

CEDRICS: When CEDRAM meets tralics

Thierry Bouche

Cellule MathDoc & institut Fourier, Grenoble

DML 2008

Towards Digital Mathematics Library

Birmingham (UK), 27 July 2008

Outline

- 1 The project CEDRAM
- 2 The CEDRAM production environment
- 3 Layout versatility
- 4 Automation
- 5 Tralics
- 6 Report on CEDRICS

The project CEDRAM

Overview

- A project for strengthening academic math research journals
- Principally aimed at supporting journals from independent or society publishers
- Although centred on those supported by CNRS, it is open to European partners
- A partnership between
 - CNRS, Ministry of research
 - The mathematical community and learned societies (SMF, SMAI, C. Sabbah)
 - Cellule MathDoc
 - Independent high quality research journals (AFST, AIF, AMBP, JTNB, MSIA. . .)

The project CEDRAM

Policy

- Published articles are original research papers scientifically checked under the responsibility of an editorial committee
- Once published, the mathematical content will remain unchanged for ever
- The navigation of the collections is free, the metadata is shared openly
- The access to the published articles is subject to a reasonable moving wall (MW)
- The articles are (physically) transferred to NUMDAM website (for long term access)
- The article's sources are archived (for long term preservation)

The project CEDRAM

Available services

The support is provided through

- The common portal (<http://www.cedram.org/>)
- An electronic publication platform instantiated for each serial (<http://<journal>.cedram.org/>)
- A set of modular productivity tools
- An export to NUMDAM
- A workflow for archiving production source files for long term preservation of the mathematical content

The project CEDRAM

Modules developed

Three main software modules have been developed or adapted to the project:

- 1 RUCHE, a software for managing the editorial process
- 2 EDBM, a database manager also generating the user interface
- 3 CEDRICS, a \LaTeX -driven journal issue production environment

The project CEDRAM

Workflow with journals

The main activity of MathDoc in the CEDRAM project is very similar to that of a print shop
(printing on *a variety of electronic supports*):

- 1 An accepted paper is prepared (structure, typography and metadata tagging) by the journal staff or subcontractors
- 2 When an issue is ready, the journal transfers all the required elements to the CEDRAM
- 3 CEDRAM returns “electronic proofs” for checking
- 4 When validated, they are exploited at the journal website and printer

The CEDRAM production environment Challenges

The main challenges were

- To keep production *manageable* by small journals
- While providing *good quality* output
- With *highly customisable* design fitting each journal's image and readership
- And *high-end features* to the users

The CEDRAM production environment

Internal workflow

Production unit: “physical volume”
(journal issue, seminar annual volume, single article)

- Starting from a \LaTeX source tree with cedram class
- Outputting all PDFs needed for print and screen
- Outputting metadata snippets in \LaTeX
- Transforming the \LaTeX metadata into XML in a dual format
- Ingesting the new volume's XML in EDBM (proof server)
- Technical and scientific validation
- Migrating the new volume to the exploitation server

The CEDRAM production environment

Principles

- 1 Any metadata is input at most once in the system, in the relevant file
- 2 Anything that is not determined by the relevant file's content should stay away from that file
- 3 Anything that can be computed *should* be computed!
- 4 Do not reinvent the wheel, do not invent exotic formats that no one will master
- 5 Stay pragmatic but avoid bottle-necks that would impact versatility of future use or quality of the output

The CEDRAM production environment Concept

A **journal**

is a set of physical **volumes**,

made of **articles**,

plus various other material, mostly constant.

The CEDRAM production environment

Core system

The production environment attaches to each level a set of files.

`cedram.cls` Variant of `amsart.cls` (with add-ons from `smfart`, `smfthm`, `smfbib`): the core of the system

Defines the input format, provides the general architecture for storing metadata, and many hooks

Operates at two different levels:

- Volume: builds a whole issue
- Article: for each article
- Special: for non-article material like TOCs

The CEDRAM production environment

Journal

A **journal** is defined by a main file (loaded from `cedram.cls`) and a set of inputs:

`cedram-JOUR.clo` Journal-wide constants (title, ISSN, layout)

`JOUR-front.tex` Front matter inputs (title page, copyright page addresses, scientific committee. . .)

