

Cahiers **GUT** *enberg*

© NORMES ET FONTES

© Daniel DARDAILLER

Cahiers GUTenberg, n° 4 (1989), p. 2-8.

<http://cahiers.gutenberg.eu.org/fitem?id=CG_1989__4_2_0>

© Association GUTenberg, 1989, tous droits réservés.

L'accès aux articles des *Cahiers GUTenberg*

(<http://cahiers.gutenberg.eu.org/>),

implique l'accord avec les conditions générales

d'utilisation (<http://cahiers.gutenberg.eu.org/legal.html>).

Toute utilisation commerciale ou impression systématique
est constitutive d'une infraction pénale. Toute copie ou impression
de ce fichier doit contenir la présente mention de copyright.

Normes et fontes

Daniel DARDAILLER

Centre de recherche Bull, Sophia-Antipolis, 06560 Valbonne. daniel@mirsa.mirsa.fr

Résumé Cet article va essayer de faire le point sur les diverses normes et standards, de droit ou de fait, que l'on rencontre dans le domaine des polices de caractères en informatique.

On verra que des noms de polices jusqu'aux types de courbes mathématiques utilisées pour représenter les images des lettres, il existe un besoin pressant de normalisation pour rendre les processus de création et d'échange de documents moins complexes qu'ils ne le sont actuellement.

Abstract *This article gives an overview of the different standards and norms, official or not, that one can meet in the digital font domain today.*

We state that from the font names to the type of the mathematical curves used to represent the letter shapes, there is an urgent need of normalization, for making the processes of document creation and interchange less difficult than they currently are.

1. Introduction

Je suis informaticien et, de plus, je manipule assez souvent des fontes. Quand je vois sur un écran d'ordinateur une police de caractères qui me plaît, je m'empresse de demander « dans quel format est-elle ? » car j'envisage automatiquement deux choses :

1. elle n'est sûrement pas dans le format de mon système favori,
2. je vais trouver un traducteur adéquat ou en écrire un !

L'utilisateur normal, *seulement* informaticien par exemple, ne peut que regretter le premier point.

De même, si je veux imprimer cet article dans une fonte particulière, sachant que j'utilise \LaTeX et donc \TeX , je vais m'arranger pour traduire les métriques de cette police en fichier `.tfm`, puis utiliser les images des caractères avec le `.dvi` généré.

Si le format TFM (*TeX Font Metric*, qui n'est d'ailleurs en rien lié à \TeX) était le standard des formats de métriques, ou, autrement dit, s'il existait un standard des formats de métriques que \TeX puisse lire, mon travail serait déjà à moitié fait.

Dans un autre registre, quand j'écris une application graphique utilisant une ressource de type police, j'aimerais pouvoir être sûr que mon programme fonctionnera encore sur un autre site. J'aimerais *décrire* la police que je veux utiliser, plutôt que la référencer par un nom local (nom de fichier la plupart du temps !).

La liste des exemples de problèmes liés à la présence d'une multitude de normes différentes est bien longue et le seul fait, pour les auteurs, d'hésiter à chaque fois entre le mot *police* ou le mot *fonte* est déjà symptomatique du présent malaise.

2. Présentation

On peut distinguer, en matière de fontes en informatique, quatre sous-domaines pouvant être intéressés par une éventuelle normalisation (on écarte, ce faisant, les mécanismes de création proprement dits ou encore les techniques d'accès aux ressources citées) :

- le nommage des fontes, leur classification selon le style (romain, italique...), le corps, la famille, etc., ainsi que leurs propriétés typographiques ;
- le type d'encodage utilisé : où est le caractère représentant telle lettre et quel est son nom ou son code ?

- la forme des caractères : type de métrique utilisée, format de stockage des images en courbes ou matrices de points ?
- et enfin, la terminologie, le vocabulaire employé dans le domaine : qu'est-ce qu'une police, une fonte, qu'entend-on exactement par caractère ?

La figure 1 nous montre l'architecture générale d'utilisation de ces futures normes dans le contexte de la création d'un document [6]. Ce modèle est lui-même l'objet d'une normalisation (première partie de ISO 9541), mais je pense, comme le montrent les exemples que je porte sur cette figure, qu'il est déjà d'usage assez courant pour le prendre comme toile de fond de cet article.

