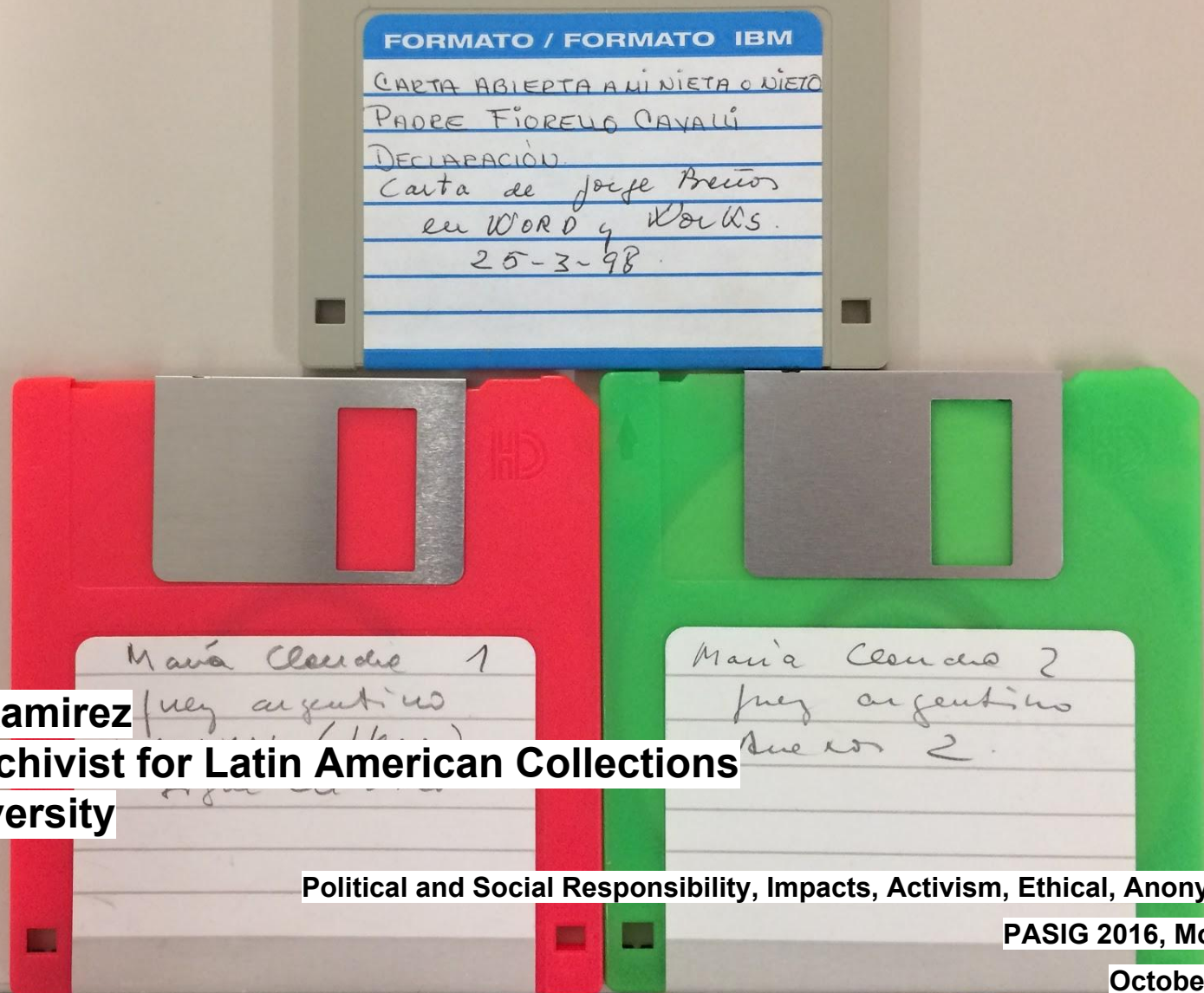


# Invisible Defaults and Perceived Limitations: Processing the Juan Gelman Files



**Elvia Arroyo-Ramirez**  
**Processing Archivist for Latin American Collections**  
**Princeton University**  
**@elvia\_arroyo**  
**#pasignyc**

Political and Social Responsibility, Impacts, Activism, Ethical, Anonymity, etc.

PASIG 2016, MoMA, NYC

October 28, 2016

**Chris Bourg's OLITA Spotlight Talk,  
Ontario Library Association Super  
Conference, 2015.**



*Used with permission; photo credit: Wayne Vanderkuil.*

From “Never Neutral: Libraries, Technology, and Inclusion,” Feral Librarian (blog), January 28, 2015.  
<https://chrisbourg.wordpress.com/2015/01/28/never-neutral-libraries-technology-and-inclusion/>

**“despite the democratizing promise of technology... the digital tools we build and provide are likely to reflect and perpetuate stereotypes, biases, and inequalities.”**

**“without active intervention we end up... classifying and arranging our content in ways that further marginalizes works by and about people of color, queer people, indigenous peoples, and others who don't fit neatly into a classification system that sets the default as western, white, straight, and male.”**