`JOUR-Sommaire.tex` TOC template: a special loaded by the previous file

`JOUR-back.tex` Back matter inputs (subscription info. . .)

`JOUR-couverture.tex` The cover template of an issue

The CEDRAM production environment

Volume

A **journal issue** or **physical volume** is defined by a \LaTeX file with class `cedram` and options `JOUR`, `Volume`.

`JOUR_2008__46-47_.tex` The volume file defines:

- Numbers (year, month, volume, issue)
- Global variables (like special issue title, specific macros)
- First article's page number
- The ordered list of articles
- Extras (editorial statements, ads. . .)

The CEDRAM production environment

Article

An **article** is a directory containing all necessary data (images, inputs, bibliography) called from one master file.

The article master file declares all article-level metadata.

Volume info and page numbers are *not* declared: they will be inherited from the issue where the article ultimately appears.

`Article/Article.tex` Article master file (language, title, author, abstracts, text, biblio, etc.)

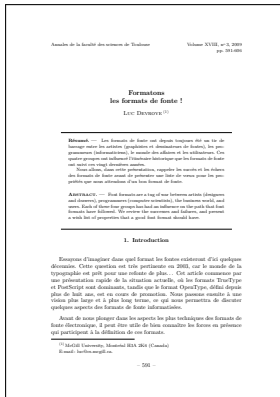
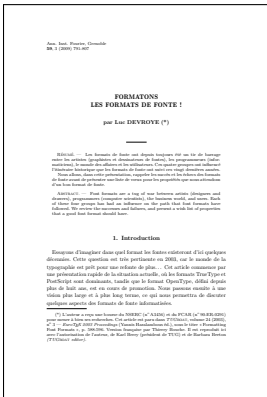
`Article/input.tex` T_EX inputs

`Article/biblio.bib` BibT_EX file

`Article/img1.pdf` Illustration

Layout versatility

Same input, varying presentation (1).



Annales de l'institut Fourier

Annales de la faculté des sciences
de Toulouse (mathématiques)

Layout versatility

Same input, varying presentation (2).

Journal de Théorie des Nombres
de Bordeaux 40 (2010), 617-631

**Formations
les formats de fonte !**

LUC DEVROYE

Résumé

Les formats de fonte ont depuis toujours été un tic de langage entre les artistes graphiques et designers de layout, les programmeurs (informaticiens), le monde des affaires et les utilisateurs. Ces quatre groupes ont influencé l'évolution historique que les formats de fonte ont eue ces vingt dernières années.

Nous allons, dans cette présentation, regarder les succès et les échecs des formats de fonte ainsi de présenter une liste de conseils pour ceux intéressés d'un bon format de fonte.

Abstract

Font formats are a tag of our business 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 existaient d'il y a quelques décennies. Cette question est très pertinente en 2003, car le monde de la typographie est prêt pour une refonte de plus... Cet article commente 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 pouvons conclure à une vision plus large et à plus long terme, ce qui nous permettra de discuter quelques aspects des formats de fonte informatiques.

1. Introduction

Essayons d'imaginer dans quel format les fontes existaient d'il y a quelques décennies. Cette question est très pertinente en 2003, car le monde de la typographie est prêt pour une refonte de plus... Cet article commente 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 pouvons conclure à une vision plus large et à plus long terme, ce qui nous permettra de discuter quelques aspects des formats de fonte informatiques.

Résumé Les formats de fonte ont depuis toujours été un tic de langage entre les artistes graphiques et designers de layout, les programmeurs (informaticiens), le monde des affaires et les utilisateurs. Ces quatre groupes ont influencé l'évolution historique que les formats de fonte ont eue ces vingt dernières années.

