
Doit-on et peut-on libérer les collections informatiques ?

Jean M. Thiéry

ModLibre.info
575 A chemin de Rastel, F-13510 Éguilles, France
Jean.Thiery@ModLibre.info

RÉSUMÉ. Les logiciels constituent un patrimoine immatériel très important. Ils doivent être conservés sous forme ouverte pour des raisons fondamentales (refus des boîtes noires) et pour des raisons juridiques (historique des règles de l'art). Ils devraient aussi être libérés pour garantir leur pérennité à long terme. L'article présente quelques exemples de logiciels à sauvegarder en priorité, dans le domaine scientifique et dans le domaine grand public, en particulier pour les micro-ordinateurs.

ABSTRACT. Software is a very important world heritage. It should be saved as open software for fundamental reasons (refusal of black boxes) or legal reasons (history of the current state of the art). Software should be liberated for durability. This article presents a list of software to be saved in priority, for science and for general use, especially on microcomputers.

MOTS-CLÉS : logiciel libre, logiciel ouvert, bioinformatique, chimie informatique, mathématiques, physique, statistiques, micro-ordinateur, station de travail, ordinateur centralisé.

KEYWORDS: free software, open software, bioinformatics, computational chemistry, mathematics, physics, statistics, microcomputer, workstation, mainframe.

1. Introduction

Les logiciels constituent un patrimoine immatériel très important qui doit être conservé avec soin. Les sources doivent rester ouvertes pour des raisons fondamentales (refus des boîtes noires) et pour des raisons juridiques (historique des règles de l'art).

Vers un Musée de l'Informatique et de la société Numérique en France ?

L'historien des sciences devrait pouvoir suivre l'évolution des logiciels qui ont permis de nouvelles découvertes. Le juriste devrait pouvoir retrouver la date certaine de telle ou telle règle de programmation.

L'expérience montre qu'il est très difficile de retrouver les sources d'un programme de plus de dix ans (et parfois moins). Les centres de calcul ont disparu, les laboratoires ont été réorganisés, les chercheurs ont changé de sujet, les sauvegardes se sont dégradées ... Dans certains cas, il ne reste plus que des supports démagnétisés, des cartes perforées ou un listing ! Sans parler de l'obsolescence technologique qui empêche de relire des supports a priori corrects !

Les logiciels libres ont plus de chance de résister à ces dégradations puisqu'ils peuvent être diffusés dans d'autres laboratoires et facilement recopiés à chaque changement technologique.

Cet article propose quelques pistes pour reconstituer un patrimoine fortement dégradé :

- pour la *chimie informatique* qui avait bénéficié, pendant quatre décennies, de l'existence du QCPE (*Quantum Chemistry Program Exchange*) ;
- pour plusieurs disciplines scientifiques (*mathématiques, statistiques, physique, chimie et bioinformatique*) ;
- et pour l'informatique quotidienne sur micro-ordinateurs.

Un Musée de l'Informatique pourrait jouer un rôle crucial pour la reconstitution et la sauvegarde de ce patrimoine.

2. L'exemple de la chimie informatique

2.1. Introduction

Toutes les disciplines scientifiques ont bénéficié de l'informatique, mais il est difficile de retrouver certaines informations pertinentes. Peu avant 1960, des chimistes ont commencé à utiliser les ordinateurs disponibles pour effectuer des calculs en chimie quantique. Ils eurent l'idée de mettre en commun leurs programmes de calculs. Richard W. Counts créa le QCPE (*Quantum Chemistry Program Exchange*) qu'il anima jusque dans les années 2000. Les informations circulèrent avec des Lettres informelles (Newsletters), puis avec un Bulletin trimestriel à partir de 1981 (ISSN 0889-7514).

Au début le QCPE était centré essentiellement sur la chimie quantique. Il s'intéressa progressivement à d'autres domaines de la chimie informatique (*computational chemistry*) : analyses de spectres (infrarouge, optique, RMN, RPE, ...), cristallographie, ... et vers les années 1990 à certains aspects de la bioinformatique.

