Le circuit des métadonnées à travers les plateformes de mathématiques

Thierry Bouche

Dynamiques de l'édition scientifique,
de l'industrie de l'information,
de la documentation
Meudon, mardi 4 novembre 2014

Sommaire

Le CEDRAM

Introduction

- Un projet CNRS (INSMI) et UJF de soutien aux revues académiques de recherche en mathématiques (principalement)
- Réalisation : cellule Mathdoc (Grenoble)
- « Académiques » : indépendants (labos. . .) ou sociétés savantes
- Un ensemble d'outils de production
- Un portail de diffusion www.cedram.org
- En projet : élargir le périmètre et le nombre de services

Le CEDRAM

Principes

- Les articles sont des travaux de recherche originaux validés scientifiquement sous la responsabilité d'un comité éditorial
- Les actes de séminaires ou conférences suivent une politique plus légère (un responsable scientifique)
- Une fois publiés, les textes ne sont jamais modifiés
- La navigation des collections est libre, les métadonnées sont ouvertes largement, elles peuvent être révisées périodiquement, au fil des enrichissements (liens. . .)
- L'accès aux articles est libre à l'issue d'une barrière mobile (BM) raisonnable
  (À ce jour : 4 revues et 5 séries d'actes en accès libre, 1 revue BM = 2 ans, 2 revues BM = 5 ans)
- Les articles sont versés dans une bibliothèque numérique de référence (libre accès à terme, voie « orange »)

Le CEDRAM

Interopérabilité

- Les sources des articles sont archivées
  (pour leur préservation)
- Les articles sont versés dans NUMDAM 1-2 ans après parution
  (pour l'accès sur le long terme)
- NUMDAM est partenaire de EuDML et bientôt GDML
  (pour une plus grande visibilité et la construction d'une bibliothèque de référence couvrant toute la discipline)
- Plus généralement : OAI-PMH (oai_dc, mini-dml, NLM/EuDML)

Le circuit idéal d'un article

Eurêka	LaTeX auteur
Prépublication	arXiv/HAL
Soumission	Choix de la revue, rapport de referee
Production	Corrections, mise aux normes
Publication	Texte et métadonnées définitifs
Indexation	MathScinet, zbMath, Crossref
Archivage	indépendant, pérenne & distribué Projet : Private CLOCKSS européen
DML	Bibliothèque de référence en libre accès à terme *Partagée, mondiale, pérenne, sans but lucratif* Enrichissements globaux : consolidation des métadonnées, liens, similarité, *math mining*, etc.

L'environnement de production du CEDRAM

Objectifs

- Un chaîne de production *maîtrisable* par de petites structures
  - Uniquement des formats habituels en entrée : LaTeX, BibTeX...
- Produisant une édition de *qualité*
  - Une structure assez universelle pour permettre des maquettes Web et papier très variables mais des fonctions de base constantes
  - Éditions papier & électronique identiques
  - Métadonnées exactes (déduites des articles)
  - (Méta)données précises et universelles (Unicode, PDF, XML, MathML...)
  - Métadonnées compatibles avec NUMDAM, OAI-PMH, faciles à transformer et exporter (dégrader...)

L'environnement de production du CEDRAM

Principes

1. Les articles et leurs métadonnées sont produits à partir des mêmes sources LaTeX
2. Une métadonnée est déclarée *une fois* au plus, dans un fichier source *ad hoc*
3. Toute métadonnée qui n'est pas pertinente dans un fichier source ne doit pas s'y trouver (exemple : tomaison ou pagination dans un article)
4. Tout ce qui peut être calculé *doit* l'être
   Pas de duplication de métadonnées pas d'heuristiques ou de copier-coller
   (*un* processus) $\Longrightarrow$ pas d'erreurs de dernière minute !
5. L'environnement doit être d'appropriation facile

L'environnement de production du CEDRAM

Les niveaux

Revue

- Les constantes de la revue : titre, ISSN, maquette...
- Les quasi-constantes : pages de titre, ours, comité de rédaction, instructions aux auteurs...

Volume

- Données bibliographiques (année, mois, tome, fascicule, lieu...)
- Le folio du premier article
- La liste ordonnée des articles
- Variables globales et extras (éditorial, pubs...)