Font Information Interchange Standard (ISO/IEC 9541) est donc la norme ISO qui couvrira le domaine des fontes informatiques (il n'existera bientôt plus, je pense, de fontes « tout court »). Issue d'un regroupement d'un groupe de travail ANSI (*American National Standards Institute*) avec un comité ISO (X3V1 et SC18), elle se divise en parties, s'occupant respectivement de l'architecture générale d'accès aux ressources de type fonte, des formats et sous-ensembles de fontes et des formes des caractères eux-mêmes. En 1989, toutes les parties ne sont pas encore disponibles.

3. Terminologie

Dans cette section, je ne parlerai que du standard ISO, bien que d'autres auteurs, et non des moindres — je pense en particulier à Richard Southall et Debra Adams dans [7] — aient aussi fourni leur part de définitions pertinentes.

Je n'en parlerai également que dans les termes anglo-saxons d'origine, laissant aux organismes officiels de traduction le soin de donner les nouvelles définitions pour la langue française (elles existent peut-être déjà).

Pour distinguer les caractères eux-mêmes et les codes de caractères, l'ISO introduit la notion de *glyph*¹. Un *glyph* est une abstraction de la lettre représentée, non associée à une forme particulière, ni à une fonte. Une *glyph representation* est l'association d'un *glyph* donné avec une forme (*glyph shape*) et une métrique précise (*glyph metric*), à l'intérieur d'une fonte donnée.

Au terme *font*, l'ISO suggère de substituer le terme *font resource* qui doit être considérée comme un regroupement de *glyph representations* avec de l'information descriptive globale à la fonte.

Je ne m'attarderai pas trop longuement sur cette nomenclature, sachons tout de même que les attributs typographiques eux-mêmes, c'est-à-dire les ressources de la fonte, font partie de l'ISO 9541 et s'organisent en noms structurés.

Par contre, il semble que le nommage des *font resources* elles-mêmes, vues de l'extérieur, ne soit pas inclu dans les travaux de l'ISO 9541. Il en va de même pour le nommage des *glyphs*, qui font l'objet d'une norme à part entière (voir la section sur l'encodage).

4. Nommage des fontes

Dans le processus de création de documents montré sur la figure 1, plusieurs entités ont besoin de référencer, de nommer les polices de caractères.

La manière de nommer les fontes uti-

¹En français, un glyphe est une ciselure, une marque en relief laissée par un outil pointu.

