des années 1970. Toutefois, ces systèmes se sont vite avérés inadéquats dès lors qu'il est devenu nécessaire d'accéder à l'information dans des CRIS multiples et hétérogènes (ou autres systèmes de bases de données) géographiquement distants : il manquait un modèle théorique sous-jacent

(qui aurait permis de faire des correspondances structurelles), des normes communes d'enregistrement des données (qui auraient permis une interopérabilité), des interfaces externes communes (qui auraient permis une intégration des environnements clients à la fois pour la saisie/mise à jour et pour l'interrogation/ l'édition de rapports). Et pire encore, il manquait les métadonnées qui auraient permis une interopérabilité instantanée si nécessaire.

Le CERIF^{NdT} propose un modèle de données complet pour l'information en R&D élaboré conjointement par les représentants des pays européens (de l'Union européenne et des Etats associés). Les premières recommandations du CERIF de 1991 ont servi de base au projet pilote ERGO [ERGO]^{NdT}. Le nouveau modèle de données du CERIF a été élaboré par le groupe de travail sur la révision du CERIF (*CERIF Revision Working Group*) [CERIF]. CERIF2000 propose un modèle de données pour les projets, organismes ou entreprises, les personnes, le financement, les manifestations, l'équipement, les installations, ainsi que pour toutes les relations entre ces différentes composantes. Ce modèle confère une grande flexibilité et prévoit des relations à la fois du type 1:n (hiérarchiques) et du type n:m (plusieurs à plusieurs ou graphes). En outre, ces relations font intervenir des rôles et sont définies dans le temps, ce qui donne un modèle de données très riche. Le modèle a pris en compte l'existence de brevets, produits et publications, mais a supposé la pré-existence de systèmes pour gérer les détails de ces types d'information. CERIF2000 a donc redéfini les CRIS en les formalisant pour proposer aux créateurs de CRIS une plate-forme stable, facilitant l'échange de données entre différents CRIS et la saisie de métadonnées décrivant de manière succincte les contenus des CRIS.

La Commission européenne a par la suite transmis la gestion et le développement du CERIF à [EuroCRIS] ; en effet, si les dernières versions du CERIF ont su répondre aux besoins des utilisateurs d'une manière structurée, le CERIF inclut désormais des données détaillées sur les entités et attributs pour la publication d'information. Les principales propositions ayant débouché sur ce modèle sont décrites dans [Je99], [JeAsRe00], [AsJe04].

4 Libre accès

La publication électronique est devenue, grâce au World Wide Web [W3C], facile et peu coûteuse. Ceci s'est traduit par une croissance explosive des archives institutionnelles, mais aussi thématiques, comme [arXiv]. L'initiative en faveur du libre accès (*Open Access Initiative* [OAI]) utilise le standard de métadonnées du Dublin Core [DC] et des logiciels de collecte de métadonnées (protocole OAI-PMH) pour relier entre elles les archives.

Comme le souligne [Be99], les deux grands défis pour le Web sont le Web sémantique, pour le rendre intelligible, et les réseaux de confiance (*web of trust*), pour le sécuriser. Le Web sémantique est en cours de construction, en grande partie par la mise en œuvre (a) de structures de données plus formelles qui se prêtent mieux à la logique de premier ordre qui implique généralement l'utilisation de métadonnées structurées et (b) d'ontologies par domaine, qui donnent des définitions des termes et indiquent les relations logiques entre les termes, qui constituent des métadonnées associées. Le réseau de confiance suppose que les documents sont protégés de toute utilisation frauduleuse et que l'entité qui détient l'information l'utilise d'une

^{NdT} Common European Research Information Format

^{NdT} European Research Gateways On-line

manière conforme à la déontologie. Ceci est garanti par des métadonnées associées restrictives liées à l'information originale.

Malheureusement, le Dublin Core n'est pas formalisé : il est donc lisible par une machine, mais n'est pas compréhensible par une machine. A bien des égards, il représente un recul par rapport au format [MARC]. Il n'a pas les propriétés du web sémantique ni des réseaux de confiance, ce qui limite sérieusement son intérêt pour le traitement automatisé, qui exigent comme condition préalable nécessaire des métadonnées formalisées. Les métadonnées sont des données sur les données ; une classification a été proposée (initialement en 1998, puis publiée dans [Je00]) qui distingue différents types de métadonnées et permet un traitement logique correct. L'application de la classification des métadonnées aux CRIS a été décrite dans [JeLoAs02]. Une version formalisée du Dublin Core a été développée progressivement dans [Je99], [JeAsRe00], [AsJe04].