Article

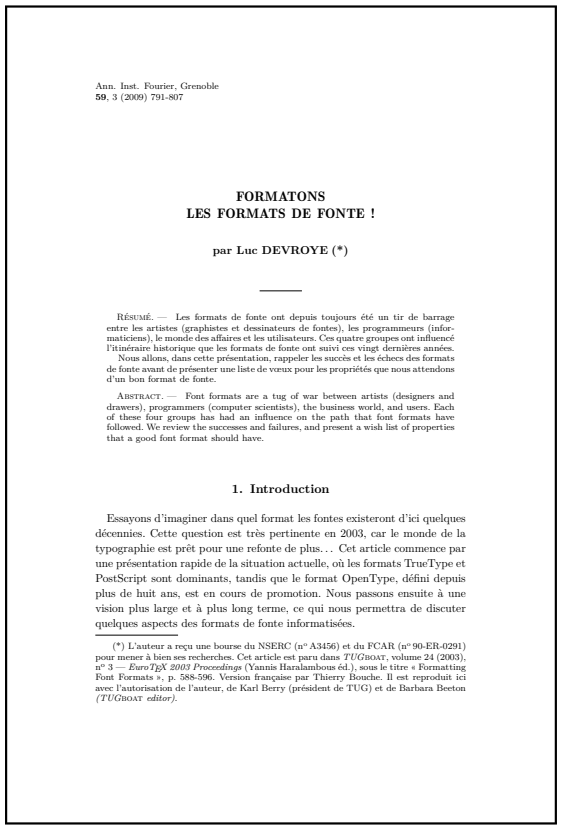
- Langue, titre, DOI, auteurs, résumés, texte, biblio, etc.

Pages

- Tous les numéros de pages sont calculés à la volée

Diversité graphique

Même source, présentation variable (1).



Ann. Inst. Fourier, Grenoble
59, 3 (2009) 791-807

FORMATONS
LES FORMATS DE FONTE !

par Luc DEVROYE (*)

Résumé. — Les formats de fonte ont depuis toujours été un tir de barrage entre les artistes (graphistes et dessinateurs de fontes), les programmeurs (informaticiens), le monde des affaires et les utilisateurs. Ces quatre groupes ont influencé l'itinéraire historique que les formats de fonte ont suivi ces vingt dernières années.
Nous allons, dans cette présentation, rappeler les succès et les échecs des formats de fonte avant de présenter une liste de vœux pour les propriétés que nous attendons d'un bon format de fonte.

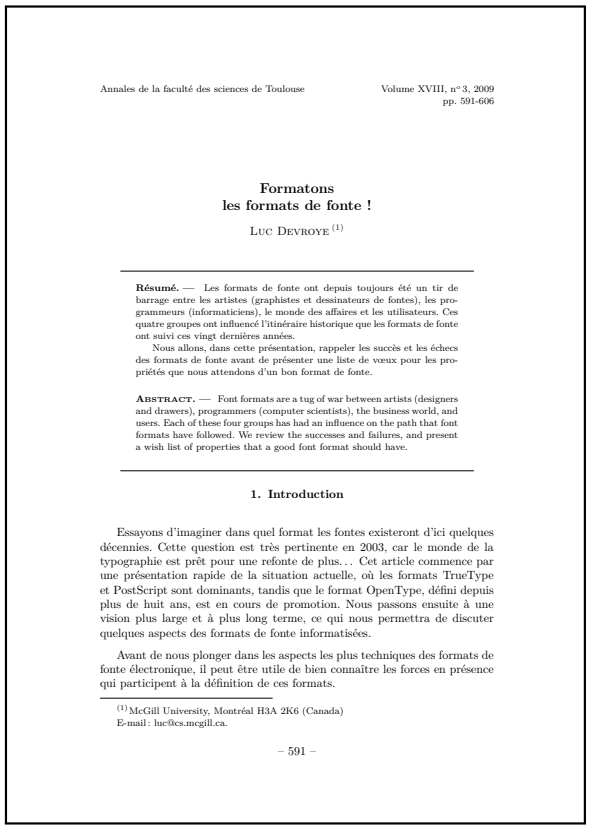
Abstract. — Font formats are a tug of war between artists (designers and drawers), programmers (computer scientists), the business world, and users. Each of these four groups has had an influence on the path that font formats have followed. We review the successes and failures, and present a wish list of properties that a good font format should have.