Nous allons, dans cette présentation, regarder les succès et les échecs des formats de fonte ainsi de présenter une liste de conseils pour ceux intéressés d'un bon format de fonte.

13

Journal de Théorie des Nombres
de Bordeaux 40 (2010), 617-631

**Formations
les formats de fonte !**

par LUC DEVROYE

Résumé Les formats de fonte ont depuis toujours été un tic de langage entre les artistes graphiques et designers de layout, les programmeurs (informaticiens), le monde des affaires et les utilisateurs. Ces quatre groupes ont influencé l'évolution historique que les formats de fonte ont eue ces vingt dernières années.

Nous allons, dans cette présentation, regarder les succès et les échecs des formats de fonte ainsi de présenter une liste de conseils pour les personnes qui sont intéressées d'un bon format de fonte.

Abstract Font formats are a tag of our business 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 existaient d'il y a quelques décennies. Cette question est très pertinente en 2003, car le monde de la typographie est prêt pour une refonte de plus... Cet article commente 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 pouvons conclure à une vision plus large et à plus long terme, ce qui nous permettra de discuter quelques aspects des formats de fonte informatiques.

Résumé Les formats de fonte ont depuis toujours été un tic de langage entre les artistes graphiques et designers de layout, les programmeurs (informaticiens), le monde des affaires et les utilisateurs. Ces quatre groupes ont influencé l'évolution historique que les formats de fonte ont eue ces vingt dernières années.

Nous allons, dans cette présentation, regarder les succès et les échecs des formats de fonte ainsi de présenter une liste de conseils pour ceux intéressés d'un bon format de fonte.

13

Annales mathématiques
Blaise-Pascal

Journal de théorie des nombres
de Bordeaux

Layout versatility

Same input, varying presentation (3).

MathematicS in Action
Vol. 3, 131 (2008)

**Formations
les formats de fonte !**
Luc Devignes *

*IAGCM University, Montreal RMA 003 (Canada)
E-mail: devignes@math.umontreal.ca

Résumé

Ce document de fonte est décrit en termes de types de caractères, de programmes (logiciels et bibliothèques), de formats, des programmes (algorithmes), de sources des lettres et de compilateurs. Ces quatre groupes ont attribué l'écriture latine que les lettres de fonte ont eues en usage moderne ancien.

Neuf lettres, dans cette présentation, suggère les caractères et les lettres des formats de fonte sans de présenter une liste de sources pour les paragraphes qui sont attribués à des formats de fonte.

Auteurs

Ces lettres ont été rédigées par un groupe de chercheurs (logiciels et bibliothèques), programmes (logiciels et bibliothèques), lettres et lettres (algorithmes et bibliothèques), sources des lettres et compilateurs. Ce travail a été réalisé en collaboration avec les lettres de fonte sans de présenter une liste de sources pour les paragraphes qui sont attribués à des formats de fonte.

1. Introduction

Ensembles d'algorithmes dans quel format les lettres existaient. Ces quelques algorithmes. Cette question est liée principalement au 2008, car le format de la typographie est peut-être une référence de plus. Cet article concerne par une présentation rapide de la situation actuelle, où les formats TeX/LaTeX et PostScript sont dominants, tandis que le format OpenType, utilisé depuis plus de huit ans, est en cours de progression. Nous pouvons envisager à une vision plus large et à plus long terme, ce qui nous permettra de discuter quelques aspects des formats de fonte informatiques.

Avant de nous charger dans les aspects les plus techniques des formats de fonte informatiques, il peut être utile de faire connaître les termes ou termes qui participent à la définition de ces termes.

Il était et est toujours, les algorithmes anciens (logiciels et bibliothèques), programmes (logiciels et bibliothèques), lettres et lettres (algorithmes et bibliothèques), sources des lettres et compilateurs. Ce travail a été réalisé en collaboration avec les lettres de fonte sans de présenter une liste de sources pour les paragraphes qui sont attribués à des formats de fonte.