# Juan Gelman Papers



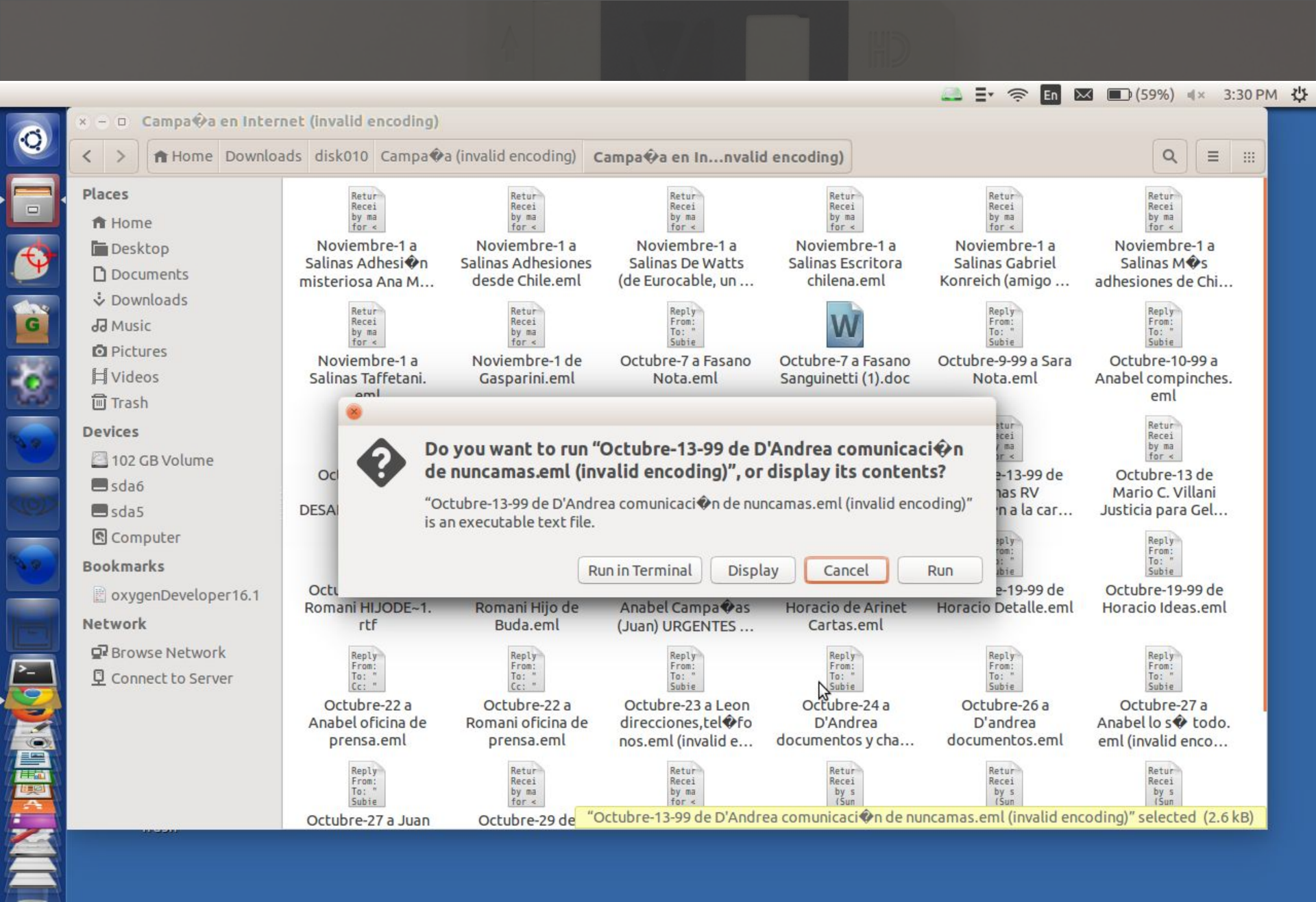
Juan Gelman Portraits; circa 1992-2006; Juan Gelman Papers, Department of Rare Books and Special Collections, Princeton University Library.

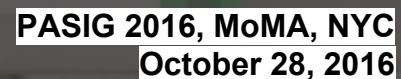
PASIG 2016, MoMA, NYC  
October 28, 2016



**Marcelo Ariel Gelman and María Claudia García Iruretagoyena; circa 1976; Juan Gelman Papers, Department of Rare Books and Special Collections, Princeton University Library.**







“the work-around that comes to mind would be to run a script on the files that replaces each non-ASCII character with an "unaccented" ASCII version. This is, I admit, not at all ideal from a preservation point of view, but you may need to **compromise on ideals for expediency.**” (Digital Curation list)

“If you are copying those files to a Linux box, then you may use **detox** (<http://detox.sourceforge.net/>)...Detox can **'clean' filenames**, removing diacritics and any **problematic character**, even recursing into directories.” (Digital Curation list)

“[This script] essentially **removes** the characters, replaces them with an underscore and logs the change.” (BitCurator list)

“Archivematica runs a filename **sanitization** script that generates a log of filename changes.” (Digital Curation list)

“To address issues with **illegal chars** in folders / file names, the following illegal chars (in WIN and \*NIX current environments) get scrubbed, and replaced with an \_.”