1. Introduction

Essayons d'imaginer dans quel format les fontes existeront d'ici quelques décennies. Cette question est très pertinente en 2003, car le monde de la typographie est prêt pour une refonte de plus… Cet article commence par une présentation rapide de la situation actuelle, où les formats TrueType et PostScript sont dominants, tandis que le format OpenType, défini depuis plus de huit ans, est en cours de promotion. Nous passons ensuite à une vision plus large et à plus long terme, ce qui nous permettra de discuter quelques aspects des formats de fonte informatisées.

(*) L'auteur a reçu une bourse du NSERC (n° A3456) et du FCAR (n° 90-ER-0291) pour mener à bien ses recherches. Cet article est paru dans *TUGboat*, volume 24 (2003), n° 3 — *EuroTEX 2003 Proceedings* (Yannis Haralambous éd.), sous le titre « Formatting Font Formats », p. 588-596. Version française par Thierry Bouche. Il est reproduit ici avec l'autorisation de l'auteur, de Karl Berry (président de TUG) et de Barbara Beeton *(TUGboat editor)*.

Annales de l'institut Fourier



Annales de la faculté des sciences de Toulouse Volume XVIII, n° 3, 2009
pp. 591-606

Formatons
les formats de fonte !

Luc Devroye [1]

Résumé. — Les formats de fonte ont depuis toujours été un tir de barrage entre les artistes (graphistes et dessinateurs de fontes), les programmeurs (informaticiens), le monde des affaires et les utilisateurs. Ces quatre groupes ont influencé l'itinéraire historique que les formats de fonte ont suivi ces vingt dernières années.
Nous allons, dans cette présentation, rappeler les succès et les échecs des formats de fonte avant de présenter une liste de vœux pour les propriétés que nous attendons d'un bon format de fonte.

Abstract. — Font formats are a tug of war between artists (designers and drawers), programmers (computer scientists), the business world, and users. Each of these four groups has had an influence on the path that font formats have followed. We review the successes and failures, and present a wish list of properties that a good font format should have.

1. Introduction

Essayons d'imaginer dans quel format les fontes existeront d'ici quelques décennies. Cette question est très pertinente en 2003, car le monde de la typographie est prêt pour une refonte de plus… Cet article commence par une présentation rapide de la situation actuelle, où les formats TrueType et PostScript sont dominants, tandis que le format OpenType, défini depuis plus de huit ans, est en cours de promotion. Nous passons ensuite à une vision plus large et à plus long terme, ce qui nous permettra de discuter quelques aspects des formats de fonte informatisées.

Avant de nous plonger dans les aspects les plus techniques des formats de fonte électronique, il peut être utile de bien connaître les forces en présence qui participent à la définition de ces formats.

[1] McGill University, Montréal H3A 2K6 (Canada)
E-mail : luc@cs.mcgill.ca.

– 591 –

Annales de la faculté des sciences de Toulouse (mathématiques)

Diversité graphique

Même source, présentation variable (2).

Annales mathématiques Blaise Pascal **17**, 13-31 (2010)

Formatons
les formats de fonte !

Luc Devroye

Résumé

Les formats de fonte ont depuis toujours été un tir de barrage entre les artistes (graphistes et dessinateurs de fontes), les programmeurs (informaticiens), le monde des affaires et les utilisateurs. Ces quatre groupes ont influencé l'itinéraire historique que les formats de fonte ont suivi ces vingt dernières années.

Nous allons, dans cette présentation, rappeler les succès et les échecs des formats de fonte avant de présenter une liste de vœux pour les propriétés que nous attendons d'un bon format de fonte.

Abstract

Font formats are a tug of war between artists (designers and drawers), programmers (computer scientists), the business world, and users. Each of these four groups has had an influence on the path that font formats have followed. We review the successes and failures, and present a wish list of properties that a good font format should have.