Les lettres et programmes existant ont influencé considérablement les formats de fonte à leur présentation. Les présentations typographiques ont été à leur tour des lettres. Au début de ces dernières années, les présentations typographiques ont été à leur tour des lettres. Au début de ces dernières années, les présentations typographiques ont été à leur tour des lettres.

Cette notice est une notice de l'ANR (n° 031201) et de l'ANR (n° 031201) pour une notice à leur sujet. Cet article est une notice de l'ANR (n° 031201) et de l'ANR (n° 031201) pour une notice à leur sujet. Cet article est une notice de l'ANR (n° 031201) et de l'ANR (n° 031201) pour une notice à leur sujet. Cet article est une notice de l'ANR (n° 031201) et de l'ANR (n° 031201) pour une notice à leur sujet.

© 2008 by the author(s). This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

1

Document de théorie spectrale et géométrie
Dédicatoire
Volume 30 (2013) 131-174

**FORMATONS
LES FORMATS DE FONTE !**
Luc Devignes

Résumé

Ce document de fonte est décrit en termes de types de caractères, de programmes (logiciels et bibliothèques), de formats, des programmes (algorithmes), de sources des lettres et de compilateurs. Ces quatre groupes ont attribué l'écriture latine que les lettres de fonte ont eues en usage moderne ancien.

Neuf lettres, dans cette présentation, suggère les caractères et les lettres des formats de fonte sans de présenter une liste de sources pour les paragraphes qui sont attribués à des formats de fonte.

Auteurs

Ces lettres ont été rédigées par un groupe de chercheurs (logiciels et bibliothèques), programmes (logiciels et bibliothèques), lettres et lettres (algorithmes et bibliothèques), sources des lettres et compilateurs. Ce travail a été réalisé en collaboration avec les lettres de fonte sans de présenter une liste de sources pour les paragraphes qui sont attribués à des formats de fonte.

1. Introduction

Ensembles d'algorithmes dans quel format les lettres existaient. Ces quelques algorithmes. Cette question est liée principalement au 2008, car le format de la typographie est peut-être une référence de plus. Cet article concerne par une présentation rapide de la situation actuelle, où les formats TeX/LaTeX et PostScript sont dominants, tandis que le format OpenType, utilisé depuis plus de huit ans, est en cours de progression. Nous pouvons envisager à une vision plus large et à plus long terme, ce qui nous permettra de discuter quelques aspects des formats de fonte informatiques.

Avant de nous charger dans les aspects les plus techniques des formats de fonte informatiques, il peut être utile de faire connaître les termes ou termes qui participent à la définition de ces termes.

Il était et est toujours, les algorithmes anciens (logiciels et bibliothèques), programmes (logiciels et bibliothèques), lettres et lettres (algorithmes et bibliothèques), sources des lettres et compilateurs. Ce travail a été réalisé en collaboration avec les lettres de fonte sans de présenter une liste de sources pour les paragraphes qui sont attribués à des formats de fonte.

Les lettres et programmes existant ont influencé considérablement les formats de fonte à leur présentation. Les présentations typographiques ont été à leur tour des lettres. Au début de ces dernières années, les présentations typographiques ont été à leur tour des lettres. Au début de ces dernières années, les présentations typographiques ont été à leur tour des lettres.

Cette notice est une notice de l'ANR (n° 031201) et de l'ANR (n° 031201) pour une notice à leur sujet. Cet article est une notice de l'ANR (n° 031201) et de l'ANR (n° 031201) pour une notice à leur sujet. Cet article est une notice de l'ANR (n° 031201) et de l'ANR (n° 031201) pour une notice à leur sujet. Cet article est une notice de l'ANR (n° 031201) et de l'ANR (n° 031201) pour une notice à leur sujet.

© 2013 by the author(s). This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

MathematicS in Action

Séminaire de Théorie spectrale et géométrie

Layout versatility

Same input, varying presentation (4).