Les archives en libre accès supposent que l'accès est gratuit au point d'entrée. Il existe deux grands modèles pour le libre accès : la « voie dorée », proposée principalement par les éditeurs, où c'est l'auteur ou son établissement qui a la charge des frais de publication. La « voie verte », en revanche, consiste à demander à l'auteur de déposer ses articles dans une archive (institutionnelle ou thématique) à peu près au moment de la publication, de manière à ce que les documents soient librement accessibles en ligne. Cela soulève des problèmes de droit d'auteur avec certains éditeurs, bien que la plupart autorisent à présent le libre accès à expiration d'un certain délai après la date de publication dans leurs revues. Bien entendu, le modèle « vert » du libre accès permet aussi le dépôt des prépublications et/ou de la littérature grise, mais ce type de documents doit bien être distingué des publications évaluées par un comité de lecture. Aujourd'hui, la majorité des éditeurs proposent un accès en ligne à leurs titres par abonnement ; le problème est que le chercheur doit utiliser différents systèmes présentant chacun une interface différente (y compris des identifiants et mots de passe différents) pour accéder aux titres des différents éditeurs. Ils sont nombreux à trouver que l'effort à fournir est trop important... et à s'en remettre à Google (ou Google Scholar), sans se soucier du manque d'exhaustivité ou de la pertinence/précision médiocre de l'interrogation.

5 Les GRID

5.1 Rappels

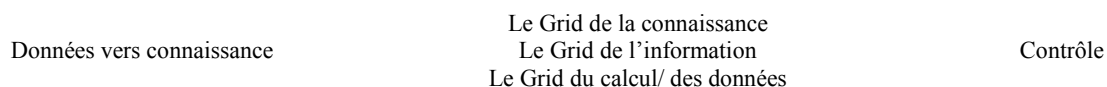


Figure 1: L'architecture GRIDs en 3 couches

Dans les années 1998-1999, la communauté des chercheurs du conseil pour la recherche du Royaume-Uni (*UK Research Council*) s'est trouvée confrontée à plusieurs problèmes informatiques. Leurs thèmes de recherche portaient sur la compréhension de la postgénomique, des changements climatiques, les études océanographiques, la surveillance et la modélisation de la pollution environnementale, la science des matériaux, l'étude des procédés de combustion, les sciences pour l'ingénieur de pointe, la conception de médicaments et la simulation et le

traitement de données dans le domaine de la physique des particules. Les chercheurs avaient donc besoin de processeurs plus puissants, de plus grandes capacités de stockage des données, de meilleurs moyens d'analyse et de visualisation, le tout associé à des outils faciles à utiliser et commandés par une interface utilisateur intuitive. Il a été demandé à l'auteur de cet article de proposer une architecture informatique intégrée.

L'architecture proposée se compose de trois couches (Figure 1). Le Grid de calcul ou de données, qui constitue la couche inférieure, comporte des supercalculateurs, des grands serveurs, des installations et des dispositifs spécialisés de stockage de volumes importants de données (par exemple, pour la réalité virtuelle, VR), le tout étant interconnecté par des réseaux à grande vitesse. Les grandes fonctions de cette couche du Grid incluent le calcul de la répartition de charge / du partitionnement des algorithmes, la résolution d'adresses de la source des données, la sécurité, la réplication et le reroutage des messages. Le Grid de l'information se superpose au Grid de calcul/ de données et résout les problèmes d'accès homogène à des ressources hétérogènes, notamment par l'utilisation des métadonnées et de logiciels intermédiaires (*middleware*). Enfin, la couche supérieure est le Grid de la connaissance qui utilise la technologie de la découverte de connaissances dans les bases de données pour générer du savoir, mais aussi pour permettre une représentation de la connaissance par les travaux scientifiques, les publications validées par les pairs et la littérature grise, cette dernière renvoyant par des liens hypertexte à des informations corroborant les hypothèses de la connaissance.

En parallèle, [FoKe98] a publié au tout début des réflexions britanniques sur le GRID une série d'articles dans un ouvrage qui s'est depuis imposé comme « la Bible du GRID ». L'idée de départ était de connecter entre eux plusieurs supercalculateurs pour générer une puissance plus grande : c'est la technique du métacalcul. Toutefois, la contribution la plus importante doit être recherchée dans les systèmes et protocoles de calcul de gestion des ressources. Le GRID correspond à la couche inférieure (couche calcul/ données) de l'architecture GRID proposée pour le Royaume-Uni.