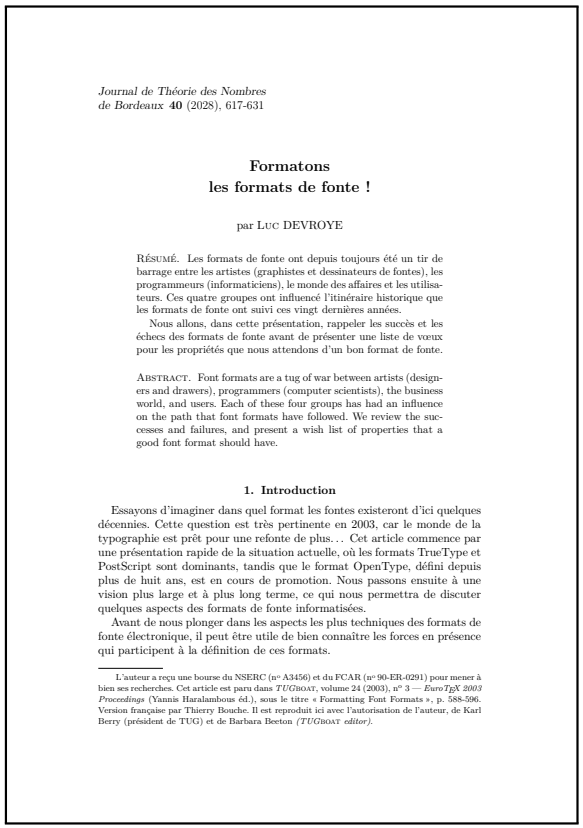
1. **Introduction**

Essayons d'imaginer dans quel format les fontes existeront d'ici quelques décennies. Cette question est très pertinente en 2003, car le monde de la typographie est prêt pour une refonte de plus… Cet article commence par une présentation rapide de la situation actuelle, où les formats TrueType et PostScript sont dominants, tandis que le format OpenType, défini depuis plus de huit ans, est en cours de promotion. Nous passons ensuite à une

L'auteur a reçu une bourse du NSERC (n° A3456) et du FCAR (n° 90-ER-0291) pour mener à bien ses recherches. Cet article est paru dans *TUG*boat, volume 24 (2003), n° 3 — *EuroTEX 2003 Proceedings* (Yannis Haralambous éd.), sous le titre « Formatting Font Formats », p. 588-596. Version française par Thierry Bouche. Il est reproduit ici avec l'autorisation de l'auteur, de Karl Berry (président de TUG) et de Barbara Beeton *(TUG*boat *editor)*.

13

Annales mathématiques Blaise-Pascal



Journal de Théorie des Nombres de Bordeaux **40** (2028), 617-631

Formatons
les formats de fonte !

par Luc DEVROYE

Résumé. Les formats de fonte ont depuis toujours été un tir de barrage entre les artistes (graphistes et dessinateurs de fontes), les programmeurs (informaticiens), le monde des affaires et les utilisateurs. Ces quatre groupes ont influencé l'itinéraire historique que les formats de fonte ont suivi ces vingt dernières années.

Nous allons, dans cette présentation, rappeler les succès et les échecs des formats de fonte avant de présenter une liste de vœux pour les propriétés que nous attendons d'un bon format de fonte.

Abstract. Font formats are a tug of war between artists (designers and drawers), programmers (computer scientists), the business world, and users. Each of these four groups has had an influence on the path that font formats have followed. We review the successes and failures, and present a wish list of properties that a good font format should have.

1. Introduction

Essayons d'imaginer dans quel format les fontes existeront d'ici quelques décennies. Cette question est très pertinente en 2003, car le monde de la typographie est prêt pour une refonte de plus… Cet article commence par une présentation rapide de la situation actuelle, où les formats TrueType et PostScript sont dominants, tandis que le format OpenType, défini depuis plus de huit ans, est en cours de promotion. Nous passons ensuite à une vision plus large et à plus long terme, ce qui nous permettra de discuter quelques aspects des formats de fonte informatisées.

Avant de nous plonger dans les aspects les plus techniques des formats de fonte électronique, il peut être utile de bien connaître les forces en présence qui participent à la définition de ces formats.

L'auteur a reçu une bourse du NSERC (n° A3456) et du FCAR (n° 90-ER-0291) pour mener à bien ses recherches. Cet article est paru dans *TUG*boat, volume 24 (2003), n° 3 — *EuroTEX 2003 Proceedings* (Yannis Haralambous éd.), sous le titre « Formatting Font Formats », p. 588-596. Version française par Thierry Bouche. Il est reproduit ici avec l'autorisation de l'auteur, de Karl Berry (président de TUG) et de Barbara Beeton *(TUG*boat *editor)*.