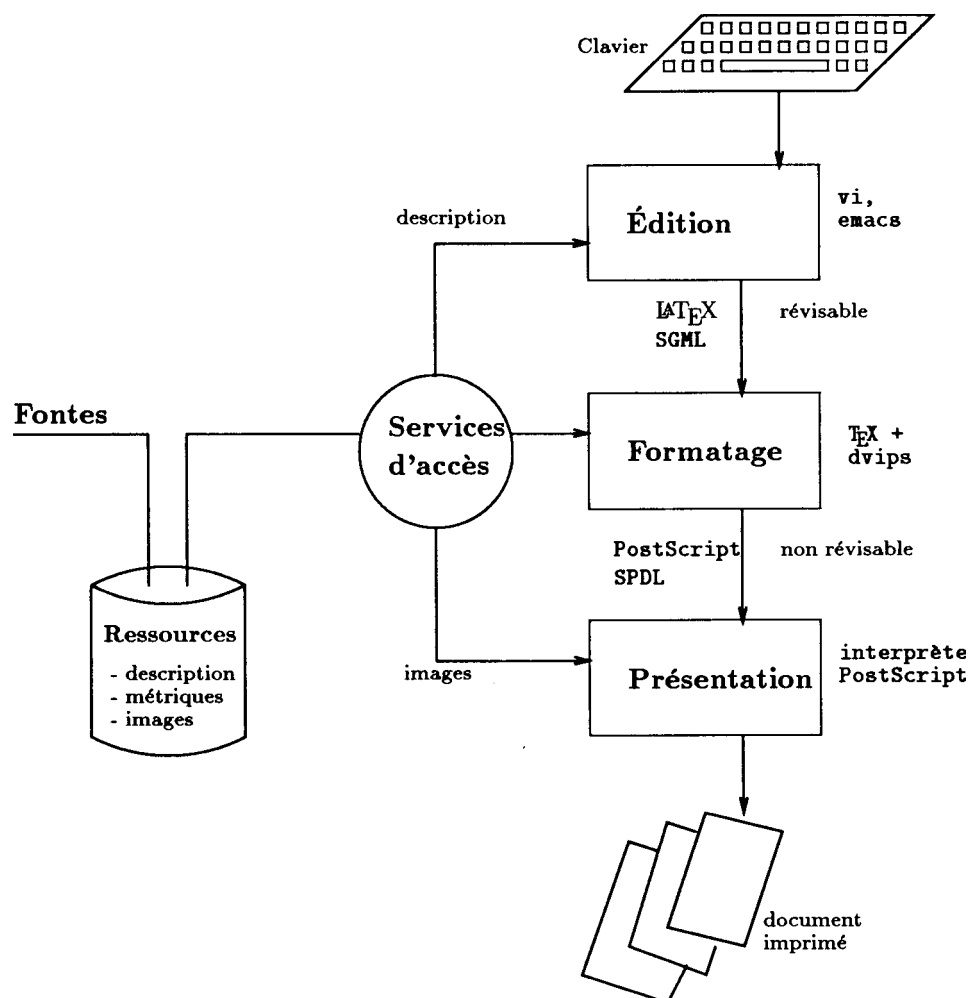


Figure 1 : Architecture d'accès aux ressources de type fonte.

lisées dans un programme informatique doit satisfaire plusieurs critères :

- être indépendante du système informatique sous-jacent, de sorte que le programme soit lui aussi indépendant ;
- être suffisamment descriptive pour permettre l'unicité de la représentation ;
- être extensible, pour pouvoir évoluer.

Plutôt que de faire la revue des systèmes de nommages « propriétaires », comme le catalogue Adobe pour PostScript (Times-Roman, Helvetica-Bold...), ou des divers systèmes de références par trop liés au programme lui-même (fonte numéro 17...), je préfère présenter le nouveau protocole de nommage du MIT/X Consortium, en vigueur dans le système de fenêtrage X11² [2] et répon-

²X11 est le nom abrégé du *X Windows System*, développé par le MIT en coopération avec la commu-

dant aux contraintes déjà citées.

Ce protocole, du nom de XLFD, pour *X Logical Font Description*, permet initialement aux clients d'un serveur X11 d'être totalement portables d'une plateforme à l'autre. Un document complet, rédigé par Jim Flowers (DEC), est d'ores et déjà disponible [8]. Je me contenterai donc d'en donner les grandes lignes.

XLFD décrit à la fois le nom des fontes et un ensemble de propriétés typographiques utilisables par le programmeur. Un nom de fonte est une chaîne de caractères (pris dans l'encodage ISO8859-1), qui décrit logiquement les caractéristiques de la police.

Exemple de nom de police :

```
-Adobe-Courier-Medium-R-Normal-\
-8-80-75-75-M-50-ISO8859-1
```

On désigne ici la police Courier normale (non graissée, ni condensée, ni oblique), digitalisée par la firme Adobe (une digitalisation de la même courbe par une autre firme amènera un nom différent) en corps 8 (8 pixels = 8 points dans ce cas puisque la résolution, 75 dpi, est approximativement égale au nombre de points typographiques par pouce, 72).

On trouve donc successivement dans le nom :

- le nom du fournisseur ou du vendeur de la police ;
- le nom de famille (Times, Helvetica...) ;
- la graisse, le degré de noirceur des caractères ;
- la pente, l'inclinaison des caractères ('I' pour italique, 'R' pour romain, ou droit...) ;
- la proportion (condensée, normale...) ;
- le style additionnel (serif, décoration...) ;
- le corps en pixel ;

nauté scientifique internationale. X, dans sa version 11, est devenu un protocole standard de communication réseau entre une application et un serveur graphique.

- le corps en décipoint typographique (10 unités = 1 point) ;
- la résolution horizontale en points par pouce ;
- la résolution verticale en points par pouce ;
- le type d'espacement de la police (proportionnel, fixe, video) ;
- la largeur ou chasse moyenne des caractères ;
- le nom de l'organisme d'encodage ;
- l'encodage de la police dans cet organisme.

Le caractère séparateur est le tiret (HYPHEN dans l'encodage ISO8859).

L'ensemble des noms d'encodages (deux derniers champs du nom de la police) n'est pas inclus dans le protocole lui-même mais est plutôt défini à partir d'une norme existante.