Le dépouillement des Bulletins du QCPE et la recherche de quelques dates complémentaires permettent de tracer un historique de la chimie informatique et des

ordinateurs disponibles. La plupart des dates ci-dessous correspondent aux dates de disponibilité dans les laboratoires, souvent plus tardives que les dates d'annonce des logiciels, des matériels et des moyens de transferts (supports magnétiques, réseaux, ...).

2.2. *Quelques dates importantes pour la chimie informatique*

1962-04 : Première *Newsletter* du QCPE citée pour son 30ème anniversaire (QCPE Bull. 1992, 12-2, p. 27).

1964 : Commercialisation du CDC 6600 considéré comme le premier super-ordinateur (http://en.wikipedia.org/wiki/Control_Data_Corporation).

1965-03 : Digital Equipment lance le PDP-8 optimisé pour l'instrumentation scientifique (<http://fr.wikipedia.org/wiki/PDP-8>).

1976 : Lancement du Cray-1 super-ordinateur à architecture vectorielle ([http://fr.wikipedia.org/wiki/Cray_\(entreprise\)](http://fr.wikipedia.org/wiki/Cray_(entreprise))).

1980-02 : Publication du catalogue de logiciels du NRCC (*National Resource for Computation in Chemistry*) (NRCC 1980). Le NRCC avait été créé pour fournir des moyens et du personnel consacrés au développement de la chimie et des sciences apparentées grâce à l'utilisation du calcul intensif.

1981-10-01 : Fermeture du NRCC. Les bandes magnétiques sont transmises au QCPE (QCPE Bull. 1981, 1-4, p. 63).

1983-02 : Le rôle de la chimie informatique pour les recherches en catalyse est confirmé par une publication dans *Science* (QCPE Bull. 1983, 3-1, p. 5).

1983-08-23 : Le *Wall Street Journal* publie un article de Ronald Alsop sous le titre « Scientists are turning to computers in search for new chemicals, drugs » (QCPE Bull. 1983, 3-4, p. 77-81).

1984-05 : Les micro-ordinateurs commencent à être utilisés en chimie informatique. « correctement configurés, ils permettent des calculs toujours plus importants ; si importants que l'on oublie qu'ils sont des ordinateurs 16 bits » (QCPE Bull. 1984, 4-2, 35). Mais ceci n'est vrai qu'avec un coprocesseur mathématique bien géré par le compilateur *Fortran* (QCPE Bull. 1984, 4-3, p. 54).

1984-08 : Le QCPE annonce la distribution de logiciels pour PC sur des disquettes 5"1/4 (QCPE Bull. 1984, 4-3, p. 54).

1984-08 : Distribution des deux premiers programmes pour IBM-PC. Ce sont les adaptations de programmes antérieurs de chimie quantique (programmes CNDO/INDO et MNDO) (QCPE Bull. 1984, 4-3, p. 75-76). Ces programmes sont écrits en *Fortran* qui restera le langage dominant. D'autres langages apparaîtront ultérieurement : *Basic*, *Turbo-Pascal*, ... et même *APL*, *C*, *Modula-2*.

1985-05 : James J. P. Stewart présente une communication sur la portabilité des programmes informatiques (QCPE Bull. 1985, 5-2, p. 51-54).

1985-08 : Donald B. Boyd résume l'enquête « Profile of Computer-Assisted Molecular Design in Industry ». La chimie informatique est en pleine expansion dans l'industrie. La moitié des ordinateurs utilisés sont des ordinateurs départementaux VAX 11 de Digital Equipment. Les logiciels proviennent du QCPE, d'universités ou de jeunes sociétés spécialisées (QCPE Bull. 1985, 5-3, p. 85-91).

1985-11 : Un éditorial résume l'histoire déjà ancienne du calcul parallèle. Le facteur limitant semble être au niveau des algorithmes, surtout en chimie quantique où les calculs sont difficilement parallélisables (QCPE Bull. 1985, 5-4, p. 113). Par contre la chimie quantique utilise beaucoup de calculs matriciels pouvant bénéficier de la vectorisation.