Journal de théorie des nombres de Bordeaux

Diversité graphique

Même source, présentation variable (3).

MathematicS in Action

MathematicS In Action
Vol. 4, 1-11 (2012)

Formatons les formats de fonte !

Luc Devroye *

* McGill University, Montréal H3A 2K6 (Canada)
E-mail address: luc@cs.mcgill.ca.

Résumé

Les formats de fonte ont depuis toujours été un tir de barrage entre les artistes (graphistes et dessinateurs de fontes), les programmeurs (informaticiens), le monde des affaires et les utilisateurs. Ces quatre groupes ont influencé l'itinéraire historique que les formats de fonte ont suivi ces vingt dernières années.

Nous allons, dans cette présentation, rappeler les succès et les échecs des formats de fonte avant de présenter une liste de vœux pour les propriétés que nous attendons d'un bon format de fonte.

Abstract

Font formats are a tug of war between artists (designers and drawers), programmers (computer scientists), the business world, and users. Each of these four groups has had an influence on the path that font formats have followed. We review the successes and failures, and present a wish list of properties that a good font format should have.

1. **Introduction**

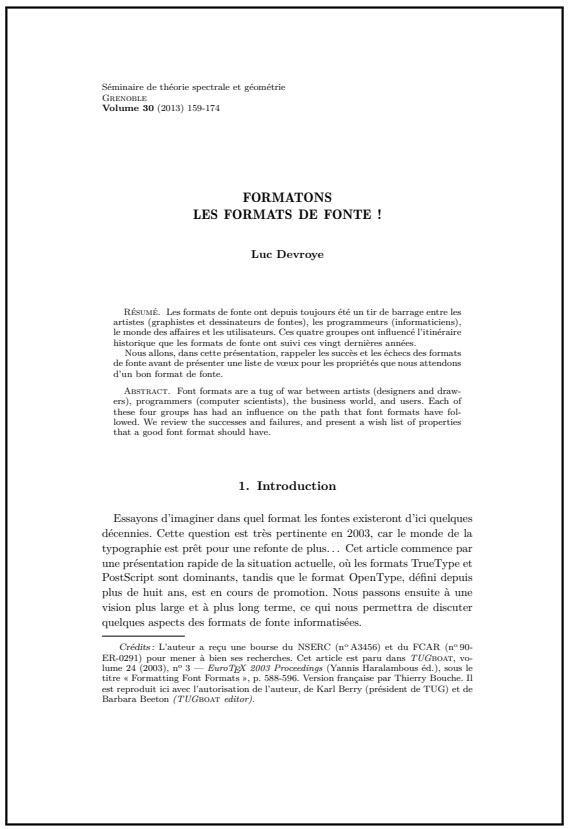
Essayons d'imaginer dans quel format les fontes existeront d'ici quelques décennies. Cette question est très pertinente en 2003, car le monde de la typographie est prêt pour une refonte de plus… Cet article commence par une présentation rapide de la situation actuelle, où les formats TrueType et PostScript sont dominants, tandis que le format OpenType, défini depuis plus de huit ans, est en cours de promotion. Nous passons ensuite à une vision plus large et à plus long terme, ce qui nous permettra de discuter quelques aspects des formats de fonte informatisées.

Avant de nous plonger dans les aspects les plus techniques des formats de fonte électronique, il peut être utile de bien connaître les forces en présence qui participent à la définition de ces formats.

D'abord et avant tout, les *utilisateurs* aimeraient disposer de formats simples et utiles, faciles à manipuler et modifier. Ils veulent à la fois mettre les mains sur les œuvres d'art créées par de grands typographes et les raffinements techniques fournis par les experts ès fontes numériques. En outre, les utilisateurs professionnels réclament un certain degré de flexibilité des fontes, de façon à pouvoir y inclure des choix personnels.