Séminaire Équations aux dérivées
partielles (Polytechnique)

Idem (recovery mode)

Layout versatility

Same input, varying presentation (5).

Journées Équations aux dérivées partielles
Éditions EMS, 7 pages (12 juin 2008)
ISBN 978-3-0391-0784-0

Formations
des formats de fonte !

Luc Duvivier

Résumé

Les formats de fonte ont depuis toujours été au lit de litage entre les artistes (graphistes et designers de fonte), les programmeurs (informaticiens), le monde des affaires et les éditeurs. Ce qui change progressivement est l'influence technologique que les formats de fonte ont eue sur ces trois domaines.

Vous allez, dans cette présentation, regarder les succès et les échecs des formats de fonte avant de présenter une liste de verrous que les programmeurs qui nous attendent d'un bon format de fonte.

Abstract

Font formats are a tag 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

Il existe un lien étroit entre les formats de fonte et les logiciels de traitement de texte. Cette relation est très particulière en 2008, car le monde de la typographie est peut-être un peu différent de ce qu'il était en 2003. Les artistes (concepteurs de fontes) et les programmeurs (informaticiens) ont joué un rôle important dans le développement de ces formats. Ce qui change progressivement est l'influence technologique que les formats de fonte ont eue sur ces trois domaines.

Vous allez, dans cette présentation, regarder les succès et les échecs des formats de fonte avant de présenter une liste de verrous que les programmeurs qui nous attendent d'un bon format de fonte.

Abstract

Font formats are a tag 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.

© 2008, Luc Duvivier. Tous droits réservés. Ce document est distribué sous licence Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International.

10 / 1

**5^e FORMATS
LES FORMATS DE FONTE !**

Luc DUVIVIER

Résumé

Les formats de fonte ont depuis toujours été au lit de litage entre les artistes (graphistes et designers de fonte), les programmeurs (informaticiens), le monde des affaires et les éditeurs. Ce qui change progressivement est l'influence technologique que les formats de fonte ont eue sur ces trois domaines.

Vous allez, dans cette présentation, regarder les succès et les échecs des formats de fonte avant de présenter une liste de verrous que les programmeurs qui nous attendent d'un bon format de fonte.

Abstract

Font formats are a tag 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

Il existe un lien étroit entre les formats de fonte et les logiciels de traitement de texte. Cette relation est très particulière en 2008, car le monde de la typographie est peut-être un peu différent de ce qu'il était en 2003. Les artistes (concepteurs de fontes) et les programmeurs (informaticiens) ont joué un rôle important dans le développement de ces formats. Ce qui change progressivement est l'influence technologique que les formats de fonte ont eue sur ces trois domaines.

Vous allez, dans cette présentation, regarder les succès et les échecs des formats de fonte avant de présenter une liste de verrous que les programmeurs qui nous attendent d'un bon format de fonte.

Abstract

Font formats are a tag 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.

© 2008, Luc Duvivier. Tous droits réservés. Ce document est distribué sous licence Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International.

Cahiers GUTENBERG n° 50 — Octobre 2008, p. 131-140

Journées Équations aux dérivées
partielles

Cahiers GUTENBERG

Automation devices

Articles

The core functionalities provided by `cedram.cls` in article mode are:

- A bunch of options for various outputs
- Automatic loading of an article config file at `\begin{document}`
- The lastpage trick
- Writing a \TeX line to an auxiliary file with all data pertaining to the article that could be used in a TOC
- Writing a \TeX structure with all metadata pertaining to the article to an auxiliary file

Compiling the article file produces the article PDF for the Web.

Automation devices

Volumes

The core functionalities provided by `cedram.cls` in volume mode are:

- A bunch of options for various outputs (Cover, offprints. . .)
- Writing a config file for each included article
- Compiling each included article or special in a subprocess (`write18`) and including it in the volume's PDF
- Storing the list of articles and assemble a volume TOC file from all article's generated lines
- Creating the volume's XML metadata from all article's generated snippets

Compiling the volume file produces a single PDF made of all inner pages of that volume.