5.2 L'architecture GRID

L'idée du GRID est de proposer un environnement informatique qui interagit avec l'utilisateur pour définir ses besoins en service, puis pour satisfaire ces besoins par le biais d'un environnement hétérogène constitué de bases de données intermédiaires, de puissance de traitement, d'installations spécifiques de systèmes de collecte et d'affichage de données, pour que l'utilisateur ait une impression d'environnement informatique homogène.

Les principaux composants (Figure 2) externes à l'environnement du GRID sont a) les utilisateurs, qui sont soit des personnes, soit des systèmes tiers ; b) les sources : données, informations ou logiciels ; c) les ressources, comme les ordinateurs, les capteurs, les détecteurs ou les installations de visualisation ou de réalité virtuelle. Chacun de ces trois grands composants est représenté activement et en continu dans l'environnement du Grid par : 1) les métadonnées, qui décrivent la composante externe et qui varie en fonction des modifications de l'environnement ; 2) un agent : qui agit pour le compte de la ressource extérieure et la représente dans l'environnement du GRID. Enfin, un des composants agit comme un intermédiaire entre les agents. Ce sont des intermédiaires qui, tout en étant des composants logiciels, agissent quasiment

de la même manière que des intermédiaires humains en négociant des contrats et des accords entre les agents. Il est donc clair que les principaux composants sont les métadonnées, les agents et les intermédiaires.

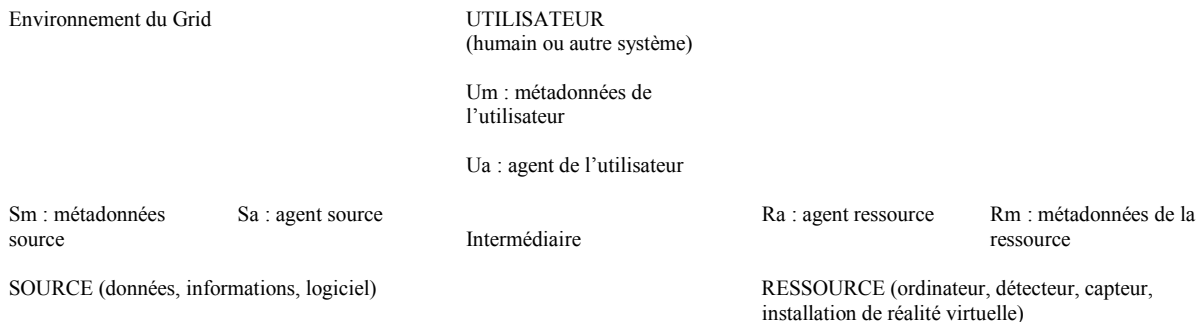


Figure 2 : Les composants du GRID

5.3 L'informatique ambiante (*ambient computing*)

Le concept de l'informatique ambiante suppose que l'environnement informatique est toujours présent et disponible d'une manière stable. Le concept de l'informatique diffuse (*pervasive computing*) suppose que l'environnement informatique est disponible partout et est présent « dans toute chose ». Le concept d'informatique mobile suppose que dispositif de l'utilisateur final peut être connecté même lorsque celui-ci est en déplacement. Dans l'acception générale du terme, l'informatique ambiante implique à la fois l'informatique mobile et l'informatique diffuse.

Une configuration classique peut comprendre : a) un micro-casque (écouteurs et microphone) pour les communications audio, connecté par une connexion locale sans fil Bluetooth à b) un assistant numérique personnel (PDA), équipé d'un petit écran, d'un clavier texte/numérique (à la manière d'un téléphone), de connexions GSM/GPRS (téléphonie mobile) pour la voix et les données, d'une connectivité LAN sans fil et de ports pour des capteurs (pour mesurer tout ce qui est à proximité de l'utilisateur final), à leur tour connectés par technologie Bluetooth à c) éventuellement à un ordinateur portable dans un sac à dos (mais sorti dès qu'il peut être utilisé dans un environnement adéquat), avec un écran normal, un clavier, un disque dur de grande capacité et une connexion GSM/GPRS, un réseau LAN sans fil, un câble LAN, et un téléphone.