Les *artistes et typographes* avaient une influence considérable sur les formats de fonte à l'ère pré-électronique. Les premiers typographes étaient à peu près tous des artisans. Au vingtième siècle, diverses avancées technologiques furent réalisées par des compagnies comme Linotype ou

L'auteur a reçu une bourse du NSERC (n° A3456) et du FCAR (n° 90-ER-0291) pour mener à bien ses recherches. Cet article est paru dans *TUGboat*, volume 24 (2003), n° 3 — *EuroTEX 2003 Proceedings* (Yannis Haralambous éd.), sous le titre « Formatting Font Formats », p. 588-596. Version française par Thierry Bouche. Il est reproduit ici avec l'autorisation de l'auteur, de Karl Berry (président de TUG) et de Barbara Beeton *(TUGboat editor)*.



Séminaire de Théorie spectrale et géométrie

Séminaire de théorie spectrale et géométrie
Grenoble
Volume 30 (2013) 159-174

FORMATONS LES FORMATS DE FONTE !

Luc Devroye

Résumé. Les formats de fonte ont depuis toujours été un tir de barrage entre les artistes (graphistes et dessinateurs de fontes), les programmeurs (informaticiens), le monde des affaires et les utilisateurs. Ces quatre groupes ont influencé l'itinéraire historique que les formats de fonte ont suivi ces vingt dernières années.

Nous allons, dans cette présentation, rappeler les succès et les échecs des formats de fonte avant de présenter une liste de vœux pour les propriétés que nous attendons d'un bon format de fonte.

Abstract. Font formats are a tug of war between artists (designers and drawers), programmers (computer scientists), the business world, and users. Each of these four groups has had an influence on the path that font formats have followed. We review the successes and failures, and present a wish list of properties that a good font format should have.

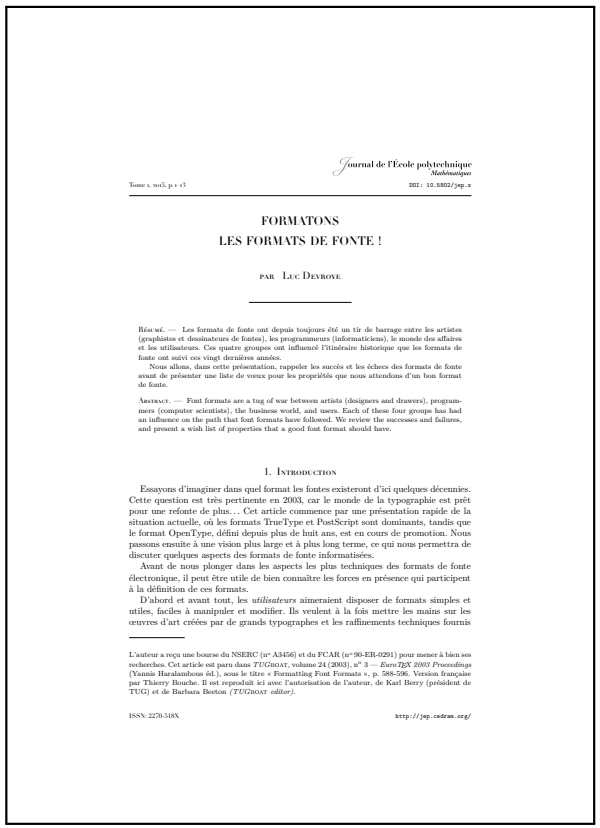
1. Introduction

Essayons d'imaginer dans quel format les fontes existeront d'ici quelques décennies. Cette question est très pertinente en 2003, car le monde de la typographie est prêt pour une refonte de plus… Cet article commence par une présentation rapide de la situation actuelle, où les formats TrueType et PostScript sont dominants, tandis que le format OpenType, défini depuis plus de huit ans, est en cours de promotion. Nous passons ensuite à une vision plus large et à plus long terme, ce qui nous permettra de discuter quelques aspects des formats de fonte informatisées.

Crédits : L'auteur a reçu une bourse du NSERC (n° A3456) et du FCAR (n° 90-ER-0291) pour mener à bien ses recherches. Cet article est paru dans *TUGboat*, volume 24 (2003), n° 3 — *EuroTEX 2003 Proceedings* (Yannis Haralambous éd.), sous le titre « Formatting Font Formats », p. 588-596. Version française par Thierry Bouche. Il est reproduit ici avec l'autorisation de l'auteur, de Karl Berry (président de TUG) et de Barbara Beeton *(TUGboat editor)*.