List of illegal chars:-

? < > : \* | " ^ (a GitHub for a tool that was mentioned in a list in response to our inquiry)



**Why did folks perceive that the diacritics of a file or folder name have to be “scrubbed” or “cleaned” in order to be “validated” by the tools and systems that we use to process born-digital archives?**




*Canta de Jorge Brea*


**Why do we have names like “detox” for tools to remove, essentially, markers of a language’s non-Englishness before we consider them “sanitized” enough for our preservation environments?**

**Why were the Spanish-language diacritic glyphs confounded with “illegal characters” in file and folder titles such as ? < > : \* | " ^?**



# File name has (invalid encoding) and CRLF issues

Name	Size	Type	Date Modified
 jag f  rst  r inte.txt (invalid encoding)	13 bytes	plain text document	Today


On Linux you sometimes you get a “” in a file name and a trailing “(invalid encoding)” in the filename. This is something that can happen when moving files from Windows to Ubuntu Linux. When uploading files to a Linux box you basically need two Linux tools to “repair” any incompatibility: “convmv” and “dos2unix”. The following commands will install them (on a Debian based Linux):


```
1 | sudo apt-get install convmv
2 | sudo apt-get install dos2unix
```

## Character encoding

To remove the “(invalid encoding)” you use the “convmv” tool. It is a tool that will convert the character encoding used in the file name. You can try the conversion of file names from different character set to UTF-8 using the following commands:

```
1 | convmv -r -f windows-1252 -t UTF-8 .
2 | convmv -r -f ISO-8859-1 -t UTF-8 .
3 | convmv -r -f cp-850 -t UTF-8 .
```

```
mv "/Desktop/cdo/disk062/Gonzalo
Dossier/Datos sobre Ricardo
Medina extrados de distintas
fuentes.doc"
```

```
"/Desktop/cdo/disk062/Gonzalo
Dossier/Datos sobre Ricardo
Medina extraidos de distintas
fuentes.doc"
```

<https://www.leaseweb.com/labs/2013/12/file-name-%EF%BF%BD-invalid-encoding/>



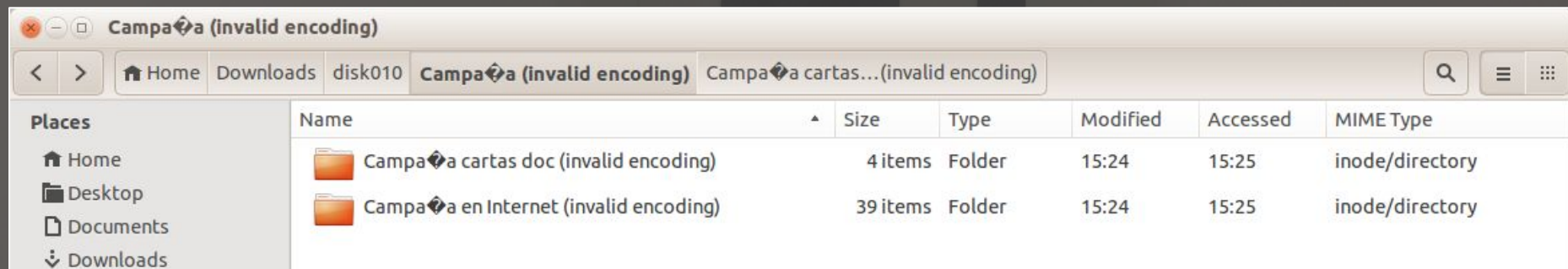
# Reflections

- What does removing or scrubbing the accent marks of a file or folder name do to the very identity of it?
- How can it potentially change the meaning behind the original intent of the creator?
- How can it potentially change the way researchers consume information?

Maria Claudia 1  
fuey argentina  
y amoros (1kws)  
figue en otros

Maria Claudia 2  
fuey argentina  
Amoros 2.





NEW

My Drive > Archival Description Group > Description and Access for Born-Digital Archival Collections (DABDAC) Workgroup > Reflections

- My Drive
- Shared with me
- Recent
- Google Photos
- Starred
- Trash

255 MB used

Name	Owner	Last modified	File size
invalid-encoding-issues-and-commands	me	Oct 17, 2016 me	—
using bagit-python in the command line	Kelly Bolding	Oct 6, 2016 Kelly Bolding	—
add-missing-file-extensions-script	Kelly Bolding	Oct 6, 2016 Kelly Bolding	—

María Cleocho 1  
fuey argentino  
y anexo 1 (1kws)  
figue en otros

María Cleocho 2  
fuey argentino  
Anexo 2.

# Conclusions

Which exceptions are we willing to make or which processes are we willing to work *through* (as opposed to work *around*), based on the perceived value or implications of the issue at hand?

Where do cultural and political implications fall on the spectrum of other pressures that archivists deal with in their everyday practices?

Maria Claudia 1  
fuey argentinos  
y mexicanos (1kms)  
figue en otros

Maria Claudia 2  
fuey argentinos  
mexicanos 2.