1986-05 : Analyse de la station de travail Apollo et référence à d'autres stations (Sun, MicroVAX II de DEC et RT-PC d'IBM) toutes bien adaptées à la chimie informatique (QCPE Bull. 1986, 6-2, p. 39-41).

1987-02 : Premières craintes sur l'évolution des *Copyrights* (QCPE Bull. 1987, 7-1, p. 3).

1987-05 : Utilisation d'un outil de calcul formel (*Macsyma*) pour la programmation en chimie informatique (QCPE Bull. 1987, 7-2, p. 69-72).

1987-08 : Une note d'Edgar Soulié « Good Programs Based on Unknown Algorithms » insiste sur la nécessité de publier les programmes ET les algorithmes (QCPE Bull. 1987, 7-3, p. 118-119).

1988-05 : Distribution des deux premiers programmes pour Apple Macintosh II. Ce sont les adaptations de programmes antérieurs de chimie quantique (programmes FORTICON8 et SCFMO écrits en *Fortran*) (QCPE Bull. 1988, 8-2, p. 102-103).

1989-02 : Version No 1 de la licence GNU GPL de Richard Stallman (2010).

1989-05 : Présentation des procédures de tests du QCPE (QCPE Bull. 1989, 9-2, p. 47-48).

1989-05 : Les nouvelles stations de travail seront des plate-formes Unix basées sur des processeurs RISC (QCPE Bull. 1989, 9-2, p. 52).

1990-02 : « On estime maintenant qu'en 1990 les revenus d'IBM provenant de systèmes personnels dépasseront ceux de la vente d'ordinateurs centralisés » (QCPE Bull. 1990, 10-1, p. 2).

1990-02 : La communication « Toward an Undergraduate Computational Chemistry Curriculum » de K. R. Fountain et Jon Salmon présente des logiciels pour l'enseignement de la chimie informatique (QCPE Bull. 1990, 10-1, p. 3-4).

1990-08 : Le QCPE commence à distribuer des logiciels avec des licences liées au statut des laboratoires. Les logiciels restent ouverts mais ils enfreignent au moins une des quatre libertés de la GPL, même pour les laboratoires « académiques » (QCPE Bull. 1990, 10-3, Annexe).

1992-05 : Le *Lawrence Livermore National Laboratory* remplace l'un de ses superordinateurs CRAY par une ferme de stations de travail IBM RS/6000 (QCPE Bull. 1992, 12-2, p. 30).

1993-02 : Compte tenu des équipements des utilisateurs, les programmes pour PC sont désormais envoyés par défaut sur des disquettes 3"1/2 au lieu de 5"1/4 (QCPE Bull. 1993, 13-1, p. 7).

1993-08 : Le téléchargement des programmes par « ftp » devient possible pour un grand nombre de correspondants du QCPE (QCPE Bull. 1993, 13-3, p. 44-45).

2.3. Peut-on libérer la chimie informatique ?

La majorité des programmes du QCPE étaient dans le domaine public car ils avaient été financés par des crédits fédéraux américains ou la recherche publique d'autres pays. Par ailleurs, les sources étaient indispensables pour permettre des adaptations avant recompilation car chaque marque d'ordinateur avait « son *Fortran* » !

À partir de 1985, la chimie informatique devint un enjeu industriel, comme beaucoup d'autres aspects de l'informatique. Des logiciels du domaine public furent intégrés dans des produits commerciaux et les sources des nouvelles versions ne furent plus disponibles.

En 1989, Richard Stallman (2010) créa la licence GNU GPL qui garantissait la pérennité du logiciel libre.

Pour libérer un logiciel du QCPE, il faudrait obtenir l'accord de tous ses ayants-droits. C'est probablement trop tard !

Par contre, il serait important de rassembler dans un Musée de l'Informatique tous les programmes du QCPE pour les raisons évoquées dans l'introduction : raisons historiques (recherches sur l'évolution de la programmation) et raisons juridiques (preuves d'antériorité). Cette logithèque est particulièrement importante car elle s'est adaptée, pendant plusieurs décennies, à l'évolution de la chimie informatique et à celle des matériels disponibles.