L'utilisateur final n'utilisera peut-être que (a) et (b) (ou bien peut-être uniquement (b) avec les écouteurs et le microphone intégrés) dans un contexte social ou professionnel, en tant que téléphone mobile ou agenda et en tant qu'espace de loisirs avec ou sans connexion à des serveurs « maison » ou un environnement informatique. Pour les environnements de travail plus traditionnels nécessitant l'utilisation d'un écran et d'un clavier, on utilisera l'ordinateur portable, probablement sans l'assistant numérique personnel. Les deux peuvent être utilisés simultanément avec des logiciels de calibration/validation de collections de données sur l'ordinateur portable et des capteurs reliés à l'assistant numérique personnel.

Une telle configuration est évidemment pratique pour les « utilisateurs nomades » (voyageurs de commerce), pour les services d'urgence, comme les pompiers ou les services médicaux, pour les hommes d'affaires, pour les responsables de production dans l'industrie, pour l'industrie de la

distribution ou de la logistique (gestion des stocks, transport, livraison), pour les chercheurs sur le terrain.... Mais aussi pour les activités de loisir comme la randonnée pédestre, les visites de musées, pour trouver un restaurant ou pour visiter un site archéologique. L'idée est d'accéder à un environnement GRID, quel que soit le lieu, le moment et la manière choisis. L'informatique ambiante, c'est la puissance du GRID à portée de main. Combinée aux CRIS et aux archives en libre accès, elle offre au chercheur, au gestionnaire de la recherche ou à l'industriel une facilité d'accès à la connaissance dont il a besoin.

6. Synthèse

En conclusion, le concept du GRID, qui était à l'origine le métacalcul, c'est-à-dire l'interconnexion de plusieurs supercalculateurs [FoKe98], a été étendu (premiers articles en interne en 1999 et publication dans [Je01]) à un véritable environnement informatique distribué, avec comme services du GRID, le concept des services Web du W3C, associés aux concepts du Web sémantique et des réseaux de confiance. L'informatique ambiante permet de disposer d'un tel environnement n'importe où et avec le dispositif de son choix [Je04].

Cet environnement constitue une plate-forme permettant un rapprochement ultime de la littérature blanche et grise, depuis les publications validées par un comité de lecture jusqu'aux rapports techniques et manuels, en passant par les prépublications commentées, le tout étant accompagné de métadonnées formalisées permettant un traitement automatisé dans les archives en accès libre. Ces données comportent des références croisées à d'autres données et à des logiciels appropriés, mais aussi, au moyen de la technologie des CRIS [EuroCRIS], à des personnes, organismes, entreprises, projets, brevets, publications, manifestations, installations et équipement. L'environnement dispose en outre d'une puissance de calcul, d'installations de sortie spécifiques (par exemple, la réalité virtuelle) et d'un contrôle dynamique des détecteurs et des données de collecte des instruments. Il prend en charge l'ensemble de la chaîne de production de la recherche.

Dans notre laboratoire, le CCLRC, nous avons mis en place :

- (a) Une archive ouverte institutionnelle comportant plus de 20 000 entrées en Dublin Core formalisé, se prêtant donc à l'interopérabilité et au traitement automatique ; le modèle de données utilise aussi les concepts du modèle IFLA [IFLA FRBR] ;
- (b) Une archive de données pour entreprise utilisant le modèle de données du CERIF (étendu pour autoriser la prise en charge de la gestion des processus métiers ainsi que les spécifications des CRIS) ;
- (c) Un environnement GRID faisant intervenir l'informatique ambiante facilitant l'interopérabilité et l'accès de l'utilisateur à l'information et aux installations de calcul.

Nous sommes en train d'intégrer toutes ces fonctions et de retravailler les processus métiers organisationnels pour parvenir à l'environnement de la recherche en ligne du futur. Un tel environnement devrait fournir une base efficace et rentable pour la gestion de la R&D.

Remerciements

Le travail sur les CRIS a été réalisé principalement en collaboration avec les collègues d'euroCRIS, que je remercie pour leur contribution. Le travail sur le CERIF et sur les relations entre l'information sur les publications et les CRIS a été réalisé pour la majeure partie en

collaboration avec Anne Asserson, de l'université de Bergen. Le travail sur les archives en libre accès a été réalisé pour l'essentiel par mon équipe du CCLRC, notamment Matthew Mascord, Catherine Jones, Brian Matthews et Heather Weaver, qui sont une source perpétuelle d'inspiration et de nouvelles idées. Le travail sur le GRID a été enrichi par des discussions avec les membres du programme e-Science du Royaume-Uni et par des discussions à un niveau plus général dans un contexte européen, principalement par le biais du projet [ERCIM]^{NdT} et du groupe [NGG]^{NdT}. La présente synthèse ainsi que l'article ont été rédigés par mes soins.