Diversité graphique

Même source, présentation variable (5).



Journal de l'École polytechnique

Journal de l'École polytechnique
Mathématiques

Tome 1, 2014, p. 1-13 DOI : 10.5802/jep.1

FORMATONS
LES FORMATS DE FONTE !

par Luc Devroye

Résumé. — Les formats de fonte ont depuis toujours été un tir de barrage entre les artistes (graphistes et dessinateurs de fontes), les programmeurs (informaticiens), le monde des affaires et les utilisateurs. Ces quatre groupes ont influencé l'itinéraire historique que les formats de fonte ont suivi ces vingt dernières années.

Nous allons, dans cette présentation, rappeler les succès et les échecs des formats de fonte avant de présenter une liste de vœux pour les propriétés que nous attendons d'un bon format de fonte.

Abstract. — Font formats are a tug of war between artists (designers and drawers), programmers (computer scientists), the business world, and users. Each of these four groups has had an influence on the path that font formats have followed. We review the successes and failures, and present a wish list of properties that a good font format should have.

1. Introduction

Essayons d'imaginer dans quel format les fontes existeront d'ici quelques décennies. Cette question est très pertinente en 2003, car le monde de la typographie est prêt pour une refonte de plus… Cet article commence par une présentation rapide de la situation actuelle, où les formats TrueType et PostScript sont dominants, tandis que le format OpenType, défini depuis plus de huit ans, est en cours de promotion. Nous passons ensuite à une vision plus large et à plus long terme, ce qui nous permettra de discuter quelques aspects des formats de fonte informatisées.

Avant de nous plonger dans les aspects les plus techniques des formats de fonte électronique, il peut être utile de bien connaître les forces en présence qui participent à la définition de ces formats.

D'abord et avant tout, les *utilisateurs* aimeraient disposer de formats simples et utiles, faciles à manipuler et modifier. Ils veulent à la fois mettre les mains sur les œuvres d'art créées par de grands typographes et les raffinements techniques fournis

L'auteur a reçu une bourse du NSERC (nº A3456) et du FCAR (nº 90-ER-0291) pour mener à bien ses recherches. Cet article est paru dans *TUGboat*, volume 24 (2003), nº 3 — *EuroTEX 2003 Proceedings* (Yannis Haralambous éd.), sous le titre « Formatting Font Formats », p. 588-596. Version française par Thierry Bouche. Il est reproduit ici avec l'autorisation de l'auteur, de Karl Berry (président de TUG) et de Barbara Beeton *(TUGboat editor)*.

ISSN: 2270-518X http://jep.cedram.org/

Cahiers GUTenberg

FORMATONS
LES FORMATS DE FONTE !

Luc Devroye

Résumé. Les formats de fonte ont depuis toujours été un tir de barrage entre les artistes (graphistes et dessinateurs de fontes), les programmeurs (informaticiens), le monde des affaires et les utilisateurs. Ces quatre groupes ont influencé l'itinéraire historique que les formats de fonte ont suivi ces vingt dernières années.

Nous allons, dans cette présentation, rappeler les succès et les échecs des formats de fonte avant de présenter une liste de vœux pour les propriétés que nous attendons d'un bon format de fonte.

Abstract. Font formats are a tug of war between artists (designers and drawers), programmers (computer scientists), the business world, and users. Each of these four groups has had an influence on the path that font formats have followed. We review the successes and failures, and present a wish list of properties that a good font format should have.

Note. L'auteur a reçu une bourse du NSERC (nº A3456) et du FCAR (nº 90-ER-0291) pour mener à bien ses recherches. Cet article est paru dans *TUGboat*, volume 24 (2003), nº 3 — *EuroTEX 2003 Proceedings* (Yannis Haralambous éd.), sous le titre « Formatting Font Formats », p. 588-596. Version française par Thierry Bouche. Il est reproduit ici avec l'autorisation de l'auteur, de Karl Berry (président de TUG) et de Barbara Beeton *(TUGboat editor)*.