Il faudrait prendre contact avec l'*Indiana University* (Bloomington, IN 47405 USA) qui avait hébergé le QCPE pendant une quarantaine d'années. Cette université ne semble plus concernée : la plupart des liens Internet sont rompus ! Même le lien souvent cité : qcpe.chem.indiana.edu .

Les archives ont-elles été transmises dans une autre institution ? Ont-elles été privatisées ?

On pourrait retrouver une partie des programmes du QCPE dans les archives de certains laboratoires de chimie théorique. Les programmes pour IBM-PC et Macintosh II étaient distribués sur des disquettes (encore lisibles dans de bonnes conditions de conservation !). Peut-on encore relire les autres supports

magnétiques ? Doit-on numériser les listes et les documentations imprimées qui auraient survécu ?

Par ailleurs, il serait souhaitable que toutes les informations disponibles soient rassemblées sur un site Internet public, accessible par tous.

3. L'exemple d'autres disciplines scientifiques

3.1. Mathématiques et statistiques

De nombreux logiciels libres ont été développés pour les **mathématiques appliquées** :

- la bibliothèque de calculs matriciels *LAPACK* ;
- la bibliothèque scientifique *GSL (GNU Scientific Library)* ;
- le logiciel matriciel *Octave* et ses bibliothèques mathématiques ;
- le logiciel matriciel et graphique *Scilab* avec ses modules spécialisés et son outil de programmation graphique *Xcos* pour l'étude de systèmes dynamiques hybrides ;
- les développements récents de *SciPy (Scientific Python)*, etc.

Les **mathématiques fondamentales** font souvent appel à des calculs symboliques :

- utilisation très ancienne de langages spécialisés comme *Lisp* ;
- existence de nombreux logiciels libres de calculs symboliques optimisés dans tel ou tel domaine (*Axiom, GAP, Maxima, PARI/GP*, etc.) ;
- émergence récente de *Sage*, un système écrit en *Python* permettant l'intégration de nombreux logiciels mathématiques libres.

Les **statistiques** et les **géostatistiques** utilisent de plus en plus le langage *R* (version libre du langage *S*) et *GRASS*, un Système d'Information Géographique pour l'analyse et la gestion des données spatiales.

Beaucoup d'autres logiciels mathématiques libres sont référencés sur le site *Framasoft.net*.

3.2. Physique et chimie

Les recherches en physique et en chimie ont souvent recours aux logiciels généraux précédents complétés le cas échéant par des bibliothèques ou des logiciels spécifiques, disponibles sur Internet lorsqu'ils sont libres.

Historiquement, des centres de recherches comme le CERN ont joué pour la physique un rôle équivalent à celui du QCPE pour la chimie informatique.

3.3. Bioinformatique

En bioinformatique (au niveau de la biologie moléculaire et des génomes) on trouve un site spécialisé pour les logiciels ouverts : l'*Open Bioinformatics Foundation* (<http://www.open-bio.org/>). Beaucoup de ces logiciels sont aussi libres.

3.4. Le rôle d'un Musée de l'Informatique

Un Musée de l'Informatique pourrait sauvegarder les versions successives des logiciels. En particulier, il pourrait conserver tous les documents nécessaires pour suivre et comprendre la libération de certains logiciels comme *Maxima* (anciennement *Macsyma*) et *Scilab*.

4. Les micro-ordinateurs

4.1. L'importance des micro-ordinateurs

Beaucoup d'utilisateurs de l'informatique ont conservé des micro-ordinateurs qu'ils avaient appréciés pour des raisons diverses. Ils ont constitué ainsi des collections qui ont un intérêt indéniable pour l'histoire fulgurante de l'informatique. Certains de ces ordinateurs fonctionnent encore. D'autres pourraient être facilement réparés en prélevant des cartes ou des périphériques sur des ordinateurs de la même génération.

Les micro-ordinateurs sont aussi très utilisés pour la recherche depuis une vingtaine d'années, comme nous l'avons vu dans la section consacrée à la chimie informatique.

4.2. La libération des micro-ordinateurs

Les logiciels de ces ordinateurs sont majoritairement propriétaires. On trouve aussi des logiciels libres, des logiciels gratuits et quelques logiciels publics.