Bibliographie

[ArXiv] www.arxiv.org

[AsJeLo02] Asserson,A; Jeffery,K.G; Lopatenko,A: 'CERIF: Past, Present and Future' in Adamczak,W & Nase,A (Eds): Proceedings CRIS2002 6th International Conference on Current Research Information Systems; Kassel University Press ISBN 3-0331146-844 pp 33- 40 2002 (available under www.eurocris.org)

[AsJe04] Asserson, A; Jeffery, K.G.; 'Research Output Publications and CRIS' in A Nase, G van Grootel (Eds) Proceedings CRIS2004 Conference, Leuven University Press ISBN 90 5867 3839 May 2004 pp 29-40 (available under www.eurocris.org)

[Be99] Berners-Lee,T; 'Weaving the Web' 256 pp Harper, San Francisco September 1999 ISBN 0062515861

[CERIF] <http://www.cordis.lu/cerif/>

[DC] http://purl.oclc.org/metadata/dublin_core/

[ERCIM] <http://www.ercim.org> [ERGO] <http://www.cordis.lu/ergo/>

[EuroCRIS] <http://www.eurocris.org>

[FoKe98] Foster I and Kesselman C (Eds). The Grid: Blueprint for a New Computing Infrastructure. Morgan-Kaufman 1998

[IFLA] <http://www.ifla.org/>

[IFLA FRBR] <http://www.ifla.org/VII/s13/frbr/frbr.pdf> (Je99) Jeffery, K G: 'An Architecture for Grey Literature in a R&D Context' Proceedings GL'99 (Grey Literature) Conference Washington DC October 1999 <http://www.konbib.nl/greynet/frame4.htm>

^{NdT} European Research Consortium for Informatics and Mathematics
^{NdT} National Grid Group

[JeAsRe00] Jeffery K.; Asserson A.; Revheim J; (2000) CRIS, Grey Literature and the Knowledge Society, Proceedings CRIS-2000, Helsinki
ftp://ftp.cordis.lu/pub/cris2000/docs/jeffery_fulltext.pdf

[Je00] Jeffery, K.G., 2000, 'Metadata': in Brinkkemper,J; Lindencrona,E; Solvberg,A: 'Information Systems Engineering' Springer Verlag, London 2000. ISBN 1-85233-317-0. 10

[Je01] Jeffery,K.G.;'GRIDs: Next Generation Technologies for the Internet' Invited Keynote Presentation (Abstract in Proceedings) Eds Wang,Y; Patel,S; Johnston,R.H; OOIS2001 Conference, Calgary August 2001, page 1, Springer, ISBN 1-85233-546-7

[JeLoAs02] Jeffery,K.G; Lopatenko,A; Asserson, A.: 'Comparative Study of Metadata for Scientific Information: The Place of CERIF in CRISs and Scientific Repositories' in Adamczak,W & Nase,A (Eds): Proceedings CRIS2002 6th International Conference on Current Research Information Systems; Kassel University Press ISBN 3-0331146-844 pp 77- 86 (available at www.eurocris.org)

[Je04] Jeffery, K.G.; 'GRIDs, Databases and Information Systems Engineering Research' in Bertino,E; Christodoulakis,S; Plexousakis,D; Christophies,V; Koubarakis,M; Bohm,K; Ferrari,E (Eds) Advances in Database Technology - EDBT 2004 Springer LNCS2992 pp3-16 ISBN 3-540-21200-0 March 2004

[Je04a] Jeffery, K.G.; 'The New Technologies: can CRISs Benefit' in A Nase, G van Grootel (Eds) Proceedings CRIS2004 Conference, Leuven University Press ISBN 90 5867 3839 May 2004 pp 77-88 (available at www.eurocris.org)

[MARC] <http://minos.bl.uk/services/bsds/nbs/marc/commarcm.html>

[NGG] www.cordis.lu/ist/grids

[OAI] www.openarchives.org

[W3C] www.w3.org

[W3Cmetadata] <http://www.w3.org/Metadata/>