Ce protocole n'est donc pas lié *stricto sensu* au monde X-Windows et pourrait donc être candidat à une éventuelle normalisation en complément de l'ISO 9541 (mais je m'avance peut-être un peu). Il a en outre l'avantage d'être dédié à l'utilisateur final, non typographe, ce qui n'était pas le cas des classifications de type Vox.

5. Encodage des caractères

Une fonte contient donc un ensemble de caractères, métrique plus image, auquel on veut accéder avec des noms logiques. Je veux imprimer un *e accentué*, et pas toujours le caractère de numéro 351. De même, si dans le monde Unix occidental, les chaînes de caractères sont généralement codées implicitement en ASCII, une application voulant garder quelque chance d'être utilisée en Asie se doit de spécifier complètement ce genre de référence (ISO8859-1 par exemple).

L'AFII (*Association of Font Information Interchange*), organisme dépendant de l'ANSI, est chargée de l'enregistrement **mondial** des noms de caractères

(*glyph identifiers*). Ces noms seront ensuite utilisés aux différents niveaux de la norme ISO 9541 et deviendront eux-mêmes une norme sous le nom ISO/DIS 10036.

Plusieurs espaces de noms existent actuellement dans la nature, on peut citer de notre côté de la planète les divers ISO-Latin-1 (8859-1), Latin-2, etc., ainsi que les noms de caractères circulant chez Adobe ou Xerox. Le système de fenêtre X11 définit, par exemple, plus de 1000 noms logiques pour désigner les touches potentielles du clavier d'un poste de travail (les *keysyms*).

L'AFII s'intéresse à la fois à cet aspect occidental mais aussi au monde asiatique (nombreuses réunions à Hong Kong, Tokyo...) qui, avec ses milliers d'idéogrammes différents, nécessite un effort tout particulier. Le genre de question posée dans ce domaine est : « Doit-on enregistrer les caractères chinois, japonais ou taiwanais dans des ensembles disjoints basés sur des standards nationaux ou essayer de les fusionner ? » On se base en général dans ces études sur des standards locaux, comme le JIS (Japon) ou l'ACCC (Chine).

Il s'agit donc ici de constituer une véritable base de données de noms de caractères normalisés, organisée généralement en alphabet et où chaque organisme existant veut se faire entendre (l'*American Mathematical Association* avec \TeX math, aussi bien qu'IBM). L'enjeu est de taille, et ceux qui ont déjà essayé de rajouter des accents à un document PostScript après avoir créé leur propre vecteur d'encodage le comprennent bien.

6. Formes des caractères

Nous abordons ici un domaine moins stable : la représentation de l'image du caractère lui-même. La troisième partie de la norme ISO 9541 qui se consacre au *Font Interchange Shape Architecture and Format* est loin d'être finie et loin d'être acceptée par le marché.

Prenons les choses dans l'ordre, en commençant par les formats d'images *bitmap* qui stockent les images des caractères considérées comme des matrices de points.

6.1. Format bitmap

Cette façon de mémoriser les images des caractères est techniquement dédiée aux périphériques de faible résolution. On se dirige donc pour son étude vers ce qui se passe sur les écrans des postes de travail.

Le format BDF (*Bitmap Distribution Format*) [5], est le standard actuel du monde X-Windows. Il semble que sa syntaxe devienne un point de convergence pour les traducteurs d'autres formats. Il existe en effet dans le domaine public une multitude de programme `XXXtobdf` avec `XXX={gf, vfont, mac, laserjet...}`.

Ce format, créé par le concepteur de PostScript, Adobe, est organisé en lignes lisibles (par l'homme, c'est-à-dire en ASCII) et comprend un en-tête et une succession de « blocs » caractères, avec leurs métriques et l'image codée en hexadécimal. Il n'est pas complet, dans le sens où il manque des tables de ligatures et de crénages, mais il devrait évoluer.

Ce format est avant tout un formalisme d'*interchange* et il existe presque autant de programmes nommés `bdftoXXX` que de traducteurs dans l'autre sens. Dans l'architecture X11, par exemple, on le compile dans un format dépendant de la

machine, SNF (*Server Natural Format*) avant de l'utiliser pour un affichage performant.