Les logiciels les plus anciens étaient souvent ouverts, mais ils ont été progressivement fermés à partir des années 80. Les logiciels libres ont été développés plus tard.

Les conditions d'utilisation des logiciels propriétaires sont très strictes. Ils sont généralement associés à un seul ordinateur. Comment le prouver au bout de quelques décennies ? Comment gérer un parc d'ordinateurs hétérogènes dans de telles conditions ? On peut espérer que les musées bénéficieront d'un assouplissement des contraintes juridiques.

L'existence de nombreux logiciels fermés pose un problème fondamental. Les musées doivent-ils se contenter de conserver des boîtes noires ? Peuvent-ils obtenir

l'ouverture des logiciels les plus anciens pour mieux enregistrer l'évolution de l'informatique ?

On pourrait libérer les ordinateurs les plus anciens en chargeant des logiciels libres à la place des logiciels originaux. Par exemple, pour beaucoup de micro-ordinateurs, on peut trouver sur Internet de nombreux logiciels libres compatibles *MS-DOS*, comme *FreeDOS* (<http://www.freedos.org/>).

L'élimination des logiciels originaux serait très critiquable car elle reviendrait à *réécrire l'histoire* ! Par ailleurs, les logiciels originaux n'ont pas tous été réécrits en libre, en particulier au niveau des pilotes. Qui aurait le courage de faire de la rétro-ingénierie pour des périphériques dépassés !

Nous avons testé la *cohabitation* entre les logiciels originaux et leurs équivalents libres en séparant bien les répertoires correspondants. En sélectionnant correctement les chemins d'accès on peut utiliser la machine en *mode original* ou en *mode libéré*.

Le *mode original* est plus satisfaisant pour les historiens des technologies. En particulier ceux qui s'intéressent à l'évolution des premiers systèmes d'exploitation se rendent vite compte que *MS-DOS* avait beaucoup à apprendre de ses concurrents (*DR-DOS*, *Prologue*, etc.). Dans cette compétition, les considérations commerciales ou industrielles l'avaient emporté sur les considérations techniques.

Le *mode libéré* est préférable pour l'enseignement et la recherche, car on évite de travailler avec des boîtes noires.

4.3. Le rôle d'un Musée de l'Informatique

Nous espérons que suffisamment de micro-ordinateurs anciens seront conservés dans des structures pérennes comme les musées et qu'ils resteront une source de réflexion et d'inspiration pour les historiens des technologies, les développeurs, les enseignants et le grand public.

5. Les stations de travail, les ordinateurs départementaux et les ordinateurs centralisés

La plupart des stations de travail avaient été développées avec des systèmes d'exploitation dérivés d'Unix. On pourrait sans doute trouver des versions ouvertes de ces systèmes.

Les principaux problèmes se situeront au niveau de la maintenance des stations : où trouver les compétences et les pièces de rechange ?

Les difficultés seront encore plus grandes pour les ordinateurs départementaux et les ordinateurs centralisés.

Peut-on envisager une simulation des systèmes d'exploitation les plus anciens ?
Peut-on simuler des démonstrations convaincantes ?

Sinon, il faudra se contenter de présenter à l'arrêt quelques ordinateurs emblématiques.

6. Conclusion

Cet article présente de nombreux souhaits pour la sauvegarde et la libération des logiciels scientifiques et des logiciels d'usage quotidien. Certaines tâches comme les sauvegardes à long terme ne peuvent être réalisées que dans un cadre institutionnel. D'autres tâches pourraient être mutualisées avec la participation des associations directement concernées.

Bibliographie

NRCC (1980). *Software catalog*. Lawrence Berkeley Laboratory, University of California.

Stallman R.M., Williams S., Masutti C. (2010). *Richard Stallman et la révolution du logiciel libre*, Eyrolles, Paris. p.171.

Biographie

Jean M. Thiéry gère le site ModLibre.info (<http://ModLibre.info/>) promouvant des ressources libres pour l'éducation et la recherche. Il participe aux associations Aful, Axul, OLPC-France et SFBT.