Automation devices

Page numbers

A page number is essentially a by-product of a completed issue, except the first page of an issue, it is set nowhere.

The trick is as simple as:

- 1 Initialise the volume's page counter before loading the first article
- 2 Write the issue info and current volume's page number to the article's config file
- 3 Compile the article file twice so that all metadata, including issue info and page numbers are in final form
- 4 Include the pages of the just generated article PDF (omitting the Web first page) in the volume PDF
- 5 `\cleardoublepage`
- 6 Start again

Enters Tralics

Tralics is a software that

- Reads \LaTeX source code and outputs an XML file
- Was developed for producing the activity report of INRIA (more than 2,500 researchers and 140 project teams)
- Current version: 2.13
- Main developer: José Grimm

Tralics

Highlighted features

Tralics includes

- A full $\text{T}_{\text{E}}\text{X}$ macros interpreter
- A versatile translator from $\text{T}_{\text{E}}\text{X}$ -encoded text character to Unicode
- A translator of mathematical formulas into MathML
- A Bib $\text{T}_{\text{E}}\text{X}$ file parser
- A number of $\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$ standard commands
- A number of commands defined by popular $\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$ packages
- A configuration file to define or refine commands

Tralics

Highlighted new features

More recently, Tralics got

- A much more comprehensive and robust support of mathematical constructs including the AMS packages
- A new mechanism for controlling the way mathematical font changes are expressed in MathML
- A “nomathml” mode where \LaTeX math code is rendered as valid XML
- A mechanism for rewinding input and interpreting it twice
- A proper treatment of typographical quotation marks, at last!

Tralics and the DML

At MathDoc, Tralics was used to generate clean and accurate XML metadata from many sources:

- Converting journal's article metadata into something currently exploitable at cedram.org
 - From BibTeX hand-made catalogue (cf. NUMDAM version)
 - From NUMDAM's OCR'd metadata with hand-written TeX code
 - And all routine production with CEDRICS
- Converting journal's article metadata into something currently exploitable at NUMDAM (i.e. in pure HTML)
 - From native CEDRAM \LaTeX +BibTeX sources, converted to HTML
 - From native publisher's TeX sources (many dialects!), to HTML
 - Also creating metadata through hand-made \LaTeX sources, to HTML
- Tralics turned the new CEDRAM production system (aka CEDRICS) into an efficient XML metadata generator for mathematical content

CEDRAM production summary

	Retrodigitisation			Born-digital recovery			Future-ready born-digital		
	1997	2000	2001	2002	2003	2004	2005	2006	2007
AIF	NUMDAM 1,811 art. 51,034 p.		Mixed strategies 359 art. 11,056 p.				CEDRICS (+B) 218 art. 6,910 p.		
AFST	NUMDAM 1,109 art. 38,400 p.			NUMDAM again			CEDRICS (R, +B) 26 art. 836 p.	CEDRICS (R, -B) 29 art. 736 p.	
AMBP	NUMDAM 177 art. 1,632 p.				CEDRICS (R, +B) 45 art. 922 p.		CEDRICS (+B) 32 art. 870 p.		
JTNB	NUMDAM 431 art. 7,700 p.				CEDRICS (R, -B) 98 art. 1,818 p.		CEDRICS (-B) 82 art. 1,550 p.		
JEDP	NUMDAM 469 art. 5,580 p.		Mixed strategies 45 art. 400 p.		CEDRICS (R, -B) 13 art. 204 p.	CEDRICS (-B) 26 art. 510 p.			
TSG	NUMDAM 266 art. 3,800 p.					CEDRICS (R, -B) 13 art. 260 p.		CEDRICS (-B) 15 art. 240 p.	
SEDP	NUMDAM 640 art. 9,400 p.		CEDRICS (R, -B) 251 art. 3,650 p.						
Total	NUMDAM metadata 4,903 art. 117,546 p.				CEDRICS metadata 848 art. 18,506 p.				