Il n'est cependant pas exclu que ce type de stockage *bitmap*, cher car obligeant à garder plusieurs versions de la même fonte pour des corps différents et idiot car ne permettant pas de manipuler la lettre facilement (rotation, homothétie), disparaisse totalement au profit des modèles *outline* ou contour. Le progrès, à la fois des algorithmes et des processeurs, permet en effet d'introduire cette technique de pointe dans des systèmes jusqu'alors obligé de se limiter à la manipulation des seuls *bitmap*.

6.2. Format contour

Ici, on décrit la forme des caractères par des courbes mathématiques et ce sont les points de contrôle de ces courbes qui sont stockés de façon permanente. Un processus de numérisation va ensuite permettre au système d'affichage de déterminer les points discrets marquant le support.

Dans les faibles résolutions, un mécanisme de *hints*, que l'on peut traduire par « indications », aide en général la digitalisation à bien faire son travail. On rajoute avec ses indications une **sémantique typographique** aux courbes mathématiques de base, du style « Cette partie est un premier jambage de N, elle doit en priorité être de la même largeur que le second jambage du même N ».

Les précurseurs de ce type de format sont, par exemple, METAFONT [4], Ikarus [3] et, bien entendu, PostScript [1].

Aujourd'hui, c'est une vraie guerre que se livrent les professionnels comme Bitstream, Adobe, Sun, Compugraphic ou The Company, avec des systèmes nommés Fontware, Type 1³, Folio, Intellifont

ou Nimbus. L'enjeu est là aussi de taille quand on réfléchit aux répercussions sur l'utilisateur.

En effet, jusqu'à l'apparition de PostScript, de sa normalisation de fait comme langage de description de page et de la possibilité qu'il offre de donner à l'imprimante une nouvelle fonte venue de l'extérieur, l'utilisateur d'un système comme une Linotype, ou une Compugraphic, était obligé de s'adresser à Linotype ou à Compugraphic pour acheter une nouvelle police.

Durant quelques années, donc, PostScript a fait la loi et la fameuse Type 1 Adobe était la norme. Les derniers épisodes de la bataille des fontes, avec le revirement d'Apple en faveur du format Royal, lâchant ainsi Adobe, pourrait bien nous ramener cinq ans en arrière avec un nouveau marché fermé, sans loi d'offres et de demandes.

L'ISO 9541-3 pourrait bien avoir son mot à dire dans ce processus, mais il faut attendre de voir le DIS (*Draft International Standard*) sur les formes pour juger.

7. Conclusion

Que ce soit dans le domaine des documents (ODA, T_EX, PostScript...), ou dans celui des applications interactives (X, NeWs, NeXt), la nécessité d'une normalisation des données relatives aux fontes est déjà une réalité.

La prise en compte par les constructeurs de l'explosion prochaine du marché informatique dans le monde « non-latin » (asiatique, arabe, indien, russe...), a pour effet de sérieusement activer les efforts dans ce sens.

si elles utilisent la mécanique d'indications (*hinting scheme*) Adobe et un type 3 si elles sont définies par l'utilisateur.

³Rappelons que les fontes PostScript ont le type 1

Références bibliographiques

- [1] Adobe Systems Inc., *PostScript Language Reference Manual*. Addison-Wesley Publishing Company, Inc.
- [2] Robert Scheffler, James Gettys, and Ron Newman, *X-Windows System*. DEC Press, 1988.
- [3] Peter Karow, *Digital Format for Typefaces*. April 87, URW Verlag Hamburg, ISBN 3-9265515-5.
- [4] Donald E. Knuth, *The METAFONT book*. Computer & Typesetting / C. Addison-Wesley Publishing Company, Inc.
- [5] Adobe Systems, Inc., *Character Bitmap Distribution Format 2.1 Distribution* dans la bande X11 du MIT.
- [6] Smura, Beaton, Savage et Griffiee, "Font Information Interchange Standard, ISO/IEC 9541". *Computer Communications*, Avril 1989, Butterworth and Co., Eds
- [7] Debra Adams et Richard Southall, "Problems of font quality assessment", in *Raster Imaging and Digital Typography* (Jacques André et Roger Hersch, eds.), Cambridge University Press, 1989.
- [8] Jim Flowers, *X Logical Font Description Conventions*, Distribution dans la bande X11 du MIT. MIT/X Consortium, 1989.
- [9] *Adobe Font Metrics Files - Specification Version 2.0*, PostScript Developer Tools & Strategies Group. November 1, 1988.