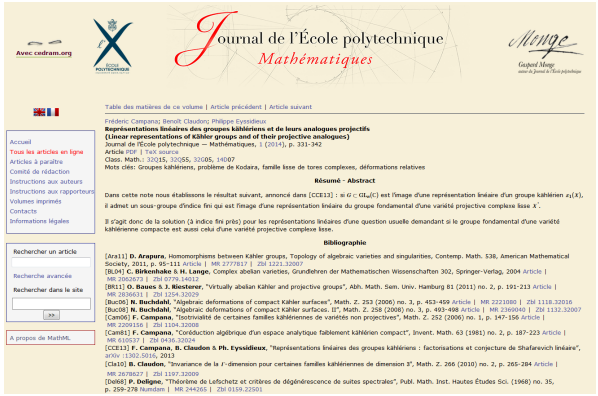
1. Introduction

Essayons d'imaginer dans quel format les fontes existeront d'ici quelques décennies. Cette question est très pertinente en 2003, car le monde de la typographie est prêt pour une refonte de plus… Cet article commence par une présentation rapide de la situation actuelle, où les formats TrueType et PostScript sont dominants, tandis que le format OpenType, défini depuis plus de huit ans, est en cours de promotion.

Cahiers GUTenberg nº 50 — Octobre 2008, p. 131-148.

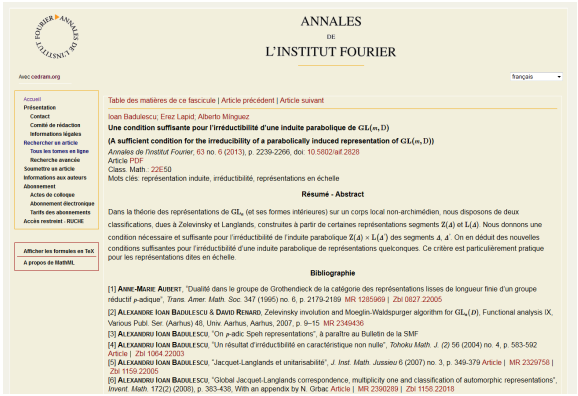
Diversité graphique

Mêmes formats, présentation Web variable.



Journal de l'École polytechnique

Confluentes Mathematici



Annales de l'institut Fourier

Annales Fac. Sci. Toulouse

Les enrichissements

- Les métadonnées comportent les biblios
- TEX↦ XML/MathML (Tralics)
- Outils de *matching* (articles et biblios) : MathSciNet, Crossref (services) ; Jahrbuch, zbMATH, mini-DML, EuDML (outils maison)
- Export NLM + plein texte vers EuDML

La vision EuDML

EuDML
The EUROPEAN DIGITAL MATHEMATICS LIBRARY

La bibliothèque numérique de mathématiques devrait s'efforcer de réunir un corpus mathématique aussi vaste que possible pour

- le rendre disponible en ligne
- en accès libre à terme,
- sous la forme d'un collection de référence,
- alimentée en continu par les nouveautés des éditeurs,
- valorisée par des outils de recherche et référencement sophistiqués,
- développée et entretenue par un réseau d'institutions

\+ et le préserver à très long terme

EuDML

Enrichissements après import

- Intégration de métadonnées (notamment en anglais) en provenance du zbMath
- Conversion des formules en MathML (OCR Infty, Maxtract → recherche de formules)
- Calcul de similitude des articles utilisant le texte intégral, les formules, les métadonnées
- Expérimental : textes accessibles

GDML

Projet en cours de définition

- Corpus mondial interopérable
- Infrastructure basse rendant possible la construction de services « sémantiques » et « computationnels »
- Format de description des structures mathématiques (objets, énoncés, preuves, définitions. . .)
- Outils de traitement automatisé de cette langue surnaturelle

$\Longrightarrow$ Activités à Washington et Séoul organisées par le CEIC

$\Longrightarrow$ Rapport du Conseil national de la recherche (NRC, Nat. Sc. Acad. USA) *Developing a 21st Century Global Library for Mathematics Research*, The National Academies Press, 2014

$\Longrightarrow$ Groupe de travail de l'union mathématique internationale

Merci !

Thierry Bouche
Directeur cellule Mathdoc (UMS 5638 UJF/CNRS)
Président EuDML initiative
Membre CEIC (IMU)
Membre GDML-WG (IMU)